Ahmet MENTEŞ

M.S. Thesis In Physics

December - 2008

STRUCTURAL AND ENERGETIC DETAILS OF THE MECHANISM OF ATPase DOMAIN OF A TYPE IIA TOPOISOMERASE

by

Ahmet MENTEŞ

December 2008

STRUCTURAL AND ENERGETIC DETAILS OF THE MECHANISM OF ATPase DOMAIN OF A TYPE IIA TOPOISOMERASE

by

Ahmet MENTEŞ

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

December 2008
Istanbul, Turkey

APPROVAL PAGE

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

Prof. Dr. Mustafa KUMRU
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.

Assist. Prof. Dr. Levent SARI
Supervisor

Examining Committee Members

Prof. Dr. Mustafa KUMRU ……..

Assist. Prof. Dr. Levent SARI ……..

Assist. Prof. Dr. Sevim ISIK ……..

Assist. Prof. Dr. Sadık GUNER ……..

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.

Assist. Prof. Dr. Nurullah ARSLAN

Director

December 2008

STRUCTURAL AND ENERGETIC DETAILS OF THE MECHANISM OF ATPase DOMAIN OF A TYPE IIA TOPOISOMERASE

Ahmet MENTEŞ

M. S. Thesis - Physics
December 2008

Supervisor: Assist. Prof. Dr. Levent SARI

ABSTRACT

The mechanism of an ATPase domain of the type IIA topoisomerase has been investigated by using a molecular dynamic method based on the force field potential. A theoretical model, Half Quadratic Biased Molecular Dynamics (HQBMD), is used to produce large scale conformational changes in the enzyme. As the passage of a DNA segment, known as the T-Segment, through the enzyme is proposed in the literature, we have analyzed how this can be possible in real time. To do this, different reaction coordinates at the lower and upper sides of the protein are chosen, and opening of the two monomers in the non-covalent complex obtained in the crystal structure is forced. First time, we have reported some correlated motions of α9 and α10 with α13. Also we have observed that the protein rotates by about 40 degrees as it opens up at the lower part. This observation supports the mechanism proposed in the literature, in which the T-Segment DNA needs to rotate after it is captured by the protein. Our structural and energetic analyses show that the most feasible opening should be the one along the gate8 at the lower part. Also, we have found out that the upper gate openings encounter with a significant potential energy barrier at the beginning of the reaction coordinate. The most striking information we got is that protein rotates in the counter-clockwise (that is right-handed)

direction in all simulation set-ups, which is needed to bring negative supercoils into the DNA. This observation nicely explains why DNAGyrase removes only positive supercoils as it is known for many years. Lastly, it is important to note that we do not get any opening at upper side of the protein as much as the diameter of the DNA (20 A) in any upper gate openings. This means that the protein needs to form the non-covalent dimer complex *after* it catches the T-segment DNA.

Keywords: E.coli GyraseB, TopoII, DNA topology, supercoiling, Molecular Dynamics Simulations, HQBMD

TİP IIA TOPOİSOMEREZLERİN ATPASE DOMAİNLERİNİN ENERJETİK VE YAPISAL YÖNDEN MEKANİZMASININ İNCELENMESİ

Ahmet MENTEŞ

Yüksek Lisans Tezi - Fizik
Aralık 2008

Tez Yöneticisi: Yrd.Doç. Dr. Levent SARI

ÖZ

Bu çalışmada, kuvvet alan potensiyeline dayanan moleküler dinamik method kullanılarak, tip IIA topoisomerezlerin ATPase kısmının mekanizması incelendi. Proteindeki yapısal değişimleri büyük ölçekli elde edebilmek için yarı quadratik eğilimli moleküler dinamik method adı verilen teorik model ile çalışıldı. Litaratürde varsayılan modelde T-Segment olark bilinen bir DNA kesiti, G-Segment olarak bilinen başka bir DNA nın içinden geçirilmektedir. Bu teorik çalışmada, literatürde varsayılan bu mekanizmayı modelledik. Bunun için, proteinin alt ve üst kısmında farklı reaksiyon koordinatları belirlendi ve non-kovalent kompleks oluşturan (zincirA ve zincirB) proteinin açılımı sağlandı. İlk olarak α9 ve α10 un bazı hareketlerinin α13 ile ilişkili olduğunu bildirdik. En önemli bulgularımızdan birisi ise, proteinin alt kısmından açılma esnasında yaklaşık olarak 40 derece dönmesidir. Bu gözlemler, litaratürde de belirtildiği gibi, DNA, protein tarafından yakalandıktan sonra, DNA'nın T-segmentinin dönmesi gerektiği mekanizmasını desteklemektedir. Ayrıca, bu dönmelerin saatin ters yönünde gerçekleştiğini gözlemledik. Bu gözlemde, DNAGyrase in neden DNA ya sadece negatif supercoil kattığını açıklamaktadır. Son olarak, hiç bir sumilasyonumuzda üst kısmın, bir DNA nın geçebileceği kadar açılmadığını bulduk. Buda, proteinin, DNA yı yakaladıktan

sonra dimer haline geçtiği göstermektedir. Bu bilgi de literatürde daha önce belirtilmiş, ve bizim gözlemlerimiz de literatür bilgileri ile uyuşmaktadır.

Anahtar Kelimeler: E.coli GyraseB, TopoII, DNA topolojisi, süperkoiling, Moleküler Dinamik Simülasyonlar, HQBMD.

ACKNOWLEDGEMENT

I would like to express my gratitude to my supervisor Assist. Prof. Dr. Levent SARI. I thank to him for his contribution, guidance, patience, experience and support throughout the research and writing of my thesis.

Also, my thanks go to Prof. Dr. Mustafa KUMRU for their valuable help.

I express my thanks and appreciation to my family and my friends especially Süleyman ÜÇÜNCÜOĞLU, Neslihan ÜÇÜNCÜOĞLU, Azize SEVİM, Doruk YILDIZTEKİN for their understanding, motivation and patience.

I should also acknowledge to TUBITAK (The Scientific & Technological Research Council of Turkey) for providing us with the funding through the project number 107T209. The project is titled as “Quantum, Classical (Molecular), and Statistical Mechanical Investigation of dynamic mechanisms of DNA-Topoisomerase systems in conjunction with the topoisomerase-targeted anti-cancer drug molecules”. Without this funding, we would not have the required hardware and software needed to carry out our research.

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
Lk	:	Linking number
Lk_0	:	Linking number in relaxed state
Tw	:	Twist
Wr	:	Writhe
ΔLk	:	Linking difference
σ	:	Supercoiling density or specific linking difference
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
b_0	:	The equilibrium bond distance
b	:	The current bond distance
θ_0	:	The equilibrium bond angle
θ	:	The current bond angle
u_0	:	Urey-Bradley 1.3 distance
q	:	Charge of atom

rmsd	:	Root mean square difference
Mg^{+2}	:	Magnesium ion
R	:	Coordinates of the final reference structure
N	:	Total number of atoms
U	:	Potential energy
E.coli	:	Escherichia coli
GyrA	:	E.coli DNA gyrase A
GyrB	:	E.coli DNA gyrase B
CPT	:	Constant Pressure/Temperature
NMR	:	Nuclear Magnetic Resonance
MD	:	Molecular Dynamics
HQBMD	:	High Quality Biased Molecular Dynamics
ATP	:	Adenine Tri Phosphate
ADPNP	:	Adenosine 5'- [ß, γ -imino] triphosphate
CTD	:	C Terminal Domain
NTD	:	N Terminal Domain
ICRF-187	:	Bisdioxopiperazine
CHARMM:		Chemistry at HARvard Macromolecular Mechanics
VMD	:	Visual Molecular Dynamics
CMU	:	Cluster Management Utility
LSF	:	Load Sharing Facility
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

CHAPTER 1

INTRODUCTION

1.1 DNA TOPOLOGY AND SUPERCOILING

Topology which can be defined as spatial properties of an object that remain unchanged under continuous deformations is important for both structure and function of DNA. Using the results of ribbon theory, Fuller applied topology for studying the properties of closed circular DNA [1]. According to the ribbon theory, other than topological characteristics of a ribbon, the linking number value (Lk), twist (Tw) and writhe (Wr) of the ribbon also play an important role. Lk is a topological invariant that doesn't depend on geometry whereas Tw and Wr and differential geometric properties. The mathematical relationship between these two quantities is defined as:

$$Lk = Tw + Wr \tag{1.1}$$

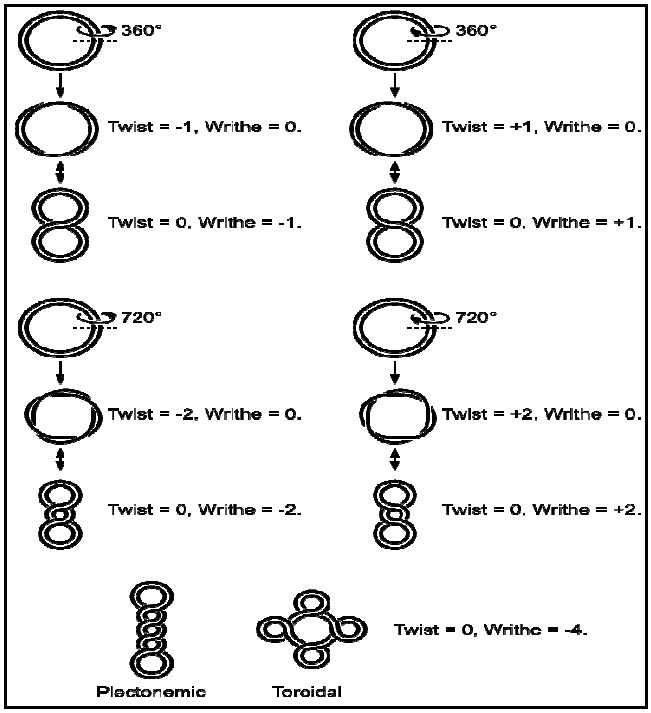
Notice that the topological property is equal to the summation of two geometric properties.

DNA is a molecule that contains the genetic information of an organism in all living cells and composed of two deoxyribonucleotide polymers wrapped around each other forming a helical structure. The two strands of the molecule are held together by weak, noncovalent interactions. This structure of DNA was first described in Watson and Crick's model where two strands of the DNA have a common central axis around which the helices are wound in a right-handed fashion.

The first studies in this field were focused on the properties of a linear DNA molecule, because, at that time it was the only type of molecule that could be extracted from cells and viruses. Scientists were later surprised to discover closed circular DNA (ccDNA) molecules in viral genomes and bacterial plasmids. Closed circular structure

refers to a DNA that doesn't have breaks in neither of two strands. Closed circular and linear DNA molecules have many distinct properties (Figure 1.1). Perhaps one of the most striking differences between these molecules is the fact that ccDNA can be knotted [1].

a)



b)

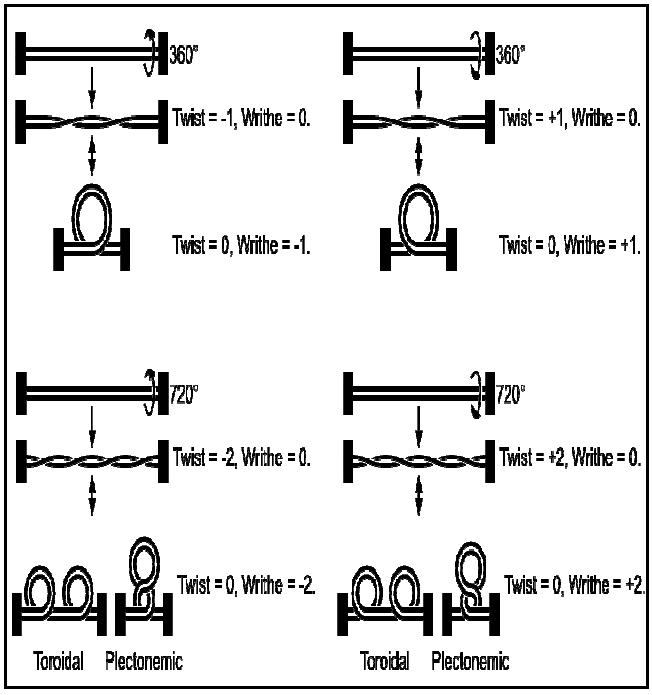


Figure 1.1 Closed circular (a) and linear (b) DNA molecules

The topology of DNA also helps us to understand the functions of DNA. ccDNA has two topological levels: the first is that DNA double helix can form different types of knots and the second is DNA supercoiling that results from interlinking of the complementary strands. It is important to understand the geometry and topology of knotting, linking and supercoiling as they play an important role in various cellular processes.

Supercoiling is crucial for compact packaging of DNA in living cells. An average human cell contains around 2 m of DNA and the size of a typical cell is 10 μm. As these numbers suggest, DNA molecule must be highly compacted in order to fit into the cell. Supercoiling provides a degree of compaction by reducing the space. At the same time, DNA must still remain available for important cellular processes such as replication and transcription even after compaction.

Supercoiling is referred as coiling of a coil. A phone cord is an example to a coiled wire and twisting of a phone cord is analogical to supercoiling (Fig. 1.2). DNA is a coiled double helix and folding of central helical axis upon itself is defined as DNA supercoiling. If no bending is present, DNA is considered to be in a relaxed state (the first picture in Fig 1.3)

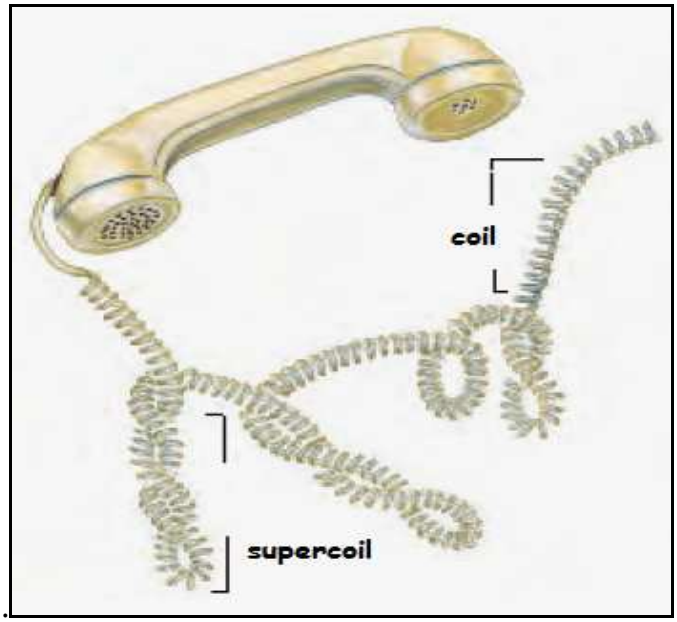


Figure 1.2 Supercoils [2].

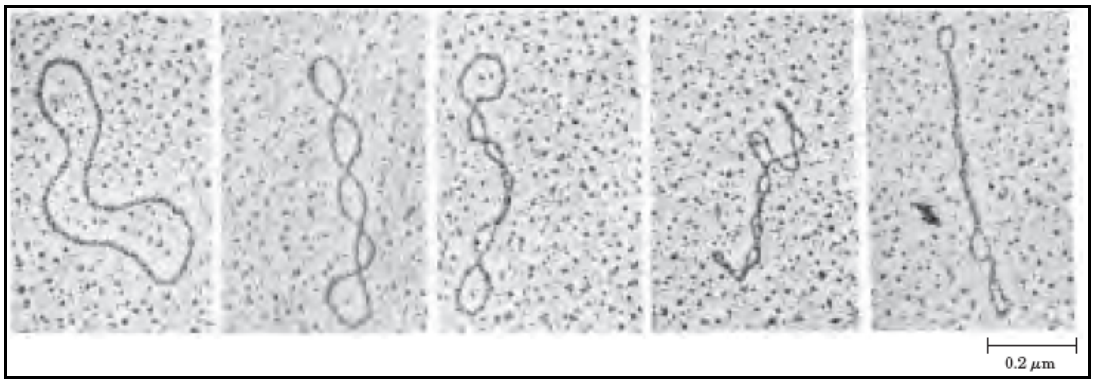


Figure 1.3 Electron micrographs of relaxed and supercoiled plasmids [2].

Topologically, a ccDNA can be defined by its linking number Lk, which is the number of times one DNA strand goes around the other stand, in the absence of supercoiling. Lk is considered as a strict topological property, because, it remains unchanged when DNA is twisted or deformed provided that both stands of the double helix remain intact. Lk is only defined for covalently closed DNA and remains constant if the molecule is covalently closed. The only way to change the Lk of a circular DNA is to break on of the complementary strands by disrupting a phosphodiester bond, pass the intact strand through the opening and then reseal the broken strand. By breaking one of the strands it is possible to untwist and separate the two strands completely. In such a case, Lk becomes undefined because of the absence of a topological bond (Fig. 1.4).

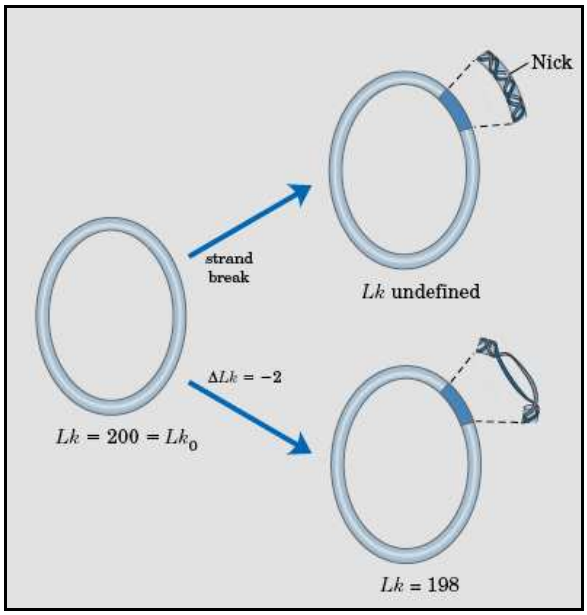


Figure 1.4 An example of linking number [2]

Linking number is defined as the sum of two structural components writhe and twist. Local inter-windings of the double strands of DNA result in nodes and crossings that are measured by the parameter named twist (Tw). Twist can be simply defined as the number of twists that double strands make around each other. Writhe (Wr) measures nodes that form coiling of helical axis. Writhe is described as the number of times the double helix coils around itself. Since these two parameters depend on the geometry of the molecule, they can be changed by deforming the DNA. Furthermore, Tw and Wr don't have to be integers and in most cases they aren't.

On the other hand, for a ccDNA Lk is always an integer and can have a positive or negative value depending on the orientation of the stands. If the strands are intertwined in a right-handed fashion, the linking number is defined as positive (+). A typical B-type DNA molecule has a (+) linking number. In a left-handed helix the linking number is negative (-). If the DNA axis contains no net bending lays flat on a plane, then this DNA is considered to be in a relaxed state (Fig. 1.3). The linking number of DNA in relaxed state is defined as Lk_0 and equals to the number of base pairs divided by 10.5. Lk_0 is used as a reference [2]. When there is a strain on DNA, Lk becomes less or more than Lk_0 and DNA undergoes a three-dimensional writhing in a process called supercoiling. Supercoiled DNA is torsionally stressed compared to the DNA in a relaxed state. The degree of supercoiling is defined by the linking difference and expressed as follows:

$$\Delta Lk = Lk - Lk_0 \tag{1.2}$$

Overwinding of DNA produces positive supercoiling and underwinding of DNA produces a negative supercoiling (Fig. 1.5). Circular DNA isolated from living cells is underwound and displays a negative supercoiling [3]. During the replication process, the unwinding of DNA strands produces a positive ΔLk or a superhelical strain. This causes an increase in ΔLk, because, Lk_0 is reduced by the separation of parental strands of the helix. A replicating DNA molecule can produce positive precatenanes, where daughter DNA molecules are interwined and have positive supercoils. Thus, during the replication process ΔLk increases by one for every 10 base pairs of replicated DNA [4]. A DNA isolated from cell is most often found in a supercoiled state.

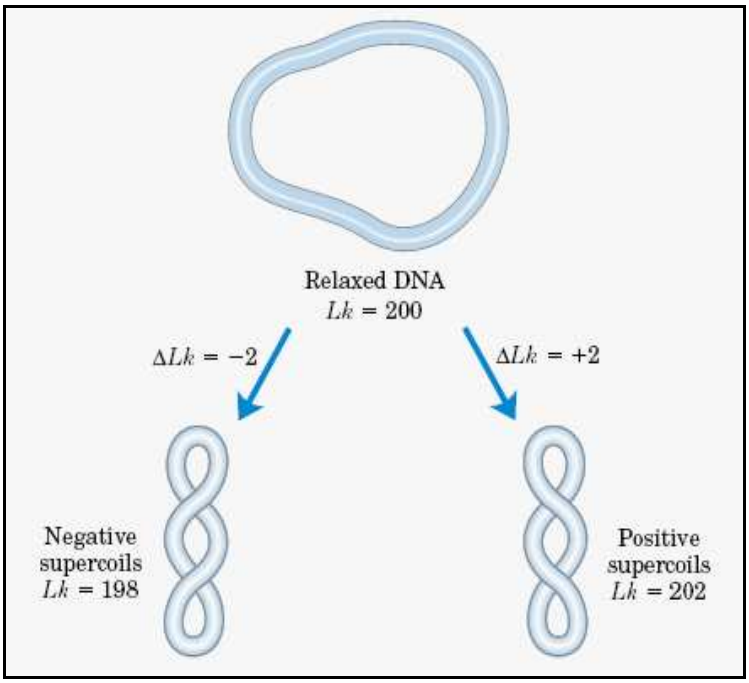


Figure 1.5 A schema about negative and positive supercoiling [2].

The topological deviation of DNA from its relaxed state can be best described by its supercoiling density σ. It is more advantageous to use the supercoiling density, because, it can be determined experimentally and doesn't depend on the length of the molecule. This term is also called a superhelical density or specific linking difference and can be described by the following formula:

$$\sigma = \frac{\Delta Lk}{Lk_0} \tag{1.3}$$

Since, DNA is a right-handed helix; it has a positive linking number. But, it is usually underwound in living cells. Therefore, Lk is smaller than Lk_0, σ is negative, and DNA is negatively supercoiled.

The only way to change the linking number is to break of the strands of DNA and then rejoin the ends. Topoisomerases are the enzymes that can increase and decrease the linking number of DNA. These enzymes are thus able to change the supercoiling degree of DNA by introducing single or double strand cuts.

1.2 TOPOISOMERASES AND THEIR CELLULAR IMPORTANCE

DNA topoisomerases are enzymes that can change the topological state of DNA without changing its primary structure by changing the linking number of the molecule. These enzymes regulate the number of topological links between two DNA strands by introducing transient single or double-strand breaks, passing the strands through each other and then resealing them rejoining them.

There exist three types of topological problems: knotting, catenation and supercoiling (Fig. 1.6). Topoisomerases are able to solve all the topological problems of DNA that arise during such processes as DNA replication, transcription, recombination or chromosome separation. These enzymes can remove supercoiling by catalyzing the swiveling and relaxation of the DNA. The function of topoisomerases is to remove the torsional stress in the DNA that forms during transcription and recombination. Additionally, topoisomerases can change the catenation and knotting by passing the DNA strands though the breaks they form.

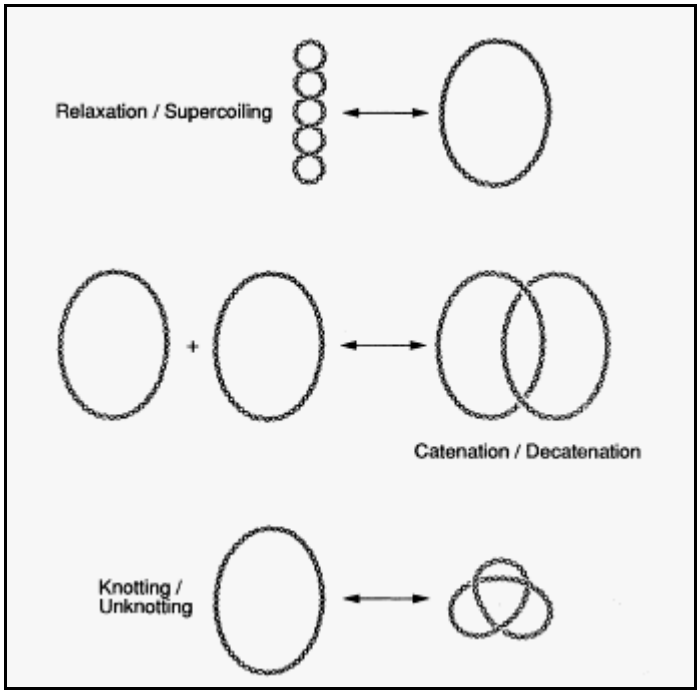


Figure 1.6 The reactions carried out by topoisomerases [4].

During the cleavage process, topoisomerases attack the phosphorus of DNA molecule with the oxygen of the tyrosyl residue forming a transient covalent phosphotyrosine bond between the cut end of the strand and the enzyme and break a phophodiester bond [5]. The topology of DNA is modified during the transient covalent intermediate period. Resealing of the broken ends is the reverse of the breaking process. The hydroxyl oxygen of DNA attacks the phosphorus of the transient phosphotyrosine link, rejoining the ends [6]. Since, the energy stored in the phosphotyrosine link is used for resealing of DNA, the cleaving and rejoining processes do not require an energy cofactor [7].

