IN SILICO INVESTIGATION OF THE MECHANISM OF RICIN-CATALYSED DEPURINATION REACTION AND DESIGN OF NOVEL RICIN INHIBITORS.

RANGA SRINATH JAYAKODY

IN SILICO INVESTIGATION OF THE MECHANISM OF RICIN-CATALYSED DEPURINATION REACTION AND DESIGN OF NOVEL RICIN INHIBITORS.

The Dissertation Presented to the

UNIVERSITY OF CAPE TOWN

In fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

By

RANGA SRINATH JAYAKODY

MSc. (University of Cape Town), BSc Hons. (Carleton University, Canada)

Supervisor: Professor Kevin J. Naidoo

Department of Chemistry

University of Cape Town

February 2012

DECLARATION

I declare that this thesis titled IN SILICO INVESTIGATION OF THE MECHANISM OF RICIN-CATALYSED DEPURINATION REACTION AND DESIGN OF NOVEL RICIN INHIBITORS is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions.

Signed by candidate

Ranga Srinath Jayakody

This thesis is dedicated to

- *My father, Mr. Ananda Jayakody (1942-2004), for teaching me how to live.*

- *My wife Anusha and my son Yasith, for giving my life its meaning.*

Acknowledgements

First and foremost I offer my sincerest gratitude to my supervisor, Professor Kevin J. Naidoo, who has supported me throughout my thesis with his knowledge and patience whilst giving me the opportunity to work in my own way. I attribute the level of my doctoral degree to his encouragement and effort and without him this thesis, too, would not have been completed or written.

I am indebted to my many colleagues at the Scientific Computing Research Unit (SCRU) for providing a stimulating environment in which to learn and grow. I am especially grateful to Chris Barnett for the informative and interesting discussions as well as for the help as the system administrator; to Jestin Mandumpal and Karl Wilkinson for proofreading, valuable suggestions and for sharing their experiences in writing up a PhD, to Kyle Fernandes for "scripts and bits", to Werner Crous for interesting discussions.

I am also thankful to Dr. Gerhard Venter for his help in many different ways in this work.

Louise Bezuidenhout at SCRU deserves a special thank for her great help in getting through the bureaucracy as well as for just being there as a fantastic person and a friend.

I thank the Scientific Computing Research Unit for the opportunity, facilities and, University of Cape Town and the National Research Foundation (SA) for funding.

I am grateful to my parents, Mrs. Rasika Jayakody and late Mr. Ananda Jayakody. They, raised me, supported me, taught me, and loved me. This work would never have been possible without their support. Thank You So Much!

I am thankful to my parents-in-law, Mrs. Chandra Ranawaka and Mr. Ranthnasena Ranawaka for their understanding and support.

Last but not least, a very special thank goes to my wife Anusha and my son Yasith who have always stood by me and dealt with all the difficulties with a smile and a great deal of patience. Without them I would be a very different person today, and it would have been certainly much harder to finish a PhD. Learning to love them and to receive their love makes me a better person. I love you both very much!

I finish with a final silence of gratitude for my life.

ABSTRACT

Ricin is a dimeric enzyme found in the castor bean plant. It is extremely toxic with a fatal dose for humans ranging from 0.1-1.0 μg/kg. This has lead to its use as a biological weapon. Cell death is caused when ricin ceases the protein synthesis by removing a specific adenine (A-4324) of the GAGA tetra loop of 28S ribosomal RNA. Despite this destructive feature, ricin has been touted as a potential therapeutic agent where applications such as immunotoxins to treat cancer, AIDS and other diseases are actively being pursued. However, the prime challenge in such applications is the non specific cytotoxicity of ricin, which cannot currently be treated due to the absence of an effective antidote.

The primary objective of this thesis is to describe the catalytic mechanism of ricin using computational reaction dynamics. For an accurate simulation of the ricin-catalysed reaction, a reasonable model of the target natural substrate is required. In chapter 4 an appropriate model for ricin's natural substrate is developed and described in the context of existing experimental data. There it is shown that the proposed 12-mer RNA loop serves as a better substrate model compared with the previously used single adenosine nucleoside. The newly proposed and validated substrate model was employed in subsequent simulations. Prior to the reaction simulation, the protonation state of the target adenine was determined. This is discussed in chapter 5, where it is shown that the natural substrate of ricin must be pre-protonated at N7 position. The protonation state of ricin's target substrate, adenine, was one of the major unanswered questions in ricin chemistry which this study addresses extensively.

In chapter 6, the mechanism of a ricin-catalysed depurination reaction is presented and discussed. Using the *Free Energy from Adaptive Reaction Coordinate Forces* (FEARCF) method the catalytic residues in the binding site, their role in the mechanism, and the Transition State of the ricin-catalyses reaction were identified. Finally in chapter 7 the design of transition state analogue inhibitors for ricin based on the identified and characterized transition state structure is described and discussed.

This work has been patent protected (RSA Patent Application No: 2011/06840, for: RICIN TRANSITION STATE ANALOGUE INHIBITORS) and I detail the process and findings in that chapter. A few novel inhibitors for ricin are been proposed and their activity against ricin and other related enzymes are discussed in this chapter.

In this thesis computational methods have successfully been used to elucidate the mechanism of the ricin-catalysed reaction and to design novel transition state analogue inhibitors for ricin. The proof of concept of a completely computational approach to discovering the catalytic mechanism, identifying the transition state and designing from this effective antidotes for enzymes such as Ricin is a major achievement of this work.

List of Abbreviations

RNA	Ribose Nucleic Acid
AIDS	Acquired Immune Deficiency Syndrome
DNA	Deoxyribose Nucleic Acid
BER	Base Excision Repair
RIP	Ribosome Inactivating Proteins
RBP	RNA Binding Protein
RRM	RNA Recognition Motif
dsRBD	double stranded RNA-Binding Domain
HIV	Human Immunodeficiency Virus
NMR	Nuclear Magnetic Resonance
IT	ImmunoToxin
RTA	Ricin Toxin *A* chain
RTB	Ricin Toxin *B* chain
FMP	Formycin MonoPhosphate
TS	Transition State
TSA	Transition State Analogue
TSAI	Transition State Analogue Inhibitor
EF	Elongation Factor
KIE	Kinetic Isotope Effects
FEARCF	Free Energies from Adaptive Reaction Coordinate Forces
QM/MM	Quantum Mechanical/Molecular Mechanical
IUPAC	International Union of Pure and Applied Chemistry
TST	Transition State Theory
SM	Statistical Mechanics
SAR	Structure Activity Relationship
LFER	Linear Free Energy Relationship
PMF	Potential of Mean Force
QM	Quantum Mechanical
MM	Molecular Mechanical
MD	Molecular Dynamics
PES	Potential Energy Surface
SE	Semi Empirical
DFT	Density Functional Theory
LCAO	Linear Combination of Atomic Orbitals
STO	Slater-Type-Orbitals
GTO	Gaussian-Type-Orbitals
DZ	Double Zeta
HF	Hartree-Fock
ZDO	Zero Differential Overlap
SCF	Self-Consistent-Field
MP	Møller–Plesset
CC	Coupled Cluster
DFTB	Density Functional Tight Binding
SCC-DFTB	Self Consistent Charge Density Functional Tight Binding
TB	Tight Binding
KS	Kohn-Sham
NSC	Non-Self-Consistent
FF	Force Field
CHARMM	Chemistry at HARvard Molecular Mechanics
MC	Monte Carlo
WHAM	Weighted Histogram Analysis Method
MSM	Minimum Substrate Model
LSM	Loop Substrate Model
IE	Interaction Energy
COM	Centre Of Mass
RMSD	Root Mean Square Deviation
SRL	Sarcin-Ricin Loop
TBS	Truncated Binding Site
UPA	UnProtonated Adenine
LUMO	Lowest Unoccupied Molecular Orbital
HOMO	Highest Occupied Molecular Orbital
ESP	Electrostatic Surface Potential
NAC	Near Attack Conformations
LO	Lead Optimisation
QSAR	Qualitative Structure Activity Relationship
FEP	Free Energy Perturbation
ADMET	Absorption-Distribution-Metabolism-Elimination and Toxicology
MCORE	Molecular Core

Contents

Background

Hydrolysis of the N-glycosidic bond in DNA and RNA nucleosides by enzymes is an interesting phenomenon with a great deal of biological importance. In certain instances, nature has employed these N-glycoside hydrolysis reactions to ensure the continuity of life. One such example is the N-glycoside hydrolytic activity of Base Excision Repair (BER) enzymes, where repair of genetic material is carried out[1]. In certain other instances, presence of N-glycosidase enzymes can be lethal. In glycobiology the action of Ribosome Inactivating Proteins RIP (a family of enzymes with N-glycosidase activity) leading to the destruction of eukaryotic cells have been central in our understanding of the role that enzymes, acting on carbohydrates, play in biology. ,RIPs enter eukaryotic cells and terminate cell life by arresting its protein synthesis[2]. Ricin is a member of the RIP family with many important applications ranging from destructive biological weapons to therapeutic immunotoxins[3]. The central hurdle that impedes our understanding of RIPs and Ricin action in particular is the absence of a catalytic reaction mechanism that draws disparate experimental observations together in one coherent model. This work successfully tackles the major challenges preventing our understanding of ricin chemistry using a battery of computational techniques that provide access to hidden troves of dynamic, electronic and structural data that are not accessible via present day experimental techniques.

CHAPTER -1

INTRODUCTION

1.1 Modifications in RNA

The presence of a significant portion of modified nucleosides derived from adenosine, guanosine, cytosine and uracil is a unique structural feature of cellular RNA[4]. To date, over 100 chemically distinct modified nucleotides have been identified in cellular RNA.[5] The chemical modifications in RNA can be very simple or complex in nature. Simple modifications yield modified RNAs such as dihydrouridinine or 7-methylguanosine whereas complex modifications yield products such as wybutosine[5].These modifications are formed when specific enzymes acting concurrently with enzymes involved in common RNA processing events such as 5'/3' nucleotide addition, 5'/3' trimming.[4]. These modifications of RNA, a large majority of which take place in the nucleus, are integral to the complex process of RNA maturation and form part of the *posttranscriptional modifications* that are central to cell growth. Not all modifications occur in the nucleus in fact key changes take place in the cytoplasm or in the process of transporting RNA through the nuclear pores.

Even though these modifications in RNA indicate that they must have important biological functions, most of those functions have not been completely understood. Some of them are considered to useful to molecular recognition. An example of this assumption is the modification of tRNA at the anticodone loop that facilitates the interaction with their corresponding mRNA codon(S). In other cases, it is assumed that *posttranscriptional modifications* play an important structural role in RNA. A well known example of this is the maintenance of the characteristic L shape of tRNA that is facilitated by RNA modifications. However, there are instances where RNA editing can be catastrophic and lethal to the cell. This is the focus of this thesis.

1.2 RNA Editing Enzymes and RNA Binding Proteins

The RNA modifying processes (constructive or destructive) are facilitated by specific enzymes. Even though there are a vast number of enzymes that are capable of editing RNA, their rate enhancement power, substrate specificity, mechanism of action and many other structural and functional features have not yet been fully understood. Nonetheless, numerous studies that have been done on this subject have provided some invaluable information in regards to these issues.

It has been understood that all RNA modifying and binding enzymes possess positively charged binding surfaces that can be employed to recognise and bind to the negatively rendered (by the its sugar-phosphodiester backbone) RNA molecules.[6] Further, it has been shown that these positively charged binding interfaces are mainly composed of ARG and LYS residues[6]. Therefore, it is obvious that the electrostatic interactions drive the substrate recognition and enhance the formation of RNA-enzyme complexes in RNA editing enzymes.[7,8] However, the acidic residues such as ASP and GLU are also abundant in those proteins.

It has been argued[9] that the water mediated hydrogen bonds are crucial in substrate binding in these enzymes in order to provide a shield from acidic residues. Besides the predominant electrostatic interactions, it has been shown[9] that the aromatic interactions (pi-stacking) between the side chains of TYR, TRP and PHE are also greatly involved in RNA-enzyme binding. Moreover, it has been proposed that these enzymes take advantage of the presence of 2'-OH group in RNA in their substrate recognition process.[10] Even though a complete understanding of substrate recognition and binding process of these enzymes is not gained yet, it has been proposed that they use an induced-fit mechanism[11] in substrate recognition.

Very often RNA is found in complexes with proteins as large ribonucleoprotein complexes such as the ribosome or in more transient interactions such as the helicase-RNA interactions[12]. Proteins that form functional units when complex with RNA are defined as RNA Binding Proteins (RBP)[8]. The RBPs are primarily built from distinguishable basic domains and repeats of these domains that are arranged in various ways in order to satisfy RBPs' diverse functional requirements[13].

Substrate specificity of the RBPs is defined by their basic binding motifs. Some such motifs are: the RNA Recognition Motif (RRM), double stranded RNA-binding domain (dsRBD), and zinc finger motif[14]. Thus far, there have been more than a thousand RBPs have been identified with known RNA-binding motifs[15].

1.2.1 Ribosome Inactivating Proteins (RIPs) - Destructive RNA Editing Enzymes

1.2.1.1 Structural features of RIPs

Ribosome Inactivating Proteins (RIPs) are a group of catalytic proteins that act on ribosomal RNA (rRNA) in a destructive manner[16]. The RIPs have a wide distribution in nature. Although, they are mainly found in plants, some RIPs are found in mushrooms and bacteria. Some of those RIPs such as ricin and abrin are famous RIPs that were well known to the mankind since more than a century ago. Interestingly, all RIPs exhibit a similar catalytic mechanism[2,11,17].

RIPs are classified in to two major classes , namely, type-1 RIPs and type-2 RIPs[18]. The type-1 RIPs consist of small polypeptide chains of approximately 30 kDa. In comparison, the type-2 RIPs are hetero dimmers. A type-2 RIP consists of a catalytic *A* chain which is similar to type-1 and a slightly larger (approximately 35 kDa) non-catalytic *B* chain[19]. The *A* chain is connected to the B chain via a disulphide linkage. The B chain in type-2 has the properties of a lectin with specificity for sugars with the galactose structure[20] and it is considered be involved in cell recognition to facilitate the entrance of *A* chain to the cell. Figure 1 schematically illustrates the type-1 and type-2 RIPs.

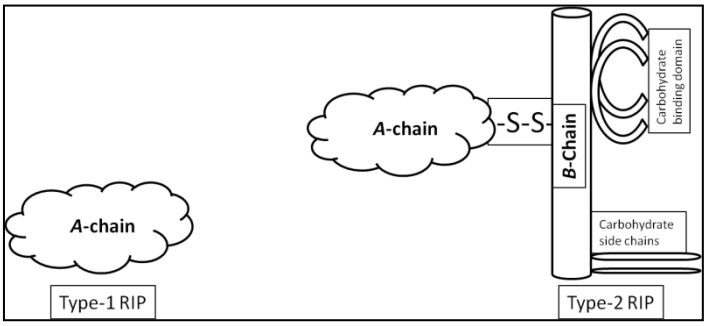


Figure 1: Type-1 and Type-2 Ribosome Inactivating Proteins. Type-1 contains a single catalytic *A* chain. In addition to the catalytic *A*-chain, type-2 contains a cell-recognizing B chain.

All members of the RIP family show a conserved catalytic activity due to the high degree of similarities found in their carboxyl-end and core sequences. On the other hand, the differences in their non-core sequences account for the diversity in other RIP activities (discussed later)[21].

1.2.1.2 Biological and Catalytic Activity of RIPs

The biological activity of type-1 RIPs was first discovered by Duggar et.al. [22]whilst investigating the inhibition of the transmission of the mosaic virus in tobacco plants. Subsequent studies have shown that many other type-1 RIPs possess antiviral properties[16,23]. It is assumed that RIPs' anti-viral activity is directly related to their de-andenylation activity on viral RNA or DNA other than their usual rRNA targets. However it has not yet been shown if the RIPs can use viral nucleic acids as a substrate. Nonetheless, the demonstrated anti-viral activity of pokeweed antiviral proteins (from the pokeweed plant) on HIV-1 RNA has produced strong evidence to support the above assumption[24,25]. Unlike type-2 RIPs, the type-1 RIPs in general do not show any toxic effect as they are incapable of entering the cells without the *B* chain.

The biological activities of Type-2 RIPs are twofold; the activities due the *A* chain and the activities due to the *B* chain. The carbohydrate binding activity shown by *B* chain of type 2 RIPs was recognised long before the recognition of their catalytic activities and power to cease protein synthesis by the *A* chain. The carbohydrate binding activity of the B chain is facilitated by multiple (2-3) active sites on the B chain[26]. The catalytic activity of chain *A* in type-2 RIPs occurs inside the cell. It has been proposed that the catalytic chains of RIPs will have 2 binding sites[27]. The first site is to identify the specific portion of the substrate RNA and the second site is to perform the specific hydrolysis reaction.

The catalytic activity of RIPs was first discovered with ricin by Endo and co-workers[28] Both type-1 and type-2 RIPs show a common depurination catalytic activity on a specific adenine in 28S ribosomal RNA (rRNA). In the ribosome, this specific adenine, A-4324, is located in a single–strand loop called sarcin/ricin loop which is located in the domain VII, about 400 nucleotides from the 3' end.

The specific adenine* is found in the GA*GA turn of the sarcin/ricin loop. Figure 2(a) illustrates the location of sarcin/ricin loop in the rRNA and the Figure 2(b) presents the NMR determined[29] 3D structure of a portion the sarcin/ricin loop.

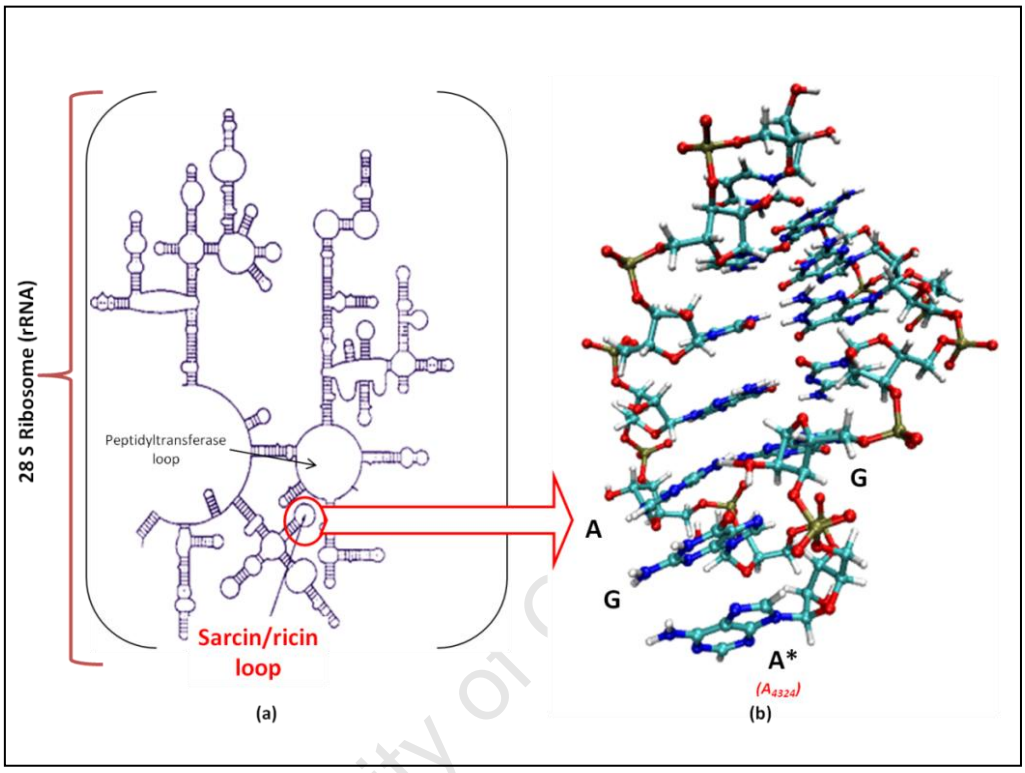


Figure 2: Natural Substrate of RIPs. (a). Sarcin/ricin loop of 28S ribosome. (b). NMR-determined structure of a portion of the sarcin/ricin loop.

The depurination reaction catalysed by RIPs is a hydrolysis reaction. More specifically, RIPs hydrolyse the *N-glycosidic* bond in A-4324 nucleoside. The products of this reaction are abasic rRNA and free adenine. Figure 3 presents a generalised mechanism for *N*-Glycosidase activity of RIPs.

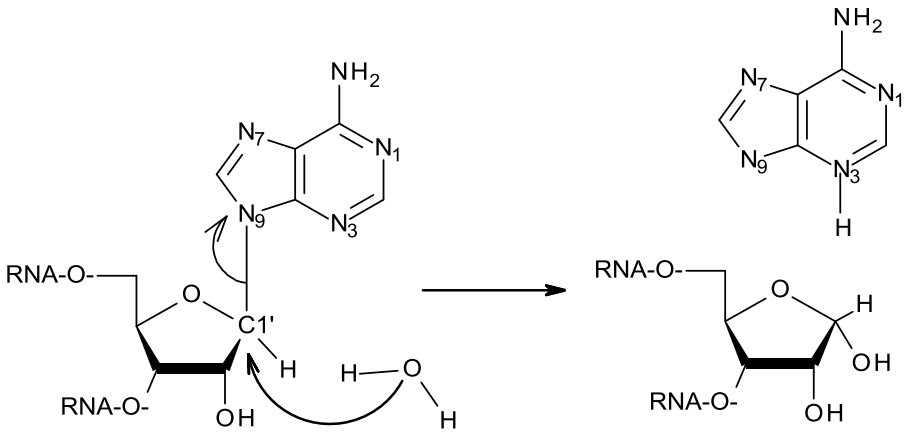


Figure 3: A generalised mechanism for *N-glycosidase activity* by RIPs.

Even though the RIPs are well known for their specific activity on A-4324, it has been shown[30] that they are capable of removing multiple nucleotides, specifically adenines from the mother loop. Therefore, RIPs are also be known as polynucleotide-adenosine glycosidases.[31] However, due to vast variation found in substrate specificity amongst different RIPs, it is yet unknown whether all RIPs have similar polynucleotide glycosidase activity.

It has been shown that RIPs are capable of performing other enzymatic activates besides their well-known glycosidase activity. These activities have only been studied relatively recently. Most of these novel enzymatic activities are related to a presumed RNase or DNase activity. Other enzymatic activities that have been reported for RIPs include phosphatase activity on lipids, phosphatase activity on nucleotides, chitinase activity and superoxide dismutase activity. However, these non-glycosidase activities of RIPs have not yet been fully explored. Table 1 summarises the enzymatic activities of RIPs.

Activity	Substrate	Product	RIP
RNA N-glycosidase	Ribosomes (animal/plant/bacterial)	Adenine	All RIP
Polynucleotide: adenosine glycosidase	Nucleic acid	Adenine	Most type-1 RIPs and some type-2 RIPs.
Ribonuclease	RNA	Cleaved RNA	A and β momorcharin
DNase	Supercoiled DNA	Linear and nicked DNA	Trichosanthin, saporin, gelonin
DNA glycosylase /AP lyase	DNA (super coiled and linear)	Adenine	Ricin , gelonin and PAP
Phosphatase	Phospholipid AMP		Ricin, Tricosanthin
Chitinase	Chitin		Trichosanthes

Table 1: Enzymatic activities of ribosome inactivating proteins.

1.3. Ricin: The Best Known RIP

Ricin is the best known RIP today and it is isolated from *Ricinus communis*(Castor bean plant) [28,32]. It was first identified and named "Ricin" after the work of Stillmark H.[33]in 1888. Identification of Ricin was an important milestone in biochemistry, as this was the first time a well-defined biological activity had been assigned to a plant protein. The toxic dose of ricin for humans is likely to be in the 0.1-1.0 µg/kg range, and it has been ranked among the most toxic substances known[32]. Enzymatic activity of ricin on 28S rRNA was identified by Endo et.al. in 1987[34], and other subsequent studies[35-39] have confirmed ricin's depurination activity on the specific rRNA adenine A-4324 , the common substrate of all RIPs.

1.3.1. Ricin- Significance and Applications

The first recorded historical application of castor bean plant is use of castor oil as a lubricator and a laxative in ancient Egypt. During World Wars I and II, castor oil was used in aircrafts as a fuel[40]. The toxin (ricin) is easily extracted from the ricin meal after extraction of the oil by using a simple salting-out procedure[40]. Ricin first gained attention of the scientific research due its lectin-like i.e. sugar-binding properties[20,41].

The work done by Paul Ehrlich in 1890 on ricin's lectin activity laid the foundation of the discipline of immunology[42]. The tumour growth inhibiting activity of the ricin *A* chain was first found[17] in 1951. Since then, ricin gained reputation as an antitumor agent which falls under the therapeutic substance class called immunotoxins (ITs). Immunotoxins consist of cell binding ligands coupled to toxins or their subunits[43]. It has been shown that ricin can be linked to so called cell–specific monoclonal antibodies to facilitate the selective killing of tumour cells[44].

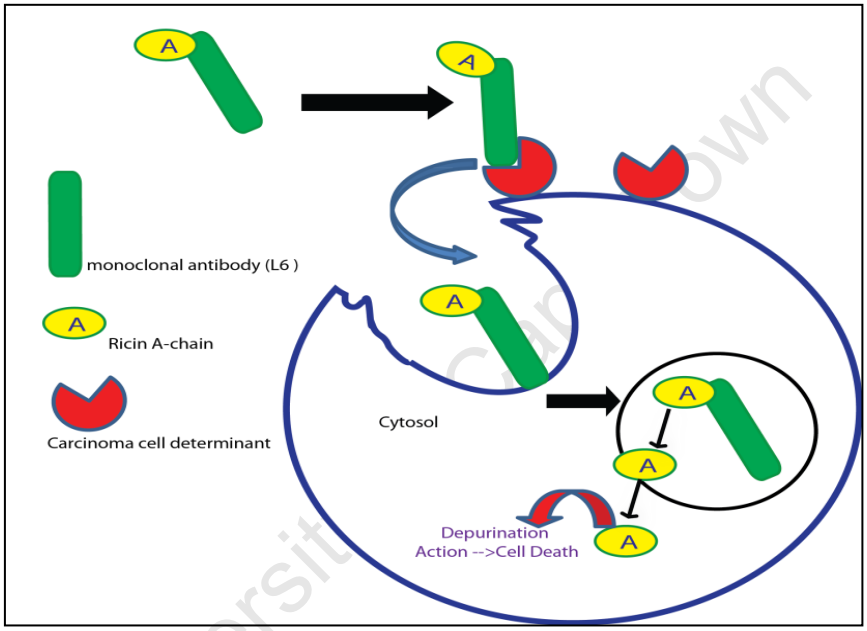


Figure 4: Immunotoxic activity of Ricin. Ricin is attached to antibody L6 which recognises a specific determinant released by cancer cells. The immunotoxin is transported in to the cell. When the toxin A chain is released to the cytosol and it invokes the deadly depurination reaction.

One such monoclonal antibody is antibody L6 that recognises a determinant that is expressed on lung, breast, colon, and ovarian carcinomas, whose presence in normal tissues is only at trace levels. The L6-ricin IT has shown selective toxicity to L6-positive H2981-T3 adenocarcinoma cells. Further studies have shown that 1 µg/ml L6-ricin could inhibit about 99.99% of H2981-T3 growth[45]. Ricin also has been successfully used as an IT in treating Hodgkin's lymphoma[43]. Figure 4 illustrates ricin's application as an IT.

Although there have been some promising results, two factors appear to limit ricin's immunotoxin efficacy: (1) lack of specificity of the antibody and (2) significant immunogenicity of the toxin moiety, which results in relatively rapid onset of refractory immunity to the therapeutic agent.

Ricin has been classified as a Class B bio threat. This is mainly due to its free availability across the globe, stability and the extremely high toxicity. It has recorded that the bio weapons were made with ricin by certain countries in the World Wars (I and II). The most famous application of ricin as bio weapon is the assassination of Georgi Markov in 1978[46]. There have been many records recently about weaponization of ricin[47].

1.3.2. Ricin – Structure, Key Residues and their Proposed Roles

Ricin toxic material makes up 1% to 5% by weight of the bean of the castor plant. The toxic protein is encoded by a small multigene family composed of approximately eight members[41]. Expression of these genes is developmentally regulated, and it is tissue specific. The toxic is synthesised in the endosperm of cells of maturing ricin seeds and it is stored in a special compartment called the protein body. The toxin is rapidly destroyed by hydrolysis reaction during the first few days of germination.

The toxic material is a 66 kD heterodimeric globular protein. Therefore, it belongs to the class of type-2 RIPs. As in any type-2 RIP, chain *A* performs the catalytic activity whereas chain *B* facilitates the transportation of the catalytic chain to its destination. The *A* chain (RTA) is released to the cell by cleaving the *A*-*B* disulfide bond, it is then internalised by endocytosis[48]. The RTA is then directed to the endoplasmic reticulum via retrograde transport[49] and subsequently is moved to the cytosol wherein it binds to a specific nucleotide sequence on the sarcin-ricin tetra loop of the 28S ribosomal RNA.

The Ricin Toxin *B* chain (RTB) has two major domains, each of which has a galactose binding site. RTB displays several Ω loops and has no regular secondary structure. Each RTB domain is made of three copies of a primitive 40 residue folding unit. In addition, it has been found that there is an assembly of sub-domain units around the pseudo threefold axis of each domain. This tight threefold binding is considered to drive the peptide folding and stabilises the structure RTB.

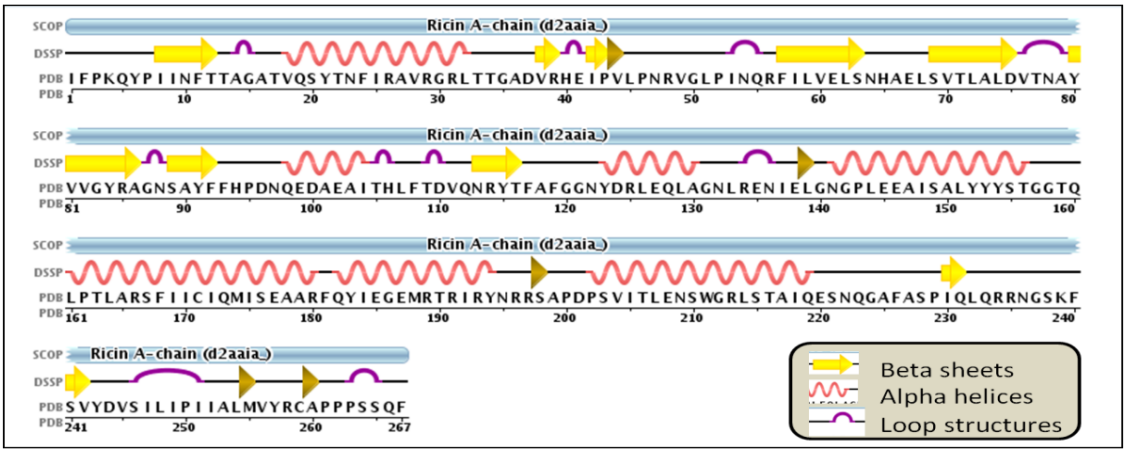


Figure 5: Amino acid sequence of ricin A chain. Residue names are given by the one letter codes. The number line corresponds to the residue numbers.

RTA has a total of 267 amino acid residues. Its sequence[50] is presented in Figure 5. In RTA , there are 7 helices that composed of 97 residues which make up to 36% of the total structure. A further 17% corresponds to beta sheets that involve 15 strands that composed of 47 residues; the remainder is composed of non-specific loop structures.

As shown in Figure 6, three main domains have been identified in the *A* chain. First 117 residues in the amino terminal form the first domain, Figure 6(A). This domain is dominated by the five-stranded β sheets. The second domain is dominated by α helices, and it is composed of residues 118-210 (Figure 6 (B)). The third domain which is composed of residues 211-267 forms a compact disk-like domain (Figure 6(C)).

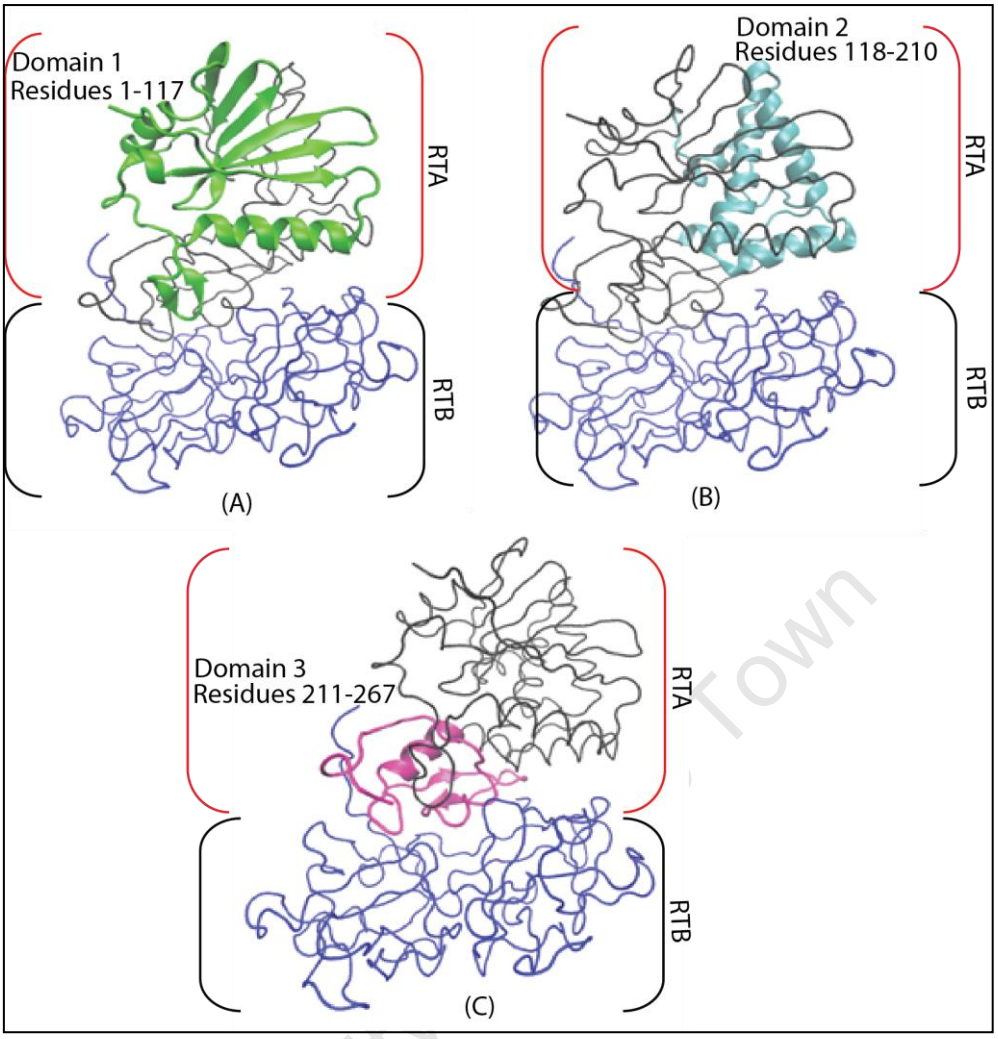


Figure 6: Ricin. The B chain is represented in blue. A chain domains 1,2,3 are Highlighted in figures (a), (b) and (c).

The crystal structure of ricin has been solved[51,52]. There are two proposed distinguishable binding sites for ricin in its open form. The first is the primary adenine specificity pocket whilst the second is a slightly larger secondary pocket. The two pockets are separated by the side chain of Tyrosine 80. The second pocket was clearly identified in model building studies and it has been proposed the role of this pocket is to accommodate a guanine base from the invariant GAGA ribosomal target sequence[53]. Furthermore, presence of this secondary pocket has been confirmed by the X-ray structure of an RTA complex with a locked cyclic nucleotide[54]. Ricin shows the common RIP feature of having a positively charged binding site which allows recognition and binding of negatively charged phosphate backbone of RNA. The ricin binding sites are shown in Figure 7.

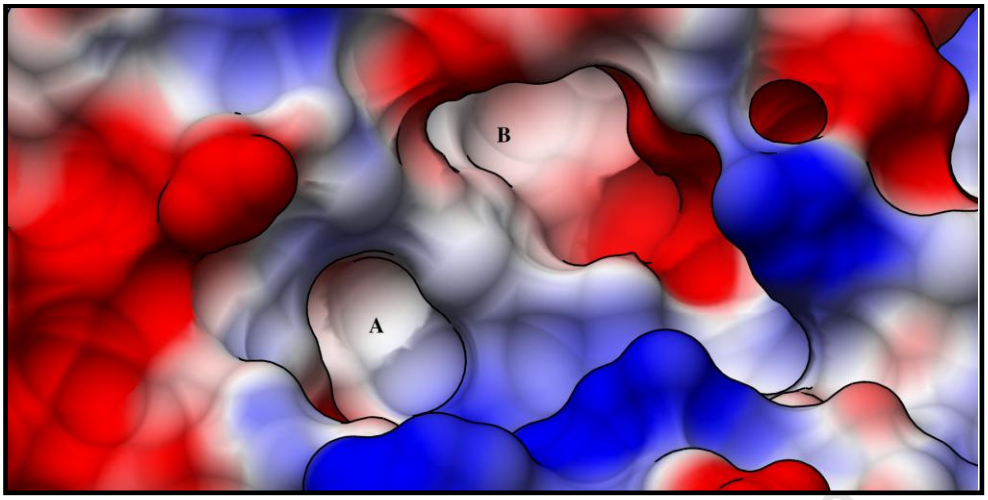


Figure 7: Electrostatic potential surface of open RTA. Adenine specific binding site is labelled A. Secondary binding site is labelled B.

X-ray crystallographic studies[53,55] and site directed mutagenesis studies[56,57] have identified several key amino acids in the ricin binding site. Among them, the most commonly reported residues are GLU177, ARG180, TRP211, TYR80 and TYR123. Even though the exact roles of these residues in substrate recognition and in catalytic mechanism have not been completely understood yet, they were assigned "plausible roles" based on the findings of the above mentioned experiments.

It is proposed[58] that TYR80 and TYR123 employ pi-stacking interactions to 'sandwich' the base moiety of target adenine in between their phenyl rings. Moreover, some kinetic studies and site directed mutagenesis experiments have suggested that these residues might also take part in the catalytic mechanism[32]. According to some previous studies[59], ARG180 is expected to play a role in ground state stabilisation (electrostatic attraction) of ribose and / or transition state destabilisation (repulsion) of the oxocarbenium ion resembling transition state (proposed) of the ricin catalysed mechanism. It has been suggested that GLU177 stabilises the transition state and/or can act as the base in the catalytic mechanism[60]. Residue TRP211 has also been proposed to play an important role in the catalytic process, where its role is defined as a non-specific participation[61]. Another residue that has been proposed to participate in the mechanism is GLU208.[62]. Moreover, various studies[54,63] have reported residues VAL81 and GLY121 as binding residues.

The proposed key amino acid residues are illustrated in Figure 8. Analysis of protein structure shows that most of these proposed key residues belong to the domain 2 of chain A. Furthermore, it shows that these residues are located at the interface of the three domains of RTA. However, understanding about the binding pocket of ricin and the specific role(s) of individual amino acids in substrate recognition/catalysis remains incomplete.

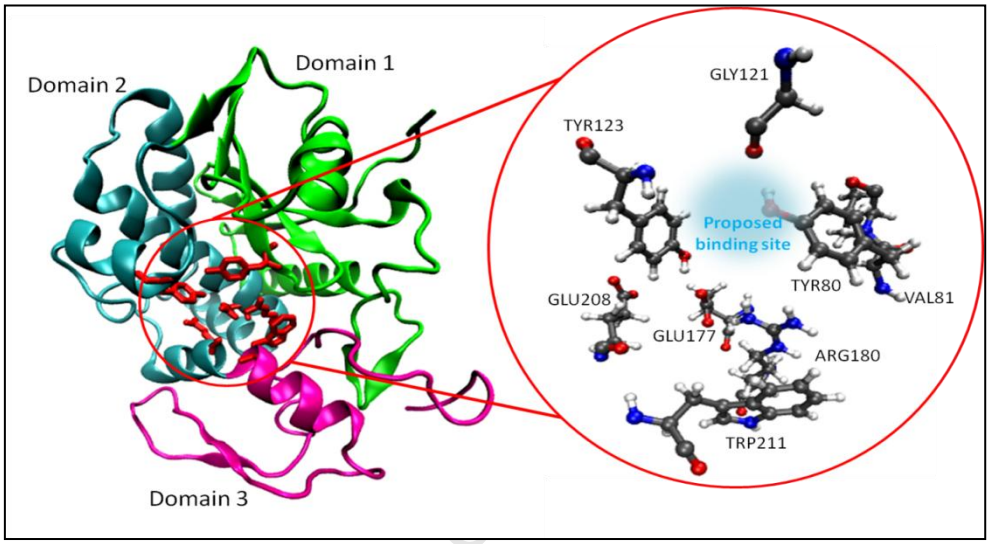


Figure 8: Proposed binding domain and key residues of ricin.

1.3.3. Inhibiting Ricin

Since ricin has gained a great deal of attention as a bio-weapon as well as a therapeutic agent, the interest of identifying effective inhibitors for ricin remains extremely high. The usage of ricin inhibitors can be twofold. Firstly, an effective inhibitor can serve as an antidote for ricin poisoning. Secondly, ricin inhibitors can be used to control the nonspecific cytotoxicity of ricin when it is used as a therapeutic agent in immunotoxins.

The most apparent approach of controlling cytotoxicity caused by ricin seems to be blocking the ricin B chain (RTB). If RTB can be inhibited, cell uptake of the toxin can be precluded. Analyses of the X-ray structures[64] have shown that RTB is composed of two related domains. Each of these domains is composed of three related sub-domains.

The galactoside binding sites are found only on one sub domain of each domain, and these sites are over 50 Å apart since they are located on the opposite ends of the protein[65-67]. Moreover, it has been found that these relatively small[68] (120–150 Å) binding sites exhibit only weak binding to galactosides with K_d values in the millimolar range[69]. It has been found that inhibition by using these sites are biologically tolerable owing their weak ligand-binding ability. Further, it has been suggested that these sites cannot be simultaneously blocked by small molecules owing to their relative locations, indicating that designing effective inhibitors for these shallow and polar galactose sites is extremely difficult. Based on the above facts, RTB is considered as a poor target for inhibitor design.

In contrast, RTA has two larger pockets that are within close proximity to each other (Figure 7) and this feature makes it possible for a molecule with two moieties to fit in to both pockets simultaneously. In addition, the presence of a deep, hydrophobic substrate-specific pocket has also made the A chain an attractive target for structure-based drug design.

As for any other enzyme, inhibitors for ricin can be classified in to three classes; (i) Substrate analogues, (ii) Transition state analogues and (iii) Product analogues. Even though inhibitors of all these types have been made and tested, yet, there is no effective inhibitor for ricin. Studies of these inhibitors however have provided invaluable clues about the ricin binding site, catalytic residues, and have made valuable predictions on the nature and mode of binding of the natural substrate and on the catalytic mechanism.

1.3.3.1. Substrate Analogues

The most popular substrate analogue inhibitor for ricin is formycin monophosphate (FMP), the structure of which is presented in Figure 9 (a). There are only few structural differences between FMP and the natural substrate of ricin (the targeted adenosine nucleoside). FMP has a C9-C1' (carbon-carbon) linkage between the base moiety and the sugar moiety, in contrast to the N9-C1' (N-glycosidic) linkage found in the natural substrate. Consequently, it has an N atom in the 8th ring position in contrast to a C atom in the natural substrate. It was expected that C9-C1' is less reactive than N9-C1' and as such, is not hydrolysed by ricin. Hence, FMP acts as an inhibitor.

Analysis of binding[53] of FMP to RTA has proposed a binding pattern for the natural substrate. The binding of adenine-like ring of FMP between the phenyl moieties of TYR80 and TYR123 has proposed a similar pi-stacked binding for the target adenine. Further, binding of FMP has proposed a strong H-bonding interaction between N3 and ARG180. It has also proposed that VAL81 will be involved in two H-bonds with N1 and NH2 group at the 6th position. Involvement of residue GLY121 in binding via H-bonds is another suggestion that has been arisen from the FMP binding. Binding of FMP to ricin is illustrated in Figure 9.(b) and Figure 9. (c). Even though FMP was found not to be effective as an a ricin inhibitor, its binding has provided strong predictions about the binding of natural substrate, where it can be assumed that the small structural changes between FMP and natural substrate make these predictions reasonable.

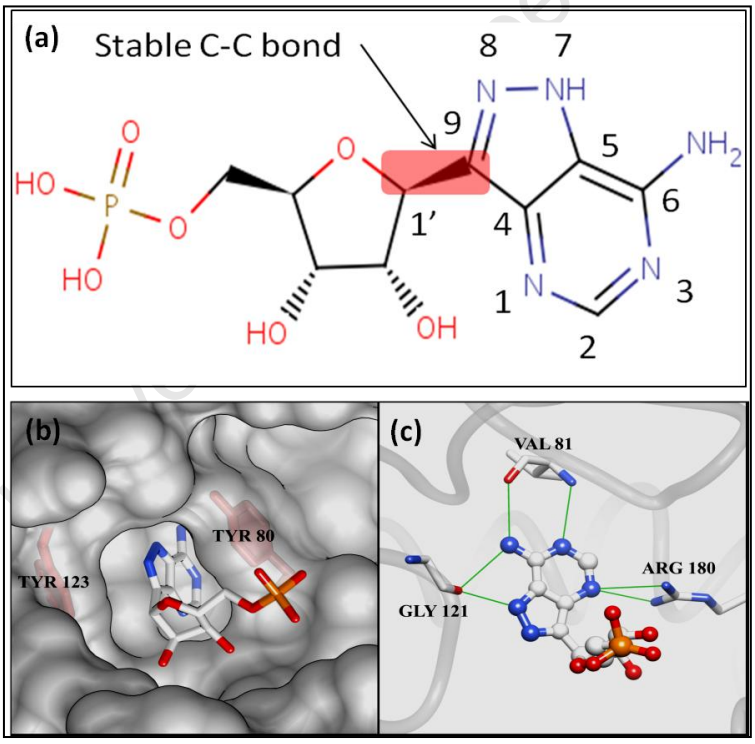


Figure 9:(a). Formycin 5'-phosphate, a substrate analogue inhibitor for ricin. (b) pi-stacking between adenine moiety with TYR80 and TYR123 (c). H-bonding interactions with VAL81, GLY121 and ARG180.

1.3.3.2. Transition State Analogues (TSAs)

Transition State Analogue (TSA) inhibitors are considered as the most powerful class of inhibitors for a given enzyme[70-73]. A dissociative-type transition state (TS) has been proposed for RTA-catalysed hydrolysis reaction[74] (discussed later in this chapter). These TS structures were characterised by the increased ribosyl–adenine distance.

In all previous attempts of ricin TSA design, it was tried to mimic this increased distance and the corresponding TS geometry by introducing linker atoms/groups between the base and the sugar. Based on the proposed Transition State structure, a few TSA inhibitors for ricin have been designed, synthesised and tested[54,75-78]. Further, those proposed transition state structures for RTA-catalysed reaction have enabled design of inhibitors for other RIPs as well[55]. However, none of these inhibitors are effective against ricin.

The latest[54] TSA for ricin is cyclic *G*(*9-DA*)*GA2-OMe* (which is referred to as 9DA hereafter) and it was designed by using 2-deoxyadenosine as a scaffold. Furthermore, it consists of a nucleotide sequence that resembles the GAGA tetra loop of the natural rRNA. However, *9-DA* was found to be an imperfect transition state mimic, as it inhibits RTA with a competitive inhibition constant (*K*i) of 300 nM at pH 4.0 where it binds only 320-fold tighter than the substrate in the same construct . Nonetheless, Binding of this TSA inhibitor has proposed reasonable binding patterns and interactions for the natural substrate.

As it was suggested by the binding of substrate analogue inhibitors, binding of TSAs have also suggested a "sandwiched" pi-stacked binding for adenine moiety of the natural substrate between TYR80 and TYR123. Further, a strong H-bonding interaction between ARG180 and N3 of the natural substrate also has been proposed. In the *9-DA* crystal structure, a potential catalytic water molecule was identified for the first time. Since this water molecule was found well positioned between GLU177 and the substrate's reactive carbon (C1'), it has been argued that GLU177 acts as the base in the catalytic mechanism to activate the attacking water molecule. Binding of *9-DA* has also suggested a protonation state for the leaving adenine in the natural catalytic mechanism. It was suggested that the leaving base might be protonated at N1 and/or N7 positions.

Besides these proposals, binding of TSA inhibitors has suggested that ricin prefers a folded GAGA tetra loop for binding and catalysis in contrast to a linear GAGA tetra loop. In conclusion, the studies on TSA have provided much useful clues about the reaction mechanism and the catalytic residues.

1.3.3.3. Product Analogues

Free adenine is one of the products of the ricin-catalysed depurination reaction. Since ricin has an adenine-specific binding pocket, there have been various attempts[32,63] to make single-ring compound inhibitors which mimic the adenine-moiety of the natural substrate, i.e. the adenine product of the reaction. The best product-analogue inhibitors thus far were made and tested by Robertus et.al.[32] However, none of these inhibitors were found to be effective against ricin. Furthermore, it has been found[79] that these product-like inhibitors are ineffective due to their poor electrostatic complimentarity with the ricin binding site. In general, product analogue inhibitors are not capable of revealing useful information about natural substrate binding and reaction mechanism due to the unique interactions between the enzyme and the expelling products that are different to substrate-enzyme interactions. For ricin, it has been found that (Naidoo K.J and Jayakody R.S. unpublished data) higher resemblance of product-like inhibitors with adenine will result in poor inhibition power. However, It has been shown[80] that high concentration (in the mM range) of free adenine can protect the ribosome from ricin via uncompetitive inhibition. From these effective concentrations, it is obvious that the free adenine and its mimics cannot be used as effective inhibitors for ricin.

1.3.4. The Proposed Catalytic Mechanism for Ricin

As mentioned before, ricin catalyses the depurination reaction at the GAGA tetra loop of 28S rRNA[81]. Similarly to other RIPs , ricin specifically targets A-4324 in this tetra loop[28]. The rRNA target site for ricin is schematically presented in Figure 10.(a). Removal of adenine 4324 destroys an elongation factor (EF) binding site[3] of the ribosomal RNA (Figure 10(b)). Consequently, the ribosome becomes immobile and incapable of continuing the mRNA decoding. This ultimately terminates the protein synthesis and causes cell death.

