Neslihan Üçüncüoğlu

M.S. Thesis In Physics

June - 2009

HUMAN DNA TOPOISOMERASE I IN COMPLEX WITH THE POISON TOPOTECAN AND COVALENTLY BONDED WITH A DNA DUPLEX

by

Neslihan Üçüncüoğlu

June 2009

HUMAN DNA TOPOISOMERASE I IN COMPLEX WITH THE POISON TOPOTECAN AND COVALENTLY BONDED WITH A DNA DUPLEX

by

Neslihan Üçüncüoğlu

A thesis submitted to

the Graduate Institute of Sciences and Engineering

of

Fatih University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Physics

June 2009
Istanbul, Turkey

APPROVAL PAGE

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

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

Assist. Prof. Dr. Sadık GÜNER ……..

Assist. Prof. Dr. Levent SARI ……..

Assist. Prof. Dr. Sevim IŞIK ……..

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.

Assoc. Prof. Dr. Nurullah ARSLAN

Director

June 2009

HUMAN DNA TOPOISOMERASE I IN COMPLEX WITH THE POISON TOPOTECAN AND COVALENTLY BONDED WITH A DNA DUPLEX

Neslihan Üçüncüoğlu

M. S. Thesis - Physics
June 2009

Supervisor: Assist. Prof. Dr. Levent SARI

ABSTRACT

Human topoisomerase I is a crucial enzyme for the cell regulating the topological state of DNA during a variety of chromosomal activity involving replication, transcription and recombination by nicking the DNA strand via transient covalent phosphotyrosine linkage. The antitumor compounds stabilize this transient complex and block further replication. *Topotecan* is a clinically confirmed drug is believed to inhibit the catalytic action by positioning in the cleavage site and acting like a DNA base pair. We modeled the relaxations of DNA in complex with human topoisomerase I and TPT molecules employing Molecular Dynamics method. We have investigated the inhibition mechanism of TPT by comparing the energetic and structural results of MD calculations of the system with and without TPT molecule. The rotation of DNA about the intact strand is achieved by half quadratic biased molecular dynamics (HQBM) which perturbs the system to accomplish the desired conformation, without affecting the constants of motion. We have conducted the same calculations for both positive and negative supercoil relaxations on two systems; with and without TPT. As a result we have found out several very important behavior of TPT molecule. The clearest one is that *TPT stops opening of the 'lips region' of the protein*, in both positive and negative supercoil relaxations. Another observation is that the DNA rotations destabilize under addition of TPT molecule, and move more towards the 'free-rotation' scheme. Also, we have observed that the effects of TPT on negative supercoil relaxation are more noticeable than the positive supercoil relaxations. It

is also worth noting that the drug molecule keeps its interactions with the protein unchanged during the DNA rotations.

Keywords: Human topoisomerase I, Cancer drugs, Topotecan, Molecular Dynamics Simulations, HQBMD.

DNA-HUMAN TOPO I KOMPLEKSİNİN KANSER İLAÇ MOLEKÜLÜ TPT İLE MEKANİZMASININ TEORİK İNCELEMESİ

Neslihan Üçüncüoğlu

Yüksek Lisans Tezi - Fizik
June 2009

Tez Yöneticisi: Yrd.Doç. Dr. Levent SARI

ÖZ

İnsan topoizomeraz I replikasyon, transkripsiyon ve DNA iplikçiklerinin birleştirilmesi gibi kromozom aktiviteleri sırasında DNA'nın ipliğini geçici kovalent fosfotirozin bağıyla keserek topolojisini düzenleyen, hücre için yaşamsal öneme sahip bir enzimdir. Tümörle savaşan ilaç molekülleri bu geçici kompleksin sabit kalmasını sağlar ve DNA'nın faaliyetlerini durdurur. Klinik testlerle etkinliği onaylanmış bir ilaç olan Topotekan'ın enzimin işlevini DNA'nın kesik kısmına yerleşip DNA baz çifti gibi görünerek durdurduğu düşünülüyor. Biz DNA molekülünün dönmesini insan topoizomeraz I ve TPT molekülü ile Moleküler Dinamik (MD) metodu kullanarak modelledik. Sistemin TPT eklendikten önceki ve sonraki durumunun MD hesaplamaları sonuçlarını kıyaslayarak TPT'nin inhibite etme mekanizmasını inceledik. DNA ipliğinin dönmesini sistemi istenilen konformasyona hareket sabitlerini etkilemeden ulaştıran bir yöntem olan Yarı Kuadratik Moleküler Dinamik (HQBM) ile sağladık. Aynı hesaplamaları hem pozitif hem de negatif süperkoil için TPT'siz ve TPT eklenmemiş iki sistemde de uyguladık. Hesplamalar sonucunda TPT molekülü ile ilgili önemli verilere ulaştık. En dikkat çekici olan TPT'nin hem pozitif hem de negatif süperkoil dönmelerinde proteinin *'dudak' kısmının açılmasını durdurması* oldu. Bir diğer gözlem ise DNA dönmelerinin TPT molekülü eklendiğinde hareketinin bozulmayıp *serbest dönmeye* daha yakın bir şekilde hareket etmesi oldu. Bunun yanında

TPT’nin negatif süperkoil dönmelerinde pozitif süperkoil dönmelerinden daha farkedilir bir etkisi olduğunu gözlemledik. Bir diğer kayda değer veri ise ilaç molekülünün DNA dönmeleri süresince protein ile olan etkileşiminin sabit kalması oldu.

Anahtar Kelimeler: İnsan topoizomeraz I, Kanser ilacı, Topotecan, Moleküler Dinamik Simulasyonları, HQBM

ACKNOWLEDGEMENT

Foremost, I would like to express my sincere gratitude to my supervisor Assist. Prof. Levent SARI. I would like to thank him for his patience, motivation, enthusiasm and immense knowledge and his guidance and support from the initial to the final level which enabled me to develop an understanding of the subject.

I also wish to thank Prof. Dr. Mustafa KUMRU for his contribution and his good-natured support.

I must acknowledge TUBITAK (The Scientific & Technological Research Council of Turkey) for offering us the funding through the project number 107T209 which 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”. This thesis would not have been possible unless we would not have the funding for the required hardware and software.

Many thanks go to all my friends including Azize Sevim, Tayyibe Bardakçı, Fatma Gözüak, Ahmet Menteş, Hüseyin Kavas, Türker Tunay and Doruk Yıldıztekin.

My parents, Meryem and Selim PARLAR deserve my highest respect and deepest gratitude for their constant support and encouragement in all the steps that I have taken over the years. My warm thanks go to my lovely sisters for always putting a smile on my face.

I owe my loving thanks to my husband, Süleyman Üçüncüoğlu for his endless love and accompaniment. I also thank to his family for their support and prayers.

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
topo I	:	Human topoisomerase I
MD	:	Molecular Dynamics
HQBMD	:	High Quality Biased Molecular Dynamics
Lk	:	Linking number
Lk_0	:	Linking number in relaxed state
Tw	:	Twist
Wr	:	Writhe
ΔLk	:	Linking difference
σ	:	Supercoiling density or specific linking difference
Tyr	:	Tyrosine
Lys	:	Lysine
Arg	:	Arginine
His	:	Histidine
CPT	:	Camptothecin
$\overset{0}{\text{A}}$	:	Angstrom
NMR	:	Nuclear Magnetic Resonance
k_b	:	Force constant for bond interaction energy
k_θ	:	Force constant for bond angle potential energy
k_Φ	:	Force constant for dihedral angle potential energy
k_ω	:	Force constant for bond improper potential energy
k_u	:	Force constant for Urey-Bradley potential energy

δ : Phase of dihedral

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

ε : Dielectric constant

α_{hqbmd} : Force constant for HQBM

ρ : Reaction coordinate

V : Volume

T : Temperature

N : Number of particles

HF : Hartree Fock

DFT : Density Functional Theory

H-K : Hohenberg Kohn

K-S : Kohn-Sham

LDA : Local Density Approximation

GGA : Generalized Gradient Approximation

B3LYP : Becke-3 term correlational functional with Lee-Yang-Parr exchange correlation functional

STO : Slater Type Orbital

CHARMM: Chemistry at HARvard Macromolecular Mechanics

VMD : Visual Molecular Dynamics

SP : Single Point

OPT : Geometry Optimization

FREQ : Vibration Frequaency

LSF : Load Sharing Facility

CMU : Cluster Management Utility

PSF : Protein Structure File

RTF : Residue Topology File

PARAM : Parameter File

PSF : Protein Structure File

CRD : Coordinate File

SD : Steepest Descent

ABNR	:	Adapted Basis Newton-Raphson
Na	:	Sodium
RMSD	:	Root mean square deviations

CHAPTER I

INTRODUCTION

1.1 DNA TOPOLOGY AND SUPERCOILING

Although the model that Watson and Crick proposed for the molecular structure of Deoxyribonucleic acid (DNA) was fairly consistent with what was already known in 1953, scientists of that time did not tend to accept it due to the inaccessibility of the bases from outside [1], [2]. DNA should preserve its topological state and organization of bases during the essential cellular processes. In that time, it seemed very difficult to prevent the adverse effects of these numerous processes to the helical superstructure of this highly dynamical molecule [3]. The objections continued until the discovery of the topoisomerases in 1971 [4] which satisfied all the doubts about the double-helix structure. That great leap elicited the further research on the cell biology. The presence of these protein based tools means that the topological problems can be resolved before making severe influences on the cell.

Before going further about these amazing proteins I will mention about the topology of the DNA.

In addition to the linear helical structure, DNA molecule is also observed in a closed circular form (ccDNA) in a great number of organisms. In ccDNA form, the chromosomes are in a highly compact and folded conformation. This packaging of DNA deforms it physically, thereby increasing its energy. DNA absorbs this torsional stress by coiling double helix over itself. Then the DNA is said to be supercoiled.

An under-wound DNA molecule (fewer turns than the relaxed state) is said to be negatively supercoiled. The helices wind about each other in a right handed path in space. An over-wound DNA molecule is said to be positively supercoiled. In this case the helices wound around each other in a left handed path.

Number of times that a strand wrapped around the other is a topological invariant and called as “linking number” (Lk). This topological bond is changed only when the phosphodiester linkage of DNA strands is broken by topoisomerases. There are two components of the linking number which are known as twist (Tw) and writhes (Wr) which have a basic relation expressed as:

$$Lk = Wr + Tw \text{ [5]} \qquad (1.1)$$

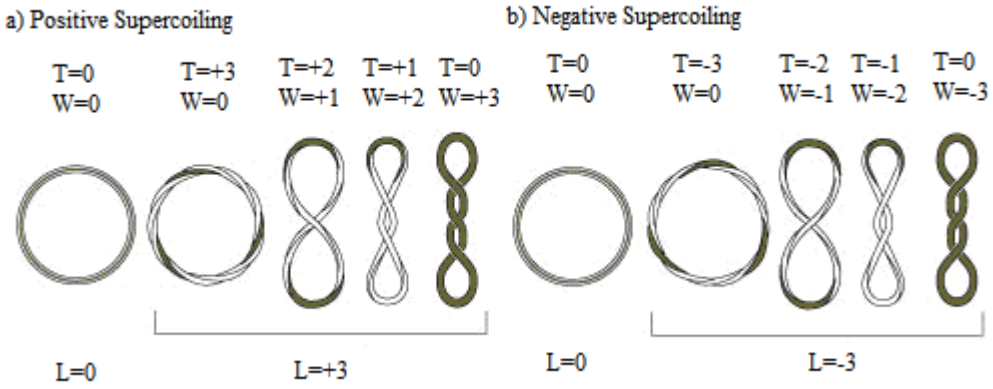


Figure1.1 The values of T and W for various positive and negative supercoils [6]

A series of application of this formula is conducted in figure 1.1.

Supercoiling density (σ) is a useful measure for the alteration of DNA topology formulated as:

$$\sigma = \Delta Lk / Lk_0 \qquad (1.2)$$

where ΔLk is the amount of the change of the linking number during the relaxation and Lk_0 is the linking number of a relaxed DNA molecule.

The supercoil formation is significant for a variety of reasons. First, through supercoiling, large volume of DNA condenses into the confined of a cell nucleus. Besides, the stored (potential) energy within the molecule is employed during DNA replication and transcription in which unwinding is needed. Finally, supercoils play a role in DNA catenation or knotting [7].

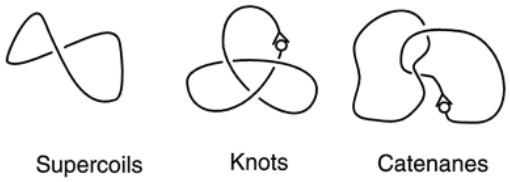


Figure 1.2 Representation of the knotting and catenation [8]

Individual strands of DNA molecule combine to form complex DNA catenae. Knotting, on the other hand involves linking of two strands in one circle. Figure 1.2 represents the three topological states of DNA molecule.

1.2. TOPOISOMERASES AND THEIR CELLULAR IMPORTANCE

DNA topoisomerases are enzymes that control and modify the topological states of DNA in living cells. DNA undergoes topological problems that stem from the replication, transcription, recombination, chromatin assembly and chromosome segregation [9]. The helical structure of DNA unwinds during these processes either permanently or temporarily. Due to the large size of the molecule in eukaryotic cells and circular form in the bacterial chromosome, DNA transactions create a superhelical tension that does not relax easily. Topoisomerases promote the relaxation by introducing or removing supercoils or formation resolution of knots and catenae [10]. They decrease the supercoil formation; that is the linking number is increased.

During the transcription, replication, recombination and chromatin remodeling processes, the tendency of DNA molecule to be as compact as possible, precludes the free movement of the strands. While unraveling the individual fibers in a rope, it will be knotted

and twisted after some time. This case illustrates the challenge that length and helical shape of the molecule impose.

All topoisomerases assure the relaxation of the topological problems by forming a transient phosphodiester bond between tyrosine residue of the enzyme and one end of the cleaved strands [11]. They utilize the C_4- oxygen of a catalytic tyrosine residue for nucleophilic attack to nick the strand and form a phosphotyrosine intermediate. The phosphotyrosine linkage is employed to reverse trans-esterfication reaction. During the breaking and maintaining the phosphodiester bond again no energy cofactor is needed [11, 12, and 13].

Topoisomerases are classified into two main groups and further subfamilies with respect to their sequence, structure and function. The main classes occur as type I and type II. The main distinction of the groups is the number of strand that is nicked. Type I enzymes act on the reactions involving the cleavage of one strand of the DNA molecule, while type II enzymes act on the reactions involving the cleavage of both strands. Type I enzymes that link to the 5´ end of the DNA molecule are defined as type IA and those which link to 3´ end are defined as type IB subfamilies [14]. Type II enzymes are further subdivided based on structural differences.

Table 1.1 Classification of topoisomerases [16]

Topoisomerase	**Subfamily Type**
Eubacterial DNA topoisomerase I *(E.coli)*	IA
Eubacterial DNA topoisomerase III *(E.coli)*	IA
Yeast DNA topoisomerase III *(S.cerevisiae)*	IA
Mammalian DNA topoisomerase IIIa (*human)*	IA
Mammalian DNA topoisomerase IIIb (*human)*	IA
Eubacterial and archaeal reverse gyrase (*sulfolobus a.)*	IA
Eubacterial reverse gyrase *(m.kandleri)*	IA
Eukaryotic DNA topoisomerase I *(human)*	IB
Poxvirus DNA topoisomerase *(vaccinia)*	IB
Hyperthrmophilic eubacterial DNA topoisomerase V (*m.kandleri)*	IB
Eubacterial DNA gyrase *(e.coli)*	IIA
Eubacterial DNA topoisomerase IV *(e.coli)*	IIA
Yeast DNA topoisomerase II *(s. cerevisiae)*	IIA
Mammalian DNA topoisomerase IIa *(human)*	IIA
Mammalian DNA topoisomerase IIb *(human)*	IIA
Archaeal DNA topoisomerase IV *(sulfolobus s.)*	IIB

E. coli topoisomerases I and III, eukaryotic topoisomerases III and archeal reverse gyrase are some of the enzymes belonging to the type IA subfamily [15]. On the other hand, type IB subfamily includes eukaryotic topoisomerases I, archeal topoisomerases V, and the poxvirus type I topoisomerases [15]. Subfamily members of some prokaryotic and eukaryotic topoisomerases are given in Table 1.

The two types of enzymes have nothing in common by means of their sequence and structure. Type II enzymes are dimeric and require ATP to pass the DNA strand through the break. These enzymes alter the linking number of a ccDNA in steps of two. On the other hand, type I enzymes are monomeric and do not consume additional energy. Only one step change in the linking number is possible in the case of type I topoisomerases [17]. A general view for two types of topoisomerases is represented in figure 1.3 and figure 1.4.

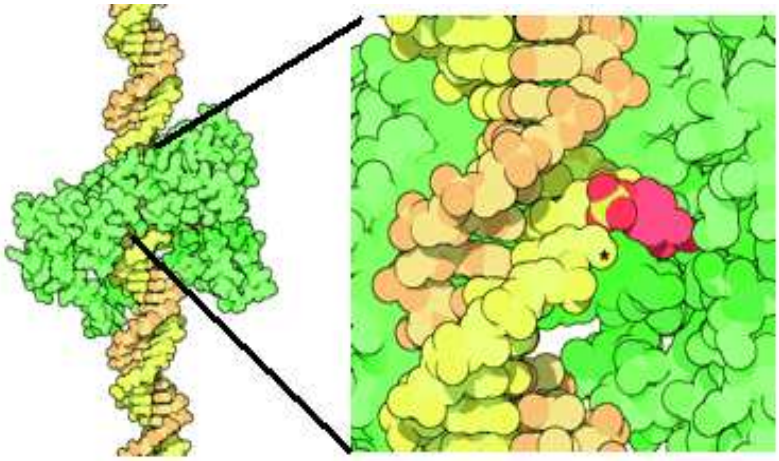


Figure 1.3 A representation of type I topoisomerases [16]

In figure 1.3 the front side of the protein (green) is removed to give a better sight. The active site of the tyrosine residue is marked with red. It cuts the DNA strand by forming a covalent bond with phosphate of the strand (marked with bright yellow and orange). One end of the gap which has a star on it becomes free to rotate around the other strand. The DNA is released after the relaxation.

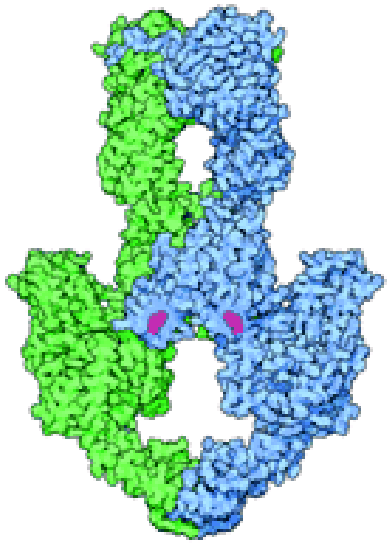


Figure 1.4 A representation of type II topoisomerases [16]

Figure 1.4 shows the dimeric structure of the type II enzymes which has holes in the middle. Differing from type I, it has two tyrosine residues forming the linkage (shown in pink) to the 5´ end of the DNA phosphate. A possible mechanism of this highly dynamic structure suggests that the DNA enters through the hole and separates to its strands there. One of the strands passes through the upper hole followed by the released strand from the middle hole.

