MOLECULAR MECHANISM OF DRUG-DNA INTERCALATION AND THE ROLE OF WATER IN MOLECULAR RECOGNITION

A THESIS

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS

OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

WILBEE D. S.

20103059

DEPARTMENT OF CHEMISTRY

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH, PUNE – 411 008

June 2015

Dedicated to…

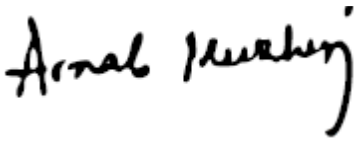
My Family Members, Teachers and Friends

भारतीयविज्ञानशिक्षाएवंअनुसंधानसंस्थान, पुणे
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH (IISER), PUNE
(An Autonomous Institution, Ministry of Human Resource Development, Govt. of India)
900 NCL Innovation Park Dr HomiBhabha Road Pune 411008

Arnab Mukherjee, Ph.D.
Associate Professor – Chemistry
www.iiserpune.ac.in/~arnab.m/

CERTIFICATE

Certified that, the work incorporated in the thesis entitled, "***Molecular Mechanism of Drug-DNA Intercalation and the Role of Water Entropy in Molecular Recognition***" submitted by ***Wilbee D. S.*** was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or institution.



Date: 29th June 2015
Pune (MH), India

Dr. Arnab Mukherjee
(Thesis Supervisor)

DECLARATION

I declare that this written submission represents my ideas in my own words and where others' ideas have been included; I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Date: 29th June 2015 **Wilbee D. S.**

Pune (MH), India **Reg: No: 20103059**

Acknowledgements

I am deeply indebted to acknowledge my heartfelt gratification at the completion PhD research work to all who have helped me in this endeavor. First and foremost, I take this opportunity to express my sincere gratitude to my research advisor, **Dr. Arnab Mukherjee**, for his excellent guidance, suggestions, encouragement and support given throughout my research work. First of all I would like to thank him for giving me an opportunity to work with him. As I was the first student of his lab, he did his duty both as a research advisor and senior at the same time. I am grateful to him for teaching me early lessons of computation and I am also indebted to him for long hours of scientific discussions which actually made me to think about how to solve a problem. The encouragement and patience shown by him throughout my research career improved my confidence and made me more comfortable and enthusiastic to do hardcore computational studies. I am deeply thankful to him for giving me enough freedom to work without any pressure throughout the research work.

It gives me profound pleasure to express my indebtedness and respect to my research advisory committee (RAC) members, **Dr. Arun Venkatnathan** (IISER Pune) **and Dr. Sudip Roy** (NCL Pune) for their valuable suggestions, appreciation and criticism given during the RAC meetings which enabled me to rectify the shortcomings of the work.

I acknowledge my deep gratitude to **Prof. K. N. Ganesh**, Director, IISER Pune, for providing excellent computational and research facilities at IISER Pune for carrying out the research work. I also thank him for giving me some opportunity to design artwork for him.

It gives me immense pleasure to extend my sincere thanks to **Dr. M. Jayakannan** (IISER Pune) for the help and encouragement given throughout my life at IISER Pune. He and his family members always gave me enough consideration that made me feel that I am not away from home.

I sincerely appreciate the collaborations with **Dr. Partha Hazra** (IISER Pune) and his lab members and I extend my sincere gratitude to the valuable suggestions given by him during the collaborative research work.

Many thanks to all computational research faculties (**Dr. Anirban Hazra, Dr. Prasenjith Ghosh** and **Dr. Srabanti Chaudhury**) for their valuable suggestions and support extended to me for the last five years at IISER Pune. Special thanks to all faculties of chemistry department especially **Dr. S. G. Srivastan, Dr. Harinath Chakrapani, Dr. Pinaki Talukdar, Dr. R. G. Bhat, Aloke Das, Dr. V. G. Anand, Dr. A. A. Natu, Dr. Mrinalini Puranik and Dr. Seema Verma** for their support and consideration.

I sincerely appreciate the help and technical support given by the members of the IT department especially **Mrs. Neeta, Mr. Suresh** and **Mr. Sachin**.

Words cannot suffice the deep gratitude I feel towards my lab mates for their love, help, support, suggestions, criticism and memorable moments given throughout my life at IISER. I extend my sincere gratitude to my labmates **Shreyas Supekar, Piyush Agarwal, Mandar Kulkarni,** and **Debasis Saha** for their everlasting care and encouragement. I deeply thank my best friends and labmates, **Reman Kumar Singh** and **Hridya V. M.,** for their continuous support, care, help, patience and scientific discussions to make my life much easier at IISER, Pune.

I greatly acknowledge my gratitude to all friends in computational chemistry for their direct and indirect support during my research work.

I take this opportunity to say my heartfelt thanks to another best friend **Minal More** and **her family members** for their continuous support, kindheartedness and encouragement.

Many thanks to my other friends, **Rohit Kumar, Niharika Joshy, Anand raj, Dharmaraja, Anurag Sunda, Biplab Jorder, Arun Tanpure, Subhamita Sarkar, Mubeena Bano, Rajeswari, Asmi, Reshma, Anju, Mahima Goel, Balamurugan, Mani, Mano, Ravi,** and **Shankar** for their presence and help at various occasions. I also thank all staff members in administration, finance, accounts, library, canteen, and transportation especially, **Mrs. Anuradha, Mr. Thushar, Ms. Naina, Mr. Mayuresh,**

and **Mr. Ankush** for their immediate help whenever I needed them. Financial support from IISER Pune is greatly acknowledged.

I would like to acknowledge my deep thanks and gratitude to my life partner, **Dr. Pramod P. S.** his love, affection, support, patience and unwavering belief in me to complete this long journey. I am also thankful to my son **Pranav W. P.** for giving me enough time to prepare for thesis defense.

Moreover, I am deeply indebted to my **Parents (K. Devadasan and D. Sasikala)**, my brothers (**Wins D. S., Winson D. S., and Windyson D. S.**), **Pramod's parents**, his brother (**Sandeep**) and **family** for their love, support, care and encouragement given during my whole life.

I remain grateful to anyone who may have either indirectly or directly helped me in anyways for the successful completion of my research work.

In the very last, I thank **almighty God** for giving me enough strength and perseverance to complete the Ph. D work.

Wilbee D. S.

Synopsis

In molecular biology, the specific non-covalent interaction between molecules, termed as ‘molecular recognition’, is vital to any biological process. Out of different kinds of such interactions, the most explored one is the ligand-macromolecule interaction where small molecules can modulate the conformation and structure of biomacromolecules thereby help to fight against various diseases. One such molecular recognition process is the interaction of small molecules with nucleic acids having potential applications in the treatment of cancer and microbial infections. Out of various binding modes of interaction of small molecules with DNA, a unique non-covalent mode of interaction is called ‘intercalation’ characterized by the insertion of planar aromatic molecule between two adjacent base pairs of a nucleic acid. This interaction causes structural changes to the DNA, thereby arresting further DNA replication and transcription leading ultimately to cell death. Therefore, the major therapeutic applications of intercalation processes lie in the treatment of cancer, microbial, and viral infections. Intercalation is also prevalent in protein-DNA interactions that help in the specific recognition of protein to DNA. An enormous number of studies were devoted to characterize the structure and dynamics of this unique binding mode experimentally. However, little progress has been made to understand the molecular mechanism of the process despite its possible significance in drug design strategy to combat cancer. Since the experimental methods are limited by techniques to explore faster intermediates and molecular pathway of the intercalation process, this thesis uses computational strategy to unveil the molecular detail of intercalation process and associated thermodynamics and kinetic aspects of one such intercalating anti-cancer agent, proflavine.

As mentioned above, intercalation process is a molecular recognition process. Any such process has three major components -- guest, host, and solvent. Water is the natural solvent in most biological reactions. It is capable of modulating the structure and function of other components. Often, water acts as the driving force in a molecular recognition process through its unique enthalpic and entropic interplay. A well-studied example of such interplay is protein-ligand interactions. Protein cavities or interfaces are

heterogeneous, occupied primarily by non-polar region (75%) and to a lesser extent by polar region (25%). The water molecules present in those cavities or interfaces form highly structured hydrogen bonded network due to exposure to non-polar regions. Sometimes even a single water molecule help to stabilize the system by forming multiple hydrogen bonding interactions with neighboring polar amino acids. The free energetically less favorable water network in the hydrophobic regions of the protein cavity or interface drives non-polar intermolecular interactions termed as 'hydrophobic effect'. During this molecular recognition process, the localized water molecules in the cavity are displaced by the ligand or by other biomolecules leading to the release of water molecules into the bulk. This results in a positive entropy change favoring the binding process if the displaced water is of lower entropy. In this context, the calculation of single water molecule entropy is important to identify the entropy of each water molecule covering the binding site. This will help in favorable ligand modification to displace only the low entropy water molecules and steer the binding process towards lower free energy. This approach is different compared to many other computational methods where the ligand is developed based on its binding to proteins, and the role of water is neglected. In many cases, this leads to an unsuccessful drug design process. In the context of the present thesis, water entropy may be important. Therefore, some part of the thesis is devoted to develop a new method to calculate single water entropy and test it on various systems including a well-studied protein-ligand binding system with a goal to use it for intercalation process in future.

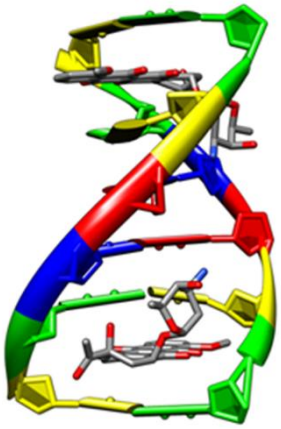
This thesis is divided into five major chapters:

1. ***Introduction Chapter:*** First chapter describes intercalation process in detail including various experimental and computational methods used to characterize various aspects of intercalation process.
2. ***Mechanism of Direct Intercalation of Proflavine to DNA:*** This chapter discusses the mechanism of intercalation of a prototype DNA intercalator proflavine, which is a simple planar molecule and an anthracycline derivative. Investigation of the direct intercalation mechanism from outside-bound state brings out the sequence of the events that take place during intercalation. This

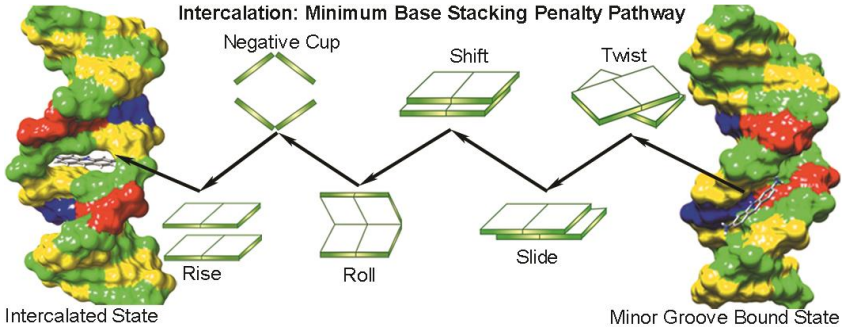
chapter answers a long-standing question about whether the intercalation happens through natural fluctuation of the DNA or the drug (proflavine) induces a cavity in DNA to intercalate.

3. ***Pathway of Intercalation Process:*** This chapter unveils the complete free energy profile and mechanism of proflavine intercalation process. It explores the structure and dynamics of deintercalation through major and minor grooves, intercalation through major and minor grooves, and dissociation from minor and major grooves of the DNA. This section also reveals the dynamics of proflavine molecule in the intercalated state. Moreover, the detailed thermodynamics and kinetic aspects are calculated and compared with experimental results. The complete free energy profile and associated kinetic pathway reveals an interesting intercalating pathway of proflavine in exquisite detail.
4. ***Structure and Dynamics of Outside Bound State of Intercalation Process:*** Fourth chapter deals with the structure and dynamics of the pre-intercalative outside-bound state of proflavine molecules to DNA at higher phosphate to dye ratio where experimentally outside-bound state is favored. The study shows that up to a very high ion concentration, the kinetics of pre-intercalative bound state formation is six orders of magnitude faster compared to that argued based on experimental kinetics. This chapter, therefore, provides a new perspective for experimental two-step intercalating kinetic results.
5. ***Entropy of Water Molecule at Single Molecular Level:*** This chapter concentrates on the development of a method to calculate the rotational and translational entropy of a single water molecule around various molecules and at bulk. This chapter also includes the calculation of the entropy of water molecules around different hypothetical solutes characterized by the hydrophobic nature of varying sizes and polar nature of positive and negative charges. This chapter also describes the application and validation of the new method of calculation of the entropy of water molecules to understand the role of water molecules in various molecular recognition processes.

We will now provide the overall summary of the whole thesis.

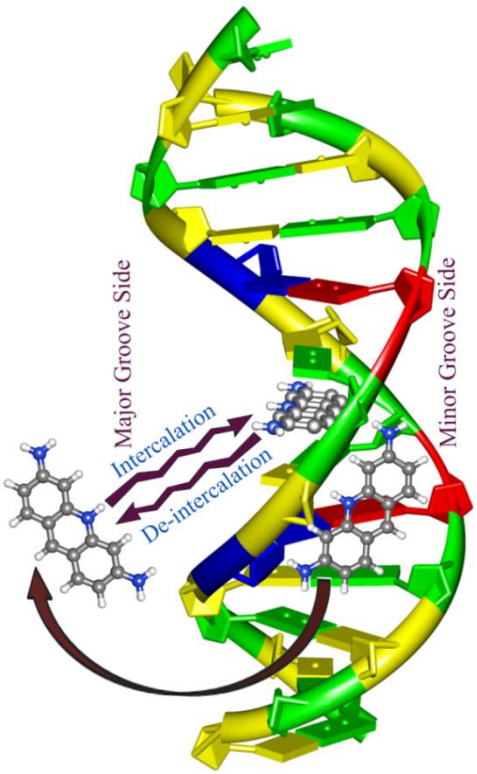


Chapter 1 describes the examples for various DNA intercalators, its classification, and biological significance. This chapter also explains various kinetic and thermodynamic experiments performed to understand the structure, energy parameters, the timescale of the formation of intercalated state and pathway of intercalation process. Chapter 1 briefly explains different computational techniques used to understand various structural aspects of the intercalated state. This chapter also explains different computational methods adapted to understand the molecular mechanism and pathway of intercalation process discussed in the forthcoming chapters.



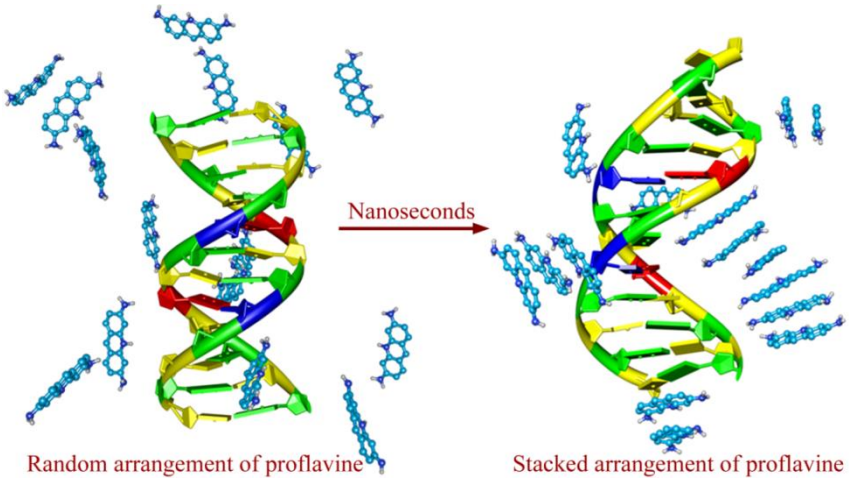
Chapter 2 addresses a long-standing debate on the intercalation mechanism and thus provides an insight to design more potent DNA intercalators. The flat aromatic intercalators have inherent fluorescent property, thus enabling various spectroscopic methods to probe into different thermodynamic and kinetic aspects of the process. The intercalation process also has two limiting hypothesis: one in which DNA base pairs open up momentarily close to the bound drug allowing it to enter and the other in which the drug induces a cavity in the DNA without requiring prior base pair opening. The intercalated state is the unique bound state (similar to the native state of a protein) while there can be several possible unbound states (similar to unfolded protein configurations). Experimentally, the overall timescale of intercalation is shown to occur in milliseconds. Therefore, direct intercalation process is extremely challenging to capture in computational studies. To understand the intercalation process at the molecular level, we chose a simple, flat, symmetric and aromatic DNA intercalator, proflavine. The simple and symmetric nature of proflavine makes the molecule equally accessible by both

minor and major groove of the DNA, thereby representing different varieties of DNA intercalators and complex intercalation pattern.With appropriate configurational restraint (to prevent dissociation) in all-atom metadynamics simulations, we captured the free energy surface of direct intercalation from the minor groove-bound state for the first time using proflavine. Mechanism along the minimum free energy path reveals that intercalation happens through a minimum base stacking penalty pathway where non-stacking parameters (Twist->Slide/Shift) change first, followed by base stacking parameters (Buckle/Roll->Rise). This mechanism defies the natural fluctuation hypothesis and provides molecular evidence for the drug induced cavity formation mechanism. The thermodynamic origin of the barrier is found to be a combination of entropy and desolvation energy.



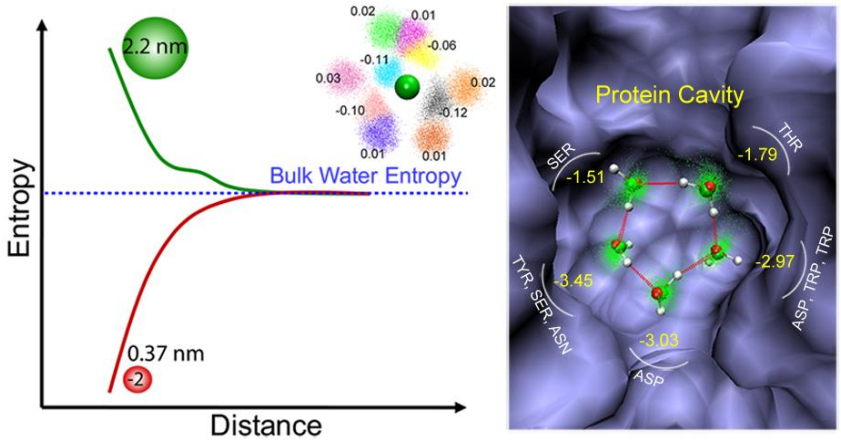
Since intercalation is a clinically relevant biophysical process, extensive kinetic and thermodynamic studies have followed since the discovery of intercalative binding mode. However, the molecular mechanism and the origin of thermodynamic and kinetic profile of the process are still not clear. Most of the kinetic studies on intercalation of proflavine, propose that the process comprises two steps: a fast bimolecular outside binding (faster than a few tenths of millisecond), followed by a slow intercalation (range of a few milliseconds). This observation was found to be independent of the nature of DNA as well as the nucleotide sequence. Based on the kinetics experiments, it was hypothesized that the outside bound state formation is a diffusion controlled process, and the dissociation of the drug from the outside-bound state is an activated process with a high free energy barrier. The timescale for overall intercalation obtained from experiments indicated that it is also a barrier driven process with high activation energy(~12 kcal/mol). Higher barrier (~15 kcal/mol) for

deintercalation of the drug makes the deintercalation even slower. ***Chapter 3*** describes the construction of free energy landscape of intercalation, deintercalation and dissociation from both the major and minor grooves of DNA using extensive all-atom well tempered metadynamics simulations, capturing both the free energy barriers and stability in close agreement with fluorescence kinetic experiments. In the intercalated state, an alternate orientation of proflavine is found with almost equal stability compared to the crystal orientation; however, separated by 5.0 kcal/mol barrier that decreases as the drug approaches the groove edges. This study provides a comprehensive picture of the intercalation process indicating that that intercalation and deintercalation of proflavine happen through the major groove side. The effective intercalation barrier increases because the path of intercalation goes through the stable (abortive) minor groove bound state, making the process a millisecond long one, in excellent agreement with experiment. Molecular origin of the higher barrier for intercalation from the minor groove side is attributed to the desolvation energy of the DNA and loss of entropy, while the barrier from the major groove, in the absence of desolvation energy, is primarily entropic.



The first step in the intercalation process for most of the intercalators is the formation of a metastable outside bound state or pre-intercalative bound state. This is also the fastest process compared to the intercalation process. However, experiments so far have failed to capture the nature of the outside bound state. The timescale of this association observed using fluorescence kinetics, is in the range of sub-milliseconds. Even though the structure of the outside-bound state at the pre-intercalative stage is not known, it is believed that the stacking between the dye molecules bound externally to the polymer chain plays an important role in the dynamics of the intercalation process. Moreover, we

have shown recently, as described in chapter 3, how a pre-intercalative bound state affects the kinetics of the subsequent intercalation process. We have also shown the mechanism of two-state binding of proflavine where it initially forms a minor groove-bound state in diffusion limited free energy downhill process, followed by intercalation through the major groove. Smoluchowski equation for downhill diffusion process at P/D ratio 22 (one drug in a twelve base pair DNA) showed that the timescale for this association should be in the order of nanoseconds which is contrary to the above-mentioned experimental values (sub-milliseconds). In ***chapter 4***, we have performed molecular dynamics simulations with multiple proflavine molecules to study the structure and dynamics of the formation of the pre-intercalative outside bound state. We show that the timescale of outside bound state formation is at least five orders of magnitude faster (in nanoseconds) than that reported in experimental kinetic studies. Moreover, we also observed the stacked arrangement of proflavine all around DNA, which is different from the experimentally predicted stacking arrangement perpendicular to the helical axis of DNA in the close vicinity of the phosphate groups. This study gives an insight into the molecular structure and dynamics of the pre-intercalative bound state – the crucial step preceding intercalation.



As mentioned above, the binding free energy in molecular recognition processes is often modulated by water where the entropy of water plays an important role, also known as the hydrophobic effect (HE).HE depends on both the size and nature of the solute. For small solutes like methane at room temperature, larger negative entropy change over smaller favorable enthalpy leads to this phenomenon. This entropy loss for water around small hydrophobic solute like methane is argued to originate from the urge of the solvent to

maintain necessary hydrogen bonds around the solute, leading to enhanced order. Since overall water entropy must have a contribution from individual water molecules constituting the solvent shell, it is therefore intriguing to address entropy of every single water molecule itself in the solvation shell or biomolecular cavity. It is therefore of particular interest to estimate the entropy of each water molecule for drug design as even a single water molecule can change the overall free energy of a molecular recognition process. Studies on various protein-ligand complexes show the presence of highly localized anchoring water molecule between protein surface and ligand. Therefore, ligand modifications are often geared towards displacing the low entropy water molecule from the binding sites to bulk to favor the binding process of ligand. The entropic behavior of individual water, however, was not addressed until recently where Friesner*et al.* calculated translational and rotational entropy of single water molecules in protein cavities and interfaces using molecular dynamic simulations. In ***chapter 5,*** we used a new approach to calculate translational and rotational entropy of the individual water molecules around different hydrophobic and charged solutes. We showed that for the small hydrophobic solute, the translational and rotational entropies of each water molecule increase as a function of its distance from the solute reaching finally to a constant bulk value. As the size increases (0.746 nm), behavior for the translational entropy is opposite; water molecules closest to the solute have higher entropy that reduces with distance from the solute. Rotational entropy for all sizes has similar increasing trend, indicating an absence of crossover in rotational entropy. This makes the crossover in total entropy (translation + rotation) happen at the much larger size (>1.5 nm). Results obtained are helpful to understand water entropy behavior around various hydrophobic and charged environments within biomolecules. Finally, we show the application of the method to calculate the entropy of individual water molecules in a biological cavity (of the streptavidin) that are replaced by ligand (biotin) binding.

Both the intercalation and water entropy has the common theme in drug design. Therefore, we conclude the thesis with some proposals towards the future work in these directions with the goal of better drug designing.

Table of Contents

Certifications **I**

Acknowledgements **III**

Synopsis **VI**

Chapter 1. Introduction **1-32**

1.1 Introduction to Molecular Recognition 2

1.2 DNA-Ligand Interaction: Intercalation 3

1.2.1 Biological Significance of Intercalation Process 4

1.2.2 Classification of DNA Intercalating Agent 5

1.2.3 Structural Characterisation 6

1.2.4 Thermodynamic Characterisation 8

1.2.5 Kinetic Characterisation 10

1.2.6 Theoretical and Computational Evidences 14

1.3 Molecular Mechanism of Intercalation 21

1.3.1 Methods Used 22

1.4 Outline 25

1.5 References 26

Chapter 2. Molecular Mechanism of Direct Proflavine-DNA Intercalation: Evidence for Drug-Induced Minimum Base-Stacking Penalty Pathway **34-54**

2.1 Introduction 36

2.2 Experimental Section 39

2.2.1 Preparation of Minor groove Bound State 39

2.2.2 Molecular Dynamic Simulation Protocol 40

2.2.3 Description of Collective Variables 41

2.2.4 Configurational Restraint 42

2.2.5. MFEP Calculation and Cluster Analysis 44
2.3 Result and Discussion 42
2.3.1 Design and Method 42
2.3.2 Time Dependence of Collective Variables 43
2.3.3 Free Energy Surfaces 44
2.3.4 Error Analysis 45
2.3.5 Minimum Free Energy Path and Changes in DNA and Energy Parameters 46
2.4 Conclusion 50
2.5 References 51

Chapter 3. Intercalation and Deintercalation Pathway of Proflavine through the Minor and Major Grooves of DNA: Roles of Water and Entropy 55-81

3.1 Introduction 57
3.2 Experimental Section 59
3.2.1 Preparation of Intercalated State 59
3.2.2 Preparation of the Minor and Major Groove Bound States 60
3.2.3 Simulation Details 60
3.3 Results and Discussions 61
3.3.1 Design of Reaction Co-ordinates 61
3.3.2 Deintercalation towards Major and Minor Grooves 62
3.3.3 Dissociation from Minor and Major Groove-bound States 66
3.3.4 Intercalation from Minor and Major Grooves 68
3.3.5 Free Energy Profile along MFEP 73
3.3.6 Comparison of Thermodynamic Profiles 74
3.3.7 Kinetics from Thermodynamics 76
3.4 Conclusion 79
3.5 References 79

Chapter 4. How Fast is the First Step of Intercalation? Structure and Dynamics of the pre-intercalative Bound State of an Anti-Cancer Agent Proflavine around DNA 82-104

4.1 Introduction 84

4.2 Experimental Section 87

4.2.1 Proflavine-DNA System and Equilibration 87

4.2.2 Proflavine in Water 87

4.2.3 Dimer Free Energy Calculation 88

4.2.4 Stacking Criteria 88

4.3 Results and Discussions 89

4.3.1 Proflavine Aggregation in the Absence of DNA 89

4.3.2 Free Energy of Proflavine Dimer 91

4.3.3 Proflavine Aggregation in Presence of DNA 92

4.3.4 Structure of the Outside-Bound State 97

4.4 Conclusion 101

4.5 References 102

Chapter 5. Single Water Entropy: Hydrophobic Crossover and Application to Drug Binding 105-146

5.1 Introduction 107

5.1.1 Role of Water in Molecular Recognition Process 107

5.1.2 Hydrophobic Effect and Water Entropy 109

5.1.3 Calculation of Entropy 113

5.1.4 Outline of Work 114

5.2 Experimental Section 114

5.2.1 Simulation Details 114

5.2.2 Methods 116

5.3 Results and Discussions 120

5.3.1 Water Structure and Inhomogeneous Solvation Theory 121

5.3.2 Single Water Translational Entropy 123

5.3.3 Single Water Rotational Entropy 127

5.3.4 Hydrogen Bond Variation 132

5.3.5 Recovery of the Entropy Loss 135

5.3.6 Application to Drug Design 135

5.4 Conclusion 140

5.5 References 141

Chapter 6. Summary and Future Directions 147-153

List of Publications 154-156

Annexure 157-187

Chapter 1

Introduction

Chapter 1: Introduction

1.1. Molecular Recognition

Molecular recognition is an elementary step in any biological process, where, specific interaction between two or more molecules happens through non-covalent bonding such as hydrogen bonding, van der Waals (vdw) interactions, π-π interactions, hydrophobic interactions, electrostatic interactions etc[1-2]. When these type of interactions are found within the same molecule, they are referred to as intra-molecular recognition exemplified by protein folding[3]. However, aforementioned interactions are most profuse between two or more molecules and known as inter-molecular recognition. Examples corresponding to such recognition process are protein-protein, protein-ligand, DNA-protein, DNA-ligand, RNA-ligand interactions[1-2, 4-5] etc. Figure 1.1 shows different types of molecular recognition processes.