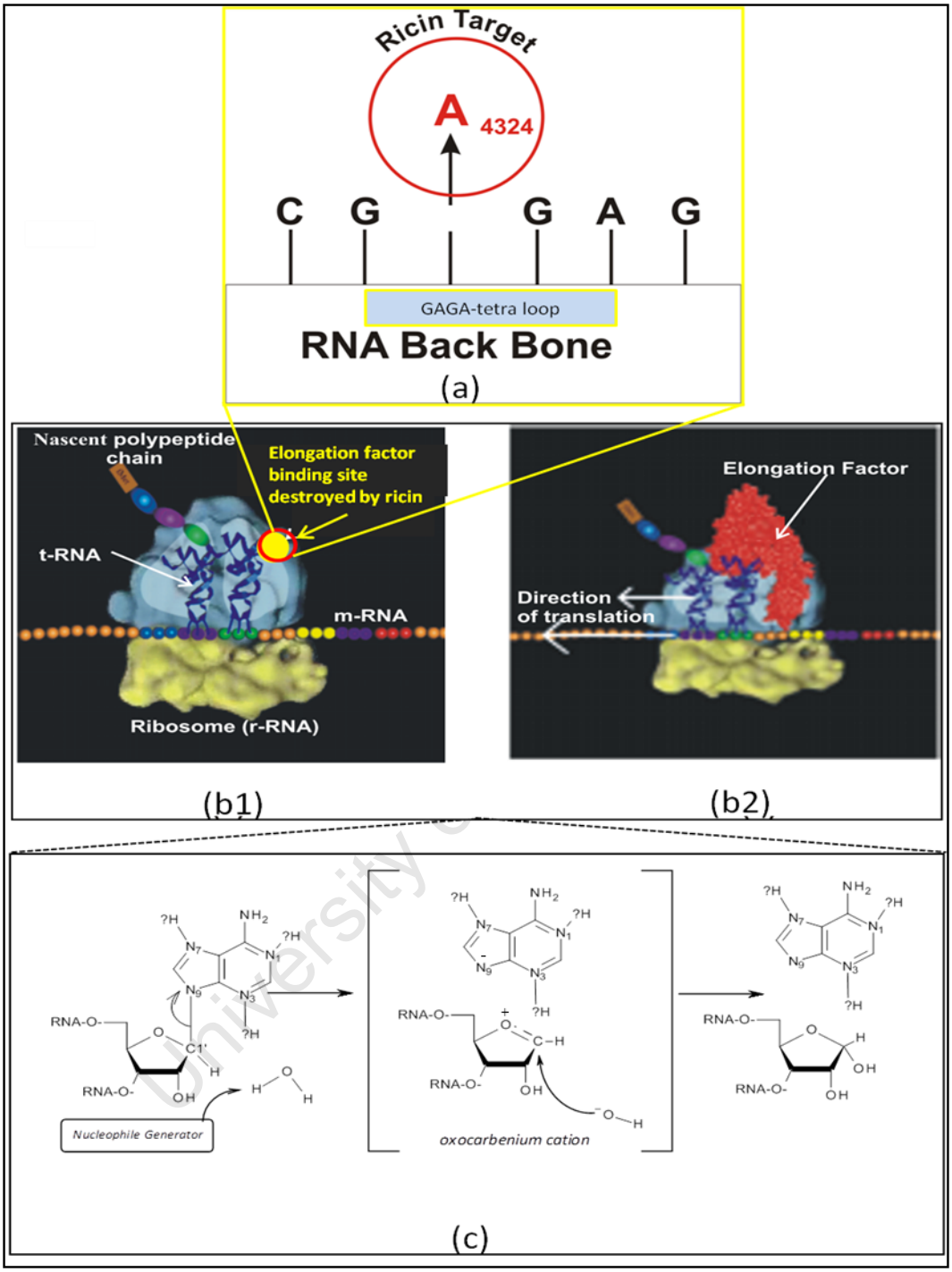


Figure 10: (a). Ricin's targeted adenine. (b). Binding of elongation factor does not occur as shown in figure (b2) when ricin destroys the EF site shown in (b1). The loss of mobility of m-RNA / ribosome terminates the protein synthesis (c). A proposed-generalised reaction mechanism for ricin-catalysed depurination.

Identification of its target adenine by ricin among millions of other nucleobases is quite intriguing, and the exact process is yet unknown. It has been suggested that ricin might use a base flipping mechanism to dock its target adenine to the active site[82]. However, the exact mechanism of the ricin-catalysed reaction is unknown. The incomplete understanding challenges into the study of the chemistry of ricin. There have been few proposed mechanisms for the ricin-catalysed reaction. Previous investigations have mainly used experimental techniques such as kinetic isotope effects (KIE), site directed mutagenesis and x-ray crystallography. From these proposed mechanisms, it is possible to represent a generalised reaction mechanism for ricin. This generalised mechanism is presented in Figure 10.(c). The individual proposed mechanisms are discussed in detail in chapter 6.

The common mechanistic features that are shared by the previously proposed mechanisms are: (a) Participation of an external water molecule in the mechanism (b) Activation of a water molecule by a binding site amino acid residue to generate the OH^- nucleophile (c) Protonation of the leaving base. (d) Formation of the oxocarbenium ion resembling Transition State (TS). The amino acid residue that activates the external water molecule and generates the nucleophile is not known yet. Therefore, it is presented as the "*nucleophile generator*" in Figure 10.(c), Various combination of protonation states have been proposed for the adenine's nitrogens (N1,N3,N7)[74]. In the generalised mechanism, protonation sites are shown with "?H" to emphasise the uncertainty of their protonation state.

It has been shown[74,78,83,84] that the depurination reaction catalysed by ricin proceeds via an oxocarbenium ion-like TS. However, experimental studies were unable to reveal the exact structure of the TS due to its extremely shot lifetime. The theoretical calculations[74] done to-date have only studied the target adenine in the gas phase. Therefore, the exact nature of the TS in the enzyme's binding site and the binding site residues that are responsible for stabilising the TS are yet to be identified. The KIE experiments have strongly suggested that the ricin-catalysed depurination reaction proceeds via an pseudo- S_N1 mechanism or $D_N{*}A_N$ mechanism via formation of an oxocarbenium cation[74,83].

1.4. Objectives

The primary objective of this thesis is to elucidate the mechanism of the ricin-catalysed depurination reaction by using computational techniques. Incomplete understanding of the reaction mechanism has resulted in failure to design an effective inhibitor for this toxin. In this thesis, an improved biasing force method, Free Energies from Adaptive Reaction Coordinate Forces (FEARCF) was used to simulate the ricin-catalysed reaction with the QM/MM protocol. Simulation of ricin's depurination reaction is discussed in chapter 6 of this thesis.

Since ricin targets a specific adenine in macromolecular rRNA, it is crucial to use a reasonable model that can adequately represent the essential features of the natural substrate. Therefore, proposing an accurate substrate model for computer simulation of the ricin's natural substrate is another objective in this thesis. This is discussed in chapter 4.

The adenine targeted by ricin can exhibit different single or multiple protonation states owing its multiple protonation sites, N1, N3 and N7. However, none of the previous studies have identified the exact protonation state of the target adenine. Since assigning the accurate protonation state is mandatory prior the reaction simulation, investigating the protonation state of ricin's natural substrate is another objective of this thesis. Chapter 5 is devoted to this problem.

It was expected to identify the Transition State (TS) of the ricin-catalysed reaction from the reaction simulations. Knowledge of the TS is crucial to design the Transition State Analogue (TSA) inhibitors, the most powerful inhibitors that can be designed for an enzyme. Designing TSA for ricin and investigating their potency against similar other enzymes was made another objective in this thesis. This is discussed in chapter 7. The basic layout of this thesis is presented in Figure 11.

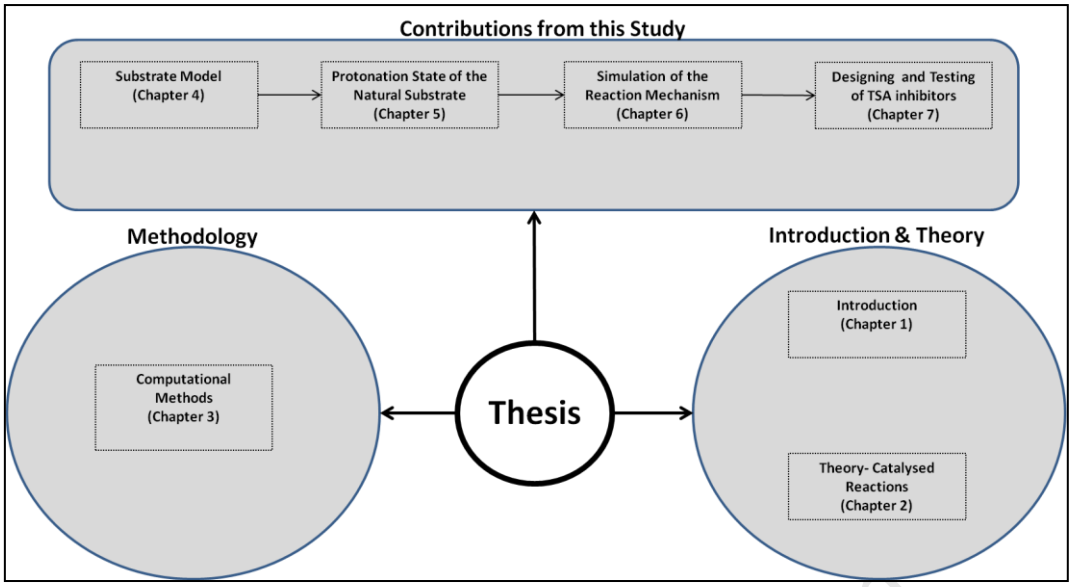


Figure 11: Layout of this thesis

1.5. References

(1) Berti, P. J.; McCann, J. A. B. *Chemical Reviews* **2006**, *106*, 506.
(2) Stirpe, F.; Battelli, M. *Cellular and Molecular Life Sciences* **2006**, *63*, 1850.
(3) Endo, Y. *Journal of Biological Chemistry* **1987**, *262*, 8128.
(4) Grosjean, H.; Benne, R. *Modification and editing of RNA*; ASM Press, 1998.
(5) Rice, P. A.; Correll, C. C. *Protein-nucleic acid interactions: structural biology*; Royal Society of Chemistry, 2008.
(6) Chen, Y. C.; Lim, C. *Nucleic Acids Research* **2008**, *35*, 29.
(7) Li, X.-P.; Chiou, J.-C.; Remacha, M.; Ballesta, J. P. G.; Tumer, N. E. *Biochemistry* **2009**, *48*, 3853.
(8) Shazman, S.; Mandel-Gutfreund, Y. *PLoS Computational Biology* **2008**, *4*, e1000146.
(9) Lejeune, D.; Delsaux, N.; Charloteaux, B. *Proteins* **2005**, *61*, 258.
(10) Hou, Y.; Zhang, X.; Jason, A. *Nucleic Acids Research* **2001**, *29*, 976.
(11) Peumans, W. J.; Hao, Q.; van Damme, J. M. *The FASEB Journal* **2001**, *15*, 1493.
(12) Robb, G. B.; Rana, T. M. *Mol.Cell* **2007**, *25*.
(13) Lunde, B. M.; Moore, C.; Varani, G. *Nat. Rev .Mol .Cell Biol.* **2007**, *8*.
(14) Chen, Y.; Varani, G. *FEBS Letters* **2005**, *272*, 2088.
(15) Sanchez-Diaz, P.; Penalva, L. O. *RNA Biology* **2006**, *3*, 101.
(16) Fiorenzo, S. *Toxicon* **2004**, *44*, 371.
(17) Stirpe, F. *ToxiconHighlights in plant toxins* **2004**, *44*, 371.
(18) Nielsen, K.; Boston , R. S. *Annual Review of Plant Physiology and Plant Molecular Biology* **2001**, *52*, 785.
(19) Olsnes, S.; Sandvig, K.; Eiklid, K.; Pihl, A. *Journal of Supramolecular Structure* **1978**, *9*, 15.
(20) Lord, J. M.; Roberts, L. M.; Robertus, J. D. *The FASEB Journal* **1994**, *8*, 201.
(21) Hartley, M. R.; Chaddock, J. A.; Bonness, M. S. *Trends in Plant Science* **1996**, *1*, 254.
(22) Duggar, B. M.; Armstrong, J. K. *Annals of the Missouri Botanical Garden* **1925**, *12*, 359.
(23) Kataoka, J.; Habuka, N.; Furuno, M.; Miyano, M.; Takanami, Y.; Koiwai, A. *Journal of Biological Chemistry* **1991**, *266*, 8426.
(24) Zarling, J. M.; Moran, P. A.; Haffar, O.; Sias, J.; Richman, D. D.; Spina, C. A.; Myers, D. E.; Kuebelbeck, V.; Ledbetter, J. A.; Uckun, F. M. *Nature* **1990**, *347*, 92.
(25) Hofmann, B.; Jakobsen, K.; Odum, N.; Dickmeiss, E.; Platz, P.; Ryder, L.; Pedersen, C.; Mathiesen, L.; Bygbjerg, I.; Faber, V. *The Journal of Immunology* **1989**, *142*, 1874.
(26) Frankel, A. E.; Burbage, C.; Fu, T.; Tagge, E.; Chandler, J.; Willingham, M. C. *Biochemistry* **1996**, *35*, 14749.
(27) Hirao, I.; Madin, K.; Endo, Y.; Yokoyama, S.; Ellington, A. D. *Journal of Biological Chemistry* **2000**, *275*, 4943.
(28) Endo, Y. *Journal of Biological Chemistry* **1987**, *262*, 5908.
(29) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H.; Pardi, A. *Journal of Molecular Biology* **1996**, *264*, 968.
(30) Barbieri, L.; Ferreras, J. M.; Barraco, A.; Ricci, P.; Stirpe, F. *Biochemical Journal.* **1992**, *286*, 1.
(31) Barbieri, L.; Valbonesi, P.; Bonora, E.; Gorini, P.; Bolognes, A.; and Stirpe, F. *Nucleic Acids Research* **1997**, *25*.
(32) Miller, D. J.; Ravikumar, K.; Shen, H.; Suh, J. K.; Kerwin, S. M.; Robertus, J. D. *Journal of Medicinal Chemistry* **2002**, *45*, 90.
(33) Stillmark, H. *PhD Thesis* **1888**.
(34) Endo, Y.; Mitsui, K.; Motizuki, M.; Tsurugi, K. *The Journal of biological chemistry* **1987**, *262*, 5908.
(35) Endo, Y.; Tsurugi, K. *Nucleic Acids Symposium Series* **1988**, *263*, 139.
(36) Endo, Y.; Tsurugi, K. *Journal of Biological Chemistry* **1988**, *263*, 8735.

(37) Frankel, A.; Welsh, P.; Richardson, J.; Robertus, J. D. *Molecular and cellular biology* **1990**, *10*, 6257.
(38) Gaigalas, A. K.; Cole, K. D.; Bykadi, S.; Wang, L.; DeRose, P. *Photochemistry and Photobiology* **1991**, *83*, 1149.
(39) Ready, M. P.; Kim, Y. S.; Robertus, J. D. *Proteins-Structure Function and Genetics* **1991**, *10*, 270.
(40) Wannemacher, R.; Hewetson, J.; P., L. *Toxinology Research* **1992**, 108.
(41) Michael, J.; Roberts, M.; Robertust, J. O. N. D. *The FASEB Journal* **1994**, *8*, 201.
(42) Ehrlich, P. *Deutsch Med Wochenschr* **1891**, *17*, 976.
(43) Schnell, R.; Borchmann, P.; Staak, J. O.; Schindler, J.; Ghetie, V.; Vitetta, E. S.; Engert, A. *Annals of Oncology* **2003**, *14*, 729.
(44) Vitetta, E.; Thorpe, P.; Uh, r. J. *Trends in Pharmacological Sciences* **1993**, *14*, 148.
(45) Schmidberger, H.; King, L.; Lasky, L. C.; Vallera, D. A. *Cancer Research* **1990**, *50*, 3249.
(46) Knight, B. *British Medical Journal* **1976**, *278*, 350.
(47) Maman, M.; Yehezkelli, Y.; Fong, I. W., Alibek, K., Eds.; Springer US: 2005, p 205.
(48) Sandvig, K.; van Deurs, B. *FEBS Letters* **2002**, *529*, 49.
(49) Lord, J. M. *Biochemical Society Transactions* **2003**, *31*, 1260.
(50) J Katzin, B.; J Collins, E.; D Robertus, J. *Proteins* **1991**, *10*, 251.
(51) Montfort, W. *Journal of Biological Chemistry* **1987**, *262*, 5398.
(52) Rutenber, E. *Proteins* **1991**, *10*, 240.
(53) Monzingo, A. F.; Robertus, J. D. *Journal of Molecular Biology* **1992**, *227*, 1136.
(54) Ho, M.-C.; Sturm, M. B.; Almo, S. C.; Schramm, V. L. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 20276.
(55) Sturm, M. B.; Roday, S.; Schramm, V. L. *Journal of the American Chemical Society* **2007**, *129*, 5544.
(56) Chaddock, J. A.; Roberts, L. M. *Protein Engineering Design and SelectionProtein Eng.* **1993**, *6*, 425.
(57) Ready, M. P.; Kim, Y.; Robertus, J. D. *Proteins* **1992**, *10*, 270.
(58) Monzingo, A. F.; Roberts, D. J. *Journal of Molecular Biology* **1992**, *227*, 1136.
(59) Roday, S.; Saen-Oon, S.; Schramm, V. L. *Biochemistry* **2007**, *46*, 6169.
(60) Yan, X.; Hollis, T.; Svinth, M.; Day, P.; Monzingo, A. F.; Milne, G. W. A.; Robertus, J. D. *Journal of Molecular Biology* **1997**, *266*, 1043.
(61) Bradley, J. L.; McGuire, P. M. *International Journal of Peptide and Protein Research* **1990**, *35*, 365.
(62) Kim, Y. *Biochemistry* **1992**, *31*, 3294.
(63) Yan, X.; Day, P.; Hollis, T.; Monzingo, a. F.; Schelp, E.; Robertus, J. D.; Milne, G. W.; Wang, S. *Proteins* **1998**, *31*, 33.
(64) Rutenber, E.; Katzin, B. J.; Ernst, S.; Collins, E. J.; Mlsna, D.; Ready, M. P.; Robertus, J. D. *Proteins: Structure, Function, and Bioinformatics* **1991**, *10*, 240.
(65) Robertus, J. D.; Monzingo, A. F. *Mini-Reviews in Medicinal Chemistry* **2004**, *4*, 477.
(66) Mlsna, D.; Monzingo, A. F.; Katzin, B. J.; Ernst, S.; Robertus, J. D. *Protein Science* **1993**, *2*.
(67) Hajduk, P. J.; Huth, J. R.; Fesik, S. W. *Journal of Medicinal Chemistry* **2005**, *48*, 2518.
(68) Laurie, A. T.; Jackson, R. M.; Q-SiteFinder. *Bioinformatics* **2005**, *21*, 1908.
(69) Wahome, P. G.; Bai, Y.; Neal, L. M.; Robertus, J. D.; Mantis, N. J. *Toxicon* **2010**, *56*, 313.
(70) Schramm, V. L. *Annual Review of Biochemistry* **1998**, *67*, 693.
(71) Pechukas, P. *Annual Review of Physical Chemistry* **1981**, *32*, 159.
(72) Lienhard, G. E. *Science* **1973**, *180*, 149.
(73) Wolfenden, R. *Accounts of Chemical Research* **1972**, *5*, 10.
(74) Chen, X. Y.; Berti, P. J.; Schramm, V. L. *Journal of the American Chemical Society* **2000**, *122*, 1609.
(75) Chen, X.-Y.; Link, T. M.; Schramm, V. L. *Journal of American Chemical Society* **1996**, *118*, 3067.

(76) Roday, S.; Amukele, T.; Evans, G. B.; Tyler, P. C.; Furneaux, R. H.; Schramm, V. L. *Biochemistry* **2004**, *43*, 4923.
(77) Tanaka, K. S.; Chen, X. Y.; Ichikawa, Y.; Tyler, P. C.; Furneaux, R. H.; Schramm, V. L. *Biochemistry* **2001**, *40*, 6845.
(78) Chen, X. Y.; Link, T. M.; Schramm, V. L. *Biochemistry* **1998**, *37*, 11605.
(79) Jayakody, R. S. *MSc Thesis* **2008**.
(80) Pallanca, A.; Mazzaracchio, R.; Brigotti, M.; Carnicelli, D.; Alvergna, P.; Sperti, S.; Montanaro, L. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1998**, *1384*, 277.
(81) Endo, Y. *Journal of Molecular Biology* **1991**, *221*, 193.
(82) Yang, X. *Natural Structural Biology* **2001**, *8*, 968.
(83) Chen, X. Y. *Journal of the American Chemical Society* **2000**, *122*, 6527.
(84) Unrau, P. J.; Bartel, D. P. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 15393.

CHAPTER - 2

CATALYSED REACTIONS

2.1. Introduction

The term "catalysis" was first introduced by Berzelius in 1836 to designate substances that increase the rate of chemical reactions without themselves undergoing any changes.[1] The IUPAC Gold Book[2] of nomenclature defines the catalyst as *"A substance that increases the rate of a reaction without modifying the overall standard Gibbs energy change in the reaction".* Enzymes are macro bio molecules (mostly protein in nature) that act as catalysts in biochemical reactions and each enzyme catalyses a specific biochemical reaction. Since enzymes have evolved for millions of years to catalyse their specific biochemical reaction, they are far more efficient than any of the man-made catalysts. Any catalytic (catalysed) reaction can be identified by the following features[1,3,4]: (i) during the catalytic cycle the mass and the chemical composition of the catalyst does not change; (ii) A very small quantity of the catalyst makes significant changes to the rate of the reaction that it catalyses. (iii) A catalyst cannot initiate a chemical reaction and it can only increase the rate of it and therefore, the overall free energy change of the reaction is not affected by the catalyst; (iv) The final state of the equilibrium is not affected by the catalyst as it alters the rate of both forward and backward reactions by the same extent. In this chapter, catalysis is discussed with a special focus on enzyme catalysis.

2.2. Activated Complexes and the Catalytic Coefficient

A general catalytic reaction can be presented by equations 1 and 2.

[1]

where, S presents the substrate, *Cat* represents the catalyst, and I represent an intermediate, which is known as the activated (catalyst-substrate) complex.

[2]

where, P represents the final products, I represent the intermediate from equation 1 and X represents some other chemical species that reacts with I (only in certain reactions) to form the products in the presence of catalysis *Cat.* The rate constants k'_1 and k_2 (in equation 1 and equation 2 respectively) are first order rate constants whereas k_1 (in equation 1) is a second order rate constant.

If the rate of conversion of I back into S and *Cat* is greater than formation of product (i.e. $k'_1 >> k_2$), the concentration of the intermediate/activated complex, [*I*] can be calculated only by considering the equilibrium given in equation 1. Therefore, the overall rate of the reaction can be given as,

[3]

Activated complexes that are similar to *I* that revert back to the starting material faster than the formation of the product are defined as Arrhenius complexes and the corresponding reactions have the rate laws similar to that given in equation 3.

However, if complex *I* is very unstable, the rate of its conversion to the products is not small compared to its rate of conversion back to the reactants. In such cases, equation 1 alone cannot be used to calculate the concentration of the complex, *I* i.e. equation 2 cannot be neglected. For such complex, its concentration has to be calculated using the steady-state-treatment.

These complexes are referred to as Van't Hoff complexes[5].

For simplicity if the second chemical species, X, is omitted from equation 2, equations 1 and 2 can be written to give a very simple catalysed reaction,

[4]

[5]

The rate of the above reaction can be given by

[6]

If the Arrhenius complex is formed in the reaction represent by equation 4 and equation 5, the concentration of *I* is given by

[7]

Where the equilibrium constant, K = k_1 / k'_1 and hence,

[8]

For the Vant Hoff type complexes, the concentration of the intermediate, [*I*] is determined by applying steady state conditions to *I*. Therefore,

[9]

[10]

And therefore, the rate is given by:

It can be seen that equation 11 (Vant Hoff) reduces to equation 8 (Arrhenius) when $k'_1 >> k_2$. Therefore, the Vant Hoff relationship is considered as the general equation. From equations 11 and 8, it is clear that the rate of the catalysed reaction is dependent on the concentration of the catalyst. Therefore, for a given concentration of catalyst, [*Cat*], the rate can be given as,

[12]

where,

for Vant Hoff complexes .

In comparison, for the Arrhenius complexes,

The constant k'' is directly proportional to the concentration of the catalyst, [*Cat*] and it is known as the *catalytic coefficient* for the catalyst.

2.3. Kinetics of Enzyme-Catalysed Reactions

Having discussed the kinetics of general catalysed-reactions in the previous section, this section discusses the kinetics of enzyme-catalysed reactions. A biochemical reaction catalysed by an enzyme *E* can be represented as[6,7],

[15]

Where S and E respectively denote substrate and enzyme and ES represents the enzyme-substrate complex. Usually, the concentration of the substrate is much larger than that of the enzyme. Consequently, only a very small portion of substrate binds to the enzyme. Therefore, the total concentration of the enzyme $[E]_T$ can be considered as the sum of the concentration of the free enzyme , ($[E]$) and the concentration of the enzyme-substrate complex ([ES]),

[16]

When the steady state approximation is applied to the enzyme: substrate complex, it yields:

[17]

or

where[and represent the concentrations of the enzyme and substrate respectively.

When $[ES]$ is expressed in terms of $[E]_T$ equations 16, 17 and 18 can be used show that:

In general, it is assumed that the rate determining step in an enzymatic catalytic cycle is the conversion step of the enzyme: substrate complex into the products. Therefore, the rate of the reaction can be written as,

Substitution of value of [ES] from equation 19 to equation 20 leads to:

where,

The ratio, , given in equation 22 is known as the Michaelis-Menten constant[8]. Consequently the scheme represented by equation 13 is referred to as the Michaelis-Menten kinetic model for enzymes. Moreover, equation 22 is known as the Michaelis-Menten equation. This equation predicts the kinetics of the enzyme that will show saturation when [S]>>k_M. When the enzyme is saturated, the rate of the reaction is known as the maximum velocity (V_{max}) and it is defined as the maximum rate that an enzyme-catalysed reaction can take place. When $[S]>>k_M$, the rate of the reaction is given by:

[23]

Further, in such cases, the reaction will be zero order with respected to the substrate concentration. However, when the substrate concentrations are relatively small, the rate becomes,

Therefore, equation 24 can be used to show the 1st order dependence of the rate on substrate concentration at relatively small values. The relationship between the rate of the reaction and the [S] is graphically represented in Figure 1.

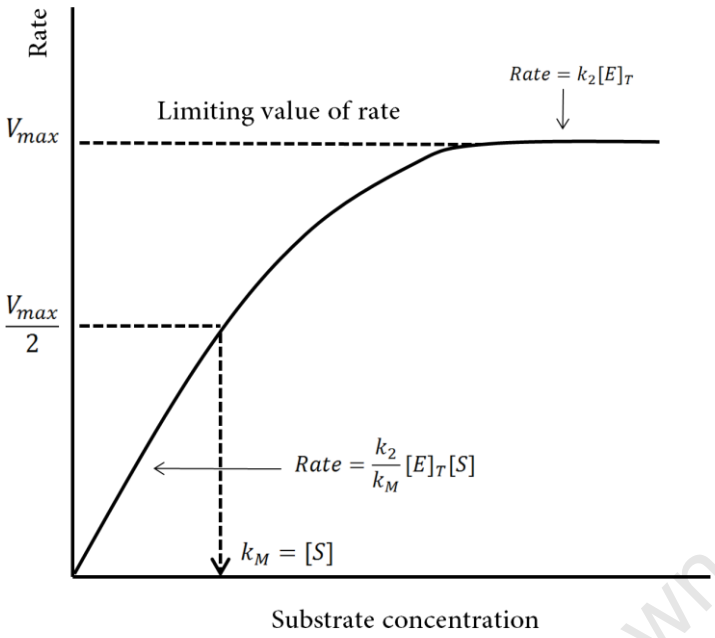


Figure 1: Dependence of the reaction rate on the substrate concentration in an enzymatic reaction[9].

As mentioned before, the rate becomes independent of [S] at high substrate concentrations. In such situations, relationships 25 and 26 become true and the rate of the reaction can be given by equation 27.

[25]

[26]

Moreover, when $k_M = [S]$, the rate of the reaction can be given by:

[28]

Consequently, k_M can be expressed as follows with respect to the reaction speed.

$$k_M = \text{the substrate concentration when the rate of the reaction is excatly half of its maximum rate} \quad [29]$$

The constant k_2 in above equations is usually referred to as $k_{cat.}$ This refers to the rate constant for converting the substrate in to the products inside the active site of the enzyme and it is also known as the *turnover* number. For unimolecular reactions it has the units of s^{-1}. Further, it is the rate constant that corresponds to the chemical step in the catalytic cycle. The values of k_{cat} and k_M are determined experimentally. The value of k_M depends on the experimental conditions such as pH, temperature, solvent, ionic strength and the nature of the substrate.

The Michaelis constant, k_M indicates the strength of binding and degree of saturation in enzyme-substrate complexation and no information about chemical transformation. The ratio of k_{cat} to k_M is defined as the specificity constant. The specificity constant is helpful in comparing competing substrates. Further, it is acts a single measure of both binding strength and the catalytic efficiency. Therefore, it is referred to as the apparent rate constant. Moreover, this ratio is generally referred to as the efficiency of the enzyme which gives the number of molecules of substrate that is converted in to products per second.

2.4. Energetics of Enzymatic Reactions

As any other catalysis, enzymes increase the rate of their corresponding reactions by providing alternative pathways for reactant to product transformations. A schematic representation of this concept is presented in Figure 2. On the alternative path a lesser amount of energy is required for the conversion of reactants to products. This is due to the lower energy barrier (activation energy) in the alternative path.

Therefore, a larger number of reaction cycles can be completed in a given time in a catalysed reaction compared to its uncatalysed counterpart. Generation of the alternative paths in enzymes is facilitated their ability to stabilise the corresponding activated complexes which ultimately lowers the activation energy.

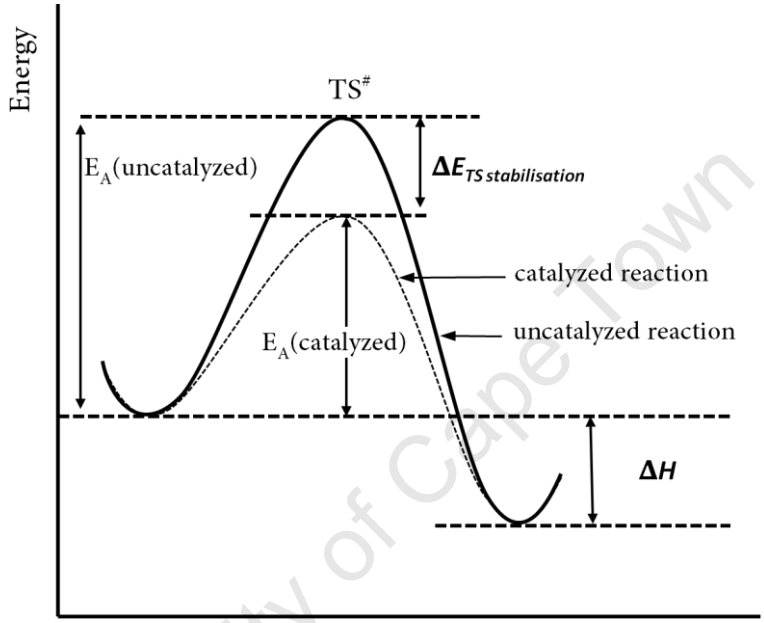


Figure 2: Energetics of a catalysed reaction vs. an uncatalysed reaction.

The above general energy profile has to be modified based on the nature of the activated complex. Reactions with Arrhenius complexes and Hoff complexes show distinct energy profiles and they are presented in Figure 3.

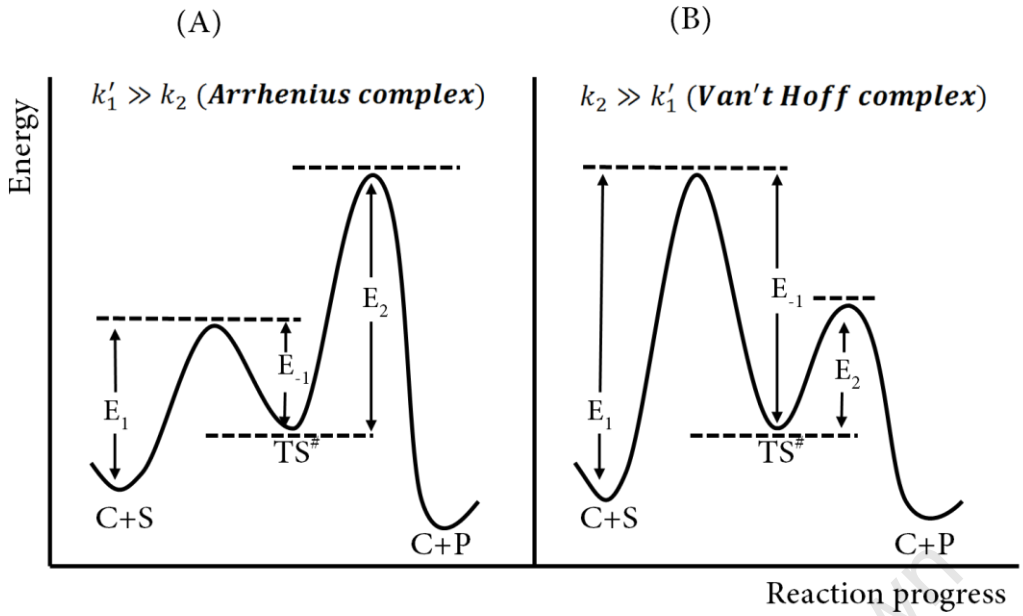
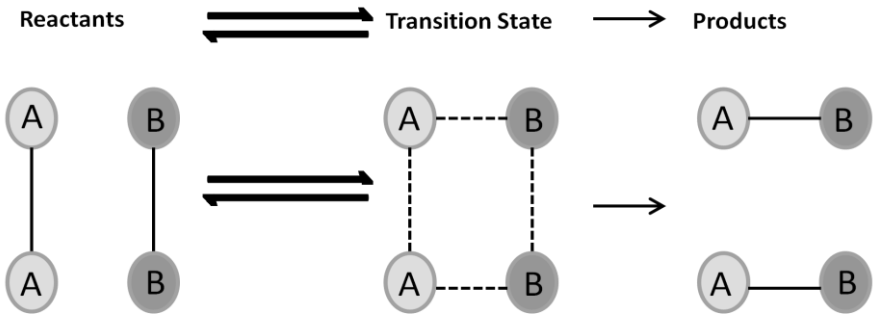


Figure 3: (A) Energy profile of a reaction with Arrhenius complex; E2 controls the rate. (B) Energy profile of a reaction with Van't Hoff complex ; E1 controls the rate.

2.5. Rates of Catalysed Reactions

The most accepted theoretical framework to study enzymatic reactions is provided by the Transition State Theory (TST) [10-13]. The TST was first formulated by Henry Eyring in 1935[14] . Later, Polanyi proposed different approaches to study TST[15] ; (1)the thermodynamic approach, (2) the kinetic-theory approach, and (3) the statistical-mechanical approach.

Even though the major forms of TST , Generalised TST (GTST) and conventional TST[16] were originally proposed for reactions in the gas phase, the current version of GTST is readily used in analyzing the enzymatic reactions. According to TST, reactant molecules are first converted into a highly unstable state called the Transition State (TS) before conversion into products. This idea is schematically presented in the Scheme 1.



Scheme 1: Schematic representation of the concept of Transition State Theory.

In TST theory it is assumed that the activated complex is an independent entity and that it is in equilibrium with the reactants. This idea is presented in equation 30.1 Equations 30.1 to equations 30.4 show the derivation of the rate expression in terms of the concentration of the transition state.

[30.1]

[30.2]

[30.3]

[30.4]

In equations 30.1- 30.4, $K^{\#}$ refers to the equilibrium constant and k refers to the experimentally determined rate constant for the conversion of reactants (A+B) to the products. The rate constant corresponding to the conversion of TS to the products is given by $k^{\#}$.

Moreover, the concentration of the TS, ($[TS]^{\#}$) can be given as in equation 31 based on equilibrium expressed in equation 30.1,

$$[TS]^{\#} = K^{\#}[A][B] \quad [31]$$

Therefore, it can be shown from the equation 30.4 and 31 that ,

$$k \quad = k^{\#}K^{\#} \quad [32]$$

At the transition state, there exists a very unstable vibrational degree of freedom which is responsible for disrupting the TS and converting it into the products (or reactants)[17,18]. Because of its very low frequency, the average energy of such vibration will be in the order of $k_B T$,where k_B is the Boltzmann constant and T is the temperature. As a result, transition states have lifetimes no longer than a single vibration. The vibrational energy of a TS can be given as,

$$E_{vib} k_B T = h\nu \quad [33]$$

Consequently, the frequency of the vibration can be given as

$$\nu = \frac{k_B T}{h} \quad [34]$$

In equations 33 and 34, h refers to the Planck's constant. It is assumed[11,12] that the rate of the transformation of the TS in to the products depends on (a) Frequency of vibration of the TS and (b) The concentration of the TS.

Therefore, the rate can be given as,

$$Rate = [TS]^{\#}\, \nu \quad [35]$$

Substitution for [TS]# from equation 31 in to equation 35 leads to :

$$Rate = \nu K^{\#}[A][B] \quad [36]$$

Substitution for the rate from equation 30.4 and substituting for ν from equation 34 to equation 36 leads to:

$$k = \frac{k_B T}{h} K^{\#} \quad [37]$$

The constant $K^{\#}$ can be expressed either by using thermodynamic functions or partition functions. The thermodynamic approach uses the fact that the equilibrium constant of a chemical reaction is related to the standard free energy change, ΔG^0 as given below.

Δ [38]

The free energy is related to the enthalpy and to the entropy via the Gibbs–Helmholtz relationship:

Δ Δ Δ [39]

From equations 38 and 39, it can be shown that,

Δ Δ Δ Δ [40]

When equation 40 is applied to the formation of the TS , the corresponding change in free energy, enthalpy and entropy are given by Δ Δ Δ respectively. Then the equilibrium constant can be given as:

Δ [41.1]

Δ Δ [41.2]

When the values of from equations 41.1 and 41.2 are substituted to equation 37, it gives arise to equations 42 and 43.

___ Δ [42]

___ Δ Δ [43]

Equation 42 and 43 are used to experimentally calculate the thermodynamic entities ($\Delta G^{\#}, \Delta H^{\#}, \Delta S^{\#}$) of activation[5].

In the statistical mechanics approach, the equilibrium constant for a simple system such as the one given in equation 44 can be given by equation 45 in terms of partition function[5] (see the following section on Statistical Mechanics).

$$aA + bB \leftrightharpoons cC + dD \qquad [44]$$

$$K_c = \frac{Q_C^c Q_D^d}{Q_A^a Q_B^b} \quad e^{-E_0/RT} \qquad [45]$$

In equation 45, Q's represents the partition function per unit volume of the molecules A, B, C and D where *a, b, c* and *d* represents the number of moles. The energy increase at absolute zero for this reaction is given by E_o. Therefore, it is considered as the hypothetical energy of activation. When the concept of equation 45 is applied to equation 37 i.e. to the activated complex / transition state, the rate constant can be written as,

$$k \quad = \frac{k_B T}{h} \frac{Q^{\#}}{Q_A\, Q_B} \quad e^{-E_0/RT} \qquad [46]$$

Where $Q^{\#}$ is the partition function of the activated complex (Transition state). In reality however, not every collision associated with required v, i.e. with required energy leads to products. This can be due to the improper alignment/arrangement of the atoms for the transformation from reactants to products or due to the interference of the rotational state of the molecule with regards to the chemical transformation. In transition state theory, these factors are taken into account by introducing a factor called transmission coefficient, γ. With γ taken in to account, equation 37 has to be re-written as,

$$k = \gamma \frac{k_B T}{h} K^{\#} \qquad [47]$$

In most cases γ is near unity and can be simply ignored. When γ is not omitted , the rate constant for a given chemical reaction at temperature T is given by[17] (based on equation 40),

— Δ [48]

where, β= and and refer to the Boltzmann constant and Planck's constant respectively. According to simple transition state theory, γ has three components[19], Γ, κ and g. Where Γ accounts for dynamical recrossing of the transition state hyper surface, κ accounts for quantum mechanical tunnelling in the reaction coordinate and g accounts for non equilibrium disturbance in phase space. The components Γ and κ are known as recrossing transmission coefficient and tunnelling transmission coefficient respectively. With respect to these components, γ is given by[20],

Γ [49]

2.6. Statistical Mechanics Approach to Reaction Rates

In Transition State Theory, Statistical Mechanics(SM) play a significant role.[21,22] Since the chemical equilibrium is a dynamic process, chemical equilibrium and reaction rates can be studied by molecular motions and the statistical distribution of the molecular speeds[23,24]. Distribution of the molecular speeds is the essence of the Maxwell-Boltzmann distribution law[25]. Therefore, the Maxwell-Boltzmann distribution law essentially draws the relationship amongst Statistical Mechanics, chemical kinetics and reaction dynamics. The basics of statistical mechanics are discussed in the next chapter whereas its applications to the reaction rates are discussed below.

When molecules react, the energy increase at the absolute zero temperature is given by the partition function, Q, per unit volume and this can be given as:

[50]

Where e_i is the energy with respect to zero point energy for a given energy state of the molecule, is the number of states corresponds to that energy level and k_B is the Boltzmann constant and T is temperature. Since includes contributions from electronic energy E_i , translational energy E_t , vibrational energy E_v and rotational energy , it can be given as,

$$e_i = E_{e_i} + E_{t_i} + E_{r_i} + E_{v_i} \quad [51]$$

Consequently, equation 50 can be given as:

$$Q = \sum_i g_{e_i} e^{-Ee_i/\,k_B T} g_{t_i} e^{-Et_i/\,k_B T} \quad g_{r_i} e^{-Er_i/\,k_B T} g_v e^{-Ev_i/\,k_B T} \quad [52]$$

this leads to

$$Q = q_e^e q_t^t q_r^r q_v^v = Q_e Q_t Q_r Q_v \quad [53]$$

Where q_e, q_t, q_r and q_v are partition functions for the electronic, translational, rotational and vibrational degrees of freedom respectively. Table 1 presents the expressions for these partition functions and approximate order of magnitude.

Motion	Degree of Freedom	Partition function	Order of magnitude
Translational	3	$(2\pi m k_b T)^{1/2}/_{h^3}$	$10^{24} - 10^{25}$
Rotational linear	2	$8\pi^2 I k_b T/h^2$	$10 \quad - 10^2$
Non linear	3	$8\pi^2 (8\pi^3 ABC)^{\frac{1}{2}} (k_b T)^{3/2}/h^3$	$10^2 - 10^3$
Vibrational	1	$(1 - e^{-hv/k_b T})^{-1}$	$1 - 10$
Restricted rotation	1	$(8\pi^2 I k_b T)^{3/2}/h^2$	$1 - 10$

Table 1: Partition functions

Based on equation 45, the equilibrium constant for formation of the activated complex can be given as ,

$$K^{\#} = \frac{Q^{\#}}{Q_A\, Q_B} \quad e^{-E_0/RT} \quad [54]$$

However, in the activated complex /transition state, one vibration (along the reaction coordinate) is quite different from the other vibrations as this vibrational degree of freedom will disappear as the TS is converted to products. Hence, to account for this phenomenon requires a correction to be made to the vibrational partition function.

With this correction, the number of vibrational degrees of freedom is given by 6 for activated linear molecules and $3(N_A + N_B) - 7$ for non-linear activated molecules. Without the correction these terms are usually given as 5 and $3(N_A + N_B) - 6$ respectively. The value of vibrational partition function is given by[5] $(1 - e^{-hv/\,k_B T})^{-1}$. It can be shown that when $v \rightarrow$zero, the vibrational partition function becomes:

lim [55]

Inclusion of equation 55 into equation 54 gives:

[56]

Therefore, the concentration of the activated complex can be given as:

[57]

By substituting the concentration of the TS, into equation 35, the rate of the reaction can be given in terms of partition functions as follows:

[58]

Experimentally determined rate is given as

[59]

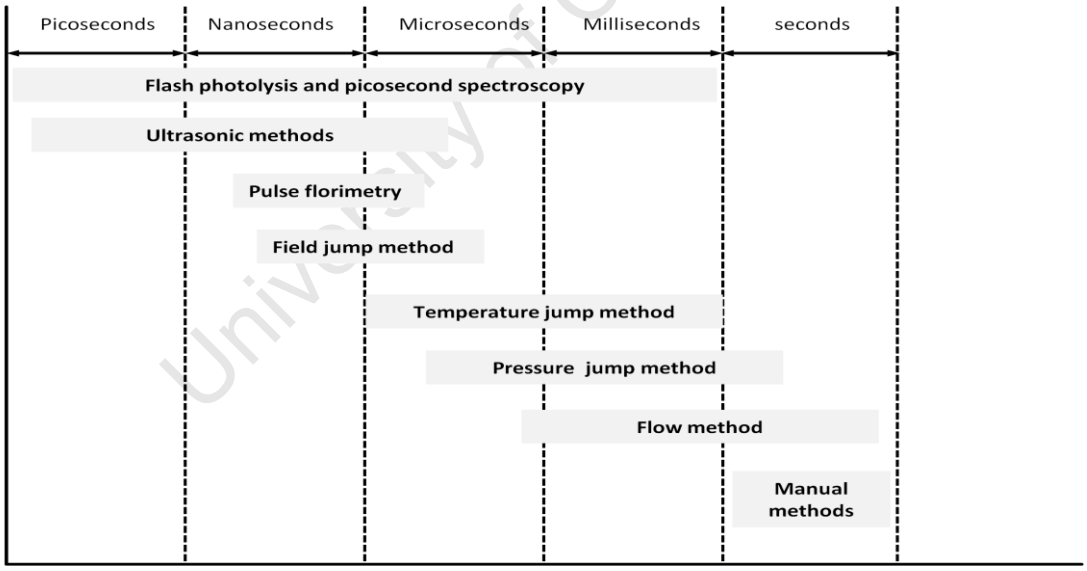
From equations 58 and 59, the rate constant can be given as:

[60]

Equation 60 expresses the rate constant in terms of partition functions.

2.7. Experimental Methods for Following Kinetics[3,25]

Enzyme activities are in general studied under steady-state conditions. Two major pieces of kinetic data that are usually obtained from these studies are: k_{cat} (turnover number) and k_M (Michaelis-Menten constant). The value of k_{cat} lies between 1 and 10^7 per second[3], S^{-1}. Therefore, the measurements must be taken in these time scales. In order to achieve that goal, methods that allow rapid mixing, observation and analysis are required. Some widely used such techniques and their time scales are schematically presented in Scheme 2. While these specific techniques provide the primary means of initiating and controlling the fast enzymatic reactions, another standard experimental technique such as spectroscopy is generally used as a secondary means to analyse and collect the data. Amongst these secondary methods, electrochemical methods circular dichroism, radioactive procedures and the spectroscopic methods are the most common[26].



Scheme 2: Time resolution of different methods that are used in studying rapid reactions[6].

2.8. Experimental Methods to Study Reaction Mechanisms

There are various experimental methods that can be used to elucidate a reaction mechanism[27]. Often, a combination of these experimental techniques is required. Some of the widely used experimental techniques for studying reaction mechanisms are discussed below.

I. Isotope Effects Experiments (IE)

In IE experiments, an atom at a selected bond is substituted with another isotope of that element and the isotope effect, i.e. the effect of isotopic substitution on the vibration of the corresponding bond is measured to determine if that bond changes in any manner during the rate determining step. The isotope effect is expressed as a ratio of the rate constants with natural abundance isotope and with the substituted isotope. When the effect of substitution changes the reaction rate and the effect is referred to as Kinetic Isotope Effect (KIE) and when it changes the position of equilibrium it is known as Equilibrium Isotope Effect (EIE). If the effect of the substitution is attributed to a bond breaking it is known as the *primary isotope effect.*

II. Substituent Effect Experiments.

The substituent effect is defined as the manner in which the reactivity of a molecule changes with respect to a change in its substituents. This information may then be utilized to propose mechanisms. The substituent effect is sometimes known as the structure-reactivity–relationship or in drug design industry it is commonly known as the structure-activity-relationship (SAR). The substituent effect is mainly due to the field effects, inductive effects, resonance effects, polarization effects, steric effects and solvation effects.

III. Linear Free Energy Relationships (LFER)

In LFER methods, it is attempted to develop quantitative relationship between the structure and the activity. A linear free-energy relationship (between two reactions) is said to exist if the same series of changes in conditions affects the rate or equilibrium of a second reaction in exactly the same way as the first reaction. Such relationships are utilized in helping to elucidate reaction mechanisms and in predicting rates or equilibria. Among many LFER methods Hammett Plots are most commonly used.

IV. Product Identification Experiments

Product identification experiments can provide valuable clues about the reaction mechanism whilst they can be used to ensure the mass balance (as states by the law con in a chemical reaction. In some cases, identification of a minor product can lead to valuable clues to the mechanism (s).

V. Trapping Methods

If long-lived intermediates are expected during the reaction, one can use some other molecules that will react with the intermediates in a predictive manner and hence trap them. This technique can be used to verify the presence of predicted intermediate(s). However, this has to be done in case-by-case manner as there is no general method to do this otherwise.

VI. Crossover Methods

These methods are used to determine if the reactants break apart to make intermediates before they recombine to form the products. This is done by isotopically labelling some of the reactants. If they break and recombine products will have a mix of labelled and non-labelled atoms otherwise they will have either labelled or non-labelled atoms.

VII. Stereochemical Analysis

These techniques can be used to verify the existence of an intermediate and/or the type of the reaction mechanism. If complete inversion of configuration is observed with the products when started with enantiomerically pure reactants, it is an indication of no intermediate and proposes S_N2 type mechanism. On the other hand complete or partial racemisation indicates the presence of a planar intermediate and a S_N1 type mechanism is likely.

VIII. Isotope Scrambling

In this method a part of the reactant is isotopically labelled and the presence of this isotope in the corresponding position of the product is then investigated. This analysis can indicate the presence of a symmetrical intermediate formed.

IX. Direct Isolation and Characterisation of Intermediates

If the intermediates are stable enough to be isolated, standard methods such as spectroscopy can be used to characterise them

X. Transient Spectroscopy

These are exceptionally fast spectroscopical methods to analyse the reaction intermediates "on-the-fly". Ultra fast laser spectroscopy is such a technique[28]. This method has extended reaction-dynamic studies into the picosecond and femtosecond time scales and allows experimental observations of transitory fragments that occur during collisions. This special method probes the short-lived activated complexes/transition states and to observe molecular reactions in progress which allows real-time viewing of energy redistributions.

2.9. Enzyme Inhibition

Enzyme inhibition is defined as a reduction in enzyme activity caused by specific binding of a ligand to a defined binding site at the enzyme[3]. Such ligands are defined as inhibitors. When this action is performed by enzyme's own product (product of the reaction catalysed by the enzyme), it is known as product inhibition. When the product is bound to the enzyme it reduces the enzyme's opportunities to bind to the substrate.

A certain degree of product inhibition is expected in all enzymes. However, it will not be pronounced in all cases. It can be shown that the Michaelis constant for the product can be given as an inhibition constant K_P after using the Haldane relationship on equation 15 with the assumption of reversible binding of the products. Therefore, K_P can be given as, [3]

[61]

Consequently the velocity of the reaction is given by :

Where, is the Michaelis constant for the substrate, is product inhibition constant, is the substrate concentration, is the product concentration and is the maximum velocity (the maximum rate) with respect to the substrate. Rearrangement of equation 62 into double reciprocal form allows one to calculate by using experimental data.

Based on the receptor site specificity, enzyme inhibition is classified as specific inhibition and non-specific inhibition. In specific inhibition, the inhibitor binds to a very specific, well identified receptor site i.e. to the binding site.

In contrast, non-specific inhibition occurs when enzyme-inhibitor interactions are not specific. These interactions may be due to unspecific surface effects, which may disturb the native enzyme structure.

Non-specific inhibitors can be identified by the following features: (i) Absence of a well defined saturation level (ii) Inability to directly compete with analogous substances (iii) Required high (< 1 mM) concentrations of the inhibitor, whereas specific inhibition shows the opposites of the above features. Figure 4 presents the comparison between unspecific binding and specific binding with respect to the saturation level.