1.3. HUMAN TOPOISOMERASE I

The focus of this thesis study is type IB topoisomerases also known as the human topoisomerases. As a summary of the aforementioned information, type I topoisomerases are monomeric enzymes transiently nicking one strand of the DNA in an ATP-independent manner. The further classification is based on the differences in the catalytic action. Type IA and type IB topoisomerases, sharing no sequence of structural similarities, performs the same functions in different ways. Type IA enzymes mostly observed in prokaryotic cells also identified in eukaryotic cells recently which are grouped as topoisomerase III [13]. These enzymes are able to relax only negatively supercoiled DNA, require Magnesium and single stranded starch of DNA and covalently attach the 5’ end of the nicked strand [18]. On the other hand, Type IB enzymes have only been found in eukaryotic cells with one exception of

vaccinia virus [19]. Type IB enzymes relax both positive and negative supercoils, do not function only on a single stranded DNA and form the phosphodiester linkage through the 3' end of the strand without any energy cofactor or metal cation needed [20].

1.3.1. Structural Features

Human topoisomerase I (topo I) is a monomeric protein consisting of 765 amino acids. Based on the limited proteolysis and the crystallographic data [21], the 91-kDa human topoisomerase I can be divided into four major regions: the highly charged NH2-terminal, the conserved core domain, a positively charged linker part, and the highly conserved COOH-terminal domain [22]. The figure 1.5 shows how the domains distribute among the residues. The core domain and the C-terminal domain which holds the active site Tyrosine (Tyr723) are the same for all the topoisomerase IB family. On the contrary, the linker and N-terminal domains may have various values for their length and sequence [22].

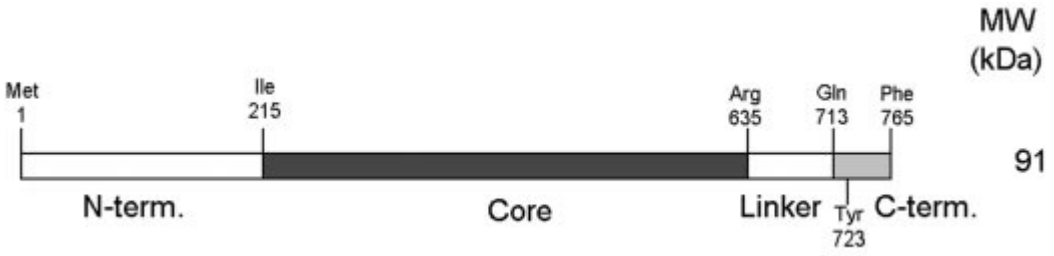


Figure 1.5 Schematic illustration of the domains among the residues [23].

The N-terminal including 214 residues of human enzyme does not have a function during the catalysis process in vitro. The amino acids contained are few hydrophobic, very disordered and highly sensitive to protease [11, 16]. The nuclear targeting signals in this domain serve in nucleolar localization through the interactions with nucleolin [24].

The core domain of 421 amino acids contains all of the residues playing role during the catalysis action except the active site tyrosine [16]. The core domain is divided into subdomains I, II and II, which will be mentioned later.

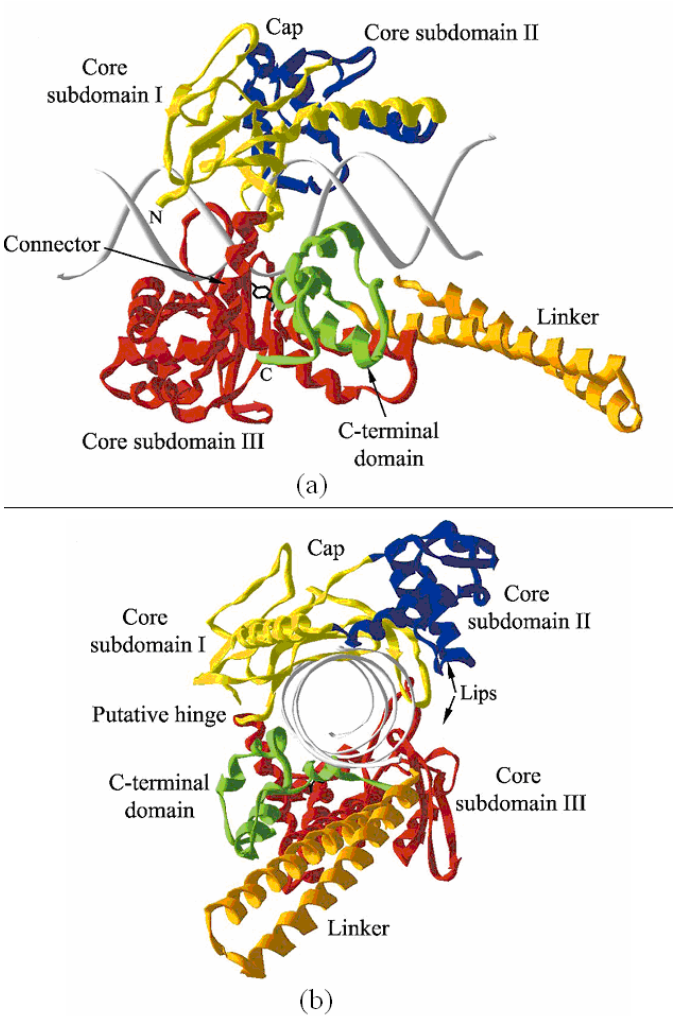


Figure 1.6 The two views of the protein for the DNA axis oriented horizontally (a) and vertically (b) [11]

The other protease sensitive region of the protein is the linker domain of 77 residues which connects the core and the C-terminal domains. Although the experimental studies indicate that the absence of this domain influences on the enzyme, this poorly conserved and positively charged domain is found not to be required for the catalysis activity in vitro [25].

The COOH-terminal domain has 57 residues including the essential nucleophilic tyrosine (Tyr 723) which forms a phosphodiester bond with the scissile strand of the cleavage site of the substrate DNA [26]. Together with the core domain the C-terminal domain (residues from 200 to 635) can be isolated to gather the full enzyme activity. As a summary, human topoisomerase I comprised of four domains only two of which are crucial for the catalytic and relaxation functions; the core and the C-terminal domain [11, 16].

The overall crystal structure of human topoisomerase I reveal a bi-lobed protein of 15-20 A^0 which is wrapped around duplex DNA for the protein-DNA phosphate interactions. The top lobe of the protein is named as “cap” which consists of core subdomains I (residues 215-232) and II (residues 233-319), The cap region includes two α-helices forming a V-shape from the front side and extends the protein 25 A^0 from the body [11]. In the figure 1.6 the two lobes of the protein from different views is shown. The other lobe, placed under the DNA, is made up the core subdomain III (residues 434-633) and the C-terminal domain. This lobe is connected to the cap through a long α-helix which is indicated as “connector” in figure 1.6.a. This below lobe includes all α-helices and three β-sheets [11]. These sheets contain the catalytic residues functioning during the cleavage and relaxation reactions. The protein clamp opens and closes via a pair of opposing loops termed as “lips” and the “hinge” [27] region at the opposing side of the lips, through the interaction of six amino acids and one salt bridge [13]. The lips region have two different clamp sites, one form His-367 to Ala-499 residues and the other from Gly-365 to Ser-534 residues which are termed as distal and proximal clamps respectively.

1.3.2. Functional Aspects of the Type IB Topoisomerases

1.3.2.1. DNA binding

Human topoisomerase I engage an intimate interaction with its DNA substrate and reveals how DNA substrate be aligned for joining covalently [16, 13]. The protein wraps around the DNA molecule by burying a total of 2500 A^{02} accessible surface area. The interaction area between the DNA and the core subdomains I and III is more spread compared to the core subdomain II and the C-terminal domain. For the covalent and the non-covalent complexes of the DNA molecule the protein-DNA contact area is identical [16].

The V-shape helices of the cap region are not favorable for the sugar-phosphate structure of the DNA because of having a high amount of positive charges. Instead the DNA joins the protein through the subdomain I residues for the cleavage. The repulsion of the positively charged helices is consistent with the free rotation model [12].

The active site of the protein gathering around the scissile phosphate is composed up of the nucleophilic Tyr723 together with the two Arginine residues (Arg488 and Arg590) and a conserved Histidine (His632). The active site is the same for all known topoisomerase IB sequences [28] and does not undergo a considerable shift during the cleavage and binding processes except His632 [16]. For both the covalent and non-covalent complexes of human topoisomerase IB and DNA an active site mechanism has been proposed involving the nucleophilic attack of the C4-Oxygen atom of Tyr723 on the scissile phosphate group. The positively charged Arg488, Arg590 and His632 were considered for the stabilization of the transient scissile phosphate. A probable function for the His632 is transferring the proton to 5'-hydroxyl group [19].

1.3.2.2. Relaxation of DNA Superhelical Tension

Although there is not a general agreement, some descriptions have been proposed on the mechanism of DNA relaxation for example the free rotation model [29]. The free rotation hypothesis suggests the relaxation of the DNA superhelical tension by just permitting the DNA duplex downstream to make free rotation around the intact strand [15]. This model proposes a less energy loss and structural damage because of not loosing the base pair interactions [12]. The rotation could be in any direction depending on the supercoil type. The tests for these mechanistic approaches based on the crystal structures of human topoisomerase I lead to a new proposal termed as "controlled rotation" [12]. In this mechanism, the rotation of the DNA duplex downstream around the intact strand is either controlled or hindered by the interactions of DNA with the cap of the enzyme and the linker domain [15]. The conformational and structural evidences support this proposal. Firstly, the charge density and the conformation of the active site of enzyme permit a controlled rotation of the DNA duplex downstream around the intact strand [13]. Secondly, the shape of the positively charged cavity of the enzyme seems to assist the rotation of the DNA duplex through interactions with helices and linker elements [13]. Finally, the v-shaped helices which are conserved in all topoisomerase IB family exhibit a rotational motion even in free space which could also facilitate the rotation of the DNA duplex [12, 13, 22].

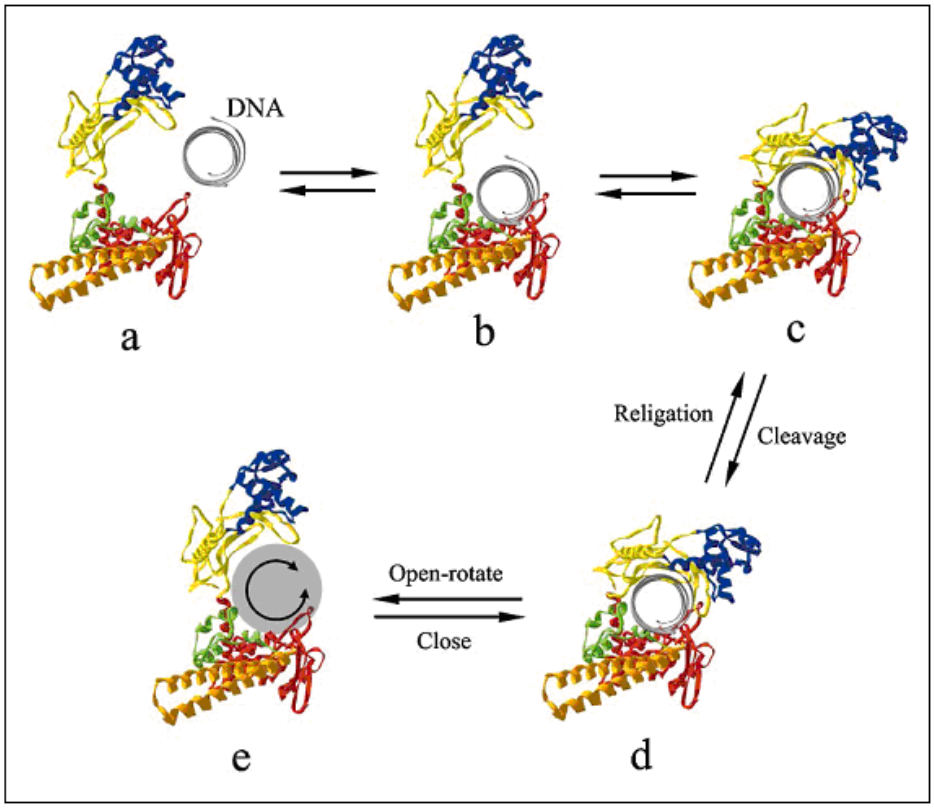


Figure 1.7 A possible mechanism for the human topoisomerase I [11].

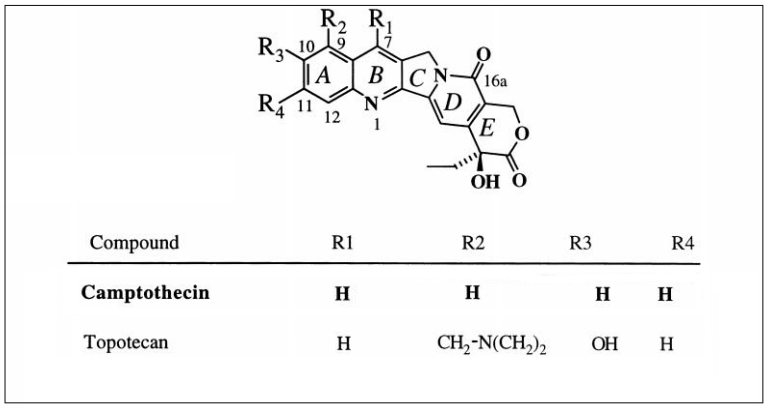
In figure 1.7, a view from the axis of the DNA duplex is represented. In the steps (a) and (b) the enzyme binds the DNA duplex and in steps (c) and (d) catalytic interactions take place. In step (e), the DNA duplex is shown to rotate in the open form of the enzyme. The DNA duplex either releases or continues to the rotation after the step (e).

1.4. INHIBITING TOPOS: ANTI-CANCER DRUG MOLECULES

The integrity of the genetic material during the transcription, replication and recombination processes is established by the DNA topoisomerases. The removal of the knots and supercoils is crucial for the cell [30]. The inhibitors of the topoisomerases are useful elements when the termination of the cell is aimed. They stabilize the transient covalent attachment and prevent the action of the enzymes.

A strong chemical agent class, Camptothecin (CPT), is one of the mostly examined compounds blocking the enzyme function by forming a ternary and stable complex with the

DNA-topoisomerase intermediate [31]. The trapped DNA-topoisomerase complexes stop the transcription and replication processes and lead to the apoptosis [32]. Although there are any other clinically tested cancer drugs, the CPTs are of strong interest to have a deeper insight of the targeted cancer chemotherapy [13].



Compound	R1	R2	R3	R4
Camptothecin	**H**	**H**	**H**	**H**
Topotecan	H	CH_2-N(CH_2)$_2$	OH	H

Figure 1.8 The CPT and its Topotecan derivative [33]

The use of CPTs is most commonly in the form of the two derivatives; Topotecan and Irinotecan which act on late-stage solid tumors of the colon, esophagus and ovaries [13]. In this study the focus will be the Topotecan derivative which is represented in the figure 1.8 which is currently in use against ovarian and small lung cancers although it's high rate of developing resistance and the transient effects.

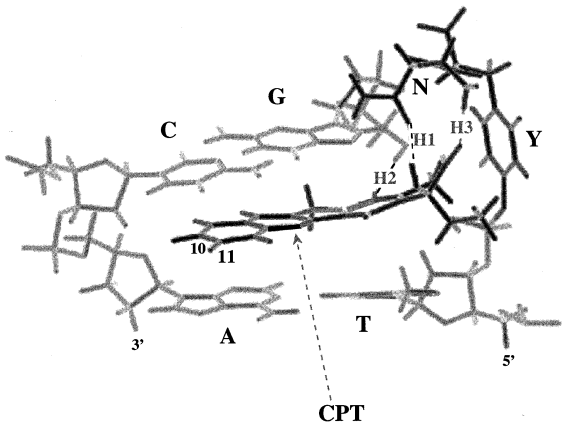


Figure 1.9 A hypothetical DNA-enzyme and CPT ternary complex [33]

The mechanism for the forming the enzyme-DNA-CPT ternary complex is still been investigated. According to the most of the reports, the molecule binds the DNA duplex through the +1 guanine which is stacked against the five-member CPT ring system [12]. In figure 1.9, a possible arrangement of the complex is modeled. The CPT molecule acts like a base pair and take place among the cleavage site.

The reported experimental data shows that the TPT molecule interacts with the protein through five amino acids: Tyr-723, Asn-722, Asp-533, Lys-532 and Arg-364. The only residue directly bonds with TPT is the Asp-533. The two residues; Asn-722 and Tyr-723 form water-mediated bonds according to the experimental data [10].

CHAPTER II

THEORETICAL APPROACH

2.1 FORCE FIELD AND MOLECULAR DYNAMICS SIMULATIONS

The computer simulations are carried out to make the structural and mechanistic analysis of assemblies of molecules. These simulations are the complement of the conventional experiments providing a microscopic approach to the complex problems. Molecular Dynamics (MD) is one of the main simulation techniques establishing the dynamical properties of the system and acting like a bridge between the theory and experiment [33].

The idea of simulating the motion of the system of molecules first appeared on the study of Alder and Wainwright with macroscopic sphere balls on 1950's [34]. The studies of Rahman and Stillinger [35] with a realistic potential carried the MD simulations to the next level. Then the first protein simulation [36] has appeared on 1977. This method is now a widely used to reach detailed information on the fluctuations and conformational changes of protein-DNA complexes and nucleic acids in the literature. The techniques are specialized to for particular problems including mixed quantum mechanical - classical simulations.

MD simulates the time-dependent behavior of a system of atoms or molecules. MD simulations employ the numerical methods instead of the analytic approach because the high number of particles and interactions between them complicate the calculations. Because of the same reason the quantum mechanical effects can be ignored. MD simulations are basically a case of the N-body problem [37] where the Newton's law is integrated to the motion of particles with the inter-particle forces included. The time-dependent manner of the

calculations provides an opportunity to work on both small and fast systems where the time is in the order of picoseconds to nanoseconds such as atomic fluctuations (0.01 to 5 Å, 10^{-15} to 10^{-1} s), rigid body motions 1 to 10Å, 10^{-9} to 1s) or large scaled motions (> 5Å, 10^{-7} to 104 s)

MD simulation is based on solutions the classical equations of motion which may be written for a simple atomic system as:

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

So the force f_i which is the derivative of the potential energy $U_{(r^N)}$ should be calculated. Here $r^N = r_1, r_2, \ldots, r_N$ represents the complete set of the atomic coordinates. Now the focus will be the description of $U_{(r^N)}$.