Topoisomerases can be classified into two groups according to the differences in their sequence, structure and functions: type I and type II. Type I topoisomerases can relieve the torsional stress in the DNA molecule and therefore play an important role in DNA replication and transcription processes. The function of topoisomerases I is to release the stress that forms during these two processes that involve unwinding of the DNA stands to give access to the information stored in the DNA. Type I topoisomerases are monometric meaning that they can only cut one strand of DNA and thus can change the linking number by the increments of one. On the other hand, type II topoisomerases are dimeric and can cut both strands of the DNA generating a passage through which another part of the DNA can be passed. Type II topoisomerases can change the linking number by the increments of two. These enzymes are very important during cell division.

Topoisomerases are further classified into four subfamilies: IA, IB, IIA and IIB. Members of each family share similar structures and mechanisms of action but, different subfamilies have several structural differences. Further classifications within the families are done according to structural differences. Subfamily members of some prokaryotic and eukaryotic topoisomerases are given in Table 1.

Table 1.1 Classification of topoisomerases [8]

Topoisomerase[a]	Subfamily type	Subunit structure	Size(s) (aa)[b]
Eubacterial DNA topoisomerase I (*E. coli*)	IA	Monomer	865
Eubacterial DNA topoisomerase III (*E. coli*)	IA	Monomer	653
Yeast DNA topoisomerase III (*S. cerevisiae*)	IA	Monomer	656
Mammalian DNA topoisomerase IIIα (human)	IA	Monomer	1001
Mammalian DNA topoisomerase IIIβ (human)	IA	Monomer	862
Eubacterial and archaeal reverse DNA gyrase (*Sulfolobus acidocaldarius*)	IA	Monomer	1247
Eubacterial reverse gyrase (*Methanopyrus kandleri*)[c]	IA	Heterodimer	A, 358 B, 1221
Eukaryotic DNA topoisomerase I (human)	IB	Monomer	765
Poxvirus DNA topoisomerase (vaccinia)	IB	Monomer	314
Hyperthermophilic eubacterial DNA topoisomerase V (*Methanopyrus kandleri*)[d]	IB	Monomer	—[e]
Eubacterial DNA gyrase (*E. coli*)	IIA	A_2B_2 hetero-tetramer	GyrA, 875 GyrB, 804
Eubacterial DNA topoisomerase IV (*E. coli*)	IIA	C_2E_2 hetero-tetramer	ParC, 752 ParE, 630
Yeast DNA topoisomerase II (*S. cerevisiae*)	IIA	Homodimer	1428
Mammalian DNA topoisomerase IIα (human)	IIA	Homodimer	1531
Mammalian DNA topoisomerase IIβ (human)	IIA	Homodimer	1626
Archaeal DNA topoisomerase VI (*Sulfolobus shibatae*)	IIB	A_2B_2 hetero-tetramer	A, 389 B, 530

[a]The sources of the best studied family members are given in the parentheses. The upper portion of the table shows the type I topoisomerases; the lower portion of the table shows the type II topoisomerases

[b]The subunit sizes of the best studied family members

[c]The only known reverse gyrase with a heterodimeric structure.

[d]The only known representative of its kind. May be present in other hyperthermophilic eubacteria.

[e]Gene not yet cloned. The purified protein has a molecular size of 110 kDa.

In our study, we tried to investigate the mechanism of gates at upper and lower sides of the type IIB topoisomerases energetically and structurally. In the figure 1.7, it is shown that type II topos attached covalently with both the 5’ and 3’ phosphates end of the cleaved strand of DNA.

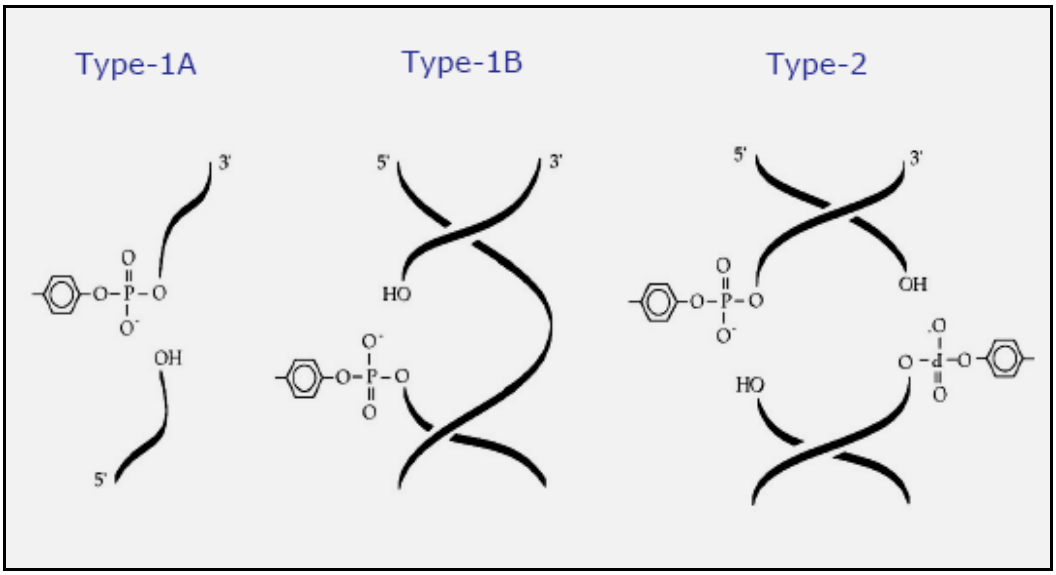


Figure 1.7 Types of topoisomerases

1.3 ATPase REGION OF TOPOISOMERASE II

Figure 1.8a illustrates the sequence alignment of the type IIA subfamily according to their homology to E.Coli DNA gyrase and shows their 3D structure. It is clear that the E.coli DNA gyrase A subunit (GyrA) is homologous to the PartC subunit and the C terminal half of eukaryotic enzymes, whereas E.coli DNA gyrase B subunit (GyrB) is homologous to the partE subunit of E.coli topoisomerase IV and N-terminal half of eukaryotic enzymes. With the exception of a 170 amino acid sequence in the C-terminal of GyrB, there is a strong sequence similarity between different type IIA enzymes. In contrast, the C-terminal domains show sequence similarities only for the first part of the tail and only among closely related species. Like the N-terminal of the type IB topoisomerases, the C-terminal tails of the type IIA topoisomerazes mainly play role in nuclear targeting and interaction with proteins [9]. The active side tyrosines are located approximately 120 amino acids away from the N-terminus of GyrA or from the B' – A' boundary, in case of single subunit enzymes (Fig. 1.8a). The ATP binding region is located within the first 400 amino acids of PartE, GyrB and the whole enzyme (Fig. 1.9). The region responsible for the interaction of two different subunits in the heterotetramers is located in the C-terminus of GyrB, next to the ATPase domain and facilitates the communication between the two halves in the single subunit enzymes [10]. A' refers to the region in the yeast enzyme that is homologous to the first 505 amino acid sequence in the GyrA and B'-A' region refers to the DNA binding and cleaving domain (Fig. 1.8a). A region located in the C-terminal tail domain of the DNA gyrase wraps approximately 140 base pairs of the G-segment DNA into a right-handed supercoil, creating the necessary substrate for negative supercoiling [11].

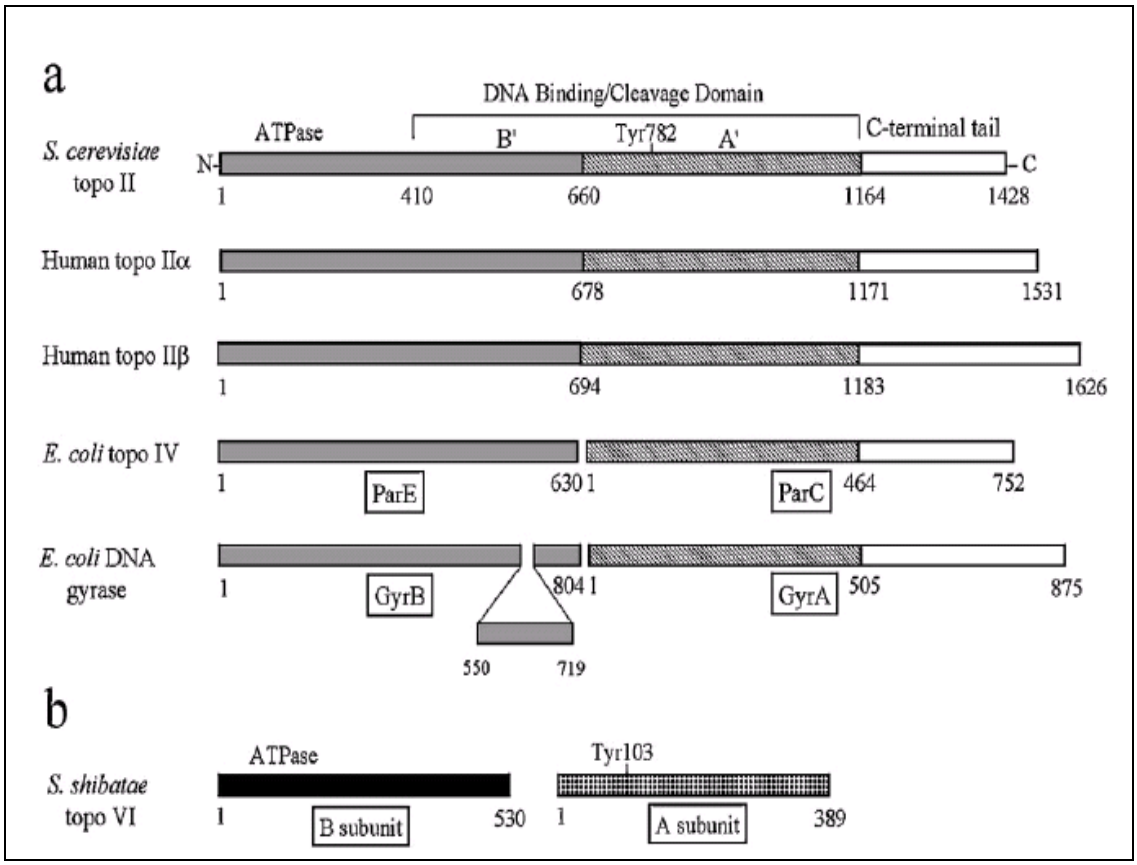


Figure 1.8 The sequence alignment of the type IIA subfamily [12]

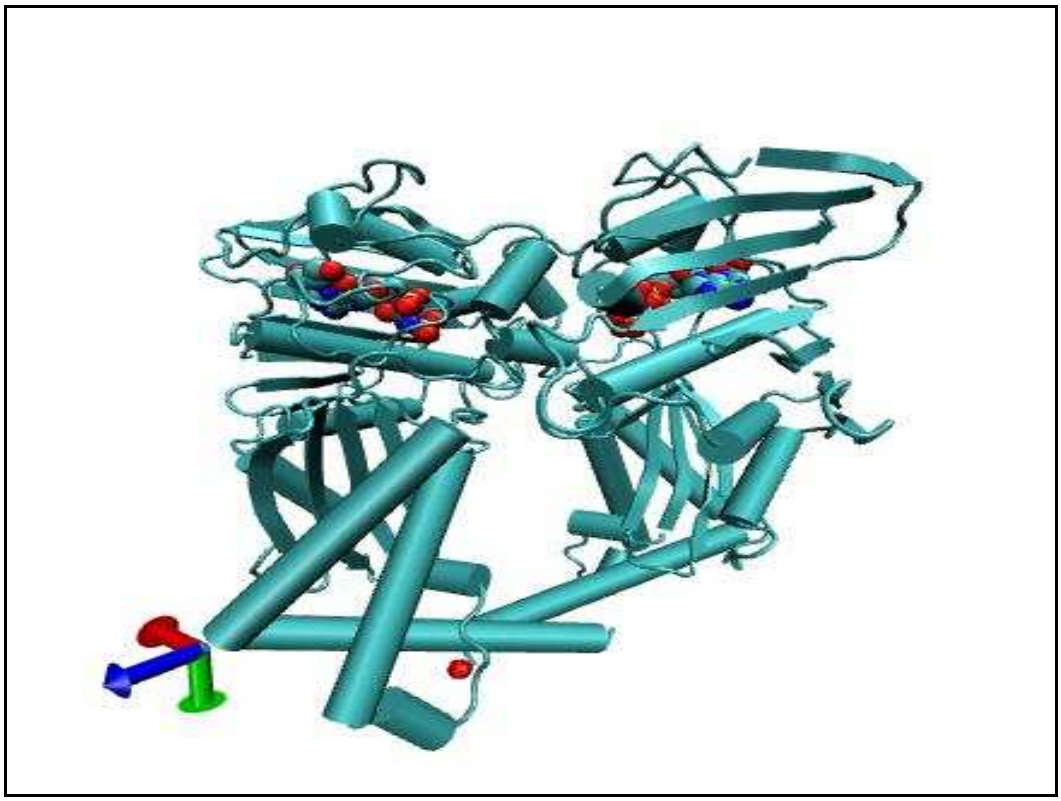


Figure 1.9 The ATPase domain of DNA GyraseB

Scientists have been able to obtain crystal structures for some fragments of subunits of E.coli DNA gyrase (Fig. 1.10). The final, complete structure of an intact type IIA topoisomarase is yet to be determined but, the currently available structures give a near-complete view about the assembly and function of this enzyme. Figure 1.9 illustrates the crystal structure of the N-terminal half-fragment of the GyrB subunit obtained in the presence of a non hydrolysable ATP analog ADPNP, to facilitate the formation of dimers [13]. In this structure, the N-terminal ATPase domain is located above the C-terminal DNA capture domain, that forms a cavity approximately 20 A^0 in diameter, just large enough for a DNA molecule to fit in it. The lower extensions of the DNA capture domains are attached to the second half of the GyrB enzyme not shown in the crystal structure (Fig. 1.10).

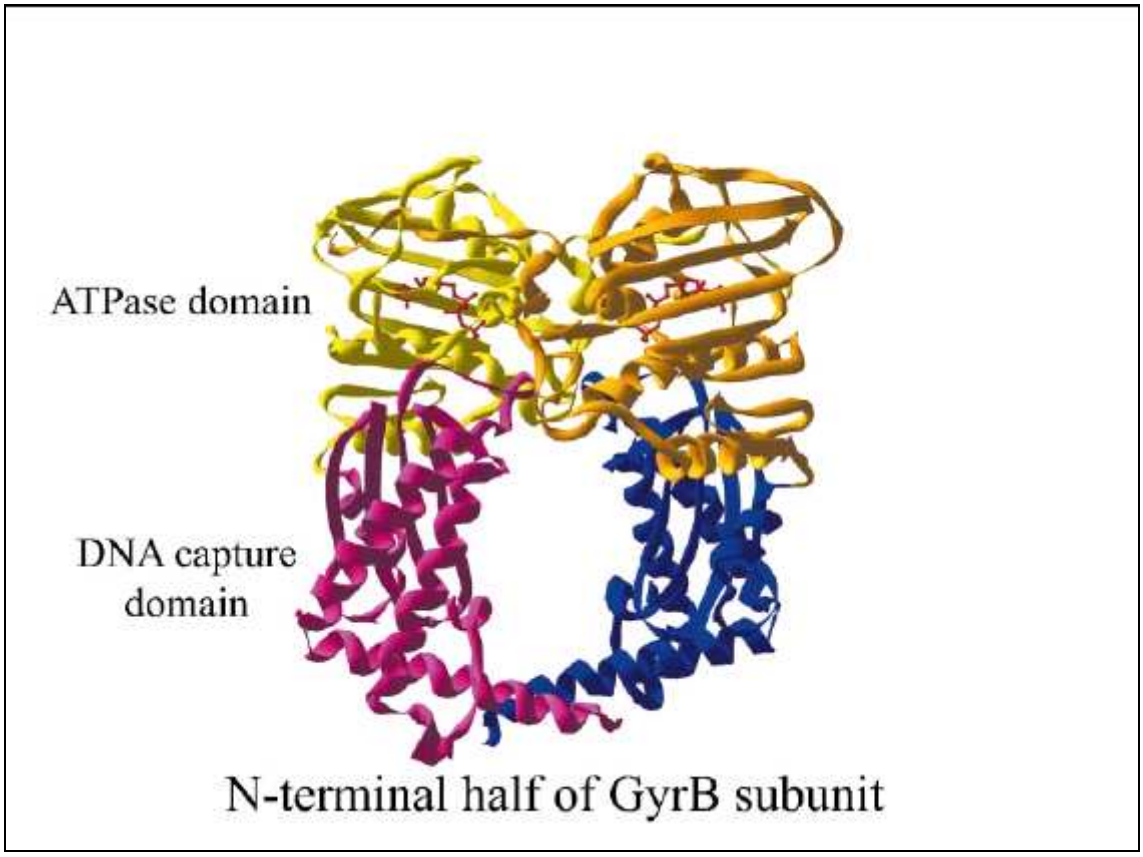


Figure 1.10 The crystal structures for some fragments of subunits of E.coli DNA gyrase [12]

1.4 THE CRUCIAL ROLE OF ATPase REGION OF TOPOISOMERASE II

Gyrase B has an ATPase activity of gyrase. While the C terminal domain (CTD) facilitates the interaction between GyrB and G segment of GyrA, the N terminal domain (NTD) of GyrB is necessary for binding and hydrolysis of ATP [14]. The GyrB-NTD is also involved in formation of the entrance for the T-segment (Fig. 1.11 part b). Coumarin drugs exert their effect by competing for the ATP binding site of GyraseB subunit [15, 16].

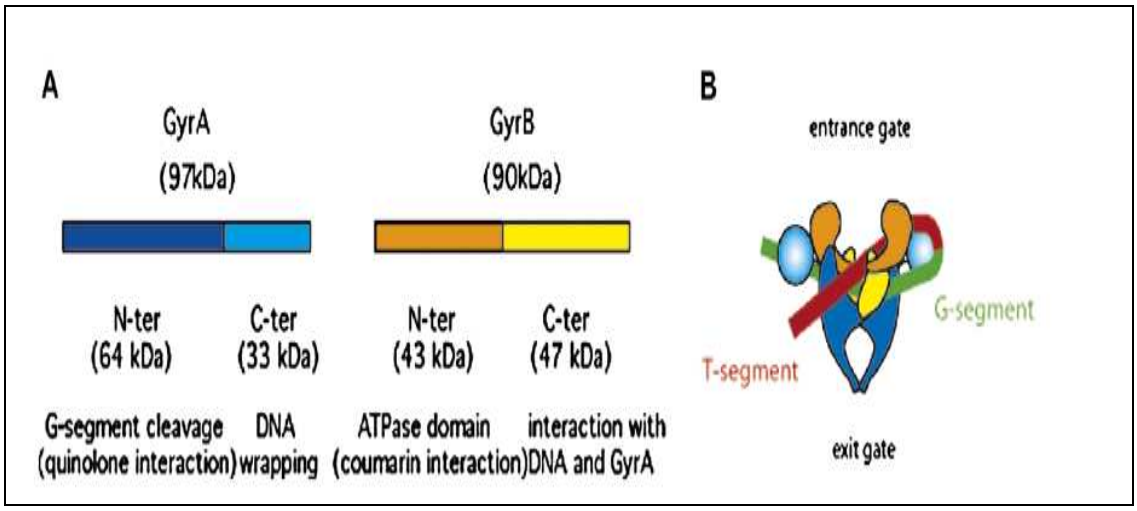


Figure 1.11 The entrance for the T-segment [16].

1.4.1 Catalytic Cycle of GyraseB together with GyraseA

The catalytic cycle model of negative supercoil formation catalyzed by gyrase is given in Figure 1.12. In the first step; gyrase binds the G-segment, through the N-terminal domain (NTD) of Gyrase A. As a second step, gyraseA-C terminal domain (CTD) forms a T-segment that introduces a positive crossing with the attached G segment by wrapping the proximal DNA segment [17, 18]. Then, ATP binds to the GyrB-NTDs. Structural studies indicate that binding of ATP results in dimerization of the GyrB-NTDs and traps the T-segment in the space formed by GyrB-CTDs [13]. Next, the ATP binding results in closure of N-gate which lead to structural changes that increases the strain on T-segment, resulting in cleavage of the G-segment by GyrA and transportation and exit of the T-segment through the gap [19]. Finally, the hydrolysis

products and the T-segment are released and the enzyme returns to its original state ready for the new cycle of catalysis. Two molecules of ATP are hydrolyzed per cycle. The timing of the ATP hydrolysis and product release is yet to be established [16].

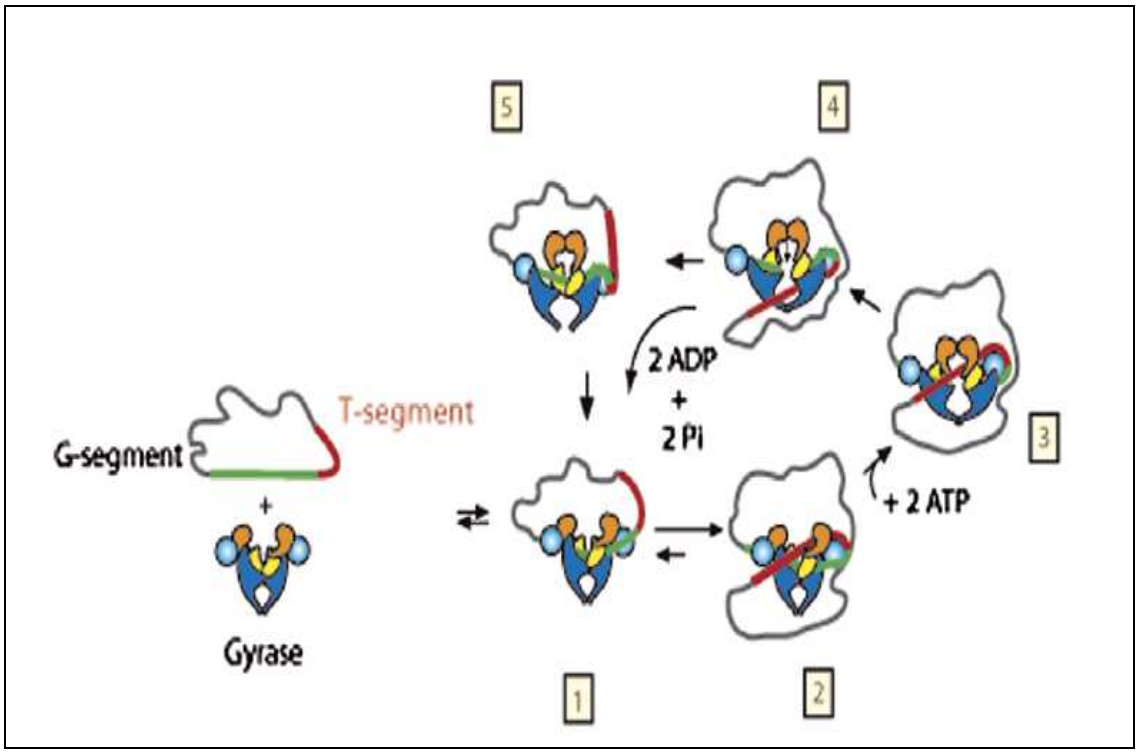


Figure 1.12 The catalytic cycle model of negative supercoil formation catalyzed by gyrase [16]

1.4.2 The Role of ATP

The exact mechanisms by which the ATP binding and hydrolysis lead to structural changes necessary for the strand passage are still unknown. For a long time it has been suggested that the mechanism of DNA supercoiling induced by ATP binding and hydrolysis catalyzed by gyrase involve the conformation changes in the gyrase-DNA complex triggered by binding of ATP. The results of DNA cleavage assays indicate that although ATP is not necessary for the cleavage reaction, binding of ATP simulates the reaction [20]. Differences were observed in the footprints of gyrase-DNA complexes in the presence and absence of ATP. Recent atomic force microscopy experiments have shown that binding of the non-hydrolysable ATP destabilizes the

wrapping of DNA [21]. The structural and ATP hydrolysis experiments indicate that binding of a nucleotide to the GyraseB induces big conformational changes, dimerization and increases the affinity of the enzyme for a second nucleotide [13]. The studies also suggest that the unavailability of the ATP binding site during dimerization might be necessary for the formation of the closed entrance gate conformation and trapping of the T-segment [13]. Although two molecules of ATP are hydrolyzed during one cycle, ATP is not needed for the cleavage process and G-segment binding. As a matter of fact, low levels of strand passage without ATP and the ATPase domain have been observed [22, 23]. It has been postulated that hydrolysis of ATP catalyzes the structural changes that result in transportation of the T-segment through the entrance in the G-segment [16].

1.5 TOPO TARGETTING ANTICANCER DRUG MOLECULES

Type IIA topos are of great biophysical interest and have been extensively studied by the biomedical community. Several antibiotics exert their effect by acting on bacterial gyrases and topo IV [24], while some anticancer drugs act by inhibiting the eukaryotic topo II [25]. One example to this class of drugs is the bisdioxopiperazines used for suppression of the carditoxic effects of anthracycline-based chemotherapies. The same class of drugs is currently in the clinical trials stage for the treatment of leukemia and some solid tumors [26]. Ultracentrifugation and glass-fiber filter studies show that bisidioxopiperazines act by locking the ATPase region of the topo II during the dimerization stage thus blocking the turnover of the enzyme [27].