Generally, a molecular recognition process involves three key components: guest, host, and solvent (water in biological processes)[6]. The driving force for the interaction between the guest and host molecules in presence of water lies in the complementarities between the molecules, first proposed by Emil Fisher in 1894 through the 'lock and key' model[7]. The factors which contribute to this complementarity are (i) shape of the molecules, (ii) various non-covalent interactions such as hydrogen bonding between the molecules, ion-ion and ion-dipole interactions, vdw interactions and (iii) hydrophobic effect imparted by water molecules due to the entropically unfavorable solvation of smaller non-polar groups.[1-2] Above factors mold the molecular recognition process complicate enough to be probed by different experimental and theoretical investigations. So, plenty of efforts are devoted to study such processes as its fidelity helps in the control of all cell exercises and chemical activity. Among the various kinds of molecular recognition processes, the association of small molecules to bio-macromolecules is of fundamental importance to drug design process to combat various ailments[2]. In this thesis, we attempt to give answers to the questions arose in one such molecular recognition process called 'drug-DNA intercalation' and understand the role of water in modulating molecular recognition processes.

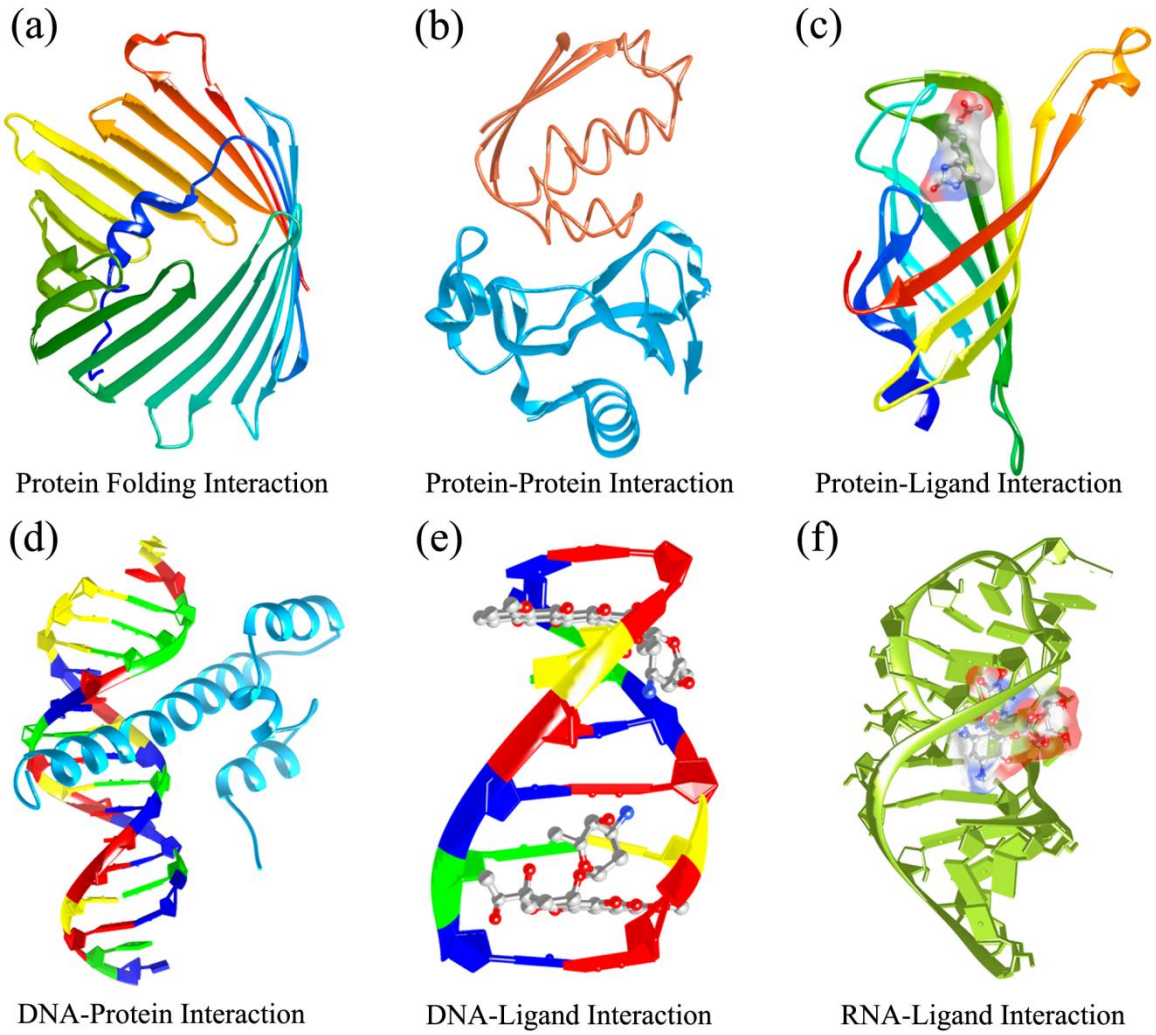


Figure 1.1. *Examples for different molecular recognition processes. Figures are made using UCSF Chimera.*[8] *PDB ID and name of the complex: (a) 4C69, Voltage dependent anion selective channel protein 1; (b)1BRS.pdb, Barnase-Barstar complex; (c)3RDO, Streptavidin-Biotin complex; (d) 1SKN, skn-1 binding domain-DNA complex; (e)1VTH, d(TGTACA)-Daunomycin complex; (f)2KXM, RNA-Ribostamycin complex.*

1.2. DNA-Ligand Interaction: Intercalation

Various small molecules have the potential to interact with nucleic acids to modulate their natural biological functions by various covalent and non-covalent interactions[9]. Covalent interaction includes the alkylation of the bases of nucleic acids by alkylating agents such as cisplatin, carboplatin etc. Different modes of non-covalent interactions are possible between nucleic acids and ligands. The different binding modes are minor groove binding, intercalation, and major groove binding as shown in Fig. 1.2. Among the different binding modes, *intercalation*, first termed by Lerman in 1961[10], is a unique

mode of interaction characterized by the insertion of small planar aromatic molecule between the adjacent base pairs of DNA[10]. It generally causes stabilization and structural changes to the DNA such as local unwinding, lengthening etc. leading to the alteration of various biological functions of DNA[10-12]. As a result, intercalation has got tremendous therapeutic implication. Intercalating agents are often used to treat various kinds of cancers, several microbial, and parasitic infections[13].

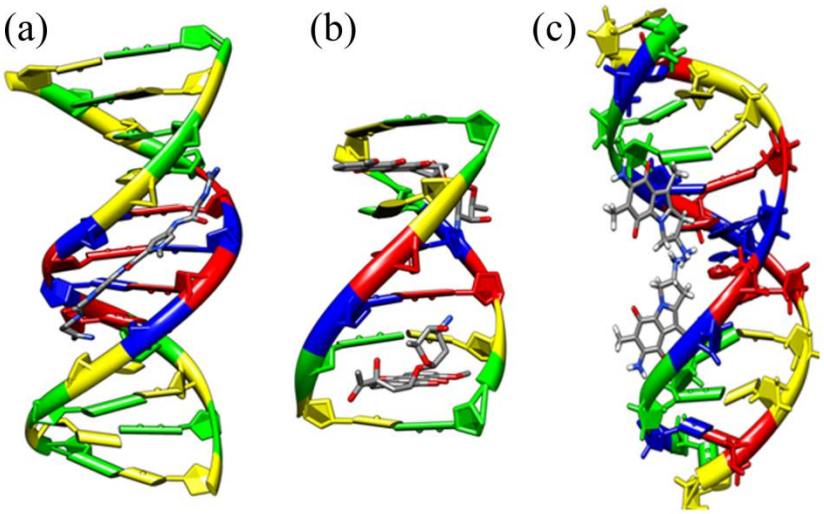


Figure 1.2.*Different modes of non-covalent interactions of ligands with DNA. (a) Minor groove binding, (b) Intercalation and (c) Major groove binding. Figures are made using UCSF Chimera.[8] PDB ID and Name of Complex: (a) 4U8A, DNA-N,N'-bis[3-(1,4,5,6-tetrahydropyrimidin-2- yl) phenyl]biphenyl-4,4'-dicarboxamide complex; (b) 1VTH, d(TGTACA)-Daunomycin complex; (c)1JO1, DNA-Decarbamoyl-2,7-diaminomitosene complex.*

Intercalation is also prevalent in protein-DNA interactions that lead to specific recognition of protein to DNA. Some proteins bend DNA towards the protein characterised by the presence of amino acids intercalated in between the base pairs.[14] A recent study has shown that the formation of specific kink in the DNA is caused by partial, not complete, intercalation of amino acid and the kinked state is thermodynamically stabilised by the presence of whole protein.[15]

1.2.1. Biological Significance of Intercalation process

Intercalation of small molecules to DNA causes the inhibition of DNA replication and transcription process[9, 16-17], thereby preventing the cell division and growth[18]. They also act by inhibiting various DNA dependent enzymes such as topoisomerase-II, DNA polymerase and RNA methylase[19-22]. Some of them are also known to cause frame shift

mutations in the DNA[23-24]. Even the DNA intercalating agents such as doxorubicin[25], actinomycin D, and ellipticine[26] can cause single strand breakage of the DNA. Various DNA intercalating agents used in the treatment of cancer are actinomycin D[27-30], daunorubicin[31-33], doxorubicin[34], idarubicin[35], mitoxantrone[36], ellipticine[37], amsacrine[38] etc. A well-known intercalating agent proflavine also has anti-cancer property but could not be used clinically because of its non-selective nature and toxicity. Efforts, however, were made to modify proflavine to make it more selective, less toxic, and potent as anti-cancer drugs[39-40]. Intercalating agents such as mepacrine and quinacrine have anti-parasitic property and are used in the treatment of malaria[41-42] and trypanosomiasis[43]. Trapaflavine, ethacridine, acriflavine, and proflavine were used as antibacterial agents[44-45]. Berberine, a DNA intercalator, has anti-fungal[46], anti-bacterial,[47] and anti-viral properties.

1.2.2. Classification of DNA Intercalating Agents

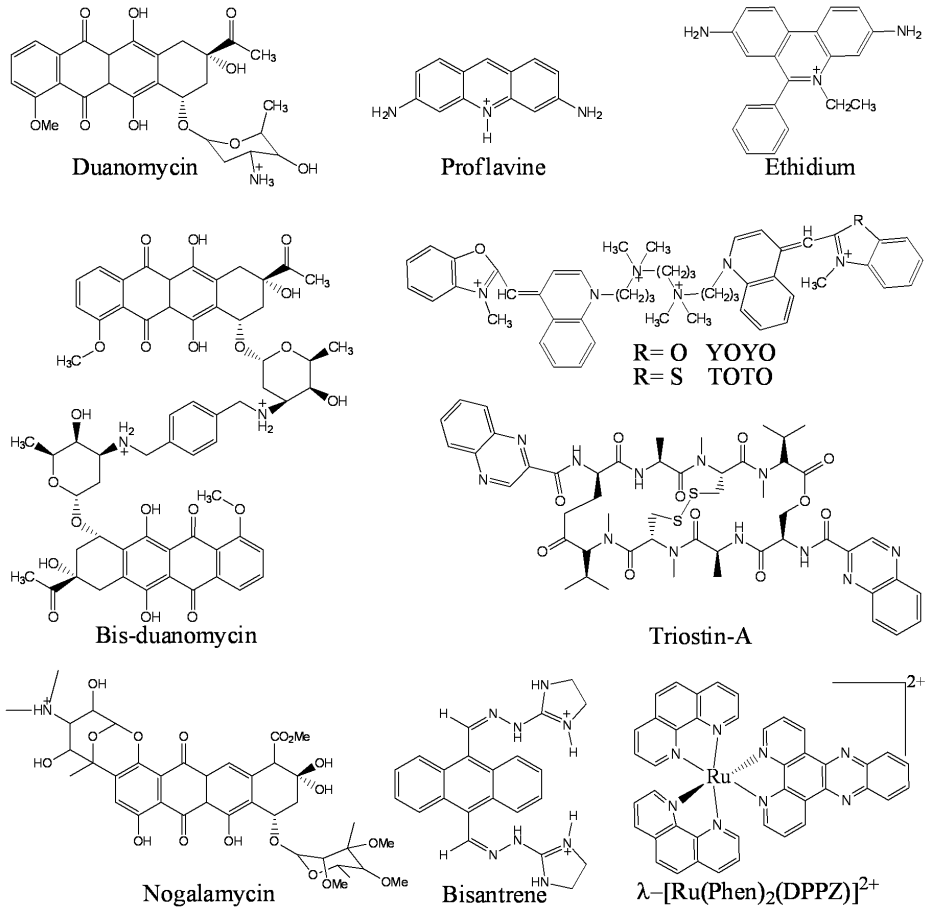


Figure 1.3. *Chemical structures of different mono-, bis- and threading intercalators.*

Based on the nature of intercalation, the intercalators are classified into mono-intercalators, bis-intercalators, and threading intercalators. In mono-intercalators (Daunomycin[48], proflavine[49]), one planar moiety inserts between the base pairs of DNA either from the major or minor groove side. But bis-intercalators consist of two mono-intercalators that insert between base pairs located at different regions of DNA. Examples are bis-daunomycin[50], echinomycin[51], triostin[52], ditercalinium[53], TOTO[54], YOYO[55] etc. In case of threading intercalators, the side chains of the planar intercalating ring pass though both minor and major groove of DNA. Nogalamycin, napthalene diimide[56], bisantrene[57] fall in this category. Chemical structures of some DNA intercalators from different category are shown in Fig. 1.3.

Based on the arrangement of long axis of planar intercalating moiety with respect to the intercalating base pairs, intercalators are again classified into parallel and perpendicular intercalators as shown in Fig.1.4. In case of parallel intercalators such as acridine orange, proflavine etc., the long axis of planar ring is parallel to the long axis of intercalating base pairs (see Fig.1.4.a)[58-60]. Perpendicular intercalators (daunomycin, nogalamycine etc.) consist of bulky substituent on one or both sides of the long axis of the planar moiety and orient themselves perpendicular to the flanking base pairs (Fig.1.4.b).[58, 60]

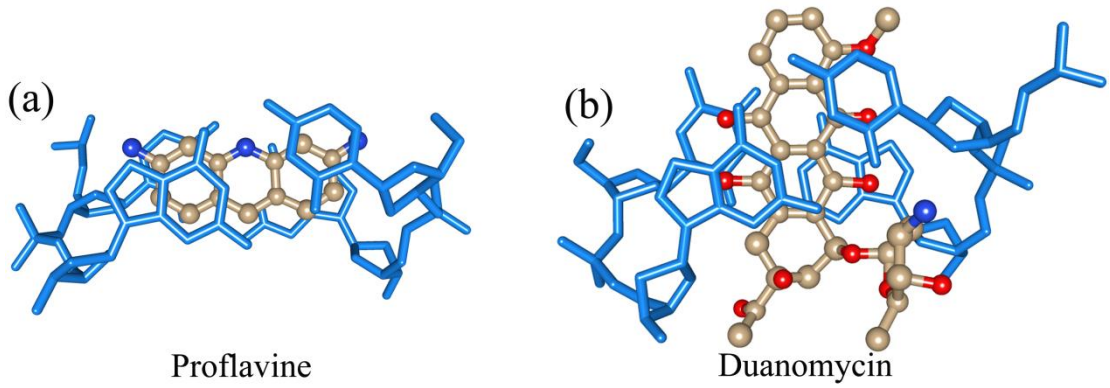


Figure 1.4. *(a) Parallel and (b) perpendicular arrangement of the ring system of intercalators with respect to the long axis of the base pairs (blue). The view is along the DNA axis.*

1.2.3. Structural Characterization

The structural characterization of intercalation process had been performed by various biophysical methods. Increase in the length of the DNA during intercalation

process (base pair separation from 3.4 Å to ~6.8 Å)[49, 61] enabled viscosity enhancement method to characterize the process[10]. Examples are acridine, proflavine and ethidium bromide[10, 62],acridine orange, acriflavine and anthracycline antibiotics[63] etc. Radioactive labeling (^{3}H (tritium) or ^{125}I) of the drug was also used to characterize intercalation as it can detect the increase in the length of DNA.[64] Intercalation also increases rigidity of the DNA, which is reflected in the increase of the melting temperature (T_m) of DNA. Increase of T_m was found for DNAs intercalated with actinomycine[65], daunomycin[66], ethidium bromide[67] etc. Similarly, the local unwinding of DNA led to the identification of the

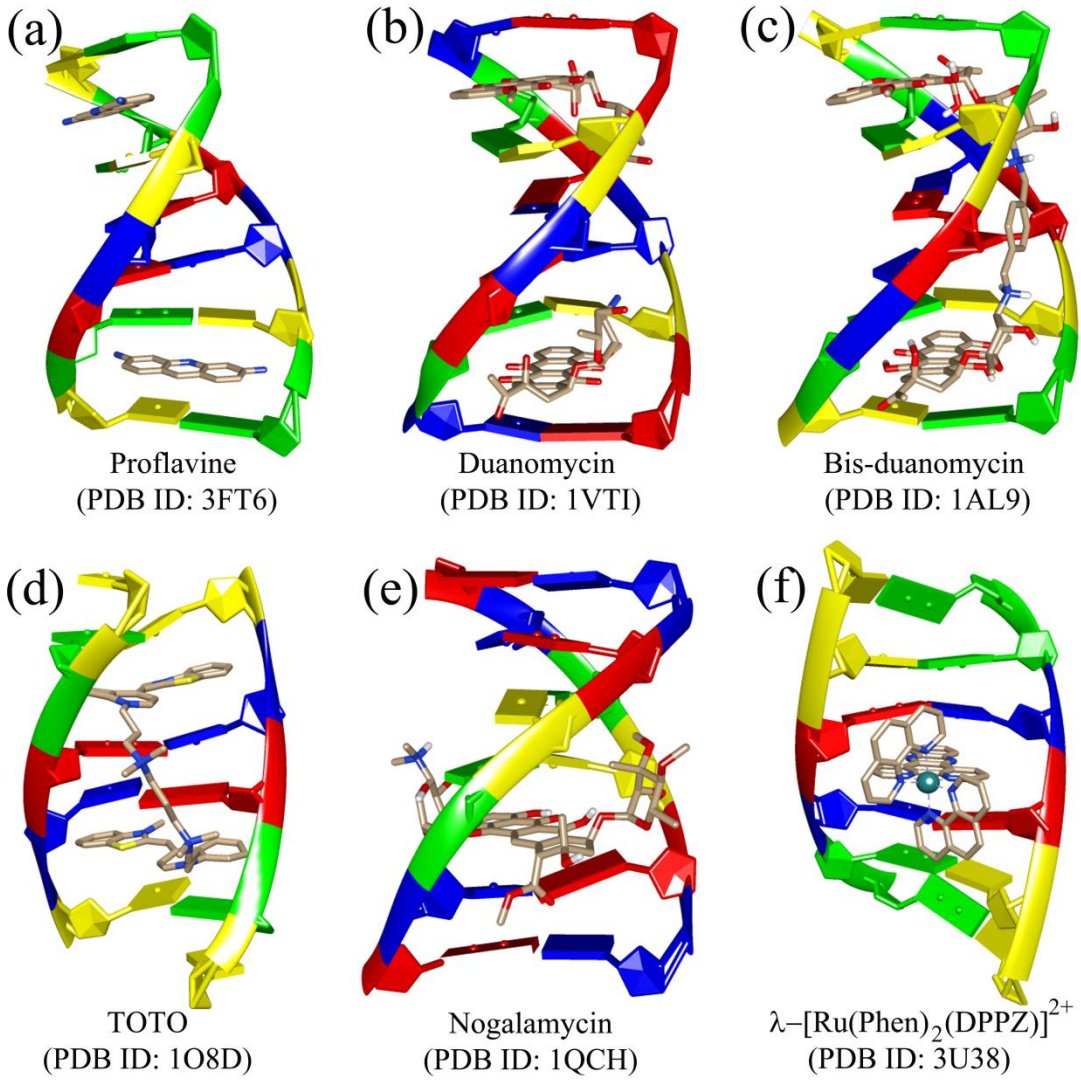


Figure 1.5. 3-*D representations of various intercalator-DNA complexes. PDB IDs and name of the intercalators are mentioned. (a) and (b) show mono-intercalators, (c) and (d) are bis-intercalators, (e) and (f) denote threading and metallo-intercalator, respectively.*

intercalated state by sedimentation method. The release of super coils in circular DNA due to unwinding decreases the sedimentation coefficient of DNA[12, 68]. The change in the size distribution due to intercalation was detected by dynamic light scattering method[69-70,71]. Various X-ray diffraction methods were able to probe into the molecular details of the intercalated state structure. Fiber diffraction study of DNA-proflavine complex[72] gave a qualitative proof of the intercalation hypothesis proposed by Lerman in 1961. It also suggested that all molecules are not intercalated, but some of them are bound outside also[64]. Single crystal X-ray techniques were focused to understand the helix unwinding, sequence specificity, and neighbor exclusion of intercalation process.[16] Crystal structures of complexes of DNA with intercalators such as actinomycin-D[73], ethidium[74], proflavine[75], daunomycin[48, 66, 75-76] and several other molecules were solved. The 3D-representations of some of the complexes are shown in Fig.1.5.

1.2.4. Thermodynamic Characterization

Thermodynamic parameters of the intercalation process were studied extensively using various methods such as equilibrium dialysis, spectroscopic (absorbance and fluorescence) titration analysis, surface plasmon resonance, isothermal titration calorimetry to understand the binding mode, affinity, specificity, contribution of enthalpy and entropy of the binding process, etc.

Various thermodynamic parameters such as $\Delta G^0, \Delta H^0$, ΔS^0 and ΔC_p^0 were calculated using different experimental techniques such as equilibrium dialysis, spectroscopic (absorbance and fluorescence) titration analysis, surface plasmon resonance, isothermal titration calorimetry, etc. for various DNA-drug complexes as listed in Table 1.1. Table 1.1 shows the thermodynamic parameters of intercalation of different intercalators with different DNA sequences. The free energy stabilities of different mono intercalators are within ~10 kcal/mol.

Different thermodynamic studies performed on various DNA-intercalator complexes had shown that the molecules have selective affinity towards specific base pairs of DNA. One of such example is DNA-acriflavine complex, where the binding affinity is more towards alternating polymer (RYR; R=Purine, Y=Pyrimidine)[77]. This binding heterogeneity was also seen in DNA-quinacrine complex[78]. However in case of

acridine orange, the intercalation happened between G-C base pairs as well as A-T base pairs.[79] Competition dialysis experiments demonstrated that daunomycin also shows sequence specificity towards G-C base pairs attaining comparatively larger value for drug bound per base pair with an increase in the G-C content of DNA. This was further confirmed by CsCl density gradients.[66] Later, a more detailed study on the specificity of daunomycin binding towards different polynucleotides illustrated that the binding constants decrease in the order poly[d(A-T)].poly[d(A-T)] > poly[d(G-C)]. poly[d(G-C)] >poly(dG).poly(dC) > poly(dA).poly(dT).[80]

Table 1.1. *The thermodynamic parameters of intercalation process of different sequences of DNA by daunomycin, ethidium bromide and acridine analogs.*

Molecule	ΔG(kcal/mol)			ΔH(kcal/mol)			ΔS(cal/Kmol)		
	a	b	c	a	b	c	a	b	c
Ethidium bromide	-9.1	-7.2	-6.9	-10.0	-1.2	-8.7	-3.0	18.0	6.0
Daunomycin	-8.0	-7.9	-9.4	-12.8	-9.1	-11.0	-16.2	-4.1	-5.4
9-amino acridine	-6.2	-6.3	-6.1	-5.1	-6.6	-4.4	3.7	-1.2	5.8
Proflavine	-6.1	-6.8	-	-5.4	-8.9	-	2.3	-7.3	-
Quinacrine	-9.1	-8.2	-8.7	-5.6	-6.0	-6.0	11.9	7.3	9.9
Methylene blue	-8.2	-8.0	-7.8	-2.4	-2.3	-2.1	19.7	19.5	19.5

Ethidium bromide: a=Poly[d(AT)].Poly[d(AT)][81], b=Poly(dA).Poly(dT)[81], c=Calf Thymus DNA[82]; Daunomycin: a=Calf Thymus DNA[66], b=$[Poly(dA\text{-}dT)]_2$[83], c= Poly(dAdC)-Poly(dGdT)[84]; 9-amino acridine: a=Calf Thymus DNA[85], b=$Poly[d(GC)_2]$[85], c= $Poly[d(AT)_2]$[85]; Proflavine: a=Calf Thymus DNA[85], b= $Poly[d(GC)_2]$[85]; Quinacrine and Methylene blue: a=Poly(dG-dC).Poly(dG-dC)[86], b=Poly(dG).Poly(dC)[86], c=Poly(dA-dT).Poly(dA-dT)[86].

Thermodynamic studies were also used to understand various factors affecting intercalation process. Thermodynamic characterization of different analogs of the intercalating agent ethidium was performed to understand the influence of amine substituents on the intercalation process. The removal of the amine group decreased the binding affinity as well as it shifted the intercalation process from enthalpy to a entropy driven one.[82] The hydration changes during the intercalation of different intercalators

ethidium, propidium, proflavine, daunomycin and 7-amino actinomycin-D were characterized using osmotic stress method. Apart from ethidium-DNA complex, all other drug-DNA complexes showed uptake of water during the complex formation.

1.2.5. Kinetic Characterization

Various experiments such as temperature jump technique, pressure jump technique, stopped flow technique and dissociation using Sodium Dodecyl Sulphate (SDS) micelle were used to understand the kinetics of intercalation process. The most important aspect of the kinetic studies is the ability of the intercalators to fluoresce and change their fluorescence intensity when intercalated. This allows experiments to monitor the overall fluorescence intensity as time progresses giving rise to time dependence of intensity, which is then fitted to different kinetic models. The methods rely on the fact that the non-equilibrium state is created much faster than the timescale of intercalation. These studies pointed towards the average mechanism of the intercalation process and suggested the complex nature of it. Fig. 1.6 shows possible pathways surmised by McGregor *et al.*[87] Some pathways would lead to the opening of base pairs followed by intercalation (steps: **c** and **a**). Other pathways are either diffusion-limited association to an "outside" bound state (step: **d**) followed by synchronized (step: **x**) or drug-assisted (step: **c'**) opening of the base pair and subsequent intercalation. However, there is a high probability that the drug binds in the wrong place of the DNA. In that case, it has to diffuse along the DNA to search for a proper intercalation site (steps: **d** followed by **s**).

Even though it is possible to devise chemical kinetic equation for each of the individual pathway, it will be really difficult to decouple contributions of each elementary step from kinetic experiments. Therefore, there were efforts to combine several different methods and different competing kinetic schemes to explain and understand the possible mechanistic pathways of the intercalation process, which also depends on the nature of the intercalators. Starting from a simple scheme of one step reaction proposed by Macgregor[88] in case of ethidium bromide, a complex multistep scheme was proposed for daunomycin by Rizzo *et al.*[89] For metal intercalators, the kinetics is even more complex.[90]

Detailed kinetic studies on the intercalation of DNA by actinomycine[91] and proflavine[11] were performed by Crothers for the first time using a variety of association and dissociation methods. While the kinetics for actinomycine was too complex to understand, he proposed a simple two step series reaction kinetics for proflavine. First step was argued to be the formation of an outside-bound state falling within the timescale of a diffusion limited process. Second step was ascribed to the formation of the intercalated state within the range of a slow activated process. However, an alternate parallel mechanism was also suggested for proflavine intercalation to DNA. Crothers argued for the series mechanism over parallel one because intercalation, which requires structural changes, could not happen directly in parallel mechanism. However, as we show in the present thesis that there are more complexities involved in the process of intercalation even for a simple drug like proflavine.