Before going further, a model must be established to determine the energy of the system with respect to the bonded structure and relative positions of the atoms. The standard way is to behave individually on the bonded and non-bonded energies [38]:

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

where the term non-bonded refers to the non-bonded atomic interactions and the bonded stands for the atoms having connected with three or less bonds. The bonded interactions are mostly less than the non-bonded interactions due to being more clear and limited in number, on the contrary the non-bonded interaction are massive due to being random and not specific [38]. Figure 2.1 shows the bonded and non-bonded potential parameters.

The traditional approach to non-bonded interaction splits the potential interactions to one-body, two-body, ..., n-body terms:

$$U_{non-bonded}\left(r^N\right) = \sum_i u(r_i) + \sum_i \sum_{j\rangle i} v(r_i, r_j) + \ldots \tag{2.3}$$

The $u_{(r)}$ term refers to an external potential field or the effect of the container. It is a general way to neglect the three or more pair interactions and focus only on the two-body interactions.

The non-bonded interactions also include the continuous differentiable pair potentials such as Lennard-Jones and Coulomb interactions. Lennard-Jones describes approximately the attractive and repulsive behavior of the separate atoms. The electron clouds of the distant atoms tend to hold each other, however for the adjacent atoms the the electron clouds overlaps and they repel each other. The most commonly used form of the Lennard-Jones potential is:

$$U_{r_{ij}}^{LJ} = \varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.4}$$

where σ_{ij} is the diameter, ε_{ij} is the depth of the potential well and r_{ij} is the distance between the particles. This model assumes the pair-wise interactions are not influenced by the surrounding electron clouds. In addition, the repulsion term, $\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}$ is not taken to be depend on the distance exponentially as in the real case to ease the calculations.

The other component of the non-bonded interactions is the Coulomb potential, representing the electrostatic interaction between the atoms.

$$U_{coul.} = \frac{q_i q_j}{4\pi\varepsilon r_{ij}} \tag{2.5}$$

where q is the charges of the atoms and ε is the dielectric constant.

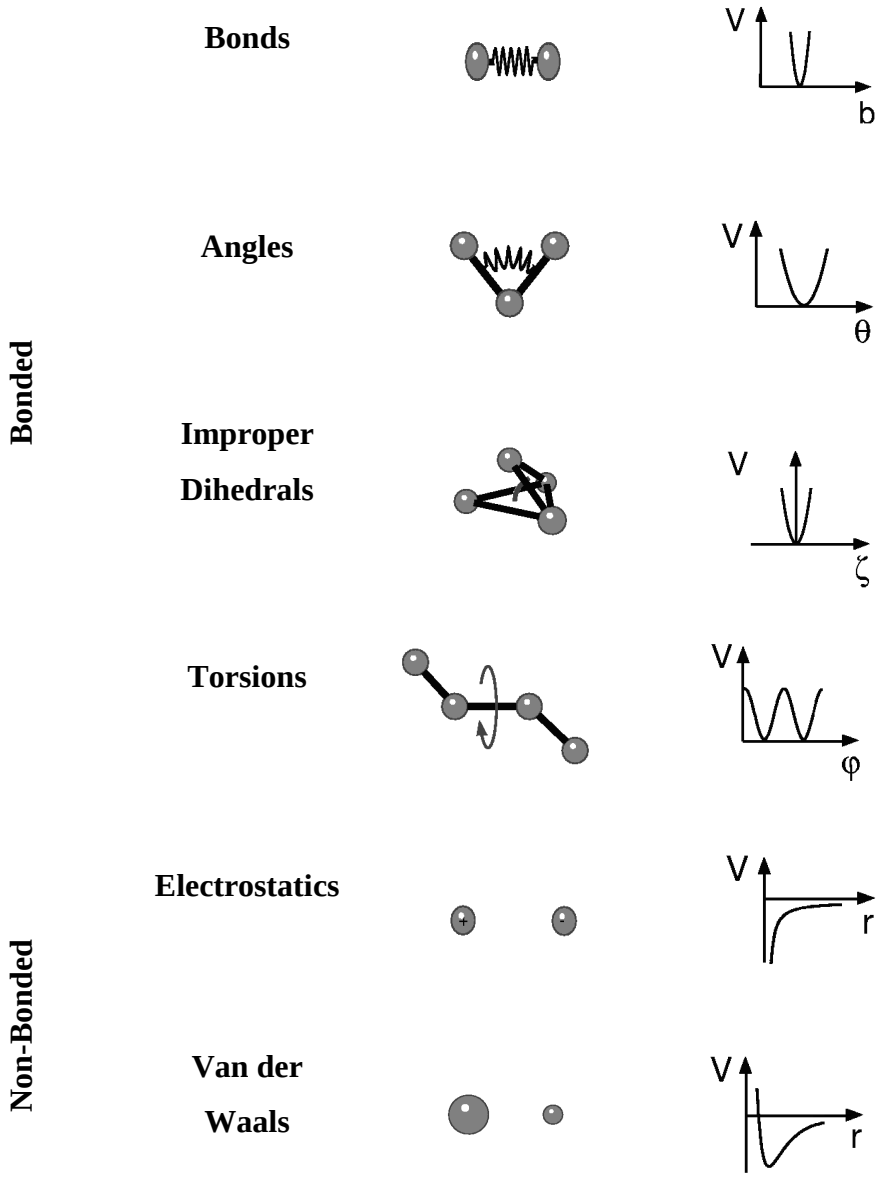


Figure 2.1 The bonded and non-bonded potentials and the graphics representing the change in the potential with respect to the distance or angle parameters [39]

The bonded potential includes the intra-molecular covalent interactions shown in the figure 2.1;

$$U_{bonded} = U_{bonds} + U_{angle} + U_{dihedral} + U_{improper} + U_{Urey-Bradley} \qquad (2.6)$$

The term U_{bonds} refers to the simple harmonic motion between two atoms connected with only one bond:

$$U_{bonds} = \sum_{bonds} k_b (b - b_0)^2 \tag{2.7}$$

where k_b is the force constant, b is the bond distance at any instance and b_0 is the bond distance at equilibrium. Both k_b and b_0 values are specific to the atoms and the constituents and are found form the experimental data such as infrared stretching frequencies or quantum mechanical calculations.

The second term in the Equation (2.6) refers to a harmonic potential depending on the angle between two successive bonds. The harmonic manner of this motion relates this case to the Hooke's Law as for the case of bonds.

$$U_{angles} = \sum_{angles} k_\theta (\theta - \theta_0)^2 \tag{2.8}$$

where k_θ is the force constant and θ_0 is the angle for the ideal case. These values are also specific to the atoms. The terms U_{angles} and U_{bonds} together describe the displacement from the equilibrium situation.

The term $U_{dihedral}$ in Equation (2.6) involves four atoms attached with three covalent bonds and applies to the rotation of the three atoms around the bond in the middle. This torsional potential is described as a periodic function;

$$U_{dihedral} = \sum_{dihedrals} k_\phi [1 + \cos(n\phi + \delta)] \tag{2.9}$$

where k_ϕ is the force constant, ϕ is the angle of dihedral and δ is the phase of the dihedral. The n term is the coefficient of symmetry around the middle bond.

$U_{improper}$ refers to the improper torsion potential which enforces the planarity of the conformations.

$$U_{improper} = k_w (w - w_0)^2 \tag{2.10}$$

where k_w is the energy constraint and the w is the angle of two planes.

These specified potential energy functions and the parameter sets which are the building blocks for the calculation of force and potential energy of a system termed as force-field [40]. The force-field parameters come from both experimental structural data and the high-level quantum-chemical calculations.

The MD simulation procedure starts with the assignment of the initial values of the coordinates and velocities of the atoms. The X-Ray crystallography or NMR data provides the coordinate values at $t = 0$.

The MD simulation method simply based on the Newton's second law leading the equation of motion,

$$F_i = m_i a_i \tag{2.11}$$

where F is the force exerted of an atom in the case of molecular motion. As the gradient of the potential energy function with respect to the internal coordinates gives the force, the force acting on each atom, i could be found by the gradient of the force-field:

$$F_{i(r)} = -\nabla_i U_{(r)} \tag{2.12}$$

Combining Equation 2.12 with 2.13 yields:

$$\frac{-dU_{(r)}}{dr_i} = m_i \frac{d^2 r}{dt^2} \tag{2.13}$$

From Equation 2.14 the equation of a motion of an object can simply be derived. However in the molecular motion the calculation becomes very difficult and expensive due to the high number of atoms. There are numeric methods developed to facilitate the calculation such as Basic Verlet Algorithm, Leap-Frog Algorithm, Velocity-Verlet Algorithm and Beeman's Algorithm.

The Leap Frog algorithm is a time-reversible, stable, simplectic and easy to conduct way of calculation. It is a modified version of Verlet algorithm. The idea for both algorithms based on the Taylor expansion of velocities and positions. The Taylor expansion of the 3rd order yields:

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

The Verlet algorithm [41] starts with positions at t and the accelerations at time $t-\Delta t$ and to predict the positions at $t+\Delta t$ where Δt is the integration step. The results obtained with an error at the order of Δt^4 for positions and Δt^2 for velocities. The verlet algorithm gives the velocity as below:

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

Leap-frog algorithm on the other hand uses the velocities at the half-step;

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

and the positions at each step;

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

The leap-frog algorithm is easier to conduct computationally and requires less storage which is an important advantage in the case of large-scale calculations. In addition, the conservation of energy is maintained which shows the stability of the integrator. However, in

this method, positions and velocities are not operated simultaneously hence to calculate the total energy at a given time step becomes impossible.

2.2 HALF QUADRATIC BIASED MD (HQBM)

HQBM is a method utilized when the macromolecule is aimed to undergo a conformational change. The molecule is subject to a very small time dependent external perturbation during the short time dynamics of the system so the conversation of the total energy is maintained. Using an extended Lagrangian can be used with the conserved macroscopic state to examine the correctness of the forces. The conformational change with an external perturbation requires two main elements; a reference or target structure which has the desired conformation and a reaction coordinate which leads from the initial to the final state [42].

The perturbation applied on the defined reaction coordinate is in the half quadratic form [43]. The reaction coordinate is chosen to be the mean square difference from the target coordinates.

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

where N is the total number of atoms, r_{ij} is the magnitude of the distance between i^{th} and j^{th} atom $\left| r_i - r_j \right|$ and R labels the final reference coordinates.

The time dependent perturbation added to the molecular potential energy to force the system to a defined configuration has the form [44]:

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

The MD simulation begins with the value $\rho_0 = \rho(0)$ at $t = 0$ and if there is a dramatic increase in a step, i.e. if $\rho(t + \Delta t) > \rho_0(t)$ the perturbation will be ineffective to the system. In this case, the value for $\rho_0(t)$ is adjusted to continue the simulation. When $\rho(t) < \rho_0$ the external perturbation acts through the reaction coordinate and permits a dramatic decrease in ρ. The α value is set by the user to limit the backward fluctuation of the system [44]. As α gets smaller the accuracy of the simulation increases due to decrease in the external impact. In figure 2.2 the change of perturbation during the motion is shown with respect to time.

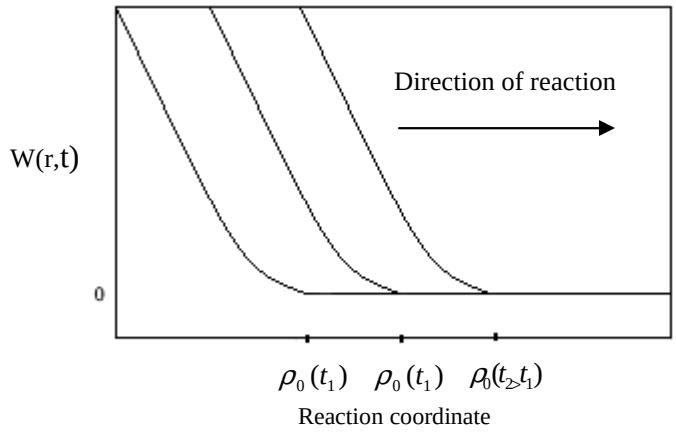


Figure 2.2 The external perturbation depending on time [44].

The effect of perturbation corresponds to a force which could be calculated by:

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

Inserting the perturbation function gives: $\rho_0(t_1)$ $\rho_0(t_{2>}t_1)$ $\rho(0)$

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

In this study this method has been used to simulate the rotation of the nicked DNA downstream with and without TPT molecule around an axis both in negative and positive relaxations. The calculations regarding a rotation around the axis have been done in every 10 degrees.

The perturbation acts on the system without affecting the macroscopic quantities. However when the perturbation energy has to deal with a free energy barrier, a transition from kinetic energy to the potential energy is observed, hence the temperature of the isolated system is decreased. Preventing such kind of situations is possible by forcing the temperature to be constant.

2.3 CONSTANT TEMPERATURE/PRESSURE MD (CPT MD)

MD simulations can be utilized using different variables characterizing the macroscopic behavior of the system. A system surrounded by a heat bath is identified thermodynamically by its volume (V), temperature (T) and number of particles (N) and defined as NVT ensemble [45].

There are many methods to control the temperature without altering the energy of the system. In this study the Berendsen method is implemented [46]. Berendsen method makes a fine adjustment on the Andersen method [47]. The system is in interaction with the heat bath and the velocities of the interacting particles change. Andersen method suggests resetting the velocities of these particles and reassigning velocities to them using Boltzmann distribution. This would, however cause severe errors during the calculations. Berendsen, on the other hand, defined a factor $\lambda(t,T)$ which adjusts the velocity shift in accordance to the temperature and time [48]. Although the canonical ensemble may be deformed with Berendsen method, it works well for the equilibration phase at a new temperature.

2.4 *Ab Initio* MOLECULAR MODELING

Molecular modeling is a term referring prediction and analysis of the structure, geometry and reactions of a molecule or a system of molecules by employing computational methods. There are three different molecular modeling methods. Molecular dynamics method which is described above models the molecular system in terms of classical mechanics. Ab initio methods rely only on the mathematical calculations of the wave functions. Finally semi-empirical methods use experimental results to perform ab-initio methods in a faster and easier way.

Ab-initio methods produce the most accurate results due to having a completely mathematical approach. The only initial values are some of the universal constants and the structural information of the system. The determination of electronic wave-functions involves complex quantum mechanical calculations. The calculations based on the solutions of the time-dependent Schrödinger equations:

$$H\Psi = E\Psi \tag{2.22}$$

In order to ease the calculations of these many-electron systems Born-Oppenheimer approximation is utilized [49]. The total energy is separated for electronic and nucleic motion in molecules.

$$\Psi_{molecule}(\vec{r}_i, \vec{R}_j) = \Psi_{electrons}(\vec{r}_i, \vec{R}_j)\Psi_{nuclei}(\vec{R}_j) \tag{2.23}$$

Since the nuclei are much heavier than the electrons, they exhibit much slower motion than the electrons. As the nuclei can be considered as static the nuclear motion can be eliminated. In this case the electronic wave functions of molecules depends on these fixed relative coordinates but not the nuclear velocities. Hence, according to this approximation, molecular systems have electrons moving quantum mechanically in the potential field of a set of atomic nuclei [50].

The ab-initio methods can be using two different theories: Hartree Fock (HF) and Density Functional Theory (DFT). There are also some hybrid methods combining the two theories. In this study, the two of the theories and a type of hybrid functional is conducted to find the optimized geometry and the force field parameters of the drug molecule; TPT.

2.4.1 Hartree Fock Method

The main idea in the HF method is to calculate the Schrödinger wave functions of the electrons one by one. One electron is treated like moving in the effect of a source of potential caused by the other electrons which are taken to be fixed. Once the more accurate orbital is defined for that electron, the calculations are repeated for all of them. Then the cycle is

processed again and again using the new defined orbitals as the initial point until the orbitals become stable which is termed as “self-consisted”.

The wave function of non-interacting particles is generally represented by the Slater determinant of occupied orbitals which is denoted with Φ.

$$\Phi(x_1,...,x_N) = \begin{vmatrix} \Phi_1(x_1) & ... & \Phi_N(x_1) \\ . & & . \\ . & & . \\ . & & . \\ \Phi_1(x_N) & ... & \Phi_1(x_N) \end{vmatrix} \tag{2.25}$$

Then the Hartree Fock energy is:

$$E = \min_{\Phi} \left\langle \Phi \quad \left| \hat{T} + \hat{U}_{ee} + \hat{U}_{ext} \right| \quad \Phi \right\rangle \tag{2.26}$$

where $\hat{T}$ is the kinetic energy of non-interacting particles, $\hat{U}_{ext}$ is the external potential and $\hat{U}_{ee}$ is the coulomb potential. Since the determinant is in the anti-symmetric form, the coulomb potential takes the form below:

$$\left\langle \Phi \quad \left| \hat{U}_{ee} \right| \quad \Phi \right\rangle = U[\Phi] + E_x[\Phi] \tag{2.27}$$

where $U[\Phi]$ is the pure electrostatic potential of the charges without any quantum effect and the second is the Fock integral which is also known as the exchange integral. The exchange integral shown in the equation 2.28 exhibits the effect of the Pauli Exclusion Principle with the orbitals occupied with double integral.

2.4.2 Density Functional Theory

Density Functional Theory has long been the fundamental for the electronic structure calculations. The main emphasis in this method is the charge as the relevant physical quantity. The method establishes highly successful descriptions of structural and electronic properties

of vast types of materials. Furthermore, the method is handled easily by the computational tools without loosing the accuracy.

Rooting back to the Hohenberg-Kohn (H-K) theorem, DFT is formalized with the Kohn-Sham calculations. The H-K theorem implies an approach to the quantum mechanical problems by replacing the potential energy function with the density function [51]. The theory states that the ground-state density $n(r)$ determines the potential up to an arbitrary constant [52]. The density can be stated as,

$$n(r) = N\int \left|\Psi(r_1, r_2, ..., r_N)\right|^2 dr_2 ... dr_N \tag{2.28}$$

with N equals to the number of electrons and Ψ is the ground-state wave function. It is also true that,

$$E[(n(r)] = \left\langle \Psi \quad \left| T + U + V \right| \quad \Psi \right\rangle = \left\langle \Psi \quad \left| T + U \right| \quad \Psi \right\rangle + \left\langle \Psi \quad \left| V \right| \quad \Psi \right\rangle \tag{2.29}$$

where T is the kinetic energy, U is the electron-electron interaction and V is the external potential. That is,

$$E[n(r)] = F[n(r)] + \int n(r)V(r)dr \tag{2.30}$$

The energy is obtained to be a functional of density. The Kohn-Sham (KS) equations on the other hand offer a solution to the kinetic energy problem [53]. The kinetic energy cannot be stated as a functional of $n(r)$ without massive calculation errors until KS system is introduced. Kohn and Sham construct a system of non-interacting electrons and employ the variation of energy. Using KS method the energy functional $E[n(r)]$ can be minimized through the variation of densities under orthonormality conditions.