Structural studies conducted on the type IIA topoisomerase/inhibitor complexes mostly analyzed the interactions between the bacterial DNA gyrase and coumarin compounds [28]. On the other hand, it is still unknown how the drugs recognize and inhibit the eukaryotic topo II enzyme. In order to elucidate the action of bisdioxopiperazine class topo II inhibitors, the structure of 45-kDa S.crevisiae ATPase attached to adenosine 5'- [ß,γ -imino] triphosphate (ADPNP) in the presence and absence of bisdioxopiperazine ICRF-187 is analyzed. The observed structures indicate

that a bisdioxopiperazine molecule acts through stabilizing the dimerized state of the ATPase region of the enzyme by binding to both promoters at the same time and bridging the dimer interface between subunits. These results explain the reason why bisidioxopiperazines class drugs specifically inhibit the ATP bound state of the enzyme and help us understand how amino acid substitutions in the inhibitor-binding region could results in drug-resistant topo II enzymes.

CHAPTER 2

THEORETICAL APPROACH

2.1 FORCE FIELD AND MOLECULAR DYNAMICS SIMULATION METHOD

Computational simulations are used in order to explain molecular systems at the atomic level. By solving the time dependent Schrödinger equation, the structure and dynamics of molecules are determined with the aid of these computational methods. However, the classical physical approaches are suitable to determine the structure and dynamics of molecules such as proteins and nucleic acids. One of these approaches is Molecular Dynamics (MD) simulations method [29].

The most accurate definition of Molecular dynamics (MD) is the science of simulating the motions of a system of particles. In the theoretical study of biomolecules, the usage of the methods of MD simulations plays an important role. To describe and understand the relationships of the structure and the function of biomolecules, MD simulations can be used. Furthermore, MD has the powerful techniques to study fast events that occur on the order of picoseconds to nanosecond time scales. The time dependent behavior of a molecular system using Newton's equations of motions can be calculated by using this computational method [30].

MD simulations method was first appeared in 1977 with the simulation of the bovine pancreatic trypsin inhibitor by McCammon *et al.* [31]. Now, these methods are coordinal used for the investigation of the structure, dynamics and thermodynamics of biological molecules and their complexes. MD simulations method is also used to determine the structures from x-ray crystallography and from NMR experiments [32].

Since the large size of biomolecules and instead empirical potential energy function, are used for the determination of the interaction energy of the particles of a

system as a function of the atomic coordinates, quantum mechanical effects can be neglected in most MD simulations. Additionally, in order to calculate the future position of the particles, the potential energy is used by MD simulations method. Generally, the energy of system is separated into two parts as bonded and non-bonded energies.

$$U_{total} = U_{bonded} + U_{nonbonded} \quad (2.1)$$

Here, energy of the covalently linked atoms is defined as bonded energy (Figure 2.1); atoms bonded with non-covalent interactions have unbonded energy (Figure 2.2).

$$U_{bonded} = U_{bonds} + U_{angles} + U_{dihedrals} + U_{impropers} + U_{Urey\text{-}Bradley} \quad (2.2)$$

$$U_{nonbonded} = U_{Lennard\text{-}Jones} + U_{Coulomb} \quad (2.3)$$

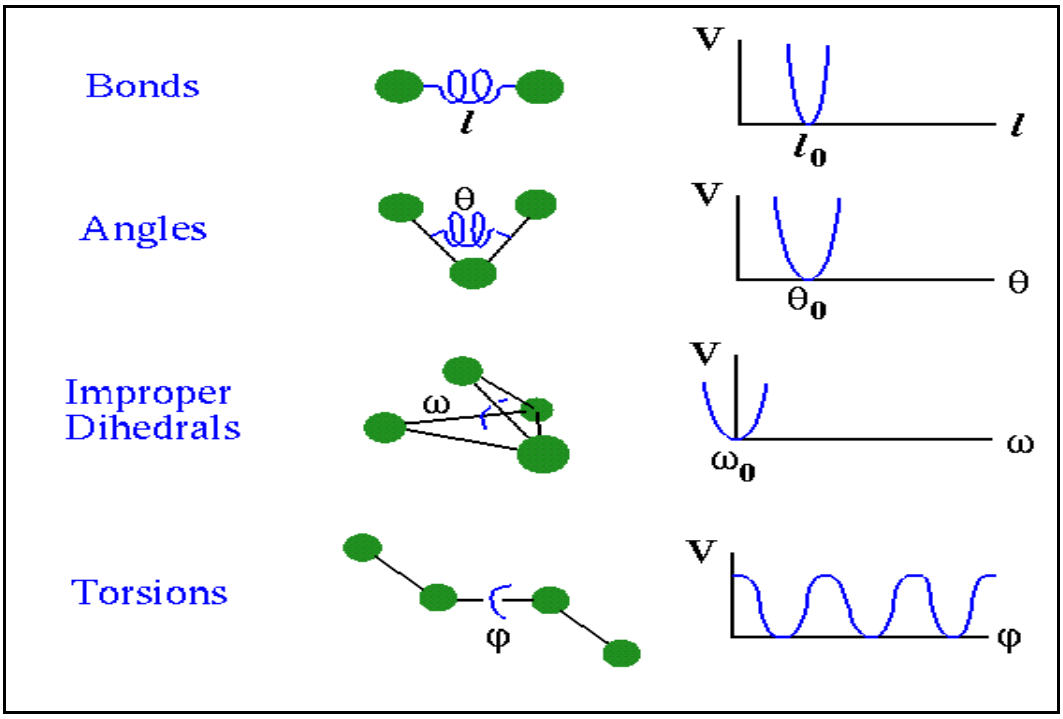


Figure 2.1 Emprical potential energy function for bonded interactions [33].

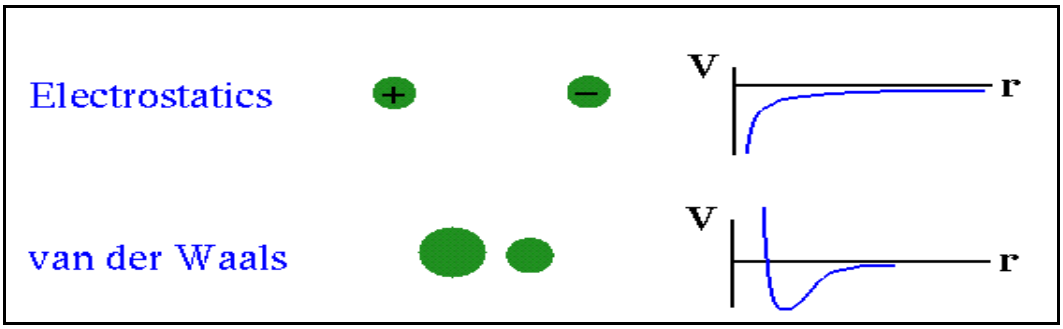


Figure 2.2 Emprical potential energy function for non-bonded interactions [33]

U_{bonds} which is the bond interaction energy is based on Hook's law and it is caused by shared electrons:

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

where k_b is the force constant, b_0 is the equilibrium bond distance and b is the current bond distance. This equation is used to estimate the energy associated with vibration about the equilibrium bond length.

U_{angles} which is the bond angle potential energy is also based on Hook's law. Like U_{bonds} , it can be defined as:

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

where k_θ is the force constant, θ_0 is the equilibrium bond angle, and θ is the current angle formed by three atoms. This equation is used for the estimation of the energy associated with vibration about the equilibrium bond angle.

Dihedral angles are formed by four atoms linearly bonded together. $U_{dihedrals}$ describes the energy due to the dihedral angle potential which is defined by a periodic function:

$$U_{dihedrals} = \sum_{dihedrals} k_{\phi}\left[1+\cos\left(n\phi-\delta\right)\right] \tag{2.6}$$

where k_{ϕ} is the force constant, *n* is the multiplicity (number of equivalent minima), ϕ is the dihedral angle value, and δ is the phase of the dihedral.

$U_{impropers}$ describes the energy due to the planer dihedral angle which is called improper and harmonic similar to the bond potential:

$$U_{impropers} = \sum_{impropers} k_{\omega}\left(\omega-\omega_0\right)^2 \tag{2.7}$$

where k_{ω} is the force constant and ω is the angle between two planes.

$U_{Urey\text{-}Bradley}$ term is also harmonic and based on Hooke's Law. It is defined as:

$$U_{Urey-Bradley} = \sum_{Urey-Bradley} k_u\left(u-u_0\right)^2 \tag{2.8}$$

with Urey-Bradley 1.3 distance, u_0.

The weak interactions between atoms are defined as Non-bonded interactions. Lennard-Jones potential points out the attraction and the repulsion force between non-bonded atoms:

$$U_{Lennard-Jones} = \sum_{nonbonded} \varepsilon\left[\left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{12}-\left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{6}\right] \tag{2.9}$$

The first term describes repulsion at short range and the second term describes attraction at long range. ε_0 is the dept of the minimum energy for the interaction, $\sigma_{i,j}$ is the finite distance at which the interparticle potential is zero and *r* is the distance between particles.

$U_{Coulomb}$ term which is the electrostatic interaction energy between charged atoms is described as:

$$U_{Coulomb} = \sum_{nonbond} \frac{q_i q_j}{\varepsilon r_{i,j}} \tag{2.10}$$

where q is the charge of atoms and ε is the dielectric constant, r is the distance between atoms i and j.

All these bonded and nonbonded equations together with the energy parameters (*e.g.* $k_b, \sigma_{i,j}$) which are required to describe the behavior of different kinds of atoms and bonds, are called force-field [34]. All of the needed values such as energy parameters, bond lengths and bond angles are usually taken from experimental structural data or from quantum chemical calculations.

Initial positions and velocities to all particles in the system should be assigned to start the MD simulations. We can read the initial coordinates of atoms at time t=0 from the crystal structure which are obtained from experimental structure. After this point, we have to determine the force-field potential to calculate the energy of a molecular structure. In order to calculate the future position and velocity of the particles, force-field potential is helpful for us. Firstly, the force acting on each atom can be expressed as the gradient of the potential energy function:

$$F(x) = -\nabla U(x) \tag{2.11}$$

Secondly, the system is moved according to the laws of Newtonian mechanics by integrating equation of motions.

$$F(x) = m\dot{V}(t) = ma\ (t) \tag{2.12}$$

The potential energy is a function of the atomic positions (3N) of all the atoms in the system. There are no analytical solutions to the equations of motion because of the complicated nature of this energy function. The solutions must be solved numerically. Numerous numerical algorithms have been developed for integrating the equations of motion. There are four different algorithms which are known: Verlet Algorithm, Leap-Frog Algorithm, Velocity Verlet Algorithm and Beeman's Algorithm.

The Velocity Verlet algorithm is a more commonly used algorithm in our researches. This algorithm yields position, velocity and acceleration at time t. It is derived from Taylor expansion. So, the velocities and positions are calculated by using these equations:

$$x(t+\Delta t) = x(t) + v(t)\Delta t + \frac{1}{2}a(t)(\Delta t)^2 \tag{2.13}$$

$$v(t+\Delta t) = v(t) + \frac{a(t)+a(t+\Delta t)}{2}\Delta t \tag{2.14}$$

where x is the position, v is the velocity (the first derivative with respect to the time)

The error of the Velocity Verlet is of the same order as the Basic Verlet algorithm. The Velocity algorithm is not necessarily more memory consuming, as it's not necessary to keep track of the velocity at every step in the simulation. The implementation scheme of Velocity Verlet algorithm is:

Firstly, position is calculated:

$$x(t+\Delta t) = x(t) + v(t)\Delta t + \frac{1}{2}a(t)(\Delta t)^2 \tag{2.15}$$

Secondly, velocity is calculated:

$$v\left(t+\frac{\Delta t}{2}\right) = v(t) + \frac{a(t)\Delta t}{2} \tag{2.16}$$

Also, $a(t+\Delta t)$ is derived from the interaction potential.

Finally, velocity is obtained:

$$v(t+\Delta t) = v\left(t+\frac{\Delta t}{2}\right) + \frac{a(t+\Delta t)}{2}\Delta t \tag{2.17}$$

The acceleration comes from the relation

$$F = -\nabla U = ma = m\, d^2x/d\, t^2 \tag{2.18}$$

However, according to this algorithm, it assumes that acceleration at time $a(t+\Delta t)$ only relies on position $x(t+\nabla t)$, and does not depend on velocity $v(t+\Delta t)$ [35].

2.2 HALF QUADRATIC BIASED MOLECULAR DYNAMICS (HQBMD)

Half Quadratic Biased MD is used to induce conformational changes in macromolecules as an external perturbation. The time dependent perturbation is used in order to introduce a very small perturbation to the short time dynamics of the system. It does not affect the conservation of the constants of motion of the system. Here, the conservation of the total energy or of the suitable conserved quantity can be used to check of the correctness of the forces when an extended Lagrangian is used. The external perturbation needs:

(1) A reference (or target) structure,

(2) A reaction coordinate which defines a “distance” from the reference structure [36].

HQBM introduce a half quadratic perturbation on a given reaction coordinate [37]. The system was biased half-harmonically, i.e. only when it moves away from the target, with an external potential of the form:

$$W(r,t) = \frac{\alpha_{hqbmd}}{2}(\rho - \rho_0)^2 \tag{2.19}$$

where α_{hqbmd} is force constant and ρ is reaction coordinate leading from the initial to the final state is given by:

$$\rho(t) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \left(r_{ij}(t) - r_{ij}^{R} \right)^2 \tag{2.20}$$

where $r_{ij} = |r_i - r_j|$ is the distance between atoms *i* and *j*, and *R* labels the coordinates of the final reference structure, and *N* represents the total number of atoms biased.

We have calculated average forces that are needed to open the gates of protein on both upper and lower sides of it around three and nine different axes respectively. The forces are obtained from the applied external potentials, as shown below:

$$F(t) = d/dr\,[W(r,t)] \tag{2.21}$$

$$F(t) = \alpha_{hqbmd}\ \ [\rho(t+\Delta t) - \rho(t)][\rho(t)]^{1/2} \tag{2.22}$$

In our research, the forces are calculated for every nine different axes which have 20 degrees between two axes on the lower side of our protein and on the upper side; the forces are calculated for three different axes like the lower side to get dynamic and energetic results about the protein.

This technique has been used extensively in our research, especially to see how the gates of protein open in a short period of time. The details of the theoretical set-up of opening of the gates are given below.

2.3 CONSTANT PRESSURE/TEMPERATURE MOLECULAR DYNAMICS (CPT MD)

A thermostat is needed to investigate the behavior of a system at a specific temperature for simulation. The usage of thermostat for simulations also avoids steady energy drifts caused by the accumulation of numerical errors during MD simulations [38]. The simulations are performed at constant temperature by using Brendsen constant temperature method [39].

Constant pressure/temperature MD method performs simulations in the constant temperature situation. This method considers that the system is thermally connected with a huge external system (a heat reservoir) and system exchanges energy with the heat reservoir. The continuous dynamics with well defined conserved quantities are the advantage of constant pressure/temperature MD method [40].

In CPT method, there are two different types of constant temperature methods:

(1) Simple method: System coupled to one heat bath,

(2) Multi-bath method: Different parts of system can be coupled to different heat baths [37].

2.4 COMPUTATIONAL TOOLS

2.4.1 Software

CHARMM (Chemistry at HARvard Macromolecular Mechanics) is one of the most popular molecular mechanics programs for biomolecules which is written by Brooks *at al.* in 1983 [40] CHARMM is widely used molecular simulation program with broad application to many- particle system which includes macromolecular energy, minimization and dynamics calculations. It provides a large suite of computational tools that encompass numerous conformational and path sampling methods, free energy estimates, molecular minimization, dynamics, and analysis techniques, and model-building capabilities. We performed all simulations by using CHARMM.

We also used Visual Molecular Dynamics (VMD), gnuplot program for plotting, and several linux shell scripts that we wrote to manipulate data and to manage the jobs other than the main software package CHARMM.

2.4.2 Hardware

We studied with a clustered supercomputer whose name is *ULUBATLI*, in our research. Ulubatli is a linux cluster machine which has a total of 52 CPUs and a total of 78 GB RAM memory. It uses Load Sharing Facility (LSF) to manage our jobs and CMU (Cluster Management Utility) for cluster managements. We have Intel Xeon 5160 dual core and 5462 quad-core CPUs. All CPUs are related to each other through a fast networking technology (up to 40 GB/s), *infiniband.* The system has also HP MSA 30 disk storage unit. Disk storage unit has a total disk capacity of 4.2 TB. Our clustered supercomputer, *ULUBATLI*, is fully financed by The Scientific & Technological Research Council of Turkey (TUBITAK) grant no: 107T209.

2.5 DATA STRUCTURE FILES

We have several fundamental data structures in the CHARMM program to have information about the molecule, its composition, its chemical connectivity, certain atomic properties, internal coordinates and parameters for the energy function and more. The topology file and the parameter file contain the necessary data for a particular class of molecules; e.g., proteins or nucleic acids. The data for our specific protein is extracted from these files and stored in the Protein Structure Files (PSF) [35].

2.5.1 Residue Topology File (RTF)

The definitions of the molecular building blocks (residues) which are used to build large molecules, for example, proteins and DNA bases are contained into the residue topology file. This file has the name of each atom type in the parameter set, its mass, hydrogen bond donors and acceptors, as well an atom's partial charge in a particular residue (amino acid). For the definition of an atom in a particular bonding situation, we used atom types. The covalent structure, for example, how the atoms are connected to one another to form amino acids, DNA bases or lipid molecules, is defined for each residue into this file [35].

2.5.2 Parameter File (PARAM)

All the necessary parameters for calculating the energy of the molecules are contained into the parameter file. It is related with the RTF file. The equilibrium bond distances and angles for bond stretching, angle bending and dihedral angle terms in the potential energy function as well as the force constants and the Lennard-Jones 6-12 potential parameters are also included in this file [35].

2.5.3 Protein Structure File (PSF)

The Protein Structure File (PSF) which is the most fundamental data structure in the CHARMM program is generated for a specific molecule or molecules. Also; it is a

state of being linked together of the information contained in the RTF file. The detailed composition and connectivity of the molecules of interest is contained in the PSF file. This file gives the total number of bonds and identifies which atoms make up a particular bond. Similarly, it is the file where triplets and quadruplets of atoms are listed for the bond angles and torsion angles respectively. Improper dihedral angles are also listed in this file. These angles are used to maintain planarity and prevent accidental chiral inversion. Additionally, several segments, which can range from a single macromolecular chain to multiple chains solvated by explicit water molecules, can be included in the PSF file.

We have to specify the PSF before any calculations can be performed on the molecule. The PSF file contains the molecular topology. However, it does not establish information regarding the bond lengths, angles, etc. Because of this reason, we need to read all of these in the parameter file to add the missing information [35].

2.5.4 Coordinate File (CRD)

The coordinate file has the Cartesian coordinates of all atoms which are obtained from x-ray crystal structures or from NMR structures, in the system. In the manner of alternatives, the Cartesian coordinates can be generated from homology modeling studies or from the library values which is contained in the topology and parameter files along with some secondary structure information. We can build the missing coordinates within the CHARMM program by using the internal coordinate facility. Additionally, we may add hydrogen atoms by using the HBUILD module in the CHARMM because these atoms are not present in x-ray crystal structures of proteins. We can accommodate two sets of coordinates in the CHARMM program because of the advantage of them for variety of calculations. For example, it can be useful for the calculation of the root mean square difference (rmsd) between a structure from the simulation and the experimental structure [35].

2.6 SYSTEM PREPERATION

2.6.1 PSF Generation

There are several steps that must be followed to set up CHARMM calculations. Firstly, we generated the PSF. Then we loaded the initial coordinate and placed hydrogen atoms. We executed some of these steps by using the *build.inp* CHARMM file. This file holds the appropriate CHARMM command to read in coordinate file as PDB format and to generate the necessary CHARMM data structures [35]. The initial coordinates of the E.Coli GyraseB were obtained from the Protein Databank (entry code 1EI1) (Figure 2.3) [12].

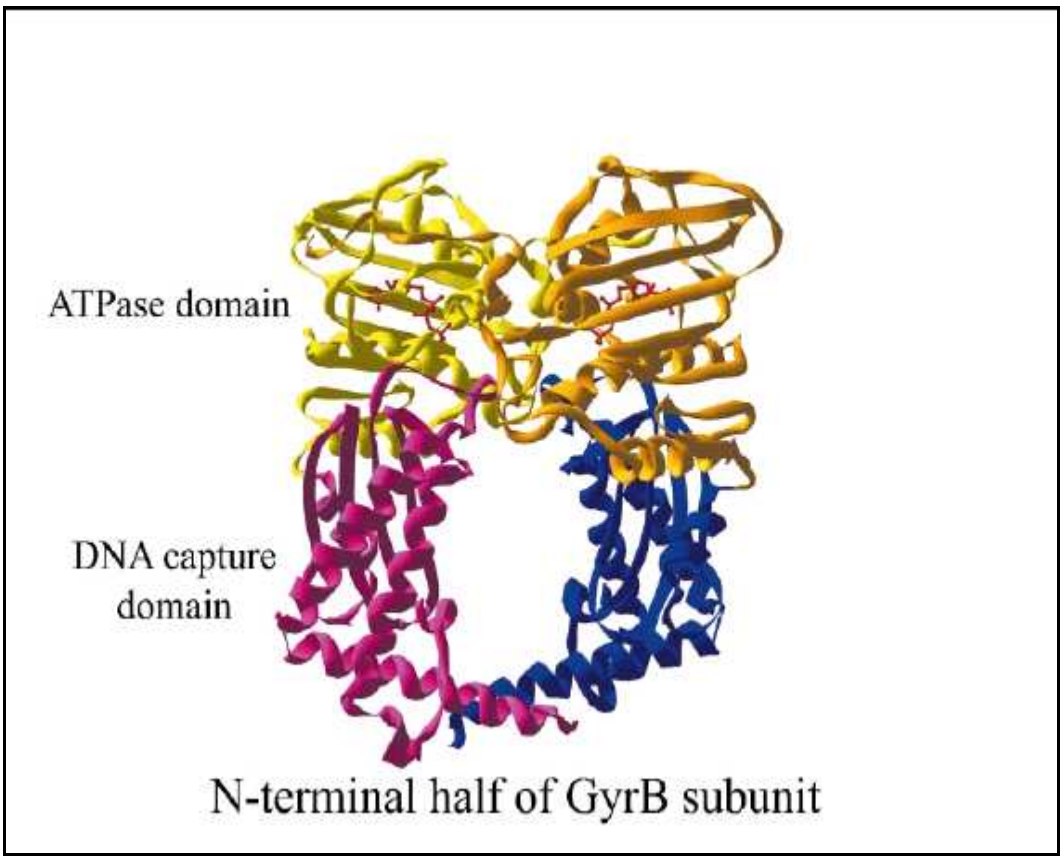


Figure 2.3 E.Coli GyraseB obtained from the Protein Databank (entry code 1EI1) [12]

At the beginning of the *build.inp* input file, we have title and commands to load the topology (RTF) and parameter (PARAM) files. After that, we read the protein sequence from the coordinate file. Then; we generated the PSF with the segment id (segid). Also, during this process, the hydrogen atoms were appended to the system by using HBUILD which is given in the input file, *build.inp*. The coordinates of the protein were printed to the output file after the PSF file was built and the coordinate file confirmed to be complete. Also, the *prt_wat.psf* and full coordinates set were written to the disk.

2.6.2 Preparation

In preparation part, the system is brought to a state where the actual dynamics can be performed. The minimization of the initial structure to relieve local strain and the heating up the system to the desired temperature are included into this step.

2.6.2.1 Energy Minimization

The minimization input file which is named as *in_minimize.inp* sets up an energy calculation. In this step of preparation part, firstly, the topology and parameter files were read into CHARMM. Then, we read the PSF file and the coordinate file which were generated previously. For the cutoff model, we used a 14 A^0 cutoff with a SWITCH truncation function on both the Van deer Waals and the electrostatic interactions. Following specification of the cutoff model, we calculated the energy and used the Steppest Descent (SD) and the Adopted Based Newton- Rapson (ABNR) algorithms for 1000 steps to minimize the energy by slightly changing the structure. (Figure 2.4) The energy minimized coordinates were then written to a new file named *system_mini.pdb*. In addition to printing out the total energy, the CHARMM program also gives us the internal energies such as the bond energy, the angle and dihedral energy terms, as well as, the non-bonded energy terms such as the Van deer Waals energy and electrostatic energy [35].