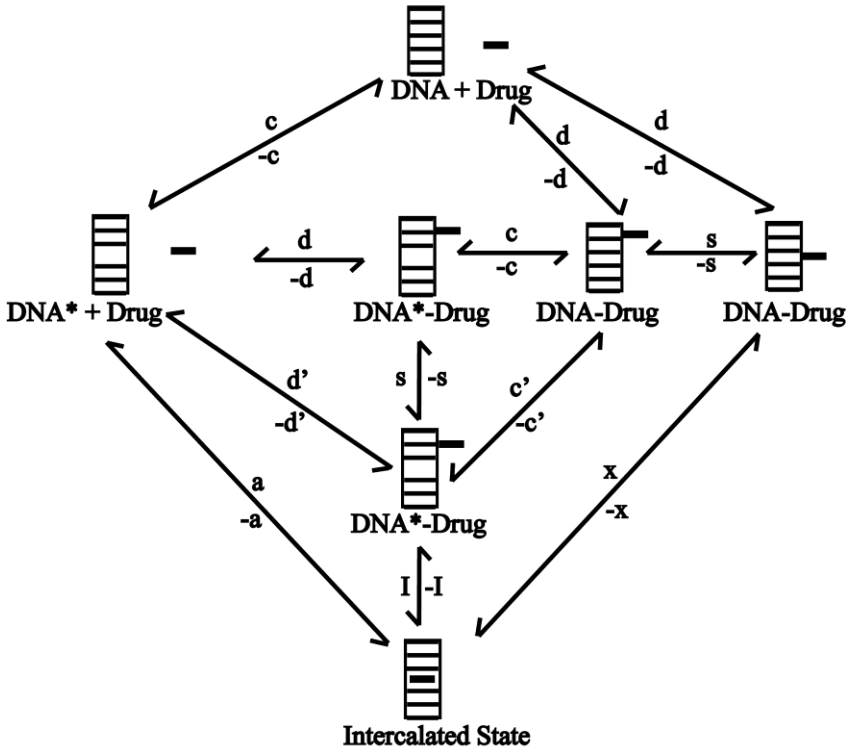


Figure.1.6. *Schematic diagram of the different possible kinetic pathways of intercalation surmised by McGregor et al.[87] Adapted with permission from Fig. 1 of Ref. 85). Copyright (1987) American Chemical Society.*

For the proflavine intercalation, Crothers *et al.*[11] proposed two limiting mechanisms for the intercalation process to happen. (a) The drug waits outside the DNA, structure of which continuously fluctuates due to the thermal environment. At a certain

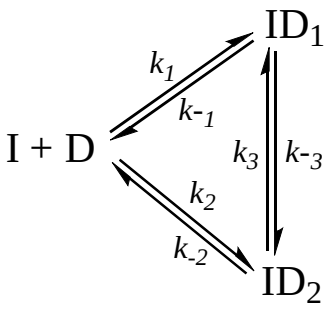
moment, the cavity size is large enough for the drug to slip in. (b) Secondly, the drug creates a cavity in the DNA and intercalates through an induced-fit mechanism. From the indirect evidence from the experiments, he supported the second mechanism. However, at that point, it was not possible to address this question in detail in the molecular level.

For certain drugs, however, even a simpler kinetic scheme (Scheme 1) explains the observation. McGregor proposed this simple model for kinetics of ethidium bromide intercalation[88],

$$I + D \underset{k_d}{\overset{k_f}{\rightleftharpoons}} ID \qquad \text{Scheme 1}$$

(Note: I represents Intercalator and D represents DNA)

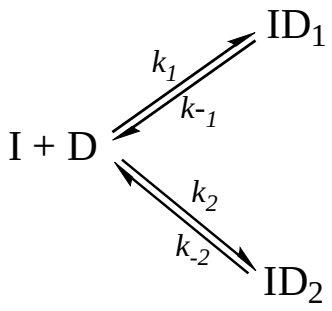
Bresloff and coworkers[92] proposed the cyclic mechanism (Scheme 2) for ethidium bromide intercalation. Two bound states form directly but they interchange amongst themselves, described as the direct transfer mechanism.



Scheme 2

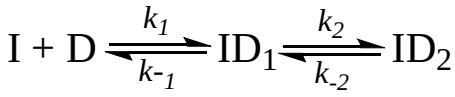
ID_1 and ID_2 are the fast bound step and slow bound step. However, the nature of states is in debate. However, recent studies have shown that direct transfer is present for ethidium intercalation in RNA and stabilization of double strand in the triple helix poly(A).2poly(U), where major groove is occupied by the third strand.[67] Therefore, it is

not clear whether the direct transfer is between different grooves of DNA or between different DNA sites.



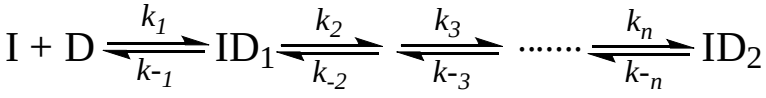
Scheme 3

The parallel mechanism shown for ethidium bromide in Scheme 3 is kinetically indistinguishable from the sequential mechanism. Using the dependence of the rate of intercalation on the ionic concentration, Wilson[93] also proposed the following (Scheme 4) two-step model for ethidium and propidium binding.



Scheme 4

Metallo-intercalators follow a rather more complicated kinetics possibly owing to their complex structure as shown below (Scheme 5). This is for the binuclear Ru-based drugs.[94]

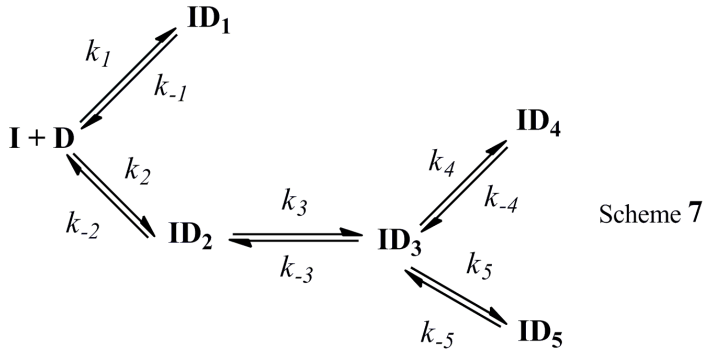


Scheme 5

Apparently, kinetics of intercalation process depends on the nature of the drug as well. Chaires [95] performed a detailed kinetic study involving intercalation of daunomycin to DNA using stopped-flow relaxation technique and proposed a simple series mechanism(Scheme 6). First step is argued to be outside-bound state, followed by intercalation in the second step and final step is the relaxation of the drug and DNA.

$$I + D \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} ID_1 \underset{k_{32}}{\overset{k_{23}}{\rightleftharpoons}} ID_2 \underset{k_{43}}{\overset{k_{34}}{\rightleftharpoons}} ID_3 \qquad \text{Scheme 6}$$

Rizzo et al[89] also found three time constants for the fluorescence intensity and fitted it to a five-step process (Scheme 7). The main difference is that the bimolecular association has two steps, one of which does not intercalate. The crystal structure of the intercalated daunomycin clearly indicates that the drug must have intercalated from the minor groove side. However, possibility of binding to the major groove also non-negligible. Therefore, ID_1 shown below could very well be a major groove-bound state, depends on the drug concentration but they could not characterize ID_1. Another important point is that they argued the first step to be diffusion limited based on the observation that the removal of methoxyl group leads to an enhancement of the rate, however, not out-side bound state as mentioned by Chaires[95].



Scheme 7

Forster *et al.*[96] used temperature jump technique and proposed only a single-step binding process. However, their main focus was on the cooperativity of the intercalation process. They proposed 3 different situations: (i) free binding site (ii) single consecutive binding site (intercalation of a drug when another one is intercalated) and (iii) doubly consecutive binding site (intercalation of a drug between two already intercalated drugs). Forster *et al.*[96] showed that adriamycin has higher cooperativity. In this thesis, we have used only one drug to intercalate into DNA. Therefore, it will apply to the first case of free binding to the DNA.

1.2.6. Theoretical and Computational Evidences

Theoretical and computational efforts to understand intercalation process are quite old. However, with the refinement of force-field, computational methods, and increase in computer efficiency, molecular level of detail in the mechanism of the process has emerged recently. As we explore the computational and theoretical aspects of intercalation in this thesis, we describe the progress in the field for different aspects of the intercalation process.

1.2.6.1. Molecular mechanical studies

The first effort to produce a molecular model of intercalated state was by Fuller and Waring[97] in 1964 who produced a hand-built model of DNA for ethidium bromide intercalation at each base pair level. Sobell used CPK (space filling model of atoms) model of actinomycin-deoxyguanosine complex to illustrate the intercalation geometry.[73] As an effort to model the backbone dihedral angles to avoid steric repulsion, computerized linked-atom modeling system was developed by Alden and Arnott.[59] The model was based on optimizing the stacking interactions, steric strain, and non-bonding interactions to accommodate the insertion of a molecule between the base pairs. They found that transition of two torsion angles to trans and sugar ring puckering to C3'-endo relieves the strain in the adjacent base pair during intercalation. It was assumed that intercalation of actinomycin and ethidium bromide happens from minor groove side, proflavine from major groove side and 9-aminoacridine from either side.[73, 98-100]. An empirical potential function calculation was done on dinucleotide-drug complex level by Nuss *et al.*[101] in 1979 to predict the structure and energetics of DNA-intercalator complex. They used the potential function of the following form,

$$V = \sum_{i \neq j}^{n} \left(\frac{q_i q_j}{\epsilon R_{ij}} + B_{ij} e^{-c_{ij} R_{ij}} - \frac{A_{ij}}{R_{ij}^6} + \sum_{k=1}^{ntor} \sum_{l=1}^{3} V_l^k cos l\varphi \right). \qquad Eqn. 1$$

They suggested that proflavine has to intercalate from major groove because of the steric reasons. Energy calculations proposed that 9-aminoacridines can intercalate from the minor groove or major groove as there is no side chain amine groups to provide steric hindrance[101]. The calculations on proflavine have shown that the intercalated structure predicted was closer to the X-ray crystallographic structure.[75] The calculated

energy (15-30 kcal/mol) for unstacking of the base pair is qualitatively reasonable to explain the intercalation barrier and the effect of solvation-desolvation energies on the intercalation process.[101] This insightful study, however, did not consider the compensation of the base unstacking penalty by the drug and also the effect of water in the intercalation process. Dearing and co-workers[102] used molecular mechanics to energy minimize different complexes of acridine orange and proflavine with nucleotide base pairs where they used the force-field (functional form of which is used current force-field) of the following form,

$$E = \sum_{bonds} k_b(r - r_b)^2 + \sum_{angles} k_a(\theta - \theta_a)^2 + \sum_{dihedrals} \frac{k_d}{2}[1 + cos(n\varphi - \gamma)] + \sum_{non-bonded} \left[B_{ij} r_{ij}^{-12} - A_{ij} r_{ij}^{-6} + \frac{q_i q_j}{\epsilon_{ij} r_{ij}}\right] \qquad Eqn.\ 2$$

Partial atomic charges for acridine orange and proflavine were calculated by CNDO/2 approximation. In order to incorporate the solvent effects, they used variable dielectric constants. The analysis predicted that acridine orange intercalation prefers a mixed C3'-endo (3'-5') C2'-endo sugar puckering with better stacking interactions with the base pairs, larger intercalation site and minimum steric strain as compared to uniform C3'-endo sugar puckering. Compared to acridine orange, proflavine showed uniform C3'-endo sugar puckering with less stacking interaction with the bases, but compensated by the hydrogen bonding interaction among it's amine groups and bases. First molecular dynamics simulation of intercalation was performed by Singh[103] *et al.* using all-atom force-field for a bis-intercalator-DNA complex. He studied a bis-intercalator triostin-A between the C-G base pairs of DNA sequence $d(CGTACG)_2$ and showed that bis-intercalator stabilizes the neighboring A-T base pairs into a Hoogsteen (HG) form with van der Waals interactions between the side chain of the drug and backbone of adenine because of the reduced minor groove width in HG base pairing. Results showed that triostin-A prefers G-C-rich DNA to A-T-rich DNA. Rao and Kollman[104] performed QM/MM studies on double intercalation of 9-amino acridines to DNA to understand the physical basis of neighbor exclusion principle proposed by Li and Crothers [11]. The double intercalated structures of 9-amino acridine into the heptamer $d(CGCGCGC)_2$ were modeled with obeying (as shown in Fig 1.7) and violating (as shown in Fig 1.8) neighbor

exclusion principle using the computer graphics program CHEM[105]. Molecular dynamics simulations have shown that the violated structures are favored over the obeyed structures because of favorable electrostatic interactions due to the reduction in the repulsion of phosphate groups which are stretched by intercalation. But the drug-helix entropy is more favorable for the obeyed structure by 2.4 kcal/mol compared to the violated structure due to more flexibility of backbone. Therefore, violation of neighbor exclusion principle is energetically feasible but entropically unfavorable due to the stiffening of the helical structure from double intercalation in the adjacent base pairs.

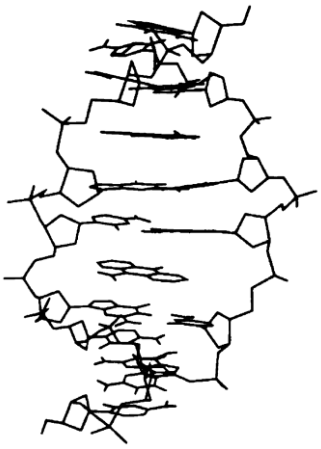


Figure 1.7. *Representation of the double intercalation of 9-amino acridine to DNA obeying the neighbor-exclusion principle. (Fig. adapted from Fig.2 of Ref.102)*

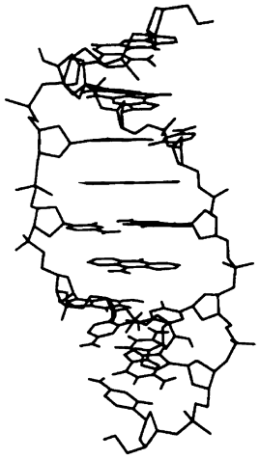


Figure 1.8. *Representation of the double intercalation of 9-amino acridine to DNA violating the neighbor-exclusion principle. (Fig. adapted from Fig.3 of Ref.102)*

Williams *et al*.[58] used two kinds of model interactions: electrostatic repulsion that winds the helix and stacking interaction that unwind it to study the exclusion principle.

During intercalation, DNA length increases. This reduces the electrostatic interaction between the adjacent base pairs. However, this is compensated by the increased stacking interaction due to unwinding of helix. So for the second intercalator to intercalate between the neighboring base pair, it has to disrupt the stable stacking interaction between the already unwound bases caused during the first intercalation. This decreases the relative stability of the second intercalator in the neighborhood of a first one. For intercalators with side chains such as daunomycin, the excluded volume of the side chain clearly determines the neighbor exclusion for the second intercalator[48]. Neidle *et al.*[106] performed modeling study for DNA-proflavine interaction using the geometry of the intercalated state from the available crystal structure and modifying it to have a decamer with one proflavine intercalated in the center. Energetic calculations thereafter showed that the conformational changes in the intercalation cause over winding of the bases which are far away from the intercalation site.

The molecular dynamics simulation studies on the 2:2 proflavine complex of dCpG nucleotides with the hydration layer[107] using Gromos force field[108] has shown the organization of water in the form of pentagonal network on the major groove and polygonal network in the minor groove as seen in the crystal structure[76]. The modeling of intercalated state of doxorubicin by molecular mechanics has shown that in the intercalated state the sugar group remains in the minor groove and the plane of the anthraquinone ring of doxorubicin is perpendicular to the base pairs CpG.[109] Later, crystal structure of doxorubicin in DNA has supported this geometry[110]. Similar study was performed by Elcock[111] on 9-hydroxy ellipticine where he showed that both parallel and perpendicular orientations are favorable in terms of stacking interactions and hydrogen bonding interactions. However, the electrostatic solvation free energy was found to be more favorable for perpendicular orientation of 9-ellipticine.

Molecular dynamic simulations on the side chain derivatives of pyrimido acridine triones (PATs) in the intercalated geometry with the duplex $d(GCGCGCGCGCGC)_2$ has shown that the position and the nature of a side chain determines the orientation of the intercalator in the intercalation cavity.[112] To understand the energy of formation of intercalation cavity by helical breathing (conformational changes in the base pair) in case of both mono- and bis-intercalators, difference in energy between uncomplexed DNA

sequence d(CGCGCGATCGCGCG)$_2$ and DNA-intercalator complex was calculated by MM-GBSA method using AMBER molecular dynamics simulation software[113]. The MM/GBSA simulations of DNA, DNA with mono-intercalated daunomycin, bis-intercalated daunomycin, intercalated bis-daunomycin, and intercalated 9-dehydroxy daunomycin performed by Trieb *et al.*[114] estimated the free energy of creating one intercalation cavity to be 32.3 kcal/mol. The subsequent cavity formation was found to be less costly. Therefore, it was argued to be originated from a cooperative cavity formation, corroborated with the previous work done by Williams et al[58]. Note that, while the cost for cavity formation is estimated[114] to be so high (32.3 kcal/mol), experimental barrier of intercalation is rather low (~15 kcal/mol)[11] possibly due to the favorable interaction between the drug and the DNA base pairs. Huge estimate in the rotational and translational entropy loss (~11-14 kcal/mol[114]) was attributed to the binding process, calculated using Sackur-Tetrode[115] formula for entropy of monatomic ideal gases. Interestingly, this study[116] pointed out the predominant formation of B_{II} conformation of the DNA at the intercalation site. Earlier study of Winger *et al.*[117] showed that de-stacking of base pairs caused the inter-conversion from B_I to $B_{II.}$ Therefore, the inter-conversion of the DNA forms may be an effect of un-stacking itself, rather than intercalation. Similarly, ellipticine intercalation also showed a conversion of the backbone from B_I state to B_{II} state. In that case, higher entropy of the DNA-ellipticine complex originates from the sugar-phosphate group rather than from the bases.[118]

1.2.6.2. Quantum chemical calculation

Along with the molecular modeling studies, which probe into the overall structural and dynamical features, quantum chemical studies were applied to understand the specific drug-DNA interactions. Pack *et al.*[119] illustrated that the specificity of ethidium towards Y(3'-5')R sequence compared to R(3'-5')Y sequence is due to the difference in energy of unwinding of DNA during the intercalation process and not due to the stacking interaction between ethidium and base pairs. Here, the stacking interactions in both the cases remained the same. They have used semi empirical quantum chemical methods PCILO and CNDO/2 for the calculations. Pack *et al.*[120] used the same methods to study the effect of outside bound proflavine on the induction of conformational

changes of the DNA to open up the base pairs before the insertion of proflavine could happen[11]. These calculations were done by taking the crystal structure geometry of dCpdG-proflavine complex[49]. The sequence specificity was probed by taking all different possible base pairs of DNA, by mutating the bases G and C in the crystal structure. The nucleotide base pairs with proflavine and without proflavine have shown that proflavine has a preference for dCpdG over dGpdC as it requires 23.5 kcal/mol more energy to unwind dGpdC as compared to dCpdG. Similarly, it has a preference for dApdT over dTpdA. The calculations on the B-DNA complex containing both intercalated and outside bound state proflavine molecules showed that the neutralization of anionic phosphate charge does not affect the sequence specificity but it lowers down the energy for opening the base pairs for intercalation to happen. This study suggested the two step intercalation process where the outside bound form stabilizes the open form of intercalation site so that the intercalation can occur.[120] The sequence specificity of daunomycin to different polynucleotide sequences such as $d(CGTACG)_2$, $d(CGATCG)_2$, $d(CITACI)_2$, $d(TATATA)_2$, $d(CGCGCG)_2$ and $d(TACGTA)_2$ were performed by Pullman[121]. Among the regular alternate sequences $d(TATATA)_2$ and $d(CGCGCG)_2$, daunomycin showed specificity towards $d(TATATA)_2$. Among the six different hexanucleotides, mixed polynucleotides $d(CGATCG)_2$ and $d(CGTACG)_2$ showed strongest affinity towards daunomycin[121] in agreement with the X-ray crystal structure of Daunomycin-d(CpGpTpApCpG)[48]. These studies indicate that the selectivity of daunomycin towards the bases should be explained in terms of three base pairs rather than two base pairs.[48, 121] Since most of the intercalators bear a positive charge, it was believed that the electrostatic interaction between the cationic intercalator and negatively charged DNA plays the most important role in the stabilization of DNA-intercalator complex[122]. Reha *et al.*[123] investigated the stacking interaction energy between different intercalators and base pairs using high level *ab initio* calculations (MP2 and dispersion corrected density functional tight-binding (DFTB-D) methods). To understand the effect of stacking energy on the vertical distance between the base pairs and intercalator, the systems chosen were AT base pair with ethidium and DAPI, and GC base pair with daunomycin and ellipticine. The study showed that none of the intercalators accumulates charge in any of the sites, instead they have a large polarizability, which clearly shows that the dispersion energy

contributes much to the DNA-intercalator stability rather than electrostatic interactions. The study also showed a dependency of Twist of base pair on the stacking energy of different intercalators by changing the Twist from 0° to 360°. This methodology was adapted by Cooper *et al.* by applying van der Waals Density Functional (vdW-DF) to study the stacking interaction between the intercalator (neutral and charged proflavine, ellipticine) and the base pair. The larger polarizabilities of all three intercalators showed that the intercalation is mostly governed by dispersion interaction. The optimization of charged and neutral proflavine in presence of C-G base pair showed that the charged proflavine has stronger binding because of the presence of electrostatic interaction. Similar calculations are done for proflavine in presence of T-A base pair and showed that the complex is less energetically stable as compared to C-G complex [124]. *Ab initio* and semi-empirical quantum chemical calculations with molecular mechanics simulations were done to understand the role of enthalpy and entropy of ethidium intercalation by applying modified Cornell *et al.* empirical force field. The DNA sequences used were $d(GCATATATGC)_2$ and $d(GCGCGCGCGC)_2$ and for high level energy calculations a minimal model was used composed of intercalator molecule and dimers TA or GC. The hydrogen bond interaction of ethidium with DNA was found to be 10% of total interaction energy and most of the contribution comes from the dispersion energy as in the case of most of the intercalators. The interaction energies for ethidium with dimer TA and GC are comparable. The free energy change of intercalation for ethidium was calculated to be -4.5 kcal/mol using QM/MM methods.[123] The intercalation process causes deformability in the neighborhood of intercalation site and in turn facilitates the insertion of a second intercalator between the neighboring base pairs.[125] The stacking of phenanthroline between A-T and G-C were studied by *ab initio* methods such as MP2. The analysis summarized that phenanthroline has a preference to G-C base pair and suggested that MP2/6-31 +G(d,p) could be used further for studying the stacking interactions[126].

1.2.7. Molecular Mechanism of Intercalation

Molecular modeling and quantum chemical studies discussed above provide various information on the structural and thermodynamic aspects of the intercalation

process[114, 127] Experimental kinetic studies using spectroscopic methods do provide the overall ensemble average picture of the mechanism, i.e., timescale of the process and existence of different intermediate states, etc.[11, 95] However, a detail molecular picture of the process remains elusive. For example, kinetic studies depict intercalation process of proflavine as a two-step process[11], ethidium bromide intercalation as a single step[88], while for daunomycin both three-step[95] and five-step process[89]. The origin of these differences was, however, unknown although it was understood that these differences must lie with the nature of the intercalators[95] and a minimum two-step[11], fast outside bound state formation and slow activated intercalation, is required for intercalation. Moreover, the limiting hypothesis of intercalation discussed earlier, i.e., whether intercalation proceeds via natural fluctuation of the DNA or by the drug-induced cavity formation could not be explained from experimental results obtained[11].

As a first attempt on understanding the molecular mechanism of intercalation, Mukherjee *et al.* chose daunomycin as intercalating drug. Crystal structure of daunomycin[48] shows that at the intercalated state daunosamine group of the drug stays in the minor groove. This indicates that the intercalation of daunomycin takes place from minor groove side of DNA (otherwise, the final structure would have been different in this case). In this study, daunomycin was de-intercalated and finally taken to a far separated state. This provided with the free energy difference between the intercalated and separated state (-12.3 kcal/mol), in close agreement with the experimental value -9.4 kcal/mol [128].

1.3. Methods Used

Mechanism of a process is governed by the system moving along the minimum free energy path (MFEP). Therefore, here we will discuss some of the methods related to the calculation of free energy profile and MFEP. Studies of intercalation mechanism used primarily three techniques: (a) umbrella sampling and (b) metadynamics (c) well-tempered metadynamics.

1.3.1. Umbrella Sampling

Umbrella sampling[129] is one of the accelerated sampling methods in molecular dynamics simulations where the inherent barrier is overcome by employing an additional bias potential (Eq. 3) along a chosen reaction coordinate (RC) or collective variable (CV).

$$V_{umb}(s) = \frac{1}{2}k(s - s_0)^2, \qquad Eqn.3$$

where, k is the force constant, s_0 is the constant CV value at which the restraint is applied, s is dynamically changing CV. A complete free energy profile is obtained by multiple simulations along different s_0 value to cover the entire reaction profile. Each simulation provides a biased distribution of s. Correct free energy ($\tilde{F}(s)$) is obtained by unbiasing (Eq. 4) it in the following way[130],

$$\tilde{F}(s) = -\frac{1}{\beta}lnN(s) - V_{umb}(s) \qquad Eqn.4$$

where, $N(s)$ is the histogram of the visited configurations in the collective variable space, $V_{umb}(s)$ is the applied bias potential, $\beta = \frac{1}{k_B T}$, where k_B is the Boltzmann constant and T is the temperature (K).

Thus, all the unbiased free energy profiles obtained for different value of s are combined together using weighted histogram analysis method (WHAM)[131-132] to get the whole free energy profile. It is assumed that all the other 3N-1 coordinates (N = no. of atoms) are faster variable that get equilibrated for a given parametric value of the reaction coordinate. The choice of the RC is based on the chemical intuition. Therefore, if the choice is wrong, free energy obtained would certainly be associated with huge errors.

1.3.2. Metadynamics

Metadynamics[133] is a relatively new but robust accelerated sampling method particularly for the computation of multidimensional free-energy surfaces. The difference from umbrella sampling is that metadynamics is a non-equilibrium sampling. However, it provides an equilibrium free energy landscape. Metadynamics is built with the ideology that more a system stays in a particular point in the phase space, more destabilized it becomes so that it can explore the neighboring points in the phase space and thereby explore the complete free energy surface (FES). Although one needs to

describe collective variables to accelerate sampling along those CV's, the exploration is rather spontaneous. Technically it is achieved by addition of Gaussian potential at certain time interval to the location of the CV at that point of time. Gaussian potentials are added till the free energy surface becomes flat and thereby exploration of FES is spontaneous. The accuracy and speed of sampling can be altered by changing two parameters such as Gaussian height and Gaussian width as well as changing the time interval between additions of two Gaussian functions. The metadynamics potential[133] ($V_G(s(x),t)$) at a particular time, t, is equal to,

$$V_G(s(x),t) = \omega \sum_{\substack{t'=\tau_G,2\tau_G,3\tau_G\ldots \\ t'<t}} \exp\left(-\frac{\left(s(x)-s(x_G(t'))\right)^2}{2\delta s^2}\right) \qquad Eqn.5$$

ω = height of the Gaussian, δs = width of the Gaussian, τ_G = frequency at which the Gaussian is added,

$s(x)$ = Conformational space described by the collective variable (CV), $s(x_G(t'))$ = value taken by the collective variable at time *t'*, t = sampling time.

Free energy surface $F(s)$ in the conformational space $s(x)$ is the opposite of the sum of all Gaussian potential.[133] So,

$$F(s) \sim -\lim_{t\to\infty} V_G(s(x),t) \qquad Eqn.6$$

1.3.3. Well-tempered metadynamics

In normal metadynamics simulation, the free energy does not converge to a definite value leading to an average error. Continuing the simulation irreversibly pushes the system to regions of physically irrelevant configurational space. To overcome this problem, well-tempered metadynamics[134] was developed which has the advantage of controlling the regions of the free energy surface which are physically relevant.

In case of well-tempered metadynamics, the Gaussian height is rescaled during the simulation according to the following equation[134],

$$\omega = \omega_0\, e^{-\frac{V_G(s(x),t)}{k_B\Delta T}} \qquad Eqn.7$$

ω_0 = initial Gaussian height, k_B = Boltzmann's constant, ΔT = a parameter with dimension of temperature.

Free energy surface is estimated as[133-134],

$$\tilde{F}(s,t) = -\frac{T+\Delta T}{\Delta T} V_G(s(x),t) \quad Eqn.8$$

$T + \Delta T$ is called CV temperature, T simulation temperature, $\frac{T+\Delta T}{T}$ is called bias factor.