2.4.3 Hybrid Methods: B3LYP

The kinetic energy of the electrons and the non-classical electron-electron exchange energy appearing in the DFT model differ form the HF theory. There is also and additional

term in DFT describing the energy related to correlated movement of the electrons having different spins. These differences require specific approaches to calculate the energy term for DFT such as local density approximation (LDA) method using the electron density or the generalized gradient approximation (GGA) using the gradients of the electron density as well. In addition to pure DFT methods, hybrid functionals as a linear combination of HF and DFT exists to calculate the exchange and correlation energy terms. In this study the method using Becke-3 term correlational functional with Lee-Yang-Parr exchange correlation functional (B3LYP) is employed.

Methods based on LDA assume a molecule in the gaseous state has a uniform electron density throughout the molecule so they do not provide satisfactory results except the solid state calculations. When the electron density calculations are combined with the gradient correlation factor, the rate of change of the density is taken into account. B3LYP method on the other hand combines the HF and DFT approach to calculate the exchange and correlation functionals. The exchange functional term combines the aforementioned methods in the way described below [54]:

$$U_{xc}^{B3LYP} = U_{xc}^{LDA} + a_0(U_x^{HF} - U_x^{LDA}) + a_x(U_x^{GGA} - U_x^{LDA}) + a_c(U_c^{GGA} - U_c^{LDA}) \quad (2.31)$$

where $a_0 = 0.2$, $a_x = 0.72$ and $a_c = 0.81$ are empirically defined values by fitting the predicted values to a set of values for ionization potentials, proton affinities, atomic energies and atomization energies.

In HF method the lack of Coulomb correlation term implies that the charges get closer to each other and the bond lengths get smaller values than it should be. Hence the binding energies increase. In addition to this systematic error the HF method is not applicable for the molecules including many atoms. On the other hand, in DFT, the systematic errors occur in the opposite direction. Thus, in this study the hybrid functional is used which produce a considerably less RMS error.

2.4.4 Basis Sets

Basis sets represent a set of numbers to initialize a computational calculation of the wavefunction. A wavefunction is a mathematical description of the interaction of electrons with each other and nucleus. Minimal basis set is used to simplify the complex Slater Type Orbital (STO) calculations. The idea is approximating a STO by combining three Gaussian functions. These minimal sets are termed as STO-nG where n refers to the number of Gaussians used. STO-nG basis sets include a specific function for each orbital in the atom. The minimal basis sets assumes each orbital is equivalent. The Split-Valence basis sets treat this problem by differentialting the core(inner) and valence(outer) electrons. The notation of Split-Valence sets are based on X-YZg format where X represents the number of Gaussians building up the core orbital functionals and the YZ points out the Gaussian function types for valence orbitals. For the Diffuse basis sets, the system is considered to be in its excited state and is indicated with a + sign. In this study the Gaussian 6-311+G(p,d) basis set is chosen which add polarization functions to the Hydrogens and heavy atoms with (p,d) sign.

2.4.5. Energy Optimization and Frequency Calculation Processes

The algorithm used in this study approximates the potential energy surface as a quadratic function and make the calculations based on Hessian matrix. The Hessian matrix, mathematically, is a square matrix having elements of the form of second-order partial derivatives of a function. In the case of optimization of the molecule, the matrix elements are the second derivatives of the energy with respect to nuclear motion in Cartesian coordinates. The nuclei are then moved in a manner which changes the energy inverse to the values of the corresponding Hessian.

The optimization process starts with an initial guess of the Hessian Matrix f_{cart} in Cartesian coordinates.

$$f_{cart} = \left(\frac{\partial^2 U}{\partial \xi_i \partial \xi_j} \right)_0 \tag{2.32}$$

where U is the potential and $\xi_{i,j}$ values refer to the displacements in the Cartesian coordinates. The first derivatives are the zero since the system is in equilibrium. This is indicated with the zero sign $\left(\right)_0$. The matrix has the form:

$$f_{cart} = \begin{bmatrix} \frac{\partial^2 U}{\partial x_1^2} & \frac{\partial^2 U}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 U}{\partial x_1 \partial x_N} \\ \frac{\partial^2 U}{\partial x_2 \partial x_1} & \frac{\partial^2 U}{\partial x_2^2} & \cdots & \frac{\partial^2 U}{\partial x_2 \partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 U}{\partial x_N \partial x_1} & \frac{\partial^2 U}{\partial x_N \partial x_2} & \cdots & \frac{\partial^2 U}{\partial x_N^2} \end{bmatrix} \tag{2.33}$$

These force constants are converted to the mass weighted Cartesian coordinates (MWC) by the formula:

$$f_{MWCij} = \frac{f_{cart}}{\sqrt{m_i m_j}} = \left(\frac{\partial^2 U}{\partial q_i \partial q_j} \right)_0 \tag{2.34}$$

where $q_{i,j} = \sqrt{m_{i,j}}\xi_{i,j}$ are the mass weighted coordinates. A matrix of f_{MWC} is formed yielding a set of 3N eigenvalues and eigenvectors. The roots are the eigenvalues give the fundamental frequencies of the molecule.

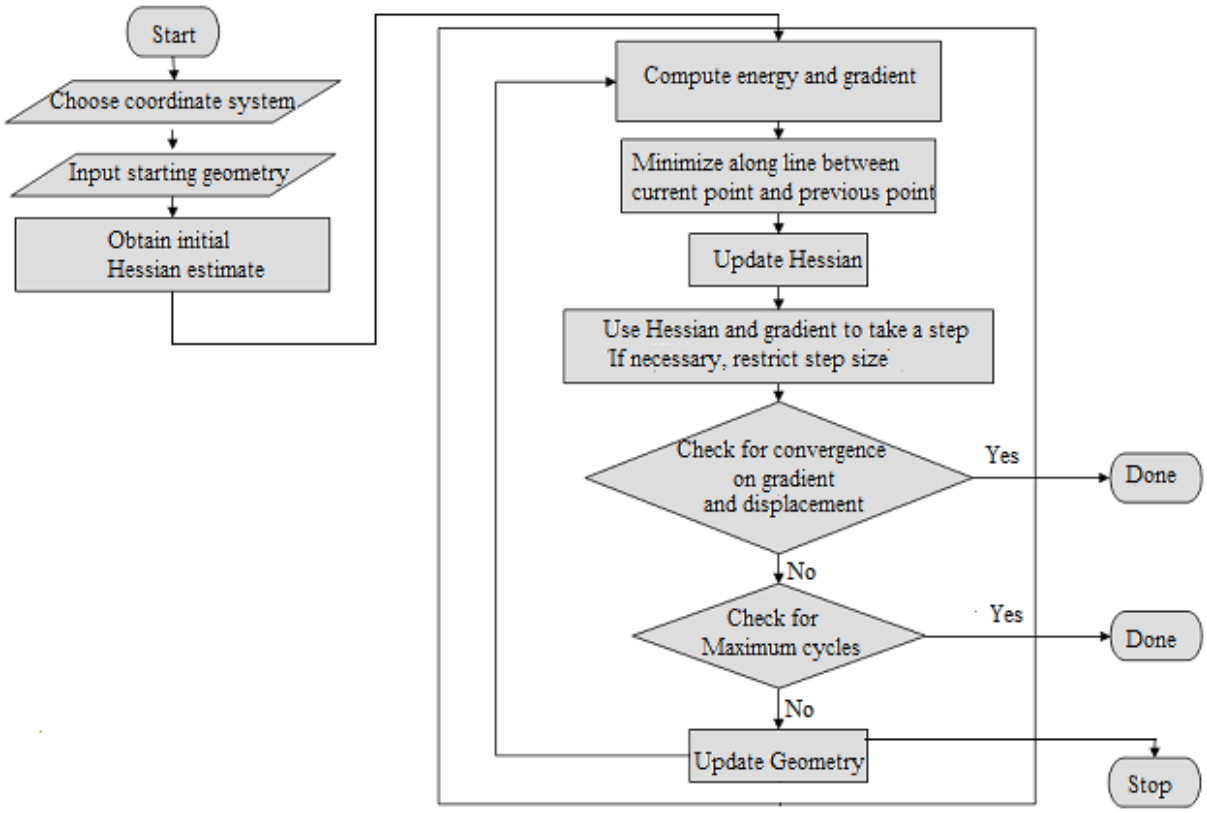


Figure 2.3 The Flowchart for the energy optimization process [55].

In figure 2.3 the flowchart of the energy optimization process is pictured. In this study internal coordinates are chosen in order to avoid problems with rigid body motion. The initial Hessian is found from the GAUSSIAN geometry optimization calculation.

2.5 COMPUTATIONAL TOOLS

2.5.1 Software

In this study, we utilized CHARMM (Chemistry at HARvard Macromolecular Mechanics) which is a widely used molecular dynamics simulation and analysis package written by Brooks at al. in 1983 [56]. CHARMM handles a wide range of molecular simulations for many-atom systems including free energy estimates, molecular minimization, dynamics calculations and model-building.

We also used the Visual Molecular Dynamics (VMD) both as graphic software and for the parametrization of TPT molecule. The Paratool plugin allows generating the force field parameters using the quantum mechanical calculation results. Using the logfiles of the QM simulations, Paratool computes force field parameters for bonds, angles dihedrals and impropers by transforming the Hessian into internal coordinates and list them in the topology and parameter files.

The graphical plotting is done via gnuplot program. Moreover, we wrote several Linux shell scripts for job managements and data manipulations.

The quantum mechanical electronic structure calculation of TPT is done using the GAUSIAN program which is a commercial quantum chemical software package. The calculations ran using B3LYP method in the 6-311+G(p,d) basis set. In Gaussian every calculation is named as job and we conducted three kinds of jobs: single point energy calculations (SP), geometry optimization (OPT) and vibrational frequency calculations (FREQ) are held for TPT molecule.

2.5.2 Hardware

We have a computer cluster with an infiniband network named as “ulubatli” having a total of 72 CPUs and a total of 120 GB RAM memory. Ulubatli uses the Load Sharing Facility (LSF) for the job management and Cluster Management Utility (CMU) for the cluster management. The CPUs are Intel Xeon 5160 dual core and 5462 quad-core which are connected through a linkage, *infiniband*, with a maximum speed of 40GB/s. The disk storage unit, HP MSA 30, has a total disk capacity of 4.2 TB. Ulubatli is funded by The Scientific & Technological Research Council of Turkey (TUBITAK). The support is documented with 107T209 as the grant number.

2.6 DATA STRUCTURE FILES

The CHARMM program requires data structure files to initiate the molecular dynamic simulations and at the end lists the results again in those files. Each type of the files includes

different kind of information about the molecule. Topology and parameter files have the information of a particular class of molecules. The protein structure files, on the other hand, contain the information of the specific molecule which consists of molecules which are defined in the topology and parameter files.

2.6.1 Residue Topology File (RTF)

This file defines some basic molecules such as water and the residues which are building blocks of the molecules such as amino acids or nucleic acids. The name of the atom types consisted with the parameter set, mass of the atom, hydrogen bonds, donors, acceptors and the partial charge of the molecule is contained in these files. In addition, the covalent structure for each residue is included to present how an atom connects to the other to form the amino acids, nucleic acids or lipid molecules [57].

2.6.2 Parameter File (PARAM)

The parameters required for energy calculation is stored in the parameter files. The file determines the bond, angle and torsion force constants and Van der Waals parameters together with the Lennard-Jones radius and the well-depth [57]. CHARMM has standard values for these parameters for nucleic acids, lipids and proteins.

2.6.3 Protein Structure File (PSF)

The protein structure file is the essential data file created internally by CHARMM. The PSF is specific to the system and holds the information in the RTF together with the detailed composition and connectivity, i.e. the topology of the molecules. Every bond, bond, torsion and improper torsion angle is listed. The information in the PSF is needed to generate Hydrogen bonds and non-bonded list [57]. The molecular entities are defined as segments which can be one or multiple macromolecular chain.

2.6.4 Coordinate File (CRD)

The Cartesian coordinates of all atoms in the system received form the x-ray or nuclear magnetic resonance crystal structures is held in CRD files. The HBUILD keyword of

the CHARMM program generates the hydrogen atoms in the system. In addition, the coordinates which could not been determined experimentally can be generated by internal coordinate calculation of the CHARMM program.

2.7 SYSTEM PREPARATION

We have two different systems to prepare; the system including the TPT molecule and the system without TPT. The procedure of the molecular dynamic calculations for each system is the same.

2.7.1 Generating the PSF

The first step of the CHARMM calculation set-up is generating the PSF. Our first input includes the commands used for reading the coordinates from the protein databank (PDB) file and generating the necessary missing coordinates. The PDB file contains the experimentally generated coordinates of crystal structure of the system. The protein databank entry code for our system is 1K4T [12]. The system initially consists of human DNA topoisomerase I in covalent complex with a 22 base pair DNA duplex. The TPT molecule will be added after the calculations without TPT is finished. CHARMM generates the required data structure files from the PDB file. Figure 2.4 represents our system in its crystal structure form. The water molecules without the Hydrogen atoms which will later be generated are shown as yellow dots.

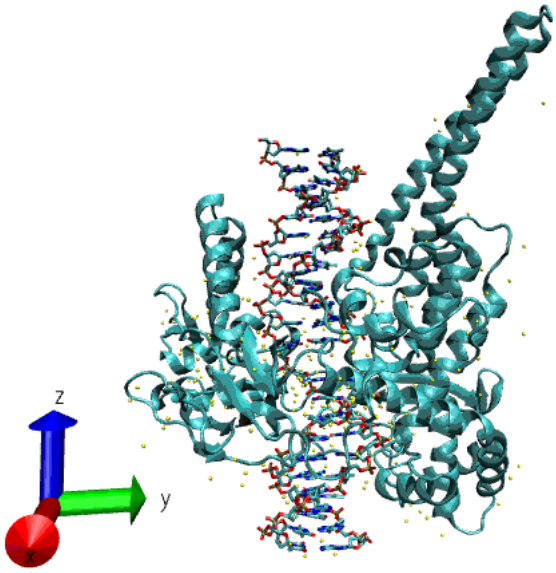


Figure 2.4 Protein Databank entry of DNA Topoisomerase IB in covalent complex with DNA duplex (1K4T)

The starting input reads the RTF and PARAM files to define the atoms listed in the PDB file. Using the appropriate CRD files five segments has generated; protein with a segment id (segid) MAIN, DNA duplex upstream with segid DNAU, DNA duplex downstream with segid DNAD, DNA duplex intact strand with segid DNAI and finally the water molecules with a segid WAT. After this definitions and generations the data structure files for the system is written in PSF, CRD and PDB format.

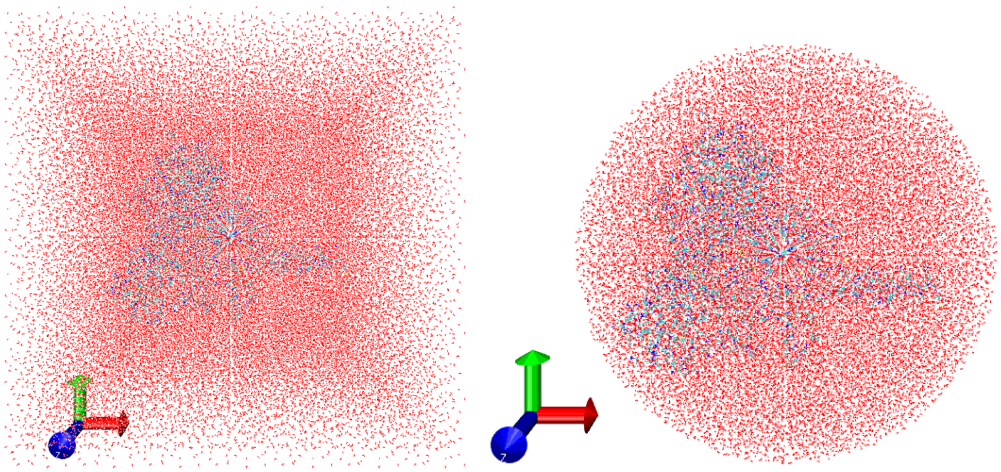
2.7.2 Preparation for Dynamic Calculations

The step after generating the system is preparing the system for the dynamic calculations. This step involves minimization and heating up the system.

2.7.2.1 Energy Minimization

The minimization input involves an energy calculation setup. In addition to the topology and parameter files, the data structure files written on the previous step is read as an input. We applied a cutoff of 14 $\overset{0}{A}$ with a SWITH truncation function on both Van der Waals and electrostatic interactions after an Adopted Based Newton-Raphson (ABNR)

minimization. ABNR minimization performs Newton-Raphson method, i.e. calculates the derivative of gradients on a subspace of the vector displacements. The saddle points of the energy surface are given by the numeric calculation of the derivative matrix and eigenvector analysis of the change in the gradient vectors. At each step the residual gradient is contributed to the next Newton-Raphson step as a steepest descent [57]. The output data structure files are written at the end.



(a) (b)

Figure 2.5 The system in the cubic (a) and the spherical (b) water box.

The next step is to put the system into the water box to simulate the conditions in the cell. The geometric shape of the water box would be sphere. In order to create a sphere water box we first generate a cubic volume and then we removed the water atoms out of the radius of $65\,\overset{0}{\mathrm{A}}$. (figure 2.5). At the end we found out that the total charge of the system is -20. We added sodium (Na^{+}) ions to neutralize the charge of the system (figure 2.6).

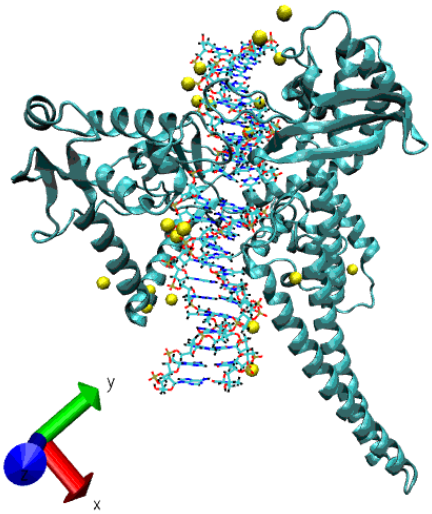


Figure 2.6 The system after the Na^+ ions (shown in yellow) added.

Before going on with the further steps the system with the water molecules and the sodium ions added is minimized both with ABNR and Steepest Descent (SD) algorithms (Figure 2.7). The SD minimization adjusts the coordinates in the negative direction of the gradient at each step of the iterative calculation.