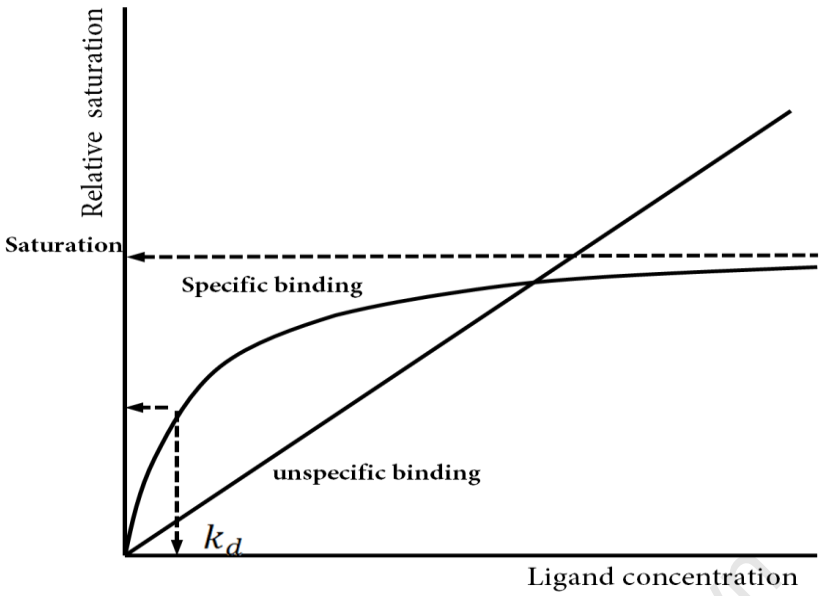


Figure 4: comparison of unspecific binding with specific binding

According to the nature of the binding process between the inhibitor and the enzyme, enzyme inhibition is classified in to two major categories; namely, reversible inhibition and irreversible inhibition. Figure 5 compares these two inhibition types in terms of enzyme activity. Dilution of the enzyme can be used as a rapid test to identify the type of inhibition. Usually on dilution, enzyme activity drops. However, in reversible inhibition, the observed drop in activity is greater than the expected drop in activity. These two categories of inhibition are briefly discussed below.

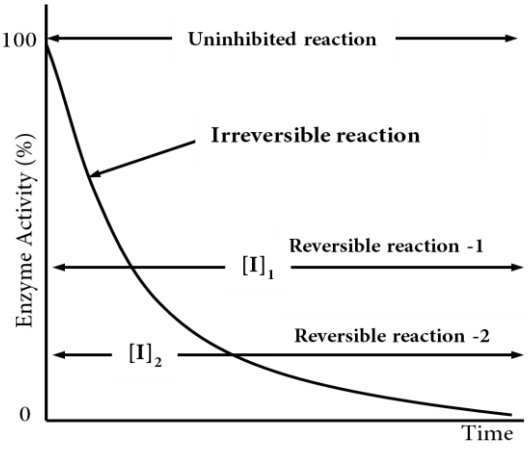


Figure 5: Reversible and Irreversible inhibition.
$[I]_1$ and $[I]_2$ refers to different inhibitor concentrations where $[I]_2 > [I]_1$.

2.9.1. Irreversible Enzyme Inhibition

Irreversible inhibition occurs when the inhibitor binds to the enzyme in a permanent fashion to reduce or to cease its activity. Irreversible inhibition can be specific or non-specific. The latter is further divided into two major classes; (i) non-specific inhibition due to covalent binding and (ii) non-specific inhibition due to non-covalent but extremely tight binding.

Inhibitors that show irreversible inhibition can be classified into two classes as: (a) suicide substrates (or substrate analogues), (b) transition state analogues. The former class of inhibitors are non-reactive mimics of the natural substrate. Therefore, the enzyme activity will be ceased upon binding of these inhibitors. The second class of inhibitors mimics the TS of the reaction catalysed by the enzyme. Therefore, these enzymes are expected to show the highest binding affinity and expect to remain bound to enzyme.

The irreversible binding of an inhibitor to an enzyme can be given by:

$$I + E \underset{k_{-1}}{\overset{k_1}{\leftrightharpoons}} EI \xrightarrow{k_2} EI_i \qquad [63]$$

The first step in equation 63 represents the initial reversible binding of the inhibitor to the enzyme. The corresponding equilibrium constants are given by $k_1$and k_{-1}. The rate constant corresponding to the conversion of initial enzyme-inhibitor complex to the inactive enzyme-inhibitor complex is given by k_2. In the presence of the inhibitor$I$, the total enzyme concentration $[\mathrm{E}]_0$ can be given as:

$$E_0 = [E] + [EI] + [EI]_i = E_a + [EI]_i \qquad [64]$$

Where (in equation 63 and 64) subscript a refers to the "active" form of the enzyme and subscript i refers the inactive form.

When the inhibitor concentration is greater than the total enzyme concentration, the time dependent formation of the inactive form, $[EI]_i$ is directly proportional to the concentration of the reversible complex, [EI] and this is given by:

Since the total enzyme concentration does not change,

Then, from equation 65 it can be shown that

can be removed from equation 65 by assuming the reversible (initial) binding. When is removed and the equilibrium constant is introduced, equation 65 becomes,

where, K and represents the inhibitor concentration. Integration of equation 68 from at time to $[EI]_t$ at time $t = t$ $(incubation\ time)$gives:

ln

If the inhibitor is competitive with the substrate A , it can be shown that

$$\ln\frac{[E]_a}{[E]_0} = \frac{k_2 t}{\left(1+\frac{K_i}{[I]}\right)\left(1+\frac{[A]}{K_a}\right)} \qquad [70]$$

whereas $[A]$ and K_a represent the concentration and the binding constant for substrate respectively.

The non-competitive equation can also be given as,

$$\log\frac{[E]_a}{[E]_0} = -\frac{k_2 t}{2.3\left(1+\frac{K_i}{[I]}\right)} \qquad [71]$$

If the time dependence of the enzyme activity in the presence of different inhibitor concentrations is plotted as $[\mathrm{E}]_a/[\mathrm{E}]_0$ against the incubation time, it will yield straight lines as shown in Figure 6.

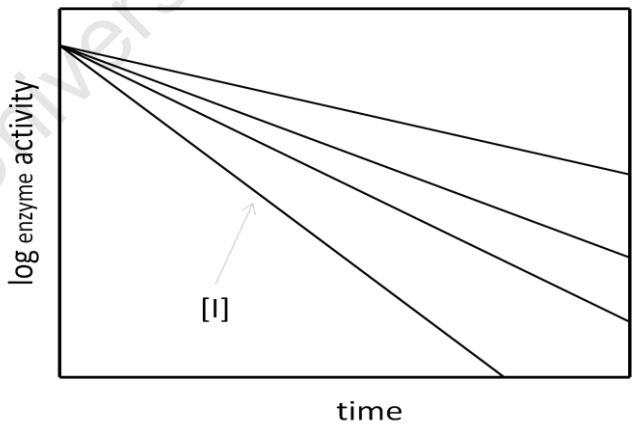


Figure 6: Time dependence of the irreversible inhibition semi-logarithmic diagram for different inhibitor concentrations.

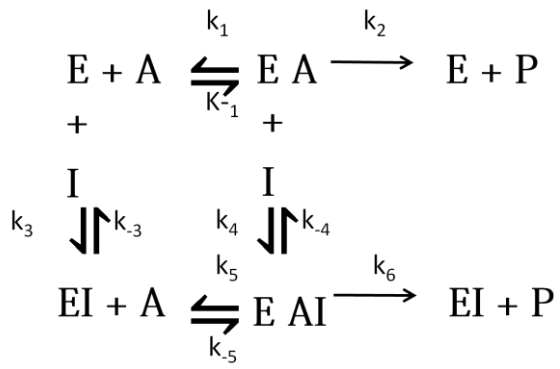
The slope of the lines given in Figure 6 is given by:

and

If the reciprocal slope is plotted against the reverse inhibitor concentration, a linear relationship is observed with y intercept of and t intercept of . Therefore, the rate constant for irreversible inhibition, is can be calculated.

2.9.2. Reversible Enzyme Inhibition

Reduction of enzyme activity by an inhibitor by binding to it in a reversible manner is called reversible inhibition. There are different ways of inhibitor-enzyme reversible binding. Therefore, different inhibition mechanisms exist. For each of these mechanisms a specific equation is required to describe the enzyme velocity. However, it is possible to a general scheme for reversible as presented in scheme 3. Scheme 3 is given under the assumptions that: (1) The inhibitor binds to the enzyme forming an complex. (2) The inhibitor can also bind to the enzyme–substrate complex, forming an complex. (3) The [EAI] complex may be inactive or active. In Scheme 3, k with (with the corresponding subscript) represents the relevant rate constants. The overall reaction mechanism contains two rate constants, $k_2$and k_6 . Starting with the steady-state differential equations for the reaction rates, an equation for enzyme velocity can be derived. This can be given be given by equation 74.



Scheme 3:A general scheme for reversible enzyme-inhibitor binding

Where, and correspond to for reaction 1 and reaction 2 (in Scheme 2) respectively. Concentration of the substrate and concentration of the inhibitor is given by and respectively. The other constants in equation 74 , and are given by :

[75]

Equation 74 gives a generalized rate equation for the binding Scheme presented in Scheme 3. According to the binding nature, reversible inhibition is further categorized and equation 72 has to be modified accordingly. Classification of reversible inhibition and the corresponding changes to equation 74 are presented in Figure 7.

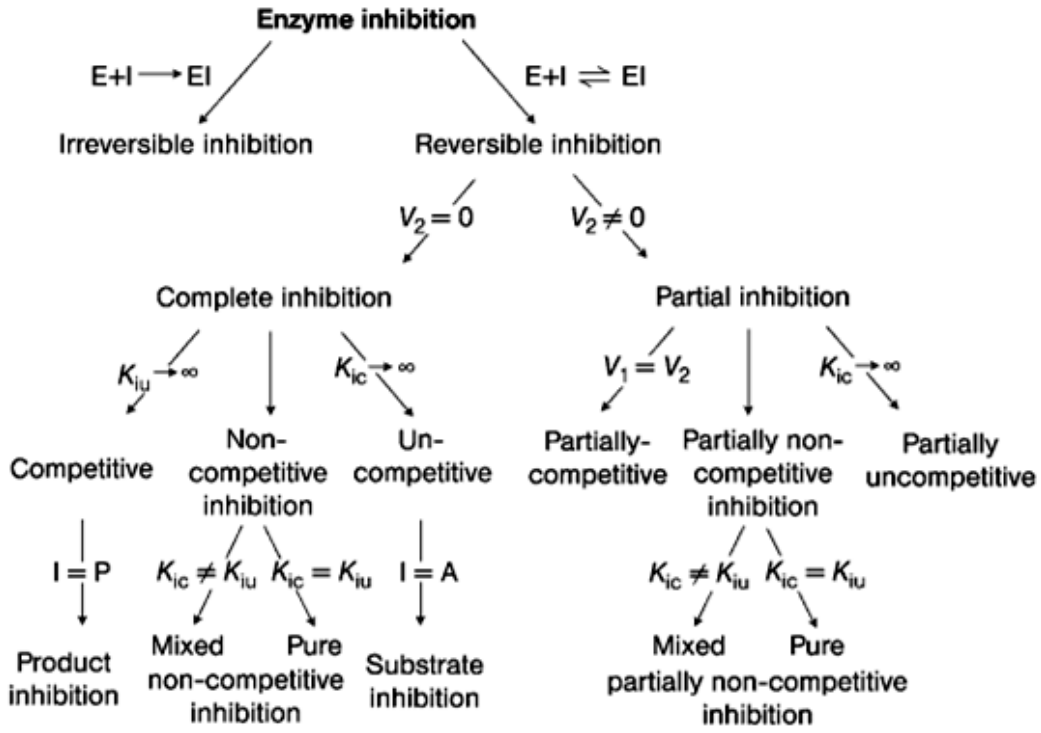


Figure 7: Summary of essential reversible inhibition mechanisms[1]

2.10. Computational Studies of Enzymatic Reactions

In equation 48 $\Delta G^{\#}(T)$ is the molar standard quasithermodynamic free energy of activation. This entity is related to so called potential of mean force (PMF). The PMF can be obtained from computer simulations based on the probability density of the regions of the phase-space that are sampled in a computer simulation. Therefore, computer simulations can be used to calculate the reaction rates and the other thermodynamic properties of a reacting system based on the principles of statistical mechanics. Furthermore, computational techniques give direct access to information such as the structure of the TS / activated complex and reaction mechanisms otherwise that is not available (or extremely difficult to obtain) experimentally. The application of computational techniques in studying enzymatic reactions is discussed in detail in chapter 3.

2.11. References

(1) Bugg, T. *Introduction to enzyme and coenzyme chemistry*; John Wiley and Sons, 2004.
(2) *IUPAC. Compendium of Chemical Terminology*; 2 ed.; McNaught , A. D.; Wilkinso, A., Eds.; Blackwell Scientific Publications. : Oxford., 1997.
(3) Bisswanger, H. *Enzyme kinetics: principles and methods*; Wiley-VCH, 2008.
(4) Page, M. I. *The Chemistry of enzyme action*; Elsevier, 1984.
(5) Upadhyay, S. K. *Chemical kinetics and reaction dynamics*; Springer, 2006.
(6) Purich, D. L. *Enzyme kinetics: catalysis & control : a reference of theory and best-practice methods*; Elsevier, 2010.
(7) Leskovac, V. *Comprehensive enzyme kinetics*; Kluwer Academic/Plenum Pub., 2003.
(8) Menten, L.; Michaelis, M. I. *Biochem Z* **1913**, *49*, 333.
(9) Leskovac, V. *Comprehensive enzyme kinetics*; Springer, 2003.
(10) Steinfeld, J. I.; Francisco, J. S.; Hase, W. L. *Chemical kinetics and dynamics*; Prentice Hall, 1989.
(11) Lasaga, A. C. *Review Mineral* **1981**, *8*, 135.
(12) Pechukas, P. *Annual Review of Physical Chemistry* **1981**, *32*, 159.
(13) Lienhard, G. E. *Science* **1973**, *180*, 149.
(14) Stearn, A. E.; Eyring, H. *The Journal of Chemical Physics* **1937**, *5*, 113.
(15) Laidler, K. J.; King, M. C. *The Journal of Physical Chemistry* **1983**, *87*, 2657.
(16) Glasstone, S.; Laidler, K. J.; Eyring, H. *The theory of rate processes : the kinetics of chemical reactions, viscosity, diffusion and electrochemical phenomena*; McGraw-Hill, 1941.
(17) Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G. *Chemical Reviews* **2006**, *106*, 3188.
(18) Truhlar, D. G.; Hase, W. L.; Hynes, J. T. *The Journal of Physical Chemistry* **1983**, *87*, 2664.
(19) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186.
(20) Pu, J.; Gao, J.; Truhlar, D. G. *Chemical Reviews* **2006**, *106*, 3140.
(21) Helrich, C. S. U. *Modern thermodynamics with statistical mechanics*; Springer, 2009.
(22) Jensen, F. *Introduction to computational chemistry*; John Wiley & Sons, 2007.
(23) Williamson, A. W. *Journal of Chemical Society* **1850**, *4*, 229.
(24) Pfaundler, L. *Annual Review of Physical Chemistry* **1867**, *131*, 55.
(25) Anslyn, E. V.; Dougherty, D. A.; University Science Books: 2006, p 1095.
(26) Fersht , H. *Structure and Mechanism in Protein Science*; 3rd ed.; W.H. Freeman and Company; New York, 1999.
(27) Anslyn, E. V.; Dougherty, D. A. *Modern Physical Organic Chemistry*; University Science Books, 2006; Vol. 69.
(28) Knee, J. L.; Zewail, A. H. *Spectroscopy* **1988**, *3*, 44.

CHAPTER - 3

COMPUTATIONAL METHODS

3.1. Introduction

Experimental techniques such as X-ray crystallography, fluorescence[1,2] infrared spectroscopy[3], NMR[4] methods and kinetic isotope effect experiments[5,6] are commonly used to investigate the dynamic behaviour of bio molecules. However, no current experimental method is capable of providing complete structural and functional information about bio molecules. Computer simulations on the other hand, are capable of providing atomistic level details of the biological/chemical systems of interest[7-9]. Computer simulations can be used to supplement the experiments or in the absence of suitable experimental techniques, they can be employed as the exclusive method of investigation. The latter is becoming increasingly popular as a result of today's high computing power, and the increasing accuracy and predictability of the programs used. An example of one such successful application is solving of the three dimension structure of a protein starting from its primary sequence using computer simulations [10].

The complexity of the system and the questions to be addressed are the primary determinants of selecting the computational technique(s). For example, to study the properties that are originating from the electronic structure, quantum mechanical (QM) methods are required. However, application of QM is limited by the size of the system. Therefore, when the structural phenomena are the only focus in a study, one can use Molecular Mechanics (MM) based methods.

In the classic view, bio molecules such as nucleic acids and proteins were considered static in nature[11]. Particularly for proteins, their x-ray crystal structures have proposed "rigid" structures for them with atoms in fixed positions[12]. However, the current understanding of these molecules reveals that their strictures are "dynamic" in character[13].

Further, it is now recognized that the atoms of these molecules are in a state of constant motion at normal temperatures. Despite this, a static picture showing the average position of the atoms is still useful for determining important structural features. The idea of fluctuations introduced in the dynamic picture leads to more accurate interpretations of functional properties. There are a vast range of experimental techniques that can be used to elucidate the structural and chemical behaviour of bio molecules.

An extremely difficult task in experiments is to evaluate the time evolution of a given chemical system. This difficult task can successfully be accomplished by means of Molecular Dynamics (MD) methods. To study the reactivity of larger systems a, hybrid Quantum Mechanical-Molecular Mechanical (QM/MM) approached is used. This chapter discusses some of the available computational methods that are relevant to simulation of biological molecules, specifically enzymes and their reactivity.

3.2. Energies of Chemical Systems

To simulate chemical systems, various mathematical models are employed with a vast range of algorithms and levels of accuracy. The most important feature of a model is its ability to obtain an accurate potential energy surface (PES) for a given chemical system; in other words, the ability to obtain a systems energy accurately as a function of its structure. For this purpose, one can use Quantum Mechanical (QM), molecular mechanics (MM) methods or so called hybrid (QM/MM) methods.

In QM methods, electrons are treated explicitly whereas in MM methods the electronic structure of the chemical system is ignored. In MM methods, empirically derived energy functions are used to describe the energetics of a given system. These classical methods allow rapid energy evaluation and conformational sampling. Under electronic-structure-methods there exist three major approaches; namely, *ab-initio* methods, Semi Empirical methods (SE) and Density Functional Theory (DFT)-based methods. These methods differ from one another depending on how they treat the electrons.

In QM/MM methods; large molecular systems are partitioned into two regions, one region is treated with QM methods whereas the other is treated with MM methods. In these methods, the boundary between the QM region and the MM region is handled with special care. One such method of dealing with the QM/MM boundary is the use of link atoms.

3.3. Electronic Structure Methods

Electronic structure methods are routinely used in chemistry to predict molecular structure, physiochemical properties and chemical reactivity. Commonly used software packages for electronic structure calculations are Gaussian[14] and GAMESS-UK[15]. In this thesis, the Gaussian package was used for static quantum mechanical calculations.

The essence of the quantum mechanics is the so called wave function, ψ which represents wave like properties of the electrons. Any given chemical system, can be described by a unique wave function. When certain operators act upon ψ, the results are the observable properties of the system. The operator that acts on ψ to return the energy of the system is called the Hamiltonian operator (H). Therefore, it can be written as:

$$H\psi = E\psi \qquad [1]$$

Equation 1 is called the Schrödinger equation and it is the key equation of quantum mechanics. The wave function ψ can be written as a function of position and time. However when the variables are separated, it gives arise to the equation 1, which is the time independent Schrödinger equation. This can be done when it is assumed that the potential energy is independent of time. Solutions to equation 1 can provide the value of ψ describing the behaviour of the electrons in atoms and molecules as well as the corresponding energies. However, the Schrödinger equation can only be analytically solved for very simple systems such as hydrogenic atoms.

The Hamiltonian operator, H has two components; the kinetic energy term and the potential energy term. In mathematical notation the Hamiltonian operator is given as equation 2 , where the first two terms corresponds to the kinetic term and the last three terms potential energy term. In equation 2 I and j run over electrons and k and l run over nuclei, m_e and m_k represents the mass of the electron and mass of the nucleus k respectively, the electronic charge is given by e and the atomic number is given by Z, the distance between particles a and b is given by r_{ab} whereas $\hbar$ represents Planck's constant divided by 2π.

$$H = -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_k \frac{\hbar^2}{2m_k}\nabla_k^2 - \sum_i \sum_k \frac{e^2 Z_k}{r_{ik}} + \sum_{i<j} \frac{e^2}{r_{ij}} + \sum_{k<l} \frac{e^2 Z_k Z_l}{r_{kl}} \quad [2]$$

For any realistic system, only an approximated solution to the Schrödinger equation can be found. Various approximations made to the Schrödinger equation give arise to different electronic structure methods.

3.3.1. The Basis Sets

Many electronic structure methods assume that the molecular wave functions can be represented as the linear combination of atomic orbitals (LCAO). The atomic orbitals can be considered as the solutions to the Schrödinger equation for the hydrogen atom, which have an e^{-r} dependence where r is the distance from the nucleus. However, integration involving these functions is extremely difficult since analytical solutions to the Gaussian-type functions can be found and in them where is a $\boldsymbol{e^{-r^2}}$ dependence, integration task is made possible by replacing the above functions with the Gaussian-type functions.

A combination of several Gaussian functions can give a closer match to the slope of the real atomic functions. They are combined in such a manner that the resulting function resembles the original atomic function. In this case a single contracted basis function is composed of one or more primitive Gaussian functions. Equation 3 represents an *S-type* basis function. Alternatively one can use the Slater-type-orbitals (STO) instead of the Gaussian-type-orbitals (GTO).

[3]

Where, N is called the degree-of-contraction and it represents the number of primitive Gaussian functions in the basis function. The weighting of the Gaussian functions is given by d_i, the contraction coefficients and it determines the contributions of the Gaussians to the basis function.

The quantities $\alpha_{i\mu}$ define the width of the Gaussians and are called the exponents. The parameters f_μ are called the scale factors and they are often set to unity. The values for all the mentioned coefficients are determined by seeking the best fit to experimental determined atomic properties such as ionization energy.

Further improvements to the basis set can be done by introducing more functions to the basis set. One way of doing it is double the number of basis functions for each atomic orbital type. This gives arise to so called *Double Zeta* (DZ) basis sets. One can then continue this process and it will lead to *Multiple Zeta* basis sets.

If the "Double Zeta approach" is taken only for the valance orbitals, it results in so called *Split-Valance* basis sets. The basis sets are further improved by adding functions corresponding to orbitals with a higher angular momentum that those are occupied. These functions are referred to as polarization functions. The selection of the basis set is determined by the nature of the problem, accuracy required and the available computing power /time.

3.3.2. Ab Initio Methods

The Born-Oppenheimer approximation dramatically reduces the complexity of the Schrödinger equation by assuming that the motion of electrons in a molecule is much faster than that of the nuclei. This means the positions of nuclei may be assumed to be fixed. This allows separation of electronic and nuclear terms in the Schrödinger equation and enables to it to be solved for fixed positions of nuclei.

3.3.2.1. The Hartree-Fock (HF) Method

The simplest of the ab initio methods is the HF method[16,17] and it is the starting point of all other ab initio methods. The essence of the HF approximation is to replace the complicated many-electron problem by a one-electron problem in which electron-electron repulsion is treated in an average way.

The simplest antisymmetric (as required by the Pauli Exclusion Principle) wave function that describes the ground state of *N*-electron system is given by a single Slater determinant,

$$|\psi_0 >= |\chi_1, \chi_2 \ldots\ldots \chi_N > \quad [4]$$

where χ represents the spin orbitals. According to the variation principle, the best wave function in the above functional form is the one which gives the lowest possible energy, therefore,

$$E_0 =< \psi_0|H|\psi_0 > \quad [5]$$

In equation 5, H corresponds to the full electronic Hamiltonian. The choice of spin orbitals in equation 4 gives it the variational flexibility. The so called Hartree-Fock equation is derived by minimizing E_0 with respect to the spin orbitals. The HF equation which is given by equation 6, is also an eigenvalue equation and it determines the optimal spin orbitals.

$$f_{(i)}\chi_{(x_i)} = \varepsilon\chi_{(x_i)} \quad [6]$$

Where $f_{(i)}$ is an effective one-electron operator and is referred to as the *Fock* operator, which can be given as:

$$f_{(i)} = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + v_{(i)}^{HF} \quad [7]$$

Where $v_{(i)}^{HF}$ is the average potential felt by the i^{th} electron due to the presence of the other electrons. This "field" depends on the spin orbitals of the other electrons i.e. the Fock operator depends on its eigenfunctions and therefore equation 6 is non linear and must be solved iteratively. The procedure for solving HF equation is referred as the self-consistent-field (SCF) method.

Electron Correlation Methods

The HF method fails to account completely for the Coulombic interactions between the electrons which causes them to repel each other. This is because the HF method treats the electrons as moving under the influence of average effect of all other electrons; in other words HF lacks the so called electron correlation. Consequently the electrons are further apart than as given by the HF method. The most commonly used methods that can handle electron correlation are Møller–Plesset (MP) perturbation theory[18], Coupled Cluster (CC) theory[19] and the Density Functional Theory[20] (DFT). Methods used in this thesis are either DFT methods or methods derived from DFT , such as Self Consistent Charge Density Functional Tight Binding (SCC-DFTB)[21]. Therefore only these methods will be discussed in details.

3.3.3. Density Functional Theory(DFT)

In its basic notation, DFT states that the energy of an electronic system can be determined by its electron density. According to the first Hohenberg-Kohn (HK) theorem, the ground sate electron density, $\rho_{(r)}$ at each point r determines the ground state properties of the atomic and molecular systems. Further, it determines the external potential, $V_{(r)}$ due to the nuclei. The total number of electrons (N) are calculated via normalization of the electron density[9]. Therefore, it can be written as:

$$\int \rho_{(r)} dr = N \qquad [8]$$

The total number of electrons (N) and $V_{(r)}$ determines the Hamiltonian of the system. If the Born-Oppenheimer approximation is made, the corresponding Hamiltonian (H_{op}) can be given as:

– — — —— [9]

In equation 9 (when expressed in atomic units), rij, r_{iA} and R_{AB} represent the electron-electron, electron-nuclei and internuclear distances respectively. The summations over A and B run over nuclei, whereas the summations over *i* and *j* run over electrons. As the energy of the system is determined by $H_{op.}$, it is related to the wave function via Schrödinger's equation.

[10]

When ψ is known, all the ground state properties including the energy can be calculated. From the equations 8 though 10 it is clear that the energy of the system, *E* is a function of ρ and can be written as:

[11]

The second HK theorem provides a variational strategy to obtain ρ by minimizing *E.* It states that the energy does not change for the optimum $\rho(r)$, provided that all time $\rho(r)$ integrates to *N.* This can be given as:

[12]

The corresponding Lagrangian multiplier is denoted with μ in the equation 12 and it can be given as:

—— [13]

Equation 13 is known as the Euler-Lagrange equation and it is considered as the DFT analogue of the time independent Schrodinger equation. In equation 13 F_{HK} is a functional (Hohenberg-Kohn functional) that contains the electronic kinetic energy , $T_{(\rho)}$ and the electron-electron interaction functional, $V_{ee}(\rho)$. Therefore, F_{HK} can be given as:

[14]

Based on equation 13, the total energy of the system can be expressed as :

$$E_{v_{(\rho)}} = \int \rho_{(r)} v_{(r)} dr + F_{HK_{(\rho)}} \quad [15]$$

If one requires the system properties, solutions to the equation 13 have to be found. This practical aspect of dealing with equation 13 was first provided by Kohn and Sham (Kohn & Sham, 1965). The new form of equation 13 shows high analogy with the Hatree equations. The new Kohn-Sham (KS) theory introduces orbitals in such a way that the kinetic energy can be calculated with easily with high accuracy.

The KS theorem uses an *N*-electron, non-interacting system as the starting reference. The Hamiltonian of this reference is given as:

$$H_{ref} = \sum_i^N \frac{1}{2}\nabla_i^2 + \sum_i^N v_i(r) = \sum_i h_{ref,i} \quad [16]$$

Whereas $h_{ref,,i}$ is given as $h_{ref,i} = -\frac{1}{2}\nabla_i^2 + v_i(r)$ where it is considered as the one-electron operator, where as the newly introduced so called *KS*-orbitals (ψ_i in equation 17) serve as the eigenfunctions. Therefore, all physically acceptable densities of the non-interacting system are given by the following equation where the summation runs over the *N* lowest energy states of h_{ref}.

$$\rho = \sum_i^N |\psi_i|^2 \quad [17]$$

The functional, ($F_{HK}[\rho]$) given in equation 14 can be further expanded and re-written as

$$F_{HK_{(\rho)}} = T_{(\rho)} + J_{(\rho)} + E_{cx_{(\rho)}} \quad [18]$$

Where the electron interaction energy is expressed in two terms, the Coulombic interaction energy term, $J_{(\rho)}$ and the electron exchange correlation term, $E_{cx_{(\rho)}}$.

The kinetic energy (T_s) functional of the reference system can now be given in terms of KS orbitals,

$$T_S(\rho) = \sum_i^N \langle\psi_i| -\frac{1}{2}\nabla^2|\psi_i\rangle \quad [19]$$

Whereas the classical Coulombic interaction term is given by:

$$J_{(\rho)} = \frac{1}{2}\int\int \frac{\rho_{(r)}\rho_{(r')}}{r-r'} dr dr' \quad [20]$$

The remainder of the total energy, the difference between the total energy and the sum of kinetic and potential energies is assigned to the term $E_{xc}[\rho]$, the exchange correlation energy. This term contains a correction to the self interaction (double-counting error) in equation 19, the non-classical part of the $V_{ee}[\rho]$and the difference between exact kinetic energy and the kinetic energy of the reference system ($Ts[\rho]$).

An effective potential is then introduced as:

$$v_{eff}(r) = v(r) + \frac{\delta J}{\delta \rho} + \frac{\delta E_{xc}}{\delta \rho} = v(r) + \int \frac{\rho(r)}{|r-r'|} dr' + v_{xc}(r) \quad [21]$$

Where, the exchange correlation potential $\nu_{xc}(r)$ is defined as:

$$v_{xc} = \frac{\delta E_{xc}}{\delta \rho(r)} \quad [22]$$

Therefore, the Euler equation(13) can be re-written as :

$$\mu = v_{eff}(r) + \frac{\delta T_S}{\delta \rho} \quad [23]$$

When the new form of Euler (23) is coupled with the normalization conditions given in equation 8, the density can be given by:

$$\rho(r) = \int \sum_i^N |\psi_i(X)|^2 d\sigma \quad [24]$$

Where, X represents the four vector-containing space and spin variables. The integration is done over the spin variable σ. The equation 24 enables one to obtain ρ(r) as it made the right hand side of the equation 24 independent of r. This will lead to the evaluation of the effective potential, $v_{eff}[\rho]$.

Once the effective potential is known (or when the ground state density (ρ_0) is known to sufficient accuracy), the corresponding orbitals can be written as:

$$\left(-\frac{1}{2}\nabla^2 + v_{eff}(r)\right)\psi_i = \epsilon_i \psi_i \quad [25]$$

Equation 25 is the starting point for further approximations in DFT and it is known as the Kohn-Sham (KS) equation. The KS equations are to be solved iteratively, just as the Hartree-Fock equations.

Various approximations are made to equations 26 to accurately incorporate the electron exchange and the correlation. This gives arise to different exchange-correlation functionals such as the popular B3LYP functional.

3.3.3.1. The Density Functional Tight Binding (DFTB) Method

Although faster than *ab initio* methods, DFT has limitations dealing with large systems. The computational scaling of DFT limits the system size[22]. When using atom-based basis sets in Hartree-Fock or Kohn-Sham equations, calculations of the matrix elements and the diagonalization of the Hamiltonian matrix requires a major computational effort. Unless *N*-ordered diagonalization algorithms are used, this process becomes the bottle neck of such calculations. One way to make these calculations more efficient is to prevent the diagonalization of the Hamiltonian matrix. The tight-binding methods (TB) use this approach and they are very popular in solid-state physics in calculating band structures.

These TB models are expected to describe the properties of "tightly-bound" electrons of the system under focus, i.e. the electrons in the TB models must be tightly bound to the atoms to which they belong. Further, their interactions with the neighbouring atoms/potentials are very limited in this model. Under these circumstances, the corresponding wave functions of the electrons of interest resemble the atomic orbitals of the free atoms that they belong to. As a result, the LCAO treatment is applicable to these systems. The TB methods usually use the LCAO representation for the Hamiltonian matrix[23]. The basic starting point in TB is tightly-*bound* electrons, where interactions are ultimately treated perturbatively.

In TB methods, empirical parameters are used as the matrix elements. These parameters are generally provided in the form of look-up tables. However, having a single lookup table becomes very challenging task due to the difficulty of finding a set of system independent parameters that are freely transferable.

The DFTB method is an approximate KS-DFT method with an LCAO type treatment for KS orbitals (5). The DFTB method is an alternative to the quantum mechanical semi empirical methods Compared to the large number of empirical parameters found in SE methods and other TB methods, the DFTB parameters (the above mentioned matrix elements) are derived from full DFT calculations. Therefore DFTB can be considered as an approximate DFT scheme. As a consequence, it inherits both strengths and weaknesses from DFT. Furthermore, as an approximate method, it is not considered to have the accuracy of full DFT. Further, due to the *tight-binding* treatment, it has been argued that the perfect transferability of the parameters is never achieved[22].

Despite the weaknesses mentioned above, DFTB has found uses such as i). In calculations of large systems ii). Accessing longer time scales iii). General trend and structure search iv). In method development (as the methods developed for DFTB can easily be converted to DFT, DFTB provides very reasonable speed for testing) (v). Use as an electronic structure method in testing ,learning and teaching (also due to the speed)[22].

The effective potential given in equation 21 can be re-written as follows by expanding the external potential term $v_{(r)}$

[26]

The tight-binding scheme (assuming tightly bound electrons) i.e. the LCAO treatment of KS orbital ,$(\psi_{(r)})$ leads to the following following expansion with the minimal local basis.

[27]

In this 'minimal' approach, one radial function is used for each angular momentum state. i.e. one for *s*-states, three for *p*-states and five for *d*-states and so forth. These minimal functions are represented with η in equation 28 and their corresponding coefficients are given by *c*.

Then the corresponding Hamiltonian can be written as:

[28]

Substitutions of these new terms back in to equation 25 yields a generalised eigenvalue problem:

[29]

The equation 29 is usually solved with matrix (Hamiltonian Matrix) diagonalization methods. However, the DFTB method takes the advantage of the following overlap matrix (given for atoms μ and ν) (6) which are calculated from the atomic orbitals:

[30]

The corresponding DFTB Hamiltonian matrix element that is calculated from the minimal LCAO basis set η is given by:

[31]

The index zero, *0* in equation 31 indicates the reference density $\rho_{(0)}$. Once $H^0_{\mu\nu}$ and $S_{\mu\nu}$ are calculated with full DFT , they are tabulated and these are the parameters found in DFTB . The total energy in DFTB therefore reads,

– [32]

– [33]

The first term in equation 32 corresponds to the sum of the KS eigenvalues. This electronic energy is only a part of the DFTB total energy. The rest is given by the second term in equation 33, that is ; the DFT double counting contributions and the core-core repulsions. These contributions are approximated as repulsive and pair wise potentials and are given for atoms α and β. The scheme given by equation 32 and 33 is three orders of magnitude faster than full DFT with a minimum-sized basis set[24].

The DFTB (and DFT) scheme explained thus far is the Non-Self-Consistent (NSC) treatment of the KS orbitals. The NSC scheme works satisfactorily if the ground state electron density is closer to the reference density $\rho_{(0)}$. In the NSC-DFTB scheme the $\rho_{(0)}$ is chosen as the superposition of the atomic densities of the neutral atoms in the system. However, problems arise with this method when applied to the atoms with large electro negativities.

When atoms combine to make molecules, there is a flow of charge (i.e. charge equilibration) between atoms due to the differences in their electro negativities. Further, this will change the shapes of atom-like densities. Therefore, in such systems where atomic electro negativities are large, the real ground state densities deviate from the guessed reference densities. Therefore this scheme is expected to work well with the systems that are composed of atoms with compatible electro negativities[25]. However, the biological organic molecules (in the presence of oxygen) in particular show charge transfers and partial back-transfers due to equilibration of electro negativities[24]. To deal with such phenomenon, a self-consistent scheme is required.

3.3.3.1.1. The Self-Consistent Charge (SCC) Extension of DFTB- the SCC-DFTB Method.

Since the most important entity of interest is the total energy, the SCC-DFTB approach begins with functional expansion of the DFT total energy. This is done by a second order expansion of the DFT total energy functional with respect to the charge density fluctuations, $\delta\rho$ around a given reference density point. The charge density fluctuation describes the deviation of the ground state density from the reference density and it is given as:

$$\delta\rho=\rho-\rho_{(0)} \qquad [34]$$

As a result of the initial expansion, the SCC-DFTB energy reads as:

[35]

The first term in equation 35 is treated with the LCAO approach and is solved as explained above. After the LCAO treatment the first term can be written as:

[36]

The second term in equation 35 (the second order term) is approximated. In this approximation Δρ is given as superposition of atomic orbitals. Therefore, it is given as:

[37]

For further simplifications, a monopole approximation is applied to $\Delta\rho$. Therefore, for a given atom α, the term Δρ can be approximated as:

$$\Delta\rho \approx \Delta q_\alpha F_{00}^{\alpha} Y_{00} \qquad [38]$$

In equation 38, q represents the corresponding effective monopole, whereas F^{α}_{00} represents the normalised radial dependence of the density fluctuation on atom α, which is approximated (constrained) to be spherical (Y_{00}).

This means the angular deformation of the charge density in second order is neglected and therefore, only charge transfer effects are treated. i.e. this neglects the change in the shape of the charge density with reference to the reference charge density in the 2nd-order expression. However, the diagonalization of Hamiltonian matrix takes care of this by this shape change (of the neutral atomic input densities) in a non-self consistent manner. After these approximations, the second term in equation 35 can be given as:

$$\frac{1}{2}\sum_{\alpha\beta}\Delta q_\alpha \Delta q_\beta \int\int' \left(\frac{1}{|\vec{r}-\vec{r}'|} + \frac{\delta^2 E_{xc}}{\delta\rho\delta\rho'}\Big|_{n_0}\right) F_{00}^{\alpha} F_{00}^{\alpha} Y_{00}^{2} \qquad [39]$$

For large distances, where, $R_{\propto\beta} = \left|\vec{r} - \overrightarrow{r'}\right| = \infty$, the exchange (XC) term vanishes and the integral in equation 39 then describes the coulomb interaction of two spherical normalised charge densities which reduces to $1/R_{\propto\beta}$.

Therefore, this term can be re-written as:

$$E^{2nd} \approx \frac{1}{2}\sum_{\alpha\beta}\frac{\Delta q_\alpha \Delta q_\beta}{R_{\alpha\beta}} \qquad [40]$$

However, for diminishing interatomic distances, $R_{\propto\beta} = \left|\vec{r} - \overrightarrow{r'}\right| \to 0$, the integral term in equation 39 describes the electron-electron interaction of a single (α) atom. Therefore, integral is approximated as:

$$E^{2nd} \approx \frac{1}{2}\frac{\partial^2 E_\alpha}{\partial^2 q_\alpha}\Delta q_\alpha^2 = \frac{1}{2}U_\alpha \Delta q_\alpha^2 \qquad [41]$$

In equation 41, the term U_α describes the change in energy of the system with respect to the addition or removal of electrons and this is known as the chemical hardness or the Hubbard parameter.

In the SCC-DFTB approach this parameter is calculated by taking the 1st order derivative of the highest occupied molecular orbital (HOMO) with respect to the occupation number.

In the above discussed limits, $R_{\propto\beta} = \left|\vec{r} - \overrightarrow{r'}\right| = \infty$ and $R_{\propto\beta} = \left|\vec{r} - \overrightarrow{r'}\right| \to 0$, the regions of chemical bonding (1-3 Å) are not taken in to account. For this purpose, an interpolating function (to interpolate between the above limits) is introduced. After various approximations and manipulations[24] this interpolating function γ, for a given atom α can be written as follows in relation to its chemical hardness.

$$\gamma_{\alpha\alpha} = U_\alpha = \frac{\partial^2 E_\alpha}{\partial^2 q_\alpha} \qquad [42]$$

The last four terms in equation 35 depend only on the reference density $\rho_{(0)}$ and can be easily calculated. These four terms in equation 35 together represents the repulsive energy (E_{rep}) contributions and it can be given as:

$$E_{rep}[\rho_0] = \frac{1}{2}\sum_{\alpha\beta} U_{\alpha\beta} \qquad [43]$$

The repulsive terms are calculated (per atom basis) by performing bond stretching calculations with full DFT[26] .

Finally, combining the LCAO treated 1st term , approximated 2nd term and calculated last four terms (E_{rep}) equation 35 can be re-written as the SCC-DFTB total energy expression:

$$E^{SCC} = \sum_{i\mu\nu} c_\mu^i c_\nu^i H_{\mu\nu}^0 + \frac{1}{2}\sum_{\alpha\beta} \gamma_{\alpha\beta}\Delta q_\alpha \Delta q_\beta + \frac{1}{2}\sum_{\alpha\beta} U[\rho_0^\alpha, \rho_0^\beta] \qquad [44]$$

3.3.4. Semi-Empirical (SE) Methods

The number of two electron integrals that is required to constructing the Fock matrix makes Hartree-Fock (HF) calculations computationally expensive . The essence of semi-empirical methods is to reduce the number of these integrals and hence reduce the computational cost. The main step in reducing the integrals is to consider only the valance electrons explicitly. Various functions are introduced to model the behaviour of the core (both electrons and nuclei) of the atom. In addition to the absence of an explicit core, in SE calculations only a minimum basis set is used for valance electrons.

The assumption of *Zero Differential Overlap* (ZDO) is the key theme in all SE methods. In ZDO all products of basis functions that are located on different atoms are neglected. The ZDO approximation leads to the following consequences:

- The overlap matrix S is reduced to a unit matrix.
- One-electron integrals involving three centres are set to zero.
- All three- and four-centre two-electron integrals, which are by far the most numerous of the two-electron integrals, are neglected.

A compensation for the above approximations is done by replacing the remaining parameters in to experimentally derived parameters. The number of neglected integrals and different parameterization procedures give rise to different SE methods. Some of the most common SE methods are:

1. Neglect of Diatomic Differential Overlap Approximation (NDDO)
2. Intermediate Neglect of Differential Overlap Approximation (INDO)
3. Complete Neglect of Differential Overlap Approximation (CNDO)
4. Modified Intermediate Neglect of Differential Overlap (MINDO)
5. Modified NDDO models
6. Modified Neglect of Diatomic Overlap (MNDO)
7. Austin Model 1 (AM1)
8. Modified Neglect of Diatomic Overlap, Parametric Methods

3.4. Force Field (FF) Methods

3.4.1. Empirical Potential Energy Functions

As mentioned before, one of the main goals in any computational simulation technique is to mimic the true potential energy surface of a given system. In the previous section it was discussed that how this task is achieved in the quantum mechanical framework by attempting to solve the Schrödinger equation. However, these calculations are very intensive and generally are not applicable to large molecular systems such as enzymes. In FF methods, explicit treatment of electrons is avoided and the electronic energy is given as parametric function of the nuclear coordinates. A collection of such functions and their associated parameters is called a Force Field (FF). Further, the motion of nuclei is also neglected in the FF methods. Hence, atoms are treated as "balls" and are handled with Newtonian mechanics. Owing the implicit treatment of electrons, all kinds of "interaction-information" have to be provided explicitly. The bonding interactions are treated as classical "springs"; therefore, molecules in FF methods are treated as "ball and spring" models.

The total FF energy of a molecular system is given as the sum of bond-stretching energy, angle-bending energy, torsional rotational energy, van der Waals energy, electrostatic interaction energy and cross–interaction-energy. The fundamental FF energy terms are illustrated in Figure 1.

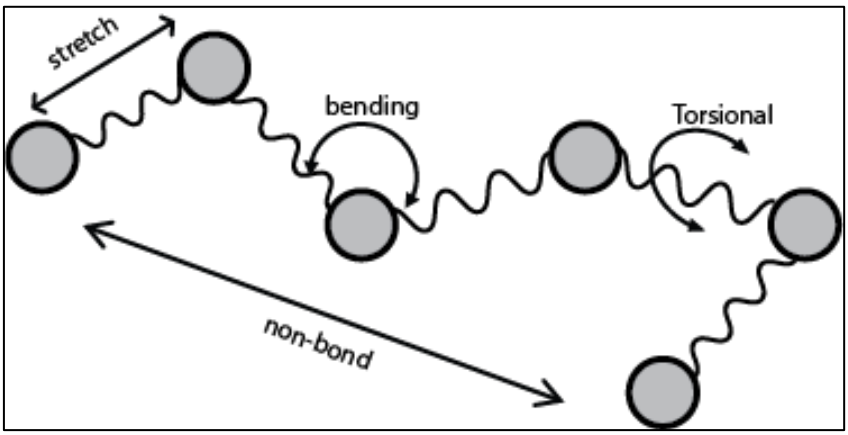


Figure 1: Schematic representation of Force Field Energy Terms

There are numerous force fields that are currently available for protein simulations. AMBER[27], CHARMM[28,29], GROMOS[30], and OPLS[31] are a few of them. In this thesis the CHARMM force field is used. The CHARMM energy function is given by equation 45.

$$V(\vec{R}) = \sum_{bonds} k_d (d - d_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} k_\chi [1 + \cos(n\chi - \delta)] + \sum_{impropers} k_\varphi (\varphi - \varphi_0)^2 + \sum_{Urey-Bradley} k_{UB} (S - S_0)^2 + \sum_{non-bonded} \{\varepsilon_{ij} [(R_{ij}^{\min} / r_{ij})^{12} - (R_{ij}^{\min} / r_{ij})^6] + q_i q_j / e_l r_{ij}\} \quad [45]$$

3.4.1.1. Bonding Terms

The equilibrium values of bond length, angle, dihedral angle, improper and Urey-Bradely (1-3) bond lengths are given by d_0 , θ_0 , χ_0 , $\varphi_0$and S_0 respectively. The terms without the subscript zero represent their values with respect to some other configuration. Likewise, k_d ,k_θ ,k_χ and k_φ are the corresponding force constants in the above terms. The bond and the angle terms are modelled using Hooke's law whereas the dihedral terms are described with a cosine function. A harmonic potential is used for improper terms. The Urey-Bradley terms are not used unless fitting of specific computational results to observable vibrational spectra is required. Bonded terms are applied to all atoms which are bonded through covalent bonds in a given system. These terms represent all bonded interactions in the system and contribute to the corresponding part of the overall potential function.

3.4.1.2. Non-bonding Terms

The non-bonded terms incorporate electrostatic and van der Waals interactions. The latter is modelled as a (12-6) Lennard-Jones interaction where in Equation 45, ε_{ij} relates to the Lennard-Jones well depth, R_{ij}^{min} is the distance at which the Lennard-Jones potential is minimum, q_i is the partial atomic charge of atom *i*, e_l is the effective dielectric constant, and r_{ij} is the distance between atoms *I* and *j*. The electrostatic terms are described as pair-wise columbic interactions.

3.4.1.3. Truncation of the Potential

As the number of atoms in a system increases, the number of non-bonded terms that need to be evaluated increases as the square of the number of atoms. Hence, evaluating all these interactions become computationally very expensive and practically nearly impossible. This problem can be dealt with by using a non-bonded cutoff. When a non-bonded cutoff is employed, the interaction energy between the atom pairs further apart than the cutoff distance is set to zero.

However, cutoffs introduce immediate discontinuity of both potential energy and the forces near the cutoff value. In order to overcome this problem, a "smoothing" technique is applied. Therefore a smooth approach of 'zero-value' is expected for the potential function. The most widely used smoothing-functions are switching and shifting functions[12].

3.5. Hybrid Methods – Quantum Mechanics/Molecular Mechanics Approach

The objective of the QM/MM hybrid-method (first introduced by Warshel et.al[32]) is to benefit both from QM and MM methods. Although MM methods can be applied to large molecular systems, these methods fail to account for any changes that occur at the electronic level. i.e. they cannot address the chemical reactivity. On the other hand, application of the QM methods to large molecular systems is prohibited by their size. Therefore, in the hybrid approach, the moieties of large molecular system in which the electronic level changes occur are treated with QM methods whereas the rest of the system is treated with inexpensive MM methods. The portioning of a large molecular system such as enzyme is illustrated in Figure 2.

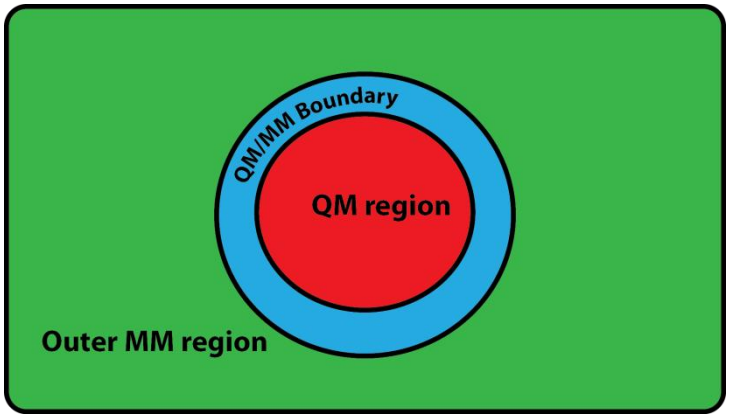


Figure 2: Partitioning of a molecular system in QM/MM calculations. The QM region is handled by expensive, accurate quantum mechanical methods, whereas the out MM region is handle by inexpensive , molecular mechanical methods. The boundary between QM region and the MM region is treated specially with an appropriate boundary scheme.

One of the main challenges in this approach is to effectively handle the boundary between the QM region and the MM region, specifically when there are covalent bonds between the two regions.

When handling the QM/MM boundary, the major problems that one has to face are: (i) The dangling bond of the QM atom must be capped. (ii) When calculating the electrostatic interactions, over polarization of the QM density, in particular, by the partial charge on MM atoms must be prevented. (iii) The bonded MM terms involving atoms from both subsystems have to be selected such that double-counting of interactions is avoided. Therefore, the boundary scheme should provide a balanced description of the QM–MM interaction at the QM/MM border. The currently available boundary schemes can be classified in to three main classes.

- **Link-atom schemes**[33]: in these schemes, an additional atomic centre (a link atom, usually a hydrogen atom) is introduced to saturate the dangling bond at the corresponding QM atom.

- **In boundary-atom schemes**[34,35]: These schemes introduce special atom types are usually referred to as a "Janus" boundary atom. This atom appears in both MM and QM calculations. In the MM region it behaves as a normal MM atom where as in the QM region it is use to cap the dangling valency.

- **Localised-orbital schemes**: these schemes use special hybrid orbitals to cap the dangling QM valency instead of introducing new atoms. In certain schemes, strictly localised bonding orbitals (SLBOs) are used, where as in certain schemes generalised hybrid orbitals (GHOs)[36] are used.

Each kind of boundary scheme has its strengths and weakness. In this thesis, straightforward and is widely used link atom method was used.

Based on how the energy expression (the overall system Hamiltonian) is formulated, the QM/MM methods can be divided in to two major schemes as: (1). Additive schemes, (2). Subtractive Schemes. The fundamental concepts of these two types are briefly presented below.