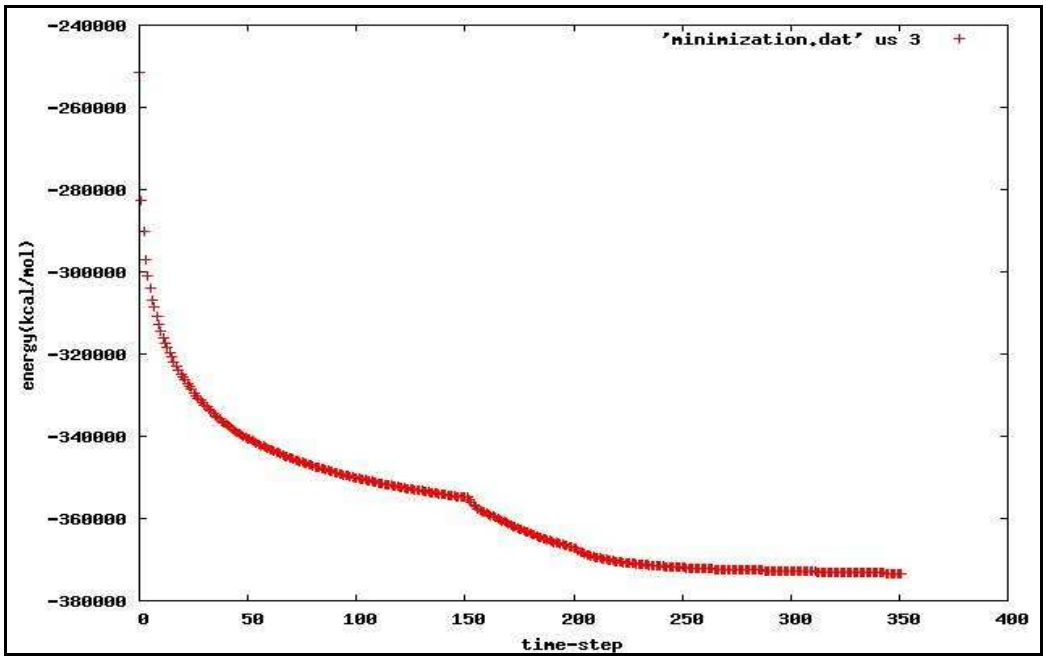
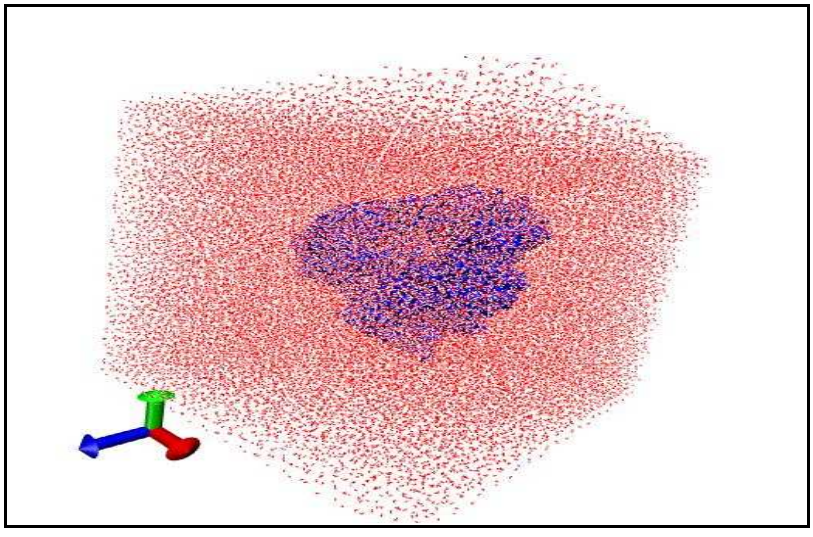


Figure 2.4 The Steppest Descent (SD) and the Adopted Based Newton- Rapson (ABNR) algorithms for 1000 steps to minimize the energy by slightly changing the structure

After this step, we put the system into water box. In this process, firstly, we created a cubic volume water box. Then, in the same input file, we got our spherical volume water box with 10 A^0 radius. (Figure 2.5) As in the previous steps, new PDB, PSF and CRD files were written. Additionally, we checked the charge of the system which is -20. So, we added 10 magnesium (Mg^{+2}) atoms into our system to neutralize it. (Figure 2.6)

a)



b)

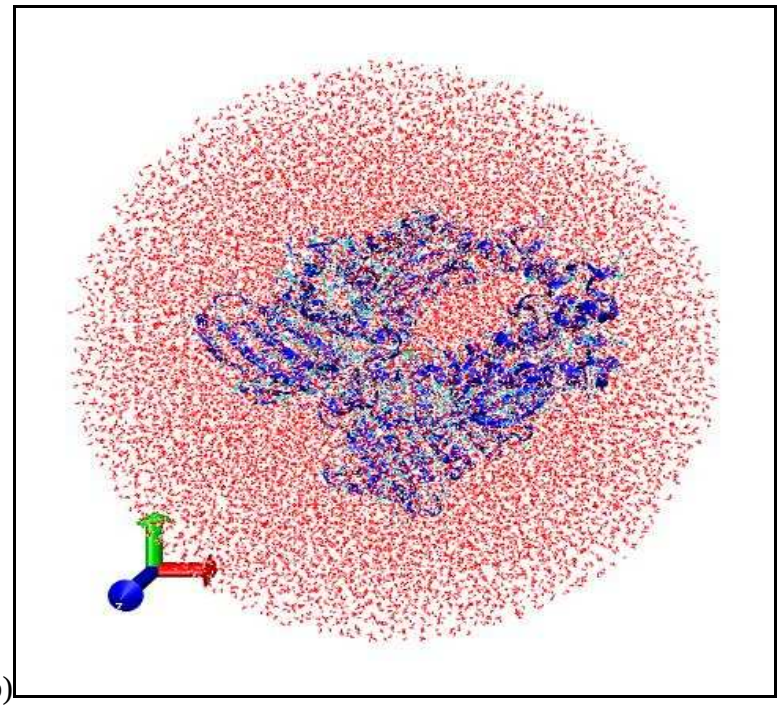


Figure 2.5 The cubic volume water box and the spherical volume water box with 10 A^0 radius

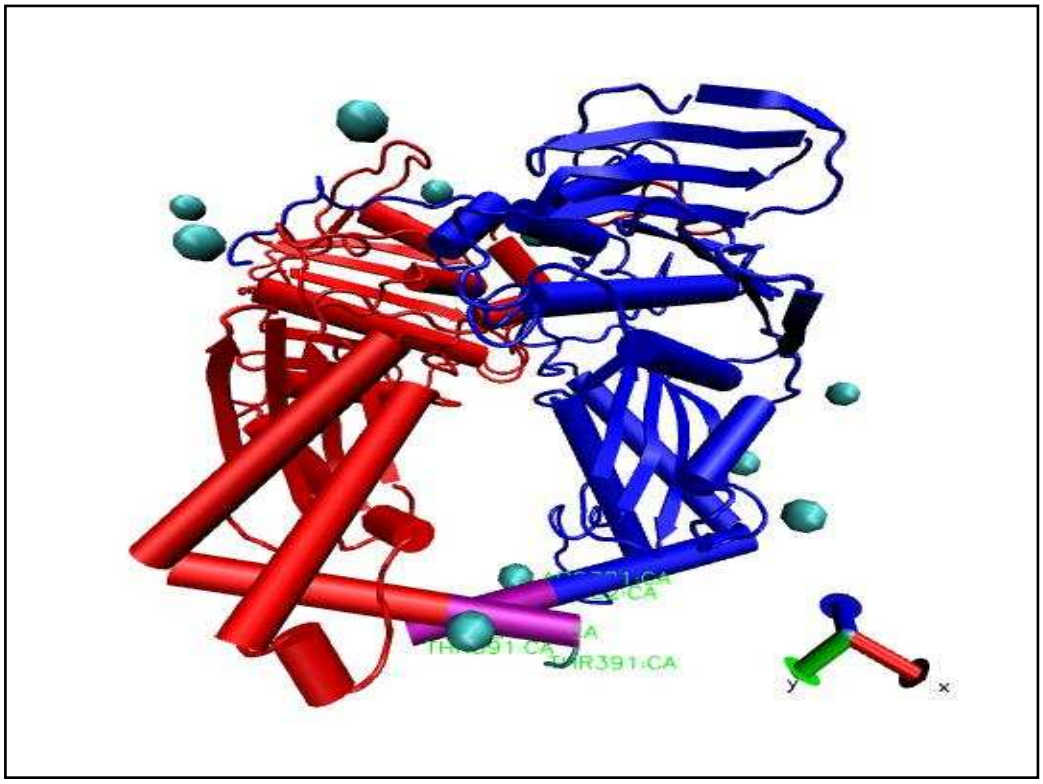


Figure 2.6 10 magnesium (Mg^{+2}) atoms in the system

2.6.2.2 Heating the protein

Here, we heated to the system to the desired temperature. Similarly, the topology, parameter and the PSF files were into CHARMM. Then, the energy minimized coordinates which was calculated in the minimization part, were used and the same cutoff model was used for the energy minimization which is specified.

We used the CHARMM DYNAmics command to run the molecular dynamics. The restart file was written at the end of this job. This file allows us to continue the simulation at another time. We used the VERLET integration algorithm to run the molecular dynamics simulation [34]. The simulation started at 0 K and the system was heated up to 300 K every 100 fs (= 0, 1 ps). (Figure 2.7) The simulation was run for 100000 steps by using a 0,001 picosecond time steps. This job took 6, 14 CPU seconds on our supercomputer machine running with LINUX.

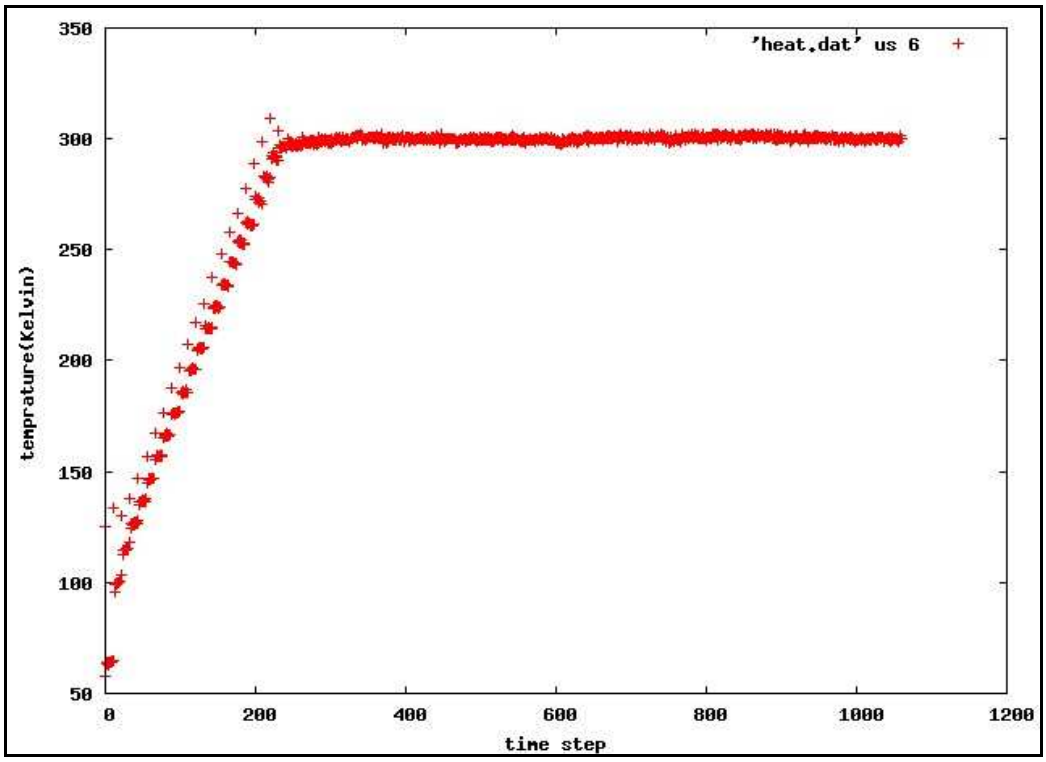
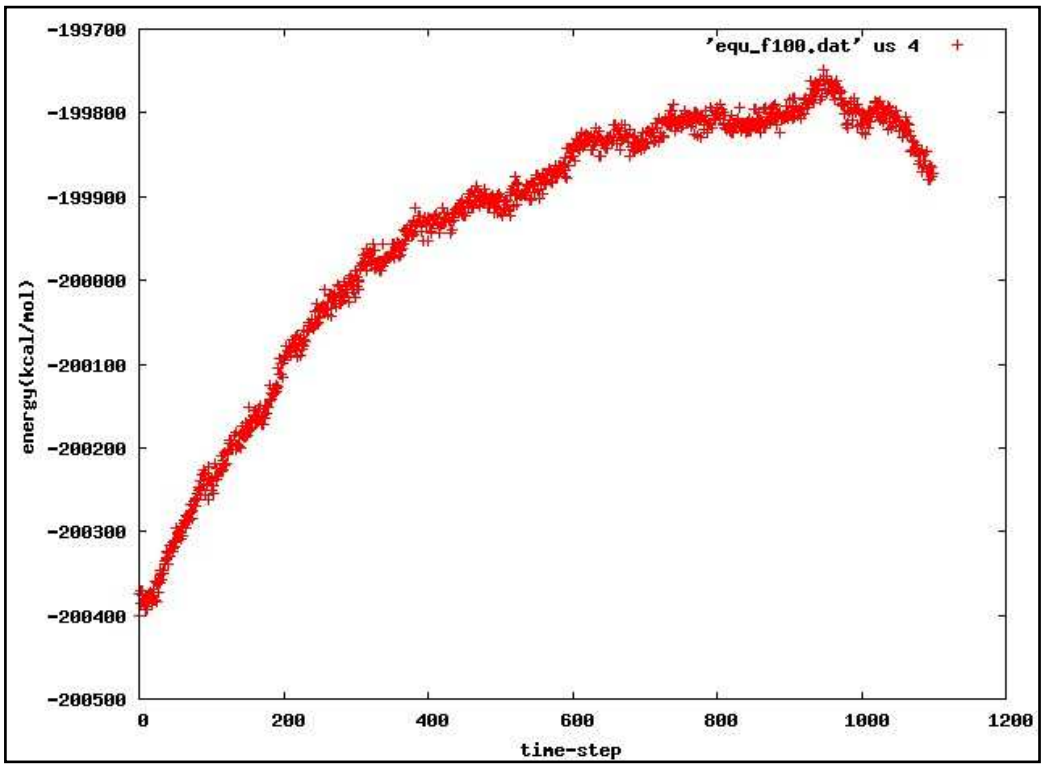


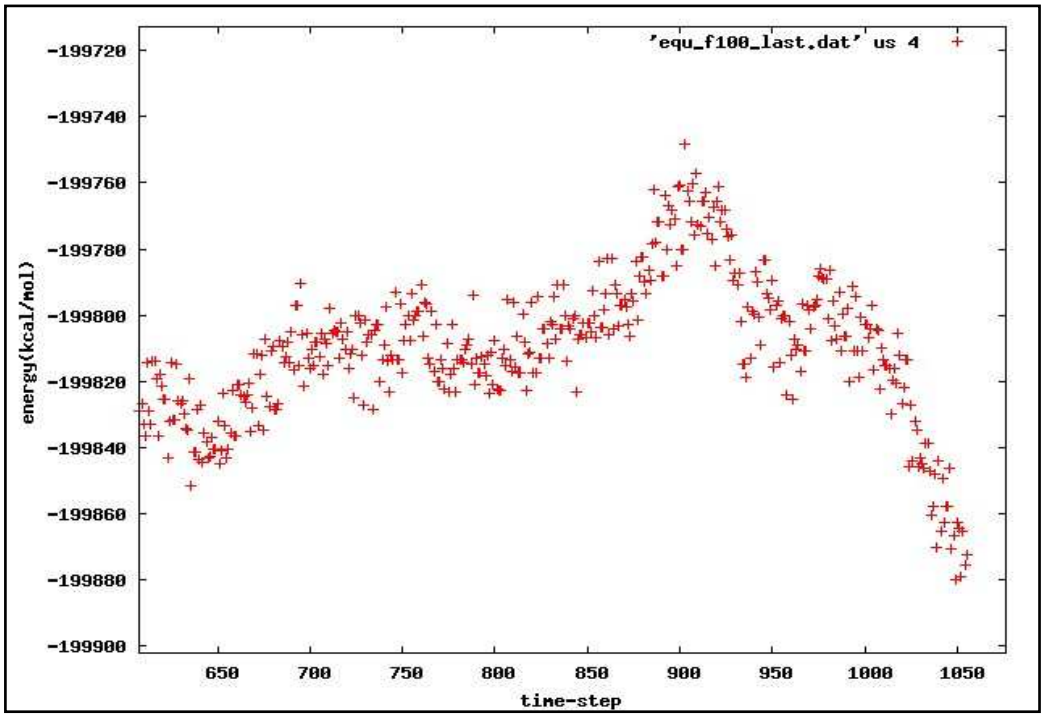
Figure 2.7 The system was heated from 0 K to 300 K

2.6.3 Equilibration

In this step, we equilibrated the system to ensure that the system is in a stable state after heating it to 300 K. In this job, our system was equilibrated for 100 ps at 300 K. The trajectory which was started in previous heating step was continued. Here, our purpose was to keep the system at the target temperature. During this part of the simulation, the temperature was checked periodically to hold it at the desired temperature. At the end of this step, we checked the fluctuation of energy and plotted the total energy, kinetic energy and potential energy as a function of time. (Figure 2.8)



a)



b)

Figure 2.8 The total energy as a function of time (a) and the details of the total energy for nearly last 20000 steps (b)

2.6.4 Preparing System for Biased Molecular Dynamics

In this study, our purpose is to investigate the mechanism of gates at upper and lower sides of the protein by employing Molecular Dynamics. We have used the HQBMD to see the large scale of movement of the protein. During the simulations, rigid bonds are constrained by using SHAKE algorithm. Also, we used Brendsen constant thermostat to control the temperature of the system.

As we have observed the openings of both sides of the protein, we tried to find the most suitable gates of movement energetically and structurally. In order to define the axes of movement, we initially determined the center of masses of whole protein, CHNA and CHNB. After this step, we totally replaced 16 dummy atoms instead of 16 CA atoms which are included in the last sequences of both CHNA and CHNB to see the opening of the gate at lower side of protein. (Figure 2.9)

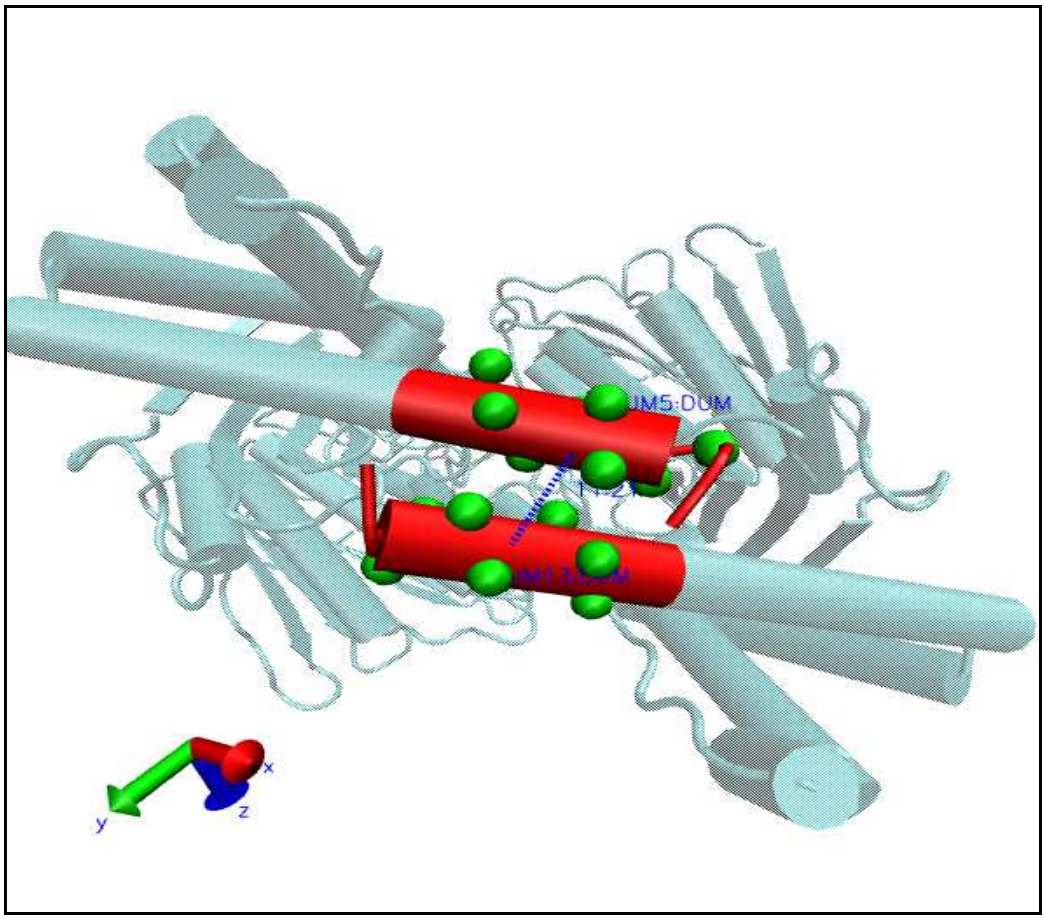


Figure 2.9 Sixteen CA atoms which are included in the last sequences of both CHNA and CHNB to see the opening of the gate at lower side of protein.

Then, these 8 dummy atoms of CHNA and CHNB were opened to get 20 A^0 gap between them to define the target (reference) state. The first opening is called as 0^0 degree opening. After the first step, the same procedure is repeated for 20^0 , 40^0, 60^0 , 80^0 ,100^0 ,120^0 ,140^0 and 160^0 degrees respectively according to the first one. (Figure 2.10)

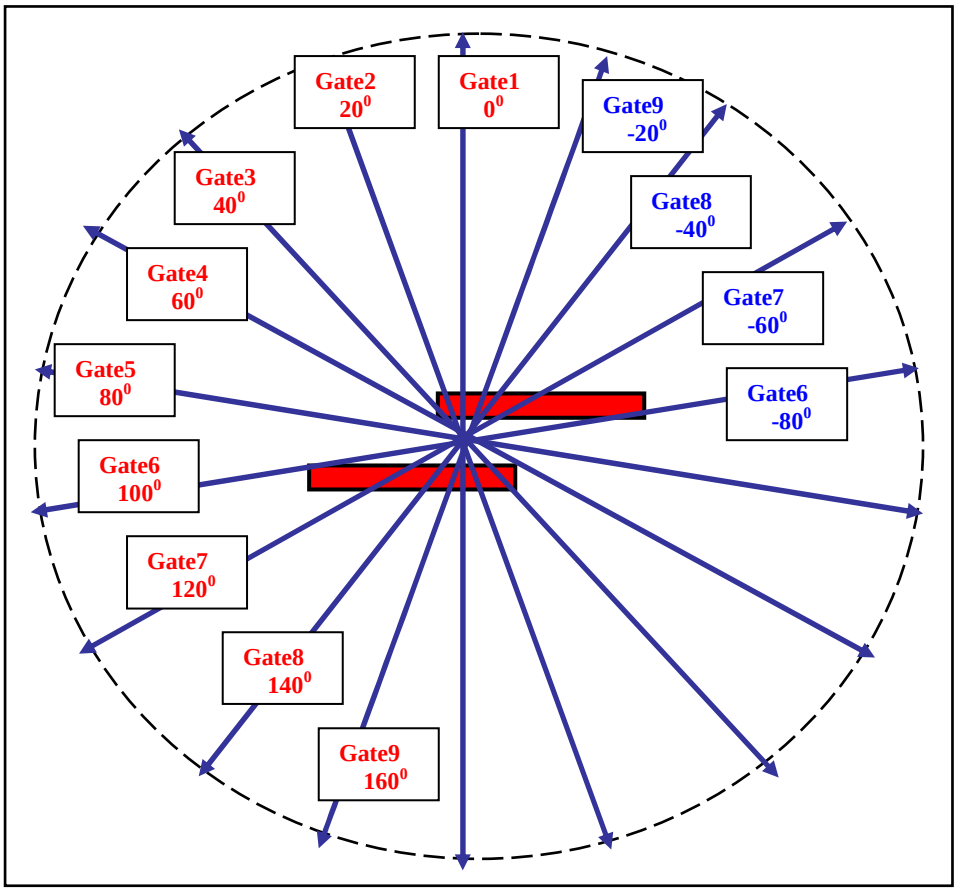


Figure 2.10 Nine different gates (gate1, gate2, gate3, gate4, gate5, gate6, gate7, gate8, gate9) at lower side of protein *(red boxes show the last sequences of both CHNA and CHNB, blue lines are gates)*

All target crd files are written after the gates are defined. In this way, we defined 9 different gates of movement to reach our purpose for the gate at lower side of protein. We did not prefer to repeat the same openings between 180^0 and 360^0 degrees because they will be symmetric of the first 9 openings.