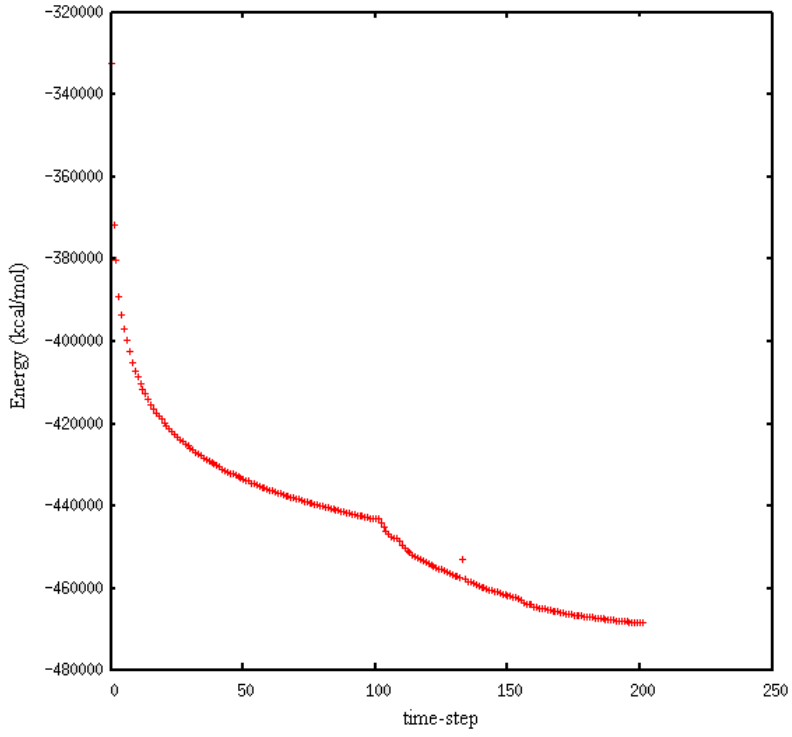


Figure 2.7 Change in the energy after 1000 step SD and ABNR minimization

2.7.2.2 Heating up the system

The minimized structure of the system is at a temperature of near to absolute zero. Heating process involves assigning initial random velocities according the Maxwell-Boltzmann distribution for low temperature and increasing gradually to the greatest velocities at predetermined time intervals.

We used a 100.000 steps leap-verlet algorithm with a time step of 0.001 ps. This heating process functions as an early equilibration also and produces the restart files containing the most recent coordinate sets to use as an input for the actual equilibration. The system was heated up to 300 K^0 (Figure 2.8) and the duration of the job was 3.78 CPU seconds at Ulubatli.

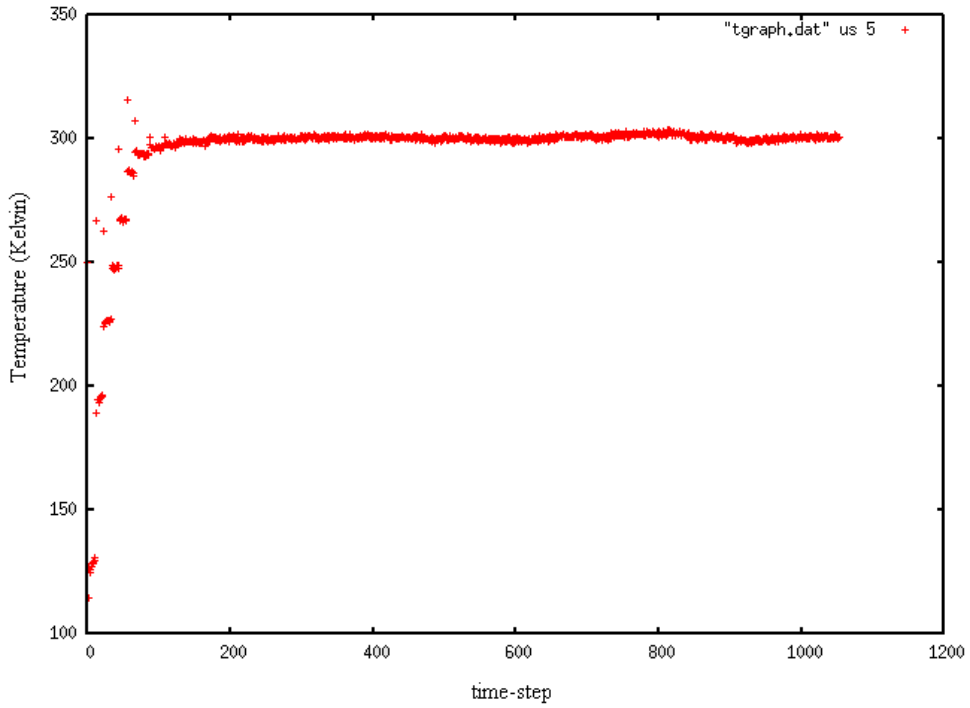


Figure 2.8 Increase in the heat of the system after applying early equilibration.

2.7.3 Equilibration

Equilibration is integrating the equations of motion during a period of time until the average temperature and the structure of the system becomes independent of time, i.e. remains stable. We ran a 100.000 step - CPT dynamics with a step size of 0.002 ps using the restart files of the early equilibration. Again a restart file is written which will be used in the final

equilibration. At the final equilibration input all the dynamic keywords are the same except the SHAKE command which is employed to restrict hydrogen bond lengths with a tolerance of 1OD - 8. SHAKE command allows concentrating on the significant frequencies by eliminating the least important figures, the hydrogen bond stretching terms. Figure 2.9 represents the change in the total energy with respect to time after the final CPT dynamics. The final equilibration takes 12.76 CPU seconds at Ulubatli.

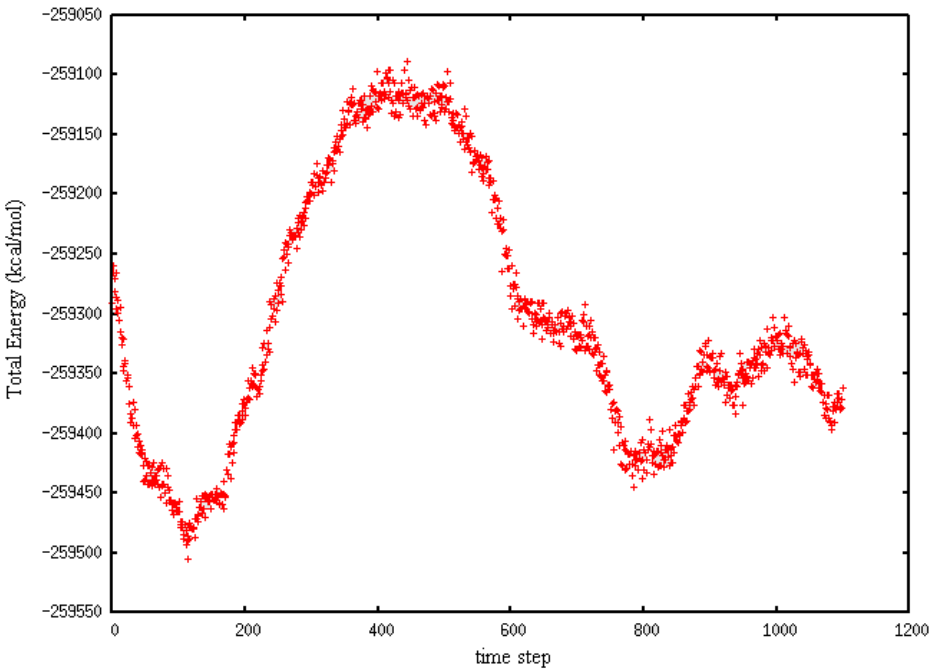


Figure 2.9 The total energy as a function of time.

2.7.4 Preparation for HQBM

We will perform molecular dynamics simulations the nicked DNA strand rotation about the intact strand both with TPT attached and without TPT. The HQBM dynamics best fits this particular simulation by establishing the conformational change without affecting the constants of the motion. The CPT dynamics of 50.000 steps ran with a step size of 0.002 ps. The first step was to define the axis of rotation. We used the most suitable axis (Figure 2.10) which is found by a previous study involving the HQBM dynamics simulations of a DNA duplex [58].

HQBM requires a reference or a target structure and a reaction coordinate which gives the distance from the reference structure. The atoms which we will apply external potential by

HQBM is the backbone atoms of the DNA downstream. We applied the external potential on -2 to -12 base pair backbone atoms of downstream and intact strand. We kept the -1 base pair free.

The HQBM is performed at every 10^0 up to 360^0 completing a full rotation hence we needed 36 different target structures. In order to obtain the target structures, we placed dummy atoms on coordinates of each 10^0 so at each step the system is forced to those coordinates.

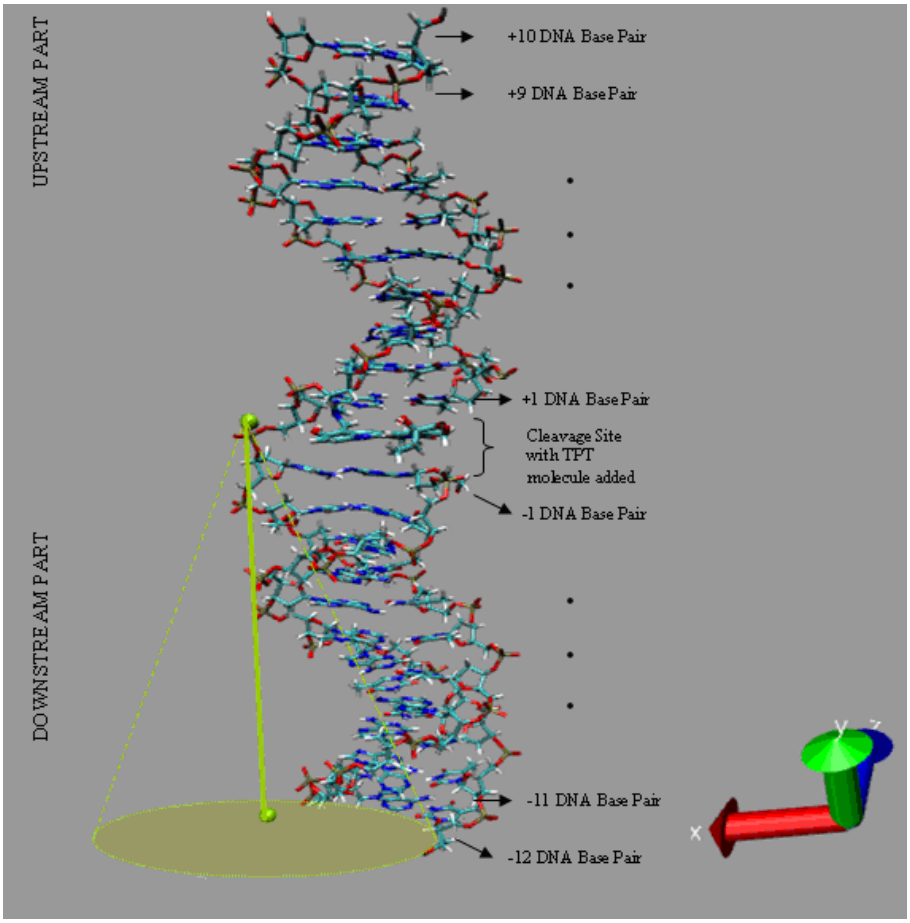


Figure 2.10 The DNA Duplex with a detailed view, the axis of rotation is shown in green.

We made a total of four simulations involving clockwise and counter clockwise rotations of the system with TPT added and without TPT.

2.8 THEORY OF THE POST-SIMULATION ANALYSES

In this section we pointed out the theory behind the measures that we used for the analysis and the geometrical approach that we built for the abstract two dimensional graphics of the path of the DNA rotation.

2.8.1 Root Mean Square Deviations

Structural comparison between two structures is held by the root mean square deviations (RMSD) of the distance between them. In the case of our study, the two structures are the ones at the initial and the final stages of the motion. If we call the initial and final Cartesian positions of an atom in the molecule as $r^{(i)}$ and $r^{(f)}$ respectively, and number of atoms as N the RMSD would be as follows:

$$d_{if} = \sqrt{\frac{1}{N}\sum_{k}\left[r_k^{(i)} - r_k^{(f)}\right]^2} \qquad (2.35)$$

where k is the index over the all atoms in the molecule. RMSD is a popular and reliable measure for the comparison of the similar structures.

2.8.2 Interaction Energies

The total energy of the molecule is defined in the Section 2.1 as the summation of the bonded and non-bonded energy terms. On of measure for the activity of the protein is the amount of interaction of the specific part of the protein with the DNA duplex, i.e. most energetic segments of the protein is said to be actively involved in the relaxation mechanism. The energy is expected to be mostly in non-bonded terms such as Van der Waals and Coulomb energies which will be mentioned in Chapter 3.

2.8.3 Rotational Path Analysis

We set up a geometric and mathematical approach in order to produce a two dimensional abstract picture of the rotational motion. figure 2.11 shows the scheme for the

formulation. The part (a) of the figure shows the full hypothetical rotation of the DNA from the top view. The HQBMD runs at every 10^0 of rotation so we have 36 different reaction coordinates $\sqrt{\rho_0}, \sqrt{\rho_1}, ..., \sqrt{\rho_{36}}$. The (b) part of the figure shows the not to scale magnification of the first 20 degrees of rotation. The pink dot symbolizes the DNA duplex downstream from the top view. We need the r value, the radius of rotation in order to visualize the rotation path. The initial reaction coordinate, $\sqrt{\rho_0}$ is given by the calculation set up as 1.53 $\overset{0}{\text{A}}$.

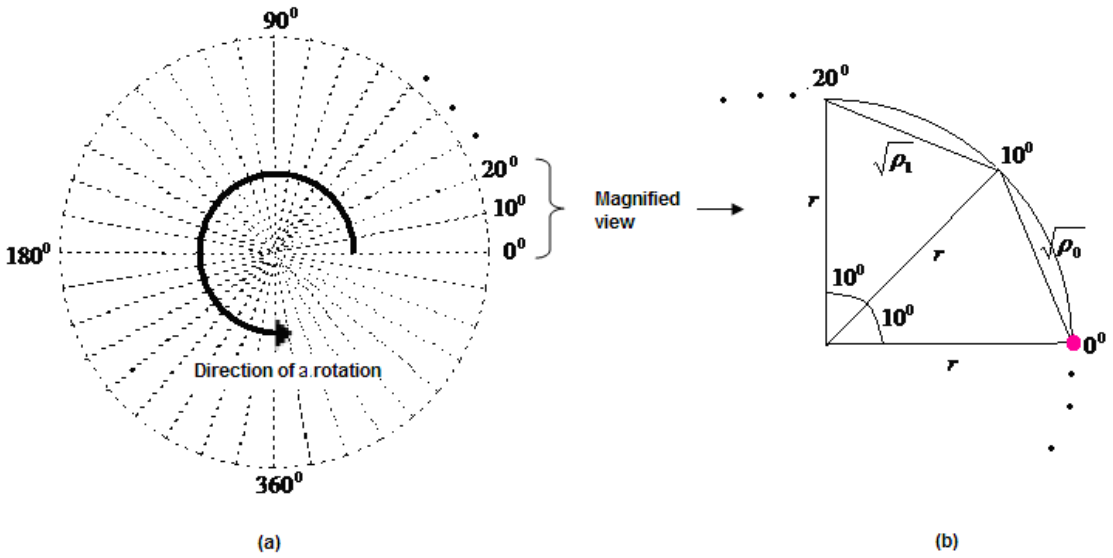


Figure 2.11 The preparation for the abstract path graph.

We calculated the value of r using the set-up value of the reaction coordinate from law of cosines:

$$\left(\sqrt{\rho_0}\right)^2 = 2(r^2 - \cos 10) \qquad (2.36)$$

which yields the radius r to be 8.86 $\overset{0}{\text{A}}$. As the calculation proceeds the next reaction coordinate is given as a result of each HQBMD step with which we could locate the DNA molecule in the Cartesian coordinates.

CHAPTER 3

RESULTS AND DISCUSSION

As we explained before, four different simulation systems has been prepared and dynamic simulations are carried out on each system. One of the main focuses of interest here is to investigate how the mechanism of supercoil relaxations is affected by the inclusion of the drug molecule Topotecan. Therefore, we have analyzed some structural and energetic behavior of these systems and compare them among each other. In the structural analyses, local deformations in each secondary structure of the protein, amount of rotations of the protein during the DNA relaxations, the opening of the 'lips' and stretching of 'hinge' regions have been analyzed. In regards to energetic analyses, the amount of forces and potential energies that we applied to bring the DNA rotations within the enzyme and the interaction energies between DNA and each secondary structure of the protein have been examined.

In addition to aforementioned analyses, the paths followed by the DNA atoms have been plotted as an abstract 2-D graph. In this graph both the hypothetical path in the case of a completely free rotation and the actual paths observed in the simulations are given. The abstract graph is realized by calculation the radius of the DNA rotation in our HQBMD set-up from the length of a 10 degree rotation, which is about 1.6 Angstrom (see equation 2.36). The mathematical details of this analysis are given before, in Section 2.8.3.

In all our discussion, we will focus on the structural and energetic behavior of $\alpha 5$ and $\alpha 6$ known as 'nose-cone' helices, linker region and openings of the 'lips' and 'hinge' regions. All these structural elements are observed with a high activity during the catalytic and relaxation mechanisms. The nose cone helices and linker domain is described before in Section 1.3.1. The helices $\alpha 5$ and $\alpha 6$ are belong to the core subdomain II and I respectively.

The linker domain consists of two α helices and the nose cone helices are charged and are thought to be electrostatic interaction with DNA downstream. The upper and the lower parts of the cap region constitute the ‘lips’ region where the upper and lower lips are belong to core subdomain I and III respectively. DNA enters through these lips and they closed in order to bind the DNA strand. Meanwhile the opposite side of the lips from residues Leu429 to Lys436 which is termed as hinge [27] is stretched and hinged. It is believed that the hinge region also plays an important role in the negative relaxation process [27].

In the energetic discussion, we computed interaction energy as being the sum of the electrostatic and Van der Waals interactions. The Coulomb interactions are seen to be almost 95 % of the total interaction, as expected. Therefore, in all analyses below, the dominant interaction comes from the negatively charged DNA backbone atoms and charged protein atoms. We also computed the potential energy of the system during the rotational motion.

In Figure 3.1, the upper and lower portions of the graph show the interaction energies of the protein with DNA duplex downstream with respect to the secondary structure organization of the systems without and with TPT respectively. The green line graphs are the initial and the bar graphs are the final interactions of the system with DNA. The secondary structure organization is shown with the colored arrows core-subdomain I with cyan, core-subdomain II with red, core-subdomain III with pink, linker domain with green and the C-terminal domain with yellow.