3.5.1. Subtractive QM/MM Schemes

If the entire system system is represented by , the qm region by ℚ, the MM region by and the link atom by, the additive-scheme QM/MM Hamiltonian $E_{QM/MM}$ is given as:

ℚ (ℚ [46]

In this scheme three calculations are performed. (i). An MM calculation of the entire system. (ii). A QM calculation of the QM region (and the link atom). (iii). An MM calculation of the inner QM region (and the link atom). The total energy of the system is then given by summing first two terms and subtracting the third, which is done as a correction to the double counting.

3.5.2. The Additive Scheme

With respect to the nomenclature introduced above, corresponding energy expression for the additive scheme is given as:

$$E_{QM/MM}(\mathbb{T}) = E_{MM}(\mathbb{O}) + E_{QM}(\mathbb{Q} + \mathbb{L}) - E_{QM/MM}(\mathbb{Q}, \mathbb{O}) \quad [47]$$

Whereas $\mathbb{O}$ corresponds to the outer region, i.e. the rest of the system excluding the QM region, which is handled by MM. In contrast to the subtraction scheme , here the MM calculations are only done for the outer region only.

Further, equation 47 has an additional $E_{QM/MM}(\mathbb{Q}, \mathbb{O})$ term. This term collectively handles the interaction energies between the two regions.

The QM/MM methods used in this thesis, CHARMM/SCC-DFTB[37] is also an additive scheme. It can be more explicitly as:

$$E^{tot} = \langle\psi|\hat{H}^{QM} + \hat{H}_{ele}^{QM-MM}|\psi\rangle + E_{van}^{QM-MM} + E^{MM} \quad [48]$$

Where the first term is corresponds to the electronic energy of the QM region which includes the electrostatic interactions with the MM part.

The second term obeys the Pauli Exclusion Principle and defines repulsions at short distances and dispersion effects at long distances. The last term corresponds to the FF energy of the MM atom.

The first component of the first term in equation 48 is solved according to the QM method used; the method used in this thesis (SCC-DFTB) was discussed earlier in this chapter. The second component, the electrostatic interactions between the QM atoms and MM partial charges are given as one electron integral. Where the corresponding Hamiltonian becomes,

$$\hat{H}_{ele}^{QM-MM} = \sum_{A\in MM}\sum_{B\in QM}\frac{Q_A Z_B}{|\vec{R}_A - \vec{R}_B|} - \sum_{A\in MM}\sum_{i=1}^{N_{el}}\frac{Q_A}{|\vec{R}_A - \vec{r_i}|} \quad [49]$$

In equation 49, Q_A is the partial charge of the atom A in MM region, Z_B represents the nuclear charge of atom B in the QM region and the N_{ele} is the number of electrons in the QM region. However, solving the above operator analytically will give arise to two– and three-centre integrals.

To avoid this problem, in SCC-DFTB, electrostatics are approximated by using the Coulombic interaction between the Mulliken charges of QM atoms and the partial charges of the MM atoms , i.e. higher level multipolar interactions between QM electrons and MM charges are neglected. Therefore:

$$\widehat{H}_{ele}^{QM-MM} = \sum_{A\in MM}\sum_{B\in QM}\frac{Q_A\Delta q^B}{|\vec{R}_A-\vec{R}_B|} \qquad [50]$$

Further improvements to this term are made by placing *1S* type Slater orbital on the MM atom and this procedure is discussed in details in the CHARMM/SCC-DFTB implementation[37].

The second term in equation 48 is described by Lennard–Jones potential as given in equation 45. The last term in equation 48 is a pure MM term and handled by the energy function given in equation 45. According to the way how the QM/MM electrostatics is handled, QM/MM schemes are classified in to three classes:

(i). Mechanical embedding schemes: where, the QM/MM electrostatics is dealt with a pure MM approach.

(ii). Electrostatic embedding schemes: where QM/MM electrostatics are handled by including the MM charges as a one electron term in the QM Hamiltonian .

(iii). Polarised embedding: where MM atoms too are allowed to be polarised by the QM charge distribution. There are various schemes under this category depending upon how the MM polarization is done.

3.6. The Bridge of Statistical Mechanics (SM)

The experimentally measured properties of a given system such as temperature are usually measured on macroscopic samples, whereas computer simulations are typically performed on relatively few numbers of particles. The observed macroscopic properties are a direct consequence of the behaviour of the microscopic particles. Statistical Mechanics describes the connection of the microscopic properties to the macroscopic properties[38-40].

The molecules that are in their energetic ground state at 0K, will be distributed among all possible energy states at a finite temperature. The relative probability (P_i) of finding a molecule i in energy state ε_i (a microstate) at a temperature T is proportional to a Boltzmann factor:

$$p_i \propto e^{-\varepsilon_i / k_B T} \quad [51]$$

Where k_B is the Boltzmann constant. If the energy of a given microstate i is given by ε_i, and the number of molecule in that state state is n_i, the conservation of the total energy (E) gives:

$$\sum_i n_i \varepsilon_i = E \quad [52]$$

If the total number of molecules is N, it can be given as:

$$\sum_i n_i = N \quad [53]$$

According to equations 52 and 53, the probability given in equation 51 can be rewritten as:

$$P_i = \frac{n_i}{N} = \frac{e^{-\beta \varepsilon_i}}{\sum_i e^{-\beta \varepsilon_i}} \quad [54]$$

Where $\beta = 1/k_B T$ and is defined as the thermodynamic temperature. This factor governs the most probable populations of states of a system at thermal equilibrium. If the proportionality constant for equation 51 is q, equation 54 becomes:

$$P_i = \frac{n_i}{N} = \frac{e^{-\beta \varepsilon_i}}{q} \quad [55]$$

And q is given as

$$q = \sum_{i=states}^{\infty} e^{-\beta\varepsilon_i} \quad [56]$$

q is defined as the partition function and it is the key concept in statistical mechanics. The partition function allows calculating macroscopic properties in SM i.e. it enables calculation of "bulk properties" from atomistic details. This normalization factor q is defined as the "partition function" because it describes how the probabilities are partitioned between available microstates based on their individual energies. Since it is given per molecule it is known as the *molecular partition function.*

When the degeneracy in energy is taken into account, equation 56 can be rewritten as follows by multiplying the sum of the exponential terms with a degeneracy factor $g_{i.}$

$$q = \sum_{i=levels}^{\infty} g_i e^{-\beta\varepsilon_i} \quad [57]$$

For a collection of N non-interacting particles, the corresponding partition function is defined as

$$Q = q^N(different, non-interacting particles) \quad [58]$$

$$Q = \frac{q^N}{N!}(identical, non-interacting particles) \quad [59]$$

If the particles are interacting, Q is calculated by taking the sum over all energy sates E_i for the whole system. Therefore, for interacting particles:

$$Q = \sum_{i}^{\infty} e^{-\beta E_i} \quad [60]$$

Since the energy states are very closely spaced, the quantum effects can often be neglected and the distribution of the states treated as continuous. Therefore, the discrete sum given in the equation 60 can be replaced with an integral over all phase-space (all coordinates (r) and momentum (p)); Only this is classical treatment of SM is discussed here. Therefore,

$$Q = \int e^{-\beta E_{(r,p)}}\, drdp \quad [61]$$

If the partition function of a system is known, all other thermodynamic properties can be calculated from it. The relationship between the internal energy U, and the partition function is given by:

$$U = kT^2 \left(\frac{\partial lnQ}{\partial T}\right)_V \quad [62]$$

From equation 62, the Helmholtz free energy A, can be calculated and once it is known, the pressure, P and the heat capacity, C can be calculated as derivatives of A and U. Consequently, other thermodynamic parameters such as enthalpy, entropy and Gibbs free energy can be calculated.

$$A = -kTlnQ \quad [63]$$

$$P = \left(\frac{\partial A}{\partial V}\right)_T = kT \quad \left(\frac{\partial lnQ}{\partial V}\right)_T \quad [64]$$

$$C = \left(\frac{\partial U}{\partial V}\right)_V = 2kT \quad \left(\frac{\partial lnQ}{\partial T}\right)_V + kT^2 \left(\frac{\partial ln^2 Q}{\partial T^2}\right)_V \quad [65]$$

$$H = U + PV = kT^2 \left(\frac{\partial ln^2 Q}{\partial T}\right)_V + kTV \left(\frac{\partial lnQ}{\partial V}\right)_T \quad [66]$$

$$S = \frac{U-A}{T} = kT \left(\frac{\partial lnQ}{\partial T}\right)_V + klnQ \quad [67]$$

In the condensed phase where the energy of the intermolecular interactions are larger than or comparable to the kinetic energy, separation of degrees of freedom is not possible.

Therefore, summing over all energy levels or integrating over all phase space is impossible. However, based on phase-space sampling it is possible to estimate the change in Q and derivatives such as $\partial lnQ/\partial T$.

When the fact that $\frac{\partial lnQ}{\partial T} = \frac{1}{Q}\,\partial Q/\partial T$ is taken in to account, equation 62 can be rewritten as:

$$U = \sum_i^{\infty} E_i (Q^{-i} e^{-\beta E_i}) \quad [68]$$

When continuous phase space formulation of the Boltzmann distribution is assumed as given by equation 61, the internal energy Ugiven in equation 68 can be written as:

$$U = \int E_{(r,p)} P_{(r,p)} dr dp \quad [69]$$

Equation 69 shows that U is a simple sum of energies weighted by the probability of being in that state i.e. that the average potential energy of the system.

Since high energy states have low probabilities and vice-versa, it can be assumed that the low energy states contribute more to the internal energy of the system. The corresponding Helmholtz free energy can therefore be given as:

[70]

In principle, U and A can be calculated by obtaining the probabilities of the states from MC or MD simulations. However, sampling entire phase space is problematic. Therefore, a representative collection of configurations can be generated the sum over all states is approximated by an average over a finite set of configurations. For M, a finite number of points, the average value of property X can be given as:

[71]

In computer simulations, an "ensemble" (a good representation of the "important" phase-space for a given property) is generated. As mentioned before, these ensembles, (collection of configurations) are calculated by using MC or MD methods. As given by in equation 72, when the property is given as an average value of the ensemble, it is referred to as the ensemble average.

Expressing system properties as ensemble averages is possible according to the *ergodic hypothesis*, which states that "the time averaged value of a property of a system (which is observed macroscopically) is equal to the ensemble averaged value (calculated by microscopic properties) of that property". Hence , the internal energy, and the Helmholtz free energy can be given as :

[72]

[73]

3.7. Simulating Dynamics

As discussed in the previous section, sampling of representative phase space is necessary to calculate macroscopic properties from atomistic level simulations. In this thesis Molecular Dynamics (MD) simulation method were used. In MD methods a series of time-correlated points in phase space is generated according to the second equation. These sequences of points are referred to as trajectories and it gives the time evolution of a system. The ensembles generated in MD simulations can be characterised by quantities such as Number of particles *(N)*, Volume *(V)*, Temperature *(T)* and Energy*(E)*. Ensembles are given labels based on the quantities that are kept constant, and this is summarised in Table 1.

N	***P***	***V***	***T***	***E***		**Acronym**	**Equilibrium**	**Name**
X		X	X			*NVT*	***A*** has minimum	Canonical
X		X		X		*NVE*	***S*** has maximum	Micro-canonical
X	X		X			*NPT*	***G*** has maximum	Isothermal-isobaric
		X	X		X	$VE\mu$	***(PV)*** has maximum	Grand Canonical

Table 1 :Constants in different ensembles and corresponding equilibrium state. *N*=number of particles; *P*= pressure; *V*= volume; *T*= temperature; *E*= energy; = chemical potential; *A*=Helmoholtz free energy; *S* = entropy ;*G* = Gibbs free energy.

In MD simulations , the heavy nuclei are treated as classical particles and dynamics are simulated by solving Newton's second equation, $F = ma$. The differential form of this equation is given by:

— — [74]

Where, U is the potential energy at position x ,m is the mass of the particle and the force F is given as the negative gradient of the potential energy. Therefore, for a given particle equation 74 can be rewritten as:

[75]

Equation 74 relates the derivative of the potential energy to the changes in position as a function of time. One can calculate a trajectory for a given system by solving equation 74, where trajectory is defined as a collection of configurations of the system as it evolves with time. Calculation of a trajectory requires the initial positions and the initial distribution of velocities for all atoms in the system. The following set of equations show how a trajectory is calculated in a MD simulation. In Equations 76 – 80 $a$is the acceleration, v is the velocity, t is the time and x is the position and the subscript 0 indicates the starting values.

$$a = \frac{dv}{dt} \qquad [76]$$

Integration of equation 76,

$$v = at + v_0 \qquad [77]$$

And

$$v = \frac{dx}{dt} \qquad [78]$$

Integration of equation 78 gives:

$$x = v.t + x_0 \qquad [79]$$

From equation 77 and 79,

$$x = a.t^2 + v_0.t + x_0 \qquad [80]$$

Equation 80 gives the value of x at time t as a function of the acceleration, a, the initial position, x_0 , and the initial velocity, v_0. Equation 74 gives the potential energy of the system as a function of time and position. If the solutions to equations 74 and 80 can be obtained for all the atoms in the system at all times, then all the positions (hence the overall configuration of the system) and the potential energy at all times can be obtained. When those two entities are known any thermodynamic property of the system can be calculated.

The equations of motion are deterministic, e.g., the positions and the velocities at time zero determine the positions and velocities at all other times, *t*. The initial positions can be obtained from experimental structures, such as the x-ray crystal structure of the protein or the solution structure determined by NMR spectroscopy.

The initial distribution of velocities is usually determined from a random distribution with the magnitudes conforming to the required temperature and corrected so that there is no overall momentum

Once the initial conditions are decided, Newton's equations must be integrated in order to generate the results. This can be done using one of many available "integrators" i.e. integrating algorithms. Many integration algorithms are available to achieve the above target of integrating the Newton's equations. Among many available, the following are the most popular integration algorithms used in MD: Verlet algorithm[41], leap-frog algorithm[42] velocity Verlet[43] and Beeman's algorithm[44]. In this thesis the leap-frog (LF) integrator was used.

In the LF algorithm the velocities are evaluated at the midpoint of the position evaluation and vice-versa.

$$r_{n+1} = r_n + v_{n+1/2}\Delta t \quad [81]$$

$$v = v_n + v_{n-1/2} + \frac{F_n}{m}\Delta t \quad [82]$$

Where r is the position, v is the velocity ,F is the force, t is the time and subscripts $n \pm 1/2$ represent the corresponding values at the mid-step time. The LF algorithm therefore, has three major steps:

1. Use the current position $r_n$to calculate the current force F_n.
2. Use the current force F_n and previous mid-step velocity $v_{n-1/2}$ to calculate the next mid-step velocity$v_{n+1/2}$.
3. Use the current position r_n and the next mid-step velocity $v_{n+1/2}$ (from step 2) to calculate the position in the next step, r_{n+1}

3.8. Boundary conditions

The number of particles that can be included in a computer simulation is very limited (usually < 10,000). This imposes limitations on the number of solvent atoms that can be incorporated in a simulation. Therefore a large fraction of molecules lie on the surface of any small sample. The molecules on the surface will experience quite different forces from the molecules in the bulk in a simulation. The surface effects are overcome by implementing boundary conditions, i.e. these conditions allow one to perform a simulation with relatively smaller number of particles in such a way that particles experience forces as if they were in a bulk fluid. There are two major categories of boundary conditions, periodic boundary conditions and non-periodic boundary conditions. In this thesis the stochastic boundary conditions, which falls under the latter category were used and it is discussed below.

3.8.1. Stochastic Boundary Conditions[45]

This approach is particularly useful in cases where only a portion of a large molecular system is needed to be simulated. A classic example for such a situation is the simulation of an isolated enzyme binding site. Here, no resources are used to simulate the whole macro molecule.[46] This goal can be achieved by partitioning the molecular system in to different volumes as illustrated by Figure 3.

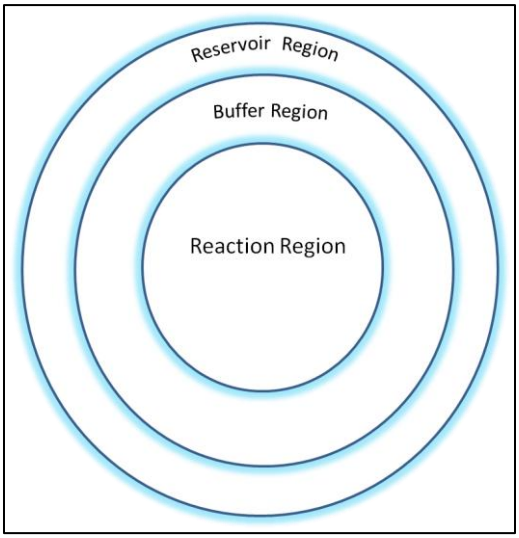


Figure3: Partitioning of the simulation system in Stochastic boundary conditions

The innermost region is called the reaction region i.e. the main region of interest and within this region normal molecular dynamics (discussed later) is performed. This region is surrounded by so-called buffer region. The outermost region in which all the atoms are fixed is known as the reservoir region and this serves as the boundary that preserves the global density and overall structure. The buffer region allows a smooth transition from the reaction regions to the boundary by means of friction a force gradient. Further, the buffer region accommodates the fluctuations in density, structure and energy that occurs in the reaction region. Within the buffer region, atoms move according to stochastic dynamics. In this thesis the Langevin equation[47] is used to generate stochastic dynamics.

3.9. Simulating Chemical Reactions

As discussed before, free energy is a central quantity in statistical mechanics from which other thermodynamic quantities can be calculated. According to the second law of thermodynamics, reaction mechanisms can be discovered from the minimum energy pathways found on reaction free energy surfaces. However, both conventional MD and Monte Carlo (MC) simulations cannot produce accurate free energies directly as they suffer from an intrinsic difficulty of calculating the system's entropy[48].

Therefore, regions that are high in energy than the available thermal energy (~3kT) are not sampled in conventional MD simulations, i.e. only a Boltzmann sampling of the phase-space is done. However, when simulating chemical reactions, high energy regions of the phase-space that are marked by rear events such as Transition States must also be sampled adequately to obtain an accurate free energy. Therefore, non-Boltzmann sampling of the phase-space is required and it can only be achieved by free energy simulations.

The most apparent way of performing non-Boltzmann sampling is by using a so-called biasing potential,$V_{(r)}$ where, by using $V_{(r)}$ sampling is biased towards the desired volume of phase space. If the potential energy of the system is $U_{(r)}$, the augmented potential $U'_{(r)}$ can be given as :

$$U'_{(r)} = V_{(r)} + U_{(r)} \qquad [83]$$

where r is the configurational space and $\beta = 1/k_B T$. The augmented potential drives the sampling away from the low energy region to obtain sufficient sampling in high energy regions. This biased-free-energy (BFE) technique gives arise to various sampling methods. The interest in the BFE methods is to find the free energy as a function of the reaction coordinate (ξ) which is a function of only the configurational coordinates (phase-space positions) q. Consequently, reaction coordinates are considered to be additional properties of the system whose extended density Ω_ξ of states can be given as:

$$\Omega_\xi(N,V,E,\xi) = \frac{1}{h^{3n}N!}\iint_{V^N} \delta[H(q)-E]\left(\prod_i \delta[\hat{\xi}_i(q)-\xi_i]\right)\mathrm{d}q \qquad [84]$$

The density of states can be related to the canonical partition function via:

$$Q(N,V,T) = \int \exp(-\beta E)\,\Omega_{tot}(N,V,E)dE \qquad [85]$$

Where the total density of states Ω_{tot} is given by

$$\Omega_{tot}(N,V,E) = \frac{1}{h^{3n}N!}\iint_{V^N} \delta[H(q,p)-E]\mathrm{d}q\,\mathrm{d}p \qquad [86]$$

Where, momentum p and position q variables are integrated over the entire system volume V^N . In equation 84 and 86 E represents the total interaction potential energy whereas H represents the system Hamiltonian with respect to the given variables. Since the Helmholtz free energy is related to the canonical partition function (equation 63), it is possible to derive free energy from the density of states using equations 85 and 63. That is, if the accurate density of states is available, all thermodynamic quantities of the system can be calculated. In calculating the accurate density of state, non-Boltzmann sampling is required to ensure adequate sampling of the entire phase-space. As mentioned before, the BFE methods can be used to achieve this.

Adaptive Umbrella Sampling Approach[49] (AUSA) is one such widely used BFE technique where the reaction coordinate is restrained using a biasing potential. In its simplest form of AUSA biasing takes form of harmonic potential centred on a region of reaction coordinate space.

Flat Histogram Approach is another sampling technique. In this approach distribution is such that all the energy levels are visited equally often i.e., when the visits to each energy level are counted, a flat visit histogram is obtained. It has been shown that calculation of free energies with this approach is highly effective in both MC[50,51] and MD[52] simulations where the latter is best known as metadynamics. The Free Energy from Adaptive Reaction Coordinate Forces (FEARCF)[53-55] used in this thesis is another flat histogram method and it is discussed below.

3.9.1. Free Energy from Adaptive Reaction Coordinate Forces (FEARCF)

FEARCF method generates forces from the probability distributions and the histograms of the reaction coordinate surface, and they are used to produce equally sampled reaction coordinate space. Unlike in usual umbrella sampling methods, this technique requires no intervention from the user other than a judicious choice of reaction coordinate and simulation length at each update of the biasing force.

In FEARCF, the reaction surface is divided into an n-dimensional grid and sampling frequency for a bin site is recorded for each simulation. A running tally of the probability density of the reaction coordinate is given by the population of this grid and it is derived from the history of the simulation to the current point. This population density is used as the input for a cubic-spline interpolation routine to generate the biasing forces for the next iteration. This biased force (reaction driving force), $V_{(\xi)}$ is then applied to the all atoms used in the reaction coordinate definition to drive the sampling away from the previously sampled regions. When these biasing forces are derived from the potential of the mean force (PMF), it is expected to sample the entire reaction coordinate space.

In this thesis two independent reactions coordinates were used. They represent the length of the breaking C1'-N9 bond and the forming C1'-O bond in the depuration reaction catalysed by ricin. These reaction coordinates can be written as:

[87]

The Hamiltonian of the reacting system can be written as:

[88]

Where is the unbiased system Hamiltonian and is the multidimensional reaction driving potential which is calculated from the sampled probability distribution

The reaction driving potential can be given as:

ln P [89]

In the simulations, the biased Hamiltonian is used instead of the unbiased of and so it generates a biased probability distribution of sampled coordinates ′ . When account for the driving potentials, this ′ can be converted to the unbiased probability distribution as below:

′ — [90]

where C is the normalisation constant , k_B is the Boltzmann constant and T is the system temperature. For a calculated unbiased probability density the corresponding PMF can be written as:

[91]

where $W(\xi)$ refers to the PMF parameterised by the above defined reaction coordinates. Once all the regions in the coordinate space have been adequately sampled, a flat histogram of probability distributions is obtained and the free energy of the system is found.

The PMF in equation 91 is calculated by using the probability distribution from the sampling of all trajectories up to the current simulation point. The best biasing potential is defined as =- where it allows sampling the entire coordinate space. The corresponding biasing force for is therefore given as:

——— [92]

A positive force increases the distance between two selected points *P1* and *P2* whereas a negative force means a decrease in the distance between them. The force given in equation 92 results in accelerations and on each molecule. Where, and are the total mass of the molecules at *P1* and *P2* which is separated by a distance represented with vector . The reaction coordinate is given as The accelerations and are then distributed equally such that the atom in the molecule at *P1* experiences a force:

—— [93]

Where is the mass of atom k.

The Cartesian force arising from the potential with respect to the reaction coordinates can be given as:

—————— ; —————— [94]

where is the force on atom in molecule 1 and is the force on atom in molecule 2.

These new forces are then applied to the atoms in the system to drive them away from the previously sampled regions to the un-sampled regions of the coordinate space. After each iteration of dynamics, the forces are recalculated and reapplied until converged, i.e until a flat histogram is obtained.

As in any other reaction simulating technique, FEARCF often involves at least two reaction coordinates and multiple simulations are usually carried out with different biasing potentials to enhance the phase space sampling. Therefore, a way to optimally combine the data from multiple simulations is required. The weighted histogram analysis method (WHAM)[56] is often used for this purpose and it is discussed below.

3.9.2 Weighted Histogram Analysis (WHAM) Method in Free Energy Calculations

As mentioned before, WHAM is used to combine data from multiple simulations. When multiple simulations for the same reaction coordinate ξ are done, each with different biasing potentials , at simulation , the unbiased probability distribution can be reconstructed from the biased one, ′ . Else, distributions can be represented as normalised histograms generated from the biased sampling data. The unbiased probability of window can be given as:

[95]

Where, is the biased probability (from window), and is the free energy from adding the biasing potential which is defined as:

[96]

In order to use all (multiple) samples in to account a linear combination of unbiased probabilities is used. This can be given as:

′

[97]

Where $ci_{(\xi)}$are the weights. Once the weights are known , equation 87 can be used to obtain the unbiased probabilities in each window. Initially , the weights are normalised.

That is:

$$\sum_{i=1}^{N} ci_{(\xi)} = 1 \quad [98]$$

The weights are however, fully defined by minimizing the variance in $p_{(\xi)}$ with respect the weights. That is,

$$\frac{\partial \sigma^2[p_{(\xi)}]}{\partial c_i} = 0 \; for \; i = 1 \ldots . N \quad [99]$$

The resulting weights can then be written as:

$$ci_{(\xi)} = \frac{n_i e^{-\beta(U_{B_{(i)}}(\xi) - f_i)}}{\sum_{j=}^{N} n_j e^{-\beta(U_{B_{(i)}}(\xi) - f_i)}} \quad [100]$$

Where n_i is the number of samples in the i^{th} window. Since f_i depends on $p_{(\xi)}$, this interdependence requires that the nonlinear system i.e. equations 91 and 92 to be solved until convergence.

$$p_{(0)}(\xi) = \sum_{i=1}^{n} \frac{N_i}{\sum_{j=1}^{n} N_j \exp\left[-\beta(U_{Bj} - f_i)\right]} p'_{(\xi)} \quad [101]$$

$$e^{-\beta f_i} = \int e^{-\beta U_{B_{(i)}}(\xi)} p_{(\xi)} d\xi \quad [102]$$

A zero change in obtained free energy $A_{(\xi)}$ means uniform probability and it marks convergence.

$$A_{(\xi)} = -k_B T ln p_{(\xi)} = 0 \rightarrow p_{(\xi)} = 1 \quad [103]$$

3.10. Solvation

Solvation is a key phenomenon in any biologically relevant chemical systems, as most of those systems are partially or fully solvated at the physiological conditions. Therefore, it is crucial to simulate the solvation accurately in computer simulations. There are two main models to incorporate the solvents in a simulation; these are are implicit models and explicit models.

3.10.1 Implicit Solvation

A large amount of computer simulation time is taken to evaluate the solvent-solvent effects. However, the solvent effects cannot be totally ignored. The implicit solvent models are originally designed to address the above issues and there are numerous implicit water models have been developed[57-59]. These models are called continuum models as they represent the solvent as a continuum medium instead of individual solvent molecules. Although these models have proved their ability to produce better results in certain types of calculations [60] (eg. solvation free energy calculations), due to the fact that they facilitate undesirable conformational changes in proteins, these methods are not generally used in protein simulations.

3.10.2. Explicit Solvation

The solvent is represented as individual or explicit molecules in this model in contrast to implicit models. Simulation of a system with only a thin layer of water / solvent around it can overcome the majority of the problems of pure implicit models. Currently, there is a wide range of explicit water models available and commonly parameters of these water models are adjusted to reproduce the enthalpy of vaporization and density of water. Some of the more common and popular explicit water models are TIP3P[61] TIP4P[61], TIP5P[62], and SPC/E[63]. These models can be classified based on the number of points used to define the model (atoms plus dummy sites); rigidity / flexibility, and polarisation effects.

The TIP3P water model is used in this thesis owing to its proven ability to mimic solvent effects in protein simulation and, more importantly, CHARMM's inherited usage of this model. All CHARMM force fields (eg. proteins, nucleic acids, lipids) have been parameterised with respect to TIP3P. It is the simplest model of explicit water and has three sites of interactions corresponding to the three atoms of the water molecule. The partial positive charges on the hydrogens are balanced by an appropriate negative charge on the oxygen atom and the van der Waals interactions between two water molecules are calculated using a Lennard-Jones function with a single point of interaction per molecule which is centred on the oxygen atom. There is no interaction calculated between the hydrogen atoms. TIP3P has a rigid geometry and the original model has been modified by placing Lennard-Jones parameters on hydrogen in CHARMM.

3.11. References

(1) Haran, G. *Journal of Physics-Condensed Matter* **2003**, *15*, R1291.
(2) Baryshnikova, E. N. *Protein Science* **2005**, *14*, 2658.

(3) Wang, T. *Biophysical Journal* **2005**, *89*, 1.

(4) Dyson , H. J. *NUCLEAR MAGNETIC RESONANCE OF BIOLOGICAL MACROMOLECULES Part C* **2005**, *394*, 299.

(5) Chen, X. Y.; Link, T. M.; Schramm, V. L. *Biochemistry* **1998**, *37*, 11605.

(6) Schramm, V. L. *The Journal of biological chemistry* **2007**, *282*, 28297.

(7) Reiher, M.; Bertini, L. U. h. b. g. c. z. b. i. t. D. x. *Atomistic approaches in modern biology: from quantum chemistry to molecular simulations*; Springer, 2007.

(8) Allen, M. P. *Computer Simulation of Liquilds*; Oxford University Press, Oxford, 1987.

(9) Jensen, F. *Introduction to computational chemistry*; John Wiley & Sons, 2007.

(10) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341.

(11) McCammon, J. A.; Harvey, S. C. U. h. b. g. c. z. b. i. O. O. Q. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press, 1988.

(12) *Computational Biochemistry and Biophysics*; Becker, O. M.; MacKerell, A. D.; Roux, B.; Watanabe, M., Eds.; CRC Press, 2001.

(13) Henzler-Wildman, K.; Kern, D. *Nature* **2007**, *450*, 964.

(14) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision A.1*; Gaussian, Inc.: Wallingford CT, 2009.

(15) M.F. Guest, I. J. B., H.J.J. van Dam, P. Sherwood, J.M.H. Thomas, J.H. van Lenthe, R.W.A Havenith, J. Kendrick, *Molecular Physics* **2005**, *103*, 719.

(16) Hartree, D. R. *The Calculation of Atomic Structures*; Wiley:New York, 1957.

(17) Fock, C. *Zeitschrift für Physik* **1930**, *61*.

(18) Møller, C.; Plesset, M. S. *Physical Review* **1934**, *46*, 618.

(19) Čížek, J. *The Journal of Chemical Physics* **1966**, *45*, 4256.

(20) Pierre, H.; Kohn, W. *Physical Review* **1964**, *136*.

(21) Kohn, W.; Sham, L. J. *Physical Review* **1965**, *140*, A1133.

(22) Koskinen, P. *Computational Materials Science* **2009**, 1.

(23) Seifert, G. *The journal of physical chemistry. A* **2007**, *111*, 5609.

(24) Elstner, M. *Journal of Physical Chemistry A* **2007**, *111*, 5614.

(25) Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. In *Physical review. B, Condensed matter* 1995; Vol. 51, p 12947.

(26) Elstner, M. *Carbon* **2006**.

(27) Weiner, P. K.; Kollman, P. A. *Journal of Computational Chemistry* **1981**, *2*, 287.

(28) Foloppe, N.; MacKerell, J. A. D. *Journal of Computational Chemistry* **2000**, *21*, 86.

(29) Mackerell, A. D.; Banavali, N. K. *Journal of Computational Chemistry* **2000**, *21*, 105.

(30) van Gunsteren, W. F.; Daura, X.; Mark, A. E. In *Encyclopedia of Computational Chemistry*; John Wiley & Sons, Ltd: 2002.

(31) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *Journal of the American Chemical Society* **1996**, *118*, 11225.

(32) Warshel, A.; Levitt, M. *Journal of Molecular Biology* **1976**, *103*, 227.

(33) Field, M. J.; Bash, P. A.; Karplus, M. *Journal of Computational Chemistry* **1990**, *11*, 700.

(34) Senn, H.; Thiel, W.; Reiher, M., Ed.; Springer Berlin / Heidelberg: 2007; Vol. 268, p 173.

(35) Antes, I.; Thiel, W. *Journal of Physical Chemistry A* **1999**, *103*, 9290.

(36) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *The Journal of Physical Chemistry A* **1998**, *102*, 4714.

(37) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *The Journal of Physical Chemistry B* **2000**, *105*, 569.

(38) Helrich, C. S. U. *Modern thermodynamics with statistical mechanics*; Springer, 2009.

(39) McQuarrie, D. A. *Statistical mechanics*; University Science Books, 2000.

(40) Chandler, D. *Introduction to modern statistical mechanics*; Oxford University Press, 1987.

(41) Verlet, L. *Physical Review* **1967**, *159*, 98.

(42) Hockney , R. W. *Methods in Computational Physics* **1970**, *9*, 136.

(43) Swope, W. C. *Journal of Chemical Physics* **1982**, *76*, 637.

(44) Beeman, D. *Journal of Computational Physics* **1976**, *20*, 130.

(45) Brooks, C. L.; Brünger, A.; Karplus, M. *Biopolymers* **1985**, *24*, 843.

(46) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press, 1988.

(47) Langevin, P. *Comptes Rendus de l'Académie des Sciences* **1908**, *146*, 530.

(48) Beveridge, D. L.; DiCapua, F. M. *Annual Review of Biophysics and Biophysical Chemistry* **1989**, *18*, 431.

(49) Mezei, M. *Journal of Computational Physics* **1987**, *68*, 237.

(50) Wang , F., G; Landau, D., P *Physical Review E* **2001**, *64*.

(51) Wang , F., G; Landau, D., P *Physical Review Letters* **2001**, *86*, 2050.

(52) Laio , A.; Parrinello , M. *Proceedings of the National Academy of Sciences* **2002**, *99*, 12652.

(53) Strümpfer, J.; Naidoo, K. J. *Journal of Computational Chemistry* **2010**, *31*, 308.

(54) Naidoo, K. *SCIENCE CHINA Chemistry* **2011**, *54*, 1962.

(55) Barnett, C. B.; Naidoo, K. J. *Molecular Physics* **2009**, *107*, 1243.

(56) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *Journal of Computational Chemistry* **1992**, *13*, 1011.

(57) Cramer, C. J.; Truhlar, D. G. *Chemical Reviews* **1999**, *99*, 2161.

(58) Simonson, T. *Current Opinion in Structural Biology* **2001**, *11*, 243.

(59) Feig, M. *Current Opinion in Structural Biology* **2004**, *14*, 217.

(60) Lazardis, C.; Karplus, M. *Biophysical Chemistry* **2003**, *100*, 367.

(61) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926.

(62) Mahoney, M. W.; Jorgensen, W. L. *Journal of Chemical Physics* **2000**, *112*, 2000.

(63) Berendsen, H. J.; Grigera, J. R. *Journal of Physical Chemistry* **1987**, *91*, 6269.

CHAPTER -4

CHOOSING A SUBSTRATE MODEL FOR RICIN

4.1. Introduction

Ricin is a member of the Ribosome Inactivating Protein (RIP) family (discussed in chapter 1) that targets the specific adenine, A-4324 in an RNA internal loop. Removal of this adenine ceases the protein synthesis[1-7] and causes cell death. In general, internal loops are very common features of RNA secondary structure[8]. Moreover, they often form binding sites for metals, proteins, and other RNAs.

The target substrate of ricin which is found in eukaryotic 28S ribosome (rRNA) is highly conserved. Further, in eukaryotes, it forms a double-helical stem and 17-base loop[9]. These stem-loop structures are formed due to the intra molecular base pairing in a single RNA strand. These local structures are often referred as hairpins or hairpin loops. The loop section of the hairpin targeted by ricin has the UCAGUAC**GAGA**GGAACC nucleic acid sequence[4]. The specific adenine, A-4324 which is targeted by ricin is found in the centre of the G**A**GA tetraloop highlighted in the above sequence. it has been reported that these tetra loop containing hairpins are extremely common in biologically active RNAs[10] and they are often referred to as GNRA loops; where, G refers to guanine, N refers to any nucleic acid base and R refers to G (guanine) or A (adenine)[10] and final A refers to adenine. Thermodynamic studies have shown that frequently occurring RNA tetra loops are more stable than other four-nucleotide loops with the same stem[11,12]. Further, it is assumed that these stable tetra loop structures provide nucleation sites to ensure desired folding of secondary and tertiary RNA structures[13].

In 28S eukaryotic ribosomal RNA, the GAGA tetra loop is a part of the elongation factor binding site. Binding of the elongation factor to the ribosome is essential to ensure the mobility of ribosome along the m-RNA in protein synthesis. When ricin destroys the elongation factor binding site by hydrolysing A-4324, ribosome becomes immobile and protein synthesis comes to a halt. The structure of the 28S rRNA stem-loop hairpin has been experimentally determined by several authors[10,14]. The 3D-structure of RNA loop structure targeted by ricin is shown in Figure 1.

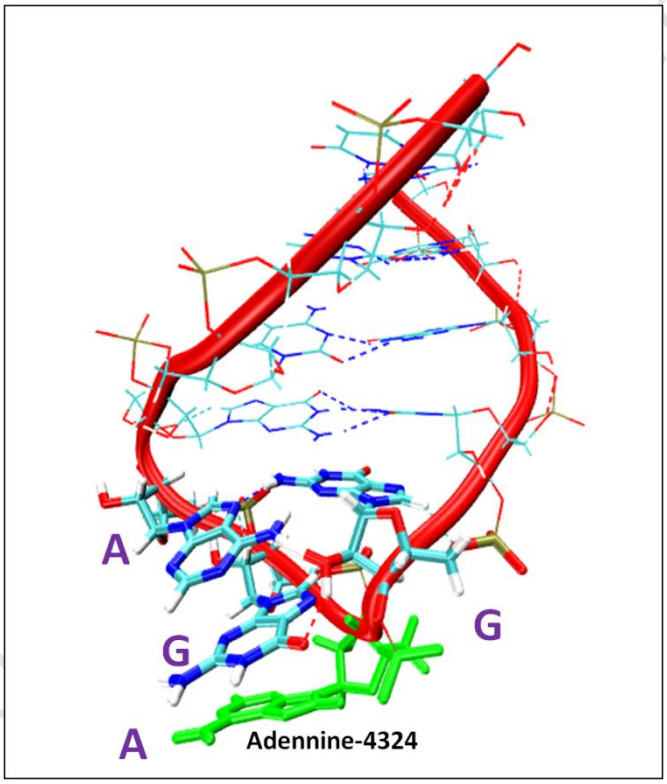


Figure 1: The Sarcin-ricin stem loop 12-mer. The GAGA tetra loop is presented in the liquorish and the targeted adenine is highlight in green. Negatively charged RNA backbone is presented in red. Intra-molecular hydrogen bonds that found between paired bases are represented with dashed lines.

4.2. Models in RNA Research

Complicated structural features of ribosomal RNA (rRNA) impose numerous challenges on studying them experimentally as well as computationally. It has been shown that experimental techniques such as selective chemical modification, fluorescence labelling and mutations are cumbersome for the whole ribosome but readily applicable to model RNAs.[15] In computer simulations, these models could play far more important roles than in experiments. This is mainly due to the computational limitations in simulating complete macro molecules such as RNA.

If one is to simulate a RNA-Protein complex, the challenge becomes more complex due to the combined structural and functional complexity arising from two macromolecules. The dynamics of such molecular associations is influenced by various kind of local motions[16] found in those macromolecules. Moreover, it makes challenges even harder as there are major differences between the localised motions of RNA and that of proteins. These differences are due to their fundamentally different local structures; where extensive stacking of aromatic rings is found in RNA, and α-helices, β-sheets and loops are mainly found in proteins.

In RNA-Protein complexes, two kinds of differences are anticipated in large scale motions that extend over few nanometers[16]. The first difference is expected between the RNA-water interactions and Protein-water interactions. This is mainly due the characteristic long-range electrostatics of RNA that is caused by its charged backbone. The second difference is due to the extended nature of the structure of RNA. Owing the above complexity in RNA-protein complexes, simulation of the chemical reactions of such complexes becomes extremely challenging. In general, these challenges are twofold; (i) selection of a suitable simulation method/(s), (ii). Selection of accurate models for simulations.

The issues related to simulations methods were discussed in chapter 3. This chapter focuses on addressing the issues related to the simulation-models. Specifically, this chapter discusses the suitability of two substrate models for the simulation of ricin-catalysed depurination reaction.

4.3. Substrate Models for Ricin

Simulating chemical reactions requires using of quantum mechanics. Therefore, the size of the "reactive-system" is of prime important in such simulations. Since ricin targets a single glycosidic bond found in an adenine, the corresponding adenosine nucleoside, shown in Figure 2a, can be considered as the most obvious substrate model for reaction simulations. Moreover, it is the minimum required substrate model for any simulation that involves the target glycosidic bond. Hereafter, this simplest substrate model is referred to as the minimum substrate model (MSM). Alternatively, one can choose a portion of the hairpin loop that is composed of multiple nucleobases as the substrate model. In such a model, hereafter referred to as the Loop Substrate Model (LSM), the number of surrounding bases to be included in the model is an important problem to be solved. Figure 2 schematically presents the idea of MSM and LSM.

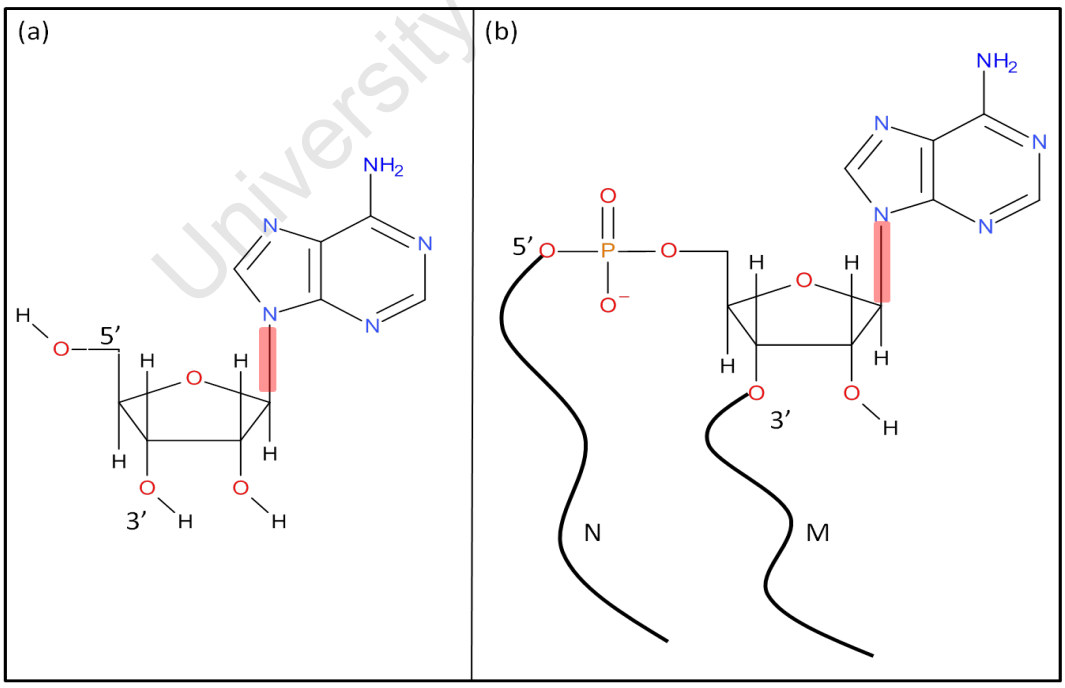


Figure 2 : (a).The minimum substrate model(MSM). 3' and 5' end are converted to hydroxyl groups to satisfy the valancies. (b) The loop substrate model (LSM). N and M represent the number of nucleobases that can be present on either side of the target adenine.

Previously reported[17] molecular dynamics simulations on similar substrate-like nucleotide sequences in ricin binding site have used additional NMR-derived distance constraints to maintain the integrity of the substrate. In order to discover an appropriate loop model that can be used without such external constrains, nucleotide sequences that are composed of total of 4, 8, 6, 10 and 12 nucleobases were investigated with classical molecular dynamics simulations. From that, it was found that any loop model that is shorter than 12 bases in total is not capable of maintaining the integrity of the base-paired loop structure (hairpin structure) without an external restraining force. Therefore, the LSM model proposed here is a 12-mer oligonucleotide sequence. The starting structure of that sequence was obtained from the NMR-solved structure of GNRA[18] hairpin loop which includes the GAGA tetra loop targeted by ricin.

The MSM model was proposed mainly due to its smaller size and expected minimal computational complexities whereas the LSM model was proposed due to its high resemblance to the natural substrate of ricin, the hairpin loop. This chapter discusses the aptness of the above proposed models, MSM and LSM for computer simulations of ricin-substrate complexes.

4.4. Simulation Details

The initial structure of ricin was obtained from x-ray crystallographic studies of substrate analogue inhibitor, formycin monophosphate (FMP) bound ricin[19]. Protonation states of amino acid residues were determined by using the WHATIF program[20]. The MSM-ricin complex was constructed by converting FMP into adenine in the FMP-ricin complex. The 12-mer LSM was constructed by truncating the NMR-determined structure of the GNRA oligonucleotide sequence[18]. The LSM was then manually docked into the ricin(from FMP-ricin complex) binding site. Prior to docking of LSM, the susceptible adenine was flipped into an extrahelical position as suggested by many previous studies.[21-25] Docking of LSM was done by placing the flipped adenine in the same orientation as FMP in the FMP-ricin complex.

After the initial setup, the ricin-substrate complexes (MSM and LSM) were solvated in a TIP3P[26] water sphere of 23.5Å radius with a buffer boundary of 20.0 Å, centred on the C1' of target adenine. The solvated systems were then subjected to geometry optimisation with Adopted Basis Newton–Rhaphson (ABNR) method for 1000 steps. After minimisation, the substrate-ricin system was carefully analysed to ensure the binding patterns suggested by the literature were present.

The minimised complexes were gradually heated from 50 to 300 K in 500ps with molecular dynamics. All dynamics simulations were performed with the CHARMM[27] program with the all-atom CHARMM27 nucleic acid force field[28,29] , and were carried out at pH 7.00 with stochastic boundary conditions[30]. The SHAKE[31] algorithm was used to fix the positions of the protons. The leapfrog integrator was used for dynamics and the non-bonded interactions were handled with group based cutoffs of 10, 12, and 14Å. For smoothing purposes, switching functions were applied to both electrostatic and van der Waals interactions. The final collection run was 10ns in duration.

4.5. Results and Discussion

4.5.1. Binding of MSM and LSM to Ricin

The amino acid residues TYR80, VAL81, GLY121, TYR123, GLU177, ARG180 and GLU208 have been previously identified[5,19] as key residues in the ricin binding site. Ligand binding diagrams for target adenine in MSM and LSM are presented in Figure 3. It was found that target adenine in LSM has demonstrated its ability to maintain the previously proposed[2,19,32] "key-interactions" with the binding site residues. These key interactions include simultaneous pi-stacking interactions with TYR80 and TYR123, strong H-bond interactions between: N1 and VAL81 , N3 and ARG180 and H62 and GLY120. In comparison, only pi-pi interaction with TYR80 and H-bonding between N1 and VAL 81 were found in MSM.

The 2D interaction diagrams of MSM shows that MSM containing ricin binding site is less crowded than the LSM containing binding site. i.e. relatively fewer interactions are present between MSM and the binding site residues. Furthermore, it is clear that most of the proposed binding site residues are not interacting with MSM. As discussed before, this may result from fast conformational changes that may occur in MSM bound ricin.

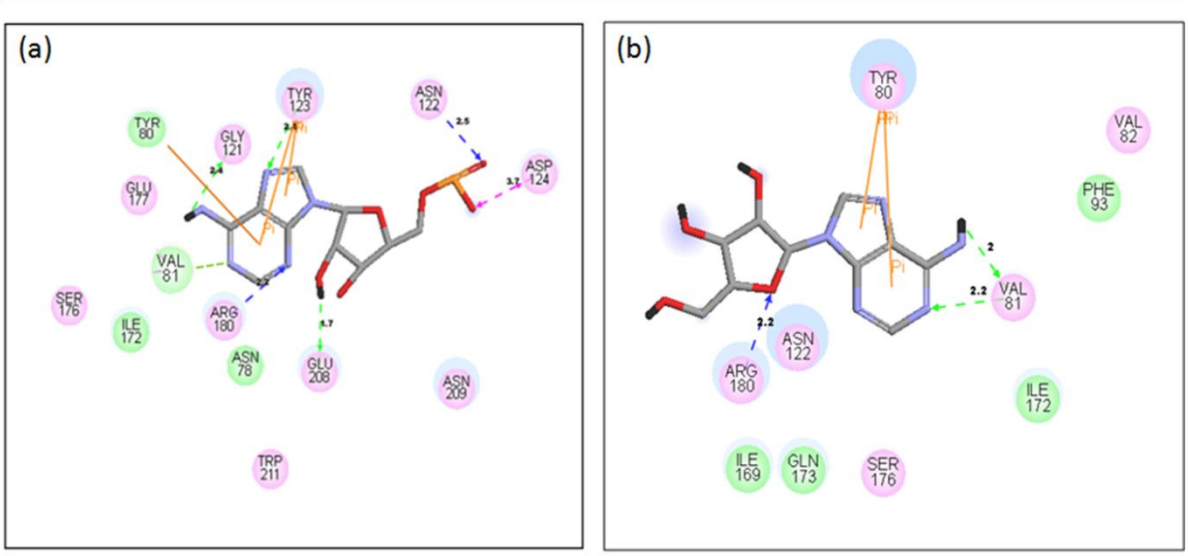


Figure 3: Binding of (a) MSM and (b) LSM to the ricin binding site. Hydrogen bonding interactions are shown in green and blue dashed lines. Amino acid residues that are highlighted in green indicates effective van der Waals interactions and the residues highlighted in magenta indicates effective electrostatic interactions. Pi-Pi interactions are indicated with orange lines.

The strength of Binding of MSM and LSM to ricin was investigated by means of Interaction Energy (IE) calculations. Interaction energies presented here are classical interaction energies and they were calculated with CHARMM force field. The corresponding IE time series plots are presented in Figure 4. (a). It was found that the total average interaction energy of MSM with ricin is -60.70 kcal/mol whereas that of LSMS is -175.90 kcal/mol. The total average energy is defined as the average of the sum of electrostatic and van der Waals interaction energies over the total simulation time of 10ns. Only adenine, was included (i.e. the rest of the loop was ignored) in the IE calculations of LSM. The analysis of IE shows that LSM is binds to ricin approximately -115.20 kcal/mol more strongly than MSM.

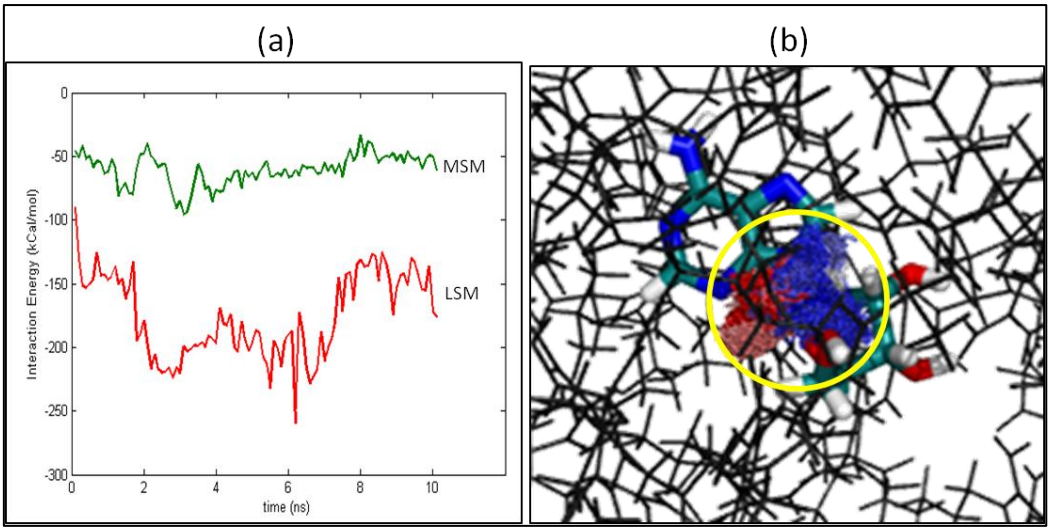


Figure 4 : (a). Interaction Energy (IE) time series of MSM and LSM with ricin. The entire protein structure was included the IE calculations. (b). Motion path of centre of mass (COM) of MSM. The yellow circle demarcates the region within the COM was moving. Motion path of COM is drawn with blue, red and orange fine lines.