1.3.4. Minimum Free Energy Calculation

The calculation of the minimum free energy path (MFEP) is done using the algorithm of Ensing et al. [135] Figure 1.9 represents contour plot of a free energy surface made by two reaction co-ordinates S_1 and S_2. The letter A represents the reactant well and letters B and C represents two product wells. Many free energy paths are possible starting from the reactant well A to the product well B. The location of minimum free energy path between two states involves finding the minima at both reactant and product well. Initial guess is done as minimum points at both A and B around which the most of the time the reaction co-ordinates fluctuates around. Initially a coarse path (represented by the points P1, P2, P3 etc.) is traced from A to B using golden section method by taking steps larger than the local bumpiness in the free energy surface. Then the path is refined by optimizing the points between the two coarse points (represented by the inset in Fig. 1.9) in the direction perpendicular to the path. This refined path is the minimum free energy path in the free energy surface made by two collective variables S_1 and S_2.[135]

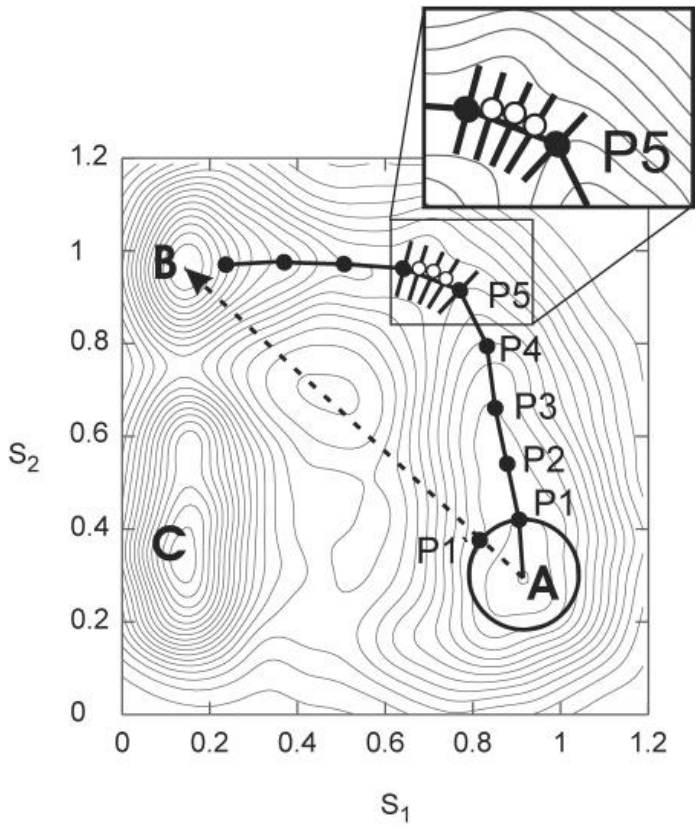


Figure 1.9. *Calculation of minimum free energy path on a 2D free energy surface made by the collective variables S_1 and S_2. (Figure adapted from Figure 1 of Ref. 135)*

1.4. Outline of the Thesis

In this thesis, in Chapter 2, we performed the direct intercalation of an anticancer agent proflavine from the minor groove of DNA revealing the molecular mechanism of the process. In the 3rd chapter, we describe the possible pathway and kinetics of the intercalation process using intercalation, deintercalation and dissociation of proflavine. Fourth chapter concentrates on the time scale of formation, nature and structure of "pre-intercalative" bound state of formation of proflavine. Fifth chapter illustrates the importance of water molecules in molecular recognition processes and estimation of water entropy.

1.5. References

1. Gilson, M. K.; Zhou, H.-X., *Ann. Rev. Biophys. Biomol. Str.* **2007,** *36* (1), 21-42.
2. Hamelberg, D.; McCammon, J. A., *J. Am. Chem. Soc.* **2004,** *126* (24), 7683-9.
3. Murphy, K. P., *Protein Structure, Stability, and Folding*. Humana Press: 2001.

4. Cleaves, H., II, Molecular Recognition. In *Encyclopedia of Astrobiology*, Gargaud, M.; Amils, R.; Quintanilla, J.; Cleaves, H., II; Irvine, W.; Pinti, D.; Viso, M., Eds. Springer Berlin Heidelberg: 2011; pp 1079-1080.
5. Baron, R.; McCammon, J. A., *Ann. Rev. Phys. Chem.* **2013,** *64* (1), 151-175.
6. Zewail, A. H., *Physical Biology: From Atoms to Medicine*. Imperial College Press: 2008.
7. Brooijmans, N.; Kuntz, I. D., *Ann. Rev. Biophys. Biomol. Str.* **2003,** *32* (1), 335-373.
8. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., *J. Comp. Chem.* **2004,** *25*, 1605-1612.
9. Waring, M. J., *DNA Modification and Cancer*. Annual Reviews: 1981.
10. Lerman, L. S., *J. Mol. Biol.* **1961,** *3*, 18-30.
11. Li, H. J.; Crothers, D. M., *J. Mol. Biol.* **1969,** *39* (3), 461-&.
12. Waring, M. J., *J. Mol. Biol.* **1970,** *54*, 247-279.
13. Chacón-García; Martínez, L. R., *Curr. Med. Chem.* **2005,** *12*, 127-151.
14. Maher, L. J., *Current Opinion in Chemical Biology* **1998,** *2* (6), 688-694.
15. Singh, R. K.; Sasikala, W. D.; Mukherjee, A., *J. Phys. Chem. B* **2015,** *119* (35), 11590-11596.
16. Berman, H. M.; Young, P. R., *Ann. Rev. Biophys. Bioeng.* **1981,** *10*, 87-114.
17. Denny, W. A., *Curr. Med. Chem.* **2002,** *9*, 1655-1665.
18. Lenglet, G.; David-Cordonnier, M. H. *J. Nucl. Acids* **2010**, 1-17.
19. Carmen Avendano; Carlos Menendez, *Med. Chem. Anticancer Drug.* **2008**, 14.
20. Hurwitz, J.; Furth, J. J.; Malamy, M.; Alexander, M., *Proc. Natl. Acad. Sci.* **1962,** *48* (7), 1222-1230.
21. Waring, M. J., *Mol. Pharmacol.* **1965,** *1* (1), 1-13.
22. Cavalieri, L. F.; Nemchin, R. G., *Biochim. Biophys. Acta - Specialized Section on Nucleic Acids and Related Subjects* **1964,** *87* (4), 641-652.
23. Müller, S., *Nucleic Acids from A to Z*. Wiley: 2008.
24. Clark, D. P.; Pazdernik, N. J., *Molecular Biology*. Academic Press: 2012.
25. Giovanni, C.; Carla, S.; Franco, Z., *Cancer Res.* **1986,** *46*, 5499-5503.

26. Ross, W. E.; Bradley, M. O., *Biochim. Biophys. Acta - Nucleic Acids and Protein Synthesis* **1981,** *654* (1), 129-134.

27. Olweny, C. L. M.; Toya, T.; Katongole-Mbidde, E.; Lwanga, S. K.; Owor, R.; Kyalwazi, S.; Vogel, C. L., *Int. J. Cancer* **1974,** *14* (5), 649-656.

28. Malogolowkin, M.; Cotton, C. A.; Green, D. M.; Breslow, N. E.; Perlman, E.; Miser, J.; Ritchey, M. L.; Thomas, P. R. M.; Grundy, P. E.; D'Angio, G. J.; Beckwith, J. B.; Shamberger, R. C.; Haase, G. M.; Donaldson, M.; Weetman, R.; Coppes, M. J.; Shearer, P.; Coccia, P.; Kletzel, M.; Macklis, R.; Tomlinson, G.; Huff, V.; Newbury, R.; Weeks, D., *Pediatr. Blood Cancer* **2008,** *50* (2), 236-241.

29. Bradof, J. E.; Hakes, T. B.; Ochoa, M.; Golbey, R., *Cancer* **1982,** *50* (6), 1070-1075.

30. Giermasz A, M. M., Nowis D, Jalili A, Maj M, Dabrowska A, Czajka A, Jakobisiak M, Golab J., *Oncol. Rep.* **2002,** *9* (1), 199.

31. Brunnberg, U.; Mohr, M.; Noppeney, R.; Dürk, H. A.; Sauerland, M. C.; Müller-Tidow, C.; Krug, U.; Koschmieder, S.; Kessler, T.; Mesters, R. M.; Schulz, C.; Kosch, M.; Büchner, T.; Ehninger, G.; Dührsen, U.; Serve, H.; Berdel, W. E., *Ann. Oncol.* **2012,** *23* (4), 990-996.

32. Samuels, L. D.; Newton, W. A.; Heyn, R., *Cancer* **1971,** *27* (4), 831-834.

33. Kantarjian, H. M.; Talpaz, M.; Kontoyiannis, D.; Gutterman, J.; Keating, M. J.; Estey, E. H.; O'Brien, S.; Rios, M. B.; Beran, M.; Deisseroth, A., *J. Clin. Oncol.* **1992,** *10* (3), 398-405.

34. Minuk LA, M. K., Chin-Yee IH, Lazo-Langner A, Bhagirath V, Chin-Yee BH, Mangel JE., *Leuk. Lymphoma* **2012,** *53* (1), 57-63.

35. Harousseau, J. L.; Reiffers, J.; Hurteloup, P.; Milpied, N.; Guy, H.; Rigal-Huguet, F.; Facon, T.; Dufour, P.; Ifrah, N., *J. Clin. Oncol.* **1989,** *7* (1), 45-49.

36. Silver, R. T.; Case, D. C.; Wheeler, R. H.; Miller, T. P.; Stein, R. S.; Stuart, J. J.; Peterson, B. A.; Rivkin, S. E.; Golomb, H. M.; Costanzi, *J. Clin. Oncol.* **1991,** *9* (5), 754-761.

37. Canals, A.; Purciolas, M.; Aymami, J.; Coll, M., *Acta Crystallogr. Sect. D* **2005,** *61* (7), 1009-1012.

38. Horstmann, M. A.; Hassenpflug, W. A.; zur Stadt, U.; Escherich, G.; Janka, G.; Kabisch, H., *Haematologica* **2005,** *90* (12), 1701-1703.
39. Bazzicalupi, C.; Bencini, A.; Bianchi, A.; Biver, T.; Boggioni, A.; Bonacchi, S.; Danesi, A.; Giorgi, C.; Gratteri, P.; Ingrain, A. M.; Secco, F.; Sissi, C.; Valtancoli, B.; Venturini, M., *Chemistry* **2008,** *14* (1), 184-96.
40. Baruah, H.; Bierbach, U., *Nucleic Acids Res.* **2003,** *31* (14), 4138-4146.
41. RP., H.-J., *Ann. Trop. Med. Parasitol.* **1949,** *43*, 345-348.
42. Valdés, A. F.-C., *Open Med. Chem. J.* **2011,** *5*, 11-20.
43. Figgitt, D.; Denny, W.; Chavalitshewinkoon, P.; Wilairat, P.; Ralph, R., *Antimicrob. Agents Chemother.* **1992,** *36* (8), 1644-1647.
44. Wainwright, M., *J. Antimicrob. Chemother.* **2001,** *47*, 1-13.
45. Browing, C. H., Gulbransen, R, Kennaway, E.L., Thornton, L.H.D, *Brit. Med. J.* **1917,** *i*, 73-78.
46. Xu, Y.; Wang, Y.; Yan, L.; Liang, R.-M.; Dai, B.-D.; Tang, R.-J.; Gao, P.-H.; Jiang, Y.-Y., *J. Proteome. Res.* **2009,** *8* (11), 5296-5304.
47. Hyeon-Hee Yu, K.-J. K., Jeong-Dan Cha, Hae-Kyoung Kim, Young-Eun Lee, Na-Young Choi, and Yong-Ouk You, *J. Med. Food* **2005,** *8* (4), 454-461.
48. Quigley, G. J.; Wang, A. H. J.; Ughetto, G.; Marel, G. V. D.; Boom, J. H. V.; Rich, A., *Proc. Natl. Acad. Sci. USA* **1980,** *77* (12), 7204-7208.
49. Neidle, S.; Jones, T. A., *Nature* **1975,** *253* (5489), 284-285.
50. Hu, G. G.; Shui, X.; Leng, F.; Priebe, W.; Chaires, J. B.; Williams, L. D., *Biochemistry* **1997,** *36* (20), 5940-5946.
51. Pfoh, R.; Cuesta-Seijo, J. A.; Sheldrick, G. M., *Acta Crystallogr. Sect. F* **2009,** *65* (7), 660-664.
52. Wang, A. H.; Ughetto, G.; Quigley, G. J.; Hakoshima, T.; van der Marel, G. A.; van Boom, J. H.; Rich, A., *Science* **1984,** *225* (4667), 1115-1121.
53. Gao, Q.; Williams, L. D.; Egli, M.; Rabinovich, D.; Chen, S. L.; Quigley, G. J.; Rich, A., *Proc. Natl. Acad. Sci. USA* **1991,** *88* (6), 2422-2426.
54. Spielmann, H. P.; Wemmer, D. E.; Jacobsen, J. P., *Biochemistry* **1995,** *34* (27), 8542-8553.
55. Günther, K.; Mertig, M.; Seidel, R., *Nucleic Acids Res.* **2010,** *38* (19), 6526-6532.

56. Torigoe, H.; Sato, S.; Yamashita, K.-i.; Obika, S.; Imanishi, T.; Takenaka, S., *Nucleic. Acids Symp. Ser.* **2002,** *2* (1), 55-56.
57. Wilson, W. D.; Ratmeyer, L.; Zhao, M.; Strekowski, L.; Boykin, D., *Biochemistry* **1993,** *32* (15), 4098-4104.
58. Lorean Dean Williams, M. E., Qi Gao and Alexander Rich, *Struct. Funct.* **1992,** *1*, 107.
59. Alden, C. J.; Arnott, S., *Nucleic Acids Res.* **1975,** *2* (10), 1701-1718.
60. Comba, P., *Struct. Funct.*. Springer: 2010.
61. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., *Nucl. Acids Res.* **2000,** *28* (1), 235-242.
62. Cohen, G.; Eisenberg, H., *Biopolymers* **1969,** *8* (1), 45-55.
63. Kersten, W.; Kersten, H.; Szybalski, W.; Fiandt, M., *Biochemistry* **1966,** *5* (1), 236-244.
64. Neville Jr, D. M.; Davies, D. R., *J. Mol. Biol.* **1966,** *17* (1), 57-IN3.
65. Waring, M. J., *Nature* **1968,** *219* (5161), 1320-&.
66. Chaires, J. B.; Dattagupta, N.; Crothers, D. M., *Biochemistry* **1982,** *21* (17), 3933-40.
67. Waring, M. J., *Biochem J.* **1974,** *143* (2), 483-486.
68. Crawford, L. V.; Waring, M. J., *J. Mol. Biol.* **1967,** *25* (1), 23-30.
69. Newman, J., *Biopolymers* **1984,** *23* (6), 1113-1119.
70. Chirico, G.; Lunelli, L.; Baldini, G., *Biophys. Chem.* **1990,** *38* (3), 201-211.
71. Nordmeier, E., *J. Phys. Chem.* **1992,** *96* (14), 6045-6055.
72. Luzzati, V.; Masson, F.; Lerman, L. S., *J. Mol. Biol.* **1961,** *3* (5), 634-639.
73. Jain, S. C.; Sobell, H. M., *J. Mol. Biol.* **1972,** *68* (1), 1-20.
74. Tsai, C. C.; Jain, S. C.; Sobell, H. M., *Proc. Natl. Acad. Sci.* **1975,** *72* (2), 628-632.
75. Neidle, S.; Achari, A.; Taylor, G. L.; Berman, H. M.; Carrell, H. L.; Glusker, J. P.; Stallings, W. C., *Nature* **1977,** *269* (5626), 304-307.
76. Neidle, S.; Berman, H. M.; Shieh, H. S., *Nature* **1980,** *288* (5787), 129-133.
77. Baldini, G.; Doglia, S.; Dolci, S.; Sassi, G., *Biophys. J.* **1981,** *36* (3), 465-477.
78. G. Baldini, B. C., G. Lucchin and A. Zedda, *Il Nuovo Cimento* **1982,** *1 D* (5).

79. Nafisi, S.; Saboury, A. A.; Keramat, N.; Neault, J.-F.; Tajmir-Riahi, H.-A., *J. Mol. Struct.* **2007,** *827* (1–3), 35-43.
80. Chaires, J. B., *Biochemistry* **1983,** *22* (18), 4204-11.
81. Breslauer, K. J.; Remeta, D. P.; Chou, W.-Y.; Ferrante, R.; Curry, J.; Zaunczkowski, D.; Snyder, J. G.; Marky, L. A., *Proc. Natl. Acad. Sci. USA* **1987,** *84* (24), 8922-8926.
82. Garbett, N. C.; Hammond, N. B.; Graves, D. E., *Biophys. J.* **2004,** *87* (6), 3974-3981.
83. Herrera, J. E.; Chaires, J. B., *Biochemistry* **1989,** *28* (5), 1993-2000.
84. Xodo, L. E.; Manzini, G.; Ruggiero, J.; Quadrifoglio, F., *Biopolymers* **1988,** *27* (11), 1839-1857.
85. Ji Hye Kwon, H.-J. P., Nataraj Chitrapriya, Sung Wook Han, Gil Jun Lee, Dong Jin Lee, and Tae-Sub Cho, *Bull. Korean Chem. Soc.* **2013,** *34* (3), 810-814.
86. Hossain, M.; Suresh Kumar, G., *Mol. Biosyst.* **2009,** *5* (11), 1311-1322.
87. Macgregor, R. B.; Clegg, R. M.; Jovin, T. M., *Biochemistry* **1987,** *26* (13), 4008-4016.
88. Macgregor, R. B.; Clegg, R. M.; Jovin, T. M., *Biochemistry* **1985,** *24* (20), 5503 - 5510.
89. Rizzo, V.; Sacchi, N.; Menozzi, M., *Biochemistry* **1989,** *28* (1), 274-282.
90. Biver, T.; Secco, F.; Venturini, M., *Coord. Chem. Rev.* **2008,** *252* (10-11), 1163-1177.
91. Muller, W.; Crothers, D. M., *J. Mol. Biol.* **1968,** *35* (2), 251-&.
92. Bresloff, J. L.; Crothers, D. M., *J. Mol. Biol.* **1975,** *95* (1), 103-110.
93. Wilson, W. D.; Krishnamoorthy, C. R.; Wang, Y. H.; Smith, J. C., *Biopolymers* **1985,** *24* (10), 1941-1961.
94. Westerlund, F.; Wilhelmsson, L. M.; Nordén, B.; Lincoln, P., *J. Phys. Chem. B* **2005,** *109* (44), 21140-21144.
95. Chaires, J. B.; Dattagupta, N.; Crothers, D. M., *Biochemistry* **1985,** *24* (2), 260-267.
96. Forster, W.; Stutter, E., *Int. J. Biol. Macromol.* **1984,** *6* (3), 114-124.
97. Fuller, W.; Waring, M. J., *Ber. Bunsenges Phys. Chem.* **1964,** *68* (8-9), 805-808.

98. Tsai, C.-c.; Jain, S. C.; Sobell, H. M., *J. Mol. Biol.* **1977,** *114* (3), 301-315.

99. Jain, S. C.; Tsai, C.-c.; Sobell, H. M., *J. Mol. Biol.* **1977,** *114* (3), 317-331.

100. Sobell, H. M.; Tsai, C.-C.; Jain, S. C.; Gilbert, S. G., *J. Mol. Biol.* **1977,** *114* (3), 333-365.

101. Nuss, M. E.; Marsh, F. J.; Kollman, P. A., *J. Am. Chem. Soc.* **1979,** *101* (4), 825-833.

102. Dearing, A.; Weiner, P.; Kollman, P. A., *Nucl. Acids Res.* **1981,** *9* (6), 1483-1498.

103. Singh, U. C.; Pattabiraman, N.; Langridge, R.; Kollman, P. A., *Proc. Natl. Acad. Sci.* **1986,** *83* (17), 6402-6406.

104. Rao, S. N.; Kollman, P. A., *Proc. Natl. Acad. Sci.* **1987,** *84* (16), 5735-5739.

105. Dearing, A., *University of California, San Francisco* **1981**.

106. Neidle, S.; Pearl, L. H.; Herzyk, P.; Berman, H. M., *Nucl. Acids Res.* **1988,** *16* (18), 8999-9016.

107. Swaminathan, S.; Beveridge, D. L.; Berman, H. M., *J. Phys. Chem.* **1990,** *94* (11), 4660-4665.

108. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., *J. Chem. Theor. Comput.* **2008,** *4* (3), 435-447.

109. Hopfinger, A. J.; Cardozo, M. G.; Kawakami, Y., *J. Chem. Soc. Faraday Trans.* **1995,** *91* (16), 2515-2524.

110. Howerton, S. B.; Nagpal, A.; Dean Williams, L., *Biopolymers* **2003,** *69* (1), 87-99.

111. Elcock, A. H.; Rodger, A.; Richards, W. G., *Biopolymers* **1996,** *39* (3), 309-326.

112. Jan Mazerski, I. A. a. S. M., *Acta Biochim. Pol.* **2000,** *47* (1), 47-57.

113. Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B. P.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 11*, University of California: San Francisco, 2010.

114. Trieb, M.; Rauch, C.; Wibowo, F. R.; Wellenzohn, B.; Liedl, K. R., *Nucl. Acids Res.* **2004,** *32* (15), 4696-703.

115. Seddon, J. M.; Gale, J. D., *Thermodynamics and statistical mechanics*. Royal Soc. of Chemistry: 2001.

116. Trieb, M.; Rauch, C.; Wellenzohn, B.; Wibowo, F.; Loerting, T.; Mayer, E.; Liedl, K. R., *J. Biomol. Struct. Dyn.* **2004,** *21* (5), 713-24.

117. Winger, R. H.; Liedl, K. R.; Rüdisser, S.; Pichler, A.; Hallbrucker, A.; Mayer, E., *J. Phys. Chem. B* **1998,** *102* (44), 8934-8940.

118. Kolar, M.; Kubar, T.; Hobza, P., *J Phys Chem B* **2010,** *114* (42), 13446-54.

119. Pack, G. R.; Loew, G., *Biochim. Biophys. Acta - Nucleic Acids and Protein Synthesis* **1978,** *519* (1), 163-172.

120. Pack, G. R.; Hashimoto, G. M.; Loew, G. H., *Ann. N.Y. Acad. Sci.* **1981,** *367* (1), 240-249.

121. Chen, K. X.; Gresh, N.; Pullman, B., *J. Biomol. Struct. Dyn.* **1985,** *3* (3), 445-466.

122. Medhi, C.; Mitchell, J. B. O.; Price, S. L.; Tabor, A. B., *Biopolymers* **1999,** *52* (2), 84-93.

123. Reha, D.; Kabelac, M.; Ryjacek, F.; Sponer, J.; Sponer, J. E.; Elstner, M.; Suhai, S.; Hobza, P., *J. Am. Chem. Soc.* **2002,** *124* (13), 3366-76.

124. Li, S.; Cooper, V. R.; Thonhauser, T.; Lundqvist, B. I.; Langreth, D. C., *J. Phys. Chem. B* **2009,** *113* (32), 11166-11172.

125. Kubař, T.; Hanus, M.; Ryjáček, F.; Hobza, P., *Chem. Eur. J.* **2006,** *12* (1), 280-290.

126. Pankaj Hazarika, B. B., Pranjali Das, Okhil Kumar Medhi, Chitrani Medhi, *J. Biophys. Chem.* **2011,** *2* (2), 152-157.

127. Baginski, M.; Fogolari, F.; Briggs, J. M., *J. Mol. Biol.* **1997,** *274* (2), 253-267.

128. Chaires, J. B.; Satyanarayana, S.; Suh, D.; Fokt, I.; Przewloka, T.; Priebe, W., *Biochemistry* **1996,** *35* (7), 2047-2053.

129. Torrie, G. M.; Valleau, J. P., *J. Comp. Phys.* **1977,** *23*, 187-199.

130. Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R., *Comput. Phys. Commun.* **2009,** *180* (10), 1961-1972.

131. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M., *J. Comp. Chem.* **1992,** *13*, 1011-1021.

132. Roux, B., *Comput. Phys. Commun.* **1995,** *91*, 275-282.

133. Alessandro, L.; Francesco, L. G., *Rep. Prog. Phys.* **2008,** *71* (12), 126601.

134. Barducci, A.; Bussi, G.; Parrinello, M., *Phys. Rev. Lett.* **2008,** *100* (2), 020603.

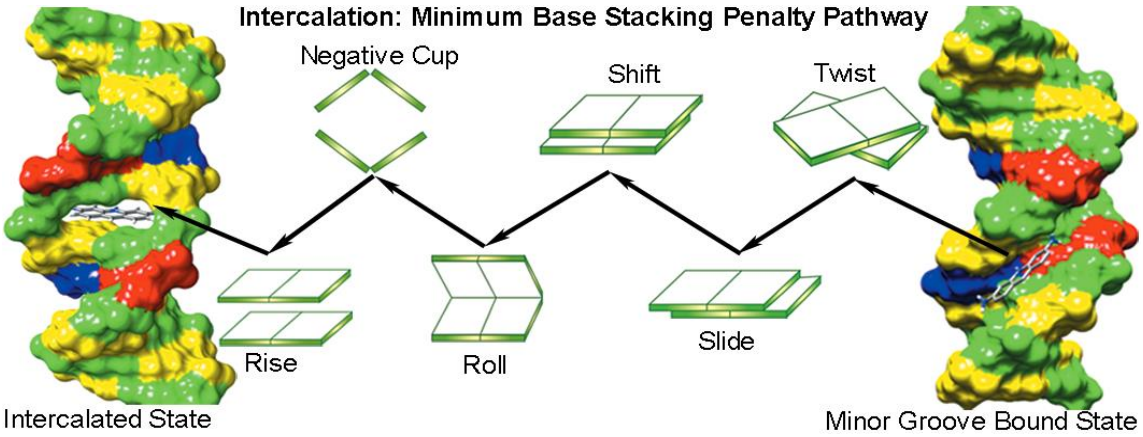
135. Ensing, B.; Laio, A.; Parrinello, M.; Klein, M. L., *J. Phys. Chem. B* **2005,** *109* (14), 6676-6687.

Chapter 2

Molecular Mechanism of Direct Proflavine-DNA Intercalation: Evidence for Drug-Induced Minimum Base-Stacking Penalty Pathway

Chapter 2. Molecular Mechanism of Direct Proflavine-DNA Intercalation: Evidence for Drug-Induced Minimum Base-Stacking Penalty Pathway

DNA intercalation, a biophysical process of enormous clinical significance, had surprisingly eluded molecular understanding for several decades. In an attempt to understand the molecular mechanism of intercalation process, a simple flat aromatic intercalator proflavine was selected as a prototype molecule. The docked minor groove bound state conformation of proflavine-DNA was selected as the outside bound state of the intercalation process. The free energy surface of direct intercalation from minor groove-bound state was captured for the first time using anti-cancer agent proflavine using all-atom metadynamics simulations with appropriate configurational restraint to prevent dissociation of the molecule from the outside bound state. Mechanism along the minimum free energy path reveals that intercalation happens through a minimum base stacking penalty pathway where non-stacking parameters (Twist->Slide/Shift) change first, followed by base stacking parameters (Buckle/Roll->Rise). The initial change in the Twist is followed by the change in both Slide and Shift. These changes are accompanied by a change in the Buckle and Roll, ultimately leading to a change in the Rise. The sequence of changes happening to DNA suggests that the mechanism of intercalation defies the natural fluctuation hypothesis and provides molecular evidence for the drug induced cavity formation hypothesis. The thermodynamic origin of the barrier is found to be a combination of entropy and desolvation energy.



2.1. Introduction

DNA intercalation[1-2], first proposed by Lerman in 1961, is the preferred mode of binding of certain anticancer and antimicrobial agents[3]. Intercalation subsequently leads to inhibition of transcription, replication and various enzymes leading to cell death.[4] Despite its clinical and biophysical implications, the molecular mechanism of intercalation remained elusive for several decades. Fluorescence properties of the intercalating drugs facilitated the several thermodynamic and kinetic studies[5-7] while structural investigations were carried out using NMR[8], crystallographic studies[9], molecular modelling[10-12], force spectroscopic methods[13] etc. The kinetic studies suggested the intercalation process of proflavine to be a two-step process[5], ethidium bromide intercalation as a single step[14], and daunomycin intercalation as both three-step [15] and five-step process[16]. Most of the kinetic studies on intercalation indicate that an intercalation process comprises *at least* two steps. The first step is the fast outside bound state formation that happens on a time scale of sub-millisecond. The second step, intercalation, happens slowly with a time scale of milliseconds. Two limiting hypothesis had emerged depicting the slow intercalation process: natural fluctuation hypothesis and drug induced cavity formation as shown in Fig. 2.1. According to the first hypothesis, the base pairs undergo natural thermal fluctuations that lead to the flipping of the base pairs towards outside. This causes the opening up of the base pairs momentarily. If the intercalating agent is sufficiently close to the gap made by the flipping of bases, then it slips inside the gap. Again, due to the thermal fluctuations, the flipped bases revert to undergo base pairing thereby stabilizing the intercalation complex. On the other hand, the drug induced cavity formation does not require prior base pair opening. Here, the ligand stays close to the DNA and induce some structural changes to the DNA to form a cavity through which the drug slips in.[5] There had been a long debate on the two mechanisms of intercalation process mentioned above. But both computational and experimental methods failed to resolve the debate as it is very difficult to capture the intermediate states of the intercalation process. Crystallographic methods were able to capture the final intercalated state. The outside bound and other intermediate states, however, remained elusive as their formation is too fast and they are less stable compared to the intercalated state. The intercalated state is the unique bound state (similar to the native state of a

protein) compared to several possible unbound states (unfolded protein). Also, the typical timescale for intercalation of small molecules is milliseconds[5]. Therefore, direct intercalation process is extremely challenging to be captured through computational studies.