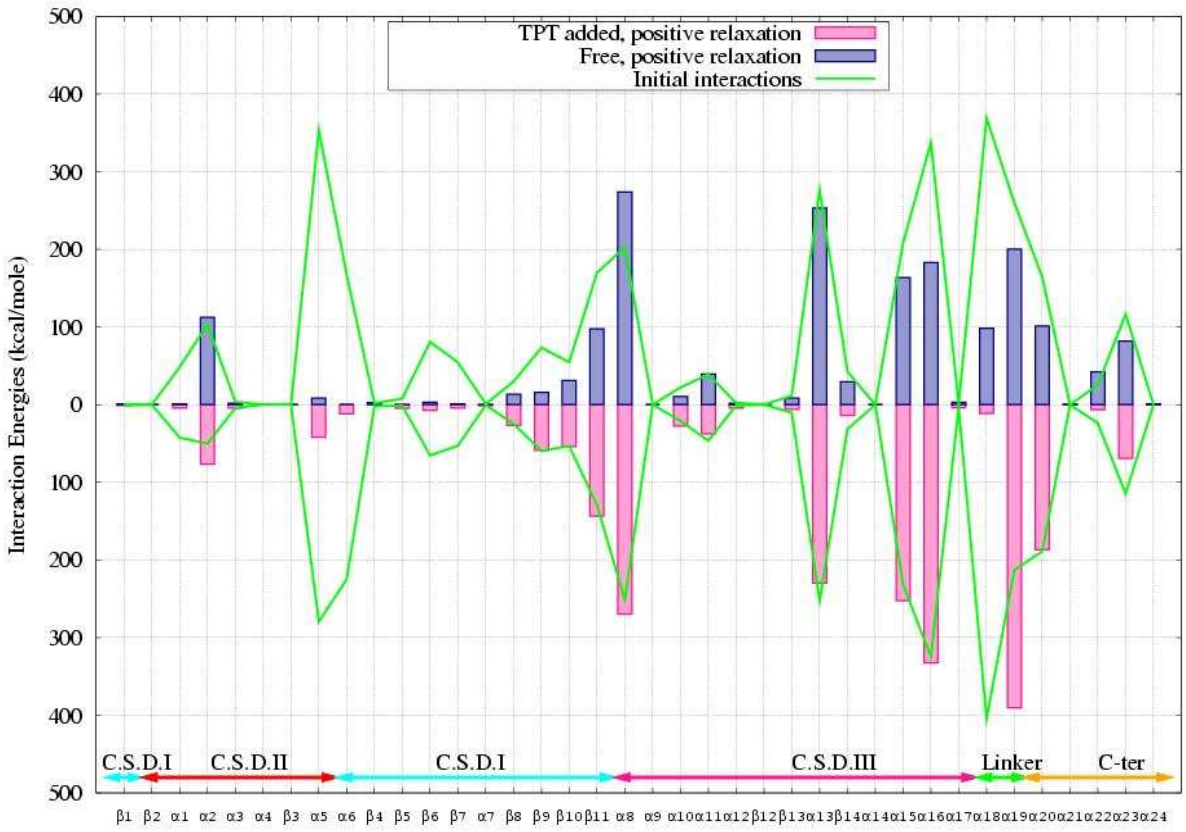


Figure 3.1 The secondary-structure based initial and the final interaction energies between the protein and DNA for the positive relaxation of the two systems.

At the initial and final stage of the positive relaxation the addition of the TPT molecule does not make any effect on the interactions as seen in the graph except for the $\alpha 19$ helix of the linker domain. Interaction energy is increased about an amount of 200 $kcal/mole$ when TPT is added to the system at the final stage for this subunit. This means that the linker activity increased for the positive relaxation.

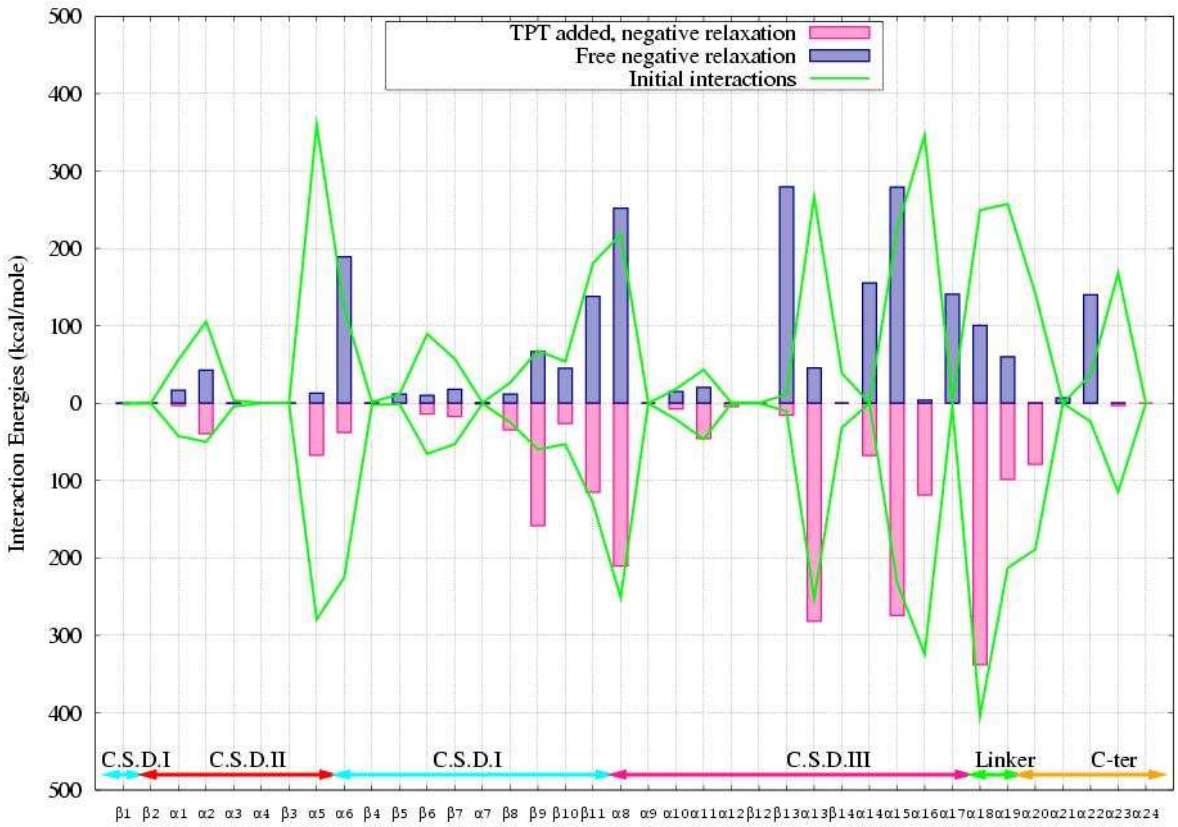


Figure 3.2 The secondary-structure based initial and the final interaction energies between the protein and DNA for the negative relaxation of the two systems.

The interaction energy analysis gives rather significant results for the negative relaxation of the systems. Figure 3.2 represents the interaction energy alterations of the TPT added and free systems with the graphs at the lower and upper side respectively. The color representations are the same with the figure 3.1, pink for the TPT added system, blue for the free system and green for the initial stages of the systems.

There is not a notable difference in the initial interactions of the both systems however the resultant interaction energies varies significantly among the two systems. The interaction energies of the TPT added system get higher than the free system of an approximate amount of 100 $kcal/mole$ at $\beta 9$ sheet, more than 200 $kcal/mole$ at $\alpha 13$ and $\alpha 18$ linker helices. The energies get lower than that of the free system about 100 $kcal/mole$ at $\alpha 6$ and about 300 $kcal/mole$ at $\beta 13$ sheet. At $\alpha 17$ and $\alpha 22$ helices the free system has an energy of more than 100 $kcal/mole$ however the TPT added system does not exhibits any interactions.

Combining figure 3.1 and 3.2 we could infer, when TPT is added to the system, the interaction energy pattern alters for negative DNA supercoil relaxation more than the positive DNA supercoil relaxation hence TPT is more effective to change the interaction energies between DNA and protein for negative relaxation.

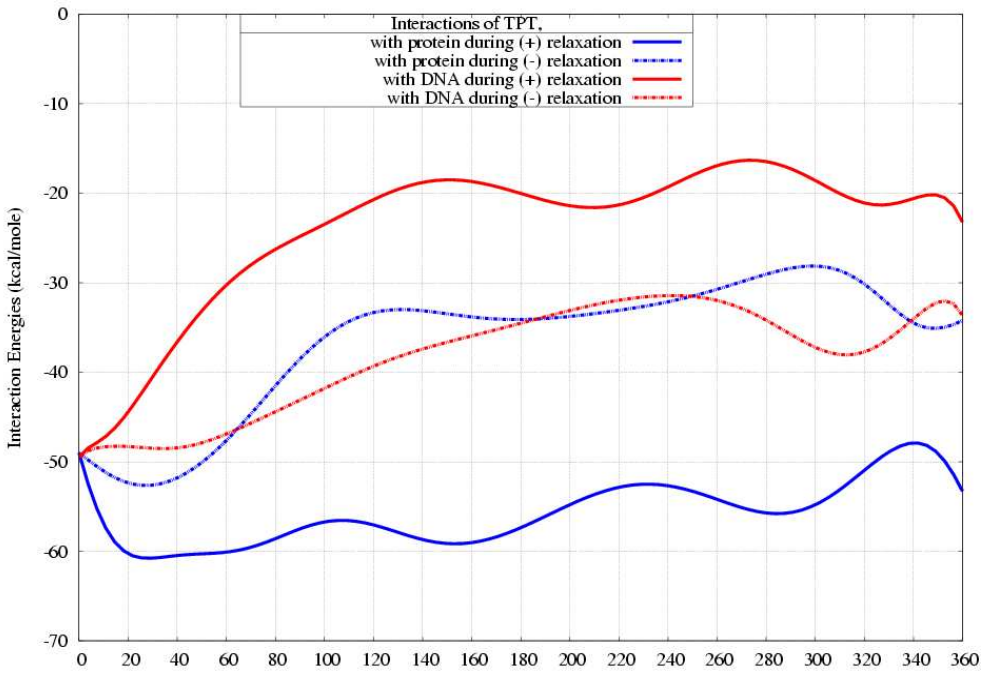


Figure 3.3 The interaction energies between the TPT and protein and between TPT and DNA strand.

Figure 3.3 shows how the TPT molecule interacts with the protein and the DNA strand during the motion. The TPT-DNA interaction energies at each 10^0 are represented with the red solid line for the positive relaxation and with the red dotted line for the negative relaxation. In a similar way, the TPT-protein interaction energies are plotted in blue color with the dotted and solid lines referring to the negative and positive relaxations respectively. The TPT-DNA and TPT-protein interactions have more or less the same pattern for the negative relaxation. On the other hand, for the positive relaxation, while TPT highly interacts with protein the interactions with DNA molecule seem to be rather low.

Comparing with the positive DNA supercoil relaxation, the negative DNA supercoil relaxation causes an increase in the interaction energies between TPT and DNA molecule and a decrease in the interaction energies between TPT and protein. This difference may be

because of the change in the distance between the TPT molecule and the DNA strand and the protein during the rotational motion. In figure 3.1 and 3.2 we observed that the protein-DNA interaction energies get disordered for the negative relaxation of TPT added system. This may be because of the increased interaction of TPT with DNA during the negative relaxation. If DNA is electrostatically held by TPT molecule more strongly during the negative relaxation, it is expected to have the interaction energy pattern between DNA and protein change more for the negative relaxation than the positive relaxation as in the figure 3.2.

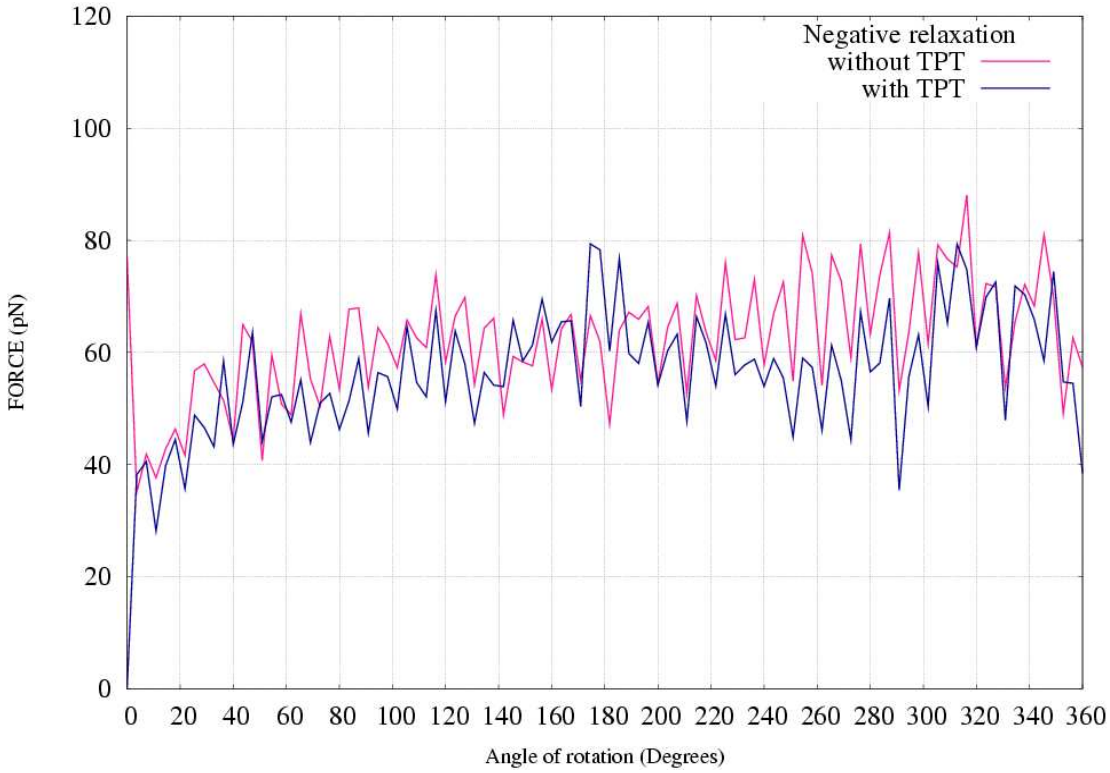


Figure 3.4 The amount of force applied during the negative relaxation

We have also calculated the forces applied through the 360 degrees of rotational motion using the equation 2.21. Figure 3.4 gives the force versus angle of rotation graph of the positive relaxation of DNA in free form and with TPT added in pink and blue colors respectively. Figure 3.5 is the same representation for the negative relaxation of DNA. The force values changes around 60 *pN* for negative relaxation, and 50 *pN* positive relaxations. This is expected, as the relaxation of negative supercoil requires the hinge region to stretch in some extent. The effect of the TPT for both relaxations is similar. Addition of TPT caused a decrease in the force values.

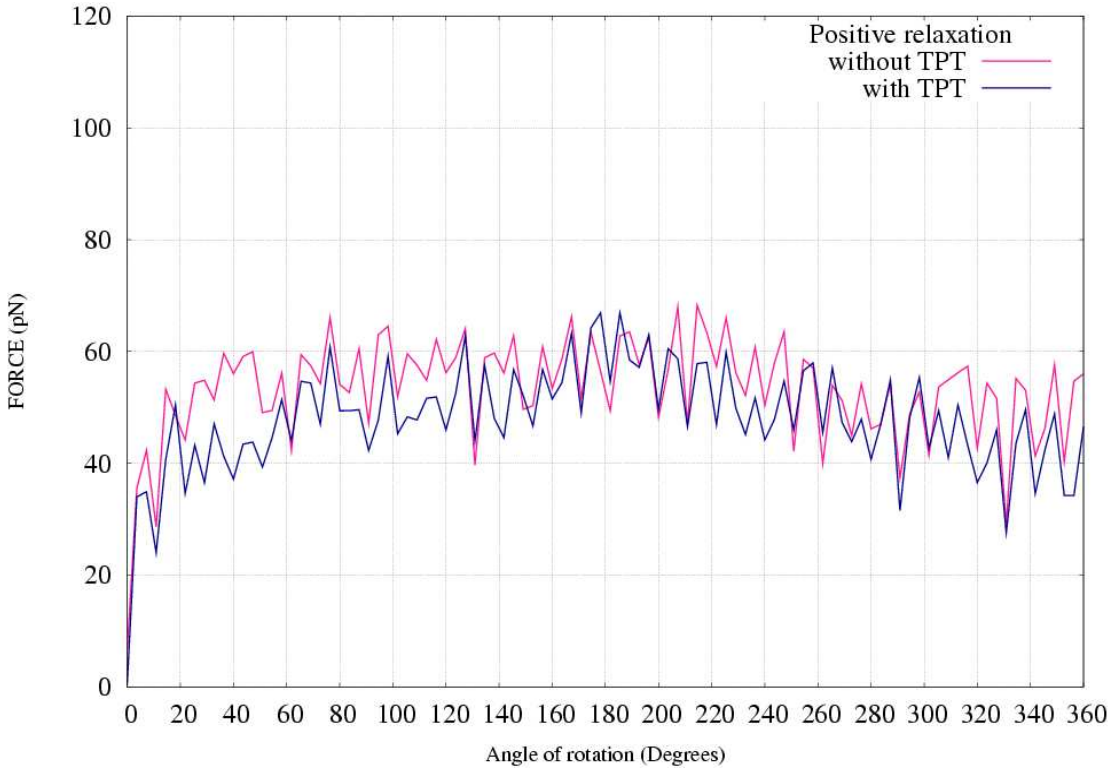


Figure 3.5 The amount of force applied during the positive relaxation

The structural analyses focus on the position variations of the atoms and molecules during the relaxations and the RMS distances between initial and final structures. We have conducted RMSD analyses, plotted the variation of the distances of the distal, proximal and hinge regions and the specific TPT-binding residues.

The root mean square deviations (RMSD) that we obtained upon superposition of the final distances of the secondary structure of the protein are plotted in figure 3.6. The blue and pink bar graphics indicate the RMSD of the systems with TPT in positive and negative relaxations respectively. The green line graphic is the RMSD of the systems without TPT in both directions. The both sides of the zero axes are positive and the unit of the y-axis is Angstrom. The domain organization is same with the figure 3.1.

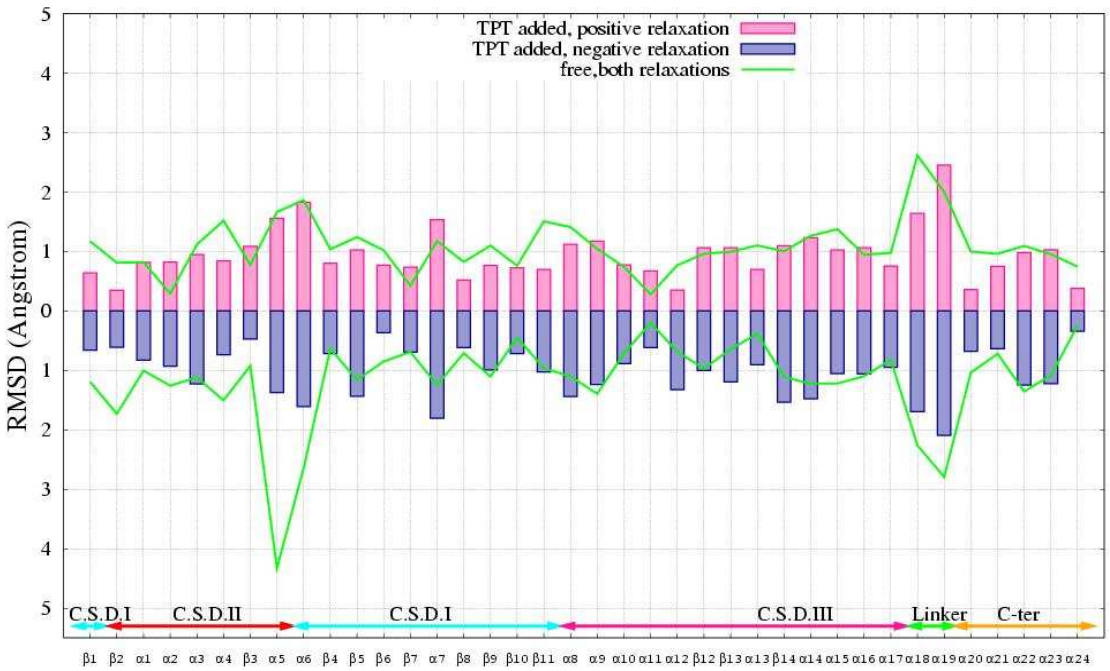


Figure 3.6 The RMSD of the secondary structure of all four systems

The most important point we see from the RMSD analyses is that TPT addition does not make any appreciable change in deformations of the protein of positive DNA relaxation, as seen in the graph. However, for the relaxation of negative DNA supercoils, we see that the TPT addition give rise to very small RMSD values for $\alpha 5$ and $\alpha 6$ which belong to an active part of the protein called as cap, as compare to the wild type case with a difference of more than $2\overset{0}{A}$. This is consistent with the figure 3.2 where the interaction energies of the secondary structure and DNA given. The energies of the cap region drop dramatically from the initial values which show that the caps region loose its activity for the negative relaxation with TPT added.