The relatively low interaction energy observed with MSM provides a strong indication of disturbed ricin-MSM interactions. In order to investigate if the low interactions were caused by diffusing of the MSM out of the binding site, its residence in the binding site was first analysed. This analysis was performed by tracing the motion path of the centre of mass of MSM (Figure 4. (b)). This analysis was not performed for LSM as free diffusion of LSM inside the binding site is not possible due to its molecular size. Inspection of motion path of the centre of mass of MSM shows its position within the binding site does not change significantly throughout the simulation. Therefore, it can be concluded that the low IE observed in MSM is not due to any migration of the MSM out of the binding site. Another possible cause of a low interaction energy is fast conformational changes of the ligand and / or the binding site.

In order to investigate the above possibilities, analyses of the root mean square deviation (RMSD) of the protein and substrates were carried out. Figure 5(a) presents the RMSD plots of the enzyme in the presence of MSM and LSM and that of the individual models. It can be clearly seen from RMSD plots that the protein undergoes continuous changes in the presence of MSM. Further, the observed maximum RMSD of 4.57 of ricin with MSM in comparison to that of 1.22 when bound to LSM clearly suggests a highly dynamic nature for MSM-bound ricin binding site, which can contribute to poor MSM-ricin interactions.

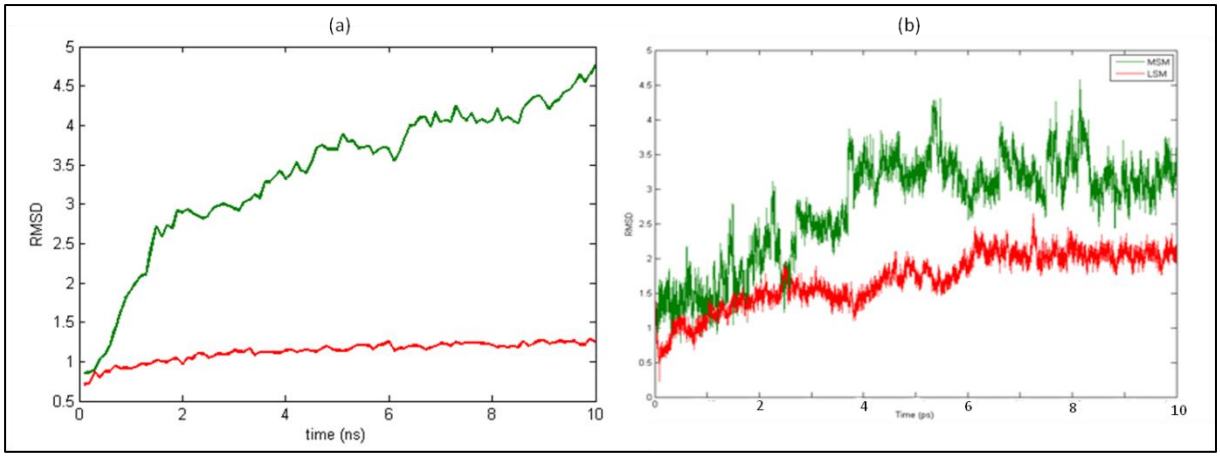


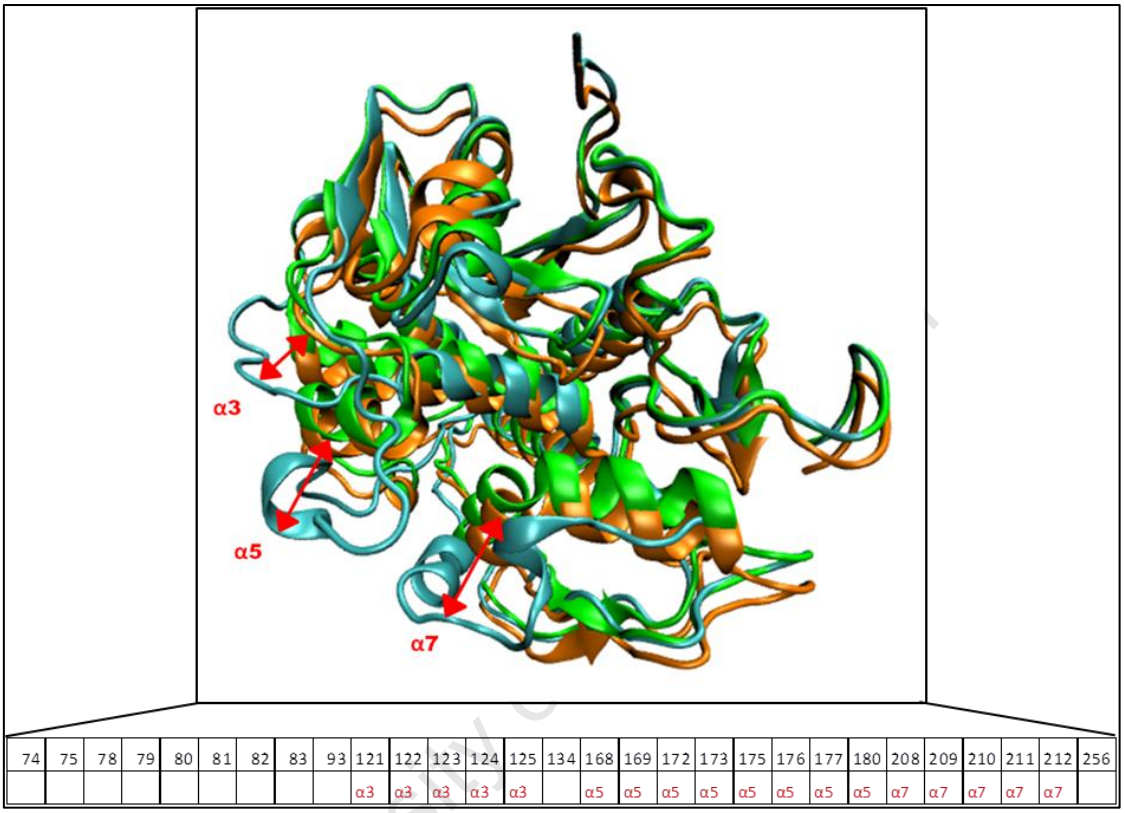
Figure 5: (a). RMSD of ricin when bound to MSM (green) and when bound to LSM(red). (b). RMSD of MSM and LSM inside the binding site.

From the RMSD plots of individual models (Figure 5.(b)), it can be seen that MSM undergoes many conformational fluctuations compared to LSM. i.e. it can be seen that the LSM model is relatively rigid during the 10ns simulation time. The relatively restricted flexibility of LSM agrees with the previously suggested[33] less dynamic behaviour of the sarcin-ricin domain in solution.

The RMSD time series plots of protein and substrate models suggest a strong correlation between conformational changes in the protein and the flexibility of the substrate model. It is assumed that the higher degree of conformational freedom of MSM induces conformational changes in the protein, and consequently it does not get properly 'locked-in' to the enzyme.

Figure 6 presents superimposed structures of MSM-bound and LSM-bound ricin after 10ns of molecular dynamics simulations on to the crystal structure of ricin. It was found that the LSM bound ricin structure is in good agreement with the crystal structure. Whereas, the MSM bound ricin showed significant conformational changes in three distinguishable regions; alpha helix-3(α3), alpha helix-5(α5) and alpha helix-7(α7). It was found that those conformational changes were mainly caused by unravelling of the above mentioned α -helices.

Moreover, it was found that the binding site residues were mainly involved in those conformational changes. The binding site residues that belong to α3, α5 and α7 are also presented in Figure 6.



74	75	78	79	80	81	82	83	93	121	122	123	124	125	134	168	169	172	173	175	176	177	180	208	209	210	211	212	256
									α3	α3	α3	α3	α3		α5	α5	α5	α5	α5	α5	α5	α5	α7	α7	α7	α7	α7	

Figure 6: conformation of: MSM bound ricin (blue), LSM bound ricin(orange) and x-ray structure of ricin (green). In the table, first raw of the table presents the proposed binding site residues and the second raw presents the α-helix that they belong to.

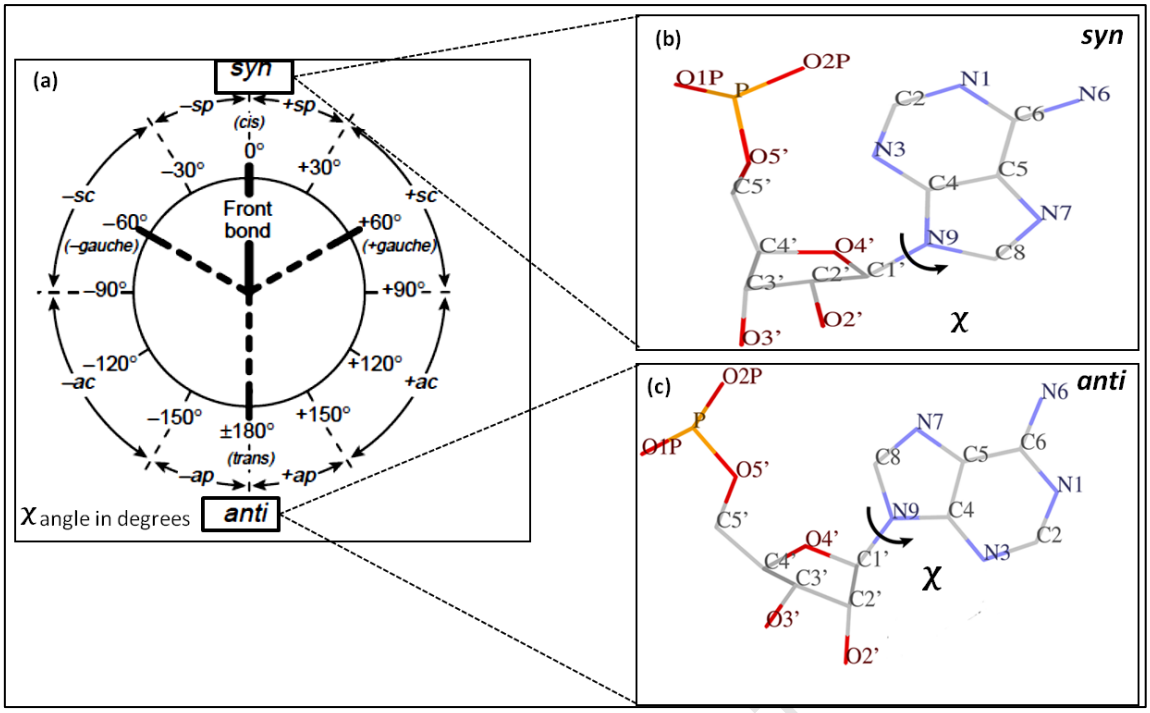
4.5.2. Conformational Behaviour of MSM and LSM in Ricin Binding Site

4.5.2.1. Rotation Around the C1'-N9 Glycosidic Bond

It is well known that the glycosidic bond of the target substrate is distorted by the glycosidase enzymes during the reaction[34]. Further, it has been proposed that twisting of the target glycosidic bonds by glycosidase enzymes affects certain functionalities that can cause enhancement of the cleavage of that bond. The most apparent of these functionalities is the increased access to the corresponding reaction centres (the two atoms between which the target bond is found) by the catalytic residues i.e. elevated exposure of the susceptible bond. This is a direct consequence of less steric hindrance that results from rotation around the bond of interest. Second functionality is the changes in electronic structure of the molecule due to glycosidic torsional rotation. Another functionality is the strain that is applied on the cleaving bond by increased intermolecular energy by torsional distortion. The effects of C1'-N9 rotation on these functionalities emphasise the importance of investigating this rotation in the proposed substrate models (MSM and LSM).

The rotation around the glycosidic bond (C1'-N9) can be described by using the so called χ angle , which is defined as the dihedral angle of atoms O4', C1', N9 and C4[35]. In nucleobases two major rotameric conformations can be identified based on the value of χ. According to the IUPAC nomenclature, they are defined as *anti* and *syn* conformations[35,36]. Scheme 1 presents the definitions of these conformations.

Switching from *anti* to *syn* or from *syn* to *anti* is defined as "base-flipping" in the IUPAC nomenclature[35]. However, the usage of term "base-flipping" is often confusingly used when referred to DNA or RNA structures. At those instances, "base-flipping" is defined as a distortion in Watson-Crick base paired structures where a target base that is normally stacked inside the duplex (two strands) is completely rotated out of the "double-helical structure" into an extrahelical position[37]. It has been proposed by previous studies[38] that nucleobase hydrolytic enzymes including ricin are utilizing a "base-flipping" mechanism to reach and identify their targets.



Scheme 1: Definition of rotamer conformations around glycosidic bond. (a).Wheel gives the different rotamer confonformations based on the χ angle. (b). *syn*(c). *anti conformation* of adenine.

The χ angle time series for MSM and LSM are presented in Figure 7(a). The average value of χ for MSM is -69.97° whereas that value for LSM is -143.56. Therefore, it is suggested that MSM prefers to be in the *–sc* rotameric conformation (Scheme 1) where as LSM prefers to be in the *–ap* conformer. Analysis of χ-space sampling shows that both MSM and LSM sample the conformers between *–ap* to *+ap* in first 2.75ns of the simulation. After 2.75 ns, MSM samples in the *–sc* region and LSM continues to sample in the *–ap* region.

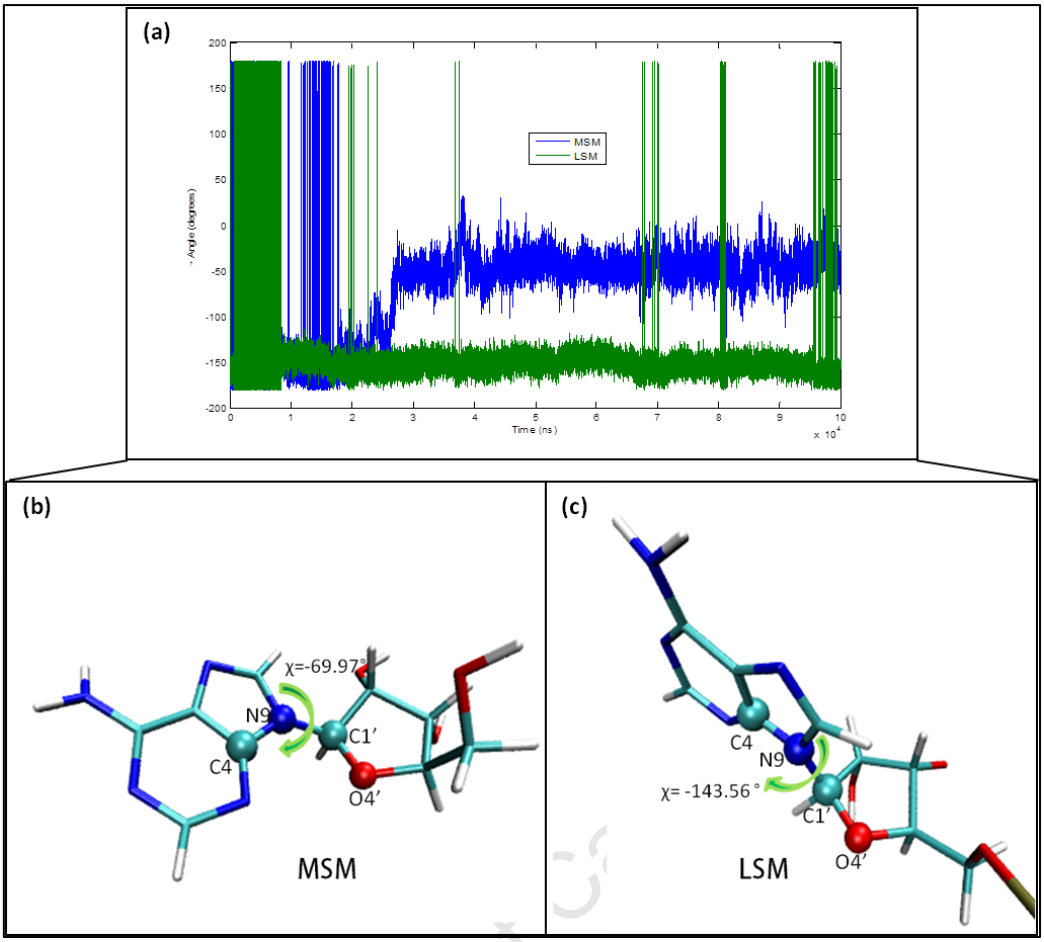


Figure 7: (a). Time series of χ angle for MSM and LSM.
(b). Average conformation of MSM. (c).Average structure of LSM.

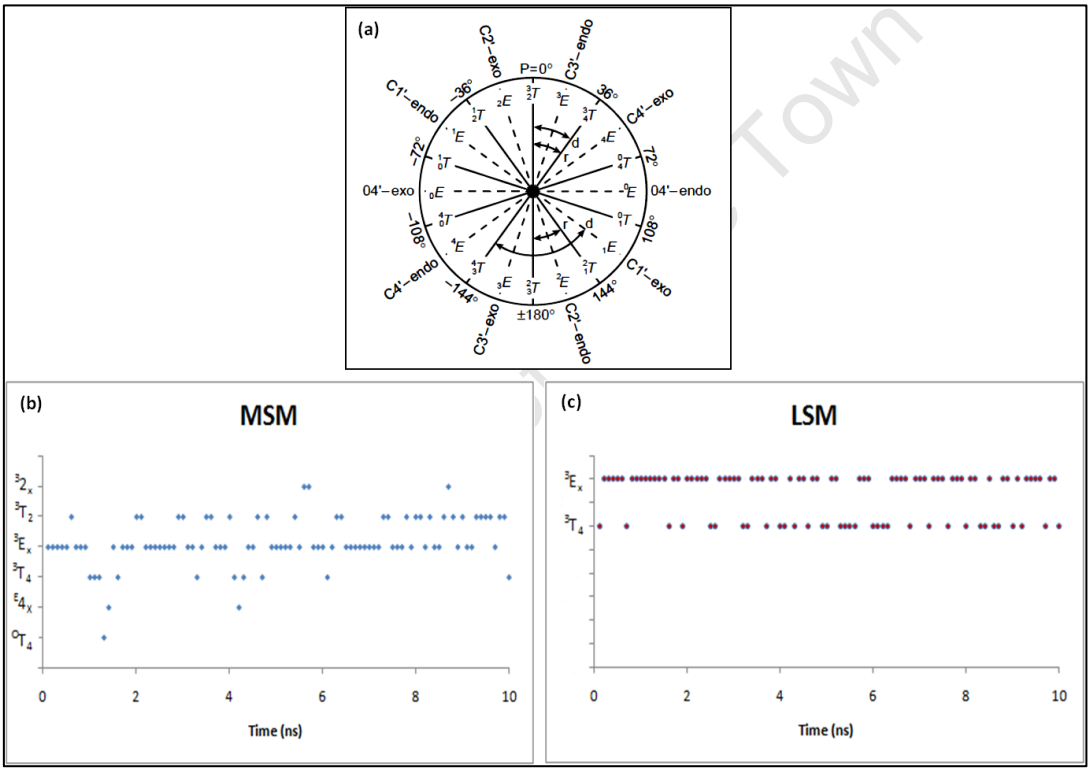
The rotameric isomers of MSM and LSM that correspond to the average χ angles are shown in Figure 7.(b) and Figure 7.(c) respectively. The rotameric conformer -*ap* of adenine in LSM corresponds to so-called "flipped-out" (from the helical RNA double strand) conformation, which is closer to the *anti* conformation. In MSM, the –*sc* conformation is closer to *syn* and it corresponds to un-flipped conformation. Since ricin is expected to employ a base-flipping mechanism in substrate recognition, with respect to the χ rotation, the flipped conformation in LSM is expected to be a better substrate conformation over the un-flipped conformation in MSM.

4.5.2.2. Puckering of the Ribose Ring

Puckering of the ribose ring in nucleobases has many important implications on their reactivity and the overall geometry. In solution, the sugar ring of nucleic acids equilibrates between two extreme forms of ring pucker neighbouring a 2'-*exo*/3'-*endo* (Northern) conformation and the opposite 2'-*endo*/3'-*exo* (Southern) conformation[39]. The IUPAC definitions of pucker conformations are presented in Scheme 2. It has been suggested that preference for any of specific pucker conformations in solution is determined by the interplay of important interactions resulting from anomeric and *gauche* effects[39,40]. Moreover, it has been proposed that when a nucleoside or nucleotide binds to its target enzyme, one pucker form is expected to dominate in the active site[41].

Pucker conformational analysis were performed for MSM and LSM to investigate the correlation (if any exists) between the substrate model and the puckering of the corresponding ribose ring. The time series of pucker conformations for MSM and LSM are presented in Scheme 2(b) and Scheme2(c) respectively. Pucker analyses found that ribose in MSM is more flexible and it equilibrates between six distinguishable pucker conformations. However, pucker time series shows that MSM prefers be in $^{3}E_{X}$ conformer over the other conformations. On the other hand, it was found that ribose puckering in LSM was very restricted. In LSM, ribose puckers between the $^{3}E_{X}$ and $^{3}T_{4}$ conformers. Moreover, in LSM the ribose ring also prefers to be in the $^{3}E_{X}$ conformer. The $^{3}E_{X}$ conformer corresponds to the C3'-*endo* conformation in the alternative pucker nomenclature scheme (Scheme 2). Further, $^{3}T_{4}$conformation (the only other conformation showed by LSM) can be considered as the upper boundary of C3'-*endo* conformation. Therefore, it can be concluded that in LSM, ribose prefers to be predominantly in C3'-*endo* conformation. This pucker behaviour is in agreement with previous suggestions[10] where it has been shown that the ricin's targeted adenine is found 75-80% in the C3'-*endo* conformation. Therefore, it can be concluded that the loop substrate model (LSM) that is proposed here is capable of producing accurate ribose pucker conformations.

It is known that, one among the many ways of optimizing the catalytic activity of glycosidase enzymes is to restrict the substrate's sugar ring in a particular conformation. This mainly allows aligning of the substrate with the catalytic residues and make susceptible carbon (C1') in line with the incoming nucleophile. Therefore, the restricted puckering of the ribose ring is considered as enhancing factor of the reactivity of the substrate. The LSM's ability to demonstrate the correct and restricted pucker conformations when compared to MSM certainly qualifies it as a better substrate with respect to the ribose puckering.



Scheme 2: (a) Definitions of ribose pucker conformations. Ribose pucker time series of : (a). MSM. (b). LSM.

4.6. Concluding Remarks

As discussed in the previous sections, the loop substrate model (LSM) was found to be capable of demonstrating the expected binding, geometrical and dynamical behaviour when compared to the minimum substrate model (MSM). More specifically, LSM's tight binding to ricin through the correct enzymatic contacts, being in the expected extrahelical flipped position, preferring C3'-*endo* ribose certainly qualifies this model as a better substrate model. Moreover, this study has clearly shown the inaccuracy of the MSM as a substrate model for ricin and it has shown that LSM can be used as an accurate substrate model for ricin.

The interaction energy between ricin and LSM was calculated to investigate the strength of binding LSM to ricin. It was found that the total (sum of van der Waals and electrostatic) average interaction energy of ricin with LSM is -652.48 kcal/mol. Further, the contribution from electrostatic to this total interaction energy was found to be -550.67 kcal/mol and it is 86% of the total interaction energy. This interaction energy calculation shows that ricin interacts with LSM predominantly via electrostatic interactions. This finding strongly agrees with previously suggested[42] non-specific electrostatic binding of ricin to its target RNA loop. Therefore, it is suggested here, that the absence of negatively charged backbone in MSM is responsible for the conformational changes occur in ricin in MSM-ricin dynamics. On the other hand, it is suggested that presence of the negative backbone in LSM allows it to tightly binds to ricin and restrict the undesired protein conformational changes.

After taking all the factors discussed above, the LSM was chosen as the model to be used in the subsequent studies. Previous studies[7] have used the standalone adenine nucleoside as a substrate model for ricin in quantum mechanical calculations. In those calculations it was attempted elucidate to transition state structure and the protonation state of ricin's substrate. This study has clearly shown that the standalone adenosine nucleoside i.e. MSM cannot be used as a reasonable substrate model for ricin. Therefore, it is argued here that previously proposed[7] gas phase TS might not be a true representation of the TS that is found in the enzyme.

4.7. References

(1) Kim, Y.; D Robertus, J. *Protein Engineering* **1992**, *5*, 775.
(2) Endo, Y.; Tsurugi, K.; Ebert, R. F. *Biochimica et Biophysica Acta* **1988**, *954*, 224.
(3) Endo, Y. *Journal of Biological Chemistry* **1987**, *262*, 5908.
(4) Endo, Y.; Mitsui, K.; Motizuki, M.; Tsurugi, K. *The Journal of biological chemistry* **1987**, *262*, 5908.
(5) Michael, J.; Roberts, M.; Robertust, J. O. N. D. *The FASEB Journal* **1994**, *8*, 201.
(6) Chen, X. Y.; Link, T. M.; Schramm, V. L. *Biochemistry* **1998**, *37*, 11605.
(7) Chen, X.-Y.; Berti, P. J.; Schramm, V. L. *Journal of the American Chemical Society* **2000**, *122*, 1609.
(8) *Molecular Modeling of Nucleic Acids*; American Chemical Society, 1997; Vol. 682.
(9) Orita, M.; Nishikawa, F.; Shimayama, T.; Taira, K.; Endo, Y.; Nishikawa, S. *Nucleic acids research* **1993**, *21*, 5670.
(10) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H. M.; Pardi, A. *Journal of Molecular Biology* **1996**, *264*, 968.
(11) Antao, V. P.; Lai, S. Y.; Tinoco, I. *Nucleic Acids Research* **1991**, *19*, 5901.
(12) Antao, V. P.; Tinoco, I. *Nucleic Acids Research* **1992**, *20*, 819.
(13) Uhlenbeck, O. C. In *Nature* 1990; Vol. 346, p 613.
(14) Correll, C. C.; Munishkin, A.; Chan, Y.-L.; Ren, Z.; Wool, I. G.; Steitz, T. A. *Proceedings of the National Academy of Sciences* **1998**, *95*, 13436.
(15) Dibrov, S. M.; Parsons, J.; Hermann, T. *Nucleic Acids Research* **2010**, *38*, 4458.
(16) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press, 1988.
(17) Olson, M. a. *Proteins* **1997**, *27*, 80.
(18) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H.; Pardi, A. *Journal of Molecular Biology* **1996**, *264*, 968.
(19) Monzingo, A. F.; Robertus, J. D. *Journal of Molecular Biology* **1992**, *227*, 1136.
(20) Vriend, G. *Journal of Molecular Graphics* **1990**, *8*, 52.
(21) Yang, X.; Gérczei, T.; Glover, L. T.; Correll, C. C. *Nature structural biology* **2001**, *8*, 968.
(22) Yang, X.; Gerczei, T.; Glover, L.; Correll, C. C. *Nature Structural & Molecular Biology* **2001**, *8*, 968.
(23) Roberts, R. J.; Cheng, X. *Annual review of biochemistry* **1998**, *67*, 181.
(24) Correll, C. C.; Yang, X.; Gerczei, T.; Beneken, J.; Plantinga, M. J. *Journal of Synchrotron Radiation* **2004**, *11*, 93.
(25) Chen, C.; Jiang, L.; Michalczyk, R.; Russu, I. *Biochemistry* **2006**, *45*, 13606.
(26) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926.
(27) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187.
(28) Mackerell, A. D.; Banavali, N. K. *Journal of Computational Chemistry* **2000**, *21*, 105.
(29) Foloppe, N.; MacKerell, J. A. D. *Journal of Computational Chemistry* **2000**, *21*, 86.
(30) Brooks, C. L.; Brünger, A.; Karplus, M. *Biopolymers* **1985**, *24*, 843.
(31) Ryckaert, J.; Ciccotti, G.; Berendsen, H. *Journal of Computational Physics* **1977**, *23*, 327.
(32) Ho, M.-C.; Sturm, M. B.; Almo, S. C.; Schramm, V. L. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 20276.
(33) Špačková, N. a.; Šponer, J. *Nucleic Acids Research*, *34*, 697.
(34) Johnson, G. P.; Petersen, L.; French, A. D.; Reilly, P. J. *Carbohydrate research* **2009**, *344*, 2157.
(35) Markley, J. L.; Bax, a.; Arata, Y.; Hilbers, C. W.; Kaptein, R.; Sykes, B. D.; Wright, P. E.; Wüthrich, K. *European journal of biochemistry / FEBS* **1998**, *256*, 1.

(36) *IUPAC. Compendium of Chemical Terminology*; 2 ed.; McNaught , A. D.; Wilkinso, A., Eds.; Blackwell Scientific Publications. : Oxford., 1997.
(37) Huang, N.; MacKerell, A. D. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* **2004**, *362*, 1439.
(38) Berti, P. J.; McCann, J. A. B. *Chemical Reviews* **2006**, *106*, 506.
(39) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
(40) Plavec, J.; Tong, W.; Chattopadhyaya, J. *Journal of the American Chemical Society* **1993**, *115*, 9734.
(41) Marquez, V. E.; Siddiqui, M. A.; Ezzitouni, A.; Russ, P.; Wang, J.; Wagner, R. W.; Matteucci, M. D. *Journal of Medicinal Chemistry* **1996**, *39*, 3739.
(42) Li, X.-P.; Chiou, J.-C.; Remacha, M.; Ballesta, J. P. G.; Tumer, N. E. *Biochemistry* **2009**, *48*, 3853.

CHAPTER -5

PRE-PROTONATION STATE OF RICIN'S TARGET SUBSTRATE

5.1. Introduction

Nucleobases adenine (A), cytosine (C) , guanine (G), thymine (T) and uracil (U) are chemically responsible for coding the genetic information in living organisms. Configurations of these molecules are presented in Figure 1. These heterocyclic aromatic molecules are classified as monocyclic pyrimidines or bicyclic purines. Both pyrimidines and purines are essentially planer molecules with limited conformational flexibility[1]. A nucleotide is formed when a base is joined from a ring nitrogen to C1' of a pentose sugar. Nucleotides are defined as phosphate esters of nucleosides.

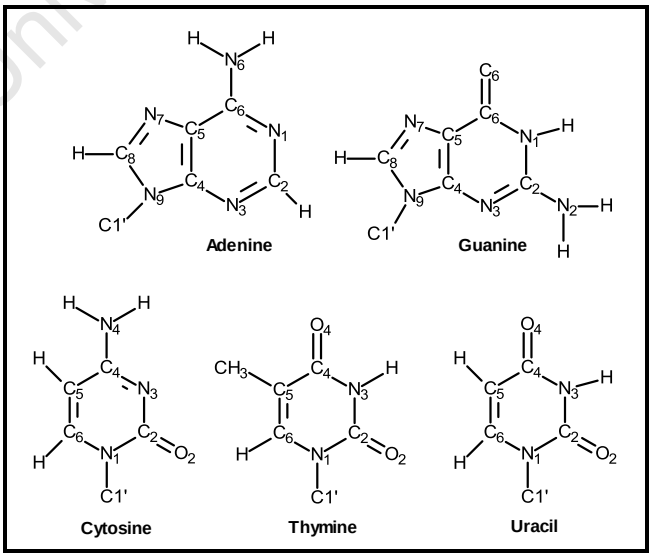


Figure 1: Configurations of nucleobases.

Nucleobases can undergo various types of chemical modifications in their natural environment[2]. Protonation of these bases is common as it occurs relatively easily and very as it significantly alters the chemistry of the parent molecule. The bases A, C and G get preferentially protonated on ring nitrogens rather than on the exocyclic amino group since that does not involve delocalisation of NH_2 lone pair in to the aromatic ring[3]. It has been suggested that the proximity of negative charges of the backbone phosphate imposes a secondary effect by rendering increasing basicity of the ring nitrogens while at the same time decreasing the acidity of the hydrogens[3]. In some bases such as adenine where multiple protonation sites are available, single or multiple protonation states can be exhibited. Owing to the presence of multiple protonation sites, nucleobases can exist in different tautomeric forms. Consequently, it gives rise to keto-enol equilibrium between 2-pyrodone and 2-hydroxypyridine and amine-imine equilibrium for 2-aminopyridine.

It has been proposed that the protonation of nucleotides can play an important role in both nucleic acid structure and catalysis[4]. Hence, phenomena such as base pairing, self-association, interaction with metal ions and molecular recognition by proteins are affected by the protonation state(s) of the nucleobases[5]. First-order dependence of the reaction rate on proton concentration[6,7] in hydrolysis reactions of purines in solution, has shown that the protonated nucleosides are the reactive species in such reactions. Moreover, those experiments have clearly demonstrated the effect of protonation on the chemical reactivity.

It has been recently shown[4] that adenines within internal RNA loops also have elevated proton affinities. Free adenine mononucleotides show pKa values around 3.5 where as adenines in loop structures show significantly high pKa values ranging from 4.8 to 5.8. It is assumed that protonation of adenine stabilizes the internal loop structure by maximizing the charge stabilization, hydrogen bond formation, and stacking interactions. It has been shown that hydrolysis and phosphorolysis reactions of N-glycosidic bond proceed via mechanisms that involve the protonation of leaving nucleobase[8]. Site directed mutagenesis studies have shown that the face-to-face pi-stacking between aromatic substrates and the aromatic enzymatic side chains promotes the protonation of substrate by elevating its pKa values[8].

Other than the contributions to structure stabilisation, protonation of nucleobases significantly facilitate the leaving group departure and enhancement of glycosidic bond cleavage in glycosidase reactions. The former is achieved by yielding an uncharged leaving group whereas latter is achieved by making the leaving base electron deficient which results in electron withdrawing from the breaking bond.

5.2. Objectives

This is chapter discusses the protonation of target adenine of ricin, A-4324 which is found in GAGA tetra loop of the sarcin-ricin loop (SRL) of 23S rRNA in the large ribosomal subunit. In order to study the mechanism of the reaction catalysed by ricin, prior knowledge of its substrate's protonation state is mandatory. Even though, protonation states of free adenine, adenosine and adenine nucleotide have been studied both experimentally[5,9] and theoretically[10-12] , protonation state of ricin's natural substrate has not been confirmed yet. In particular, this chapter attempts to address the effect on chemistry, conformation and binding of A-4324 with respect to its different protonation states.

5.3. Pre-protonation of Target Adenine; Background

Adenine has three potential sites for protonation. According to nucleic acid nomenclature, theses sites are identified at N1, N3 and N7 positions (Figure 1). Protonation of these sites can result in strong electrostatic interactions or hydrogen bonding with the enzyme's binding site. Moreover, it has been suggested that such interactions can cause electron withdrawing effects from the ring's pi-system and hence can facilitate the glycosidic bond cleavage[13]. Further, previous studies have shown that in the acid catalyzed hydrolysis of adenine nucleoside involves a multiple protonation strategy[7].

However, the exact sites of protonation, the source of protons and the mechanism of protonation still remain an unsolved problem in ricin chemistry. Various studies have suggested different answers to this problem and the rest of this section discusses the current status of the problem.

Some inhibition studies[13] have suggested a double protonation at N3 and N1 sites in hydrolysis of adenosine by ricin. Protonation of N3 by ARG180 has also been suggested by the crystallographic experiments[14] of substrate analogues and it has been confirmed by the site directed mutagenesis studies[15] and theoretical studies[12].

Despite the above mentioned evidences to support the protonation at N3, some studies have proposed protonation of target adenine at N1. In particular, kinetic isotope effect (KIE) studies by Schramm et al and theoretical studies by Bene[16] et al have shown the possibility of protonation at N1. However, it has been suggested the the energetic contributions from protonation of N1 to the glycosidic bond cleavage are not significant when compared to the contributions from N1 and N7 double-protonation[17].

Protonation of N7 has also been proposed by several studies[13]. Inversed isotope effects observed at N7 have suggested that protonation of N7 facilitates the catalytic reaction leading the reactants to the transition state[13]. It has been proposed[18] that ASP96 is responsible for protonating N7. Studies done by Versees et al[8] have suggested that protonation at N7 facilitates the leaving group departure in purine nucleoside hydrolysis. Further, ^{15}N NMR experiments done on ATP prototypes have reported the presence of a mixed population of N1 and N7 protonated species in solution. Some studies have also shown that the N1 protonation is clearly favoured in the gas phase whereas the N7 protonation is favoured in the aqueous phase by 0.5 kcal/mol for a particular ATP derivative 8-NH2- adenine. These studies have reinforced the idea that N7 and N1 are “competitive” sites for protonation in adenine.

5.4. Simulation Details

5.4.1. Dynamics Simulations

Hybrid QM/MM dynamics simulations were employed to investigate the binding of different tautomers of target adenine to the ricin binding site. Initial structure of the target GAGA tetra loop was obtained from NMR studies[19]. For ricin, the starting structure was obtained from crystallographic studies of substrate analogue formycin monophosphate (FMP) [20]. Protonation states of amino acid residues were determined using the WHATIF program[21].

Since automated docking of the 12-mer RNA loop to ricin binding site was not possible due to the large number of atoms, it was manually docked to ricin. The binding pattern of FMP was used as guidance in the docking process. Entire substrate-ricin system was then subjected to geometry optimization with Adopted Basis Newton–Rhaphson (ABNR) method. After minimization, the substrate:ricin system was carefully analysed to ensure the correct binding pattern as suggested by the literature. Then the system was gradually heated from 50 to 300 K in 100ps. Prior to the QM/MM simulations, the substrate-ricin complex was equilibrated for 10ns using classical Molecular Dynamics (MD) simulations.

All simulations were performed with CHARMM[22] program with the all-atom CHARMM27 nucleic acid force field[23,24] and were carried out at pH 7.00 with TIP3P[25] water model using stochastic boundary conditions[26]. A water sphere of 23.5Å radius was used with a buffer boundary of 20.0 Å. The water sphere was centred on the C1' of target adenine. The SHAKE[27] algorithm was used to fix protons. The leapfrog integrator was for dynamics and non-bonded interaction were handled with group based cutoffs of 10, 12, and 14Å. For smoothing, purposes switching functions were applied to both electrostatic and van der Waals interactions.

A snapshot of the system taken after 10ns from MD simulations was taken as the starting structure for the subsequent QM/MM simulations. The QM region was composed of reaction moieties of amino acids TYR80, TYR123, GLU177, ARG180 target adenosine (with 3 the' phosphate group) and the catalytic water molecule (Figure 2). The QM/MM boundary was treated with the link-atom approach[28] available in CHARMM program. An approximate density-functional method SCC-DFTB-D[29] was used to treat the QM region and the MM region was handled with the CHARMM27 nucleic acid force field[23,24].

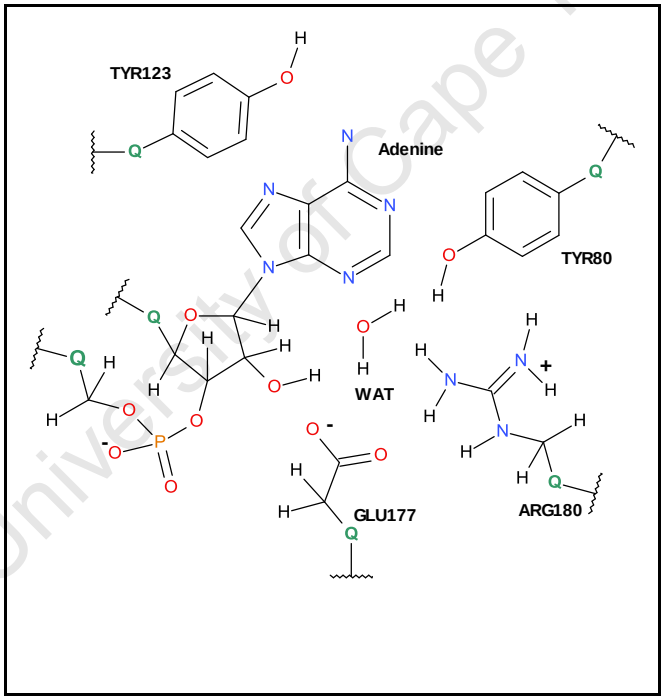


Figure 2: QM region of the QM/MM setup. Link atoms are indicated with green letter Q.

5.4.2. Electronic Structure Calculations

All static electronic structure calculations were performed with the Gaussian 09 programme[30]. The starting structures for QM interaction energy calculations were obtained by minimizing snapshot-structures form 1 nS long QM/MM simulations. Energy minimization was done with SCC-DFTB-D level of theory for 1000 steps. This was done to ensure that the selected "snap-shot" structure is in a low energy conformation. A truncated binding site (TBS) and the ligand (L) were then clipped off from the original ricin-substrate complex. The truncated binding site was composed of all amino acid residues within 3Å radius of the ligand. The dangling valencies were capped with hydrogen atoms. All calculations were carried out in the gas phase with MO6-2X/631-G(d) level of theory[31].

When calculating the enzyme-substrate interaction energy, it is not practical to use *ab-initio* calculations they are very demanding in terms of time and computational recourses for large systems. The binding of ligands to the ricin binding site involves a great deal of Pi-Pi type (Pi-stacking) interactions. Therefore, relatively inexpensive Density Functional Theory (DFT) calculations are also not suitable for this particular system due to DFT's inability to account for the dispersion effects. Therefore, in this study , meta-hybrid functional, MO6-2X[31] with dispersion was used in this study to calculate the quantum mechanical interaction energies . Specially this functional is considered to perform well in calculating the geometries and energies of large bio molecules such as DNA and RNA[32] .

The interaction energy between the ligand, L the truncated binding site, is given as,

[1]

where, $IE_{(L:TBS)}$, $E_{(TBS)}$ and $IE_{(L)}$ are the single point energy of the ligand-binding-site complex, free TBS and L respectively.

5.5. Results and Discussion

The results are presented in a comparative fashion for unprotonated adenine (UPA) and protonated adenine at N7 position (N7HA). Selection of these two protonation states and elimination of protonation at N1 and/or N3 is discussed first. As revealed from QM/MM dynamics simulations, the binding patterns of UPA and N7HA to ricin are discussed in the following section. Figure 3(a) and 3(b) illustrate the binding of UPA and N7HA respectively. Table 1 presents the average distances between the corresponding ligands and the key residues in the binding site. These interacting-distances were taken in to account with the corresponding binding orientations to propose the nature of binding each ligand.

Interactions		UPA		N7HA	
Description	ID	Dist (A)	Type	Dist	Type
TYR80 - ADE	PP1	5.04	Pi-Pi(S/T)	3.65	Pi-Pi(S)
TYR123-ADE	PP2	4.48	Pi-Pi(S/T)	5.51	Pi-Pi(S)
ARG 180(HH12)-N3	HB4	1.82	H-Bonding	2.80	H-Bonding
ARG 180(HH22)-N3	HB4’	2.74	H-Bonding	3.26	H-Bonding
GLY121(O)-H61	HB5	2.44	H-Bonding	4.21	
TYR123(HN)-N7	HB6	2.57	H-Bonding	4.60	
VAL 81(O)-H62	HB3	3.05	WK-HB	1.91	H-Bonding
VAL 81 (HN)-N1	HB2	2.51	H-Bonding	2.19	H-Bonding
TYR123(HN)-O4'	HB7	4.24	-	2.54	H-Bonding
GLY121(O)-N7/**(HB1)**	HB1	4.06		2.24	H-Bonding
TYR123(OH)-N7	ES1	6.63		3.41	Strong electrostatic

Table 1: Proposed binding interactions for UPA and N7HA based on distance analysis.

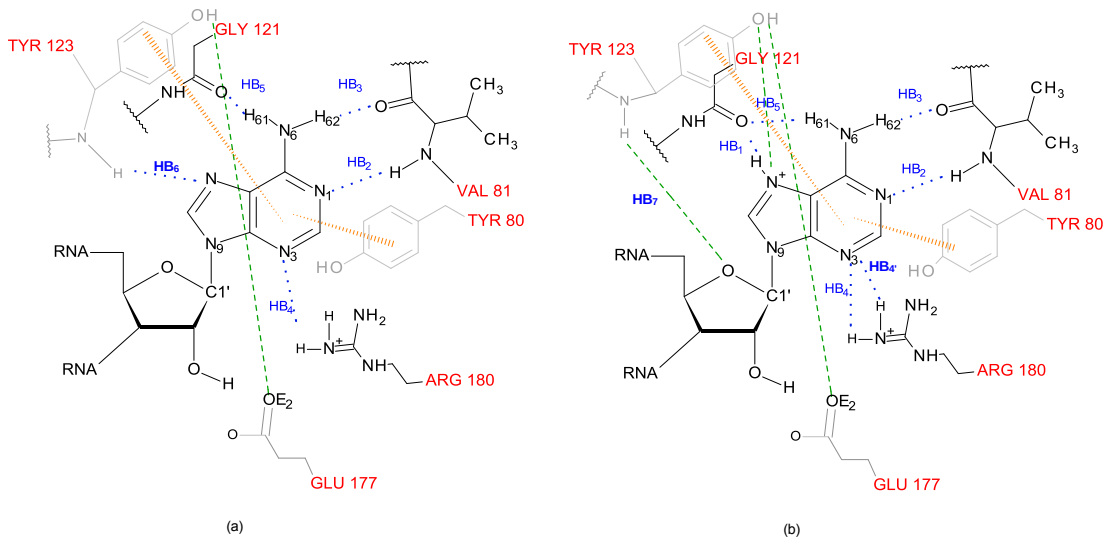


Figure 3: Binding pattern of: (a). UPA , (b). N7HA. HB_n represents the hydrogen bonds.

5.5.1. Binding of UPA and Elimination of Protonation at N3 and N1

Short distances from the residues TYR80 and TYR123 (5.04A and 4.58A respectively) to UPA suggests a Pi-Pi type of interaction between UPA and these residues. Further, trajectory analysis showed that such pi-pi association between TYR80 and UPA can vary from so called "stacked" to "T-shape" pi-interacting conformations.

Analysis of QM/MM dynamics trajectories suggested a network of hydrogen bonds (H-bonds) that consist of five H-bonds between UPA and ricin. These H-bonds are shown in Figure 3(a). The short donor [ARG180(HH12)] – acceptor [N3] distance of 1.82Å suggests a very strong H-bond interaction (HB_4) between these two atoms. In general, H-bonds with very short donor-acceptor distances are expected to show some covalent character due to the orbital overlaps and charge transfer processes[33].

Therefore, it can be assumed that the positively charged guanidinium group of ARG180 can impose an electron withdrawing effect on reacting adenine and hence facilitate the cleavage of target glycosidic bond. Moreover, it has been clearly shown[34] that these short-range low-barrier H-bonds can significantly contribute to the transition state stabilization.

Based on the observations made in this study and the previous findings (section 5.2), it is argued here that the target adenine cannot be pre-protonated at N3 position as pre-protonation at N3 will jeopardize the strong H-bonding interactions between ARG180 and N3. Moreover, if N3 is pre-protonated, there will be a strong electrostatic repulsion between the positive charge on N3 and the positively charged guanidinium group of ARG180. Therefore, pre-protonation of N3 will result in poor binding of the substrate and low reactivity.

H-Bonding interaction, HB_2 is proposed between N1 and backbone amide hydrogen of VAL81 with an average donor-acceptor distance of 2.51Å. If the substrate is pre-protonated at N1 position, it will require an H-Bond acceptor instead of a donor in the position of NH group of VAL81 for effective binding. The nearest H-bond acceptor is the backbone oxygen of VAL81; however, aligning with this acceptor is not possible mainly due to the tertiary structure of the protein. VAL 81 belongs to beta strand 6 which is in anti parallel orientation with beta strand 5 and 7. These three beta strands are not flexible as they from a stable beta sheet. Therefore, any rearrangements of VAL81 required by protonation of N1 are not allowed by the protein structure. Moreover, reorientation of the adenine moiety to interact with the backbone oxygen of VAL81 is also not possible due the restricted vertical movement of the base in the direction of C1'-N9 bond. Other than VAL81 backbone oxygen, no other potential accepter was is found in the close proximity of N1 position. Therefore, it is proposed here that N1 must not be pre-protonated.

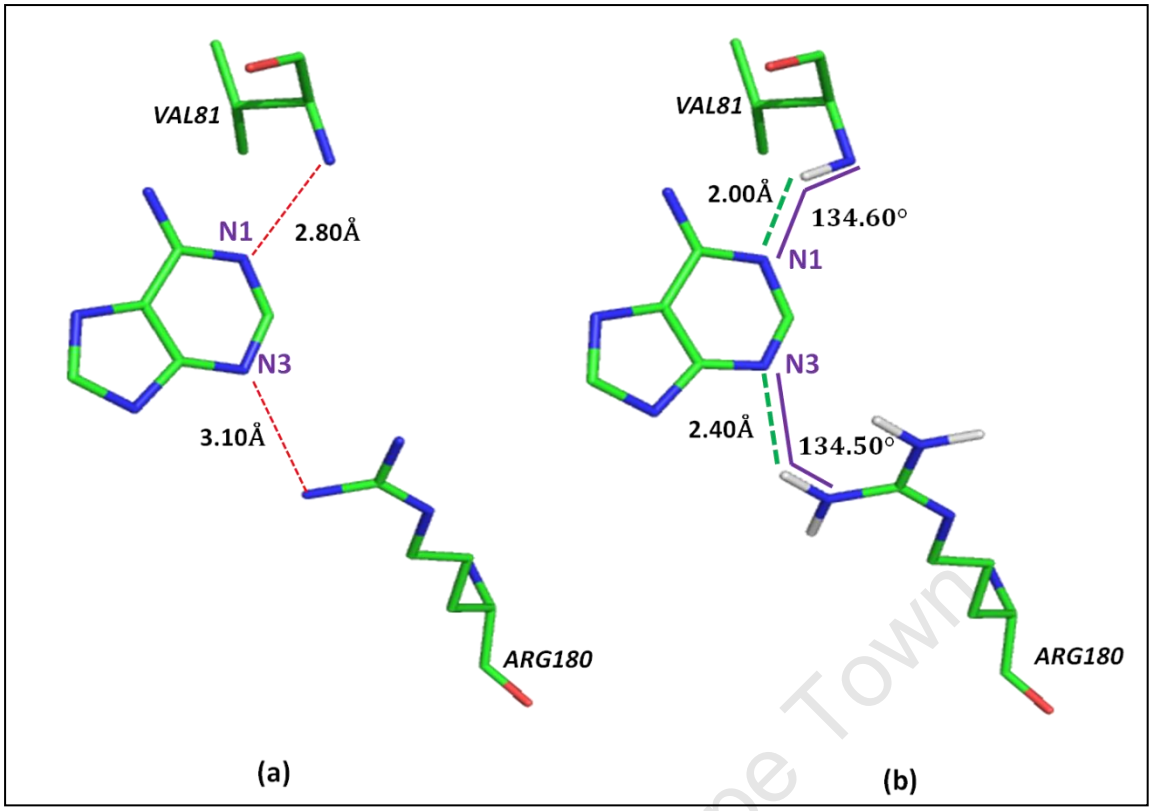


Figure 4: (a). x-ray structure of free adenine in ricin binding site. (b). x-ray structure of free adenine in ricin binding site with protonated VAL81 and ARG180. Hydrogen bonds that are deduced from the distance analysis are indicated with green dashed lines. Donor-acceptor distances are given in angstroms and H-bonding angles are given in degrees.