For opening of the gate at upper side of the protein, we firstly defined the centers of masses of the first three helixes on CHNA and CHNB. These centers of masses are exactly on the CA atoms whose segid numbers are 18, 91,122 on both CHNA and CHNB. Then, 6 dummy atoms were replaced instead of these CA atoms. (Figure 2.11)

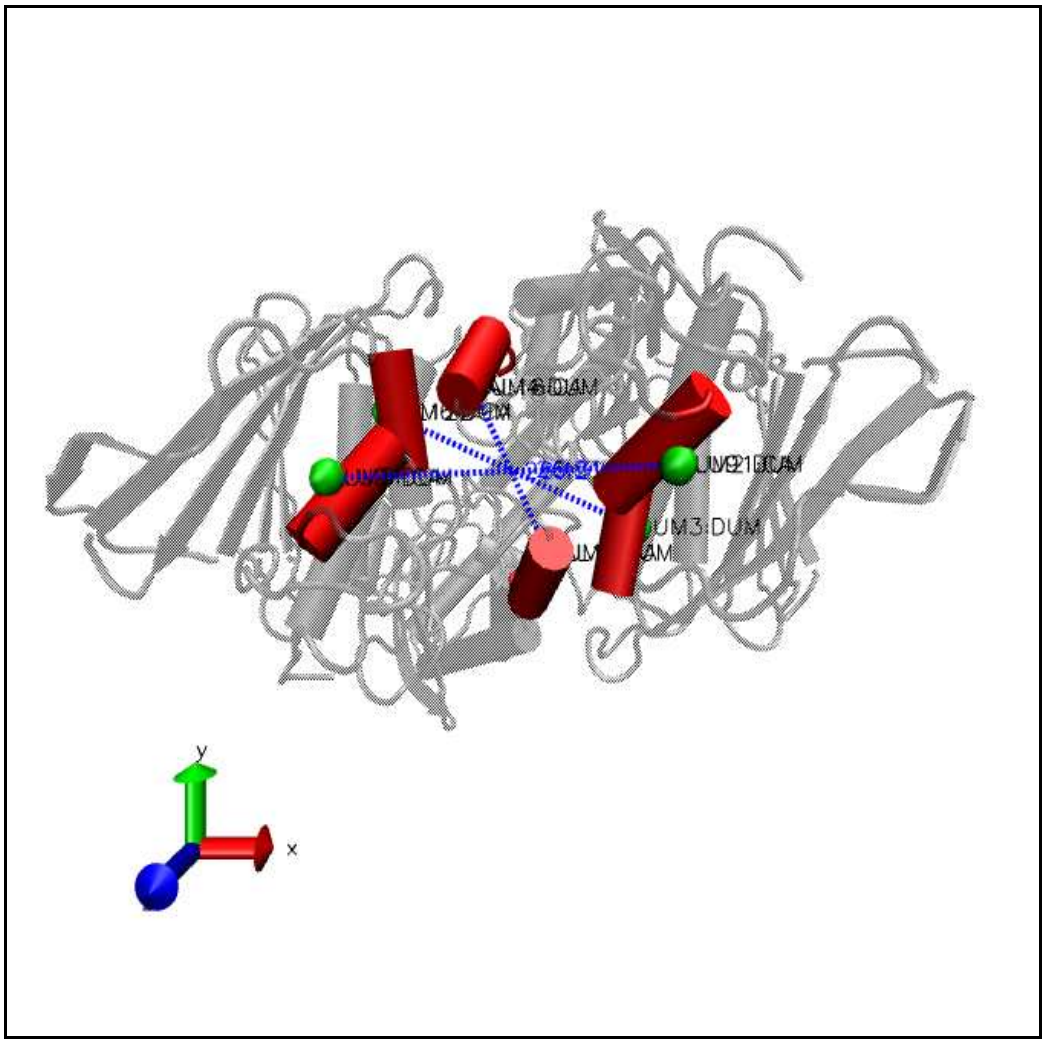


Figure 2.11 Six dummy atoms replaced instead of these CA atoms whose segid numbers are 18, 91,122 on both CHNA and CHNB

Like the lower side of the protein, these 6 dummy atoms were opened symmetrically to determine the target state in order to get at least 20 A^0 gap between them. So, we also have 3 different gates at upper side of protein. (Figure 2.12)

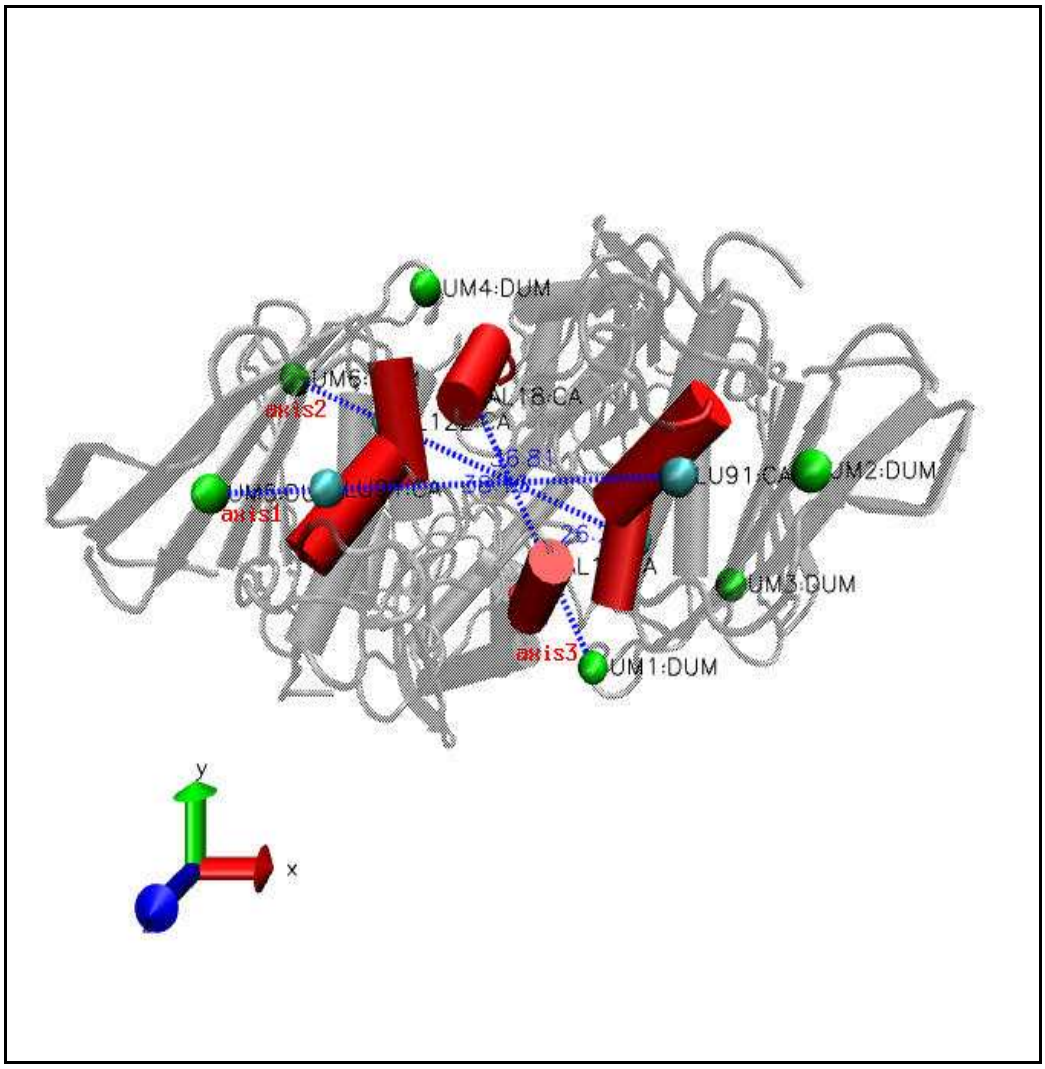


Figure 2.12 Three different gates (gate1, gate2, gate3) at upper side of protein

All the axes on both upper and lower sides of protein are nearly perpendicular with the axis which passes on the center of whole protein. After the definition of gates, HQBMD are employed with α_{hqbmd} which is chosen 1 kcal/mol/A^4 to understand the mechanism of openings of the gates at both upper and lower sides of protein energetically and structurally. Also, we employed HQBMD for the upper side in order to understand the mechanism of openings of the gates better. In his time, we used α_{hqbmd} as 10 kcal/mol/A^4 .

CHAPTER 3

COMPUTATIONAL AND THEORETICAL RESULTS

3.1 STRUCTURAL RESULTS

In the structural analysis, we have derived the root mean square deviations (RMSD) between the initial and final structures of each secondary structure of the protein. This analysis will give us information about the internal deformations in the protein, while the gates are opening. The results of RMSD analyses are given in figures from 3.1 to 3.15.

Secondly, the distance between each amino acid of the two chains are also analyzed, as this will give us a direct measurement of the amount of opening between the two chains. The distances are calculated form the Cα s of the corresponding amino acids. The results are given in figure 3.16 to 3.19.

As an important point, the amount of rotation in the protein as the DNA is passing through it is an important point given in the literature. Therefore, we have looked at the amount of rotation during our simulations. The rotations are calculated as superimposing the final geometry to the initial one, by minimizing the RMSD deviations. The relative rotation of the two helices, α13 of each chain, with respect to he upper part of the protein is also a focus of interest. Therefore, we have analyzed the rotation of each monomer as well as the rotation of the lower gate with respect to the upper one. The results that we obtained are given in table 3.1 to 3.6.

First, the aforementioned data derived from the simulations are given, the discussions on the figures will be in the next chapter.

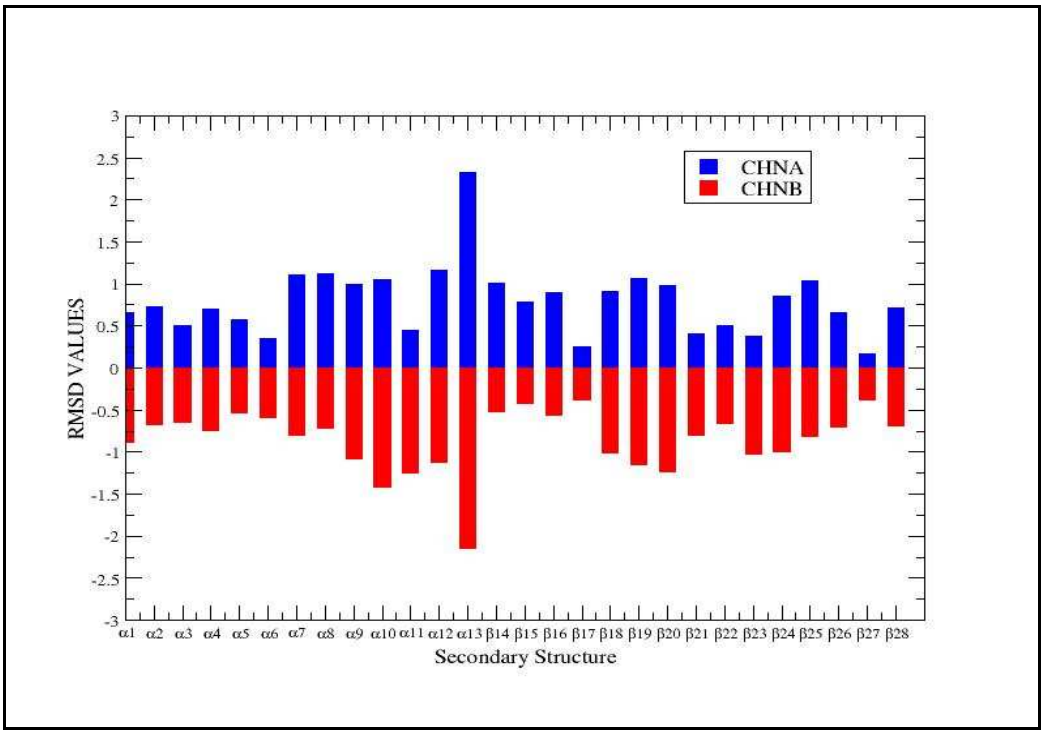


Figure 3.1 Rmsd (root mean square deviation) between the final and initial geometries of each secondary structure, as the opening is at the lower part of the protein along the gate1.

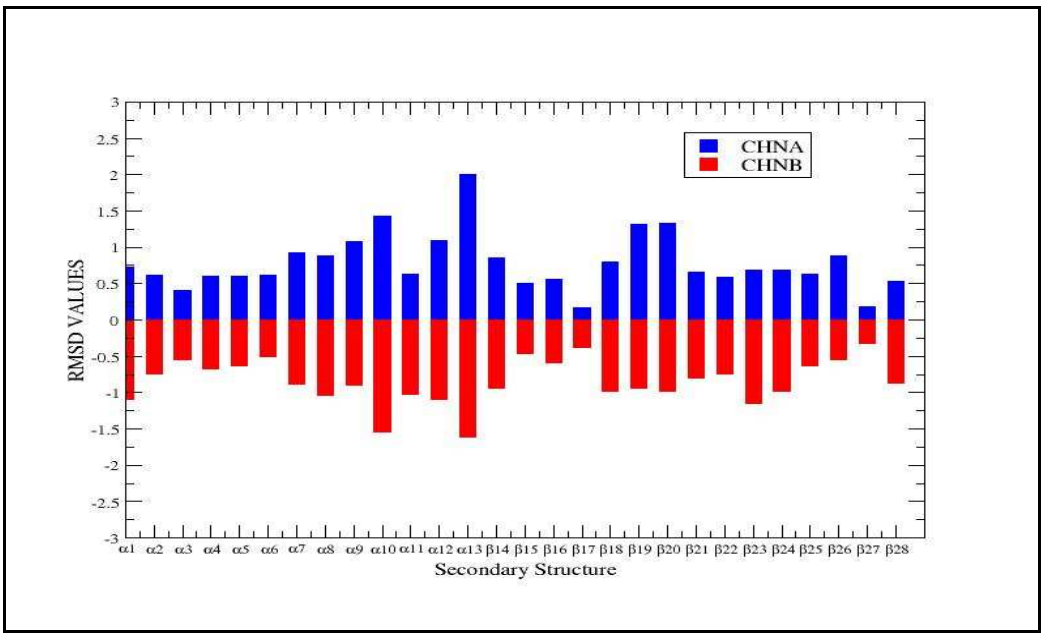


Figure 3.2 Rmsd (root mean square deviation) between the final and initial geometries of each secondary structure, as the opening is at the lower part of the protein along the gate2.

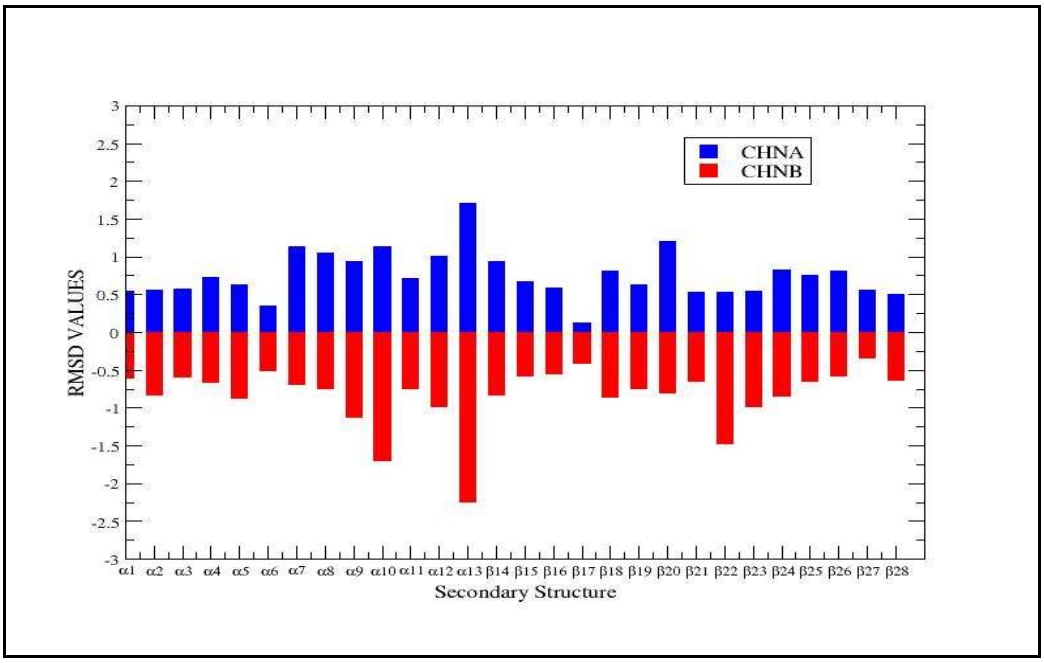


Figure 3.3 Rmsd (root mean square deviation) between the final and initial geometries of each secondary structure, as the opening is at the lower part of the protein along the gate3.

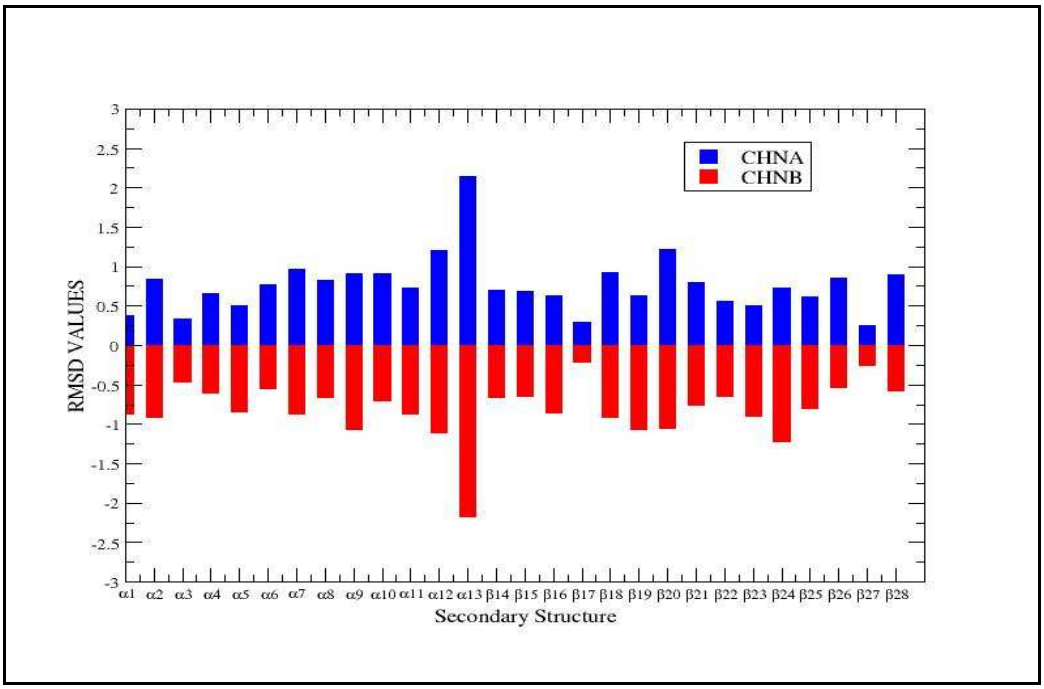


Figure 3.4 Rmsd (root mean square deviation) between the final and initial geometries of each secondary structure, as the opening is at the lower part of the protein along the gate4.

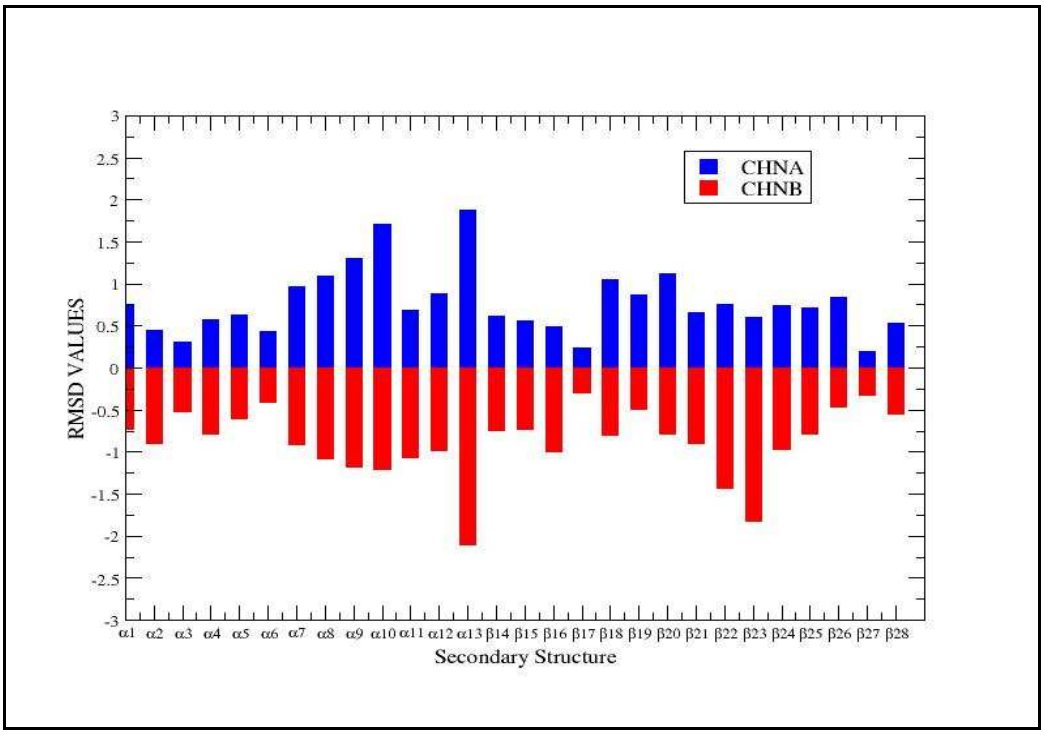


Figure 3.5 Rmsd (root mean square deviation) between the final and initial geometries of each secondary structure, as the opening is at the lower part of the protein along the gate5.

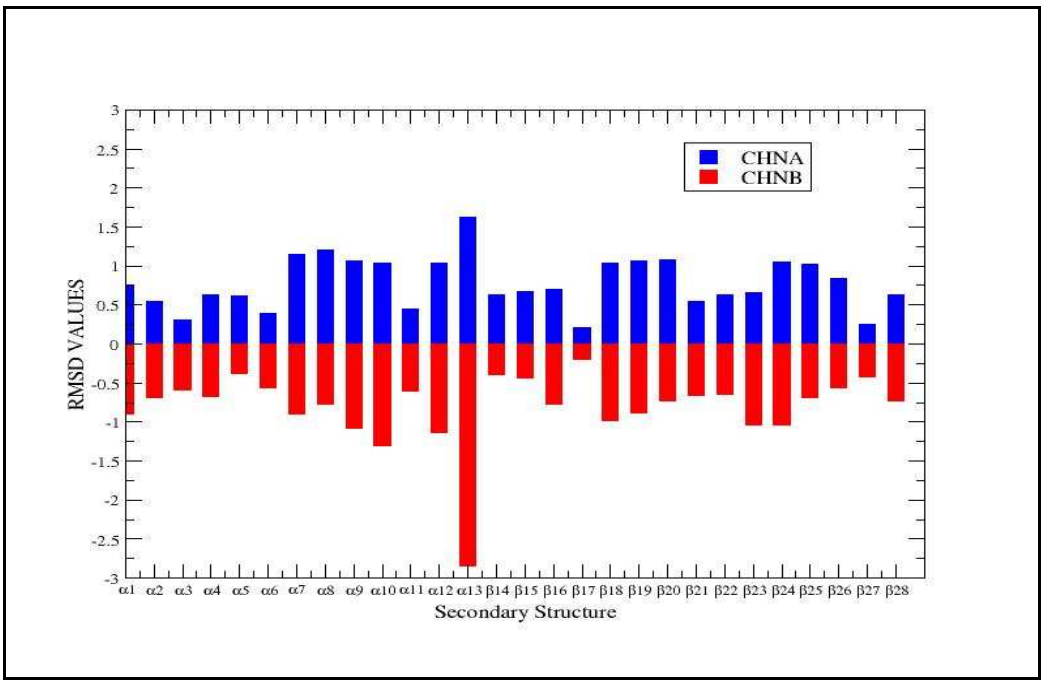


Figure 3.6 Rmsd (root mean square deviation) between the final and initial geometries of each secondary structure, as the opening is at the lower part of the protein along the gate6.

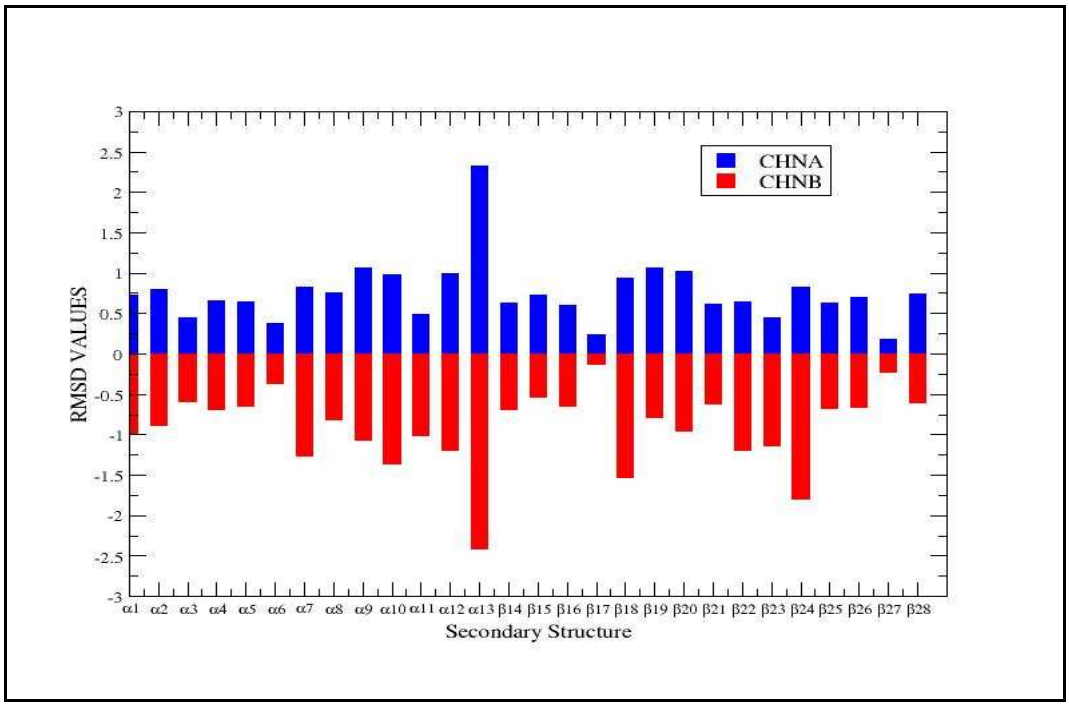


Figure 3.7 Rmsd (root mean square deviation) between the final and initial geometries of each secondary structure, as the opening is at the lower part of the protein along the gate7.