As described in Section 1.3.1 the experimental studies [10] indicate some segments of the protein actively involve in the relaxation mechanism. These two opposing segments are called as ‘hinge’ and ‘lips’ through where the DNA takes its position. The ‘distal’ and the ‘proximal’ clamps are the components of the lips region. These important regions constitute a significant part of our analyses in order to gain an insight on the effect of the TPT in the mechanism. Previous studies indicate that the lips region opens during the positive relaxation and hinge region stretches during the negative relaxation. Figure 3.7 and figure 3.9 show the

amount of opening and stretching of these particular protein segments through the rotation of negative and positive relaxations respectively. The distance values of distal, proximal and hinged clamps are represented in green, red and blue respectively. The dotted lines show the system without TPT i.e. wild type and the solid lines show the system with TPT.

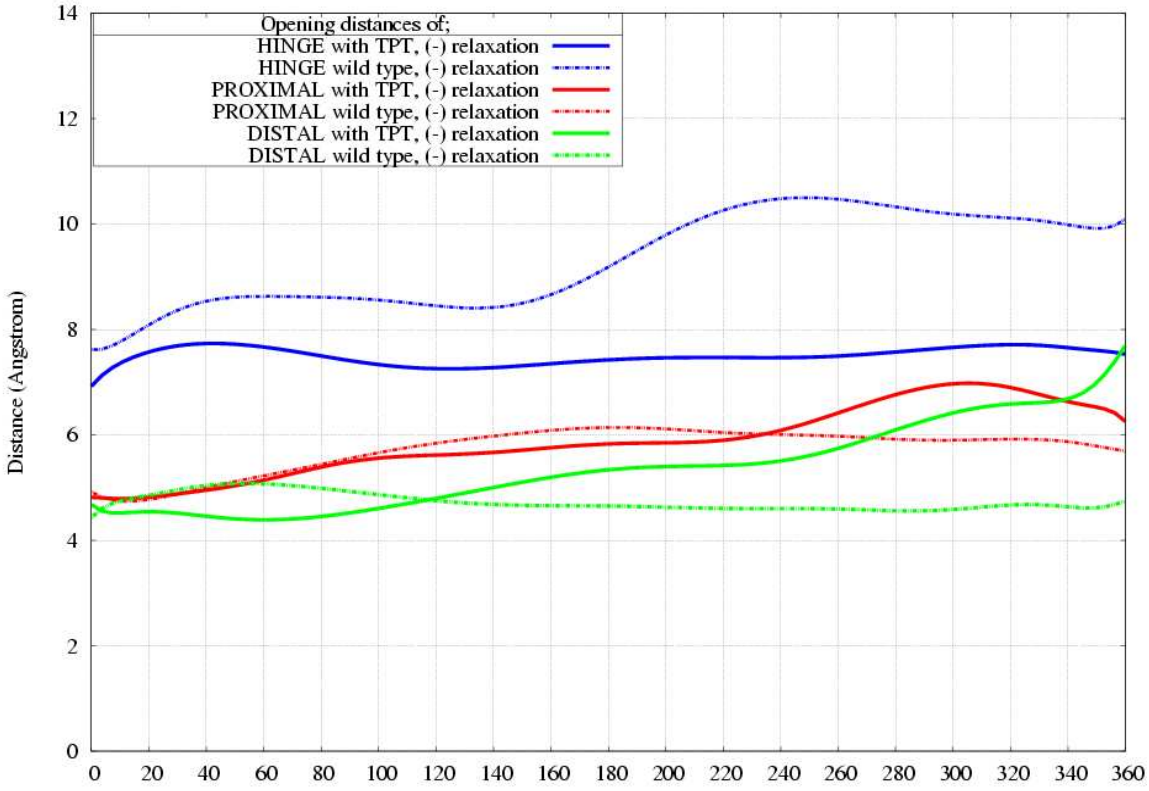


Figure 3.7 The amount of opening of the ‘gates’ of the negative relaxation

The addition of TPT mostly affected the stretching of the hinge region for the negative relaxation in figure 3.7 with an approximate amount of 3 Angstrom decrease. The distal and proximal regions open a little bit more than the wild type. So the hinge region stops stretching during the negative relaxation when TPT added which is consistent to the energetic data. The $\alpha 8$ helix where one of the hinge residues belongs lowers its interaction energy by 100 $kcal/mole$ for the TPT added system during negative relaxation. Figure 3.6 is also supports the decreased activity of hinge during negative relaxation in the system with TPT by showing the decreased RMSD value for the $\alpha 8$ helix.

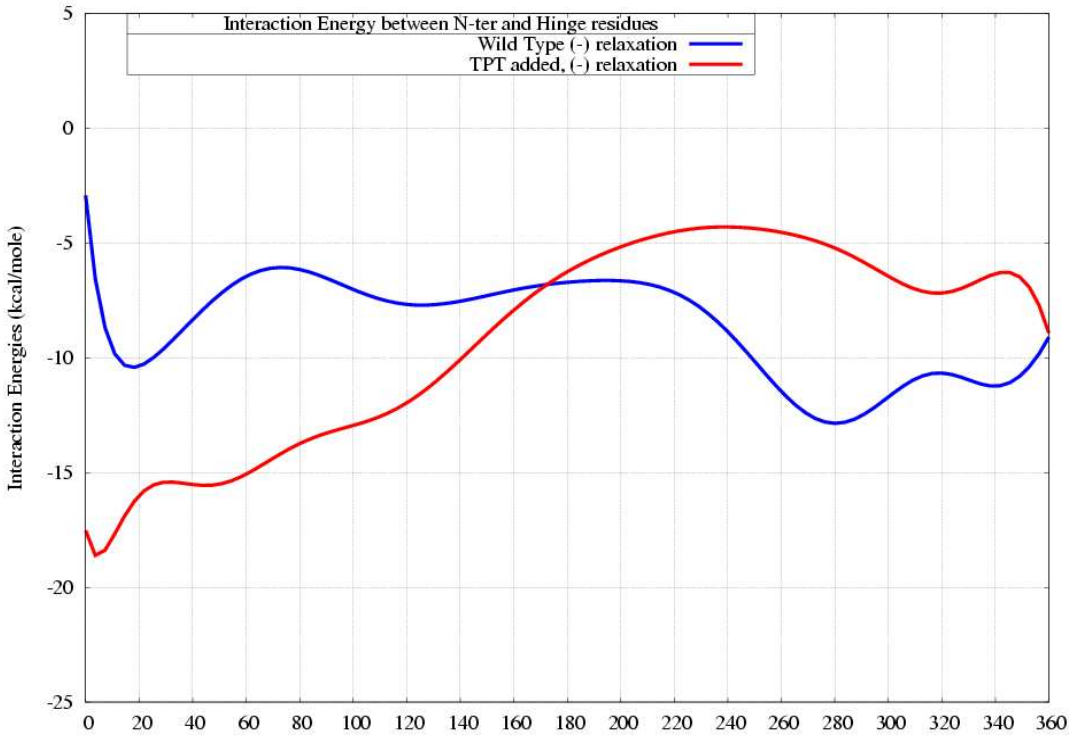


Figure 3.8 The interaction between N-terminal domain and Hinge region

The hinge region stretches about 12 A in the relaxation of negative supercoils for the wild type mechanism [27]. However, we have observed here that the same 'hinge' region does not stretch both for the wild type and for the systems with TPT molecule. Therefore, this distinct behavior can not be attributed to the existence of TPT, as the *wild type* here also does not let hinge stretching. When we investigate the possible reasons for this, we have realized that the additional 12 residues in the N-terminal region we have here now (it was absent in the set-up of Sari et. al. [27] interact directly with the hinge region. As seen in the figure 3.8 the hinge region has an appreciable amount of interaction energy with the hinge region, and this interaction is preserved more or less during the DNA rotations. Also we have observed that these new 12 residues of N-terminal region have many atoms which are quite close to those of the hinge region. Therefore, we have concluded that the N-terminal region of the protein is important in the relaxation mechanism of negative supercoils, a result which is also deduced indirectly by Sari et al. before.

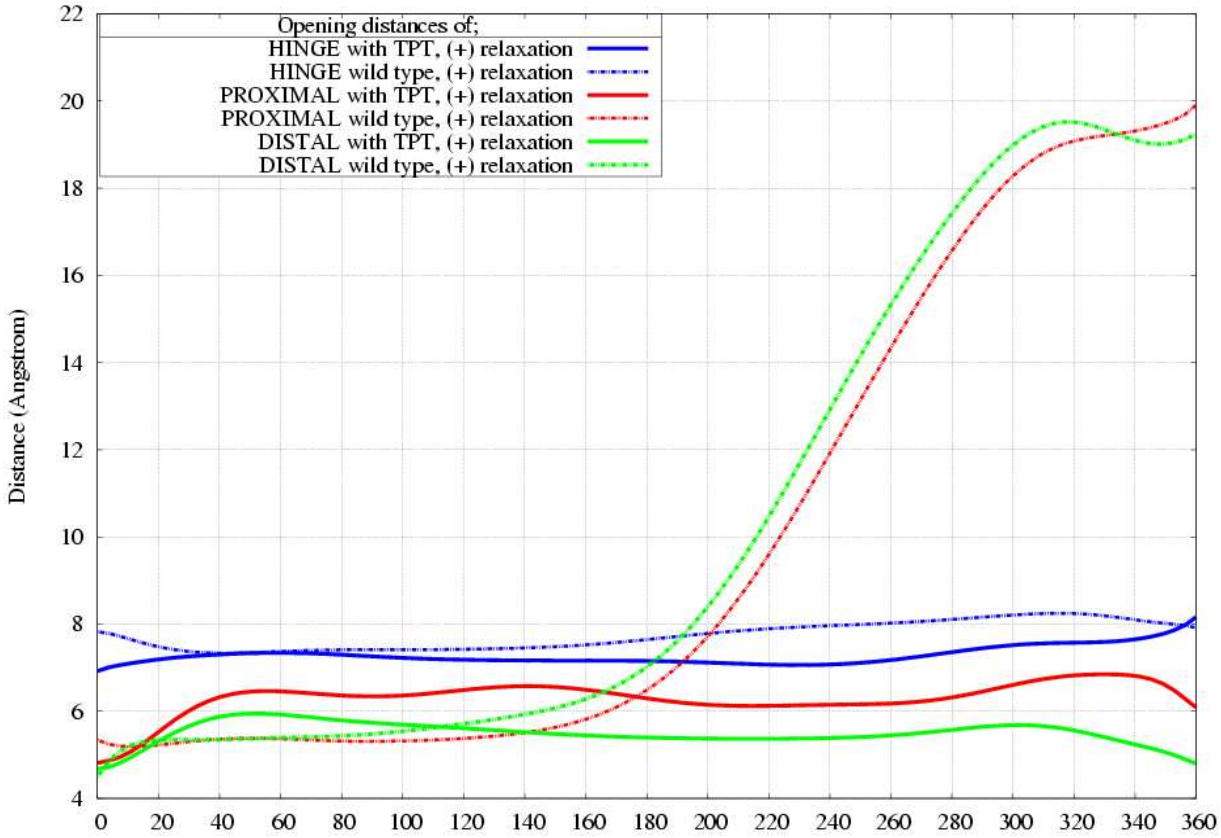


Figure 3.9 The amount of opening of the ‘gates’ of the positive relaxation

Figure 3.9 show that TPT addition has a weak impact on the activity of the hinge region for the protein during the positive relaxation. However the distal and proximal regions exhibit one of the most significant results of the analyses, they turned out to be inactive in the TPT added system. The large, approximately 15 Angstrom distance shows that the lips does not open during the TPT added positive relaxation. Lips region opens during DNA binding and unbinding. The addition of TPT to the system diminish the activity of the enzyme hence reduce the control of it over the relaxation mechanism.

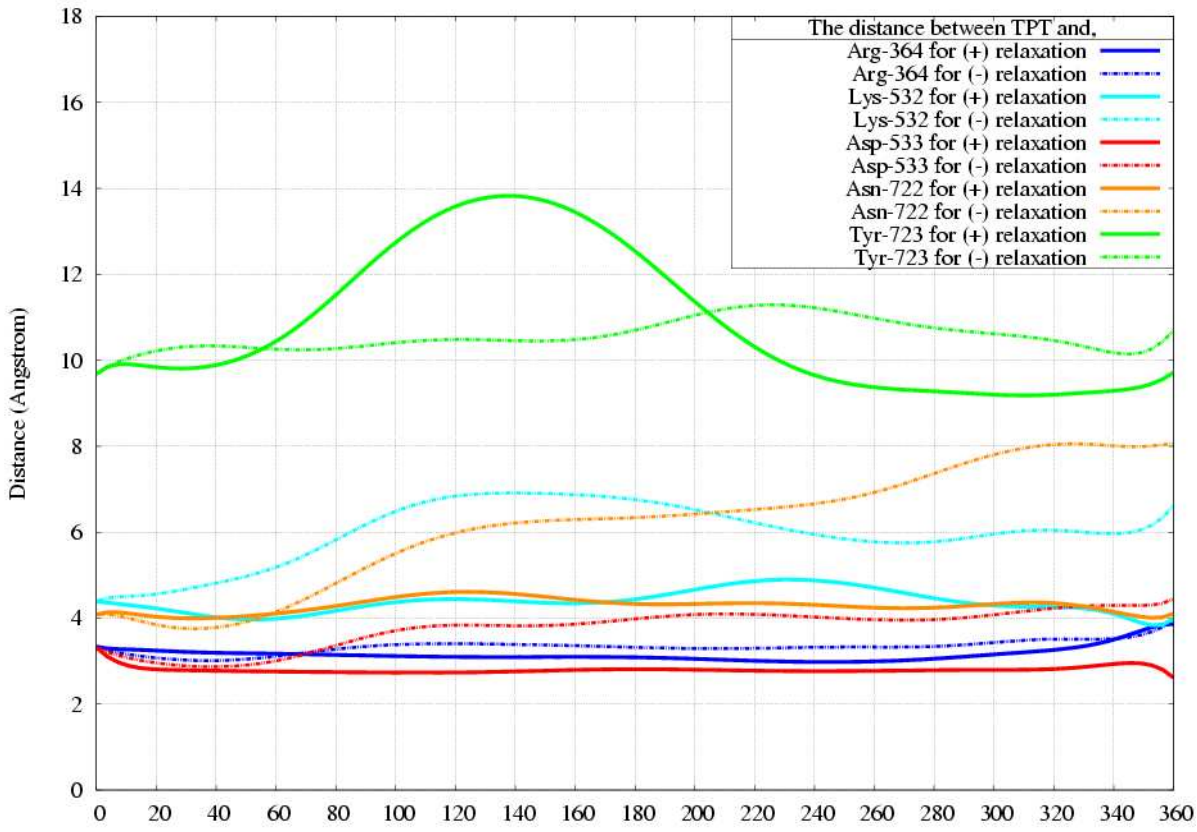


Figure 3.10 The distance between the TPT and the residues interacting with it.

The previous experimental examinations as stated in Section 1.4, there observed to be 5 residues which are bonded to the TPT molecule. The only direct hydrogen bond occurs between the Arg-364 and TPT. The rest is constructed through the mediating atoms like the oxygen atoms in water molecules. The variation of distance of these residues between TPT could give a clue about the action of TPT during the both relaxations mechanism. Figure 3.10 where these distances are plotted shows Arg-364 in blue, Lys-332 in cyan, Asp-533 in red, Asn-722 in orange and Tyr-723 in green. The dotted lines are for negative and solid lines are for positive relaxation. The distances of residues do not differ significantly for different relaxations except for the Asn-722. The Asn-722 residue is next to an active part of the protein, in the C-terminal domain. The distance between Asn-722 and TPT gets larger for the negative relaxation.

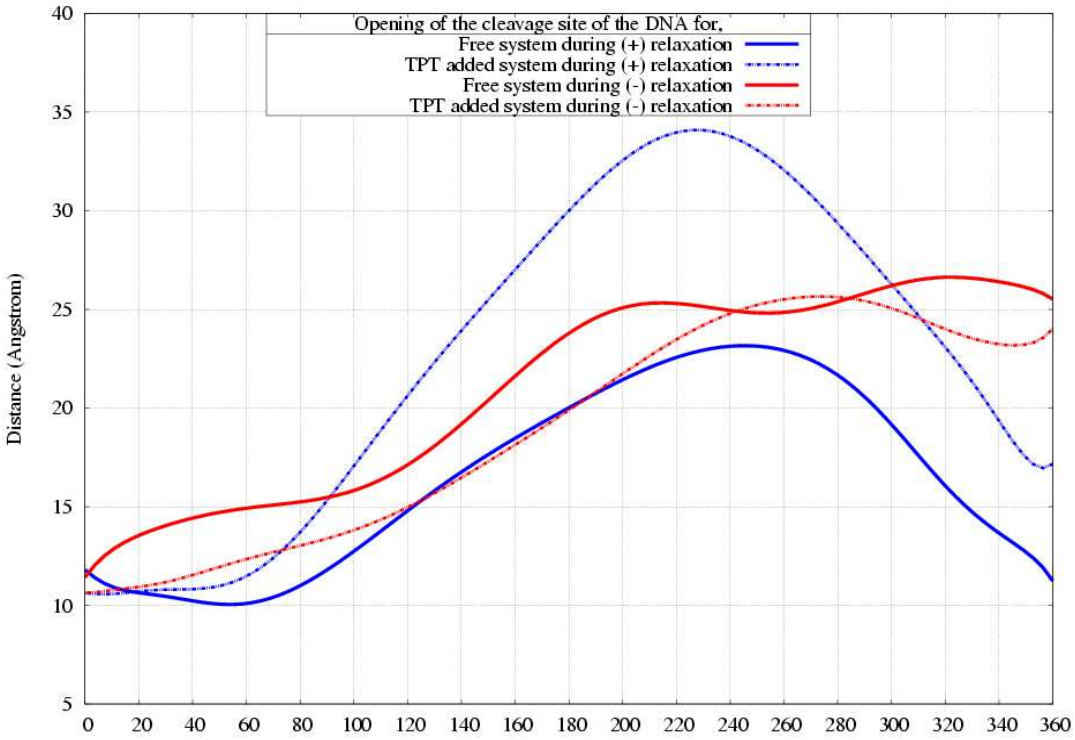


Figure 3.11 The rotation of DNA downstream in terms of distance.