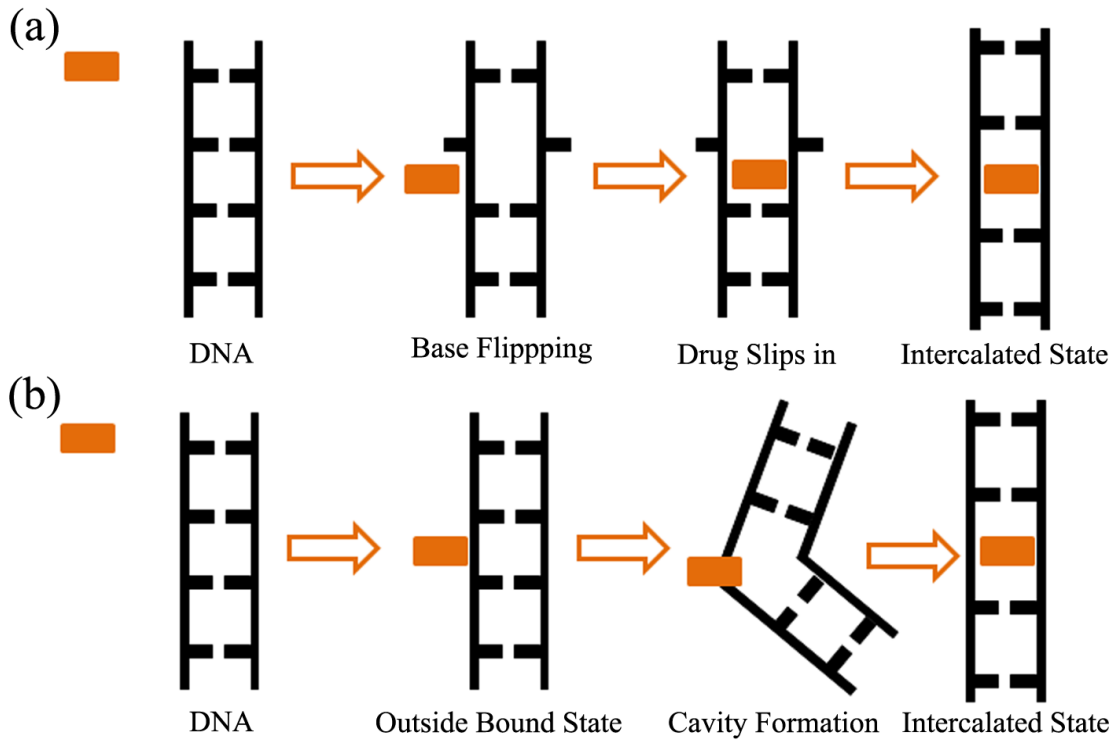


Figure 2.1: *Schematic representations illustrating two limiting hypotheses of intercalation process. (a) Represents the natural fluctuation hypothesis and (b) represents drug induced cavity formation hypothesis. The black vertical lines represent the backbone of double helical DNA and the horizontal lines represent the bases. The orange colored box represents the flat intercalator.*

In a first attempt to understand the molecular mechanism of intercalation, Mukherjee *et al.* chose an anticancer compound, daunomycin, as an intercalating agent. Crystal structure of daunomycin[17] confirmed the formation of intercalated state and showed that in the intercalated state, the daunosamine group of the drug stays in the minor groove. From the intercalated geometry, it was suggested that the intercalation of daunomycin took place from minor groove side of DNA (otherwise, the final structure would have been different in this case). To understand the mechanism of intercalation, Mukherjee *et al*[18] performed the deintercalation of daunomycin from the intercalated

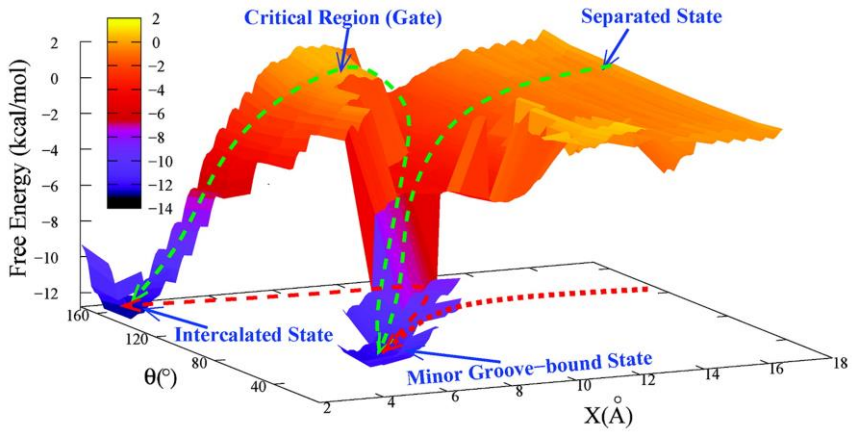


Figure.2.2: *Potential of mean force (PMF) of umbrella sampling simulations along the reaction coordinate X and θ. The completely dissociated state (X ~ 16 Å) is considered as reference zero free energy value. The green dashed line with an arrow (the red projections on to the X-θ plane) shows the schematic representation of the most probable path from completely separated → minor-groove bound → intercalated state. Figure adapted with permission from Ref.18 Copyright (2008) American Chemical Society.*

geometry. The deintercalation process was chosen because it is similar to the protein unfolding where one state can go to many states while intercalation is like protein folding where many possible states have to go to one particular native state. They used two collective variables X (distance with vector nature) and θ (angle) separately to de-intercalate the drug from the DNA using umbrella sampling method. The combined free energy surface is shown in Fig.2.2. Figure 2.2. shows the overall mechanism of intercalation of daunomycin where binding of the molecule to the minor groove of DNA by a diffusion limited process is followed by the slow active intercalation process with a barrier 12 kcal/mol that qualitatively matches with the experimental values. But the FES showed in Fig.2.2 was a combination of two different pathways: deintercalation from intercalated state to form minor groove bound state and dissociation of minor groove bound state. Therefore, a direct path between the minor groove-bound to intercalated state, i.e., the path of intercalation was not obtained. Emergence of a third step has been shown recently through conformational sampling of the daunomycin and DNA.[19] While both the studies provide significant understanding towards molecular detail of the process, a successful intercalation event that would suggest pre-intercalative mechanistic details, however, has not been achieved. The recent force-spectroscopic study indicated that intercalation and deintercalation are distinct processes and can occur in different

timescales[13]. So, one of our aims was to understand the molecular mechanism through direct intercalation rather than deintercalation. By looking at the complex nature of daunomycin molecule, it is predictable that it is not suitable for computationally challenging direct intercalation process for the first attempt. So we chose a simple molecule proflavine for setting the computational protocol which could be further extended to daunomycin and other DNA intercalators.

The present chapter deals with the construction of free energy surface of intercalation of an anticancer[20] and antimicrobial agent[21] proflavine directly from the outside bound state using computational methods. To understand the intercalation process at the molecular level, we chose proflavine because it is a simple, flat, symmetric and aromatic DNA intercalator. The simple and symmetric nature of proflavine makes the molecule equally accessible by both minor and major groove of the DNA thereby, representing different varieties of DNA intercalators. By using appropriate configurational restraints on collective variables or reaction co-ordinates, we achieved direct intercalation of proflavine from minor groove bound state using extensive all-atom simulations with well-tempered metadynamics[22] method. The present work also addresses a long-standing debate on the intercalation mechanism and provides insight to design better and more efficient intercalators.

2.2. Experimental Section

2.2.1. The Preparation of Minor groove Bound State.

A twelve base pair DNA, $(GCGCTCGAGCGC)_2$, was created using nucleic acid builder (NAB)[23] program. AMBER99/parmbsc0 force-field[24-25] has been used for the DNA. Since proflavine exists as a cation in the physiological system, protonated form of proflavine was optimized followed by Merz-Kolmann charge calculation[26] using Hartree-Fock theory with 6-31G* basis set using Gaussian03[27]. Antechamber module of AMBERTools[28] was used for a restrained electrostatic potential charge (RESP)[29] calculation and generation of general amber force field (GAFF)[30]. Finally, the co-ordinates and topology were converted to GROMACS format using an amb2gmx.pl program[31].

Minor groove bound state was created by docking using AutoDock software[32-33]. The AMBER charges of the nucleic acid atoms and quantum chemically derived charges

for proflavine were kept unchanged during docking process. During the docking proflavine was kept flexible and DNA was kept rigid. The grid was generated on whole BDNA with grid points of 96, 104 and 96 in orthogonal directions respectively with a spacing of 0.54 Å and search was performed using Genetic algorithm[32]. All the 10 structures obtained after docking were minor groove-bound states, amongst which the best docked structure with binding score -6.25 kcal/mol was selected as initial configuration of minor groove-bound state (X= 2.0 Å and φ= 76.9˚).

2.2.2. Molecular Dynamic Simulation Protocol

The drug-DNA system was solvated by 11763 TIP3P water molecules[34] in a large cubic box of length 71.23 Å. Twenty-two sodium ions and one chlorine ion were added to neutralize the system. Final system of 36096 atoms was energy minimised using steepest descent method[35] followed by heating up to 300K using Berendsen thermostat[36] with coupling constant 0.2 ps and a harmonic restraint of 25 kcal/mol/Å^2 on heavy atoms of DNA and proflavine. After the energy minimization, the restraints were slowly reduced to 0.5 kcal/mol/Å^2 during a series of six simulations of 50 ps each at constant temperature 300 K and pressure 1 bar using Berendsen thermostat and barostat[36], followed by energy minimization for 1000 steps. Then a final unrestrained equilibration for 1 ns was done at constant temperature 300 K and pressure 1 bar using Nose-Hoover thermostat[37-38] with coupling constant 0.2 ps and Parrinello-Rahman barostat[39] with coupling constant 0.2 ps. Time step of each simulation was kept at two fs. The electrostatic interaction was treated using PME (Particle Mesh Ewald)[40] with cut-off at 10 Å and van der Waals cut-off was taken to be 10 Å.

Well-tempered metadynamics[22] was performed thereafter. It is a smoothly converging and tunable free energy method, which offers possibility of controlling the regions of FES that are physically meaningful to explore and provides a unified framework of standard metadynamics and unbiased standard sampling.[22] We used bias factor 15 and added Gaussian height 0.048 kcal/mol at 2 ps interval for intercalation and dissociation processes. Widths (σ) of the Gaussians were varied for different collective variables depending on the natural fluctuation observed in different states. For intercalation and dissociation, we used σ = 0.5 Å for X and σ = 0.2 radians for φ. GROMACS[41] molecular dynamics program package has been used to carry out the

simulations. A separate program patch, PLUMED[42], was used with modification to incorporate our collective variables to perform well-tempered metadynamics simulations.

2.2.3. Description of Collective Variables

For a complex molecular mechanism involving 3N (N is the number of atoms), degrees of freedom, proper choice of the CVs (along which the reaction is monitored) is crucial. Here, we used two different CVs (X,φ) shown schematically in Fig. 2.3. X denotes the distance of proflavine from the DNA approximately perpendicular to the helical axis. X is positive for displacement of the drug towards the minor groove side and negative for displacement towards the major groove side. Second collective variable φ measures deviation of the drug's position from the body-fixed vector, $\hat{b}$. Therefore, φ measures the movement of the molecule along the helical axis of the DNA. φ can vary from 0° to 90° in the minor groove (+ve X) and 90° to 180° in the major groove (-ve X) by construction. The selected collective variables help to distinguish between intercalated state and outside bound state. Here, proflavine at the intercalated state is characterized by both X and φ close to zero. Figure 2.3.b show the initial configuration of the minor groove-bound state. The collective variable values are also shown. The higher value of X (>0 nm) and φ (>0^0 and <90^0) represents the minor groove bound state.

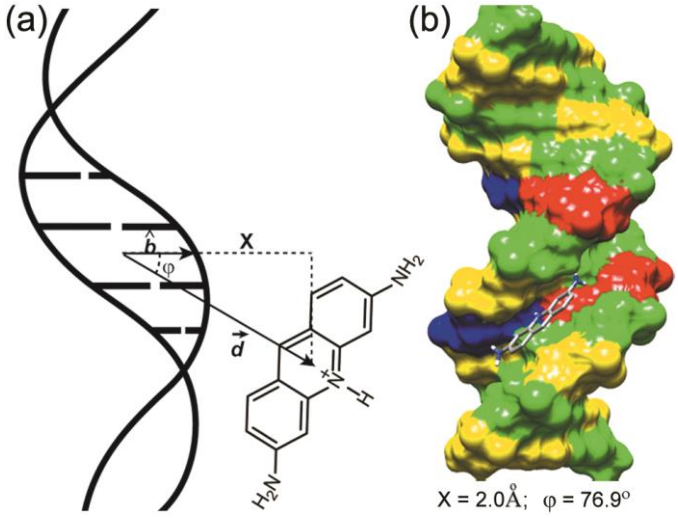


Figure 2.3.: *Schematic picture of the collective variables and initial drug-DNA configuration. (a) $\hat{b}$ is the unit vector from center of mass (COM) of the intercalating base pairs (IBP) consisting of four bases (C6,G7,C18,G19) to the COM of G7 and G19 ribose sugars that lie more towards the minor groove. $\vec{d}$ is the vector from the COM of IBP to the COM of proflavine $X = \hat{b}.\vec{d}$, $\varphi = Cos^{-1}(\hat{b}.\vec{d}/|\vec{d}|)$. (b) Initial configuration of the docked minor groove-bound state. Note the values of X and φ.*

2.2.4. Configurational Restraint

To capture the direct intercalation of proflavine from minor groove bound state, configurational restraints were applied on the collective variables to prevent the escape of molecule to the completely dissociated state. For simulations, a quartic potential[42] of the following form was used to apply configurational restraint,

$$V_{wall}(S) = \kappa(s - s_0)^4, \qquad (1)$$

where, κ represents the force constant taken to be 6 kcal mol^{-1}Å^{-2} in this study. s is the collective variable; s_0 is the cut-off value where the potential is applied.

2.2.5. MFEP Calculation and Cluster Analysis

We calculated the minimum free energy path (MFEP) for each process using the algorithm of Ensing et al.[43] Subsequently, we collected the structures that fall within a small deviation $(dX, d\varphi)$ round each point, spaced to avoid overlap, along the MFEP. Then we performed cluster analysis of the configurations collected around each point using RMSD cut-off 1.0 Å. Middle structures of the biggest cluster for each point were selected and the corresponding trajectory was made for each point in MFEP. All the subsequent calculations of average properties, standard deviations of DNA parameters and energy components (plotted Figures 2.7, 2.8 and 2.9) were performed on those trajectories containing the members of the best cluster at each point in MFEP.

2.3. Results and Discussion

2.3.1. Design and Method

We performed well-tempered metadynamics simulation from the minor groove-bound state using two reaction coordinates X and φ (Fig.2.3). X denotes the distance of proflavine from the DNA approximately perpendicular to the helical axis. φ measures displacement of the drug along the helical axis of the DNA. Therefore, φ can distinguish between intercalated (φ ~ 0°) and minor groove-bound state (φ ~ 90°).

Expectedly, well-tempered metadynamics simulation starting with the docked minor groove-bound state (MNS) along X and φ resulted either the drug going to the end of the DNA or in dissociation of the drug to separated state. Proflavine molecule sampled the minor groove region; moved along the DNA and got stabilized at the end of the 12 bp

DNA. To prevent it from going to the end of the DNA, we used a configurational restraint on the collective variable φ at 80^0. But then, after sampling the minor groove region the molecule got dissociated from the DNA. To achieve intercalation, we now placed another configurational restraint along the distance between the intercalating base pairs (IBP) of DNA and the center of mass of the drug at 10 Å, apart from the usual restraint on φ, to prevent dissociation of the drug. This reduced the configurational freedom of the molecule and thus prevented it from sampling the uninteresting part of the free energy surface. The restraints were placed far away from the DNA so that FES for intercalation was not affected.

2.3.2. Time Dependence of Collective Variables

Figure 2.4.a shows the time evolution of collective variables (X and φ) for dissociation of proflavine from MNS and the corresponding free energy surface is shown in Fig. 2.5.a. Up to 1.6 ns, drug stayed in the minor groove indicated by the values of the collective variables (φ ~ 70˚and X ~ 3.0 Å). Subsequently, the drug started sampling along the DNA in the minor groove indicated by low X (~ 0) and high φ (~ 90˚) up to 4.2 ns.

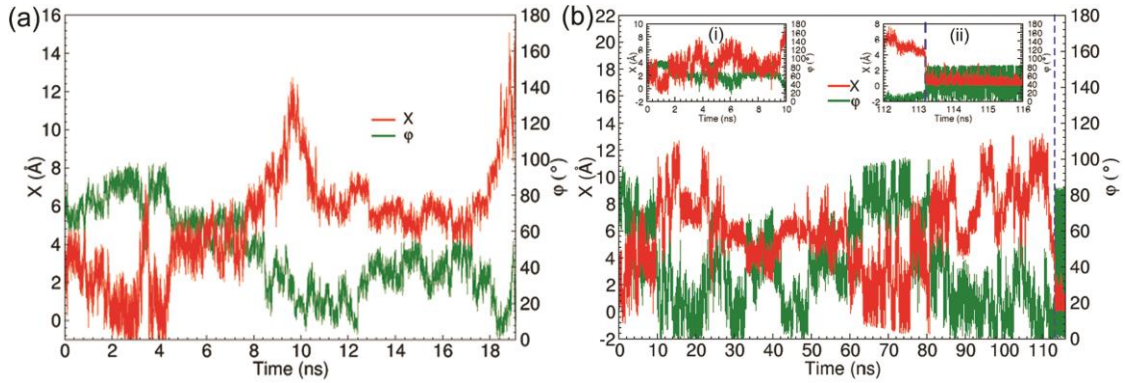


Figure 2.4. *Panel (a) and (b) shows the time dependence of the collective variables X and φ for dissociation and intercalation from the minor groove bound state (MNS), respectively. The inset figures in panel (b) show the magnified portion of the trajectory during the initial stages of simulation and at the transition state (TS), represented by a vertical dashed line. The complete dissociation from MNS happened at 19.1 ns and the intercalation from MNS happened at 113.2 ns.*

Because of the presence of soft wall potential on φ at 80˚, the drug then came back to X~ 4.0 Å and φ~ 60˚ at 4.4 ns and started sampling that region of the minor

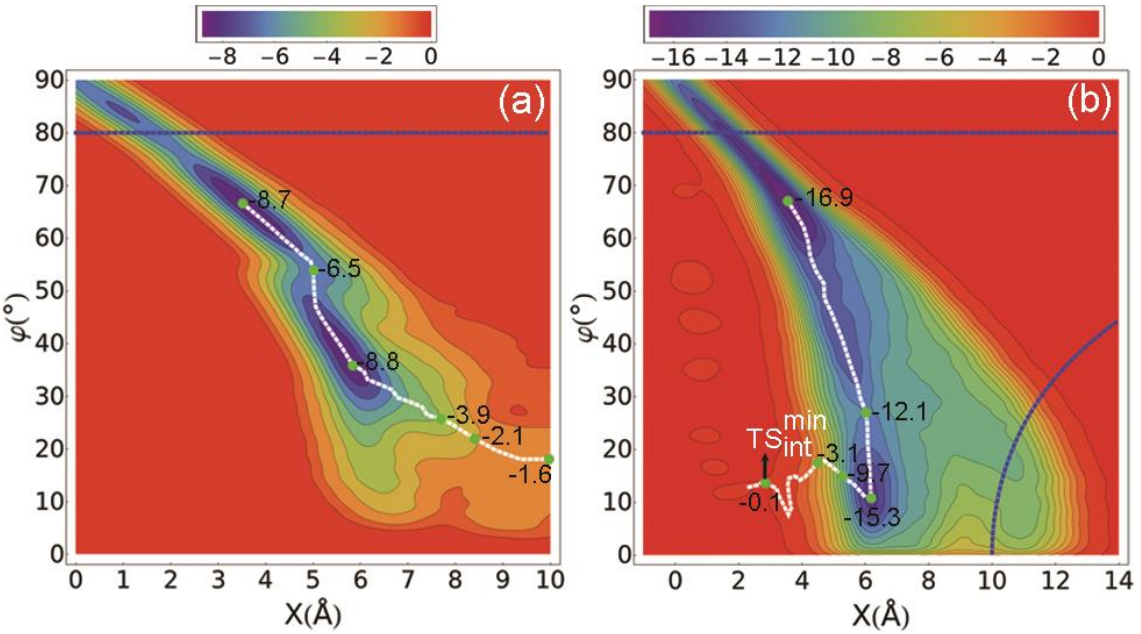
groove up to 7.8 ns. Later, X started increasing and φ started decreasing up to 10.2 ns. The low value of φ (~ 0˚) and the high value of X (~ 12.0 Å) indicate that the drug moved away from IBP perpendicular to the helical axis of DNA. At 12.4 ns, the drug again came back to the minor groove at X~ 6.0 Å and sampled the region up to 18 ns. From there, it started going out of the minor groove as indicated by the large value of X.

The time evolution of collective variables (X and φ) for intercalation of proflavine from MNS is shown in Fig. 2.4.b. The corresponding free energy surface is shown in Fig. 2.5.b. The initial and final parts of the trajectory are zoomed-in the inset (i) and inset (ii), respectively. Initially, φ is ~ 80˚ and X is ~ 3.0 Å indicating that the drug was in the minor groove but not close to IBP (φ is small close to IBP). Then φ decreased to zero after 20 ns simulation; however, X increased to a large value (~ 12 Å) indicating that the drug was along the IBP, but far away from the DNA. Then drug was pulled back to the minor groove (influenced by the soft potential placed at 10.0 Å on the distance between IBP and drug, $|\vec{d}|$ as described in Methods). The low value of X and high value of φ between 65 to 75 ns indicate that drug moved along the DNA and tried to sample different regions of the minor groove. From 110 ns the drug started coming close to IBP as shown by the decrease in both X and φ. The TS of intercalation is at X= 2.8 Å and φ= 13.6˚ indicated by the dotted line at 113.2 ns. After that, X remained low because drug was intercalated.

2.3.3. Free Energy Surfaces (FES)

Figure 2.5.a shows the FES for dissociation from the minor groove bound state along the collective variables, X and φ. Configurational restraint was used (at φ ≥80˚ for MNS) to prevent the drug from going to the ends of DNA. Free energy of the MNS (ΔG_{MNS}) was -8.8 kcal/mol, in reasonable agreement with experimental estimate -6.8 kcal/mol[6]. FES for intercalation process through minor groove is shown in Figure 2.5.b. Configurational restraints (shown as blue dotted lines) were used (at φ ≥80˚ and $\vec{d}$ ≥10.0 Å) to prevent the complete dissociation of molecule from DNA. Intercalation barrier from minor groove was thus found to be ($\Delta G_{MNS}^{\#,int}$) 16.9 kcal/mol which is in good agreement with experiments (12.5 to 15.1 kcal/mol)[6,44-45]. Simulation for intercalation process was stopped once the drug crossed the transition state (TS) after allowing

sufficient time to re-cross. Continuing the simulation would have populated the intercalated state until the drug de-intercalates again. This deintercalation process has not



been studied here as seen from Fig. 2.5.b where the stable intercalated state is not present.

Figure 2.5. *Free energy surface of (a) dissociation and (b) intercalation along X and φ. Free energy values are indicated by the labeled (kcal/mol) color bar on the top of each figure. White lines indicate the calculated minimum free energy path (MFEP). The green dots denote some representative points along MFEP, the corresponding structures of which (in Fig. 2.5.b) are shown in Fig. 2.9. The transition state for intercalation process is labeled. Blue lines indicate the configurational restraint.*

2.3.4. Error Analysis

Since the empirical error estimation for metadynamics simulation is not applicable in the well-tempered metadynamics simulation adopted here (see Methods in Chapter 1), we have performed additional well-tempered metadynamics simulation starting with different initial velocity distributions (which will follow a different dynamical trajectory), followed by equilibration using different metadynamics parameters to recalculate the free energy surface of intercalation shown in Fig. 2.6. Here, we used bias factor 20 and added Gaussians of 0.048 kcal/mol height at 1 ps interval at 300 K. Free energy barrier of intercalation obtained from this calculation is 17.2 kcal/mol (taking minor groove-bound state at -18.0 and transition state at -0.79 kcal/mol), in close agreement (16.9 kcal/mol) with the one as shown in Fig. 2.5.b.

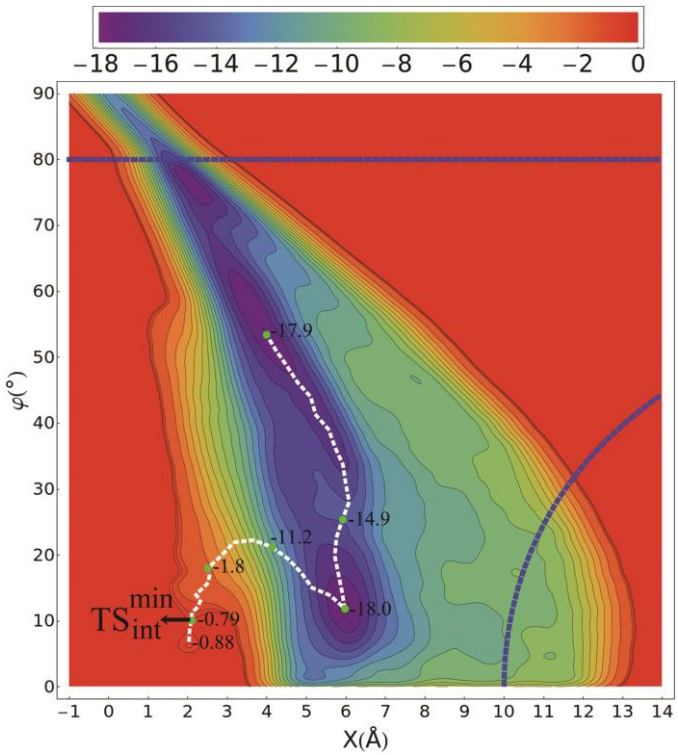


Figure 2.6. *Free energy surface of intercalation along X and φ. Free energy values are indicated by the labeled (kcal/mol) color bar on the top the figure. The white line indicates the calculated minimum free energy path (MFEP). The green dots denote some representative points along each MFEP. The transition state for intercalation process is labeled. Blue lines indicate the configurational restraints.*

2.3.5. Minimum Free Energy Path and Changes in DNA and Energy Parameters

In a multi-dimensional FES of a process, minimum free energy path (MFEP) provides most favourable mechanistic picture statistically. Therefore, we calculated the minimum free energy path for intercalation from the minor groove bound state using the algorithm by Ensing *et al.*[43], previously applied to understand dissociation mechanism of minor groove-bound drugs[46]. Structures were collected around discrete points along MFEP (white lines in Fig. 2.5) and subsequently cluster analyses[47] were performed. The average structural and energy components were calculated from the members of the biggest cluster (Fig. 2.7) and plotted against MFEP index (steps along MFEP). The MFEP value at 0 was taken to be the TS and the index increased towards the MNS. Energy parameters were calculated for a subsystem consisting of intercalating base pairs (IBP), drug and the largest number of water in the first hydration shell (taken to be within 3.4 Å from the heavy atoms) at any point along the MFEP (to conserve mass for the

subsystem). Here we assume that the part of the system uninvolved in the intercalation will not contribute to relative energy change.