X-ray structure of ricin complex with adenosine[35] provides clear evidences to support the above argument of exclusion of pre-protonation of N1 and N3. In this structure, free adenine is found in the ricin binding site after the reaction. Figure 4 illustrates that the alignment of backbone amide of group VAL81 with N1 and alignment of ARG180 with N3 found in the x-ray structure cannot occur if N1 and N3 were pre-protonated. Because, under physiological conditions, both backbone amide nitrogen of VAL81 and the guanidinium group of ARG180 must be protonated. By analyzing the relevant interatomic distances (Figure 4(a) and (b)), it can be inferred that there exist strong H-bonds between N1 and VAL81 and N3 and ARG180 in the x-ray structure. These H-bonds can be clearly visualized after protonation of the protein (Figure 4(b)) and they reinforce the argument of not pre-protonating N1 and N3.

In UPA, the NH_2 group at 6^{th} position donates two hydrogen bonds, HB3 and HB5 to backbone oxygens of VAL81 and GLY121 respectively. The remaining hydrogen bond, HB6 is found between the backbone amide hydrogen of TYR123 and N7 of adenine.

As discussed above, binding pattern analysis of UPA along with the experimental evidences strongly suggested that there exists only a minimal probability (if not nil) of protonating the positions N3 and /or N1. Hence, a selection has only to be made from unprotonated adenine, (UPA) and pre-protonated adenine at N7, (N7HA). Selection of N7HA as ricin's target substrate over UPA is discussed in the rest of this chapter.

5.5.2. Binding of Pre-protonated Adenine, N7HA

N7HA follows a similar binding fashion as UPA to bind to ricin. i.e. it binds predominantly via a H-bonding network similar to that of UPA and with pi-stacking type interactions with TYR80 and TYR123. Nonetheless, some major differences can be found between the binding patterns of UPA and N7HA.

A rearrangement of ricin binding site was observed up on the binding of N7HA to optimize the H-bonding network. Movement of GLY121 contributed the most to that rearrangement. In the rearranged binding site, backbone oxygen of GLY121 forms a strong H-Bond with the hydrogen on N7; this H-bond is indicated as HB1 in Table 1 and Figure 3(b). Since it does not belong to any secondary structure, it is possible for GLY121 to rearrange. Further, it was found that the rearrangement of GLY121 was facilitated by a slight uncoiling of bend 7 in the ricin-tertiary structure, which is composed of residues PHE117 to GLY120. Moreover, this rearrangement allows the backbone oxygen of GLY121 to hydrogen bond to H61 (HB5) and to H7 simultaneously. From these evidences, it is clear that the favourable binding interactions are not penalized by the pre-protonation of N7. The rearrangement of the binding site also gives arise to a novel hydrogen bond (HB7) between the backbone amide hydrogen of TYR123 and O4' of the ribose ring. Hydrogen bonds HB2 and HB3 are suggested for N1 and H_{62} with VAL81 as discussed in UPA. As in UPA, a strong H-bonding interactions (HB4 and HB4') are suggested between N3 and ARG180.

Analysis of the dynamics trajectories showed that TYR80 and the adenine moiety of the substrate are engaged in a "stacked", face-to-face Pi interaction throughout the simulation time. The average distance between the centre of geometry of the phenyl moiety of TYR80 and the centre of geometry of the six membered ring of adenine was found to be 3.65 Å and this distance is much shorter than the corresponding distance (5.04Å) found in binding of UPA. It is argued here that the strong electrostatic interaction between the positive charge on N7 and the hydroxyl oxygen of TYR123 is responsible for the observed short distance. Average distance between hydroxyl oxygen of TYR123 and N7 was found to be 3.41 Å Furthermore; cation-pi type of interaction can also be proposed between N7HA and TYR80 due to the short distance between them.

In total, binding of N7HA involves five very effective H-bonds, one strong pi-stacking interaction and a strong electrostatic interaction whereas binding of UPA uses only four effective H-bonds. Therefore, binding of N7HA is expected to be much stronger than the binding of UPA.

5.5.3 Findings from Electronic Structure Calculations

Quantum mechanical interaction energy calculations were carried out in order to further investigate the above suggested enhanced-binding of N7HA. The interaction energy between N7HA and the the truncated binding site in the gas phase was found to be 63.12 kcal/mol favourable compared to UPA. This extremely strong interaction of N7HA with the binding site confirms the enhanced binding of N7HA that was suggested in the previous section.

Fridgen *et.al* [25] proposed that a larger population N7HA is expected be found in the cellular environment compared to its UPA counterpart. Calculated (with M06-2X/6-31G(d,p) level of theory) dipole moment of of UPA is 19.24D where as that of N7HA is 34.79D. As the solutes with larger dipoles are more stable in water, this result strongly agrees with Fridgen's argument and suggests that N7HA will be more stable in solution compared to UPA.

Moreover, because of its specific location in the GAGA tetra loop, the target adenine is overexposed to the environment (solvent) in comparison to the other adenines in the same region. Therefore, it can be assumed that this additional exposure will increase the probability of pre-protonation at N7.

5.5.4. Effect of protonation on glycosidic bond cleavage

To investigate the effect of pre-protonation at N7 on the cleavage of ricin-targeted glycosidic bond (CN-bond), one dimensional potential of mean force (PMF) calculations were carried out. The reader is referred to Chapter 2 for details on PMF calculations. Free energy profiles of glycosidic bond stretching for UPA and N7HA are presented in Figure 5. The C1'-N9 bond distance was used as the reaction coordinate in the PMF calculations.

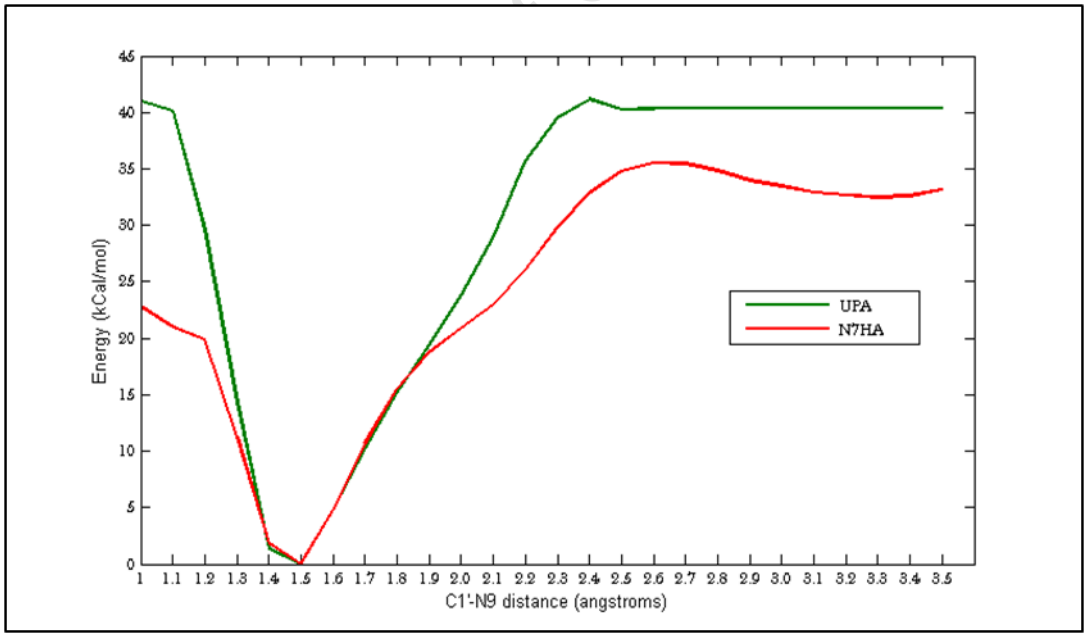


Figure 5: Free energy profiles of glycosidic bond stretching for UPA and N7HA.

In both UPA and N7HA, energy required to stretch the CN-bond increases with the distance as expected. From the corresponding equilibrium bond lengths of 1.40Å for UPA and 1.5Å to up to 1.80Å both curves follow a similar trend. However, beyond 1.80Å, UPA and N7HA show completely different free energy profiles from one another. Free energy profiles clearly show that the stretching of the CN-bond in UPA requires more energy than in N7HA. Further, it shows that stretching of C-N bond beyond 2.4Å in UPA is extremely difficult.

Maximum of the N7HA curve is located at 2.60Å with a value of 35.55 kcal/mol. In the region of from 1.80Å to the maximum, N7HA does not show the linearity as shown by the UPA curve as a consequence of the low-energy region that is located around 2.20Å. The substrate forms that are found in this region are about 9.50 kcal/mol stable than the corresponding forms in UPA. The N7HA substrate forms in the ~2Å region can be considered as "dissociated-forms" of the substrate, where the covalent existence of the glycosidic bond is minimal. Presence of these stable dissociated substrate forms in N7HA is an indication of their stabilisation by the protein binding site. In UPA, these low-energy conformations are not observed and hence it can be concluded that the corresponding dissociated UPA forms are not stabilized by the ricin binding site.

It is proposed here that the dissociation of the C-N bond is also affected by the nucleophilicity of the leaving base. The dissociated base of N7HA will have an overall neutral charge where as the corresponding UPA counterpart will have a single negative charge overall. Consequently, the negatively charged leaving base in UPA will act as an extremely powerful nucleophile (due to the negative charge) to attack the forming carbocation on C1' due to the C-N bond dissociation and will attempt to reform the cleaving bond. In N7HA, neutrally charged leaving base is a relatively very weak nucleophile and its attack on forming carbonation is not expected to be vigorous. Therefore, it is suggested here that the less nucleophilicity of the leaving base in N7HA facilitates the cleavage of C-N bond.

After the maximum, N7HA curve loses its linear-shape and becomes almost parallel to the distance (X) axis where a minimum is found at 3.3Å. The shape of N7HA curve shows that after 2.60Å, separation of the base moiety from the sugar becomes energetically less expensive. The plateau (from 2.4Å to 3.5Å) in UPA plot indicates a region with no sampling. During the total of 3ns simulation time, it was unable to sample the above region with respect to the selected coordinate. Simulation time of 3ns was selected since N7HA gave a converged PMF profile within that period of time. The last point that was sampled in the UPA curve is at 2.4Å and stretching of the CN-bond up to this distance requires 41.24 kcal/mol whereas in N7HA stretching of the C-N bond up to this limit requires only 32.87 kcal/mol. As discussed above, dissociation of the C-N bond in N7HA is energetically inexpensive than in UPA. Therefore, it is concluded here that the protonation of the target adenine at N7 has a favourable direct consequence on the C-N bond cleavage.

5.5.5. Protonation at N7 and Molecular Orbitals

Chemical reactions and the reactivity of the molecules can be understood by taking the interactions between the empty and filled orbitals of the reacting species in to account[26]. According to the Koopman's theorem[27], the electron affinity is defined as the negative value of the energy of the Lowest Unoccupied Molecular Orbital (LUMO). Therefore, to predict the nucleophillic susceptibility of a given reactive centre, analysis of LUMO energy of the reacting species can be used [28]. However, there is no direct measure of nucleophilicity, because the ranking of nucleophiles is totally disturbed when the reference electrophile is changed and vice versa[27]. Therefore, analysis of LUMO cannot be used as the sole argument for assigning the reactivity; nonetheless, it provides a reasonable initial guess about the reactivity when the same attacking species (same nucleophile) is involved.

When comparing similar chemical systems that are at the risk of nucleophillic attack by the same nucleophile, the system that has the lower LUMO energy will react first due to the smaller energy gap between the Highest Occupied Molecular orbital (HOMO) of the nucleophile and the LUMO of the target molecule which enhances overlap of orbitals.

The LUMO energies for UPA and N7HA were obtained from single point energy calculations. The starting structures were obtained as mentioned in the methods section. Calculations were carried out with MP2/6-31G(d,p) level of theory with the Gaussian 09 package in the gas phase. Energy of the LUMO of UPA is 5.1801 eV where as that of N7HA is 0.097414 eV. Therefore, N7HA LUMO is 5.08269 eV lower in energy than the LUMO of UPA. As discussed above, this indicates that the N7HA is more vulnerable to nucleophillic attack when compared to UPA.

Figure 6 shows the LUMOs of UPA and N7HA. Inspections of LUMOs of these two molecules reveal that N7HA has a "diffused" LUMO around the reactive carbon, (C1') compared to UPA. Therefore it provides an additional preference (other than being low in energy) for the incoming nucleophile to choose N7HA over UPA. i.e. the LUMO of N7HA will be more accessible to the incoming nucleophile.

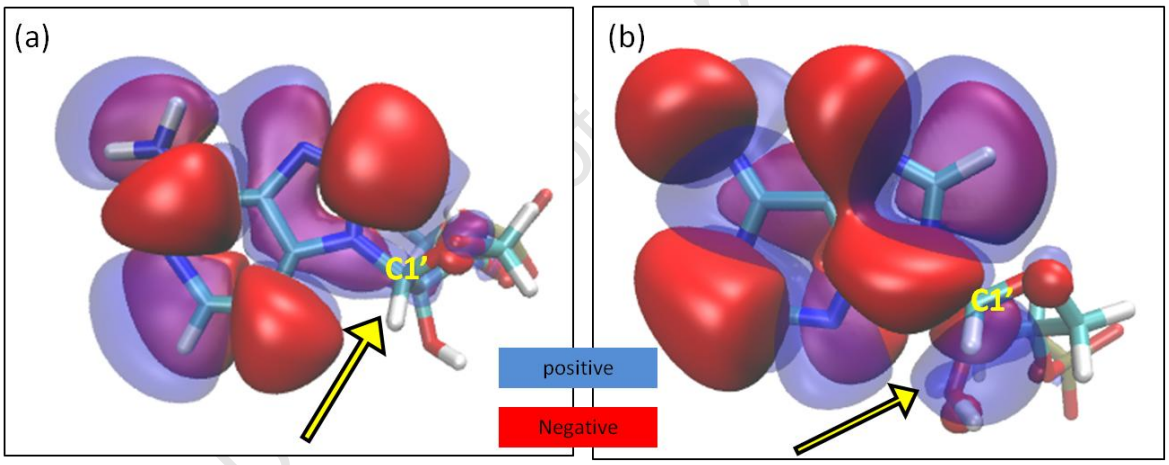


Figure 6: Illustration of LUMO of UPA (a) and N7HA (b). Orbitals are plotted at iso value of 0.008 eV. Arrow heads indicate the diffused LUMO of N7HA around C1' and its non-diffused counterpart of UPA. Colour codes refer to the phases of the orbitals.

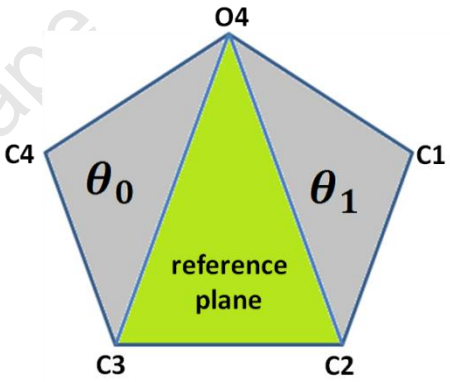
This study has clearly shown that pre-protonating at N7 position has a direct effect on the energy and the shape of the LUMO of target adenine. As discussed above, these effects strongly suggest an increased reactivity at C1' in N7HA compared to that in UPA. Therefore, it is argued argued here that the N7HA is a better substrate candidate for the ricin-catalysed reaction.

5.5.6. Effect of N7 Protonation on Ribose Ring Puckering

It was found in this study that the ribose in UPA and N7HA show distinguishable puckering conformation when inside the ricin binding site. It is proposed here that the H-bonding interaction between O4' and TYR123(HN) is the cause for specific puckering observed in N7HA. As discussed in the binding of N7HA, this H-bonding interaction is a result of the rearrangement of the ricin binding site to accommodate N7HA. Whereas in UPA, this specific H-bond is absent and hence it shows a different pucker conformations.

Pucker analysis were performed on the QM/MM dynamics trajectories of UPA and N7HA. Scheme 1 presents the IUPAC pucker nomenclature references that was used in this study.

Canonical conformation	θ_0	θ_1
$^{O}T_1$	-13.10	-33.89
T_4	-42.16	13.21
$^{2}T_3$	34.50	-34.50
$^{1}T_2$	-13.21	42.16
$^{4}T_0$	33.89	13.11
$^{1}T_0$	13.10	33.89
$^{4}T_3$	42.16	-13.21
$^{3}T_2$	-34.50	34.50
$^{2}T_1$	13.21	-42.16
$^{0}T_4$	-33.89	-13.11
E_2	-24.88	40.00
E_1	0.00	-39.90
E_0	24.50	24.50
E_4	-39.50	0.00
E_3	39.50	-24.90
^{2}E	24.88	-40.00
^{1}E	0.00	39.90
^{0}E	-24.50	-24.50
^{4}E	39.50	0.00
^{3}E	-39.50	24.90
P	0.00	0.00



Scheme 1: Canonical pucker conformations for five membered rings. Figure shows the reference plane and the rotatable planes for the ribose ring. Table presents the triangular decomposition of pucker coordinates in degrees that correspond to ideal IUPAC canonical conformations.

The results of the pucker analyses are presented in Figure 7. It was found that UPA prefers to be in the $^{3}T_{4}$ conformer whereas N7HA prefers to be in the $^{0}T_{4}$ conformer. In UPA , $^{3}T_{4}$ was found to be 72% of the total counted conformations and $^{0}T_{4}$ was found to be only 1%. In N7HA, 69% was in $^{0}T_{4}$ where as $^{3}T_{4}$ was only 1%. In both UPA and N7HA about another 25% (21% and 26% respectively) was found to be in $^{4}E_{x}$ conformation. Therefore, it can be concluded that both UPA and N7HA have a similar probability around 25% to be in $^{4}E_{x}$ conformation. These ring conformers cannot however be compared with the proposed near-transition-state ring conformers, wherein O4', C1' and C2' are expected to be on the same plane (to enhance the formation of planer oxocarbenium cation) as the simulations carried out here are not meant to simulate the reaction.

However, from these results, it can be predicted that the ribose of N7HA with $^{0}T_{4}$ conformer to be a better candidate for nucleophillic attack. Because, this specific $^{0}T_{4}$ conformer dramatically reduces the steric hindrance in the region below the plane of the ribose ring. Further, due to this specific pucker conformation, the ring oxygen will have minimal interference with the incoming nucleophile. Here it is proposed that $^{0}T_{4}$ pucker conformer of the ribose ring is a prerequisite for nucleophillic attack on C1' from the bellow the plane of ribose ring. It is presumed that when the nucleophile becomes close, < 2 Å to C1' , the puckering of the sugar ring will dramatically change from the $^{0}T_{4}$ towards so called O4', C1' ,C2'-coplanar conformer. Figure 7(e) and 7(f) illustrate the relationship between the available space for the incoming nucleophile (around C1') and the pucker conformation of the ribose ring.

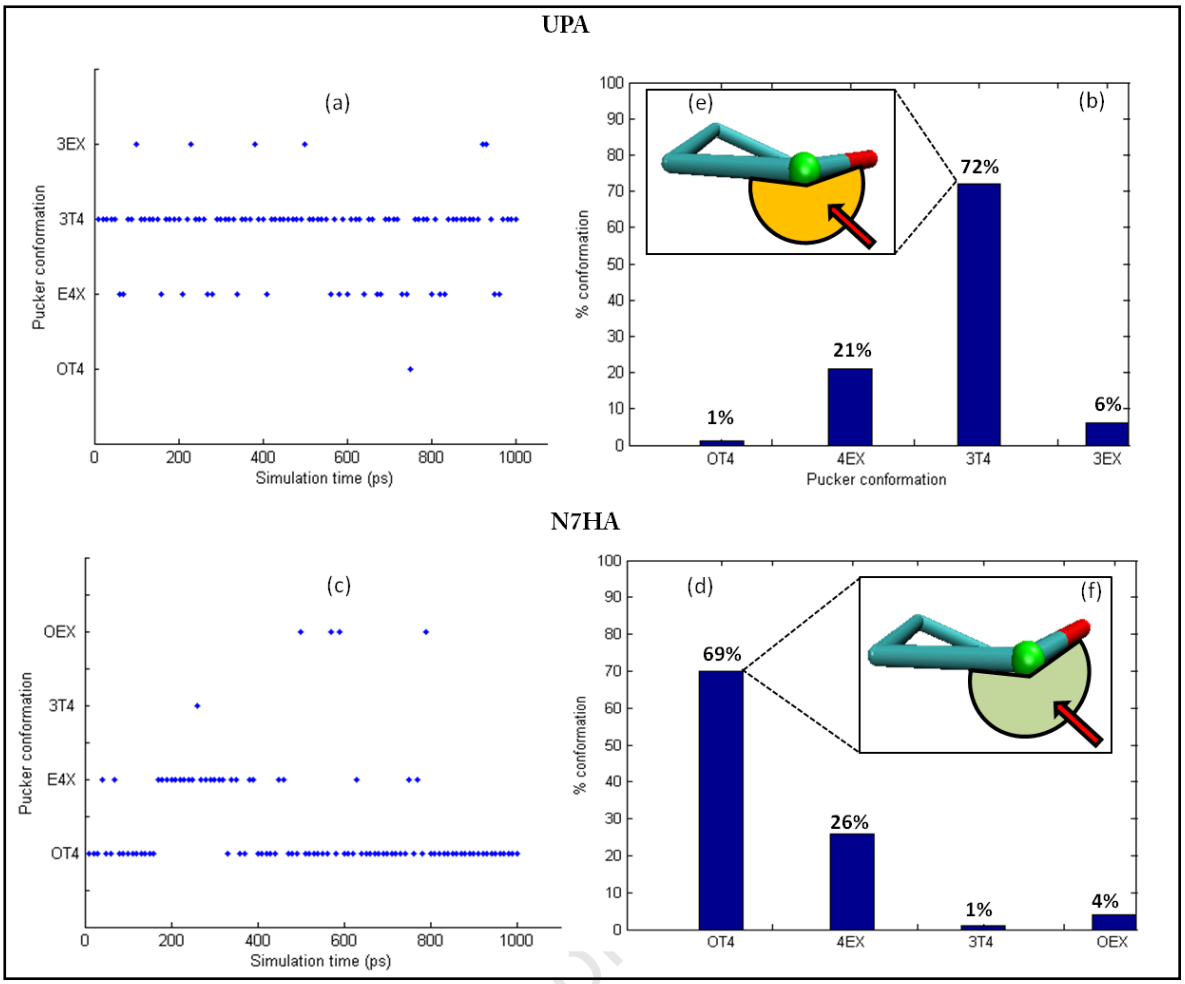


Figure 7: Preferred ribose pucker conformations in UPA and N7HA. Frames (a) and (c) present the pucker time series. Histograms are presented in frames (b) and (d). (e) $^{3}T_{4}$ conformer in UPA has more steric hindrance, less available space for the incoming nucleophile (red arrow) as indicated by the orange shaded area. (f). $^{O}T_{4}$ conformer in N7HA has less steric hindrance , more available space for the incoming nucleophile as indicated by the green shaded area.

5.6. Concluding Remarks

It is argued here that the possibilities of pre-protonating target adenine at N3 and/or N1 positions can be pre-empt as these protonation states will jeopardize the initial binding of the substrate to the ricin binding site. Based on the findings of this study, it is proposed that the pre-protonated adenine at N7 position is a better candidate by far compared to the unprotonated adenine. However, the mechanism of pre-protonating the target adenine has still to be investigated.

Quantum mechanical interaction energy calculations showed that N7-protonated adenine (N7HA) binds much stronger to ricin compared to the unprotonated adenine(UPA). Further, Protonation of adenine at N7 has a direct positive influence on the bond dissociation as shown by the PMF calculations done in this study. Analysis of the molecular orbitals of these two species (UPA and N7HA) showed that N7 protonated adenine is more vulnerable to nucleophillic attack than the unprotonated adenine. Moreover, it was found that N7-protonated adenine has more favourable ribose pucker conformer for nucleophillic attack than the unprotonated adenine.

Considering all these factors, it is concluded here that the N7HA (pre-protonated adenine at N7 position) is the most appropriate substrate candidate for the reaction catalysed by ricin. However, as proposed by many previous studies there exists the possibilities of protonating the other sites (N1, N3 or N9) during or after the reaction. These possibilities are still need to be investigated. Conclusively, this study clearly answered a long-standing question about the pre-protonation state of the target adenine in the depurination reaction catalysed by ricin.

5.7. References

(1) Voet, D.; Voet, J. G.; W, P. In *Fundamentals of Biochemistry*; John Wiley & Sons,Inc: 1999, p 663.
(2) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I. *Nucleic acids: structures, properties, and functions*; University Science Books, 2000.
(3) Blackburn, G. M.; Royal Society of Chemistry, U. h. b. g. l. b. i. h. d. l. *Nucleic acids in chemistry and biology*; RSC Pub., 2006.
(4) Ravindranathan, S.; Butcher, S. E.; Feigon, J. *Biochemistry* **2000**, *39*, 16026.
(5) Rajabi, K.; Theel, K.; Gillis, E. a. L.; Beran, G.; Fridgen, T. D. *The Journal of Physical Chemistry. A* **2009**, *113*, 8099.
(6) Zoltewicz, J. a.; Clark, D. F.; Sharpless, T. W.; Grahe, G. *Journal of the American Chemical Society* **1970**, *92*, 1741.
(7) Garrett, E. R.; Mehta, P. J. *Journal of the American Chemical Society* **1972**, *94*, 8542.
(8) Versées, W.; Loverix, S.; Vandemeulebroucke, A.; Geerlings, P.; Steyaert, J. *Journal of molecular biology* **2004**, *338*, 1.
(9) T Major, D.; Laxer, A.; Fischer, B. *The Journal of Organic Chemistry* **2002**, *67*, 790.
(10) Walker, G. a.; Bhatia, S. C.; Hall, J. H. *Journal of the American Chemical Society* **1987**, *109*, 7629.
(11) Rios-Font, R.; Rodríguez-Santiago, L.; Bertran, J.; Sodupe, M. *The Journal of Physical Chemistry. B* **2007**, *111*, 6071.
(12) Touboul, D.; Bouchoux, G.; Zenobi, R. *The Journal of Physical Chemistry. B* **2008**, *112*, 11716.
(13) Roday, S.; Amukele, T.; Evans, G. B.; Tyler, P. C.; Furneaux, R. H.; Schramm, V. L. *Biochemistry* **2004**, *43*, 4923.
(14) Monzingo, A. F.; Robertus, J. D. *Journal of Molecular Biology* **1992**, *227*, 1136.
(15) Frankel, A.; Welsh, P.; Richardson, J.; Robertus, J. D. *Molecular and cellular biology* **1990**, *10*, 6257.
(16) Bene, J. E. D. *Journal of American Chemical Society* **1983**, *1673*, 367.
(17) Chen, X. Y.; Berti, P. J.; Schramm, V. L. *Journal of the American Chemical Society* **2000**, *122*, 1609.
(18) Huang, Q.; Liu, H.; Tang, Y.; Jin, S.; Wang.Y *Biochemistry* **1995**, *309*, 285.
(19) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H. M.; Pardi, A. *Journal of Molecular Biology* **1996**, *264*, 968.
(20) Monzingo, A. F.; Robertus, J. D. *Journal of Molecular Biology* **1992**, *227*, 1136.
(21) Vriend, G. *Journal of Molecular Graphics* **1990**, *8*, 52.
(22) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187.
(23) Mackerell, A. D.; Banavali, N. K. *Journal of Computational Chemistry* **2000**, *21*, 105.
(24) Foloppe, N.; MacKerell, J. A. D. *Journal of Computational Chemistry* **2000**, *21*, 86.
(25) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926.
(26) Brooks, C. L.; Brünger, A.; Karplus, M. *Biopolymers* **1985**, *24*, 843.
(27) Ryckaert, J.; Ciccotti, G.; Berendsen, H. *Journal of Computational Physics* **1977**, *23*, 327.
(28) Field, M. J.; Bash, P. A.; Karplus, M. *Journal of Computational Chemistry* **1990**, *11*, 700.
(29) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *The Journal of Physical Chemistry B* **2001**, *105*, 569.

(30) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision A.1*; Gaussian, Inc.: Wallingford CT, 2009.
(31) Zhao, Y.; Truhlar, D. G. *Theoretical Chemistry Accounts* **2007**, *120*, 215.
(32) Zhao, Y.; Truhlar, D. G. *Accounts of chemical research* **2008**, *41*, 157.
(33) Scheiner, S. *Hydrogen Bonding: A Theoretical Perspective*; Oxford Univ. Press: Oxford, 1997.
(34) Cleland, W.; Kreevoy, M. *Science* **1994**, *264*, 1887.
(35) Weston, S. A.; Tucker, A. D.; Thatcher, D. R.; Derbyshire, D. J.; Pauptit, R. A. *Journal of Molecular Biology* **1994**, *244*, 410.

CHAPTER -6

THE MECHANISM OF RICIN-CATALYSED DEPURINATION REACTION

6.1. Introduction

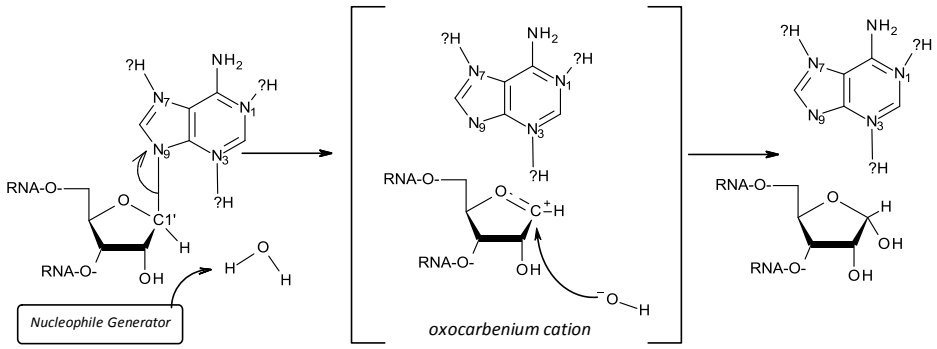
Identifying effective inhibitors for ricin remains a major challenge due to the poor understanding of its mechanism of action[1,2]. Modelling the catalytic activity of ricin holds the most promise for the designing and development of novel ricin inhibitors. Understanding the mechanism of the enzyme activity is useful on several counts but here I mention two important ones. Firstly, knowing the mechanism of action leads to the identification and ranking of, reaction rate affecting, catalytic residues in the binding site. The reaction can then be tuned from that of total inhibition to maximum enhancement. Secondly, knowledge of the molecular reaction mechanism provides insight into the structural details of the transition state. This can lead to the design and development of Transition State Analogue (TSA) inhibitors. Enzymes show extremely high affinity to the Transition States (TS)/intermediates of the reactions that they catalyse[3-5]. Therefore, inert (with less or no reactivity) molecules that mimic the chemical, electronic and geometrical features of the TS are considered as highly effective inhibitors. Efficient TSA's are a critical to the development of designer pharmaceuticals as will be discussed in the following chapter. The structural, conformational and electronic character of the transition state is not visible to the enzymologist from his/her experimental arsenal (i.e., spectroscopic, diffraction or kinetics experiments). This is because the TS exists for a brief (few femtoseconds) moments along the reaction trajectory. This is where computer simulations come into their own providing comprehensive atomic-level insight into TS.

Dynamic QM/MM reaction dynamics simulations as described in Chapter 3, besides revealing the mechanism, provides a trove of information about the catalytic residues, specific interactions that stabilise the TS and the role played by surrounding solvent (water). Despite gaining the initial coordinates from crystal structures enzyme simulations can provide answers to the questions that are often impossible to address experimentally.

In this chapter I present details and results of the chemical reaction catalysed by ricin as found from the FEARCF computational reaction dynamics method. Mechanism of action, reaction free energy surface, transition state properties and the key catalytic residues of ricin-catalysed depurination reaction are shown and discussed here.

6.2. Overview of Previously Proposed Mechanisms

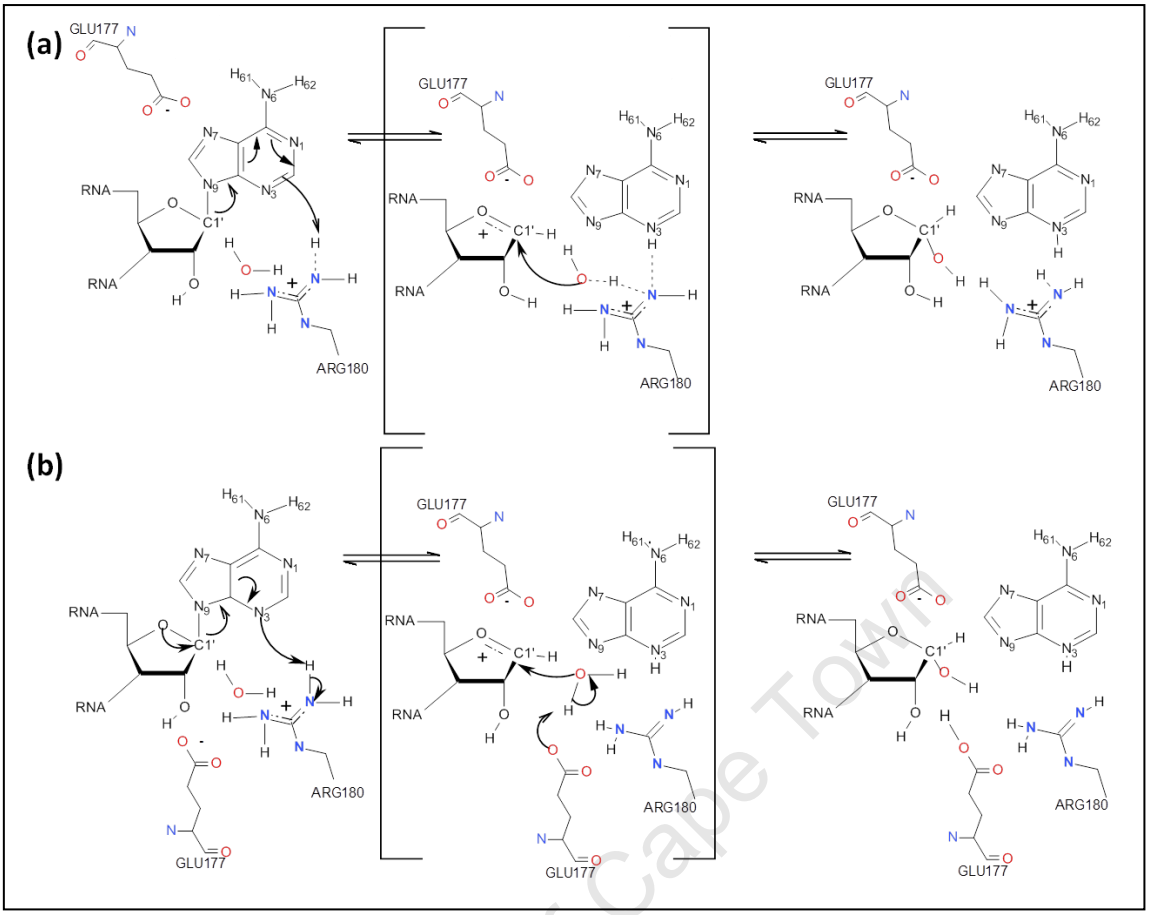
Although a complete understanding of the mechanism action of the reaction catalysed by ricin has not been gained hitherto, a few plausible mechanisms have been proposed. All of those proposed mechanisms share some common mechanistic features. Such features can be used to draw a "generalised" reaction mechanism, which is presented in Scheme 1. The common features that are found in previously proposed mechanisms are: (a). The participation of a water molecule in the mechanism, (b). The activation of the water molecule generating an OH^- nucleophile, through its deprotonation by a catalytic domain amino acid residue (labelled "*nucleophile generator*" in Scheme 1), (c). The protonation of the leaving base, (d). Formation of the oxocarbenium ion on the sugar moiety resembling the TS.



Scheme 1: A generalised reaction mechanism for ricin-catalysed depurination reaction.

Previously it was suggested that one or more ring nitrogens (N1,N3,N7) should be protonated to promote the cleavage of the C1'-N9 bond[6-8]. However, the exact protonation state of the leaving base remained uncertain until this study determined (in chapter 5) that it must be protonated at the N7 position. In the generalised mechanism, protonation sites are shown with "?H" to emphasise the uncertainty of their protonation state. There is convincing experimental and theoretical evidence[1,6,9,10] supporting the presence of an oxocarbenium ion-like TS in the ricin-catalysed reaction. However, as I have previously stated the limitations in the experimental methods prevented the precise structure of the TS from being discovered. A major shortfall of previous computational modelling studies[6] is the unrealistic nature of the model. The target adenine was modelled in the gas phase ignoring the enzyme and solvent making the model a poor facsimile of the true TS as observed in the enzyme catalytic domain.

There are two widely accepted mechanisms that share common mechanistic features as laid out above, which I now very briefly present. In the first proposed mechanism[11], *mech-1* (Scheme 2(a)), the N3 atom of the base is protonated by ARG180 to enhance the cleavage of C1'-N9 bond. The resulting oxocarbenium ion, i.e. the TS is stabilised by its ion pairing interaction with GLU177. The incoming water molecule is activated by ARG180, which abstracts a proton producing the OH^- nucleophile that then attacks the oxocarbenium ion to complete the reaction. This mechanism was proposed based on site-directed mutagenesis experiments[12] and the crystal structures of substrate analogues bound to ricin[13].



Scheme 2: Proposed Mechanisms for the ricin-catalysed depurination reaction.(a). ***mech-1: ARG180 activates the catalytic water and protonates the base at N3.*** **(b).** ***mech-2: ARG180 protonates the base at N3 and GLU177 activates the catalytic water molecule.***

In the second proposed mechanism[14], *mech-2* (Scheme 2(b)), protonation state of the base and the source of proton are identical to that of *mech-1*. i.e. the base is protonated by ARG180 at N3. In *mech-1* , the oxocarbenium ion is stabilised by forming an ion pair with GLU177. However, in *mech-2*, GLU177 acts as the *nucleophile generator* activating the incoming water by abstracting a proton from the catalytic water resulting in the formation of glutamic acid and a hydroxyl ion. The reaction is completed when the attacks C1'.

Evidence from Kinetic Isotope Effects (KIE) experiments [15] suggests that the GLU177 is positioned to activate the water rather than playing the role of stabilising the oxocarbenium cationic intermediate. These mechanisms are speculative made up from scattered, mutagenesis, KIE and diffraction experimental results of which the conclusions compete with each other.

Collectively they provide invaluable clues to the composition of the catalytic domain, the protonation state of the base, the nature of the transition state and plausible amino acid residues that could act as the nucleophile generators.

6.3. Identification of the Important Binding and Catalytic Residues

Numerous x-ray crystallographic experiments[16-22] on substrate analogues, small-molecule inhibitors and transition state analogues have identified a number of the key amino acid residues in the ricin catalytic domain. These are shown in Figure 1.(a). in a 3D arrangement within the catalytic domain. The exact roles of these residues are not completely understood. However, to contextualise the computational reaction dynamics results presented in this chapter I summarised the current understanding of the action of TYR80, TYR123, ARG180, GLU177, GLY121 and VAL81 in the ricin catalytic domain.

π Stacking: It has been proposed that TYR80 and TYR123 play a crucial role in binding the natural substrate to the ricin catalytic domain. Several x-ray crystallographic experiments of inhibitors[23], substrate analogues[13] and transition state analogues[15] have suggested that the target adenine will bind in a "sandwiched" fashion between TYR80 and TYR123. The proposal is that there are strong π-stacking interactions between TYR80, adenine and TYR123 that hold the triad together and fix the position of the reacting substrate for activation and reaction with the nucleophile, the hydroxyl ion. This hypothesis came about through an crystallographic analysis of the substrate analogue formycin monophosphate, FMP on binding to ricin. The "sandwiched-binding" of FMP is shown in Figure 1.(b). Furthermore, the crystal structure[24] of adenine-ricin complex (where adenosine was converted in to free adenine after the reaction) also provides strong evidence for the suggested "sandwiched-binding" of the natural substrate.

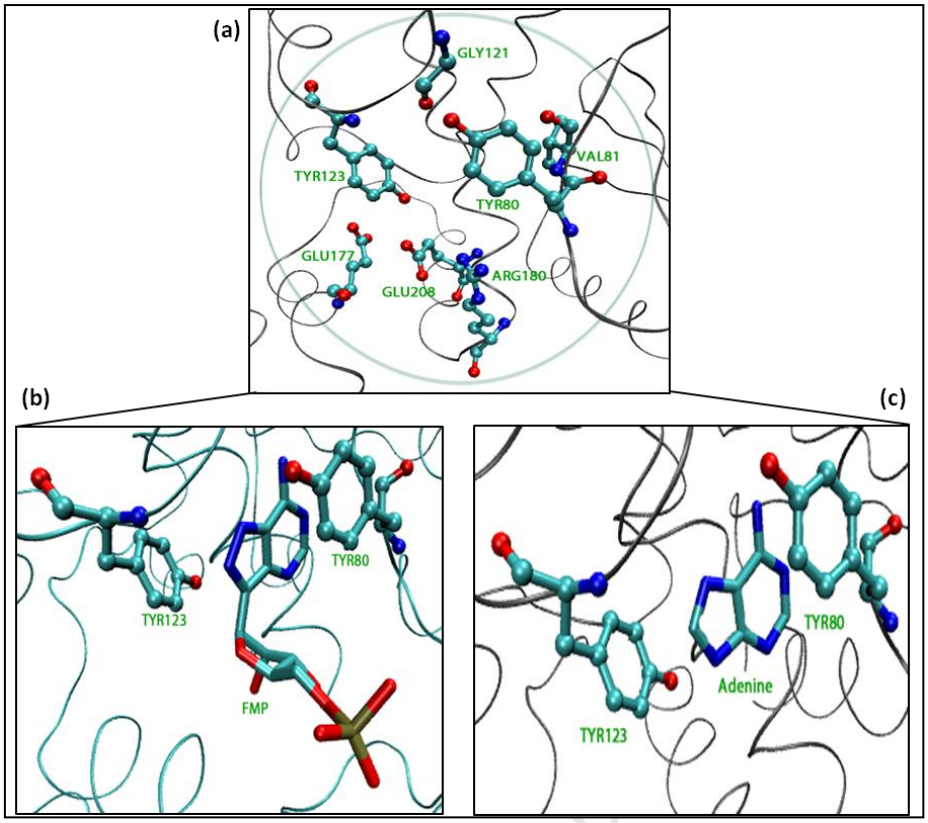


Figure 1: (a). 3D catalytic domain of ricin. Key amino acids are TYR80, VAL81 , GLY121, TYR123, GLU177, ARG180 and GLU208. (b). Binding of substrate analogue FMP. Base moiety of FMP is pi-stacked between TYR80 and TYR123. (c). Binding of free adenine to ricin. Base is pi-stacked between TYR80 and TYR123.

A recent study[25] has proposed an alternative mechanism to general acid catalysis of nucleobase hydrolysis, wherein it was shown that leaving group activation is facilitated by pi-stacking interactions between the catalytic domain residues and the heterocyclic aromatic nucleobases. It was proposed that these strong pi-pi interactions elevate the pKa values of the ring nitrogens of leaving base; wherein protonation of the leaving base is done by using water as the proton donor in the absence of an acidic amino acid residue. As a consequence of protonation of the leaving base, cleavage of C1'-N9 bond is expected to be energetically less demanding and efficient. Even though, protonation of the leaving base has been proposed[6,26,27], none of the catalytic domain residues have been identified as the proton source. This is mainly due to the absence of suitable acidic amino acid residues in the proximity of protonation sites of the base. Therefore, the phenomenon of pKa elevation of the target adenine by aromatic pi-stacking with TYR80 and TYR123 can be expected to be true for ricin.

Hydrogen bonding and base protonation: ARG180 has also been proposed by various studies as a key residue in the ricin catalytic domain. While several roles have been proposed for ARG180 the most widely accepted one is that of a hydrogen bond donor to the N3 position of adenine[11]. that the argument is that the strong H-bond character between N3 and ARG180 may facilitate complete or partial protonation of N3 atom in the course of the catalytic cycle[20]. Therefore, it is assumed that the developing oxocarbenium-ion-like transition state is indirectly stabilised by enhanced departure of the leaving adenine due to the protonation of the N3 atom. However, it has been shown that mutation of this residue reduces the k_{cat} by only 500 fold[21].

Therefore, some authors have argued that the effect of ARG180 is substantial but it is not essential for the catalytic mechanism. It has been argued that the protonation of N3 by ARG180 will diminish as it forms an ion pair with GLU177 at physiological pH[28]. Despite these discrepancies about the role of ARG180, catalytic mechanisms have been proposed for ricin based on protonation of N3 by ARG180 (see proposed mechanism 1).

Other than the direct effects on the catalytic mechanism, mutation studies have suggested that ARG180 also plays a role in determining folding of the protein and its solubility[29]. This effect on the tertiary structure is assumed to be mainly due to its location (on an alpha helix) and the positive charge on the guanidinium group. In order for the charged guanidinium group to access the solvent, it is assumed that a special folding pattern is implied on the tertiary structure. However, the exact role of ARG180 still remains unclear.

Several studies[15,19,23] have identified GLY121 and VAL81 as binding residues. According to those findings, the backbone amide hydrogen of VAL81 is expected to donate a strong hydrogen bond to the N1 position of the base, where as the backbone oxygen of GLY121 is expected to hydrogen bond with NH2 hydrogens at 6th position of adenine. It is assumed that these strong H-bonds play a role not only in the initial binding of the natural substrate, but also in stabilising the leaving adenine group in the reaction.

Oxocarbenium ion stabilisation: Mutagenesis and kinetic studies of GLU177 have suggested that this residue plays a predominant role in the rate-limiting step of the enzymatic mechanism[30]. Moreover, It has also been proposed that GLU177 plays a major role in stabilising the proposed oxocarbenium cation-like transition state[20]. In addition, some studies have suggested that GLU177 together with ARG180 coordinate the catalytic water[29]. Therefore, GLU177 can be considered to act as a base in the catalytic mechanism. This proposal is strongly supported by the crystallographic studies of guessed TSAs bound to ricin[15].

6.4. Computational Details

In chapter 4, it was shown that the substrate-containing RNA loop must at least consists of 12 bases in total to maintain the integrity of its stem-loop structure. However, these coordinates were not readily available therefore conformational NMR studies of GNRA tetraloops[31] was used to derive a starting structure of the target GAGA tetra loop. The crystallographic coordinates of the substrate analogue formycin monophosphate (FMP) (PDB ID: 1FMP)[13] was used to generate initial coordinates for ricin. The protonation states of the amino acid residues were determined using the WHATIF[32] program. The 12-mer RNA loop was then manually docked into the ricin catalytic domain. The target adenine was oriented in the catalytic domain using the orientation of FMP as a guide. Following this the substrate-ricin system was subjected to energy minimisation with the Adopted Basis Newton–Rhaphson (ABNR) method. The minimised system was gradually heated from 50K to 300 K in 100ps. To obtain an equilibrated Michaelis complex, a 10 ns long classical MD simulation was carried out with the heated system. The classical simulations were performed using CHARMM program with the all-atom CHARMM27 nucleic acid force field[33,34]. Simulations were carried out at pH 7.00 with TIP3P[35] water model using stochastic boundary conditions[36].

To solvate the substrate-enzyme complex, a water sphere of 23.5Å radius was used with a buffer boundary placed at 20.0 Å. The SHAKE[37] algorithm was used to fix the protons during the dynamics. The target adenine was pre-protonated at N7 position using the protocol as laid out in chapter 4. After 1ns of QM/MM dynamics, a snapshot of the system was taken as the starting structure for reaction simulations. All reaction simulations were performed using the FEARCF method[38-40] (see chapter 3 for details on FEARCF).

In QM/MM simulations, the QM region was composed of reaction moieties of TYR80, TYR123, GLU177, ARG180, target adenosine (including the 3' phosphate group) and a catalytic water molecule. The rationale for this selection of amino acid residues is discussed in section 6.3. The QM/MM boundary was treated with the link-atom approach[41] as implemented in CHARMM. Simulation setup for the reaction dynamics was identical to the setup for classical MD simulations as discussed before. The quantum mechanical region was treated with the SCC-DFTB level of theory[42] where as MM region was treated with the CHARMM force field.

The QM region and the Reaction Coordinates

Previously compelling evidence supporting the critical value of TYR80, TYR123, ARG180 and GLU177 in the catalytic reaction was presented. Therefore these amino acids were included in the reactive quantum mechanical region. It has been shown that larger QM regions yield more accurate results in QM/MM simulations[43]. However, in this study, only the minimum reactive moieties of the above mentioned amino acids were treated quantum mechanically to access practical simulation times.

The catalytic water molecule (WAT) was identified from the preparatory equilibrium 10ns classical MD simulation of the ricin bound 12-mer RNA loop. This water molecule is located on the opposite side of the sugar ring to the adenine moiety (Figure 2.(b)). The orientation of this water and analysis of its distance to GLU177 and C1' clearly showed its stay between C1' and GLU177 during the pre-QM/MM dynamics simulation. From WAT, the average distances were 4.00Å and 3.25Å to GLU177 and C1'.

Moreover, WAT was found to be in a very favourable orientation for activation by GLU177 and for the attack on C1'. This orientation of WAT agrees well with previous[11] x-ray crystallographic findings .

Previous studies[15] done on transition state analogue inhibitors support the above proposal for the orientation of catalytic water. Therefore, in the subsequent QM/MM simulations and in the reaction simulations, this water molecule was included in the QM region and was treated as the catalytic water. The QM region of the ricin-substrate complex is presented in Figure 2(a).