In addition to the behavior of the protein with the addition of the drug molecule, we have also checked how the rotation of DNA is affected. Figure 3.11 represents the rotation of DNA downstream with the distances between the -1 and +1 DNA base pairs. The blue and red colors refer to the free and drug added systems respectively. The dotted lines imply the negative relaxation. The rotation causing the widest opening of the DNA duplex downstream, with approximately 35 Angstrom is the negative relaxation without the TPT molecule. The opening of the DNA for positive relaxation of the same system stops at a value of about 20 Angstrom. The system after the addition of TPT exhibits the same opening manner for the DNA openings which is about 25 Angstrom for both relaxations.

To sum up, for the positive relaxation, DNA downstream opens more than the negative relaxation and after TPT addition the DNA downstream opens less than the negative relaxation. Figure 3.3 where the interaction energy between the TPT and DNA duplex is plotted is consistent with the figure 3.10. For the negative relaxation the interaction energy is higher than the positive relaxation case. This may be due to the less amount of opening of the negatively supercoiled DNA. Since the distance between DNA and TPT do not increase as

large as the positive relaxation, the TPT can interact with DNA more easily than the positive relaxation.

Table 3.1 The data regarding to the orientation of the systems after both relaxations

	Angle	**Axis of Rotation**		
		i	**j**	**k**
(+) relaxation	25.35	-0.373548	-0.172937	-0.911348
(+) relaxation with TPT	24.75	0.515693	-0.349464	-0.782263
(-) relaxation	26.72	0.533227	-0.163211	0.830079
(-) relaxation with TPT	25.29	0.380822	0.148441	0.912656

We have information of the final orientation of the protein in all the systems by calculating the total angle of rotation and the Cartesian coordinates of them at the final stage of the rotational motion. Table 3.1 lists these values for both systems in both relaxations. The addition of TPT for both relaxations alters the orientation of the protein only about 1 Angstrom. In addition the two systems do not demonstrate notable differences in terms of the final position of protein.

In order to see a representative top-view of the motion we drew path graphics. The details of the preparation of these graphs are given at Section 2.8.3. Figure 3.10 (a) and (b) gives the top-view geometric representation of the rotational path of both systems during both relaxations. The green path is the hypothetical full rotation of the DNA downstream. The pink and blue paths refer to the systems with drug molecule added and the free system respectively. The radius of the path is calculated with equation 2.36 as 8.86Angstrom.

Figure 3.12 and 3.13 shows that the addition of the drug molecule TPT does not inhibit the rotation action during negative and positive relaxations. Actually, the system with TPT achieves a rotation more closely to the hypothetical path for both relaxations.

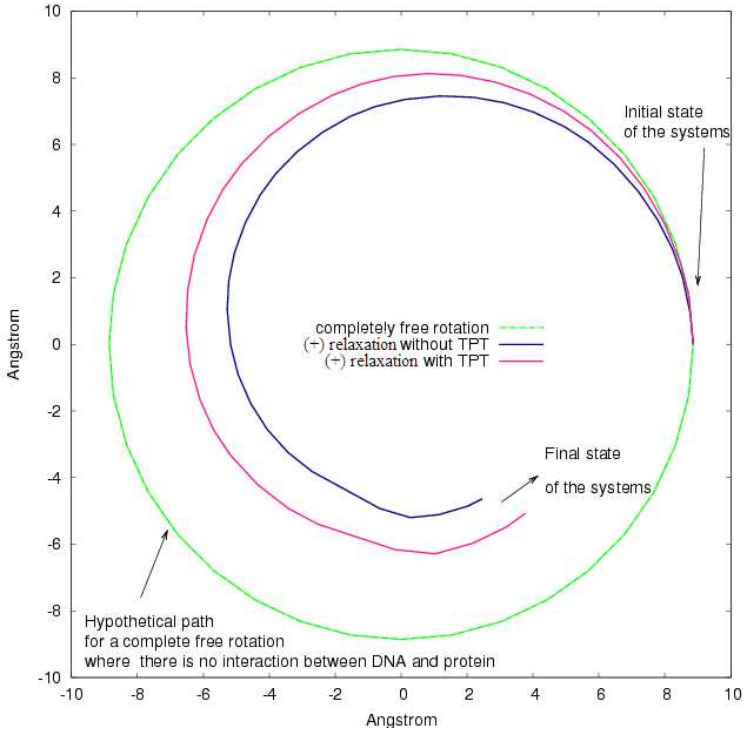


Figure 3.12 The paths of the motion during the positive relaxations for both systems.

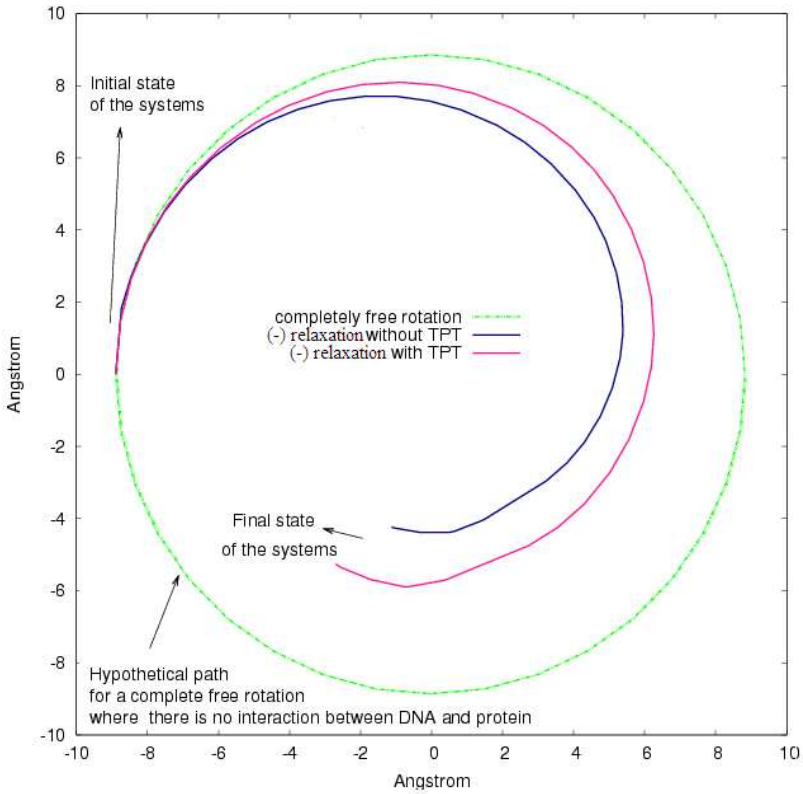


Figure 3.13 The paths of the motion during the negative relaxations for both systems.

We see from the graphs of the paths of the downstream DNA during the rotations, the addition of TPT molecule de-stabilizes the DNA rotations, and causes system to behave more like ‘free-rotation’ rather than the ‘controlled-rotation’ scheme. This is consistent with the above analyses. The free rotation demands less amount of external force than the controlled rotation hence the control of the enzyme via applying external force is not taken into account. Figure 3.4 and 3.5 where the amount of external force during the both relaxations imply that the both relaxation require less amount of external force hence for the TPT added system, the control of the enzyme fall off. Another observation that supports the free-like rotation of the DNA after the TPT addition is the behavior of the lips and the hinge regions. Figure 3.8 and 3.9 points out a loose of activity of these regions after the TPT addition hence the system becomes more free form the control of the enzyme.

In summary, we observed a variety of different and similar behaviors of the systems in our analyses. The most note-worthy results which will be handled in chapter 4 in detail are: During TPT added positive relaxation the hinge does not stretch differing from the wild type. On the other hand, TPT added negative relaxation prevents the opening of the lips and the activity of the caps. The system demonstrates a more free-like rotation after the TPT addition. The TPT addition affects the negative DNA supercoil relaxation more than the positive DNA supercoil relaxation.

CHAPTER 4

CONCLUSIONS

We have found out the following main results in our study:

a-) One of the most important results we got is that, in the relaxations of positive supercoils, the 'lips region' of the protein opens up about 12 Angstrom for the wild type protein (in consistent with Sari et al. [27]), while it does not open up in the system with the anti-cancer drug molecule TPT. The control of the enzyme on the relaxation mechanism lessens and the unbinding process of TPT may be disturbed. This may cause the TPT molecule trap within the enzyme.

b-) The rotation of DNA within the TPT-added enzyme becomes more like 'free-rotation' rather than 'controlled-rotation' scheme. This result is in accordance with most of the analyses presented in Chapter 3. The effect of the protein to the DNA duplex downstream rotation decreases after the TPT addition. The anti-cancer drug exhibits its function not by stopping the removal of the supercoil, instead it allows the relaxation however limits the action of the enzyme and stabilizes the DNA-TPT-enzyme complex. The drug molecules inhibits the enzyme by either preventing the DNA-protein binding or trapping the transient complex. This result supports the experimental studies indicating that the Topotecan poisons the cell by trapping the protein [61].

c-) Interaction energy pattern between the DNA and secondary structures of the protein changes when we add the TPT. This effect is seen to be more dominant in the case of relaxations of negative supercoils. The protein becomes more sensitive to the drug during the negative relaxation of the DNA.

d-) The hinge region of the protein interacts directly with the 12 residues of the N-terminal domain, and this interaction prevents hinge stretching in the relaxation of negative supercoils. The hinge region was observed to stretch about 12 A before by Sari et al [27] wherein they did not have these 12 residues in the simulation set-up. The experimental study of Knudsen et.al. [59] supports the effect of the N-terminal domain on the hinge stretching during the negative DNA supercoil relaxation. Figure 4.1 shows the residues that we believed to have an interaction and affect the relaxation mechanism.

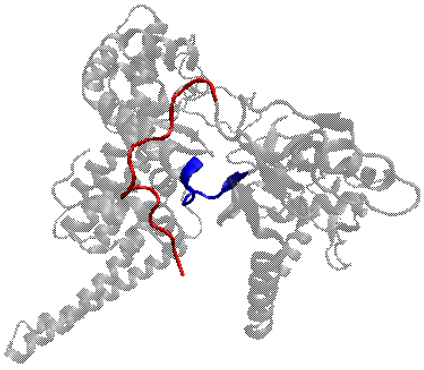


Figure 4.1 The recently added N-terminal domain residues in red and the hinge residues in blue.

e-) Amount of forces need to bring DNA rotations within the TPT-added systems is found to be somewhat smaller than the wild type protein. This observation supports our finding that addition of TPT shifts the DNA rotations to more ‘free rotations’. The fewer amounts of external forces imply the decreased control of the enzyme on the DNA duplex.

f-) TPT has several direct interactions with the protein, as given in the crystal structure [19]. We have observed that these interactions remain unchanged during both positive and

negative supercoil relaxations. The position of TPT molecule is also stable during the rotational motion. Figure 4.2 gives the initial and final states of the (a) negative, (b) positive relaxation in red and blue respectively.

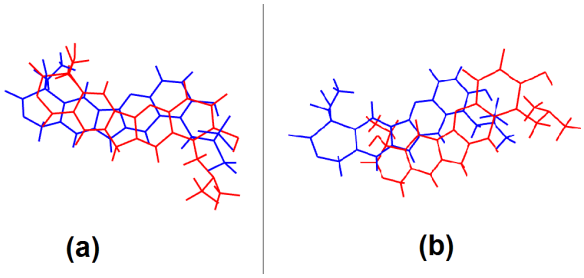


Figure 4.2 The change of the position of TPT molecule

g-) Addition of TPT into the system gives rise to important changes in the interactions of 'nose-cone' helices ($\alpha 5$ and $\alpha 6$) and linker helices ($\alpha 18$ and $\alpha 19$) with DNA. The changes are predominantly on $\alpha 5$ from hinge region and $\alpha 18$ from linker for the relaxations of negative supercoils, while on $\alpha 19$ from linker region for the relaxations of positive supercoils. This supports the conclusions of Sari et al. that the alpha6 and alpha18 are in control of negative DNA supercoil relaxations.

h-) The amount of rigid rotation of the protein during DNA rotations are found to be around 25 degree. This value was observed to be around 80 degree in the previous study. As the current set-up of the systems relies on the continuous velocities in each 10 degree biased MD, the current value should be more reliable.

REFERENCES

[1] J.D.Watson, F.H. Crick, *Nature*, Vol. 171, pp. 737-738, April 1953.

[2] T. Viard, C. Bouthier, *Biochimie*, Vol. 89, pp. 456-467, April 2007

[3] K. D. Corbett, J. M. Berger, *Current opinion in structural biology*, Vol. 13, pp. 15-22, 2003

[4] J.C. Wang, *J. Mol. Biol.*, Vol. 55, pp. 523-533, 1971

[5] F.B. Fuller, *Proceedings of the National Academy of Sciences USA*, Vol. 75, pp. 3557–3561, 1978

[6] Calladine, C.R., et.al., *Understanding DNA*,Elsevier Academic Press, 1998

[7] W. R.Bauer, F. H. C. Crick, J. H. White, *Scientific American*, Vol. 243, 118-133. 1980

[8] V. Rybenkov, C. Ullsperger, A. V. Vologodskii, N. R. Cozzarelli, *Science*, Vol. 277/5326, pp. 690–693, 1997

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

[10] Staker et.al., *PNAS*, Vol. 99, pp.15387-15392, 2002

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

[12] L.Steawart, M.R.Redinbo, X. Qui, W.G.J. Hol, J.J.Champoux, *Science*, Vol. 279, pp. 1534-1540, 1998

[13] M.R.Redinbo, W.G.J. Hol, J.J.Champoux, *Current Opinion in Structural Biology*, Vol. 9, pp. 29-36, 1999

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

[15] J. M. Berger, *Biochimica et Biophysica Acta*, Vol. 1400, pp. 3-18, 1998

[16] M.R. Redinbo, L.Steawart, P.Kuhn, J.J. Champoux, W.G.J. Hol, *Science*, Vol. 279, pp. 1504-1513, 1998

[17] M. R. Redinbo, L. Stewart, J.J. Champoux, W. G. J. Hol, *J. Mol. Biol.*, Vol. 292, pp. 685-696, 1999

[18] A. Das et. al., *Nucleic Acids Research*, Vol. 30, pp. 3794-802, 2002

[19] M.R. Redinbo, J.J. Champoux, W.G.J. Hol, *Biochemistry*, Vol. 39, 6832-6840, 2000

[20] J.J. Champoux et. al., *The Journal of Biological Chemistry*, Vol. 271/13, pp. 7602–7608, 1996

[21] N. Lue, A. Sharma, A. Mondragon, J. C. Wang, *Structure*, Vol. 3, pp. 1315, 1995

[22] B.R. Knudsen et.al, *J. Mol. Biol.*, Vol. 336, pp. 93–103, 2004

[23] L. Stewart, G.C. Ireton, J.J. Champoux, *J. Biol. Chem.*, Vol. 274/46, pp. 32950–32960, 1999

[24] A. K. Bharti, M. O. Olson, D. W. Kufe, E. H. Rubin, *J. Biol. Chem.Vol.* 271(4), pp.1993-7, 1996

[25] L. Stewart, G.C. Ireton, J.J. Champoux, *J Mol Biol*, Vol. 269, pp. 355-372, 1997

[26] J.F. Carey, S.J. Schultz, L. Sisson, T.G. Fazzio, J.J. Champoux, *PNAS*, Vol. 100(10): 5640–5645, 2003

[27] Sari L. and Andricioaei I., Nucleic *Acids Research*, Vol. 33(20), pp. 6621-6634, 2005

[28] D. Esposito, J.J. Scocca, *Nucl. Acids Res.*, Vol. 25, pp.3605-3614, 1997

[29] J. J. Champoux, J. C. Wang, N. R. Cozzarelli, *Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY*, pp. 217–242. 1990

[30] M. Binaschi, F. Zunino, G. Capranico, *Stem Cells*, Vol. 13(4), pp. 369-79, 1995

[31] B. R. Knudsen et. al., *Biochemistry*, Vol. 37, pp. 10815-10827, 1998

[32] Redinbo, M.R. et. al., *J. Mol. Biol.*, Vol. 339, pp. 773–784, 2004

[33] D. Strumberg et.al., *Biochim. Biophys. Acta*, Vol. 1400, pp. 83–105, 1998

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

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

[36] Karplus, M. et.al., *Nature (Lond.)*, Vol. 267, pp. 585, 1977

[37] Gu, Y., *"FBGA Acceleration of Molecular Dynamics Simulations"*, PhD Thesis, Boston University, 2008

[38] Noon, W.H., *"Molecular Dynamics Simulations of Bio-Fullerene Interfaces"*, PhD Thesis, Rice University, 2004

[39] Bert de Groot and Phil Biggin, *Introduction to molecular dynamics simulations*, Unpublished Material, http://www.dddc.ac.cn/embo04/lectures/Bert/other/intro.ppt

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

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

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

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

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

[45] Ciccotti G. et.al., *Mol. Phys*., Vol. 101, pp. 6, 765–778, 2003

[46] H. J. C. Berendsen et.al. , *J. Chem. Phys*., Vol. 81(8), pp. 3684-90, 1984

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

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

[49] M. Born and R. Oppenheimer. *Ann. Phys. (Leipzig)* , Vol. 84 (20), pp. 457, 1927

[50] Guy A. et.al., *J. of Physics*, Vol. 68:1, pp. 69-79, 2000

[51] P. Hohenberg and W. Kohn, *Phys. Rev*. Vol.136, 1964

[52] Kieron Burke et. al. *The ABC of DFT* , 2007

[53] W. Kohn and L. J. Sham. *Phys. Rev*., Vol. 140, pp. 1133, 1965

[54] A.D. Becke (1993) *J. Chem. Phys.* 98: 1372–1377

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

[57] Stote, R. et.al., *CHARMM Tutorial,* http://vit-embnet.unil.ch/MD_tutorial/

[58] Sevim,A., *"Investigation of energetic and mechanistic pathways in the rotation of a nicked DNA"* PhD Thesis, Fatih University, 2008

[60] Knudsen, B.R. et.al., *Nucleic Acid Research*, Vol. 35(18), pp 6170-6180, 2007

[61] Larsen, A.K., et.al., *Pharmacol Ther.*, Vol. 99, pp. 167-181, 2003

[62] Webb, M.R. and Ebeler, S.E., *Anal Biochem.*, Vol. 321, pp. 22-30, 2003