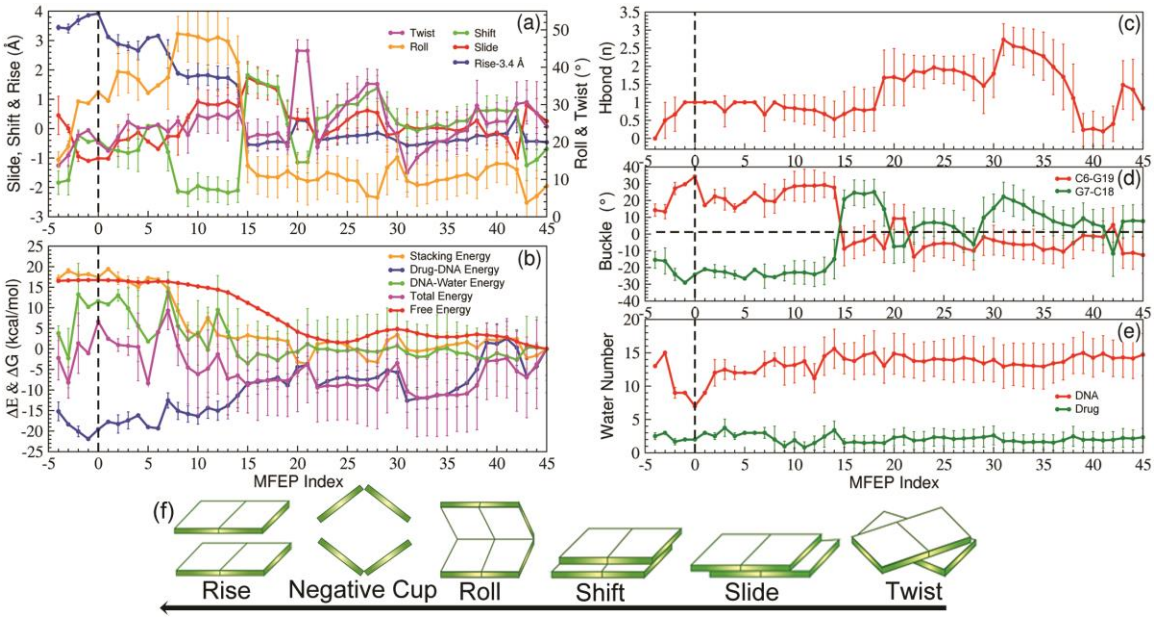


Figure 2.7. *The analysis of intercalation process from the minor groove-bound state. (a) DNA base pair step parameters for C6G7 base pair calculated using Curves+[48]. The vertical black dashed line represents the transition state. (b) Relative energy parameters (sum of vdw and coulombic) taken to be zero at the MNS (MFEP index 45). Stacking energy is taken to be the energy between C6 and G7 base pairs. The total energy is the energy of the subsystem comprising IBP, the drug and the largest number of water at any point along the MFEP within 3.4 Å from the heavy atoms of IBP-drug complex. Each point represents the average value of the members of biggest cluster around points along MFEP. (c) Number of hydrogen bonding interactions between the amine groups of proflavine and the intercalating base pair. (d) Change in the Buckle of base pairs C6-G19 and G7-C18 along MFEP (e) Number of water molecules around the drug and IBP along MFEP within 3.4 Å from the heavy atoms. (f) Different base pair parameters of DNA and the direction of the arrow show the sequence of change in the parameters. Standard deviations are shown.*

Moving along the MFEP of intercalation from the minor groove-bound state, Figure 2.7.a shows the various DNA base pair step parameters and Figure 2.7.b shows various average energy parameters for the intercalating base pair (IBP) C6G7. The variation in step parameters reflects a sequential process that follows a minimum base-stacking penalty pathway correlated with the position of the drug. The Twist is the first DNA parameter to change attaining lowest value at MFEP index 31. Here the drug forms maximum hydrogen bond with the DNA base pairs (Fig. 2.7.c) stabilizing Coulombic

energy (Fig. 2.8.b). From MFEP index 45 to 21, the stacking interaction energy (Fig. 2.8.a) between the IBPs is not changing much because the drug stays in the minor groove. But from MFEP index 21 onwards, van der Waals interaction energy increases because as the drug slides in, the IBPs move far apart from each other. But from 14 to 11 Coloumbic interaction energy decreases due to positive Roll as described in Fig. 2.8.a. Due to the opening of the base pairs at a high positive Roll, the base ring atoms of upper base pair 7G:18C forms electrostatic interactions with sugar-phosphate atoms of lower base pair 6C:19G resulting in the decrease of Coloumbic energy from MFEP index 14 to 11. Fig. 2.8.b shows the interaction energy between the drug and IBPs of DNA. Initially, there is Coloumbic energy stabilization because of hydrogen bonding between the amine groups of drug and the IBPs of DNA (Fig. 2.7.c). From 41 to 38 Coloumbic energy increases as the drug loses its hydrogen bonding with IBPs. Again, the Coloumbic energy decreases which is reflected by the regaining of the hydrogen bonding interaction of the drug with the IBPs. But as the drug start intercalating, the Coloumbic energy increases and the van der Waals energy decreases and the drug starts getting stabilized because of stacking interaction with the IBPs. The IBP-drug interaction energy compensates the decrease in vdw stacking interaction between the IBPs (Fig. 2.8.a).

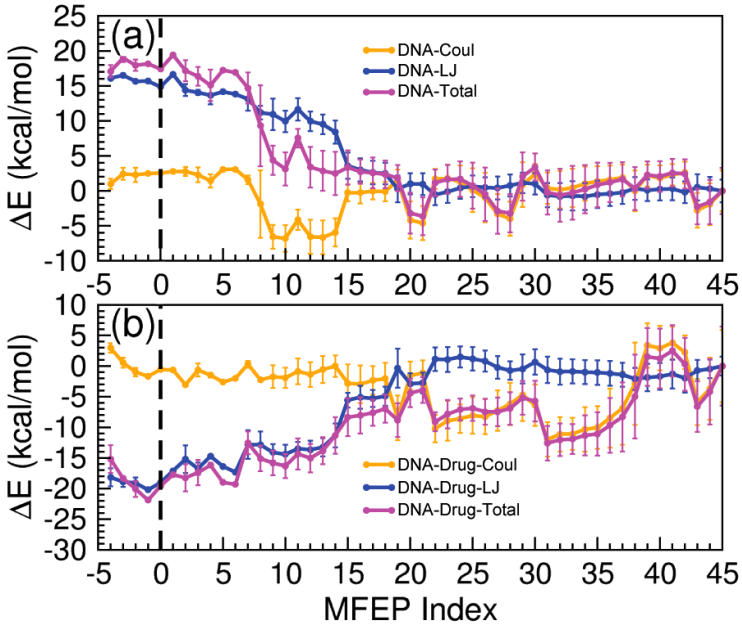


Figure 2.8. *The analysis of different energy components of intercalation process along MFEP. (a) stacking energy between the intercalating base pairs (IBPs) of DNA showing the changes in the Coloumbic and van der Waals interaction energies. (b) interaction energy between IBPs of DNA and the drug.*

Change in Twist is followed by an increase in Slide and Shift reaching a maximum at MFEP 15. This happens through hydrogen bonding of two amine groups with the sugar oxygen (see structure 18 of Fig. 2.9). The Twist in general and Shift and Slide to a certain extent do not cause base-stacking penalty[49]. This is also seen from the unchanged stacking energy (combined van der Waals (vdw) and coulombic energy of the IBP) up to 18 in Figure 2.7.b. Largest shift is followed by a sudden increase of Roll at 14 (thus releasing the linear force into an angular motion), where the drug rotates and inserts like a flap between the base pairs (see video S1). Here base stacking energy increases, but is compensated by the interaction with the drug (Fig. 2.7.b). In the final phase of intercalation, Roll and Slide decreases followed by a change in buckle from positive to negative cup (see Fig. 2.7.a and Fig. 2.7.d). The presence of proteins and drugs in the intercalation cavity is known to cause buckling in the DNA[50]. Buckle in negative cup reflects as a sudden increase in Rise. However, the Buckle does not cost base-stacking[51] and, therefore, serves here as a remedial measure to avoid energetically costly Rise. At the final stage, Rise increases resulting in the increase of stacking energy. At this stage, drug-DNA interaction energy decreases stabilizing the total interaction energy. A major energy contribution to the free energy barrier comes from the desolvation energy (Fig. 2.7.b) due to presence of less number of water molecules in the first hydration shell of the IBP (Fig. 2.7.e). The rest of the contribution to the free energy barrier may be attributed to entropy.

Figure 2.9 shows some representative structures of the intercalation process along the minimum free energy path (MFEP). The configuration at 45 represents the initial minor groove bound state. From there the molecule samples the different regions of the minor groove represented by some of those states at MFEP 30 and 24. Thereafter, the molecule started making some structural changes to the DNA as shown by configuration at 18. The initiation of cavity formation is shown by the configuration at 14. MFEP at zero represents the transition state for the intercalation process where it shows a clear well-defined cavity.

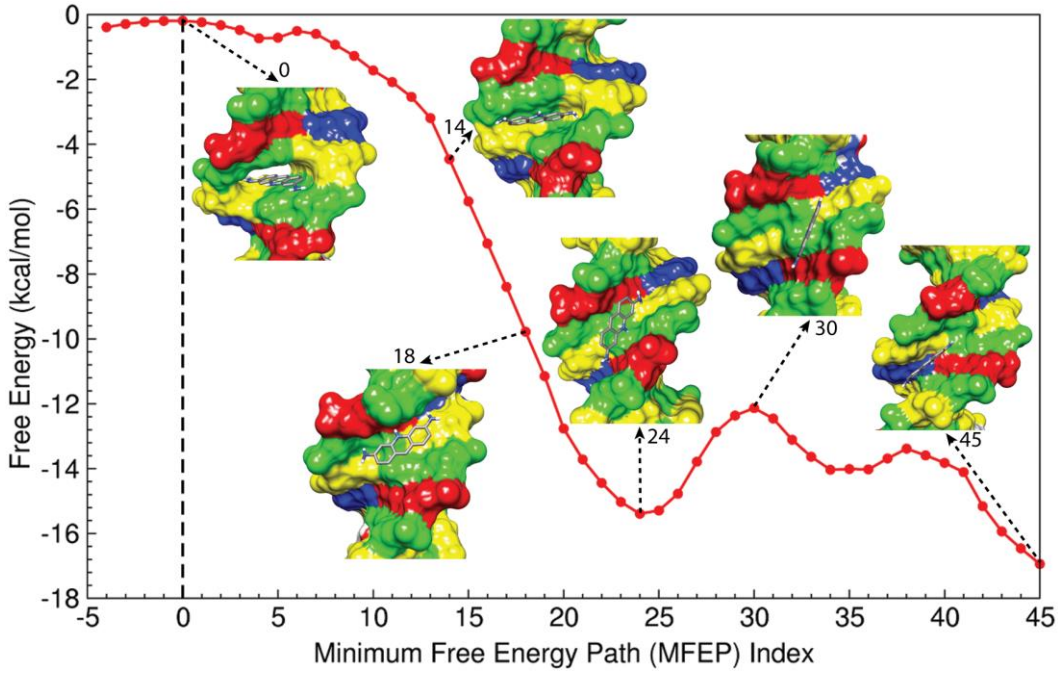


Figure 2.9. *Representative structures along the minimum free energy pathway (45➔0) of intercalation process. The red dotted line represents the minimum free energy pathway. The vertical dotted line represents the transition state for the intercalation process. Structures are drawn in chimera[52]. MFEP indices for individual structures are also mentioned*

2.4. Conclusion

In summary, this study uses configurational restraint to capture a direct intercalation event successfully in molecular detail for the first time. Analysis of DNA parameters and various energy components along the minimum free energy pathway indicates that intercalation of proflavine adopts a minimal base-stacking penalty pathway (Twist➔Shift/Slide➔Buckle/Roll➔Rise) induced by the drug proflavine. Therefore, these results suggest that the mechanism of intercalation for proflavine defies the natural fluctuation hypothesis[53] and supports a drug induced cavity formation mechanism[5]. It is interesting to note that the importance of Roll angle opening for deintercalation pathway was shown before using a different drug daunomycin[18], although the reason for Roll angle opening was not clear. Moreover, correspondence between DNA structural changes and various energy parameters reinforces the minimum base-stacking penalty pathway mediated intercalation process. Therefore, a similar mechanism (the exact order of parameter variation may differ depending on the structural features of the drug) may be applicable for a general intercalation process. We find here that water plays a major role

in controlling the energy component of the intercalation barrier while the entropy component is most likely dependent on the drug. This requires further investigation using configurational entropy as employed in drug-DNA binding[54] and intercalation process[55]. Moreover, a complete mechanistic pathway for intercalation and deintercalation through major and minor groove pathway will provide a deeper understanding of the intercalation process.

2.5. References

1. Lerman, L. S., *J. Mol. Biol.* **1961,** *3*, 18-30.
2. Berman, H. M.; Young, P. R., *Ann. Rev. Biophys. Bioeng.* **1981,** *10*, 87-114.
3. Hurley, L. H., *Nat. Rev. Cancer* **2002,** *2* (3), 188-200.
4. Brana, M. F.; Cacho, M.; Gradillas, A.; De Pascual-Teresa, B.; Ramos, A., *Curr. Pharm. Design* **2001,** *7* (17), 1745-1780.
5. Li, H. J.; Crothers, D. M., *J. Mol. Biol.* **1969,** *39* (3), 461-477.
6. Ramstein, J.; Ehrenberg, M.; Rigler, R., *Biochemistry* **1980,** *19* (17), 3938-3948.
7. Bereznyak, E.; Gladkovskaya, N.; Khrebtova, A.; Dukhopelnikov, E.; Zinchenko, A., *Biophysics* **2009,** *54* (5), 574-580.
8. Tang, P.; Juang, C. L.; Harbison, G. S., *Science* **1990,** *249* (4964), 70-2.
9. Neidle, S.; Jones, T. A., *Nature* **1975,** *253* (5489), 284-285.
10. Nuss, M. E.; Marsh, F. J.; Kollman, P. A., *J. Am. Chem. Soc.* **1979,** *101* (4), 825-833.
11. Neidle, S.; Pearl, L. H.; Herzyk, P.; Berman, H. M., *Nucleic Acids. Res.* **1988,** *16* (18), 8999-9016.
12. Swaminathan, S.; Beveridge, D. L.; Berman, H. M., *J. Phys. Chem.* **1990,** *94* (11), 4660-4665.
13. Paik, D. H.; Perkins, T. T., *Ang. Chem. Int. Ed.* **2012,** *51* (8), 1811-1815.
14. Macgregor, R. B.; Clegg, R. M.; Jovin, T. M., *Biochemistry* **1985,** *24* (20), 5503 - 5510.
15. Chaires, J. B., *Biopolymers* **1985,** *24* (2), 403-19.
16. Rizzo, V.; Sacchi, N.; Menozzi, M., *Biochemistry* **1989,** *28* (1), 274-282.

17. Quigley, G. J.; Wang, A. H. J.; Ughetto, G.; Marel, G. V. D.; Boom, J. H. V.; Rich, A., *Proc. Natl. Acad. Sci. USA* **1980,** *77* (12), 7204-7208.
18. Mukherjee, A.; Lavery, R.; Bagchi, B.; Hynes, J. T., *J. Am. Chem. Soc.* **2008,** *130* (30), 9747-9755.
19. Wilhelm, M.; Mukherjee, A.; Bouvier, B.; Zakrzewska, K.; Hynes, J. T.; Lavery, R., *J. Am. Chem. Soc.* **2012,** *134* (20), 8588-8596.
20. Denny, W. A., *Current Medicinal Chemistry* **2002,** *9*, 1655-1665.
21. DeJong, E. S.; Chang, C.-e.; Gilson, M. K.; Marino, J. P., *Biochemistry* **2003,** *42* (26), 8035-8046.
22. Barducci, A.; Bussi, G.; Parrinello, M., *Phys. Rev. Lett.* **2008,** *100* (2), 020603.
23. Macke, T.; Case., D. A., *Modeling unusual nucleic acid structures*. N.B. Leontes J. SantaLu- cia, Jr. ed.; American Chemical Society: Washington, DC, 1998.
24. Wang, J.; Cieplak, P.; Kollman, P. A., *J. Comput. Chem.* **2000,** *21*,1049-1074.
25. Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M., *Biophys. J.* **2007,** *92* (11), 3817-3829.
26. Besler, B. H.; Merz, K. M.; Kollman, P. A., *J. Comput. Chem.* **1990,** *11* (4), 431-439.
27. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; 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.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.;

Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A., *Gaussian 03, Revision C.02*. 2003.

28. Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., *J Comput Chem* **2005,** *26* (16), 1668-88.
29. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollmann, P. A., *J. Am. Chem. Soc.* **1993,** *115* (21), 9620-9631.
30. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., *J. Com. Chem.* **2004,** *25*, 1157-1174.
31. Sorin, E. J.; Pande, V. S., *Biophys. J.* **2005,** *88* (4), 2472-93.
32. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., *J. Comput. Chem.* **1998,** *19* (14), 1639-1662.
33. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., *J. Comput. Chem.* **2009,** *30* (16), 2785- 2791.
34. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., *J. Chem. Phys.* **1983,** *79*, 926-935.
35. W.H Press; B.P, F.; S.A., T.; W.T, V., *Cambridge University Press* **1992,** *2nd Edition*.
36. Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; DiNola, A.; Haak, J. R., *J. Chem. Phys.* **1984,** *81* (8), 3684-3690.
37. Nose, S., *Mol. Phys.* **1984,** *52* (2), 255-268.
38. Hoover, W. G., *Phys. Rev. A* **1985,** *31* (3), 1695-1697.
39. Parrinello, M.; Rahman, A., *J. Appl. Phys.* **1981,** *52* (12), 7182-7190.
40. Tom Darden, D. Y., Lee Pedersen., *J. Chem. Phys.* **1993,** *98* (12), 10089-10092.
41. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., *J. Chem. Theor. Comput.* **2008,** *4* (3), 435-447.
42. Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R., *Comput. Phys. Commun.* **2009,** *180* (10), 1961-1972.
43. Ensing, B.; Laio, A.; Parrinello, M.; Klein, M. L., *J. Phys. Chem. B* **2005,** *109* (14), 6676-6687.

44. Ramstein, J.; Leng, M., *Biophys. Chem* **1975,** *3* (3), 234-240.

45. C Ciatto, M. D. A., G Natile, F Secco, M Venturini, *Biophys J.* **1999,** *77* (5), 2717-2724.

46. Vargiu, A. V.; Ruggerone, P.; Magistrato, A.; Carloni, P., *Nucleic Acids Res.* **2008,** *36* (18), 5910-21.

47. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E., *Ang. Chem. Int. Ed.* **1999,** *38* (1-2), 236-240.

48. Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K., *Nucleic Acids Res.* **2009,** *37* (17), 5917-5929.

49. Hunter, C. A.; Lu, X.-J., *J. Mol. Biol.* **1997,** *265* (5), 603-619.

50. Victor, J.-M.; Ben-Haïm, E.; Lesne, A., *Phys. Rev. E* **2002,** *66* (6), 060901.

51. Christopher A, H., *J. Mol. Biol.* **1993,** *230* (3), 1025-1054.

52. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., *J. Comput. Chem.* **2004,** *25*, 1605- 1612.

53. Macgregor, R. B.; Clegg, R. M.; Jovin, T. M., *Biochemistry* **1987,** *26* (13), 4008-4016.

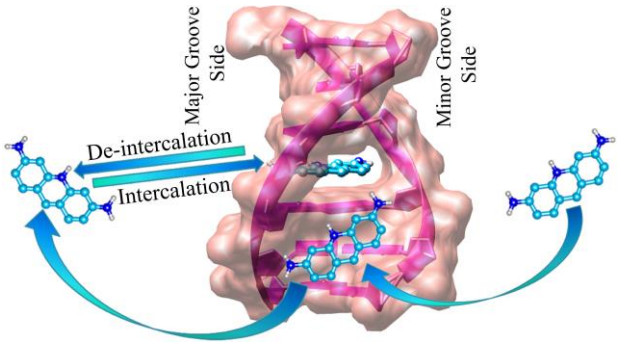
54. Harris, S. A.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A., *J. Am. Chem. Soc.* **2001,** *123* (50), 12658-12663.

55. Mukherjee, A., *J. Phys. Chem. Lett.* **2011,** *2* (24), 3021-3026.

Chapter 3

Intercalation and Deintercalation Pathway of Proflavine through the Minor and Major Grooves of DNA: Roles of Water and Entropy

Chapter 3: Intercalation and Deintercalation Pathway of Proflavine through the Minor and Major Grooves of DNA: Roles of Water and Entropy



Extensive kinetic and thermodynamic studies had followed since the discovery of intercalative binding mode. However, only recently the molecular mechanism of intercalation process received attention. As demonstrated in the previous chapter, the use of extensive computational methods achieved direct intercalation of proflavine from minor groove bound state for the first time. Further investigations revealed the molecular mechanism of intercalation that happens through drug induced cavity formation. However, the origin of thermodynamic and kinetic profile of the process is still not clear. Here we constructed the free energy landscape of intercalation, deintercalation and dissociation from both the major and minor grooves of DNA using extensive all-atom metadynamics simulations, capturing both the free energy barriers and stability in close agreement with fluorescence kinetic experiments. In the intercalated state, an alternate orientation of proflavine was found with almost equal stability compared to the crystal orientation; however, separated by 5.0 kcal/mol barrier that decreases as the drug approaches the groove edges. This study provides a comprehensive picture in comparison with experiments, which indicate that intercalation and deintercalation of proflavine happen through the major groove side, although the effective intercalation barrier increases because the path of intercalation goes through the stable (abortive) minor groove bound state, making the process a millisecond long one in excellent agreement with experiment. Molecular origin of the higher barrier for intercalation from the minor groove side was attributed to the desolvation energy of the DNA and loss of entropy, while the barrier from the major groove, in the absence of desolvation energy, is primarily entropic.

3.1. Introduction

Although intercalation[1], a unique binding interaction between ligand and DNA, is being explored as one of the drug design strategy to combat a major ailment- cancer and other microbial infections[2-3], the understanding of molecular mechanism and the pathway of intercalation process remained elusive. Only recently, efforts were put to understand the molecular mechanism of the intercalation process[4-5]. Previous chapter illustrated the molecular mechanism of intercalation process in general using a simple planar aromatic intercalator, proflavine. But the kinetic aspects of intercalation process[6-12] remained elusive. Various kinetic experiments suggested the complex nature of the intercalation process and predicted possible pathways of the process. First kinetic experiment on proflavine-DNA intercalation was done by Crother et al in 1969. The oscilloscope records of relaxation curves of proflavine-DNA interaction following a rapid temperature perturbation of the equilibrium are shown in Figure 3.1. Two timescales were obtained from the bi-exponential fit of the time-dependent curve (Fig. 3.1). So, they proposed that the process comprises two steps: the first step is a fast bimolecular outside binding (τ_1, faster than a few tenths of millisecond), followed by a slow intercalation (τ_2, range of a few milliseconds).[6]

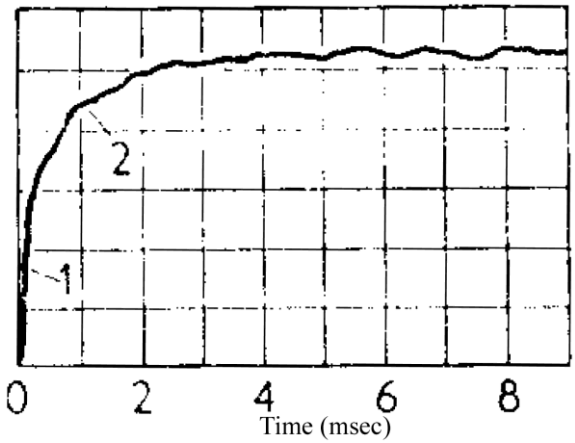
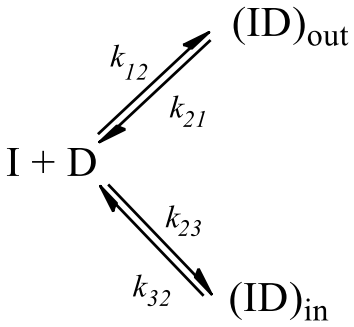


Figure 3.1. *The relaxation profile for intercalation of proflavine to Calf Thymus DNA. Reprinted from publication from Ref. 6, Copyright (1969), with permission from Elsevier.*

Several other kinetic studies also proposed the two step process for intercalation of proflavine to DNA.[7-8, 10, 13] However, there is a high probability that the drug binds in the wrong place of the DNA as shown in Fig. 1.6 of Chapter 1. In that case, it has to

diffuse along the DNA to search for a proper intercalation site leading to multiple steps for intercalation process[14]. Based on the two step reaction, Crothers[6] proposed a serial pathway for the intercalation of proflavine to DNA as shown in Scheme 1. In serial pathway, the drug forms outside-bound state (ID_{out}) and the drug causes some structural changes to the DNA to aid the intercalated state (ID_{in}) formation. However, an alternate pathway (parallel pathway) was also suggested as shown in Scheme 2. As per scheme 2, intercalation can happen between free DNA and proflavine directly.

$$\mathrm{I + D} \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} (\mathrm{ID})_{\mathrm{out}} \underset{k_{32}}{\overset{k_{23}}{\rightleftharpoons}} (\mathrm{ID})_{\mathrm{in}} \qquad \text{Scheme 1}$$



Scheme 2

Note: I corresponds to intercalator, D denotes DNA, and ID is the intercalator-DNA complex. Out and in corresponds to outside-bound state and intercalated state respectively.

Crothers argued for the series mechanism (Scheme 1) over parallel one (Scheme 2) because intercalation requires structural changes that could not happen directly as indicated in Scheme 2. The kinetic study by Ramstein and Leng showed that intercalation of proflavine happens in two steps irrespective of the nature of DNA as well as the base sequences[13]. Based on the kinetics experiments, the following assumptions were emerged: the outside bound state formation is a diffusion controlled process and the activation energy for the dissociation of the drug from the outside bound state is quite large. The slowness of intercalated state formation indicates that it is a highly barrier driven process and high activation energy (~15 kcal/mol)[6, 15] for deintercalation of the drug makes the deintercalation even slower.

In the previous chapter, we discussed the successful intercalation of proflavine from minor groove bound state and revealed the molecular mechanism of the

intercalation process.[5] But being a small and symmetric molecule, proflavine could intercalate through the major groove also. Moreover, from the static configuration of proflavine in the intercalated state obtained through X-ray crystallography, it is not possible to predict the pathway of intercalation, ie, whether the intercalation has happened though the minor groove side or the major groove side. So, a comparison in thermodynamics and kinetics of intercalation and deintercalation pathway through both grooves is necessary to get the complete picture. This was not possible by standard experimental methods as well. Therefore, the pathway of intercalation of proflavine remained elusive till date. So this study focuses on the capture of stable states and barriers within experimental bounds, however, providing a detailed molecular understanding of the intercalation process. Here, we have used well-tempered metadynamics simulation method[16-17] using multiple collective variables (CV) to calculate the free energy landscapes of six processes -- intercalation, deintercalation and dissociation through both the major and minor grooves of the DNA. Along the minimum free energy path, energetics have been calculated and compared with the free energy to gain entropic insight. We show here that proflavine intercalates and de-intercalates through the major groove side. However, intercalation precedes a minor groove-bound state, which makes the process millisecond long as observed in experimental kinetics.[26]

3.2. Experimental Section

3.2.1. Preparation of the Intercalated State (IS)

The intercalated crystal structure (PDB ID 3FT6[18]) contains two proflavine molecules intercalated between two top and bottom base pairs of a hexamer complex $d(CGATCG)_2$ in which amine groups face the major groove side of the DNA. Keeping the geometry of the lower two base pairs flanking the proflavine, a twelve base pair DNA was created using nucleic acid builder (NAB)[19] program. The final DNA has a sequence $d(GCGCTCGAGCGC)_2$ where proflavine is intercalated between C6 and G7 base pairs. We used more stable sequence, $d(GCGC)_2$, at both ends of DNA to increase overall DNA stability and to prevent flanking of base pairs at the ends during molecular dynamic simulation process. We used AMBER99/parmbsc0 force-field[20-21] for DNA. The preparation of force-field parameters of proflavine is described in detail in our previous

study[5]. The co-ordinates of proflavine were obtained from X-ray structure of proflavine in the intercalated state. The topology and co-ordinates of both DNA and proflavine were converted to GROMACS format by using an amb2gmx.pl program[22].