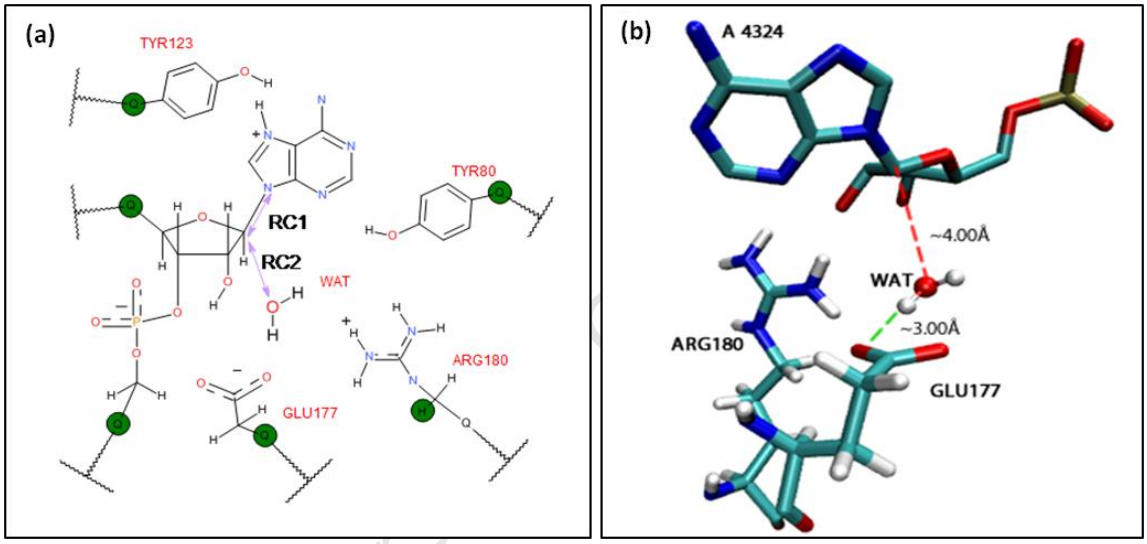


Figure 2. (a).The quantum mechanical region and selected reaction coordinates. ***"Q" represents the link atoms. Purple arrows represent the selected reaction coordinates, RC1 and RC2. (b). Orientation of the catalytic water molecule (WAT).***

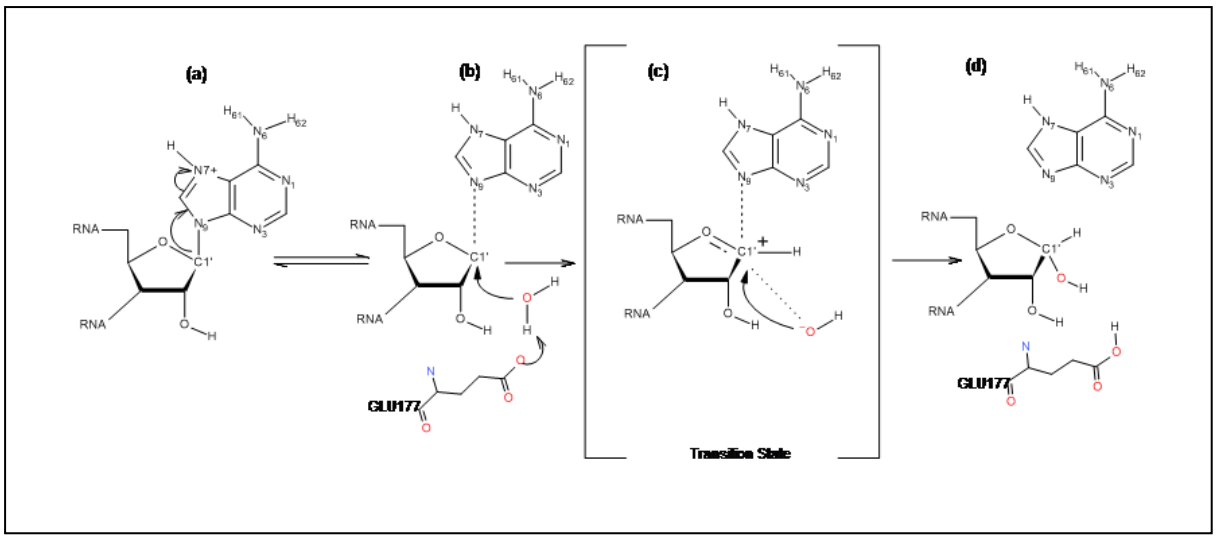
Selection of the reaction coordinates is critical step in the process of simulating chemical reactions. Even though the FEARCF method allows the selection of multiple reaction coordinates that produces a multidimensional reaction volume[38,39], choosing more than three reaction coordinates becomes problematic to visualise and the results complex to interpret. With a choice of three reaction coordinates requires a four-dimensional-representation for $W(\xi_1,\xi_2,\xi_3)$ is required where one of the dimensions is used to represent the free energy. A selection of two reaction coordinates allows a representation that chemists are familiar with.

However, reducing the number of reaction coordinates imposes a limit on the number of events/chemical steps that one can investigate in a single simulation. Therefore, this study was confined to the rate limiting steps in the reaction catalysed by ricin. These steps are a) the cleavage of the susceptible C1'-N9 bond and b) the formation of the bond between C1' and incoming OH^-. Consequently, the distance between C1' and N9 was chosen as the first reaction coordinate (RC1) and the distance between C1' and the oxygen of incoming OH^- was chosen as the second reaction coordinate (RC2). Deprotonation of the catalytic water molecule to generate the attacking nucleophile was estimated to occur without assistance of the additional biasing forces. The selected reaction coordinates are shown in Figure 2(a).

6.5. Results and Discussion

6.5.1. *The Reaction Mechanism*

The mechanism of the ricin-catalysed RNA depurination reaction discovered in this study is presented in Scheme 3. There are four distinguishable steps in this mechanism. They are: (a). the reversible cleavage of the C1'-N9 bond, (b). the activation of the catalytic water molecule and generation of the OH^- nucleophile, (c). the formation of the cationic TS, (d). Permanent departure of the adenine base, attack by the OH^- at C1' and formation of the final products. These important steps of the reaction mechanism are discussed below.



Scheme 3. Mechanism of the ricin-catalysed adenine hydrolysis. (a). eversible cleavage of the C1'-N9 bond (b). Activation of the catalytic water molecule and generation of the OH^- nucleophile, (c).The transition state (d) Final products.

Reversible cleavage of the C1'-N9 bond

Analysis of the C1'-N9' bond length clearly showed a reversible fashion in its cleavage. In the C1'-N9 bond distance time series plot (Figure 3.(a)) the red dashed vertical line shows the point in simulation time at which the C1'-OH bond is just formed, whereas the solid horizontal purple line represents the experimentally determined [31] C1'-N9 equilibrium bond distance of 1.47Å. The sudden increase of C1'-N9 bond after 676.50 ps is due to the permanent departure of the adenine base upon the completion of the reaction. After 50ps the C1'-N9 bond starts to vibrate at a higher amplitude resulting in large deviations from the equilibrium value. If the distance between C1' and N9 is greater than 2 Å, it is considered as the corresponding atoms are not covalently bonded.

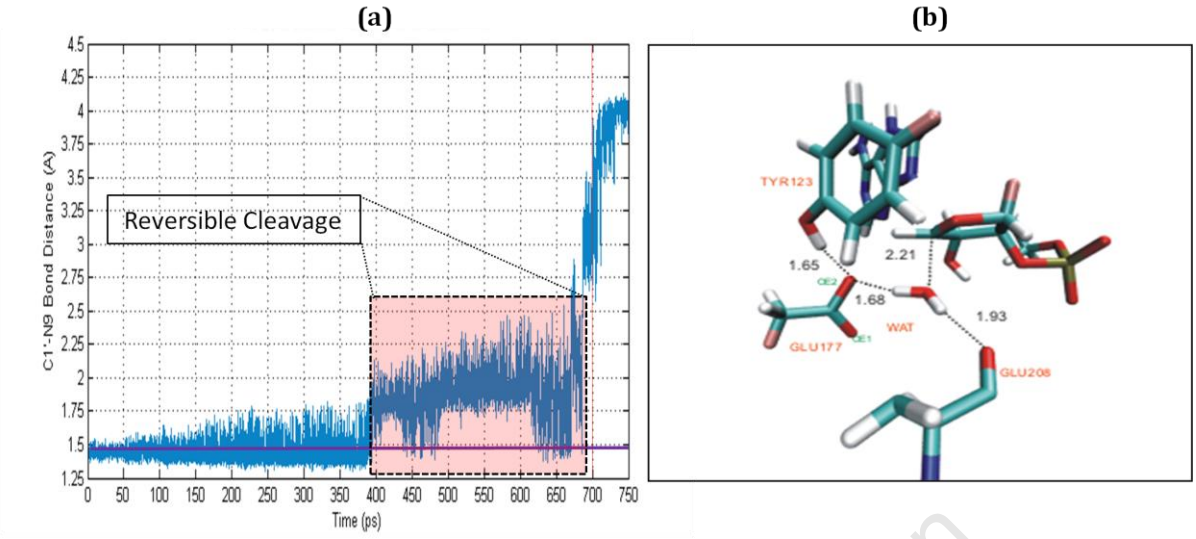


Figure 3. (a). Reversible Cleavage of the susceptible C1'-N9 bond. The time series is discontinued after 750ps for clarity. Green solid line represents the C1'-N9 equilibrium bond distance. Point at which the new C1'-OH bond is formed is represented by the dashed vertical red line. (b). Activation of the catalytic water (WAT).

It is argued here that the apparent reversible cleavage of the C1'-N9 i.e. the reformation of the breaking bond during the period of 380-680ps is due to the nucleophillic attack by the leaving base on the forming carbocation. The increased electron density on N9, that results from the C1'-N9 cleavage enhances the nucleophilicity of the leaving base. Therefore, it can be presumed that this phenomenon of reversible C1'-N9 cleavage may cause delays in the departure of the leaving group. Consequently it may impose a penalty on the overall rate of the reaction. The discovered reversible cleavage pattern of the C1'-N9 bond is in complete agreement with the evidence provided by previous studies[2] that suggested the possibility of such a cleavage pattern.

Activation of the water molecule:

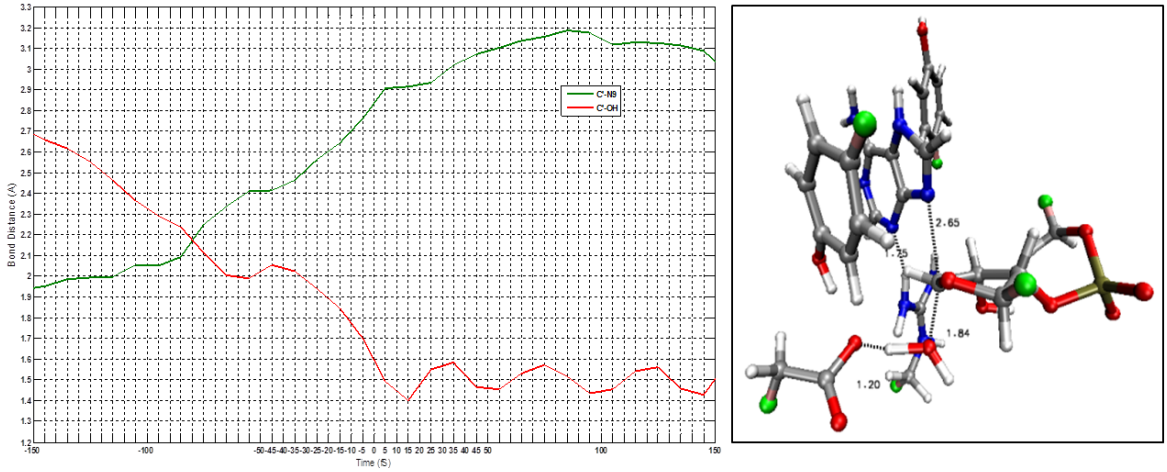
The incoming water molecule is activated by GLU177, when a proton is abstracted onto oxygen OE2 thereby converting glutamate into glutamic acid (Scheme 3.(b)). Consequently, the embryonic nucleophile (OH^-) is generated. It can be clearly shown that the activation of the water / generation of the nucleophile process has a very low activation energy barrier that is around 3kcal/mol as it occurs without any external forces. Various studies[1,2,13,18,30] have suggested a role for GLU177 as the base in catalytic mechanism, the findings from this study provide evidence for those suggestions.

Analysis of reaction trajectories strongly suggested a direct involvement of GLU208 in coordinating the incoming water molecule. The suggestion emanates from i) the shortening of the average distance (i.e., 2.50Å) observed between GLU208 and the incoming water molecule, ii) their mutually favourable orientation for strong H-bonding interactions and iii) the conservation of GLU208 throughout the ribosome inactivating protein, (RIP) family.

The Transition State

Locating the Transition State (TS) of a given chemical process is a challenging task. The classical Transition State Theory (TST) defines the TS as a first order saddle point where it is a minimum all directions but a maximum in the direction of the selected reaction coordinate. This definition of the TS was used to locate / identify the TS of the reaction studied here.

The TS was located using the free energy surface combined with the time series of the breaking bond (C1'-N9) and the forming bond (C1'-OH). Once the FEARCF method locates the reaction path it then produces the converged reaction surface leading to a flat histogram. The result is that the biasing force are increased and enhancing the sampling of the free energy surface to achieve equal sampling. Consequently, with greater biasing forces, i.e. in later stages of the simulation, produces trajectories that are deviate from the true chemical and physical reaction paths. Thus, the very first trajectory (out of multiple parallel reaction trajectories) that showed the entire chemical transformation from the reactants to the products was chosen to investigate the reaction mechanism. From this and the reaction surface all mechanistic details and data are derived.



(a) (b)

Figure 4: (a). Distance time series of the breaking and the forming bond. (b). The Transition State of the ricin-catalysed reaction.

Figure 4.(a) shows the time series plot of the breaking bond and the forming bond. The corresponding bond lengths are given in angstroms on the Y axis where as the X axis represents a total selected simulation time of 300fs. The zero on the time axis corresponds to the point at which the products are first formed. Formation of the products is identified by the formation of a novel C1'-OH bond, i.e., when C1'-OH reaches 1.60Å and is considered covalently formed.

For a pure S_N2 reaction, the crossover point of the time series of the forming and that of the breaking bond demarcates the TS. For this reaction, this cross over point is located at bond lengths C1'-N9 =C1'-OH=2.18Å, and on the free energy surface, and this point belongs to the reactant region. Therefore, cross-over point of does not signify the TS for this reaction. Hence, it was not chosen as the "zero-point", the reference point on the time axis of the time series plot (Figure 4(a)). Instead, the initial C1'-OH bond formation point was chosen as the reference point. The negative values on the time axis refer to the times prior to formation of the products and the positive values refer to the times after the formation of the products.

When the C1'-OH bond has just formed, i.e. at time zero with bond length of 1.60Å, the distance between C1' and N9 is 2.85Å, at distance the C1'-N9 bond can be considered as covalently non-effective. That is, the formation of C1'-OH bond and the cleavage of C1'-N9 bond are asynchronous. This kind of bond breaking and formation clearly suggest that the corresponding reaction mechanism is not pure S_N2 type.

Further, it can be seen that the bond forming process for the C1'-OH bond takes about 35fs. During this time period, the C1'-N9 bond is broken and C1'-N9 distance continues to increase from 2.45Å to 2.85Å. Therefore, it is argued here that the TS species of this reaction has a minimum life time of 35fs, i.e. of several bond vibrations. This observation is in excellent agreement with the results of the previous kinetic isotopes effects experiments[2]. Further, these findings strongly support the previously proposed[6,26] stepwise D_N*A_N mechanism. However, this cationic species is not considered as a true reaction 'intermediate" as in a S_N1 reaction, as there is no minimum found on the surface in the corresponding region. Therefore, it is argued that this cationic form is rather an "extended-transition state".

The C1'-OH bond starts to become effective only at -35fs, that is when the bond distance gets shorter than 2.0Å. At this point, the C1'-N9 bond distance is 2.45Å. these observations suggest that the TS for this reaction must be located within the region where the C1'-N9 distance is between 2.45Å and 2.85Å, and the C1'-OH distance is between 2.0Å and 1.6Å. Therefore, this region of the reaction potential energy surface was analysed to locate the first order saddle point that corresponds to the TS.

The transition state for the ricin-catalysed RNA hydrolysis reaction was identified at C1'-N9 (distance)=2.65Å and C'-OH (distance)=1.84Å (Figure 4(b)). The TS comprise of three fragments. They are: (a). the incoming nucleophillic OH^-, (b). the ribocation, (c) the leaving base. It was found that the TS has a cationic nature and it has a finite lifetime. The cationic nature was confirmed by the relatively high positive charge of +0.30e on C1' at the TS, in comparison to the corresponding charge of -0.41e in the reactants. These Mulliken atomic charges were calculated in the gas phase with B3LYP/631-G+(d, p) level of theory.

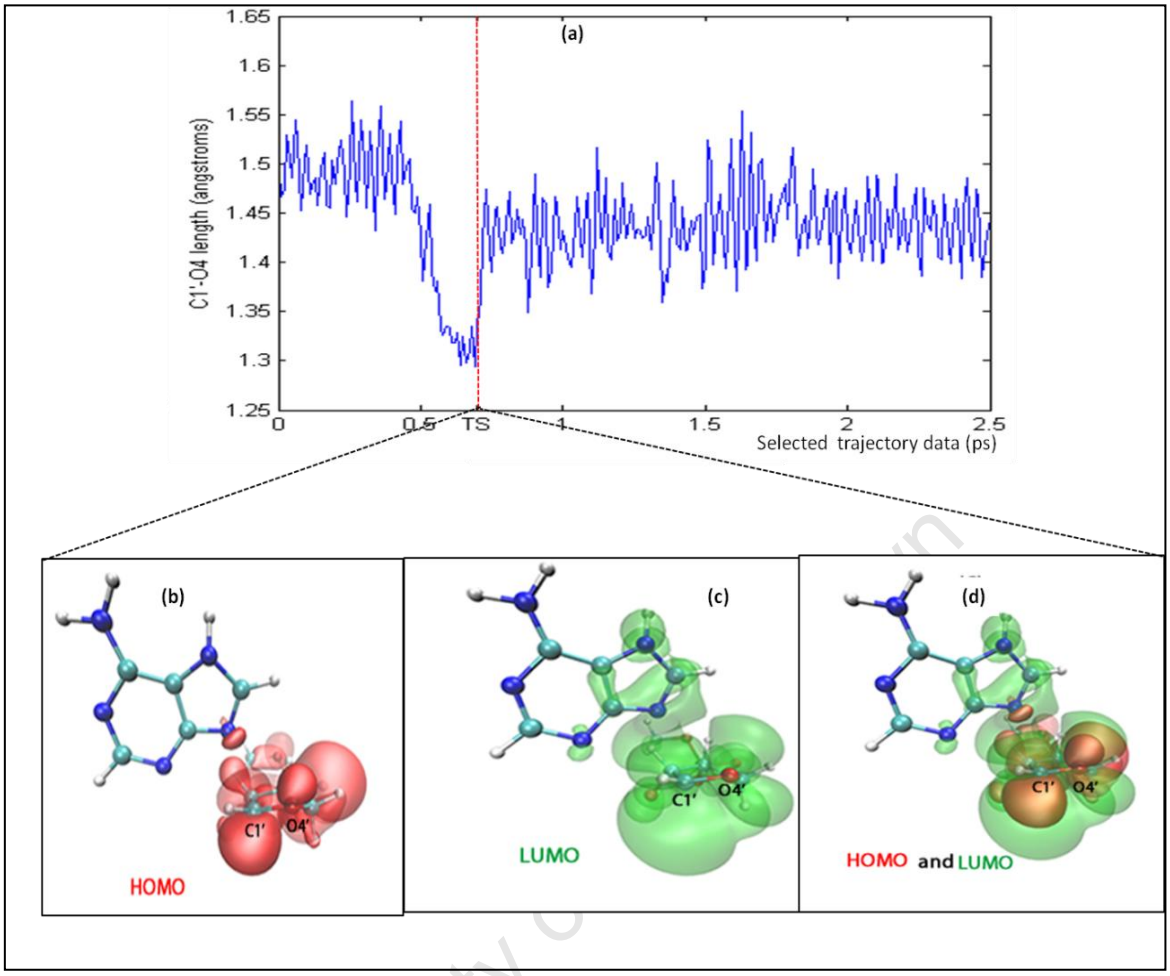


Figure 5: (a). C1'-O4' bond distance time series. (b). HOMO, (c). LUMO , (d). HOMO and LUMO of the transition state.

Note: both positive and negative phases of the LUMO and HOMO are coloured in a single colour for clarity.

The positive charge on C1' at the TS is primarily stabilised by delocalisation of electrons between C1' and O4'. The time series of O4'-C1' bond is given in Figure 5(a). At the TS, the O4'-C1' bond length is 1.32Å where as its equilibrated length in the starting structure was 1.51Å. This reduction in the bond length clearly indicates the double bond character of the C1'-O4' bond at the TS. The frontier molecular orbitals, HOMO and LUMO at the TS are shown in Figures 5(b) and 5(c) respectively.

Inspection of the orbital distribution shows that there is a substantial concentration of HOMO at the reactive C1'. Therefore, it is assumed that this delocalised electron density contributes dramatically towards the stabilisation of the cationic TS. Further, a considerable degree of HOMO and LUMO overlap (Figure 5 (d)) at C1' was identified.

Therefore, it is argued here that the effective overlap of HOMO and LUMO further enhances the stabilisation of the cationic TS. The net effect of the overlap of HOMO and LUMO, and the electron delocalisation is reflected in the shorter bond length of the O4'-C1' at the TS. In addition to the "self-stabilisation" via HOMO-LUMO overlap, it is further argued here that GLU177 provides a certain degree of stabilisation to the cationic TS via electrostatics, and the short distance of 3.29Å observed between C1' and GLU177 at the TS provides clear evidence to support this argument.

Figure 6. (a) and 6(b) show the Electrostatic Surface Potential (ESP) of ricin and the newly identified TS respectively. Figure 6(c) presents the ESP of the previously proposed TS by Schramm et.al[2]. The ESP of the new TS clearly shows that it has many negatively rendered sites than the sites that are positively rendered. Hence, it shows a great compatibility with the positively rendered catalytic domain (Figure 6. (a)). However, the previously proposed TS (Figure 6(c)) is predominantly positively rendered, and it is assumed that omission of the incoming OH^- from the TS has resulted in such positively rendered ESP. Analysis of Electrostatic Surface Potentials clearly suggests a very high electrostatic nature for binding of new TS to the ricin catalytic domain due to their mutual electrostatic compatibility. On the other hand, previously proposed TS is not electrostatically compatible with the ricin catalytic domain. Moreover, a very high degree of electrostatic repulsion is expected between that TS and the ricin catalytic domain. The previously proposed TS that was obtained from the gas phase energy minimisations, and it significantly differs in ESP from the currently proposed TS. Hence, the previously proposed TS is considered to be inaccurate with respect the catalytic mechanism that occurs in the enzyme.

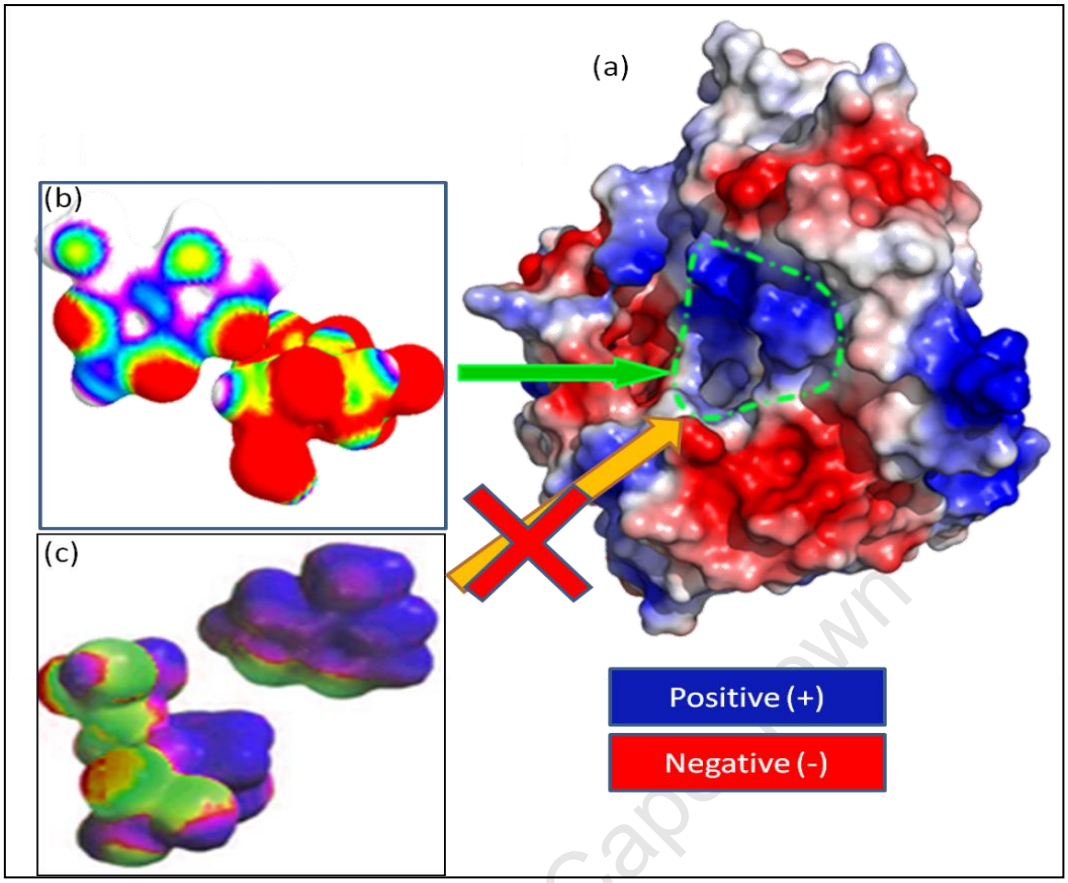


Figure 6 : Electrostatic surface potential of: (a) ricin ; generated with the CHARMM force field charges in the gas phase and by using the APBS[44] program. Green dashed line demarcates the active site.(b). the transition state in the gas phase; calculated using the GAUSSIANO3 program with B3LYP/6-31G+(d,p) level of theory. (c) TS proposed by Schramm et.al, the colour scheme of the original figure has been modified to match the current scheme.

At the TS, numerous hydrogen bonds were identified between the TS structure and the enzyme. The distance between HH12 of ARG180 and N3 of the leaving base was found to be 1.75Å at the TS. This relatively short distance suggests an extremely strong covalent–type hydrogen bonding interaction between ARG180 and N3. Therefore, it can be presumed that this interaction withdraws electrons from the pi system of the leaving base to the positively-charged guanidinium group of ARG180 while making the leaving group less nucleophillic.

As a result, reformation of N9-C1' becomes extremely difficult and it allows the incoming OH- an easy attack on C1'. Further, it is argued here that the torque arising from the ARG180-N3 intermolecular force (H-bond) enhances the cleavage of the C1'-N9 bond. At the TS, GLY121 (treated MM) accepts two simultaneous H-bonds from H61 and H7 of the leaving base. The distance between backbone oxygen of GLY121 and H61 and H7 are 3.33Å and 2.26Å respectively. These hydrogen bonds are also expected to enhance the leaving group departure and the glycosidic bond cleavage.

At the TS, puckering of the ribose ring mainly shows OT4 (C4'-exo) character (Figure 7(c). This pucker conformation at the TS however, does not agree with the previously proposed[2] C3'-endo conformation for the TS ribocation. It is argued here that the previously proposed transition state structure determined by a gas phase energy minimisation cannot be compared with the currently proposed TS that was obtained from the reaction simulation carried in the solvated protein.

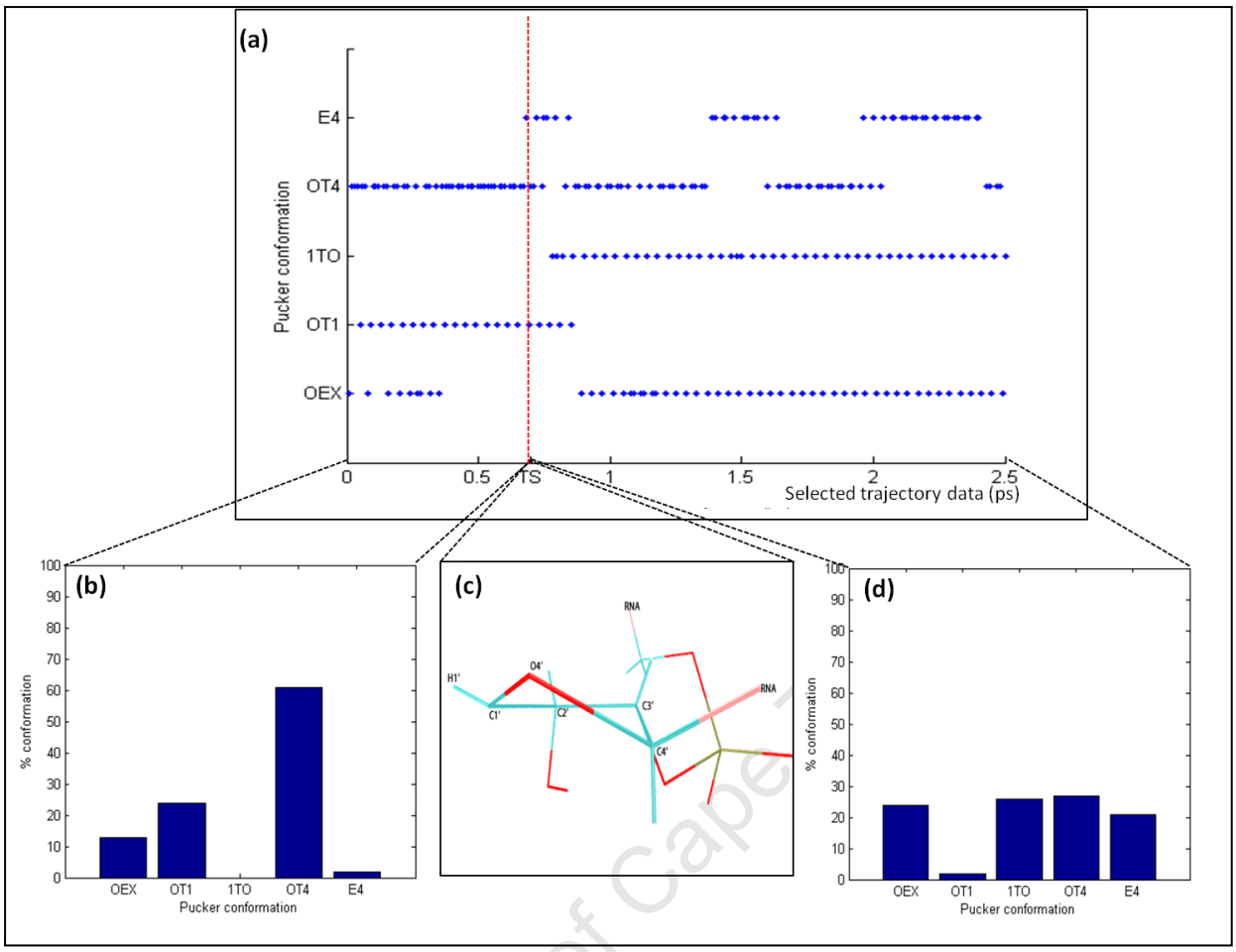


Figure 7: Pucker analysis of the reaction trajectory (a). Ribose pucker conformation time series. Pucker population histograms (b). In the reactants. (d). In the products. (c). Ribose pucker at the TS.

In the Near Attack Conformations (conformations immediately prior to the TS) , the sugar puckering is predominantly (61%) $^{O}T_4$ (Figure.(7b)). In the products however, there is an even distribution of $^{O}E_x$, $^{1}T_O$, $^{O}T_4$ and E_4 pucker conformations (Figure 7.(d)). This distribution of pucker conformations has clearly shown that ricin restricts the ribose puckering in target adenine. This reduction in the pucker degrees of freedom in the reactants (especially in the NACs) facilitates the formation of the carbocation and enhance attack by the nucleophile.

In the previously proposed TS with C3'-endo pucker conformation, the atoms O4',C1' , C2' and H1' were considered to be coplanar with a pure sp^2 character at C1'. However, in the TS found in study, a pure sp^2 planer geometry at C1' is not observed. In current TS, H1' is found to be not coplanar with O4',C1' and C2' and it lies about 35° above the plane of those three atoms.

It was found that the atomic (Mulliken) charge on C1' was -0.040e in the O4'-C1'- C2'-H1' coplanar conformation , where as the charge on C1' in the currently proposed TS (where O4'-C1'- C2'-H1' is non-coplanar) is +0.30e. These charges clearly show that the O4'-C1'- C2'-H1' coplanar conformation is not cationic in nature and therefore is not ready for the attack by negatively charged OH^-. On the other hand, the TS proposed here shows a high reactivity towards the incoming nucleophile with an increased positive charge on C1'. Therefore, when its effect on the steric hindrance and the charge distribution is taken in to account, it is argued here that the ribocation at the TS in the ricin-catalysed reaction will show predominant C4'-exo character in its pucker conformation.

The Role of GLU177

GLU177 is identified as the most important residue in the ricin catalytic domain where it is assigned the role of base in the catalytic mechanism. Furthermore, it is proposed here that GLU177 (along with GLU208) is also involved in coordinating the catalytic water to generate the attacking OH-. Moreover, the distance of 3.29Å between GLU177 and C1' at the TS suggests that GLU177 could also contribute to the stabilisation of the cationic TS via electrostatic interactions.

It was found that active oxygen (OE1) of GLU177 is strongly H-bonded to the hydroxyl hydrogen of TYR123 and guanidinium hydrogen (HH1) of ARG180 simultaneously. The H-bond distances were 1.77Å and 2.03Å for TYR123 and ARG180 respectively. It is argued here that this special arrangement of TYR123-GLU177-ARG180 triad is enzyme's strategy to prevent any covalent interactions between GLU177 and the substrate or the TS during the mechanism. GLU177 can be expected to covalently bind to the TS which could easily be driven by the large favourable electrostatic interactions in the absence of the strong H-bonds with TYR123 and ARG180. Given such argument, it is proposed here that TYR123 plays an important role in controlling the reactivity of GLU177 in the mechanism. Therefore, it is concluded that TYR123 has a significant role beyond a "binding-residue".

Formation of the Final Products

The reaction is completed when the highly reactive cationic TS is attacked by the incoming OH- nucleophile. The LUMO of the TS is highly concentrated around C1' and it is highly diffused below the plane of the ribose ring (Figure 8 (d)). This clearly indicates that the nucleophile must attack C1' from below the plane of the ribose ring, but not from above it. An attack from the top will follow a retaining mechanism. However, such attack from the top is impossible in this reaction mainly due to the steric crowding and the absence of a long-lived reaction intermediate. It is argued here that the strong electrostatic attraction between the positively charge on C1' and the negatively charged OH- drives the last step in the ricin-catalysed reaction.

6.5.2. *Energy Facts and the Reaction Surface*

Figure 8. (a) and 8.(b) present the free energy surface of the reaction catalysed by ricin. Due to the complexity of the free energy surface and other technical difficulties (discussed later), computationally determined (from the free energy surface) barrier was 16.85 Kcal/mol greater than the expected (calculated from k_{cat} and k_M) barrier height of 15.65 kCal/mol. Moreover, simulating the reaction in water was not successful due to the current difficulties in handling QM water molecules. However, from the 1D-PMF calculations (Figure 8.(c)) , it was found that the cleavage of target glycosidic bond in ricin catalytic domain requires 24.55 kCal/mol less energy compared to the cleavage in water. This clearly demonstrates the efficiency of ricin in cleaving the C1'-N9 bond and indicates an extremely high efficiency in the overall reaction.

The free energy surface presented here was successfully reproduced many times with various simulation times and with multiple simulations where the number of simulations per single PMF iteration was as many as 72. This reproducibility test confirmed that there exists no other feasible reaction mechanism for RNA depuration by ricin, other than the mechanism presented here.

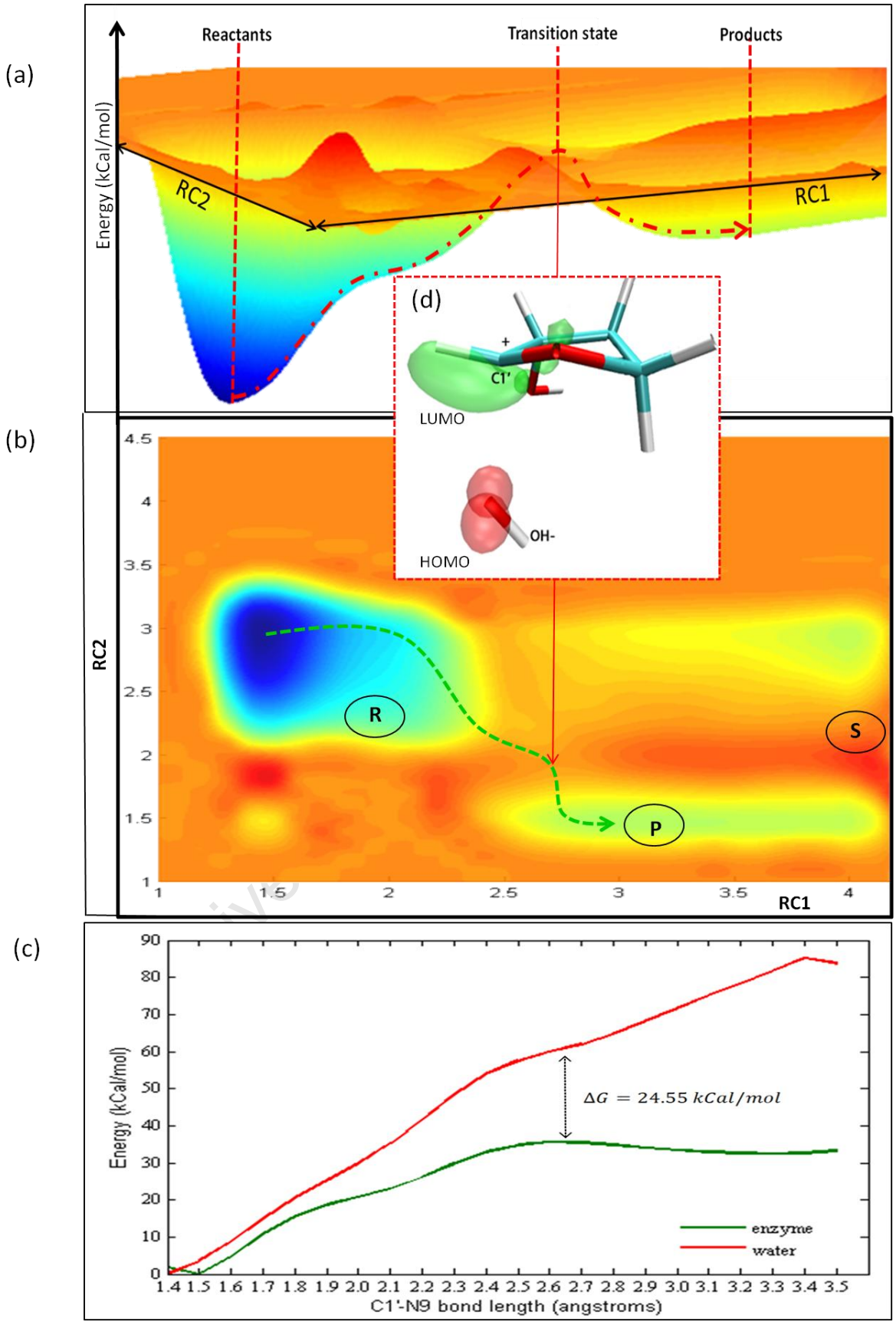


Figure 8: (a), (b). Free energy surface of the ricin-catalysed reaction. (c). 1D Free energy profile of the C1'-N9 bond cleavage. Green: in ricin , Red: in water. (d). Molecular orbitals of the oxocarbenium cation and the incoming nucleophile.

The free energy surface shows a total of three wells; namely **R, P** and **S**. As a result of these three wells, the relative geometry of the surface is somewhat complex. Well **R** corresponds to the reactants and well **P** corresponds to the expected products. Whereas, the third well, **S** is mainly due to the chemistry /post-reaction modifications of the products. Analysis of certain reaction trajectories showed that formation of this well was also contributed by the undesired side reactions of the cationic intermediate species i.e. the TS. This is particularly possible in later simulation times when the biasing force is relatively high. Studying about the complex behaviour of the products and the alternative /side reactions of cationic intermediate was beyond the scope of this study, and it must be investigated separately.

Other than the complexity of the free energy surface, there are various other factors that can affect the barrier height in a free energy simulation. The SCC-DFTB method was used in this study because of its proven ability[42,45-49] to simulate biologically relevant systems and their reactions. However, some authors have argued that the non-ab initio (NAB) i.e. the semi-empirical and approximated methods are typically qualitatively reasonable but quantitatively inaccurate[50] . Further, it is known that the accuracy of the NAB-QM/MM MD and free energy simulations may be limited by the approximations in their descriptions of the breaking and formation of chemical bonds as well as interactions. In addition, parameters that are used in such methods have a significant influence on the results. It has been proposed[51] that Specific Reaction Parameters (SRPs) must be used to circumvent this problem in NAB methods. The SRPs are obtained by adjusting existing parameters to experiments or to a higher level of theory[52] for the specific chemical system under focus. Moreover, the quantum mechanical effects such as tunnelling and zero point energy corrections which are not included in the NAB methods also contribute largely to the inaccuracies[52].

In addition to these general deficiencies of the NAB method, the SCC-DFTB has certain ingrained deficiencies. One such problem that is particularly important to this study is its inaccuracy in estimating accurate barrier heights[53-55]. SCC-DFTB does not accurately model dispersion interactions.

Even though, the dispersion forces are taken in to account via an empirical correction, their real quantum mechanical meaning and the influence is included. A great deal of pi-pi interactions are expected between the target adenine and the catalytic domain residues (TYR80 and TYR123). Therefore, it is clear that a true quantum mechanical description of dispersion is required to model the TYR80-ADENINE-TYR123 triad. Hence, poorly modelled interactions of this triad might have affected the reaction barrier.

Over and above the previously mentioned method-related deficiencies, certain choices made in setting up the simulation-system might also have affected the results. The major choices include the SCC-DFTB parameters, the methods to use to treat the QM/MM boundaries, the distance of the QM/MM boundary cut-offs to the reaction centre and the size of the QM region. Below discussed are some of these choices made in this study that might have influenced the obtained results.

It has been shown that larger QM regions in QM/MM simulations lead to more accurate results[43]. In the reaction studied here, if it was not limited by the number of QM atoms, GLY121 and VAL81 should have included in the QM due to their expected strong H-bonding interaction with leaving adenine. Since VAL81 and GLY 121 were not included in the QM region, it is assumed that it has made departure of the adenine more difficult and might have affected the barrier height.

The location of the link atoms with respect to the reaction region is another factor that can affect the accuracy of the energies obtained. The link atom at C4' (Figure 2.(a)) is only 4.5Å away from the reactive C1'. Inclusion of this link atom at C4' was unavoidable due to the absence of C-O or P-O "cut" in the current SCC-DFTB-link atom scheme.

It is of prime importance to incorporate the correct solvent effect in an enzymatic reaction. For example, the proton wires and proton hopping play significant roles in reaction mechanisms. In order to reasonably address these phenomena, a large number of water molecules have to be included in the QM region. However, due to the restrictions on the number of QM atoms, and difficulties in handling multiple QM waters due to their diffusion, only a single water molecule was treated quantum mechanically. Therefore some discrepancies are expected in the results due to poorly modelled solvent effects.

As mentioned before, the mechanism proposed here was confirmed by reproducing the free energy surface multiple times, although the above discussed technical issues have produced a higher reaction barrier.

6.6 Concluding Remarks

For the first time, this study has shown a feasible mechanism of action for the ricin-catalysed RNA depurination reaction via computational methods. This study has answered many unanswered questions in ricin chemistry, and therefore, it is considered as a major milestone in this area of research. Orientation and activation of the catalytic water molecule, amino acid residues that are directly involved in the mechanism, nature of the transition state were amongst the key issues that were attempted to address in this study. Moreover, the mechanism proposed here agrees well with the previous experimental and theoretical suggestions about the reaction mechanism. The knowledge gained about the ricin-catalysed reaction from this study should broadly be useful in designing an effective antidote for this deadly poison.

6.7 References:

(1) Chen, X. Y.; Link, T. M.; Schramm, V. L. *Biochemistry* **1998**, *37*, 11605.
(2) Chen, X.-Y.; Berti, P. J.; Schramm, V. L. *Journal of the American Chemical Society* **2000**, *122*, 1609.
(3) Schramm, V. L. *Accounts of chemical research* **2003**, *36*, 588.
(4) Schramm, V. L. *The Journal of biological chemistry* **2007**, *282*, 28297.
(5) Schramm, V. L. *Annual review of biochemistry* **2011**, *80*, 703.
(6) Chen, X. Y.; Berti, P. J.; Schramm, V. L. *Journal of the American Chemical Society* **2000**, *122*, 1609.
(7) Bene, J. E. D. *Journal of American Chemical Society* **1983**, *1673*, 367.
(8) Huang, Q.; Liu, H.; Tang, Y.; Jin, S.; Wang.Y *Biochemistry* **1995**, *309*, 285.
(9) Chen, X. Y. *Journal of the American Chemical Society* **2000**, *122*, 6527.
(10) Unrau, P. J.; Bartel, D. P. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 15393.
(11) Michael, J.; Roberts, M.; Robertust, J. O. N. D. *The FASEB Journal* **1994**, *8*, 201.
(12) Ready, M. P.; Kim, Y.; Robertus, J. D. *Proteins* **1992**, *10*, 270.
(13) Monzingo, A. F.; Robertus, J. D. *Journal of Molecular Biology* **1992**, *227*, 1136.
(14) Anslyn, E. V.; Dougherty, D. A. *Modern Physical Organic Chemistry*; University Science Books, 2006; Vol. 69.
(15) Ho, M.-C.; Sturm, M. B.; Almo, S. C.; Schramm, V. L. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 20276.
(16) Yang, X.; Gérczei, T.; Glover, L. T.; Correll, C. C. *Nature structural biology* **2001**, *8*, 968.
(17) Yan, X. J.; Hollis, T.; Svinth, M.; Day, P.; Monzingo, A. F.; Milne, G. W. A.; Robertus, J. D. *Journal of Molecular Biology* **1997**, *266*, 1043.
(18) Ready, M. P.; Kim, Y. S.; Robertus, J. D. *Proteins-Structure Function and Genetics* **1991**, *10*, 270.
(19) Olson, M. a. *Proteins* **1997**, *27*, 80.
(20) Monzingo, A. F.; Robertus, J. D. *Journal of Molecular Biology* **1992**, *227*, 1136.
(21) Day, P. J. *Biochemistry* **1996**, *35*, 11098.
(22) Endo, Y. *Journal of Biological Chemistry* **1987**, *262*, 5908.
(23) Yan, X.; Day, P.; Hollis, T.; Monzingo, a. F.; Schelp, E.; Robertus, J. D.; Milne, G. W.; Wang, S. *Proteins* **1998**, *31*, 33.
(24) Carra, J. H.; McHugh, C. A.; Mulligan, S.; Machiesky, L. M.; Soares, A. S.; Millard, C. B. *Bmc Structural Biology* **2007**, *7*.
(25) Versées, W.; Loverix, S.; Vandemeulebroucke, A.; Geerlings, P.; Steyaert, J. *Journal of molecular biology* **2004**, *338*, 1.
(26) Chen, X. Y.; Link, T. M.; Schramm, V. L. *Biochemistry* **1998**, *37*, 11605.
(27) Endo, Y.; Tsurugi, K.; Ebert, R. F. *Biochimica et Biophysica Acta* **1988**, *954*, 224.
(28) Roday, S.; Amukele, T.; Evans, G. B.; Tyler, P. C.; Furneaux, R. H.; Schramm, V. L. *Biochemistry* **2004**, *43*, 4923.
(29) Frankel, A.; Welsh, P.; Richardson, J.; Robertus, J. D. *Molecular and cellular biology* **1990**, *10*, 6257.
(30) Chaddock, J. A.; Roberts, L. M. *Protein Engineering Design and Selection*

Protein Eng. **1993**, *6*, 425.
(31) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H. M.; Pardi, A. *Journal of Molecular Biology* **1996**, *264*, 968.
(32) Vriend, G. *Journal of Molecular Graphics* **1990**, *8*, 52.
(33) MacKerell, A. D.; Banavali, N. K. *Journal of Computational Chemistry* **2000**, *21*, 105.

(34) Foloppe, N.; MacKerell, J. A. D. *Journal of Computational Chemistry* **2000**, *21*, 86.
(35) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926.
(36) Brooks, C. L.; Brünger, A.; Karplus, M. *Biopolymers* **1985**, *24*, 843.
(37) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *Journal of Computational Physics* **1977**, *23*, 327.
(38) Strümpfer, J.; Naidoo, K. J. *Journal of Computational Chemistry* **2010**, *31*, 308.
(39) Naidoo, K. *SCIENCE CHINA Chemistry* **2011**, *54*, 1962.
(40) Barnett, C. B.; Naidoo, K. J. *Molecular Physics* **2009**, *107*, 1243.
(41) Field, M. J.; Bash, P. A.; Karplus, M. *Journal of Computational Chemistry* **1990**, *11*, 700.
(42) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Physical Review B* **1998**, *58*, 7260.
(43) Stanton, C. L.; Kuo, I.-F. W.; Mundy, C. J.; Laino, T.; Houk, K. N. *The journal of physical chemistry. B* **2007**, *111*, 12573.
(44) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proceedings of the National Academy of Sciences* **2001**, *98*, 10037.
(45) Elstner, M. *Theoretical Chemistry Accounts* **2005**, *116*, 316.
(46) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *The Journal of Physical Chemistry B* **2001**, *105*, 569.
(47) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S. n.; Kaxiras, E. *The Journal of Chemical Physics* **2001**, *114*, 5149.
(48) Barnett, C. B.; Wilkinson, K. a.; Naidoo, K. J. *Journal of the American Chemical Society* **2010**, *132*, 12800.
(49) Barnett, C. B.; Naidoo, K. J. *The journal of physical chemistry. B* **2010**, *114*, 17142.
(50) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. In *Journal of the American Chemical Society* 1985; Vol. 107, p 3902.
(51) Rossi, I.; Truhlar, D. G. *Chemical Physics Letters* **1995**, *233*, 231.
(52) Truhlar, D. G.; Gao, J.; Alhambra, C.; Garcia-Viloca, M.; Corchado, J.; Sánchez, M. L.; Villà, J. *Accounts of chemical research* **2002**, *35*, 341.
(53) Sattelmeyer, K. W.; Tirado-Rives, J.; Jorgensen, W. L. *The journal of physical chemistry. A* **2006**, *110*, 13551.
(54) Tubert-Brohman, I.; Guimarães, C. R. W.; Jorgensen, W. L. *Journal of chemical theory and computation* **2005**, *1*, 817.
(55) Sattelmeyer, K. W.; Tubert-Brohman, I.; Jorgensen, W. L. *Journal of Chemical Theory and Computation* **2006**, *2*, 413.

CHAPTER -7

TRANSITION STATE ANALOGUE DESIGN AND NOVEL INHIBITORS FOR RICIN

** Intellectual Property Status*

The novel ricin inhibitors and the novel transition state analogue design protocol presented in this chapter are covered under the patent licence: ***RSA Patent Application No: 2011/06840,*** *titled ,* ***RICIN TRANSITION STATE ANALOGUE INHIBITORS,*** *by* ***Jayakody, R.S. and Naidoo, K.J.*** *The reader is referred to this patent for detailed structures of the inhibitors and for the details of the design protocol.*

7.1. Transition State analogues

It has been proposed by Linus Pauling[1] that the powerful catalytic action of an enzyme is due the tight binding between the enzyme and the Transition State (TS) species of the reactions catalysed by them. Enzymes are capable of achieving remarkable catalytic rate enhancements because they significantly lower very high TS energetic barriers of the reaction as observed in solution[2]. The binding energies of TS binding to the enzymes are caused by the realignment of the enzyme-substrate contacts as the enzyme and the substrate rearrange their conformation and structure in TS complex[3]. By invoking the idea of Pauling on TS binding, Wolfenden[4] proposed that the chemically stable molecule that resembles the TS state would be expected to bind to an enzyme tighter than its natural substrate. These chemically stable TS mimics are known as Transition State Analogues (TSAs). In general , TSAs bind to the enzyme 10^2-10^4 tighter than the natural substrate[5,6].

Design of TSAs is a promising way of making powerful inhibitors[2,3,7,8]. To achieve this goal, one requires a complete (structural, conformational and electronic) fingerprint of the transition state. Kinetic Isotope Effect (KIE) experiments are the most commonly method to understand the nature of the TS. This method was first invented by Bigeleisen et.al[9]. by tracing the relationships between the isotopic substitution and the subsequently altered reaction rates. In the recent past, this approach has been extensively used by Vern Schramm[2,3,7,8,10,11] to propose reaction mechanisms and design TSAs for various enzymes. The KIE experiments have been detailed in chapter 3.

Another promising experimental technique that can be used to elucidate chemical reaction mechanisms and corresponding TS is ultrafast laser spectroscopy[12]. However, much effort is needed experimentally to obtain sufficient information about the TS, which is extremely short-lived with a lifetime of few femtoseconds.