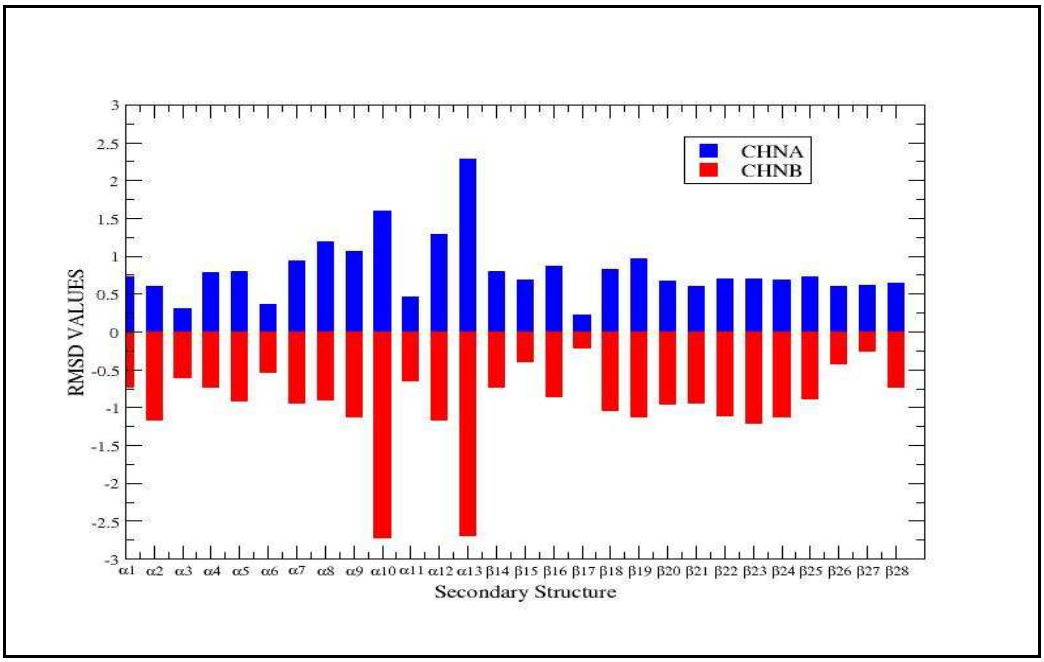


Figure 3.8 Rmsd (root mean square deviation) between the final and initial geometries of each secondary structure, as the opening is at the lower part of the protein along the gate8.

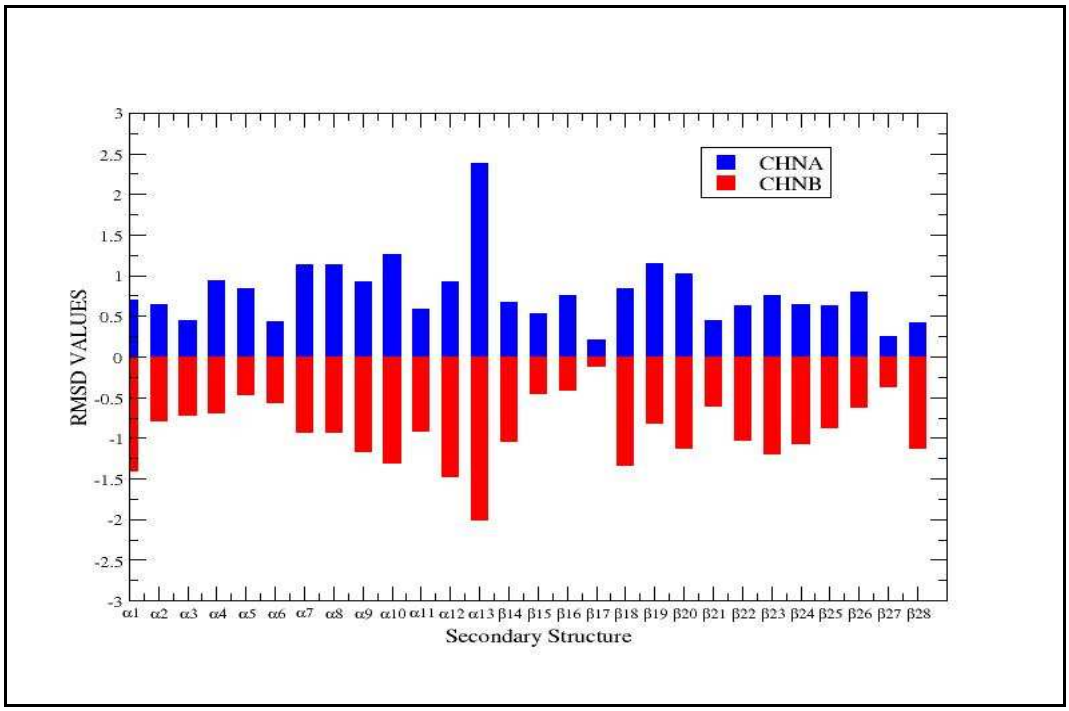


Figure 3.9 Rmsd (root mean square deviation) between the final and initial geometries of each secondary structure, as the opening is at the lower part of the protein along the gate9.

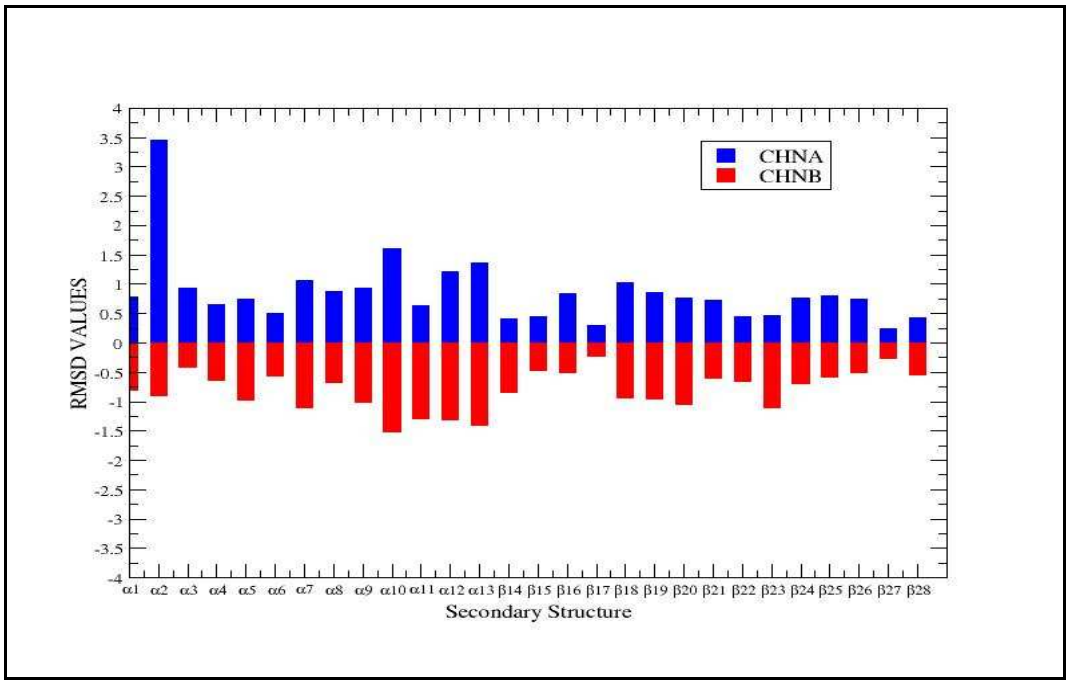


Figure 3.10 Rmsd (root mean square deviation) with α_{hqbmd} 1 kcal/mol/A^4 between the final and initial geometries of each secondary structure, as the opening is at the upper part of the protein along the gate1.

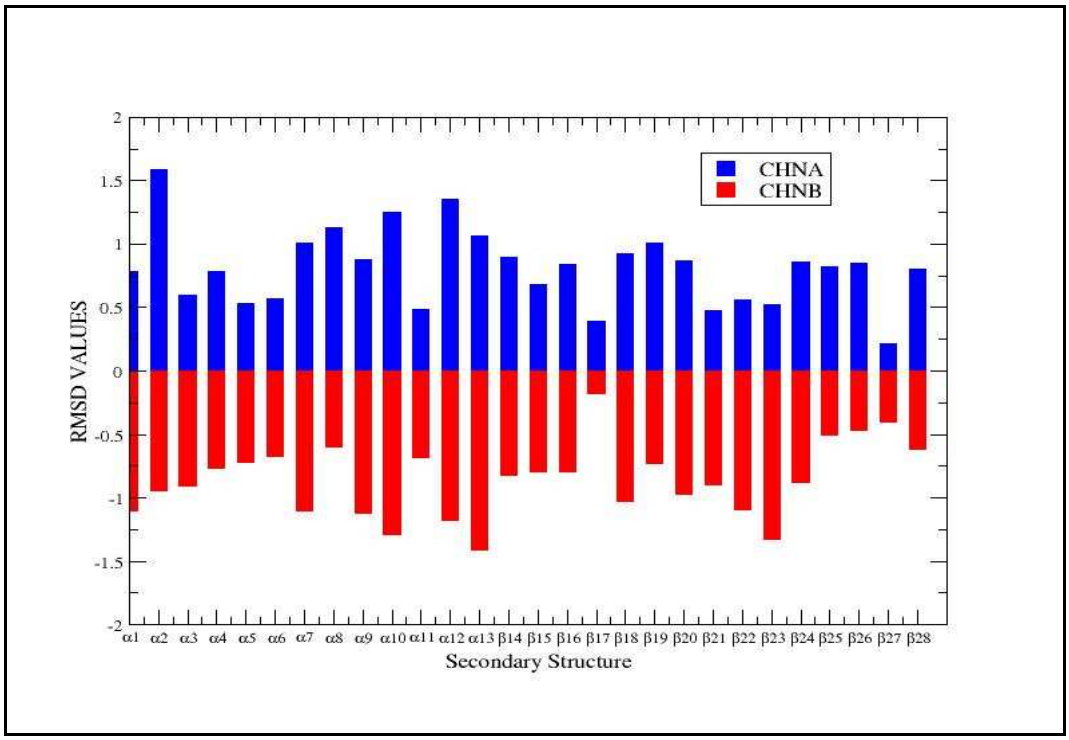


Figure 3.11 Rmsd (root mean square deviation) with α_{hqbmd} 1 kcal/mol/A^4 between the final and initial geometries of each secondary structure, as the opening is at the upper part of the protein along the gate2.

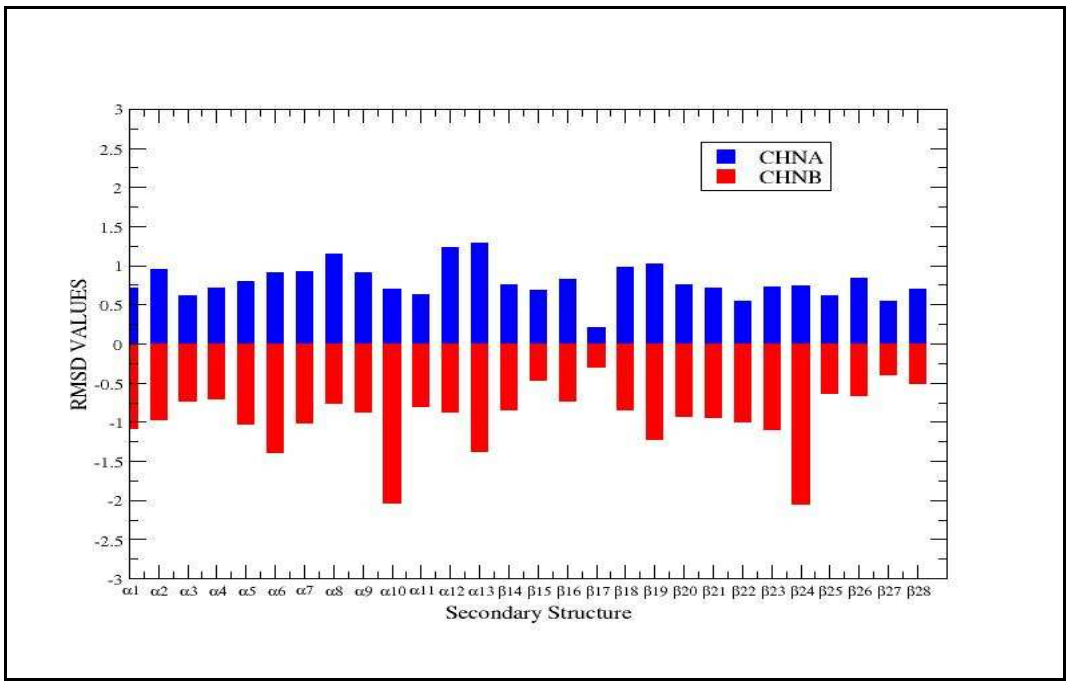


Figure 3.12 Rmsd (root mean square deviation) with α_{hqbmd} 1 kcal/mol/A^4 between the final and initial geometries of each secondary structure, as the opening is at the upper part of the protein along the gate3.

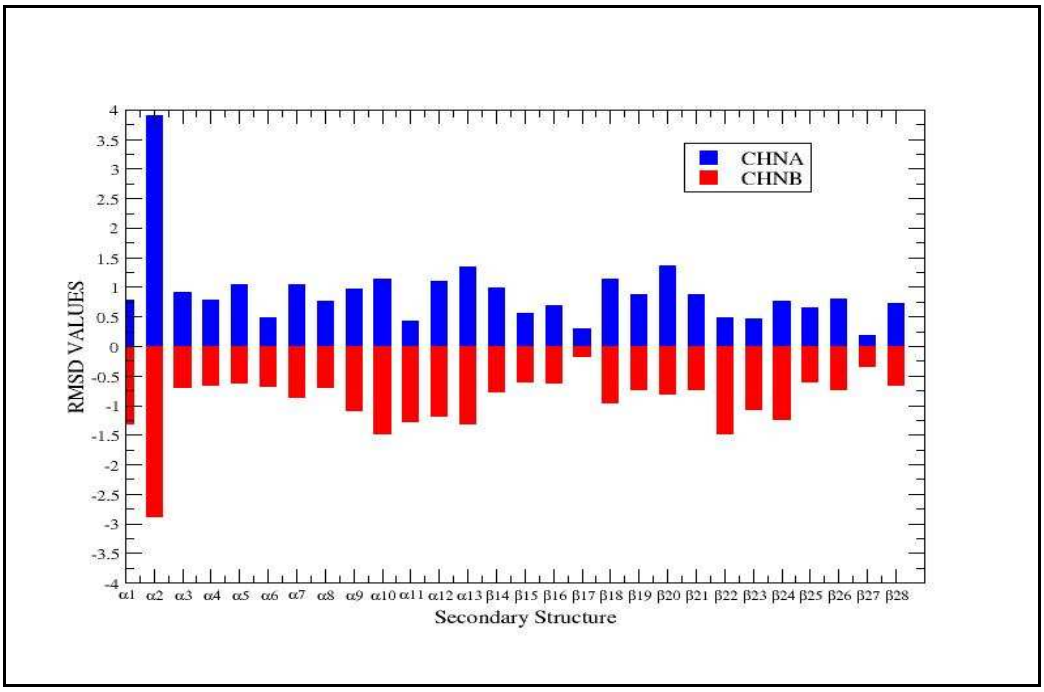


Figure 3.13 Rmsd (root mean square deviation) with α_{hqbmd} 10 kcal/mol/A^4 between the final and initial geometries of each secondary structure, as the opening is at the upper part of the protein along the gate1.

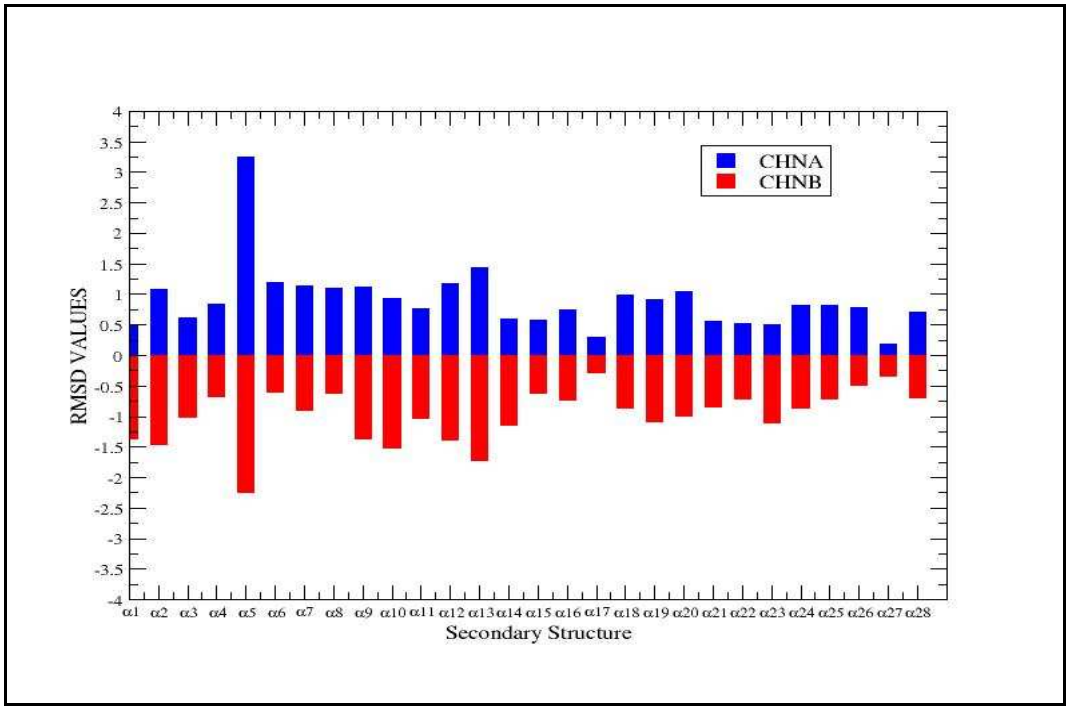


Figure 3.14 Rmsd (root mean square deviation) with α_{hqbmd} 10 kcal/mol/A^4 between the final and initial geometries of each secondary structure, as the openning is at the upper part of the protein along the gate2.

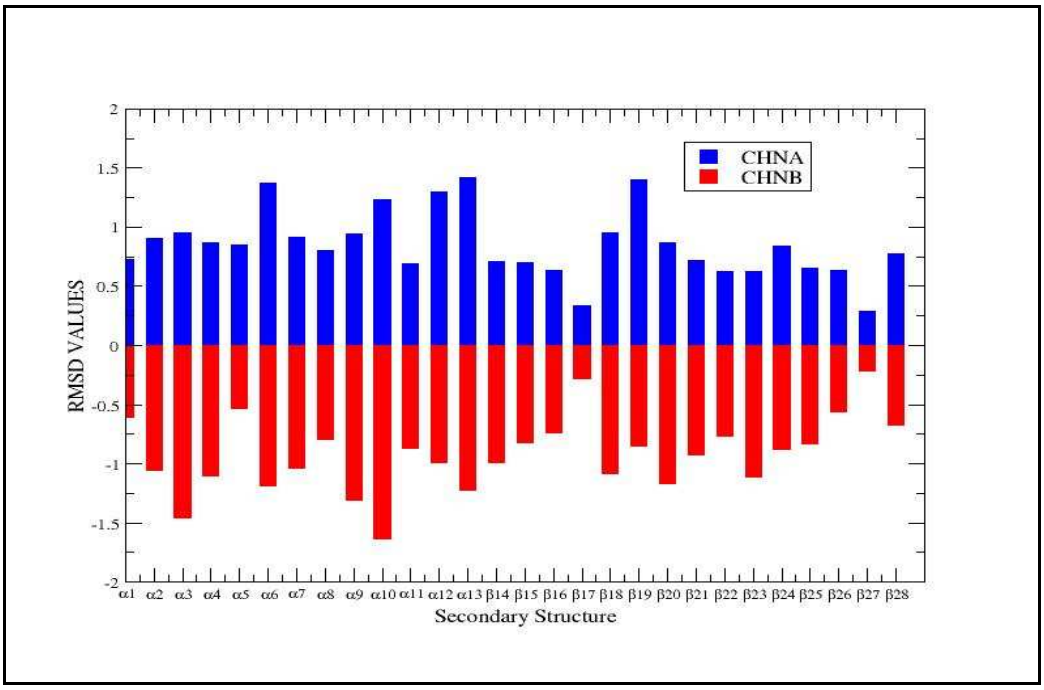


Figure 3.15 Rmsd (root mean square deviation) with α_{hqbmd} 10 kcal/mol/A^4 between the final and initial geometries of each secondary structure, as the openning is at the upper part of the protein along the gate3.

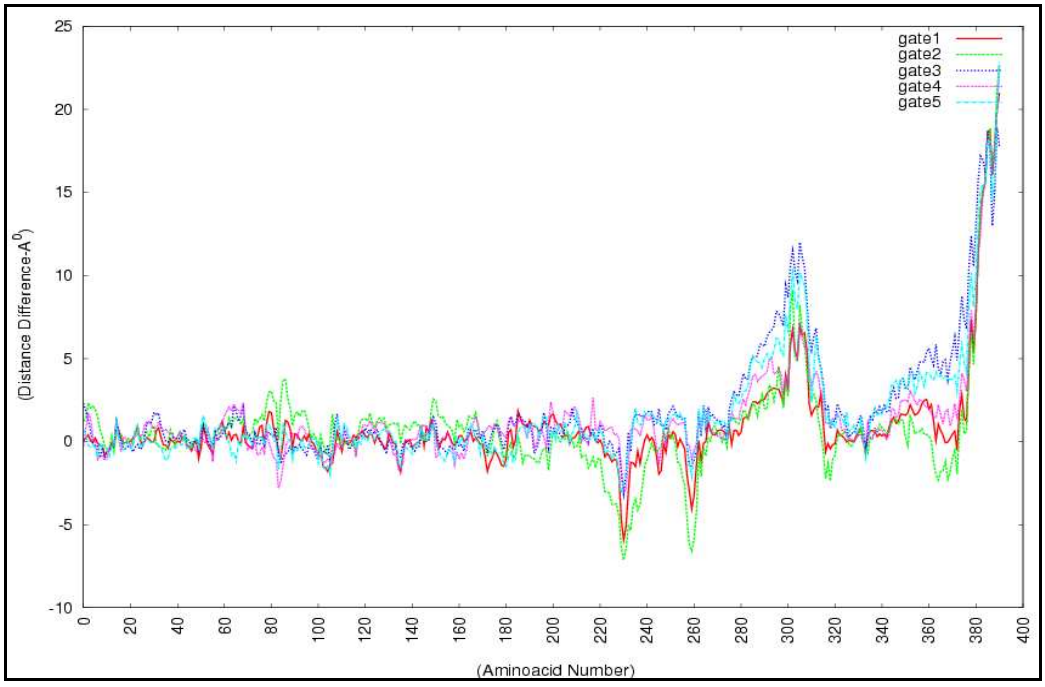


Figure 3.16 The separation between the same amino acids of the two chains at lower part of the protein, as the opening is along gate1 (red), gate2(green), gate3(blue), gate4(purple), and gate5(cyan).

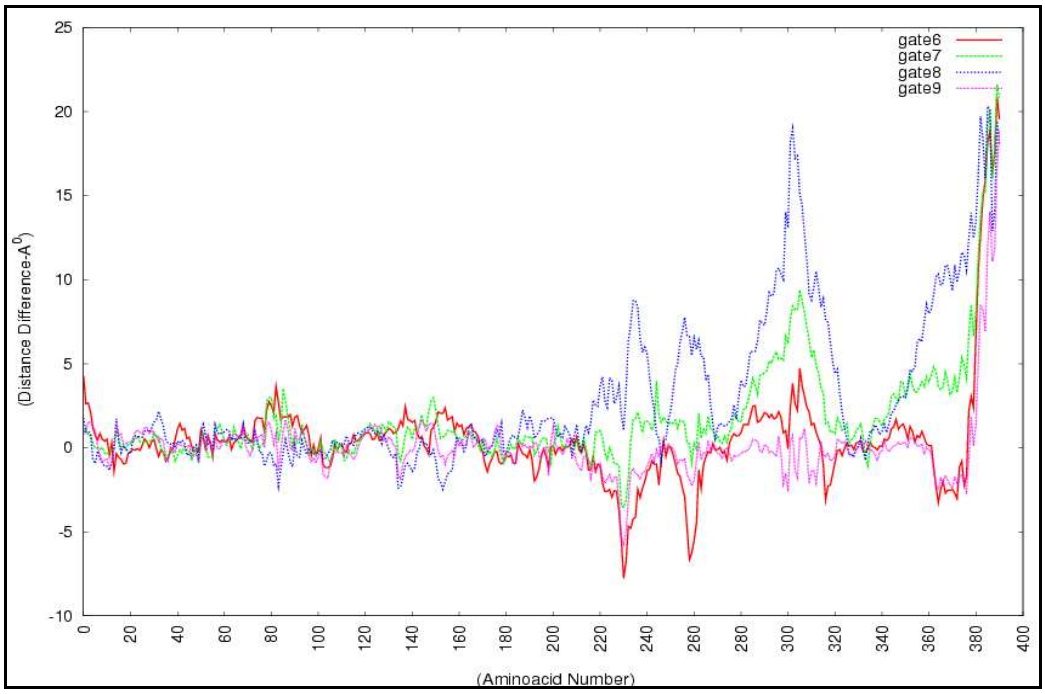


Figure 3.17 The separation between the same amino acids of the two chains at lower part of the protein, as the opening is along gate6 (red), gate7 (green), gate8 (blue), and gate9 (purple).