3.2.2. Preparation of the Minor and Major Groove-bound States

We made a twelve base-pair B-DNA having same sequence as in the IS using NAB[19]. Minor and major groove bound states were made by docking proflavine on to DNA using AutoDock software[23]. Docking protocol is described in detail in Chapter 2. Initial docking gave only minor groove bound states of which we have taken the DNA-proflavine complex, as minor groove bound state, with lowest binding energy -6.25 kcal/mol which is having initial configuration X = 2.0 Å and φ = 76.9˚. Since we did not get any major groove bound state during docking, we took the most stable minor groove bound state complex as the receptor and docked another proflavine on to it. After docking we got the major groove bound state with binding energy -3.68 kcal/mol and with initial configuration as X= -1.0 Å and φ= 98.8˚.

3.2.3. Simulation Details

Proflavine-DNA system was initially solvated with water (TIP3P water model[24]) in a large cubic box (7.123nm x 7.123nm x 7.123nm). 22 Na^+ ions and 1 Cl^- ion were added into the system to neutralize the negative charge on DNA and positive charge on proflavine. Final system was then energy minimized followed by a series of position restraint and equilibration steps.[4] Finally an unrestrained equilibration was done for 1 ns at constant pressure (1 bar, using Parrinello-Rahman barostat[24]) and temperature (300 K, using Nose-Hoover thermostat[24]) with a coupling constant 0.2 ps. Electrostatic interaction was treated using PME (Particle Mesh Ewald)[24]. Detailed simulation protocol is described in Chapter 2. We used well-tempered metadynamics[17] for all calculations. For intercalation, deintercalation and dissociation of proflavine from minor groove bound state; bias factor used was 15 and Gaussian height added was 0.048 kcal/mol at 2 ps interval, except for dissociation of proflavine from major groove bound state where we used a bias factor 8. Based on the natural fluctuation of collective variables in different states, we used different widths (σ) of the Gaussians for CV in various processes. We

used σ = 0.5 Å for X and σ = 0.2 radians for φ for intercalation and dissociation processes. For deintercalation processes, we used σ = 0.5 Å for X and σ = 1.0 radians for θ. GROMACS[25], a molecular dynamics simulation package coupled with a separate program PLUMED[26], was used for all simulations with the modification on PLUMED package to incorporate new CVs. Calculation of MFEP (Minimum Free Energy Path) and Cluster analysis are described Chapter 2.

3.3. Results and Discussions

3.3.1. Design of Reaction Co-ordinates

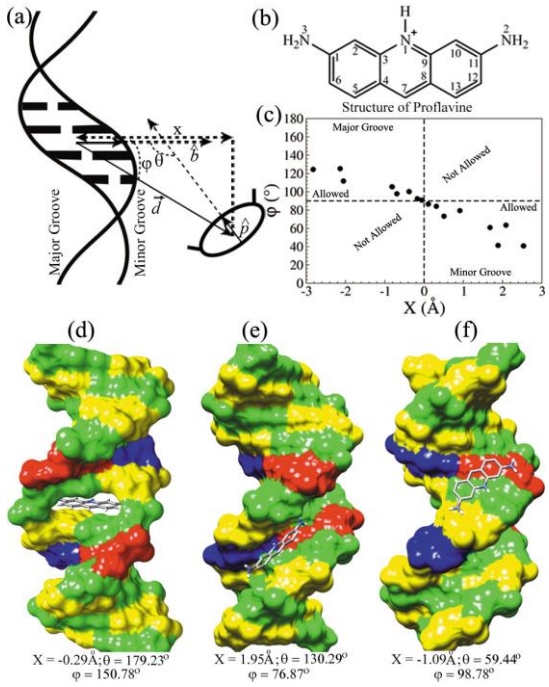


Figure 3.2: *Schematic representation of the CV and initial drug-DNA configurations. (a) Pictorial representation of X, θ and φ constructed with three vectors,* $\hat{b}$, $\vec{d}$ *and* $\hat{p}$. $\hat{b}$ *is the unit body-fixed vector that runs from the center of mass (COM) of the intercalating base pairs, IBP, (C6, G7, C18, G19) to the COM of 3' side ribose sugar groups (G7 and G19). Therefore,* $\hat{b}$ *points to the minor groove of the DNA.* $\overrightarrow{d}$ *denotes the vector from the COM of IBP to the COM of proflavine. So |d| provides a measure of distance between the drug and the DNA.* $\hat{p}$ *provides orientation of amine groups of the drug. It is the vector pointing to the central nitrogen atom (N1) of proflavine from its opposite carbon atom (C7) (see Fig. 3.2b).* $X = \hat{b}.\overrightarrow{d}, \theta = Cos^{-1}(\hat{b}.\hat{p})$, *and* $\varphi = Cos^{-1}(\hat{b}.\hat{d})$, $\hat{d} = \vec{d}/|\vec{d}|$.*(b) Chemical structure of proflavine. (c) {X,φ} values for the base ring atoms of four intercalating bases (C6, G7, C18, G19) which show that first and third quadrant are non-allowed regions due to geometrical constraints. (d), (e) and (f) show the initial configuration of the intercalated, minor groove-bound and major groove-bound states, respectively (see Methods for details) with the values of CV. Surface representations of DNA are drawn using UCSF chimera[27].*

The choice of proper collective variables is crucial to study a complex process like intercalation. Here, we used three different CV (X,θ,φ) shown schematically in Fig. 3.2a. X denotes the distance of proflavine from the DNA perpendicular to the helical axis and described in detail in Chapter 2. θ measures the orientation of the drug's amine groups (Fig. 3.2b) with respect to the DNA. Therefore, if the amine groups of proflavine in the intercalated state (IS) point to the major groove side as in the crystal structure, θ is ~ 180° and if they point towards the minor groove side, θ is ~ 0°. Third collective variable φ measures the movement of the drug along the helical axis of DNA and described in detail in Chapter 2. Due to constraints imposed by the design of the collective variables, only second and fourth quadrants of {X, φ} diagram are accessible because any point on the minor groove side cannot have φ value more than 90°, and any point on the major groove cannot have φ value less than 90°. This is shown in Fig 3.2c where we plotted {X, φ} values of the base ring heavy atoms involved in intercalation. Figure 3.2d, e and f show the initial configuration of the intercalated, minor groove-bound and major groove-bound states. The values of the CV indicate that these CV's can distinguish among the above states and therefore the states in between.

3.3.2. Deintercalation towards Major and Minor Grooves

To understand the pathway and kinetic-thermodyamic parameters of intercalation process, we performed deintercalation of proflavine using well-tempered metadynamics simulation along the collective variables, X and θ starting with the equilibrated intercalated configuration in crystal geometry (Fig. 3.2d). In the IS, the third CV, φ does not provide useful information as it varies widely between 0˚ (while X is +ve) and 180˚ (while X is –ve) and φ is important to understand the displacement of the drug along DNA axis which is not feasible in the intercalated state. So φ was not taken into consideration during the deintercalation process. We started the simulation with the crystal structure configuration[18] of proflavine where, the amine groups of the proflavine molecule face towards the major groove side of DNA.

The time evolution of collective variables, X and θ for deintercalation through major and minor grooves are shown in Fig. 3.3A and Fig.3.3B respectively. The trajectory (Fig. 3.3A) shows that in both cases, proflavine samples different orientations

(ranges from 0° to 180°) of amine groups with respect to the grooves of DNA(see the insets of both graphs). But initially, the sampling of X was confined to the stable free energy well of the intercalated region (X~±2 Å). However, after sampling both orientations (beyond 6 ns), fluctuation in X increased to a higher value (±4 Å). After 55 ns, the fluctuation was mainly in the major groove side and X changed towards a more negative value. Finally, the drug exited DNA through the major groove side at 56.4 ns as shown by the larger negative X. At the transition state (TS), θ is close to 40˚ (Fig. 3.3A inset (b)) indicating that the drug de-intercalates with the amine groups facing DNA (i.e., amine groups de-intercalates later).

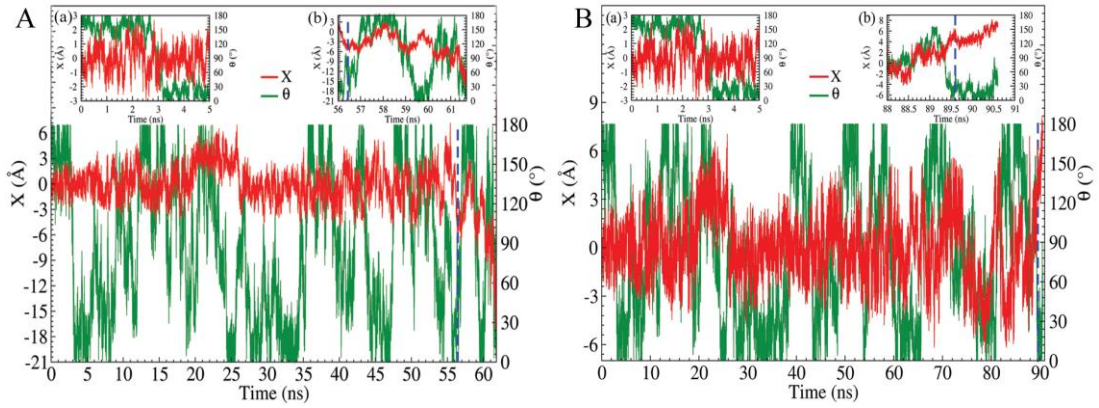


Figure 3.3. *Panel A and B shows the time dependence of the collective variables X and θ for deintercalation through major groove side and minor groove side, respectively. The inset figures show magnified portions of the trajectory during the initial stage of simulation and at the transition state (TS), represented by a vertical dashed line.*

In case of deintercalation through the minor groove side, the fluctuation is same as that of deintercalation through the major groove side. After filling up of the free energy wells corresponding to two different orientations, the drug came out through the minor groove side at X~ 5.7 Å and θ~ 33.7˚ after 89.5 ns of simulation time. The initial value of θ is high indicating that the orientations of the amine groups face the major groove side (Fig. 3.3B inset (a)). At the TS (Fig. 3.3B, inset (b)), low θ value (~ 33.7˚) indicates that the drug de-intercalated with amine groups of proflavine facing the water.

Initially, proflavine started sampling the intercalated state with the amine groups facing towards the major groove side, represented by θ value ~ 173˚ as shown in Fig. 3.4a. After sampling the well for 5.0 kcal/mol, the drug changed its orientation such that

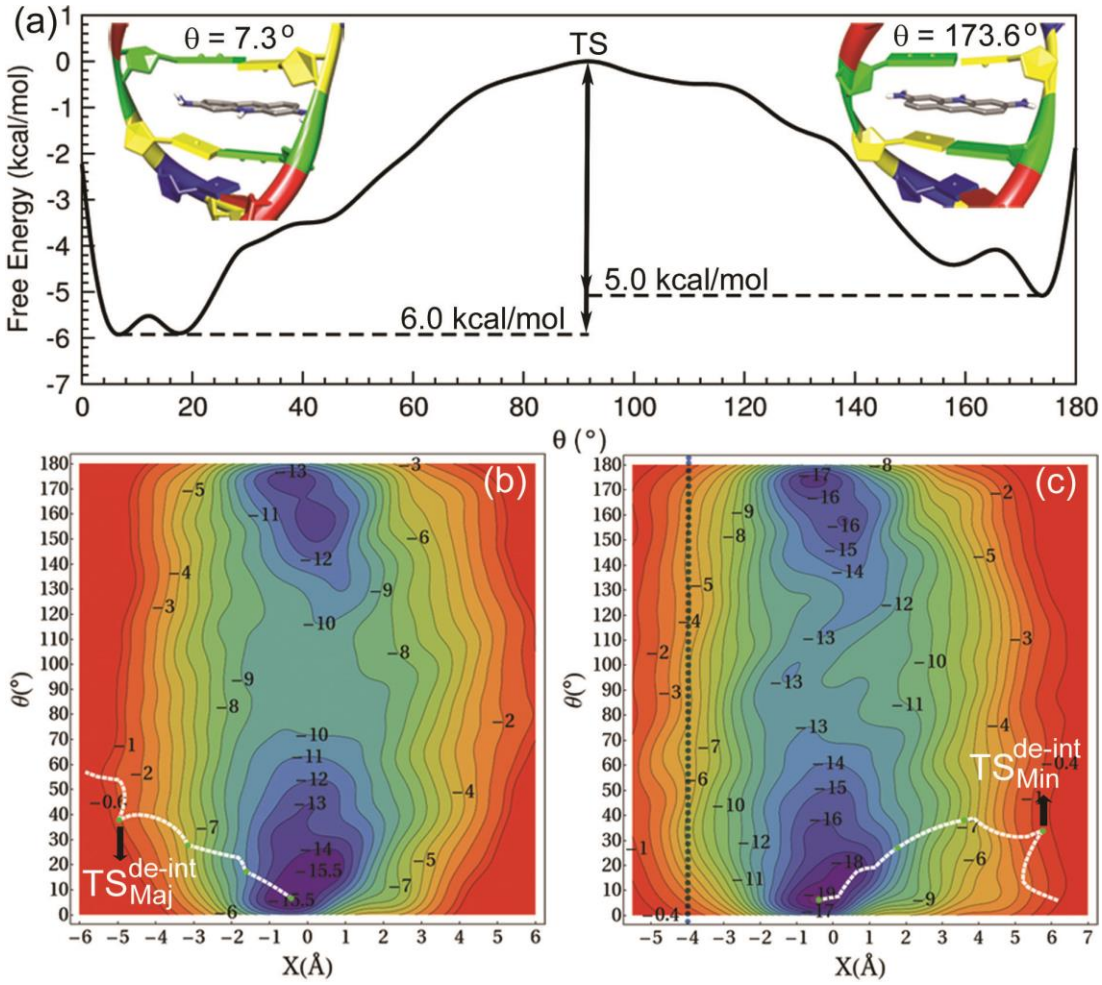


Figure 3.4. *Free energy surfaces (FES) of deintercalation against X and θ. (a) Free energy profile of proflavine rotation in the intercalated state (|X| < 1 Å) plotted against θ. The transition state (TS) for rotation is chosen as reference value zero. It shows two stable minima with a barrier of 5.0 kcal/mol from the crystal (right) orientation with θ = 173.6˚ to the other orientation (left) with θ = 7.3˚. (b) FES of deintercalation towards major groove side. (c) FES of deintercalation towards minor groove side with configurational restraint (dashed blue line) along major groove direction. Contours are labeled with corresponding free energy values in kcal/mol. White dotted lines represent the calculated minimum free energy paths.*

the amine groups face towards the minor groove side corresponding to θ value ~ 7˚ (See Fig. 3.4a). In this orientation, drug samples up to 6.0 kcal/mol. After sampling the wells, the sampling of θ became uniform. Thus we obtained a free energy profile of rotation of the proflavine in the intercalated state as shown in Fig. 3.4a with the representative configurations for each orientation corresponding to θ = 173.6˚ and 7.3˚ (minimum at lower θ value is rather broad). This also indicates the possibility of alternate orientation

of proflavine in the intercalated state, which is slightly more stable than crystal orientation at the simulation condition (infinite dilution), however, separated by a barrier of 5.0 kcal/mol between crystal and opposite orientation. Intercalating part of the DNA for both the stable proflavine orientation in the intercalated state resembles the equilibrated configuration of the crystal structure as shown in Fig. A1 of Annexure 3. However, the absence of the other orientation of proflavine in the intercalated state in the crystal structure may be because of the different experimental conditions such as temperature 293 K, drug concentration, presence of cobalt hexamine for duplex stability, PEG buffer etc[16]. These external conditions may favor only one orientation.

It turned out that the orientational barrier of proflavine in the intercalated state is still small compared to the barrier for deintercalation process. The sampling of free energy surface provided the estimate of deintercalation free energy barrier towards major groove ($\Delta G_{Maj}^{\#,de-int}$) as 14.7 kcal/mol, in excellent agreement with experimental values (11.7-14.8 kcal/mol) as shown in Table 3.1. Figure 3.4b shows the free energy surface (FES) for the deintercalation along X and θ towards major groove side. Minimum free energy path (MFEP) calculation located the transition state (TS) at X = -5.0 Å and θ = 37.9˚. The small value of θ indicates that the deintercalation happens by amine groups facing minor groove side, i.e., amine groups exit the DNA later.

To capture deintercalation transition through the minor groove, a configurational restraint[26] was placed at X= -4.0 Å (1 Å before the TS) to prevent the escape of proflavine through the major groove side. This led the drug to de-intercalate through the minor groove (Fig. 3.3B) with a barrier, $\Delta G_{\mathrm{Min}}^{\#,\mathrm{de-int}}$, as 18.5 kcal/mol. Figure 3.4c shows the free energy surface, with TS at X = 5.7 Å and θ = 33.9˚. Here, small θ indicates that the drug exits by pointing amine groups towards the minor groove, i.e., amine groups exit the DNA first.

Although we find distinct orientations during deintercalation process along minimum free energy path, FES shows that the barrier for proflavine orientation (along θ) decreases as the drug de-intercalates towards either major or minor groove side until X ~ ± 2.0 Å, beyond which there is no significant barrier between these different orientations. Therefore, the deintercalation could happen with any orientation. This is counter-intuitive as the presence of amine groups and crystal orientation of proflavine suggest[28] that

intercalation and deintercalation should happen through major groove where amine groups should face major groove side to avoid a steric clash. Present calculation shows that size of the amine groups poses no such constraints. Table 1 provides a comparison with several experimental estimates. The deintercalation barrier matches closely with the values obtained for deintercalation through the major groove indicating that deintercalation most likely would happen through the major groove.

3.3.3. Dissociation from Minor and Major Groove-bound States (MNS and MJS)

The dissociation of the proflavine from minor groove-bound state is illustrated in Chapter 2. Here, we study the same process from major groove-bound state as well and compare the free energy surfaces of both the processes in more detail. The initial equilibrated configuration of minor groove bound state is shown in Fig. 3.2e, while for major groove bound state it is shown in Fig. 3.2f. Further, we performed well-tempered metadynamics simulations against X and φ. While X provides a measure of distance of the drug along the body fixed vector, φ denotes the displacement along DNA axis. Therefore, large |X| (both positive and negative value) must denote dissociation. This is not true if we use radial distance to measure dissociation because large distance between the drug and the intercalating base pair can be achieved if the drug simply moves along the DNA axis (note that, distance * cosφ = X). Configurational restraints were used here in both cases (at $\varphi = 80^{\circ}$ for minor and $\varphi = 100^{\circ}$ for major) to prevent the drug from going to the ends of DNA. We did not use θ here as CV because sampling of θ in the out-side bound state is facile. Therefore, average value of θ can be obtained from a particular part of the FES. Thus, we could restrict the FES to two dimensions only.

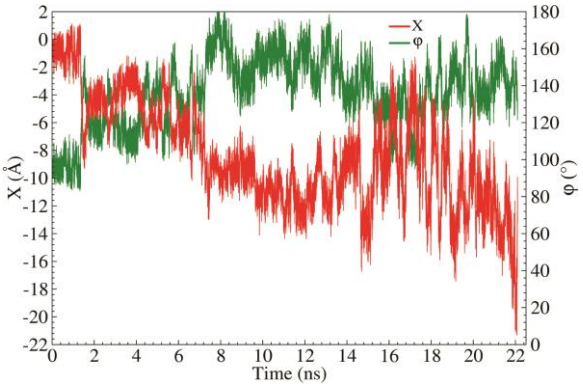


Figure 3.5. *Time variation of collective variables X and φ for the metadynamics trajectory of dissociation process from major groove-bound state.*

The time evolution of collective variables (X and φ) for dissociation of proflavine from MJS is shown in Fig. 3.5. Initially, the drug was close to backbone shown by a low φ (~ 90º) and low X (~ 1 Å) up to 1.2 ns. From backbone, drug moved towards the groove by increasing both X and φ. From 9 to 13 ns, drug stayed away from IBP shown by larger negative X (~ -12.5 Å) and larger φ (~ 160º). From 17 ns onwards, the drug started going away from IBP to outside indicated by larger negative X. We stopped the simulation at 22 ns when the drug was separated.

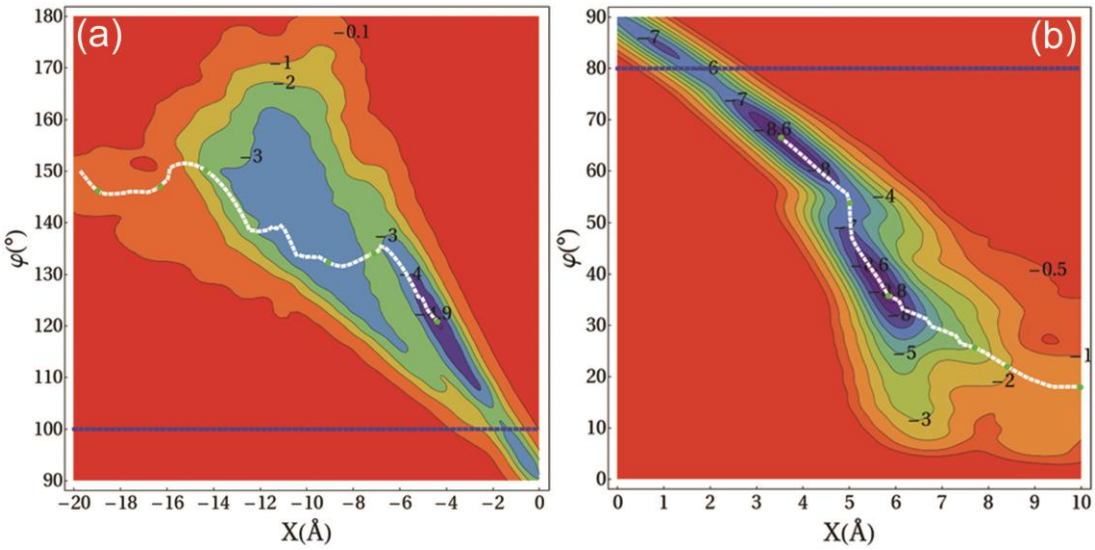


Fig. 3.6. *Free energy surfaces (FES) for system starting from groove-bound states along X and φ. (a) dissociation from major groove-bound state (b) dissociation from minor groove-bound state. Contours are labeled with corresponding free energy values in kcal/mol. White dotted lines represents the calculated minimum free energy paths (MFEP). The blue lines show the position of the configurational restraint. Note that, φ <90° in the minor groove and φ >90° in the major groove is the natural restriction (see Fig. 3.2c) due to construction of this angle from the body-fixed vector.*

Free energy stability (ΔG_{Maj}) of major groove bound state was found to be -4.7 kcal/mol. We have previously shown that dissociation of the drug starting from the minor groove provided -8.8 kcal/mol stability for the minor groove bound state (ΔG_{Min}).[5] FES for dissociation from the major and minor groove-bound states are shown in Figs. 3.6a and 3.6b, respectively. As expected from preliminary docking studies, major groove bound state was found to be significantly less stable than the minor groove bound state. The stability of outside-bound states (major and minor) could be compared with the experimental values of the "outside-bound" state (-4.1 to -6.8 kcal/mol in Table 1).

Interestingly, the experimental values lie mostly within the calculated values for major and minor groove bound state indicating the ensemble average picture of the experimental results.

Table 1: Comparison between experimental and theoretical estimates of free energy stabilities and barriers for intercalation, deintercalation processes and external bound states.

Energy	Experimental								Theoretical	
(kcal/mol)	a	b	c	d	e	f	g	h	Major	Minor
$\Delta G^{\#,int}$	12.5	12.6	11.2	12.2	13.8	12.2	15.1	-	8.8	16.9
$\Delta G^{\#,de-int}$	14.8	14.3	11.7	13.9	14.4	14.4	-	-	14.7	18.5
ΔG_{ext}	-6.8	-7.0	-5.9	-4.1	-4.1	-	-6.0	-6.5	-4.7	-8.8
ΔG_{int}	-9.0	-8.7	-6.4	-5.8	-4.7	-	-	-7.9	-10.6	-10.4

a,b (T=25˚C, 0.1M Na^+)[7]; c (T=10˚C, 0.016M Na^+)[6]; d (T=25˚C, 0.2M Na^+)[6]; e (T=25˚C, 0.1M Na^+)[8]; f (T=17˚C, 0.1M Na^+)[9]; g (T=25˚C, 0.1M Na^+)[13]; h (T=25˚C, 0.02M Na^+)[10]; Theoretical (T=27˚C, 0.1M Na^+). $\Delta G^{\#,int}$=Free energy barrier for intercalation, $\Delta G^{\#,de-int}$=Free energy barrier for deintercalation, ΔG_{ext}=Free energy of externally bound state, ΔG_{int} =Free energy of intercalated state.

3.3.4. Intercalation from Minor and Major grooves

Deintercalation is like protein unfolding process, where one state goes to any of many accessible states. Intercalation process is on the other hand like protein folding process where many states go to one (or few) state. Therefore, achieving intercalation from groove-bound states using computational method is non-trivial. Recently we could successfully intercalate proflavine from the minor groove-bound state using configurational restraint and showed that the molecular mechanism of intercalation proceeds through minimum base-stacking penalty pathway[5]. Here we used the similar technique to intercalate proflavine from the major groove-bound state and obtained a comparative free energy profile for intercalation from both sides.

The time evolution of collective variables (X and φ) for intercalation of proflavine from the major groove side is shown in Fig.3.7. The inset figures (Figs. 3.7a and 3.7b) show initial and final stages of the simulation. The vertical blue dotted line shows TS of

intercalation from major groove side at 43.5 ns. Initially, X is small (~ 1 Å) with φ ~ 100˚ indicating that the drug is close to the backbone of IBP. After 3.2 ns, X started increasing in the negative direction and φ also started increasing. At 20 ns, the drug was at X~ -13 Å and φ~ 180˚ showing that drug moved out of major groove perpendicular to the helical axis of DNA. After that, drug went towards IBP influenced by the soft wall potential on the distance between IBP and the center of mass of the drug ($|\vec{d}|$) (see methods for details) at 11.5 Å. Likewise, the drug sampled the major groove and finally intercalated through the TS at 43.5 ns (inset Fig. 3.7b) with X~ -3.1 Å and φ~ 164˚. Once intercalation had happened, X remained low (~ 0).

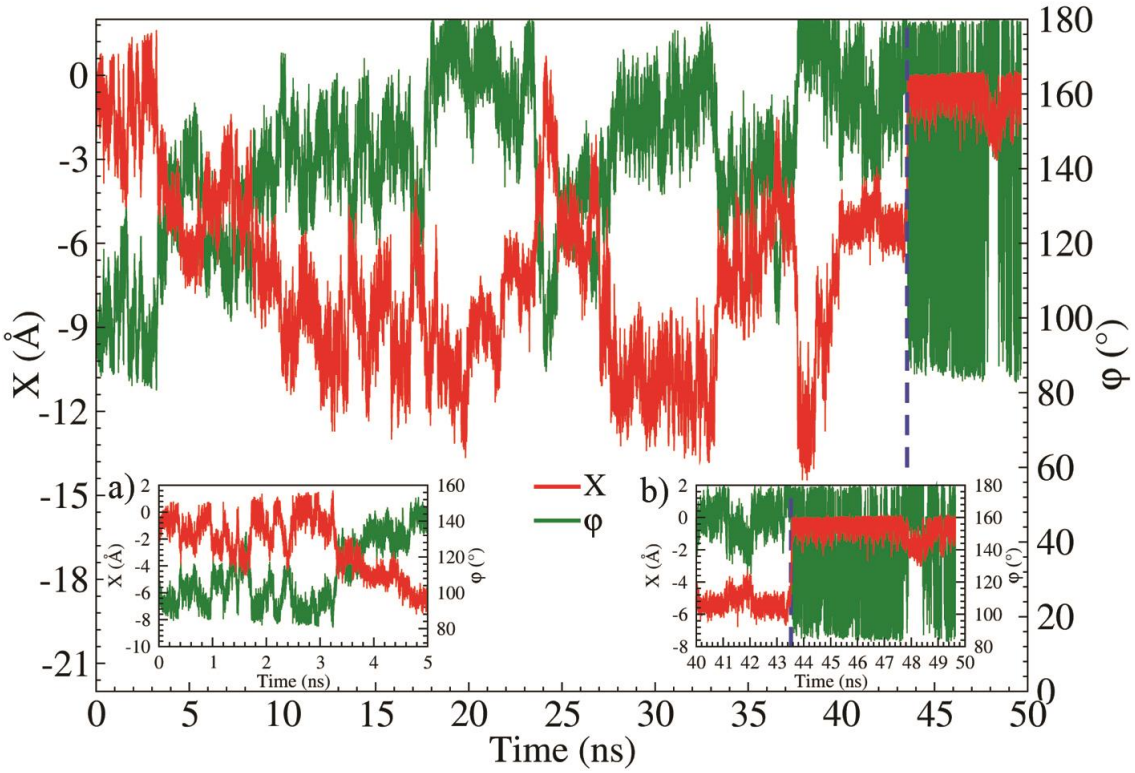


Figure 3.7. *Time variation of collective variables X and φ for the metadynamics trajectory of intercalation process from the major groove-bound state.*

FES's for intercalation process through major and minor grooves are shown in Figs. 3.8a and 3.8b. In order to achieve intercalation, we placed configurational restraint on the distance between the intercalating base pairs (IBP) and COM of the drug $(|\vec{d}|)$ at 11.5 Å for the major groove in the similar manner as was done in case of intercalation from minor groove-bound state. We maintained the usual restraint on φ as used for

dissociation processes. Minimum free energy path calculation captured intercalation barrier from major groove, $\Delta G_{Maj}^{\#,int}$, as 8.8 kcal/mol. This barrier is much smaller than that we got for intercalation from minor groove-bound state, $\Delta G_{Min}^{\#,int}$, earlier as 16.9 kcal/mol[5].Table 1 provides a comparison of the intercalation barrier with experimental values (11.2 to 15.1 kcal/mol), which lie between the estimate of barrier either directly through direct major or minor groove sides.