Computer simulations on the other hand, can provide detailed atomistic level information about a reacting chemical system which is not usually accessible by experimental techniques. This includes the structure and the other chemical-electronic features of the TS. One can use a suitable computational technique to simulate the chemical reaction, from which the TS can be identified. For details on computer simulations of chemical reactions, the reader is referred to chapter 3.

7.2. Energetics of Transition State Analogue Inhibitor Binding

Figure.1 shows the comparison between the energy profile of Transition State Analogue Inhibitor (TSAI) binding, and that of the catalysed and uncatalysed reactions[3]. It is expected to observe a mismatch of substrate recognition features when the enzyme is presented with a TSAI. Therefore, in general , binding of a TSAI is recognised by rapid weak binding followed by slow tight binding[13].

In Figure 1 , $[E:I]^{\#}$ represents product of the rapid weak binding of the transition state analogue $I^{\#}$, to the enzyme E. Energetically more difficult tight binding which follows the rapid weak binding is represented by the formation of the stable-$EI^{\#}$ complex from $[E:I]^{\#}$ complex. Upon the binding to the enzyme, TSAIs will invoke unique conformational changes in the protein by initializing the interactions that are similar to the ones found between TS and enzyme. However, the stable bond at the reaction centre of the TSAI will prevent the initialisation of the catalytic cycle. Further, due to the incompletion of the reaction, there is no escape for the stable-$EI^{\#}$ complex from the enzyme via the product expulsion pathway. In such inhibitor-enzyme complexes, the energy of the TS stabilisation ($\Delta\Delta G^{\#}_{TS\ stabilization}$) is converted in to binding energy for the transition state inhibitor, ($\Delta\Delta G - I^{\#}_{binding}$). Furthermore, owing to the tight-binding of TSAI, it is not possible to convert the stable-$EI^{\#}$ complex back to the $[E:I]^{\#}$ complex.

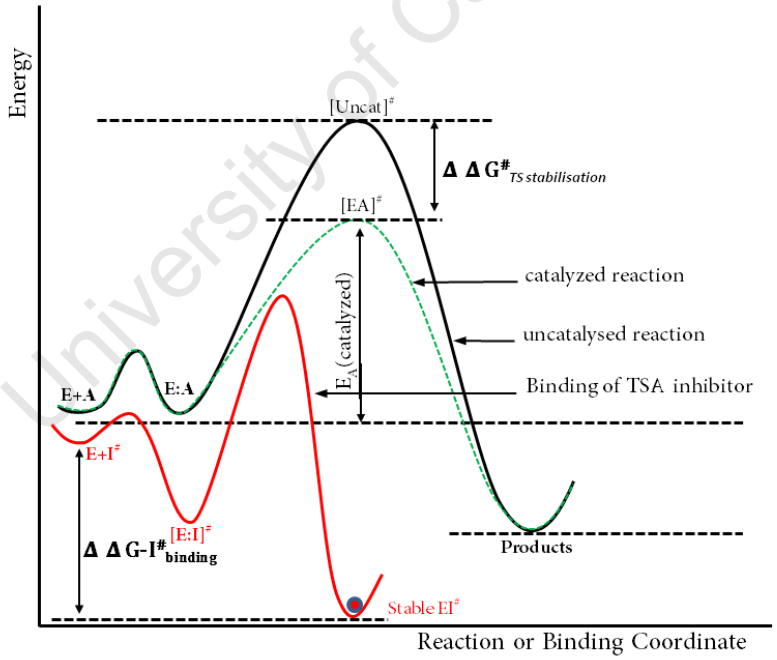


Figure 1: Energy profile of TSAI binding Vs. energy profiles of catalysed and uncatalysed reactions.

It has been argued[14] that the natural substrate of an enzyme actively induces certain catalytic conformational changes in it. However, the efficiency of inducing TS-enzyme like conformational changes in the enzyme by the TSAI is less than that of enzyme's natural substrate. Owing this longer induction process, TSAIs show relatively slow onset of action [14]. The kinetics of TSA-type inhibitors are discussed in details in chapter 2.

7.3. Design of Transition State Analogues

It can be assumed that the objectives in TSAI designing are threefold. Firstly, the TSAI needs have a compatible geometry (size and shape) with the natural substrate of the enzyme in order to access the protein binding site. Secondly, a TSAI should be able to mimic the hydrogen bonding interactions between the real TS and the enzyme. Thirdly, a prospective TSAI inhibitor is expected to have a very closer mimic of the electrostatic surface potential of the real TS to ensure the presence of important polar interactions.

However, the theoretical binding affinity (expected value) is not achieved by any of the currently available TSA inhibitors. Because, not any stable molecule can completely mimic the geometric and electronic features of an unstable transition-state structure, especially with respect to the bond forming and bond breaking regions. In spite of this, when applied correctly TSA design provides an effective conceptual basis for inhibitor design[15].

7.3.1. Geometry of Transition State Analogues

There is often a considerable difference between the geometry of the TS and that of the the ground state of the natural substrate. Many of these geometrical transformations involve interchange between different hybridisation forms of carbons i.e. sp^2 and sp^3. Many cases of TSAI design, take the advantage of this difference. To achieve this goal, stable replacements for cationic (sp^2) carbons and tetrahedral adducts (sp^3) have been used extensively[15].

When designing a TSAI, to arrive at the desired geometrical features, one can use the natural substrate as a scaffold. This can then be modified according to the available information (specially the breaking and forming bond distances, corresponding angles and dihedrals) about the transition state structure. As mentioned earlier, these modifications can be done by introducing different hybridisations, new atoms and chemical groups. Along with these modifications, one has to ensure that non-reactive (stable) bonds are introduced at the reaction centre.

7.3.2. Forces in TSA Binding

As mentioned above, understanding of the forces that are involved in inhibitor-enzyme binding is of prime importance in designing any effective inhibitor. Further, in the essence of structure-based inhibitor designing, attempts are made to maximise these interactions and there is no exception for this in TSAI designing.

The interactions that can present in inhibitor-enzyme binding are the same type of interactions that stabilise the protein tertiary structure. These interactions include hydrogen bonds, hydrophobic interactions, electrostatic interactions, ion-dipole interactions, dipole-dipole interactions, charge-transfer interactions and cation-π interactions. All of these interactions contribute enthalpically to the inhibitor-enzyme binding free energy. On the other hand, certain interactions such as hydrophobic effect contribute entropically to the binding free energy.

The hydrophobic effect can be considered as a stabilisation arising from transfer of hydrocarbon surface of the inhibitor out of water to the nonpolar interior of a protein.[16] It has been proposed that the hydrophobic effect arises less from attraction of the hydrophobic surfaces for each other, but more from the favourable change in free energy as ordered water molecules surrounding the hydrophobic surface are released in to the bulk solvent.[17,18]. Consequently, this effect has a direct correlation with the surface area that is desolvated. Therefore, when designing inhibitors for hydrophobic (non-polar) binding sites, caution must be exercised to have the correct hydrophobic surface on the inhibitors.

As the van der Waals (vdW) interactions between elements in the first and second rows of the periodic table is relatively insensitive to the nature of the atoms involved, it does not change significantly on replacing solvent-ligand contacts with solvent-solvent or ligand-protein contacts. Therefore, vdW interactions often do not play significant role in inhibitor–protein binding. Hence, it has the minimum influence on the inhibitor design process.

The electrostatic and the hydrogen bonding (H-bonding) interactions are considered to be contributing significantly to the inhibitor-protein binding process. It has been suggested that a single H-bond can contribute 4-6 kcal/mol to the binding free energy[19]. Therefore, preserving the sites that can invoke the TS-like H-bonding pattern is extremely important in TSAI design. It has been shown that electrostatics plays a crucial role in TS stabilisation[20]. Therefore, it is of prime importance to mimic the electrostatic features of the real TS in the TSA inhibitors. To achieve this goal, one can make the use of electrostatic potential surfaces.

7.3.3. The electrostatic Surface Potential

As a consequence of chemical reactivity, the electronic charge in a given molecule is redistributed. Therefore, the TS structures will have distinguishable unique electronic charge distribution pattern from the corresponding reactants and the products. This unique charge distribution pattern is used when designing TS mimics. The Electrostatic Surface Potential (ESP) is a reflective map of electron/charge distribution in a chemical species.

The electrostatic potential is a scalar field, hence it is anywhere in space.[21] The operator for the electrostatic potential is where vector is the position vector of the charge distribution that gives an electrostatic potential at any point . Therefore, the electronic part of the electrostatic potential is given by,

——— [1]

where, represents the electron density at distance . The total electrostatic potential of a given system is given by the superposition of the classical electrostatic potential of the nuclei and the above given (equation 1) electrostatic potential of the electrons. To calculate these potentials, one has to use accurate electronic structure (QM) calculations. Once the total electrostatic potential is known, it is possible to calculate the corresponding electric field. The electric field is defined as the first derivative of the electrostatic potential as given by equation 2.

[2]

The ESP is generated by calculating the interaction between the electrostatic potential at individual points of the electron iso-density map/ the van der Waals surface with an external point-positive charge. Based on the outcome of the interaction, i.e. repulsive or attractive, the corresponding point on the surface is coloured with a specific colour. The electrostatic potentials have the units of energy.

If one is to mimic the charge distribution pattern of the TS, the ESPs can be used to achieve that goal. Once the geometrical mimicry is obtained (as discussed before), the ESP of the TSAI can be compared with that of the real TS. A higher degree of similarity in ESPs reflects a higher degree of similarity in charge/electron density distribution.

7.4. Transition State Analogues as Drug Candidates

The novel inhibitors that are designed according to the above procedure can be considered as potential lead compounds. Satisfactory results of the initial screening tests allow the novel drug candidates to enter the lead optimisation phase of drug design /drug discovery pipeline. A modern-day *in silico* drug discovery cycle is illustrated in Figure 2.

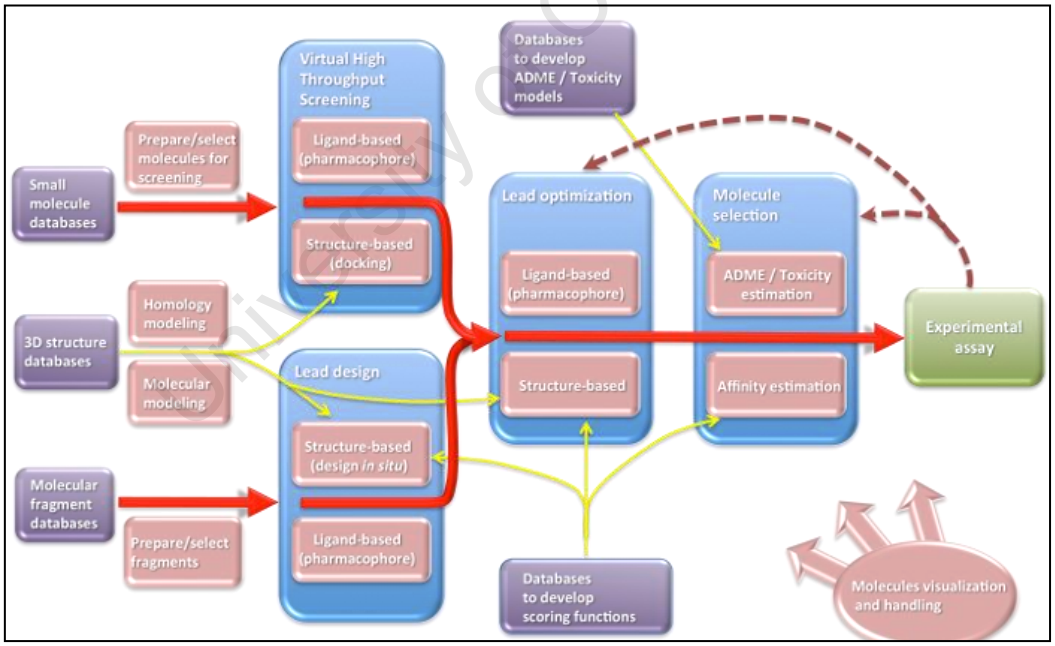


Figure 2: *In silico* drug design pipeline as applied in several commercial and academic laboratories.

7.4.1. Initial Screening with Molecular Docking

Molecular docking is the most widely used method in screening the drug candidates in modern drug discovery protocol. In the process of molecular docking, possible binding geometries of a molecule with a macromolecule are studied. The docking process is composed of two parts: a search algorithm and a scoring algorithm. The search algorithm ensure the sufficient sampling of the degrees of freedom of the ligand-macromolecule in order to include as many as possible true binding modes, where as the the scoring algorithm represents the thermodynamics of ligand-receptor binding interactions which is used to distinguish the true binding mode(s) from all others explored binding conformations [22].

The search problem is computationally very expensive. Therefore, there have been many proposed solutions to this problem. Those methods include docking with molecular dynamics, docking with Monte Carlo algorithms[23,24] and docking with simulated annealing[25-27]. Moreover, certain search protocols that consider molecular flexibility use rotamer search[28-30], distance geometry[31], and genetic algorithm[32] methods. To make the search tractable for processing a large set of molecules, the molecular components are often treated as rigid objects. Based on this approximation, docking can be done with either of systematic searching[33], pattern recognition[34,35], graph theoretical[36,37] or superposition techniques.[38]

Scoring functions are classified in to three major classes; (i)Force field functions , (ii). Empirical functions, (iii). Knowledge-based functions. Further, there is a fourth class where it employs mixed features of the above functions. Standard force field energy calculations only attempt to compute the enthalpic interactions. Therefore, such calculations are relatively fast and computationally inexpensive. On the other hand, methods such as Free Energy Perturbation (discussed later) in which the free energies are calculated are computationally expensive. Therefore, the non-bonded part of a force field function is usually modified with empirical terms, capturing some of the entropy and solvent effects, and the resulting scoring function is parameterised against experimental binding data[39] for optimisation.

The entropy terms are generally structural descriptors, such as the number of hydrogen acceptors and donors, number of torsional degrees of freedom. The torsion angles are given special attention in scoring functions, since it is argued that the ligand-enzyme binding causes fixing of torsion angles from which there is a constant loss of entropy for each entity. A general sample function is given by equation 3.

$$\Delta G_{scoring} = a_1 \Delta E_{vdw} + a_2 \Delta E_{ele} + a_3 \Delta G_{rot} + a_4 \Delta G_{H-bond} + a_5 \Delta G_{sol} + \cdots \quad [3]$$

where, $a_i (i = 1,2 \ldots 5)$ represent the corresponding weighting factors which are fitted to actual binding data for a specific protein-ligand system. Developing scoring functions is an active field of research and no fixed general scoring function has been developed yet. The accuracy of a docking calculation depends mainly on the chosen scoring function and the search algorithm. With carefully chosen docking scoring function and a sampling algorithm, comparable inhibitors can be reasonably ranked.

7.4.2. Further optimisation and screening

Molecules that pass through the initial binding test proceed to the lead optimisation (LO) phase. The early stages of LO-phase are focused on enhancing the binding affinity and the selectivity, where the nature of selectivity depends on the therapeutic area. In some areas (treatment for acute diseases) rapid and short-term intervention of the drug is required where as in other areas (treatment for chronic diseases) slow and long-term intervention is required[40]. If one is to optimise the TSA inhibitor, almost no changes to the molecular core can be done as it needs to mimic the core of the TS. However, changes to the periphery of the core can be done while preserving the desired features. Such small changes sometimes result in remarkable changes in the activity[40].

7.4.2.1. Qualitative Structure Activity Relationship (QSAR)

The QSAR analysis can be performed on a drug candidate in order to further improve it. QSAR is based upon the assumption that the difference in physiochemical properties accounts for the changes in biological activities of compounds. This idea was first introduced by Hansch and co-workers[41,42], and it is primarily based on the linear free energy relationships and the Hammett equation.

The QSAR methods can be used to predict the biological activity of a given molecule based on 'molecular descriptors'. Therefore, one can attempt to gain the desired biological activity(s) by sensibly modifying the structural parameters. The quantitative relationship between the structural parameters and the biological activity is given by multiple linear analyses,

– [4]

where is drug concentration , is a constant, is regression coefficient and is a molecular descriptor. These descriptors can be parameter such as Hansch π parameter (hydrophobicity), the Hammett σ parameter (electron-donating or accepting properties), or molar refractivity (MR) parameter (description of volume and electronic polarisability).

In silico Structure Activity Relationship methods, which are generally represented as (Q)SAR, make use of the above mentioned QSAR methods, simple *structure-activity-relationship* (SAR) and existing data. A (Q)SAR in general has three parts; the data (activity) to be modelled , the data that is used for such modelling and a method to formulate a model. The most important applications of (Q)SAR are : the rational identification of new leads with desired pharmacological or biological activity, optimisation of pharmacological and biological activity and identification of undesired effects.

7.4.2.2. Calculating Binding Free Energies

Experimentally, binding free energies of two different inhibitors to the same enzyme can be obtained from two independent experiments. However, obtaining absolute free energies from computer simulations is nearly impossible due to the sampling issues (as discussed in chapter 3). Instead, one can calculate the relative binding free energies for two more given inhibitors by using the Free Energy Perturbation (FEP) [43] method and the thermodynamics cycle given below (Figure 3) .

Even though this a computationally rigorous process, when applied to the correct situations, can yield very promising results. FEP calculations are of particular use when the ligand molecules are very similar in structure. Therefore, FEP can be used to rank inhibitors with subtle structural changes.

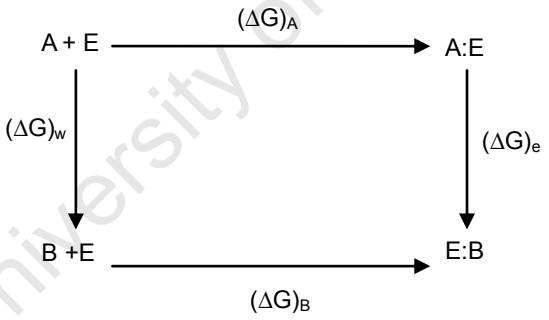


Figure 3 : A Thermodynamic cycle to calculate the relative binding energies. A and B are two different ligands binding to the enzyme E.

If A and B are two different inhibitors, the relative free energy for their association with enzyme E can be calculated by using the above thermodynamic cycle. A is computationally mutated to B in water for ΔG_w and the same mutation is repeated in the solvated binding site of E for ΔG_e. The relative free energy of binding is given by,

$$\Delta\Delta G_{binding} = \Delta G_B - \Delta G_A = \Delta G_e - \Delta G_w \quad [5]$$

7.4.3. The ADMET Process

It is a well known fact that over 50% of drug failures are due to the issues related to Absorption-Distribution-Metabolism-Elimination and Toxicology (ADMET)[44]. The role of experiments in the ADMET phase has been taken over by computational procedures to a great extent. Further, it has been shown[44] that *in silico* ADMET process has removed the bottlenecks in *in vitro* and *in vivo* methods and has provided a cost-effective , fast screening procedure. *In silico* process employs unique models to model each component: absorption, distribution, metabolism, elimination and toxicology. The reader is referred to reference 44 for further details *on in silico* ADMET process. The *in silico* predictions are then verified with experimental assays for promising candidates.

7.4.4. Pre-clinical and Clinical Trials:

Upon completion of the basic development phases which is usually demarcated by ADMET screening, the successful drug candidates enter the pre-clinical trials. Pre-clinical trials primarily involve testing on animals. The successful candidates will then enter the last phase of clinical trials, which will be done on a selected group of people.

7.5. Transition State Analogue Inhibitors for Ricin

Designing molecules that can mimic the transition state is a very promising way of making novel inhibitors[3,7,8]. The state of art of Transition State Analogue Inhibitor (TSAI) design was discussed in the previous sections. The rest of this chapter discusses the application of those techniques in designing TSAI for ricin via computational means.

The x-ray crystal structures of Ricin[45], Abrin[46] , Shiga[47] and Saporin[48] confirm that they share a common binding site as they are members of the same Ribosome Inactivating Proteins (RIP) family. It has been suggested[47,49-51] that the key amino acid residues in the binding site of these proteins include two TYRs , one ARG and one GLU.

These residues are found to be highly conserved amongst these members of the RIP family. Corresponding residue numbers of these residues for ricin , abrin, saporin and shiga are presented in Table 1.

Enzymes abrin, saporin and shiga catalyse the same chemical reaction that is catalysed by ricin and this leads to the assumption that TS states of the reactions catalysed by these enzymes will show similar chemical, geometrical and electrostatic features to the TS observed in the reaction catalysed by ricin. Extremely high similarity found in the binding sites of these enzymes and the presumed ricin-like TS suggests a common inhibitor for these enzymes. Owing to that, it also discussed here the applicability of ricin TSAI to: abrin, shiga and saporin.

A novel protocol for designing TSAI is also proposed here. The new protocol is expected to provide an efficient way to design/propose true TS analogue inhibitors. It uses the knowledge of TS, the advantage of Virtual Screening (VS) and the concept of electrostatic compatibility. This protocol is discussed next.

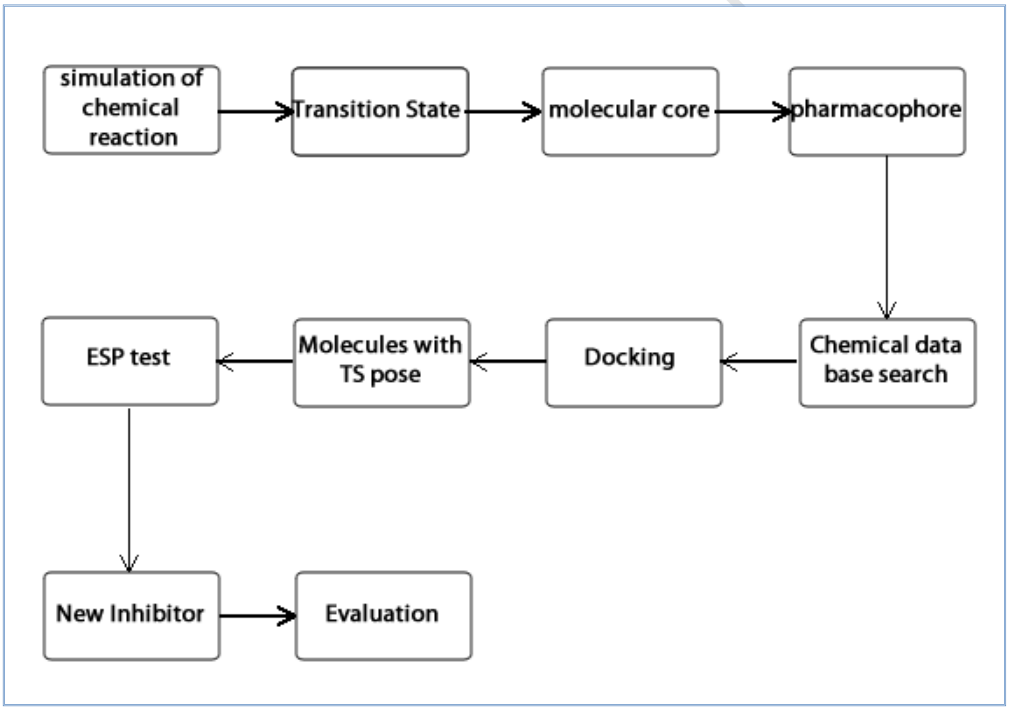
Catalytic Site Residues	Abrin	Shiga	Saporin	Ricin
ARG	ARG167	ARG170	ARG177	ARG180
GLU	GLU164	GLU167	GLU174	GLU177
TYR	TYR74	TYR77	TYR73	TYR80
TYR	TYR113	TYR114	TYR123	TYR123

Table 1. the active residues of Abrin, Shiga, Saporin and Ricin. Proposed catalytic site of these enzymes are composed of highly conserved ARG, GLU , TYR and TYR tetrad .

7.5.1. TSAI Designing Protocol – the Basics

Scheme 1 presents the protocol used in this study to design TASI for . The starting point of this novel inhibitor design protocol is the transition state(TS) structure. The transition state of ricin-catalysed reaction was discussed in chapter 6 and its structure is given in Figure 4.

It was found that at the TS , the distance of the breaking C1'-N9 bond is 2.65Å where as the distance of forming C1'-OH bond is 1.84Å. Simple structural analysis of this TS suggests that there must be an adenine (base) resembling moiety (𝔸), a ribocation resembling moiety (ℝ) and an OH^- resembling moiety (𝕆) in any inhibitor that is expected to mimic the TS.



Scheme 1 : Transition State Analogue Design Protocol as developed in the Scientific Computing Research Unit.

*** This protocol is presented in the patent license: RSA Patent Allocation No. 2011/06840 (Jayakody R.S , Naidoo, K.J) , written permission must be obtained from the authors prior to any usage of this protocol.**

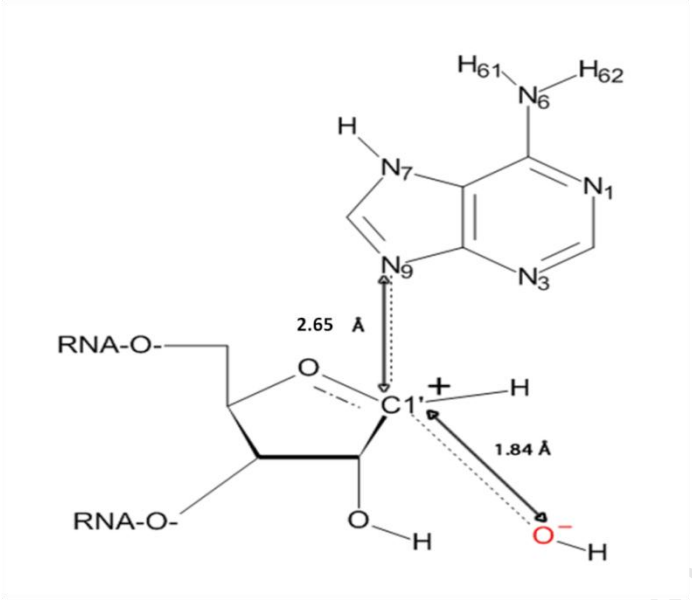


Figure 4: Transition State of the ricin-catalysed depurination.

In this study, incoming OH^- nucleophile is also included in the TS structure and it makes the TS proposed here significantly differ from the previously proposed[11,52] ricin TS structures. The significance of including OH^- in the TS was evaluated by comparing the corresponding electrostatic potential surfaces.

Electrostatic Surface Potentials (ESP) of possible TS structures i.e. TS without the OH^- nucleophile and the TS with the OH^- nucleophile are shown in Figure 5. The ESPs were calculated in gas phase with B3LYP/6-31G+(d,p) level of theory. As it can be seen from Figure 5, there are significant differences between the ESPs of two TS structures. The TS without the OH^- (Figure.5(A))can be considered as overall positive, where as the TS with OH^- can be as considered overall negative (Figure 5 (B)). The latter is expected to get highly stabilised by the positively rendered ricin binding site (see chapter 6 for more on this argument). Therefore in this study, TS given in Figure 5(B) was used.

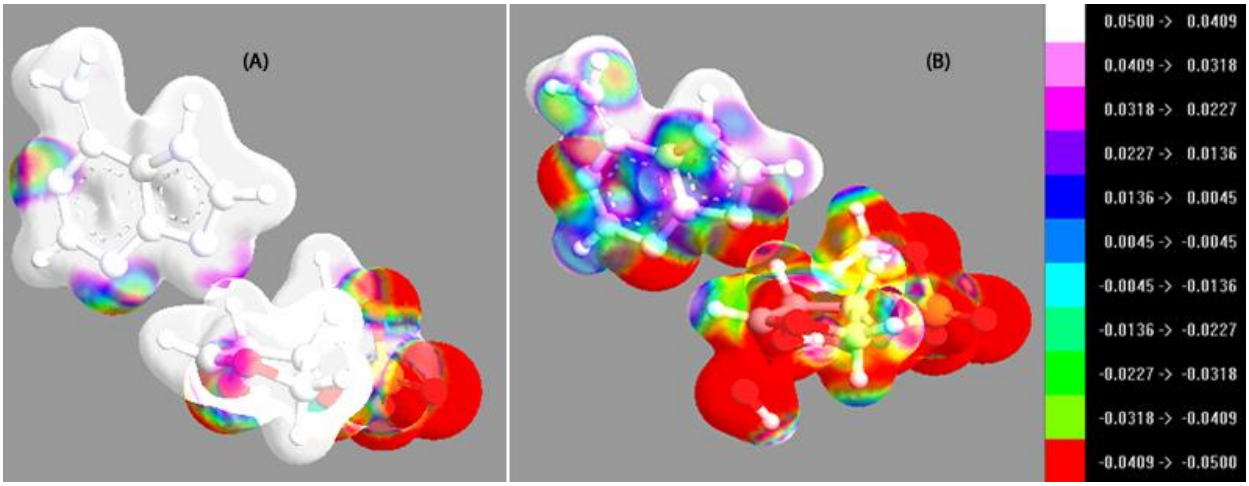


Figure 5: Significance of incoming OH^- nucleophile on the EPS of TS. (A). TS without the incoming OH^-. (B). TS with the incoming OH^-.

As mentioned earlier in this chapter, the first objective of TSAI design is to find molecules that can mimic the size, geometry and the electro-chemical properties of the real TS. To achieve this goal, a "Molecular Core" (MCORE) was introduced in this study. To ensure the preserving of electrostatic properties of the TS, ESPs were used to guide the design of MCORE.

7.5.2. TSAI Designing Protocol- the Details

Because of the patenting formalities, the reader is referred to the patent : RSA Patent Allocation No. 2011/06840 , Jayakody R.S , Naidoo, K.J., RICIN TRANSITION STATE ANALOGUE INHIBITORS, for details on the design protocol / methods.

7.6. Results and Discussion

7.6.1. The Novel Inhibitors for Ricin

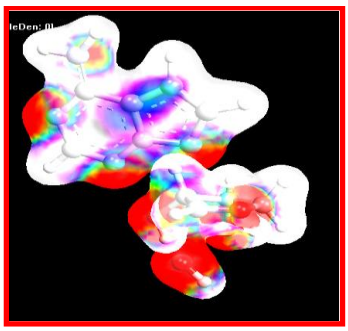
All newly designed inhibitors have same the core structure, as they were designed by using the MCORE as a scaffold. As mentioned before, in all inhibitors, there is an adenine resembling moiety, 𝔸, a ribocation resembling moiety, ℝ, and a linker "X" connecting 𝔸 and ℝ. Once again, due to the patenting formalities, the reader is referred to the above mentioned patent licence for the exact structures of the novel inhibitors. Hereafter this inhibitor are referred to as SCRU-RTAI- 1, SCRU-RTAI- 2, SCRU-RTAI- 3, SCRU-RTAI- 4, SCRU-RTAI- 5 and SCRU-RTAI- 6, where SCRU stands for Scientific Computing Research Unit, the laboratory at which this research was carried out and RTAI stands for Ricin Toxin A-chain Inhibitor.

7.6.2. Binding of Novel Inhibitors to Ricin

The ESPs of the novel inhibitors are presented in Table 2 and the initial docking results of them for binding to ricin are given in Table 3. It should be stressed here that the results presented here are preliminary results, and therefore, further investigations might be required to produce the ultimate ranking of the inhibitors. The docking results show that the new inhibitors bind stronger to the ricin binding site compared to some of the existing inhibitors. The binding of novel inhibitors were compared to that of the most widely reported ricin inhibitors; Formycin-5'-Monophosphate (FMP) a substrate analogue, 9-oxoguanisine (9OG), 7-deazaguanasine (7DG) and Guanine (product-like inhibitors). Further, the results present here have clearly demonstrated the expected correlation between an inhibitor's ability to mimic the TS-ESP and its binding affinity to the target enzyme.

The analysis of binding patterns reveals that the binding of adenine resembling moiety (𝔸) of these inhibitors resembles the binding of the adenine moiety in the natural substrate. The similarity among these binding patterns can be easily identified by the characteristic H-bonding interactions and pi-stacking interactions.

The H-bonding interactions N3-ARG180 , N1-VAL81 , H7-GLY121, and GLY121-H62 were observed in all inhibitors as well as in the natural substrate. The featuring pi-stacked binding of adenine moiety between TYR80 and TYR123 in the natural substrate was also observed with all new inhibitors.



ESP of the transition state.

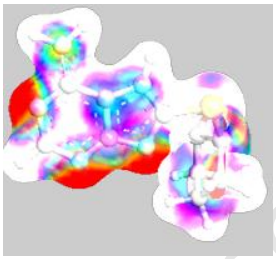
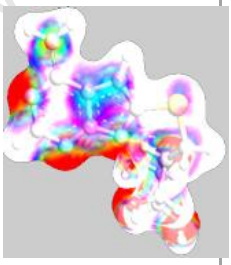
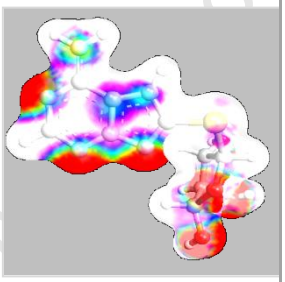
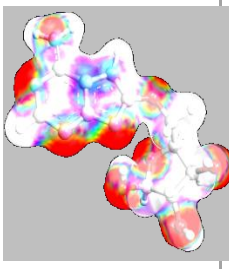
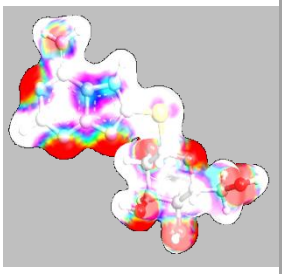
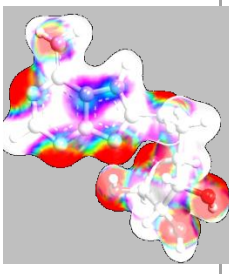
Novel Transition State Analogues Inhibitors			
Inhibitor	**ESP**	**Inhibitor**	**ESP**
SCRU-RTAI- 1		SCRU-RTAI- 4	
SCRU-RTAI- 2		SCRU-RTAI- 5	
SCRU-RTAI- 3		SCRU-RTAI- 6	

Table 2: ESP of Novel Ricin Inhibitors. SCRU stands for Scientific Computing Research Unit, laboratory at which this research was carried out and RTAI stands for Ricin Toxin A-chain Inhibitor.

In inhibitors SCRU-RTAI-1, SCRU-RTAI-2, SCRU-RTAI-3, SCRU-RTAI-4 and SCRU-RTAI-5, the hydroxyl hydrogen of TYR123 donates a H-bond to the corresponding hydrogen bond acceptor atom in the linker X. This H-bond interaction is considered as an additional advantage of having a H-bond accepting atom in 𝔸-ℝ link. Hence it is proposed that any future inhibitor should have atoms of H-bonding accepting elements in the linker to promote such favourable H-bonding interactions.

As discussed above, all inhibitors show very similar binding in regards to . Therefore, it is argued here that the differences in the binding pattern in ℝ are accountable for the observed differences in the binding scores. From the analysis of the binding pattern of the ribose-like moiety (ℝ) and the docking score, it is proposed here that the presence of hydroxyl groups at 1st , 2nd and 3rd positions on ℝ enhances the binding to the enzyme. The corresponding amino acid residues that can strongly H-bond with these OH groups are GLU177, GLU208 and ASN209 respectively. Therefore, presence of OH groups at 1st , 2nd and 3rd positions on ℝ is recommended as a required feature in any ricin inhibitors to be designed.

It was observed the inhibitors with a longer link (SCRU-RTAI-4 and SCRU-RTAI-5) show the highest binding score. It is argued here the flexibility in the longer linkage enhances the alignment of the –OH groups on ℝ for effective H-bonding. Therefore, a two-atom-long-link is proposed for any future ricin inhibitors that could be designed from the proposed MCORE.

The best inhibitor invented in this study is SCRU-RTAI- 4 , which shows a binding score of -12.34. Although, further studies are required to rationalise its high potency, it is argued here that any potent ricin inhibitor must have 5-membered sugar resembling moiety with hydroxyl groups at 1, 2 and 3 positions. Additionally, it must also have a longer (two atom) “X” link.

Known Inhibitors			
Inhibitor	***Docking score***	***IC 50 (nM)***	
FMP	-9.69		
9OG	-5.54	4.00E+05	
Guanine	-4.28	9.00E+05	
7DG	-2.30	2.80E+06	
Novel Transition State Analogues Inhibitors			
New Inhibitor	***Docking Score***	***New Inhibitor***	***Docking Score***
SCRU-RTAI- 1	-11.33	SCRU-RTAI- 4	-12.34
SCRU-RTAI- 2	-11.66	SCRU-RTAI- 5	-12.26
SCRU-RTAI- 3	-12.05	SCRU-RTAI- 6	-11.00

Table 3: Docking Score of novel ricin inhibitors for binding to ricin. For comparison, docking scores of some known inhibitors and the corresponding IC_{50} values are also presented.

7.6.3 Binding of Novel Inhibitors to other Rips

Abrin , Shiga and Saporin are members of the RIP family that are very similar to ricin. The similarities are seen in the primary sequence as well as in their tertiary structure. From their x-ray crystal structures (Figure 6), it can be seen that they have exactly the same binding site as ricin. Particularly, the proposed ARG-GLY-TYR-TYR catalytic tetrad of these enzymes has an identical orientation to that of ricin. Therefore, it is argued here that the novel inhibitors should bind effectively to these enzymes.

Molecular docking calculations were used to investigate the binding of these new inhibitors to Abrin, Shiga, and Saporin. In docking calculations, the crystal structures of the enzymes were used. All docking calculations were performed with high precision docking algorithms. The results of docking calculations are presented in Table 4.

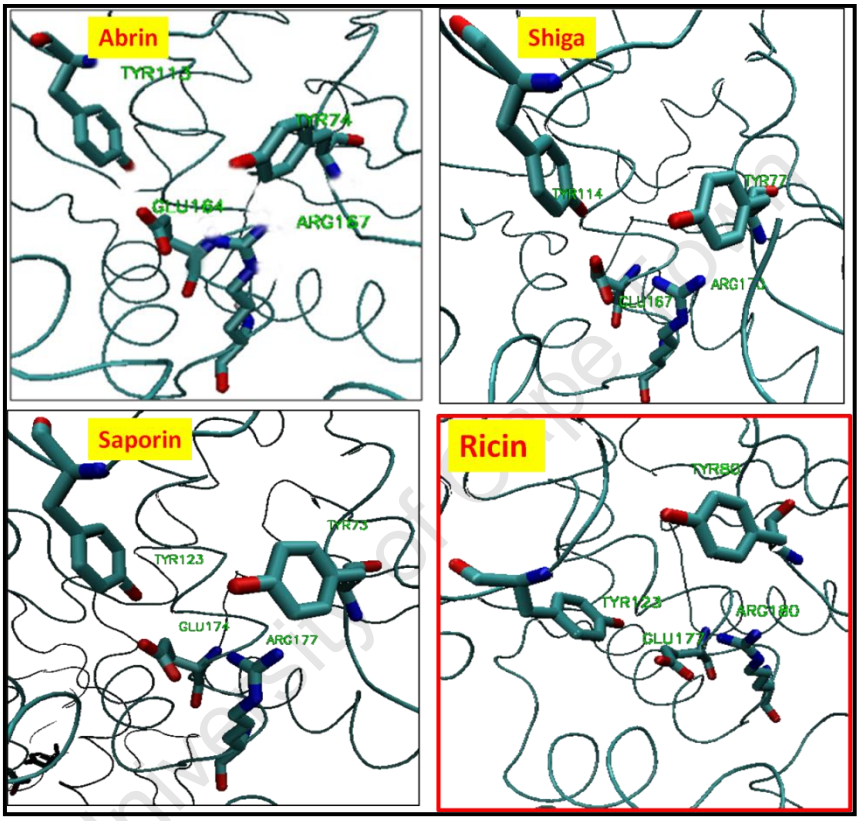


Figure 6: Proposed Catalytic Residues of Abrin, Shiga, Saporin and Ricin

Inhibitor	Enzyme		
	Abrin	Shiga	Saporin
	Docking Score (Kcal/mol)		
SCRU-RTAI-1	-6.96	-8.70	NP
SCRU-RTAI -2	-6.62	-7.73	-6.55
SCRU-RTAI -3	-6.86	-7.09	-6.60
SCRU-RTAI -4	-6.53	NP	-8.50
SCRU-RTAI -5	-6.67	NP	-7.71
SCRU-RTAI -6	-7.01	NP	-6.27

Table 4: Binding of novel ricin TSAIs to Abrin, Shiga and Saporin. NP indicates the absence of a pose that is similar to the poses of the TS for ricin.

It can be seen that that inhibitors SCRU-RTAI-2 and SCRU-RTAI-3 are capable of binding to Abrin, Shiga and Saporin with a pose similar to that of the ricin-TS. Further, these two inhibitors bind to ricin with a docking score of -11.66 kcal/mol and -12.05 kcal/mol respectively. Therefore, it can be concluded that these two inhibitors are "universal inhibitors" for the members of the RIP family studied; namely, ricin, abrin, shiga and saporin. Structure analysis of SCRU-RTAI-2 shows that it has the greatest similarity to the TS structure of the ricin-catalysed reaction. Therefore, the finding of this study provide clear evidences to support the argument of applicability of a TSA inhibitor for one enzyme in a family of enzyme, to the other members of the same family.

However, relatively low docking scores for new inhibitors were observed for these enzymes compared to ricin. It is assumed that this might be due to the fact that these enzyme structures were not relaxed in the solution. Therefore, it is expected to see improved binding scores with the relaxed enzymes structures of these enzymes and further studies are required to for this purpose.

7.7. Concluding Remarks

Transition State Analogue inhibitors for ricin were proposed in this study. Those inhibitors were docked in to the ricin binding site using extra precision docking algorithms. The results showed that the novel inhibitors proposed in this study have a very high binding score compared to some of the previously proposed inhibitors.

A novel approach was proposed in this study to designing the TSA inhibitors. Designing of a pharmacophore based on a molecular core that resembling the TS, has not been reported before. This study has clearly demonstrated that the ability of new molecules to mimic the ESP of the identified TS can be use to guide the designing of new TSA inhibitors. Further, application of these inhibitors to other similar enzymes suggests that TSA inhibitors for one enzyme can be used to inhibit the other enzymes that catalyse the same chemical reaction. The TSA inhibitors proposed here differ from the previously proposed TSAIs in such way that there is no linkage between the N9 position of the adenine resembling moiety and the C1' position of the sugar resembling moiety in any of the newly proposed inhibitors.

Inhibitors SCRU-RTAI-2 and SCRU-RTAI-3 showed desired binding to abrin, shiga, saporin and ricin. Therefore, these two inhibitors are proposed as "common" inhibitors for these enzymes. Further, SCRU-RTAI-2 is proposed as the best candidate for further developments as it has the greatest similarities to the TS of the reaction catalysed by ricin.

7.8. References:

(1) Pauling, L. *American Scientist* **1948**, *36*, 51.
(2) Schramm, V. L. *Annual review of biochemistry* **2011**, *80*, 703.
(3) Schramm, V. L. *Annual Review of Biochemistry* **1998**, *67*, 693.
(4) Wolfenden, R. *Accounts of Chemical Research* **1972**, *5*, 10.
(5) Fersht , H. *Structure and Mechanism in Protein Science*; 3rd ed.; W.H. Freeman and Company; New York, 1999.
(6) Jackson, M. B. *Molecular and cellular biophysics*; Cambridge University Press, 2006.
(7) Schramm, V. L. *Accounts of chemical research* **2003**, *36*, 588.
(8) Schramm, V. L. *The Journal of biological chemistry* **2007**, *282*, 28297.
(9) Bigeleisen , J.; Mayer, M. G. *Journal of Chemical Physics* **1947**, *15*, 261.
(10) Schramm, V. L. *Current Opinion in Chemical Biology* **2001**, 556.
(11) Ho, M.-C.; Sturm, M. B.; Almo, S. C.; Schramm, V. L. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 20276.
(12) Zewail, A. H. *Science* **1988**, *242*, 1645.
(13) Merz, K. M.; Ringe, D.; Reynolds, C. H. *Drug Design: Structure- and Ligand-Based Approaches*; Cambridge University Press, 2010.
(14) Morrison, J. F.; Walsh, C. *Molecular Biology* **1988**, *61*, 201.
(15) Mader, M. M.; Bartlett, P. A. *Chemical Reviews* **1997**, *97*, 1281.
(16) Mader, M. M.; Bartlett, P. A. *Chemical Reviews* **1997**, *97*, 1281.
(17) Kyte, J. *In Structure in Protein Chemistry;*; Garland Publishing,Inc.: New York, 1995.
(18) Tanford, C. *The Hydrophorbic Effect*; Wiley-Inter-science: New York, 1980.
(19) Morgan, B. P.; Scholtz, J. M.; Ballinger, M.; Zipkin, I.; Bartlett, P. A. *Journal of American Chemical Society* **1991**, *113*, 297.
(20) Warshel, a. *The Journal of biological chemistry* **1998**, *273*, 27035.
(21) Murray, J. S.; Sen, K. D. *Molecular electrostatic potentials: concepts and applications*; Elsevier, 1996.
(22) EWING, T. J. A.; KUNTZ, I. D. *Journal of Computational Chemistry* **1996**, *18*, 1175.
(23) Caflisch, A.; Niederer, P.; Anliker, M. *Proteins* **1992**, *13*, 223.
(24) Nola, A.; Roccatano, D.; Brendensen, H. J. C. *Proteins* **1994**, *19*, 174.
(25) Goodsel, D. S.; Olson, A. J. *Proteins***1990**, *8*, 195.
(26) Moon, J. B.; Howe, W. J. *Proteins* **1991**, *11*.
(27) Abagyan, R.; Totrov, M.; Kuznetsov, D. *Journal of Computational Chemistry* **1994**, *15*, 448.
(28) Leach, A. R.; Kuntz, I. D. *Journal of Computational Chemistry* **1992**, *13*, 730.
(29) Leach , A. R. *Journal of Molecular Biology* **1994**, *235*, 345.
(30) Mizutani, M. Y.; Tomioka, N.; Itai, A. *Journal of Molecular Biology* **1994**, *243*, 1994.
(31) Smellie, A. S.; Crippen, G. M.; Richards, W. G. *Journal of Chemical Information and Computer Sciences* **1991**, *31*, 386.
(32) Judson, R. S.; Jaeger, E. P.; Treasurywala, A. M. *Journal of Molecular Structure* **1994**, *114*, 191.
(33) Pang, Y. P.; Kozikowski, A. P. **1994**, *Journal of Computer Aided Molecular Design*, 683.
(34) Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. *Proceedings of the National Academy of Sciences* **1992**, *89*, 2159.
(35) Fischer, D.; S. L. Lin; Wolfson, H. L.; Nussinov, R. *Journal of Molecular Biology* **1995**, *248*, 459.
(36) Kuntz, I. D.; Blaney, J. M.; Oatley, R. *Journal of Molecular Biology* **1982**, *161*, 269.
(37) Miller, M. D.; Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P. *Journal of Computer-Aided Molecular Design* **1994**, *8*, 153.

(38) Bohm, H. J. *Journal of Computer-Aided Molecular Design* **1994**, *8*, 623.
(39) Jensen, F. *Introduction to computational chemistry*; John Wiley & Sons, 2007.
(40) Hubbard, R. E.; Chemistry, R. S. o. *Structure-based drug discovery: an overview*; Royal Society of Chemistry, 2006.
(41) Hansch, C.; Muir, R. M.; Fujita, T.; Maloney, P. P.; Geiger, E.; Streich, M. *Journal of American Chemical Society* **1963**, *85*, 2817.
(42) Fujita, T.; Iwasa, J.; Hansch, C. *Journal of American Chemical Society* **1964**, *86*, 5175.
(43) Zwanzig, R. W. *Journal of Chemical Physics* **1954**, *22*, 1420.
(44) Testa, B.; Turski, L. *Virtual ADMET assessment in target selection and maturation*; IOS Press, 2006.
(45) Montfort, W. *Journal of Biological Chemistry* **1987**, *262*, 5398.
(46) Tahirov, T. H.; Lu, T. H.; Liaw, Y. C.; Chen, Y. L.; Lin, J. Y. *Journal of Molecular Biology* **1995**, *250*, 354.
(47) Fraser, M. E.; Fujinaga, M.; Cherney, M. M.; Melton-Celsa, A. R.; Twiddy, E. M.; O'Brien, A. D.; James, M. N. G. *Journal of Molecular Biology* **2004**, *279*, 27511.
(48) Savino, C.; Federici, L.; Ippoliti, R.; Lendaro, E.; Tsernoglou, D. *FEBS Letters* **2000**, *470*.
(49) Fiorenzo, S. *Toxicon* **2004**, *44*, 371.
(50) Stirpe, F.; Battelli, M. *Cellular and Molecular Life Sciences* **2006**, *63*, 1850.
(51) Michael, J.; Roberts, M.; Robertust, J. O. N. D. *The FASEB Journal* **1994**, *8*, 201.
(52) Chen, X. Y.; Link, T. M.; Schramm, V. L. *Biochemistry* **1998**, *37*, 11605.

CHAPTER -8

CONCLUSIONS

The primary objective of this thesis was to elucidate the mechanism of the depurination reaction catalysed by toxic ricin. FEARCF, a flat histogram free energy method was used to produce the free energy surface of this reaction. For the first time, the mechanism of ricin-catalysed reaction was discovered in this study. Therefore, it is believed to be a major milestone in ricin chemistry. The newly proposed mechanism agreed exceptionally well with the evidence from the previously performed experiments.

A novel substrate model for ricin was proposed in this study. This model, the Loop Substrate Model (LSM) was compared with a previously used model, the Minimum Substrate Model (MSM). The comparison study clearly showed that the MSM cannot be used in simulations due to its inability to mimic the natural substrate accurately. In comparison the LSM demonstrated the expected binding pattern to ricin and the puckering of sugar ring that was consistent with previous experimental studies. Moreover, it was found that the binding of LSM is dominated by electrostatic interactions, where this finding is in very good agreement an electrostatically-driven binding of the targeted GAGA loop to ricin.

In this study it was shown that the target adenine by ricin must be pre-protonated at the N7 position prior to the initialisation of the catalytic mechanism. Further, it was shown that the target adenine cannot be pre-protonated at N1 position for the reaction to proceed. Moreover, the protonation of N7 convincingly promoted C1'-N9 bond cleavage and on development of the anomeric carbon (C1') as a nucleophilic centre. Therefore this study successfully addressed the question of the extent and nature of pre-protonation of ricin's natural substrate.

The transition state structure of the ricin-catalysed reaction was clearly identified in this study. Its molecular structure, conformation and electron density features were used to design transition state analogue inhibitors (TSAI) for ricin. A novel TSAI designing protocol was also proposed. In this novel protocol, a molecular core (MCORE), which is a true, mimic of the TS is used to design a pharmacophore. When searching databases for hit molecules, resulting molecules were supposed match both MCORE and pharmacophore criteria simultaneously. It was shown that this searching strategy reduces the number of hits resulting from database search dramatically. Therefore, it was concluded that the proposed novel protocol of database searching is more effective than the conventional pharmacophore-based search. In order to further filter the hit molecules, their electrostatic surface potentials were compared with that of the transition state. This ESP test ensures the new hit molecule's ability to mimic the electrostatic features of the TS. Moreover, the importance of electrostatic complimentarity between TS and the TSAI was clearly shown in this study.

The inhibitors designed in this study were tested against ricin and other similar enzymes. Initial docking results showed that the novel inhibitors are highly effective against ricin than the existing inhibitors. Moreover, it has been shown that those novel inhibitors are also effective against other similar enzymes, shiga, saporin and abrin.

In summary, the contributions from this study to the field of ricin chemistry are fourfold. Firstly, a substrate model for ricin's natural substrate is proposed. Secondly, the pre-protonation state of ricin's target adenine has been clearly identified. Thirdly, and most importantly, the mechanism of ricin-catalysed reaction was discovered. Finally, novel ricin inhibitors (transition state analogues) were discovered and their potency against ricin was proved.