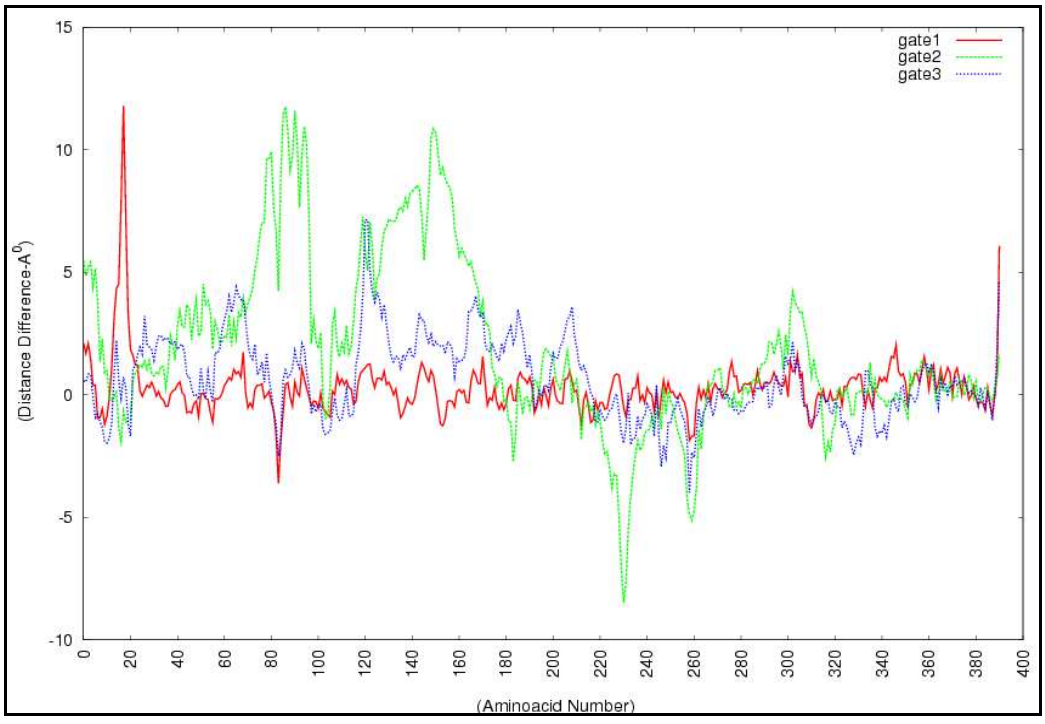


Figure 3.18 The separation between the same amino acids of the two chains at upper part of the protein with α_{hqbmd} 1 kcal/mol/A^4, as the opening is along gate1 (red), gate2 (green), and gate3 (blue).

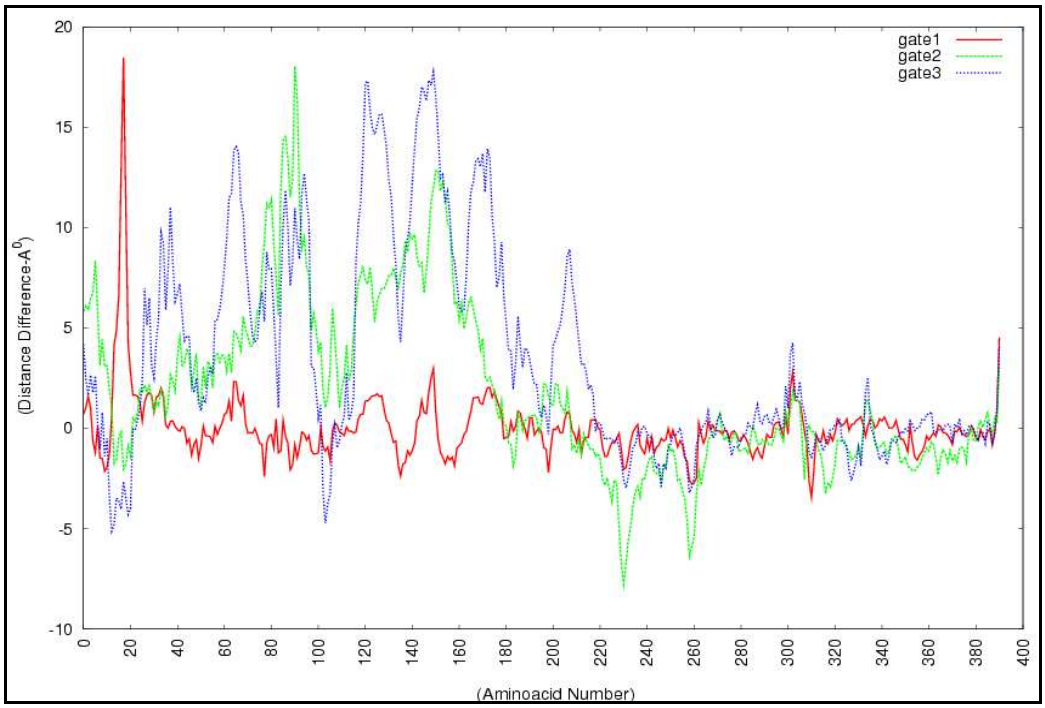


Figure 3.19 The separation between the same amino acids of the two chains at upper part of the protein with α_{hqbmd} 10 kcal/mol/A^4, as the opening is along gate1 (red), gate2 (green), and gate3 (blue).

3.2 ENERGETIC RESULTS

To compliment the structural analysis, we have also analyzed the simulations energetically. The first thing that we looked at is how the interaction energy between the two chain changes as the gates are opening. Therefore, we have presented the interaction energies, which are calculated to be the total non bonded interactions, in figure 3.20 for gate1, gate2, gate3, gate4, and gate5, and in figure 3.21 for gate6, gate7, gate8, and gate9. The same analysis is also done for the upper openings, and the results are given in figure 3.22 and in figure 3.23.

We have additionally looked at the amount of external potential energy that is applied to bring the desired changes in the protein. This will tell us something about the relative easiness for openings through different structural paths. The results we obtained are given in figure 3.24 and figure 3.25 for lower part openings, while they are given in figure 3.26 and 3.27 for the upper openings.

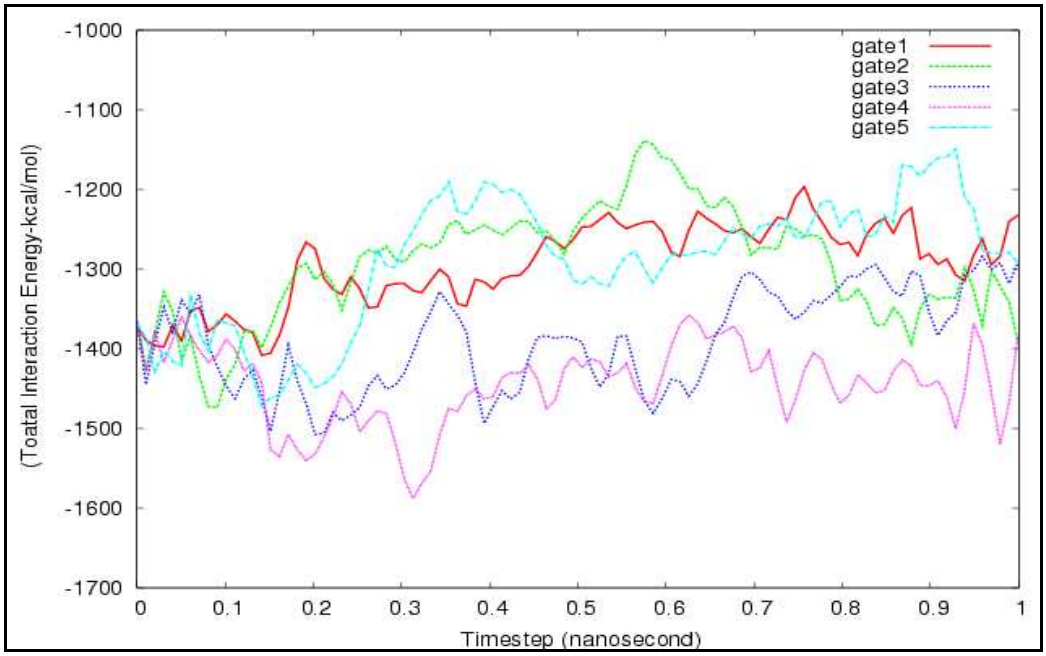


Figure 3.20 Total interaction energy (Electrostatic + Van Deer Walls) between the two chains at lower part of the protein, as the opening is along gate1 (red), gate2 (green), gate3 (blue), gate4 (purple), and gate5 (cyan).

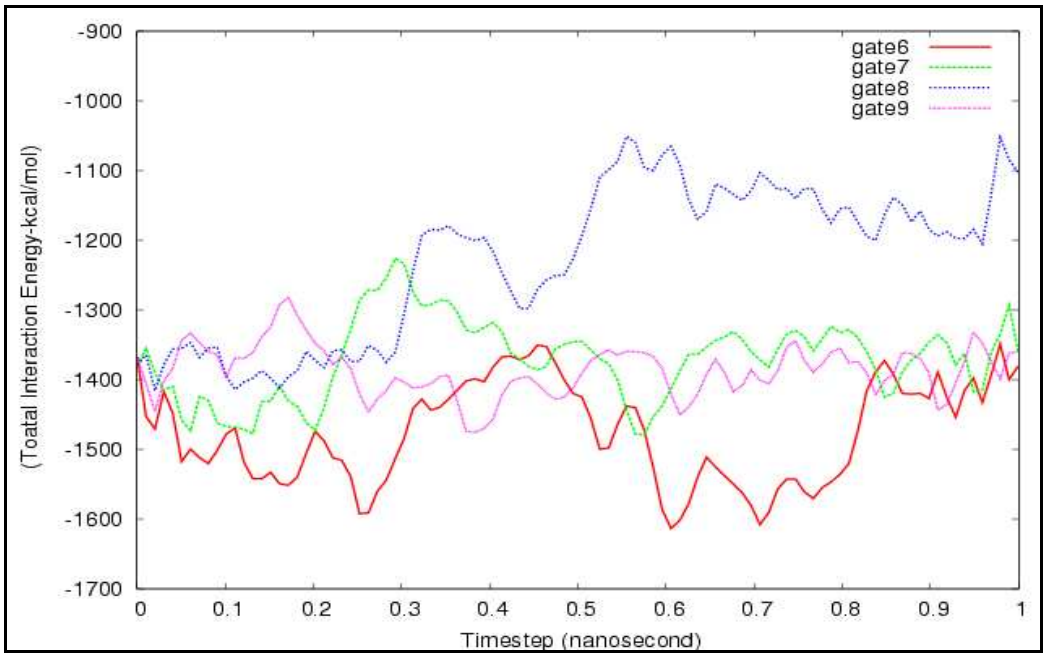


Figure 3.21 Total interaction energy (Electrostatic + Van Deer Walls) between the two chains at lower part of the protein, as the opening is along gate6 (red), gate7 (green), gate8 (blue), and gate9 (cyan).

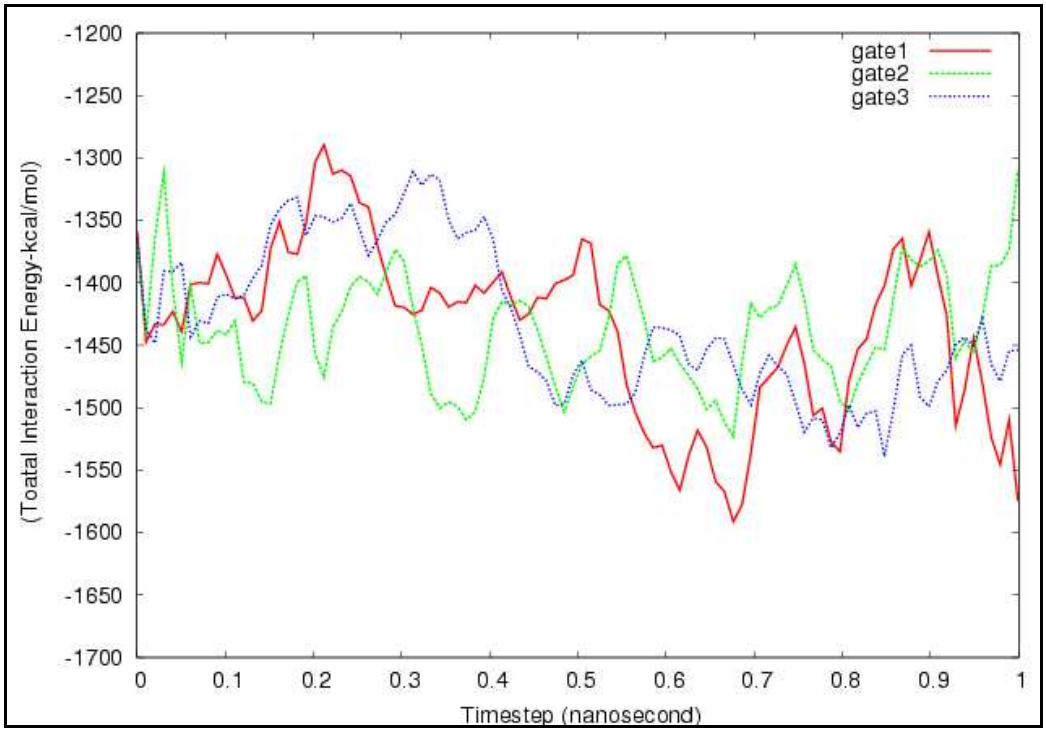


Figure 3.22 Total interaction energy (Electrostatic + Van Deer Walls) between the two chains at upper part of the protein with α_{hqbmd} 1 kcal/mol/A^4, as the opening is along gate1 (red), gate2 (green), and gate3 (cyan).

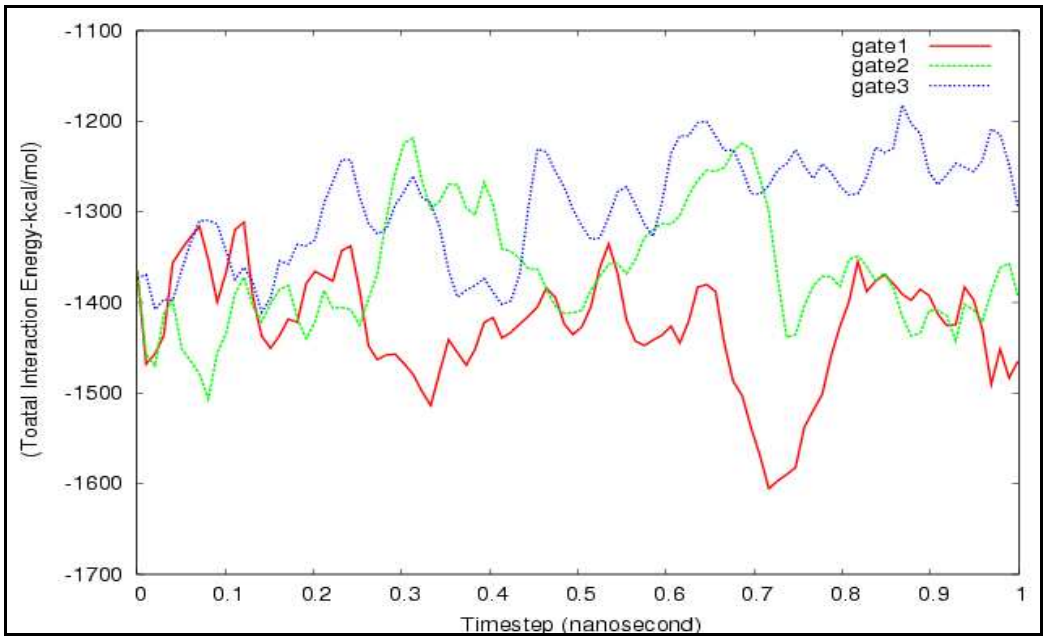


Figure 3.23 Total interaction energy (Electrostatic + Van Deer Walls) between the two chains at upper part of the protein with α_{hqbmd} 10 kcal/mol/A^4, as the opening is along gate1 (red), gate2 (green), and gate3 (cyan).

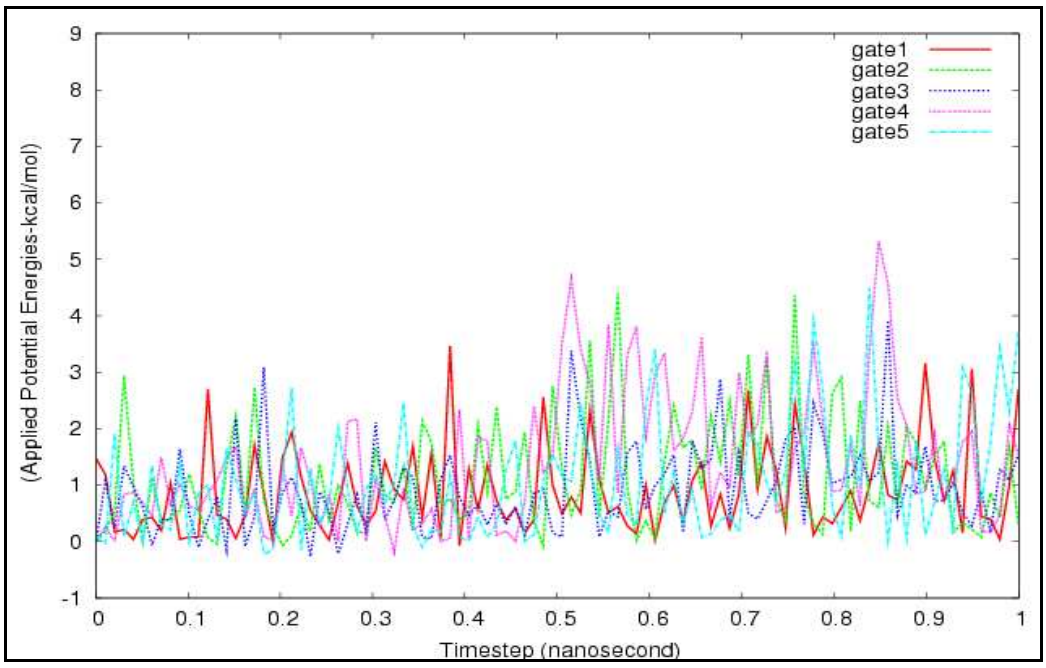


Figure 3.24 External potential that we applied as a function of time. The opening is at the lower part of the protein, along gate1 (red), gate2 (green), gate3 (blue), gate4 (purple), and gate5 (cyan).

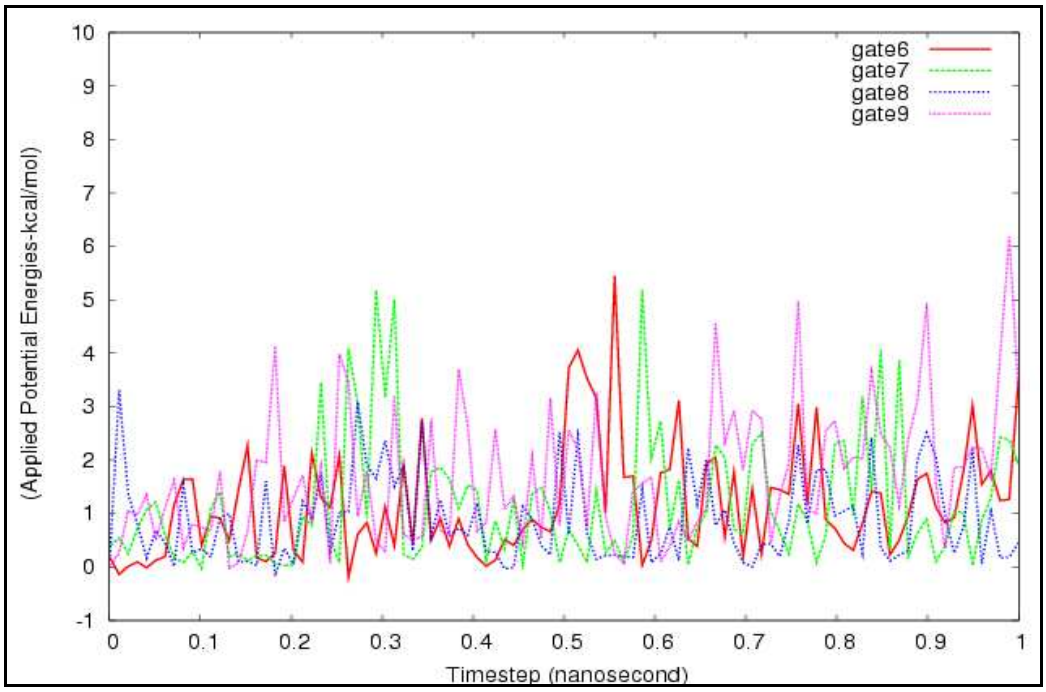


Figure 3.25 External potential that we applied as a function of time. The opening is at the lower part of the protein, along gate6 (red), gate7 (green), gate8 (blue), and gate9 (cyan).

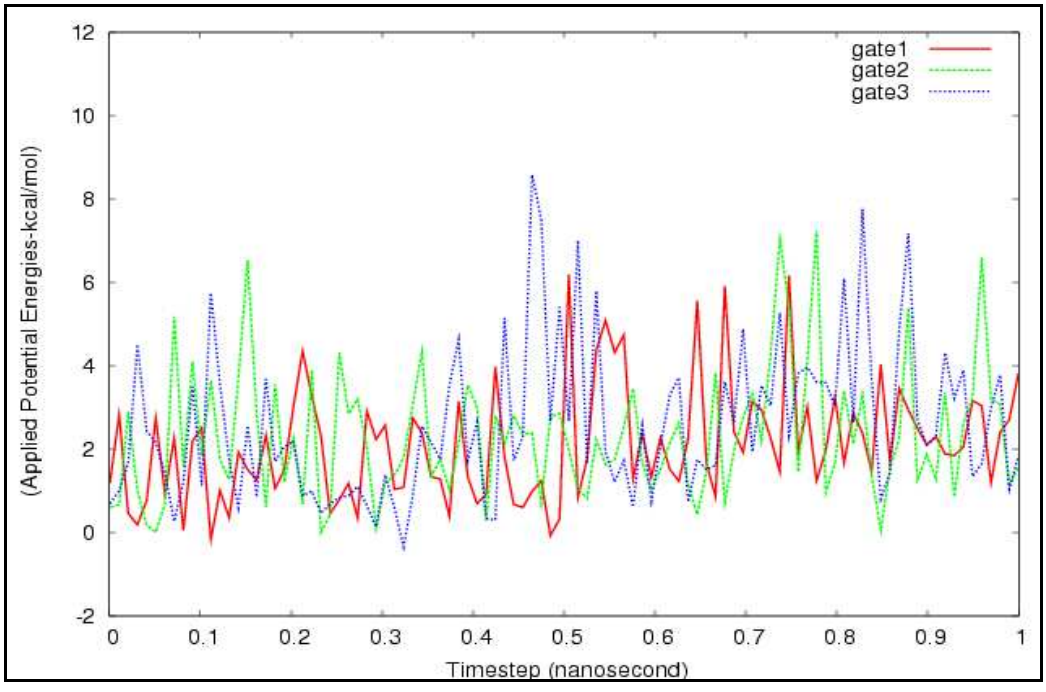


Figure 3.26 External potential that we applied as a function of time. The opening is at the upper part of the protein with α_{hqbmd} 1 kcal/mol/A^4, along gate1 (red), gate2 (green), and gate3 (cyan).

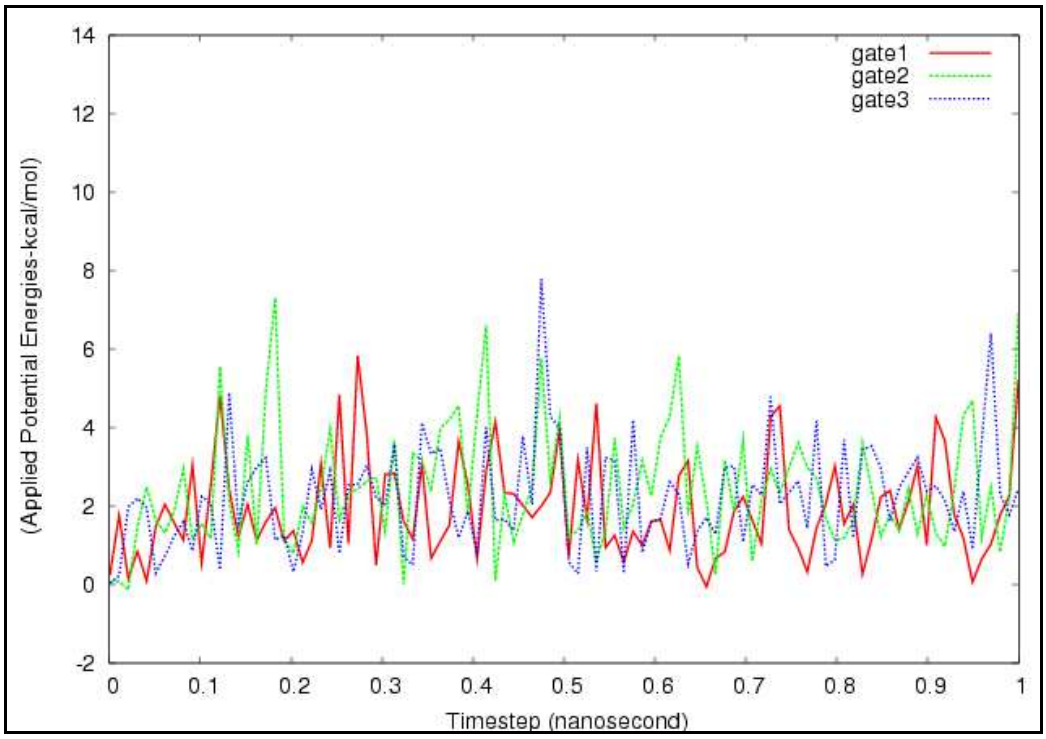


Figure 3.27 External potential that we applied as a function of time. The opening is at the upper part of the protein with α_{hqbmd} 10 kcal/mol/A^4, along gate1 (red), gate2 (green), and gate3 (cyan).

CHAPTER 4

RESULT AND DISCUSSION

We will discuss the structural and energetic results, considering each single amino acid, secondary structure of the protein, as well as the domain motions. Therefore, we have tabulated the secondary structures as we get them from the crystal structure (pdb id: 1EI1). The table 4.1 lists these secondary structures.

Table 4.1 All residue numbers of alpha and beta subunits for the secondary sturucture of the protein

alpha subunits	residue number	beta subunits	residue number
alpha1	5-9	beta1	10-11
alpha2	15-21	beta2	57-63
alpha3	22-27	beta3	67-73
alpha4	32-52	beta4	99-100
alpha5	88-96	beta5	126-135
alpha6	117-125	beta6	138-145
alpha7	182-197	beta7	148-158
alpha8	219-231	beta8	163-170
alpha9	278-299	beta9	201-206
alpha10	300-306	beta10	212-215
alpha11	310-315	beta11	240-246
alpha12	341-363	beta12	249-258
alpha13	363-389	beta13	263-269
		beta14	272-273
		beta15	318-325

The RMSD analyses show that the most deformations are seen in α13 for the openings at the lower gate. This is normal as we apply the external potential to this subunit. Although not perfect, there is symmetry between the RMSDs of the chainA and those of the chainB, as we expect since the two chains are identical in amino acid sequences. As we compare RMSDs corresponding to openings through different gates, we see a clear distinction for gate8, in which the reaction coordinate for opening is rotated by 140 degree. In this set up, we see that α10 has also similar RMSDs to α13, where the amount of deformations is around 2.75 A^0 (see the figure 3.8). For all of the other secondary structures, RMSDs are less than 1.5 A^0, which can be considered normal under the thermal motion at T=300 Kelvin. Therefore, gate8 produces extra

deformation in α10 (see the figure 3.8) while the other systems keep this secondary structure almost un-deformed.

In the upper gate openings, when the force constant of the external potential is 1 kcal/mol/A^4, the only large deformation is seen to be in α2 (as seen in figure 3.10), which around 3.5 A. For the openings along gate2 and gate3, we do not see any appreciable deformation in the secondary structures. However, when we increase the force constant of the external potential to 10 kcal/mol/A^4, we see that the openings through gate1 and gate2 produces large deformations in the units that we apply the potential (α2 in gate2 and α5 in gate3, see the figure 3.13 and 3.14). It is interesting that we do not see any appreciable deformations at all in opening along gate3 (see the figure 3.15).

We analyzed the distance difference of each amino acid of both ChainA and ChainB as the gates are opening at both the lower and the upper parts of the protein. Figure 3.16 and 3.17 show the separation of each amino acid in the openings for nine different gates at the lower side. For the opening of the lower gate, we applied the external potential to the amino acids from 382 to 390. The main target of these openings at the lower part of the protein is to get 20 A^0 separation between two helices of ChainA and ChainB. In our analysis, we reached this purpose for the lower part (see the figures 3.16-3.17). Unexpectedly, we also see that the distance between the two α10 (which includes nearly 10 amino acids between 300-310) increases about 12-13 A^0, and two α8 subunits gets close to each other by about 7-8 A^0 (see figure 3.16 and 3.17). Other than these subunits, the loop between β12 and β13 shows about separation or closures of about 5-6 A^0, which can be attributed to the free motion of the loop. The striking point here is that we see a correlated motion of α10 with α13 in *all* simulation set-ups. This correlated motion may indicate that these two subunits have correlated functionalities in the overall mechanism.

In the upper gate openings, when the force constant of the external potential is 1 kcal/mol/A^4, figure 3.18 shows the distance differences for the gate1, gate2, gate3. For the gate1, it is clearly seen that there is 12 A^0 separation between two helices which include the amino acids of α2. This is expected as we apply the external potential to the α2. For gate1, we do not see any other apriciable amount of opening. For the gate2, there is nearly 10 A^0 separation for the amino acid sequences which are located between

80-100 and 140-160. Here, although we push just only the α5, it is seen that β5, β6, and β7 also get separated by about 8-10 A^0. There is no considerable separation for the gate3, when the force constant is 1 kcal/mol/A^4. There is approximately 10 A^0 separation only around the amino acid at 120 (see the figure 3.18).

On the other hand, when we increase the force constant of the external potential to 10 kcal/mol/A^4, we clearly see quite much separations between the amino acids of the two chains. In all three gates, we see separations around 15-20 A^0. Through gate1, the separation is around α2. For gate2, there is a normal separation between the two α5, because of the external potential that we applied. Also, we see that there are separations for the β5, β6, and β7. But, for the gate3, although we push just only the α6, it is seen that α4, β3, β4, β6, β7, β8 and α8 also get separated by about 8-18 A^0. Therefore, we can say that there is an appreciable separation for all amino acid sequence located between amino acid numbers 20 and 200, along the gate3 (see the figure 3.19).

It is important to note that, in gate3 opening in the upper side of the protein, if the force constant of the external potential is 1 kcal/mole/A^4 the separation between the two chains is not observed. But when we shift the force constant to 10 kcal/mol/A^4, wee see a clear separation. This means that we have a potential barrier along the gate3, and the small force canst is not enough to pass through it. To further analyze this interesting situation, we have calculated the forces that are applied (from the derivative of the potential). We see the barrier is somewhere at the beginning of the reaction coordinate. As seen in figure 4.1, the applied force (when the force constant is 10 kcal/mole/A^4) is about two times larger than that of the other case where the force constant is 1 kcal/mole/A^4.

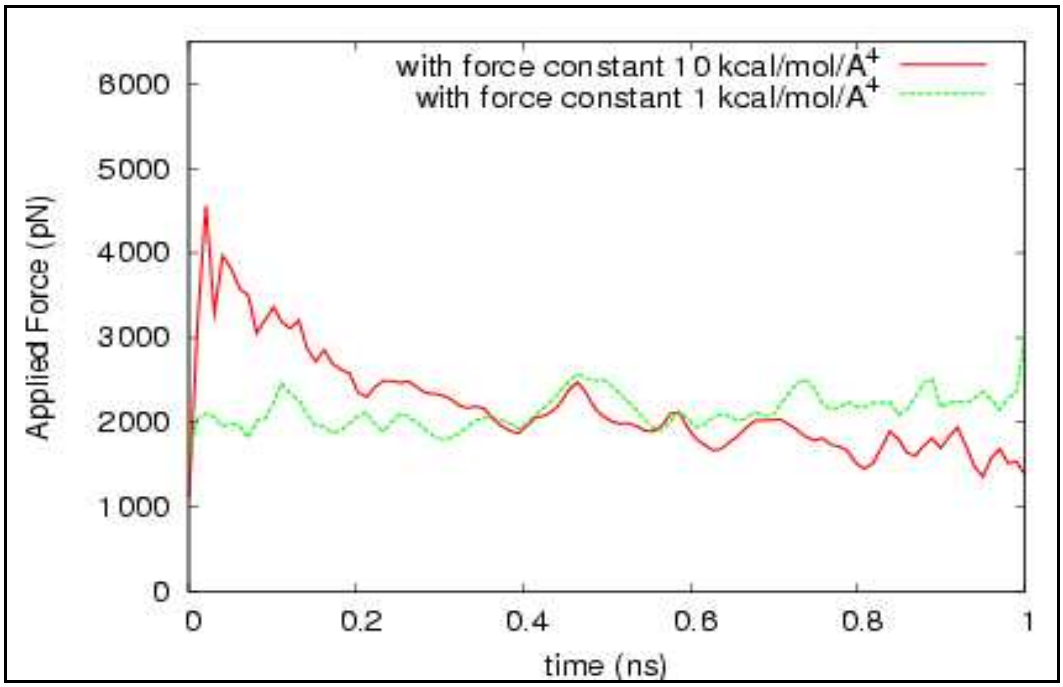


Figure 4.1 Applied forces with the force constant 1 kcal/mole/A^4 and with the force constant 1 kcal/mole/A^4

In the third part of the structural analysis, we have analyzed the rotations of ChainA, ChainB and overall protein for the different gate openings at the lower and the upper parts of the protein.

In the lower gate openings, we have nine different openings. In regards to the protein rotation, it is important to note that the overall rigid rotation of the system that is caused by our way of applying the external potential needs to be eliminated from the observed rotations. When we do this, we have observed that gate6 (100^0 rotation) corresponds physically a -80 degree rotation. Similarly, gate7 corresponds to -60 degree rotation, gate8 corresponds to -40 degree, and gate9 corresponds to -20 degree rotation.

The largest rotation that we see is 39.65 degree that happens to be along the gate8. The amount of rotations, both in the chains and in the overall protein, is given in table 4.2. Because, it is proposed in the literature that there should be 40 degree rotation of the protein for the passage of T-segment DNA through the G-segment, once the DNA is captured. Therefore, when we investigate the rotations of ChainA, ChainB and overall protein, we see that the value of rotation along gate8 is perfectly matches that proposed in the literature[41]. The protein rotations at each gate are shown in Figure 4.2. As seen

in the figure, we have two maximums, one for the positive rotations and one for the negative rotations. The maximum value of the positive rotations is seen to be 21.97 degree that belongs to opening along gate3.

Table 4.2 Rotations of ChainA, ChainB, and overall protein for the nine different gate openings at the lower part of protein. Rotations are obtained by superimposing final geometry onto initial one, as the total rigid rotation of the system is removed.

	CHNA	***CHNB***	***PROTEIN***
GATE1	**15.21**	**13.05**	**15.98**
GATE2	**2.51**	**4.83**	**6.07**
GATE3	**18.60**	**19.72**	**21.97**
GATE4	**13.77**	**13.59**	**15.53**
GATE5	**10.67**	**10.71**	**13.63**
GATE6	**-4.94**	**-5.19**	**-2.39**
GATE7	**22.36**	**20.18**	**23.95**
GATE8	**37.86**	**37.19**	**39.65**
GATE9	**26.73**	**27.77**	**27.39**

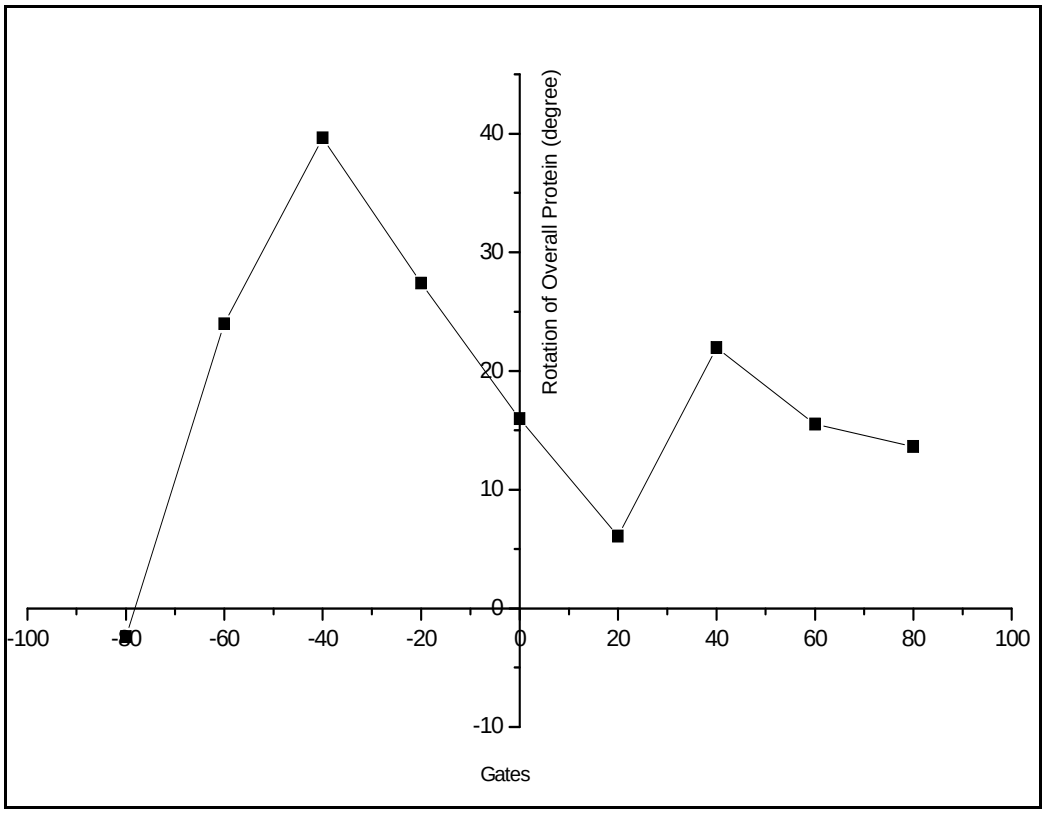


Figure 4.2 The protein rotations at each gate of lower openings

In the upper gate openings, when the force constant of the external potential is 1 kcal/mol/A^4, all rotations for different gates are changing between 8-10 degrees (see the table 4.3). For the gate1 and gate3, the rotation values which are 8.6 and 8.7 degrees respectively, of overall protein are nearly the same. The rotation of overall protein is 7.3 degree for the gate2. When the force constant of the external potential is 10 kcal/mol/A^4, we observe that the rotation of overall protein is changing between 5-8 degrees for the gate1 and gate2. But, we do not see appreciable rotation for the gate3. When we only look the rotations of ChainA and ChainB; we can easily say that rotation values are changing between 5-12 degrees (see the table 4.4). In brief, although we see protein rotations around 30-40 degrees for the lower gate openings, we see only about 8-9 degree rotations for the upper gate openings.

Table 4.3 Rotations of ChainA, ChainB, and overall protein for the three different gate openings at the upper part of protein with α_{hqbmd} 1 kcal/mol/A^4. Rotations are obtained by superimposing final geometry onto initial one, as the total rigid rotation of the system is removed.

	CHNA	***CHNB***	***PROTEIN***
GATE1	**8.54**	**8.58**	**8.60**
GATE2	**10.13**	**10.11**	**7.34**
GATE3	**8.69**	**9.09**	**8.66**

Table 4.4 Rotations of ChainA, ChainB, and overall protein for the three different gate openings at the upper part of protein with α_{hqbmd} 10 kcal/mol/A^4. Rotations are obtained by superimposing final geometry onto initial one, as the total rigid rotation of the system is removed.

	CHNA	***CHNB***	***PROTEIN***
GATE1	**5.96**	**5.83**	**5.15**
GATE2	**11.80**	**8.81**	**7.47**
GATE3	**8.18**	**8.51**	**0.68**

Also, we analyzed the rotation of the lower part of the protein with respect to the upper part of the protein (see tables 4.5, 4.6, 4.7). For the lower gate openings, the biggest value of rotation difference is around 7 degree for the gate2, gate6 and gate9. But, this difference is lower than 1 degree for the gate8 (see the table 4.5). Overall, we can say that we do not observe any appreciable amount of relative rotation of the lower parts of the protein with respect to the upper part, as we have a maximum value of 10.29 degree rotation.

Table 4.5 Rotations differences of upper and lower parts of protein for the nine different gate openings at the lower part of protein. Rotations differences are obtained by superimposing final geometry onto initial one, as the total rigid rotation of the system is removed.

	UPPER	***LOWER***	***DIFFERENCE***
GATE1	**15.76**	**20.30**	**4.54**
GATE2	**5.76**	**12.87**	**7.11**
GATE3	**21.82**	**24.99**	**3.17**
GATE4	**15.39**	**18.12**	**2.73**
GATE5	**13.38**	**18.21**	**4.83**
GATE6	**-2.72**	**4.89**	**7.61**
GATE7	**23.87**	**25.59**	**1.72**
GATE8	**39.69**	**38.87**	**-0.82**
GATE9	**27.08**	**34.55**	**7.47**

Table 4.6 Rotations differences of upper and lower parts of protein for the three different gate openings at the upper part of protein with α_{hqbmd} 1 kcal/mol/A^4. Rotations differences are obtained by superimposing final geometry onto initial one, as the total rigid rotation of the system is removed.

	UPPER	***LOWER***	***DIFFERENCE***
GATE1	**8.55**	**10.64**	**2.09**
GATE2	**7.16**	**18.31**	**11.15**
GATE3	**8.77**	**9.53**	**0.76**

Table 4.7 Rotations differences of upper and lower parts of protein for the three different gate openings at the upper part of protein with α_{hqbmd} 10 kcal/mol/A^4. Rotations differences are obtained by superimposing final geometry onto initial one, as the total rigid rotation of the system is removed.

	UPPER	***LOWER***	***DIFFERENCE***
GATE1	**5.01**	**9.67**	**4.66**
GATE2	**7.28**	**15.96**	**8.68**
GATE3	**0.41**	**10.70**	**10.29**

In the upper gate openings, we have observed that the gate2 has the biggest rotation difference between the upper and the lower parts of the protein for the force constant of the external potential is 1 kcal/mol/A^4. For the gate3, this value is lower than 1 degree (see the table 4.6). When we increase the force constant of the external potential to 10 kcal/mol/A^4, we do not see any appreciable change of rotation differences for the gate1 and the gate2. But, for the gate3, this rotation difference is 10.29 degree as it is 0.76 degree for the force constant 1 kcal/mol/A^4.

In the energetic results, we looked how the total energy (Electrostatic + Van Deer Walls) between two chains changes as the gates are opening. Energy analysis shows us that the interaction energies are increasing as a function of time. Physically, this means that the electrostatic interaction between atoms is decreasing and the distance between these atoms get larger. Especially, we observe that the most considerable increase of the interaction energy is for the gate8 among the lower gate openings (see the figure 3.21). The interaction energy changes between -1600 and -1200 kcal/mol for the lower gate openings except gate8. For the gate8, it reaches bigger values than -1200 kcal/mol (see the figures 3.20-3.21).

In the upper gate openings, when the force constant of the external potential is 1 kcal/mol/A^4, we can clearly say that the interaction energies are decreasing for all gates (see the figure 3.22). That is, all atoms get closer through tree gate openings. But, when we increase the force constant of the external potential to 10 kcal/mol/A^4, the interaction energy for gate3 is changing between -1400 and -1200 kcal/mol as the other

two gates have no considerable changes (see the figure 3.23). The only opening where we see the interaction energy between the chainA and chainB is decreasing as the gate opens up is the opening along gate8. This observation confirms the fact that gate8 is the most probable opening scheme, as we must see such decrease in the interaction energy once the two chains get separated.

Also, we analyzed the external potentials that we applied as a function of time. Even if the external potentials have some peaks in time, the changes in potential for all gates are nearly the same. For the lower gate openings, the change of the external potentials is between 0-4 kcal/mol (see the figures 3.24-3.25). At the upper part of the protein, when the force constant of the external potential is 1 kcal/mol/A^4, the external potentials that we applied are changing between 0-8 kcal/mol (see figure 3.26). But, for the force constant 10 kcal/mol/A^4, the external potentials are between 0-7 kcal/mol (see the figure 3.27).

CHAPTER 5

CONCLUSIONS

In conclusion, we find out the following main results out of our study.

a-) The most striking information we got is that protein rotates in the counter-clockwise (that is right-handed) in almost all simulations. This is very nicely in agreement with the fact that DNAGyrase introduces only negative supercoils and to do this it needs to rotate counter-clockwise. This selectivity of the rotation is quite interesting, we observe that in our simulations.

b-) Our energetic and structural results show that the most feasiable opening of the protein is along the gate8, at the lower side of the enzyme.

c-) We do not get any opening at upper side of the protein as much as the diameter of the DNA (20 A) in any gate opening. This means that the protein needs to form the non-covalent dimmer complex *after* it catches the T-segment DNA.

d-) We *do not* see any appreciable deformations at all in opening along gate3.

e-) We see a *correlated motion of α10 with α13* in all simulation set-ups. This correlated motion may indicate that these two subunits have correlated functionalities in the overall mechanism.

f-) For the gate3, although we push just only the α6, it is seen that α4, β3, β4, β6, β7, β8 and α8 also get separated by about 8-18 A^0. Therefore, we can say that there is an appreciable separation for all amino acid sequence located between amino acid numbers 20 and 200, along the gate3.

g-) We see a potential barrier in opening of the upper part of the protein. This barrier is somewhere at the beginning of the reaction coordinate.

h-) When we investigate the rotations of ChainA, ChainB and overall protein, we see that the value of rotation along gate8, as it is the best opening, perfectly matches that proposed in the literature.

i-) Although we have observed protein rotations around 30-40 degrees for the lower gate openings, we found out only about 8-9 degree rotations for the upper gate openings.

j-) Overall, we can say that we do not observe any appreciable amount of relative rotation of the lower parts of the protein with respect to the upper part, as we have a maximum value of 10.29 degree rotation.

k-) The only opening where we see the interaction energy between the chainA and chainB is decreasing as the gate opens up is the opening along gate8. This observation confirms the fact that gate8 is the most probable opening scheme, as we must see such decrease in the interaction energy once the two chains get separated.

REFERENCES

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

[2] D.L.Nelson, M. M. Cox, *Principles of Biochemistry*

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

[4] L. Postow, N. J. Crisona, B. J. Peter, C. D. Hardy, N. R. Cozarelli, *PNAS* 2001, 98- 8219

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

[6] J. C. Wang, *Nature Reviews* 2002, 3- 430

[7] M. R. Redinbo, L. Stewart, J. J. Champoux, W. G. J. Hol, *J. Mol. Biol.* 1999, 292 -685

[8] M. R. Redinbo, L. Stewart, P. Kuhn, J. J. Champoux, W. G. J. Hol, *Science* 1998, 279-1504

[9] S.E. Mirski, J.H. Gerlach, S.P. Cole, *Exp. Cell Res.* 1999, 251: 329-39

[10] L. Bjergbaek, P. Kingma, I.S. Nielsen, Y. Wang, O. Westergaard et al., *J. Biol Chem.* 2000, 275:13041-48

[11] J.C. Wang, *Q.Rev. Biophys.* 1998, 31: 107- 44

[12] J. J. Champoux, *Biochemistry* 2001, 70: 369- 413

[13] D.B. Wigley, G.J. Davies, E.J. Dodson, A. Maxwell, G. Dodson, *Nature* 1991, 351: 624-29

[14] K. Miwzchi, M.H. O'Dea, M. Gellert, *Proc. Natl. Acad. Sci. USA.* 1978, 5960-2963

[15] M. Gellert, M.H. O'Dea, T. Itoh, J. Tomizowa, Proc. *Natl. Acad. Sci. USA.* 1976, 4474-4478

[16] M. Nollmann, et al. *Biochemie* 2007, 490-499

[17] L.F. Liu, J.C. Wang, *Cell* 1978, 979- 984

[18] J. Gore, Z. Bryant, M.D. Stone, M. Nolmann, N.R. Cozzarelli, C. Bustamante, *Nature* 2006, 100-104

[19] N.L. Williams, A.J. Howells, A. Maxwell, *J. Mol. Biol.* 2001, 969- 984

[20] K.E. Scheirer, N.P. Higgins, *J. Biol. Chem.*1997, 27202- 27209

[21] J.G. Heddle, S. Mitelheiser, A. Maxwell, N.H. Thomson, *J. Mol. Biol.* 2004, 597- 610

[22] P.O. Brown, N.R. Cozzarelli, *Science* 1979, 1081-1083

[23] K.N. Kreuzer, N.R. Cozzarelli, *Cell* 1980, 245- 254

[24] C. Levine, H. Hiasa, K.J. Marians, *Biochim. Biophys. Acta* 1998, 29-43

[25] D.A. Burten, N. Osheroff, *Biochim. Biophys. Acta* 1998, 139-154

[26] D.D. Von Hoff, *Semin. Oncol.* 1998, 31-36

[27] S. Chang, T. Hu, T.S. Hsieh, *J. Biol. Chem.* 1998, 19822-19828

[28] V. Lamour, L. Hoermann, J.M. Jeltsch, P. Oudet, D. Moras, *Acta Crystalogr.* 2002, 1376-1378

[29] B. J. Alder, T. E. Wainwright, *J. Chem. Phys.* 1959, 31, 2, 459

[30] W. B. Streett, D. J. Tildesley, G. Saville, *Mol. Phys.*, 1978, 35, 3, 639

[31] A. McCammon, B. R. Gelin, M. Karplus, *Nature* 1977, 267- 585

[32] M. Karplus, J. A. McCammon, *Nature Structural Biology* 2002, 9- 646

[33] http://cmm.info.nih.gov/intro_simulation/node15.html

[34] MacKerell et al. *J. Phys. Chem. B.* , 1998, 102- 3586

[35] http://vit-embnet.unil.ch/MD_tutorial/

[36] http://www.charm.org

[37] M. Marchi, P. Ballone, *J. Chem. Phys*., 1999, 110, 8- 3697

[38] P. H. Hünenberg, *Adv. Polymer. Sci.* 2005, 173- 105

[39] Anderson, *J. Chem. Phys*., 1980, 74, 2384

[40] B.R.Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comp. Chem.* 1983, 4- 187

[41] L.Brino, A. Urzhumtsez, M. Mousli, C. Bronner, A. Mitschler, P. Oudet, D. Moras, *J. of Bio. Chem.* 1999, 9468-9475