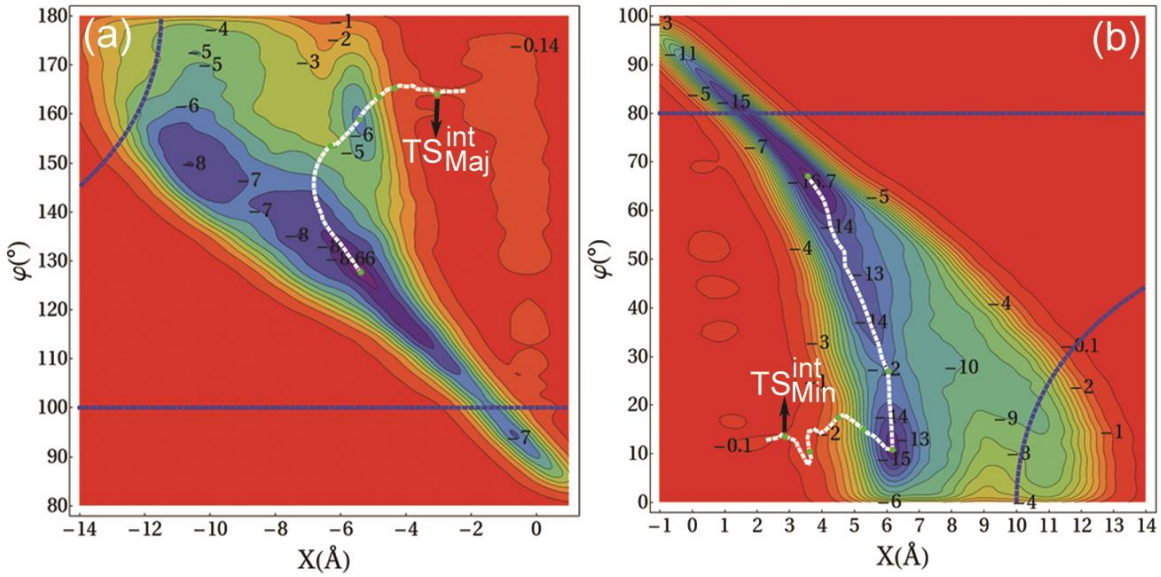


Fig. 3.8. *Free energy surfaces (FES) for system starting from groove-bound states along X and φ. (a) intercalation from major groove-bound state (b) intercalation from minor groove-bound state. Contours are labeled with corresponding free energy values in kcal/mol. White dotted lines represents the calculated minimum free energy paths (MFEP). The blue lines show the position of the configurational restraints.*

In both cases the configurational restraints were far from the TS's to affect it in each case. TS for intercalation along the minor groove was at X = 2.8 Å, φ = 13.6˚, whereas from major groove side it lies at X= -3.1 Å, φ = 164˚. It is important to mention that the simulation was stopped once the drug crossed the TS from either side after allowing sufficient time to re-cross. Continuing the simulation would eventually result in deintercalation, which has already been studied in detail using the collective variables X and θ.

In the previous chapter, we had shown that intercalation happens through the minimum base stacking penalty pathway. Similarly, intercalation through major groove

also follows minimum base stacking penalty pathway indicating the possibility of this as the general mechanism of the intercalation process. Figure 3.9 shows the changes in the DNA parameters and energy components along the MFEP of intercalation from major groove side of DNA. Similar to the intercalation from the minor groove, here also untwisting was the first DNA parameter to change at MFEP index -65 (Fig. 3.9a), without costing any stacking energy (Fig. 3.9g). Twist starts increasing from -63 as proflavine (FLV) forms more H-bonds (Fig. 3.9i) with IBP, also apparent from change in the θ to lower value at -61 (Fig. 3.9i), and pulls the C6:G19 base pair causing increase in its Shift (Fig. 3.9c).The free energy decreases slightly at -56 (Fig. 3.91) where both the amine groups of proflavine form H-bond with backbone sugar groups. From this point onward, negative Slide starts and reaches minimum at -48 (Fig. 3.9b), where both the two intercalating base pairs experience negative shift (Fig. 3.9c) (possibly to avoid unfavorable negative slide in CG base pairs or being pushed by the drug) resulting in the decrease in Roll (opening of base pair towards major groove side) at that next instant at -47 (Fig. 3.9d). Buckle changes from positive to negative cup at -52 (Fig. 3.9e) resulting in the increase in Rise (Fig. 3.9f). Stacking interaction decreases from -52 onwards with sudden jump around -47 (Fig. 3.9g). This is compensated again by increased drug-DNA interaction (Fig. 3.9h). *Note that the changes described above are along the minimum free energy pathway, therefore the observations above are statistically viable and not the outcome of a single observation.* Changes in different DNA parameters during the intercalation from major groove suggest a drug induced cavity formation mechanism from the major groove side also. However, in contrast with minor groove intercalation, the number of water molecules around major groove IBP atoms does not change appreciably (Fig. 3.9j), reflecting in little desolvation cost near TS at -41 (Fig. 3.9k), and subsequently results in no overall energy cost for intercalation (Fig. 3.9l) from major groove side. Intercalation mechanism through major groove is shown in Supplementary Information movie S2. Analysis of θ shows that intercalation from major groove happens by inserting the amine groups first into the DNA (See Fig. 3.9i). Different DNA and energy parameters of the deintercalation process through the major and minor groove are shown as Fig. A2 and A3 in Annexure-3. The same for dissociation from minor and major groove are shown as Fig. A4 and A5 in Annexure-3.

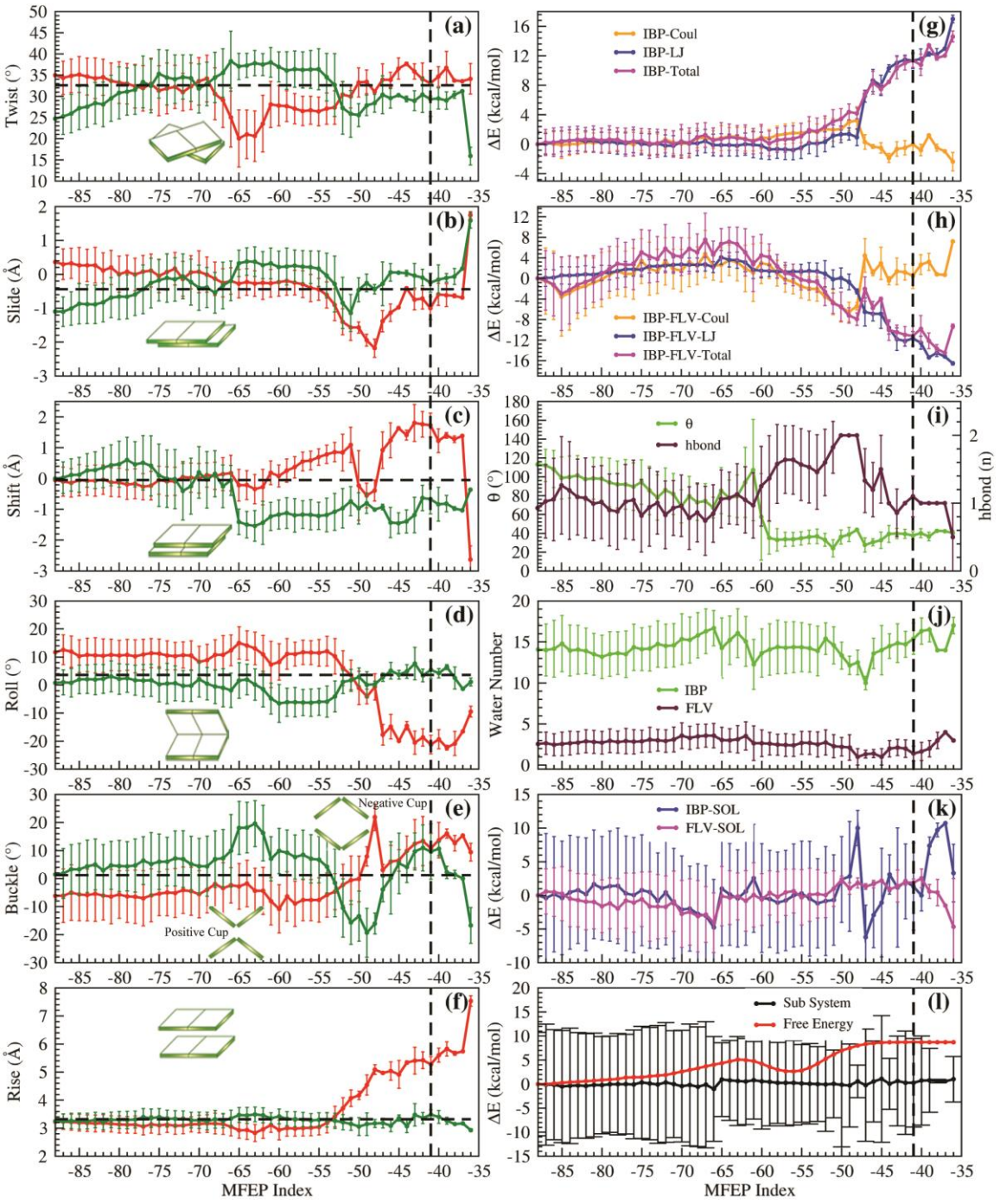


Figure 3.9. *Analysis of intercalation process from major groove-bound state. The vertical dashed line represents TS from MJS (MFEP index -41). The energy values are taken to be zero at major groove bound state (MFEP index -88). Each point shows the average values obtained from the members of the biggest cluster of the structures collected around discrete points along the MFEP. The standard deviation is shown. Left panel shows DNA base pair step parameters (red C6G19 and green G7C18), calculated by Curves+[29]. The horizontal dashed line represents the average value of normal B-DNA parameters[30]. Right panel shows various relative energy parameters along MFEP in (g),*

(h) which are taken to be zero at the MJS. (i) average θ value and number of hydrogen bond (hbond) between IBP and FLV, (j) number of water within 3.4 Å from heavy atoms of IBP and FLV, (h) interaction between IBP and FLV with water within 3.4 Å from heavy atoms of IBP-FLV complex, (l) Total energy of the subsystem comprising IBP, FLV and largest number of water along the entire MFEP within 3.4 Å from the heavy atoms of IBP-FLV complex.

3.3.5. Free Energy Profile along MFEP

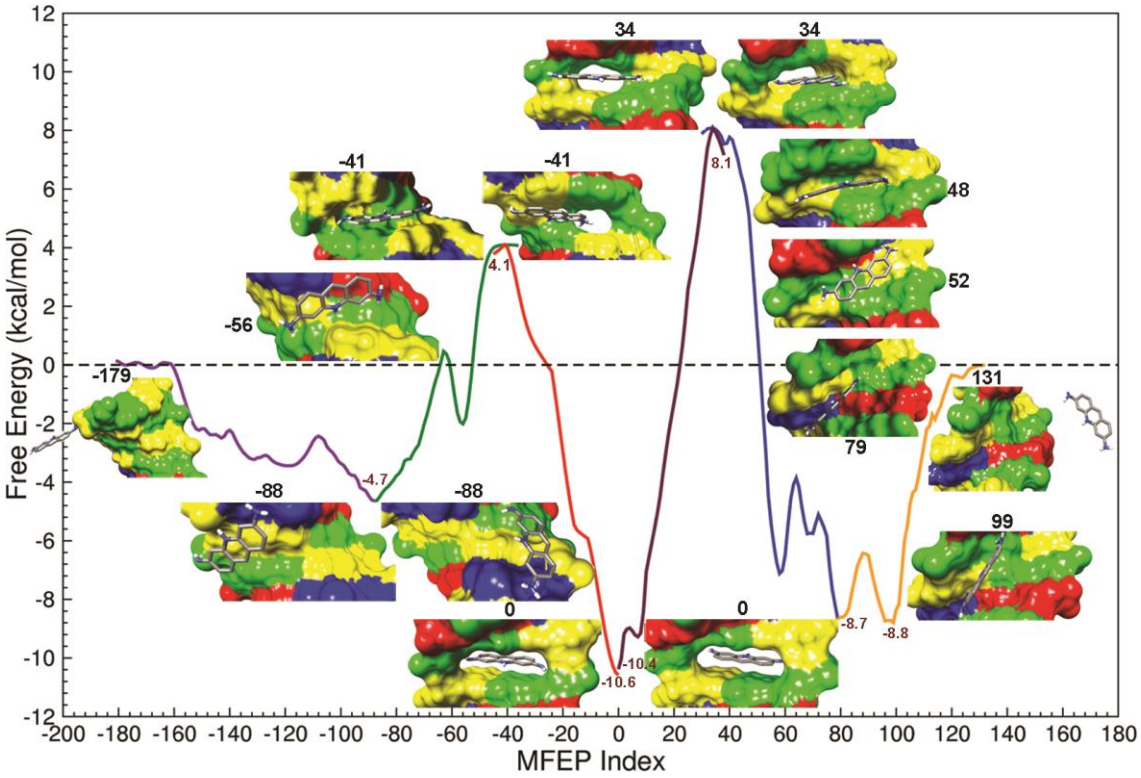


Figure 3.10. *Combined minimum free energy path (MFEP) for intercalation, deintercalation and dissociation processes. MFEP for each process is shown in different color for clarity. MFEP index of the intercalated state (IS) is taken to be 0. Positive and negative MFEP index correspond to the minor and major groove sides, respectively. The difference between each point has a radius of 0.05 in (X, θ) or (X, φ) coordinate. MFEP indexes are joined at 79, 34, 0, -41, and -88. The MFEP index for each process starting from right are: 132➔79, separated to minor groove-bound state;79➔34, minor groove bound state to transition state for intercalation from minor groove bound state ($TS^{Min \rightarrow IS}$); 34➔0, TS for deintercalation towards minor groove ($TS^{IS \rightarrow Min}$) to the IS; 0➔-41, IS to TS for deintercalation towards major groove ($TS^{IS \rightarrow Maj}$); -41➔-88, TS for intercalation from major groove-bound state ($TS^{Maj \rightarrow IS}$) to major groove-bound state; -88➔-181, Major groove bound state to the separated state. Representative configurations of DNA-drug complex at important places along the MFEP is shown.*

All six processes mentioned above were combined in Fig. 3.10 along the MFEP to obtain a comprehensive picture of proflavine intercalation to DNA. We had combined the

MFEPs and joined them to match the highest (TS's) and lowest (stable states) free energy points with the separated states taken to be zero. The path is drawn by taking MFEP index zero at IS. MFEP index increases towards the minor groove side and decreases towards the major groove side. MFEP index had been considered as a measure of steps taken on a multi-dimensional surface from IS towards major and minor groove side. The stability of the IS was calculated from the differences of intercalation and deintercalation barrier heights (as practiced in kinetic experiments) from minor groove side as, $\Delta G_{int} = \Delta G_{Min} + \Delta G_{Min}^{\#,de-int} - \Delta G_{Min}^{\#,int}$ (-10.4 kcal/mol) and from major groove side as, $\Delta G_{int} = \Delta G_{Maj} + \Delta G_{Maj}^{\#,de-int} - \Delta G_{Maj}^{\#,int}$ (-10.6 kcal/mol). Note that, the independent calculations of the free energetic stability of the intercalated state from two different pathways (major and minor groove) need to match due to the restriction from microscopic reversibility (see later). Therefore, the close agreement indicates the robustness of the method and sampling. It is important to mention here that one often obtains an estimate for the free energetic stability from forward and backward rate constants in kinetic experiments. However, it is rare to see in computational study because of the difficulty to obtain an accurate estimate for all the processes. Error estimation is done through additional FES calculation (See Annexure-3 Fig. A6).

3.3.6. Comparison of Thermodynamics Profiles

To capture the molecular origin of the thermodynamics of intercalation process of proflavine, in Fig. 3.11., we plotted the relative free energy and potential energy changes for deintercalation, dissociation and intercalation processes along the MFEP (see method for the calculation of energy profile and the estimate of the errors). For deintercalation through major groove, the free energy barrier is 14.7 kcal/mol and the energy barrier is 11.1 kcal/mol (Fig. 3.11a). So, the difference 3.4 kcal/mol may be attributed to entropy (multiplied by temperature). Similarly deintercalation free energy barrier through minor groove is 18.5 kcal/mol, while energy barrier is 23.2 kcal/mol (Fig. 3.11b). Therefore, entropy cost is 4.7 kcal/mol. For both of these processes, energy barriers originate from drug-DNA interactions. So the deintercalation barrier is mostly due to de-stacking energy, corroborated by a recent quantum calculation of proflavine with base pairs[31]. Recent studies using daunomycin[32] and ellipticine[33] had shown that intercalation

enhances DNA entropy. Therefore, the small loss in entropy due to deintercalation could be attributed to the loss of DNA entropy. Note that during deintercalation process, the configurational freedom of the intercalator or water does not change much.

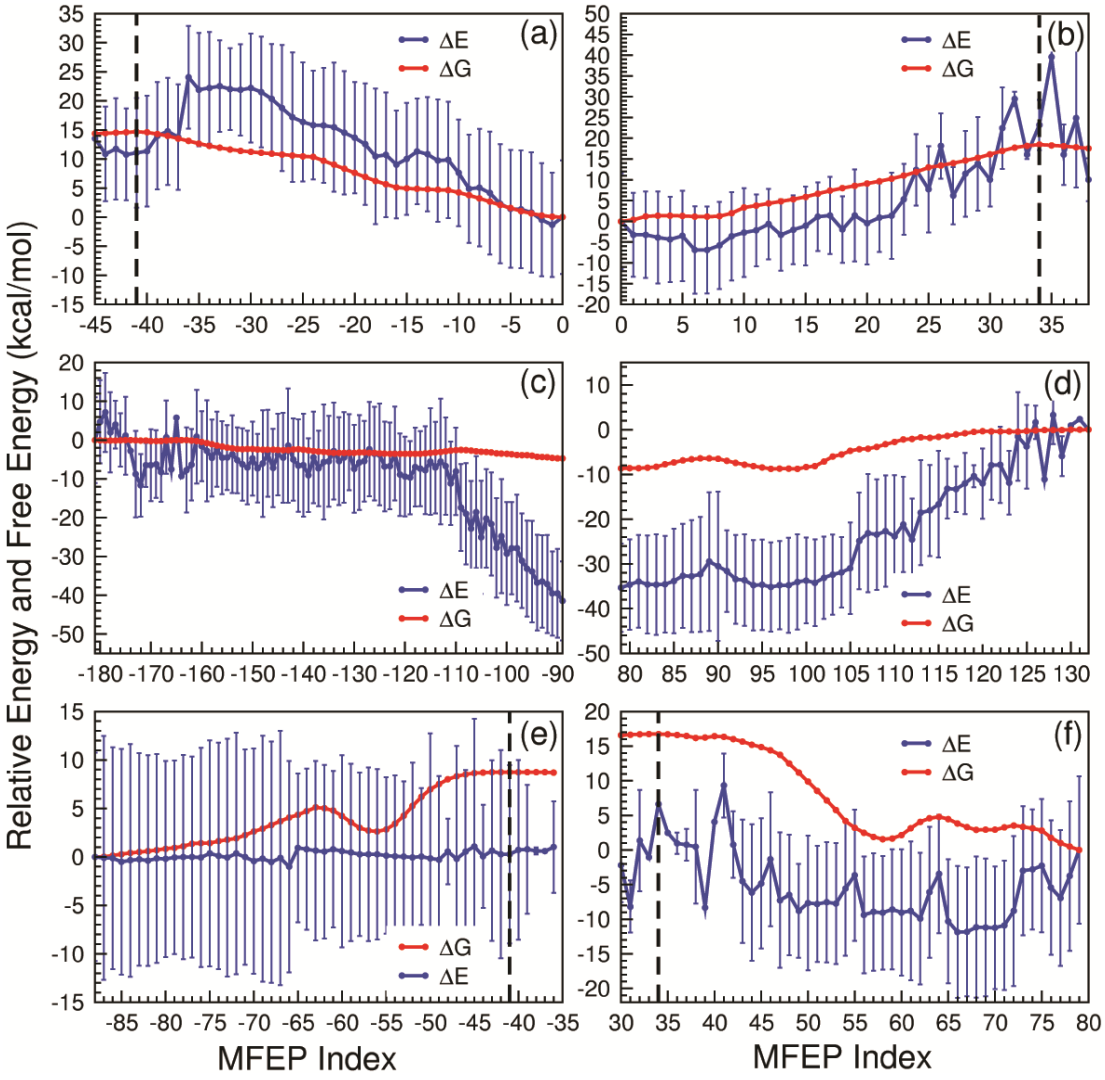


Figure 3.11. *Comparison of relative energy and free energy along MFEP for deintercalation, dissociation and intercalation processes. (a) Deintercalation through major groove side (b) Deintercalation through minor groove side (c) Dissociation from major groove bound state (d) Dissociation from minor groove bound state (e) Intercalation from major groove bound state (f) Intercalation from minor groove bound state. Red line shows relative free energy and blue line shows the total relative energy of the sub-system comprising intercalating base pairs (IBP), proflavine (FLV) and largest number of water along the MFEP within 3.4 Å from the heavy atoms of IBP-FLV complex. The vertical dotted line indicates the location of TS for different processes. To calculate relative changes in free energy and energy, they are set at 0 at the IS for (a) and (b); at completely dissociated state for (c) and (d);at the groove-bound states for (e) and (f). Error bars indicate standard deviation.*

Fig. 3.11c and 3.11d show the comparison of free energy and energy of the binding processes to major and minor groove-bound states respectively. Major groove bound state is stable by -4.7 kcal/mol of free energy and -41.5 kcal/mol of energy, indicating large negative entropy change -36.8 kcal/mol (Fig. 3.11c). For dissociation from minor groove bound state (Fig. 3.11d), free energy is -8.8 kcal/mol and energy is -35.3 kcal/mol, indicating again a large entropy change of -26.5 kcal/mol. Therefore, we find that external binding is entropically unfavourable, similar to recent observation in detailed entropy estimate of daunomycin[32], contrary to general assumption that minor groove-bound states are entropically favourable while intercalated states are enthalpically favourable[34-35].

The comparison of free energy and potential energy for intercalation process is extremely interesting. For intercalation through minor groove (Fig. 3.11f), the free energy barrier is 16.9 kcal/mol, whereas energy barrier is 6.7 kcal/mol, indicating an entropy barrier of 10.1 kcal/mol. For intercalation from major groove side (Fig. 3.11e), the free energy barrier is 8.8 kcal/mol and the energy barrier is 0.3 kcal/mol. Therefore, entropy component is similar for intercalation from both grooves. However, due to the absence of energy cost, the barrier of intercalation through the major groove is less than the barrier through the minor groove. Dissecting the energy component in detail, we find that the desolvation energy, which is the primary contributor to the free energy cost for intercalation[5] from the minor groove side, is not present in case of intercalation from major groove side (see Fig. S.9k) making intercalation barrier mostly entropic. Since the entropy cost is same for both the intercalation pathway, we think that the entropy cost originates from the loss of entropy of the drug. However, the further calculation is required to establish this. Only a detailed extensive entropy calculation for the drug, DNA and water along the path is required to establish that and it is beyond the scope of present calculation.

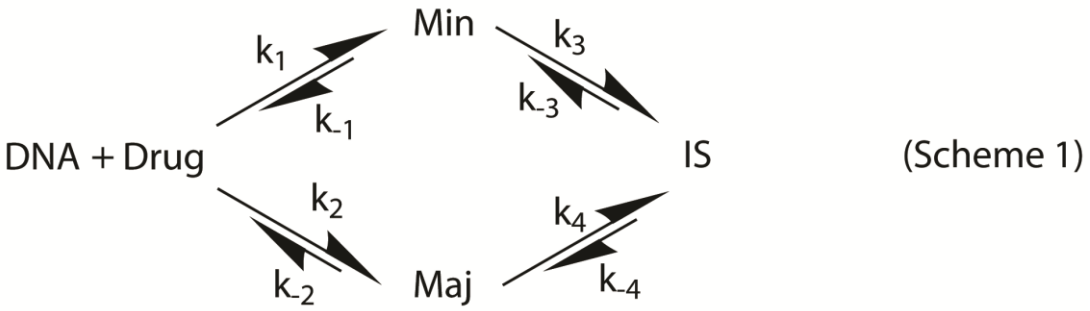
3.3.7. Kinetics from Thermodynamics

The effect of abortive pre-intercalative bound state in the intercalation process can be understood from the following kinetic analyses where we propose a kinetic scheme (Scheme 1) of the intercalation process as a two step one, as in experiment, however,

both the major and minor groove pathways are considered from our result of the free energy profile. As the outside bound state formation is a diffusion limited process[6], the rate constants for association to either major (k_2) and minor groove (k_1) are calculated from Smoluchowski rate of diffusion limited reaction by using the following equation[36],

$$k = \frac{4\pi N}{1000} R_{dp}(D_d + D_p) \qquad \text{Eq. (1)}$$

where, N = Avogadro number, R_{dp} = radii of reactants (reaction radius) taken to be 5 Å[36], D_d = Diffusion coefficient of DNA ignored because it is negligible as compared to small molecule proflavine, D_p = Diffusion coefficient of proflavine taken to be $7.5*10^{-6}$ cm^2/sec[37]. All other rate constants (k_{-1}, k_{-2}, k_3, k_{-3}, k_4 andk_{-4}) are calculated using the equation, $\left(\frac{k_B T}{h}\right) e^{\left[-\frac{\Delta G^{\#}}{k_B T}\right]}$. Note that, the requirement of microscopic reversibility is satisfied if one of the rate constants is obtained from the rest, known as Onsager reciprocal relation[38]. For Scheme 1, we can write the following relation using detailed balance for all the elementary steps, $\frac{k_4}{k_{-4}}\frac{k_2}{k_{-2}} = \frac{k_1}{k_{-1}}\frac{k_3}{k_{-3}} = e^{(-\Delta G_{int}/k_B T)}$. Therefore, the estimate of the free energy of the intercalated state should be same from both minor and major groove pathway from the requirement of microscopic reversibility, which is maintained in our estimate as mentioned before.



(Scheme 1)

Scheme 1: Possible intercalation pathways of proflavine

Fig. 3.12a and Fig. 3.12b show the kinetics of free DNA, major groove-bound state, minor groove-bound state, and intercalated state (IS), taking the initial concentration of the drug and DNA as 1.0. Free DNA quickly forms the groove-bound states by binding to both major and minor grooves. However, due to lower stability, the population in the major groove is less compared to the minor groove (Fig. 3.12a). The inor groove-bound state subsequently decays resulting in accumulation of the intercalated

state in a slow (milliseconds) process (Fig. 3.12b). To understand the complete pathway of intercalation, we decoupled the kinetic scheme (Scheme 1) in two competing pathways (Fig. 3.12c): (i) direct intercalation from minor groove-bound state and (ii) indirect pathway through major groove-bound state, and compared their kinetic profile (Fig. 3.12c). In both pathways we took the initial concentration of minor groove bound state as 1.0. Fig. 3.12c shows that the kinetic profile of the indirect pathway through major groove side mirrors the coupled complete kinetic profile, while the direct pathway through minor groove is extremely slow. Note that in the absence of the stable minor groove-bound state, the intercalation process will be in microseconds as shown in Fig. 3.12d. Therefore, although intercalation of proflavine happens through the major groove-side, the timescale is determined by the stable pre-intercalative minor groove-bound state.

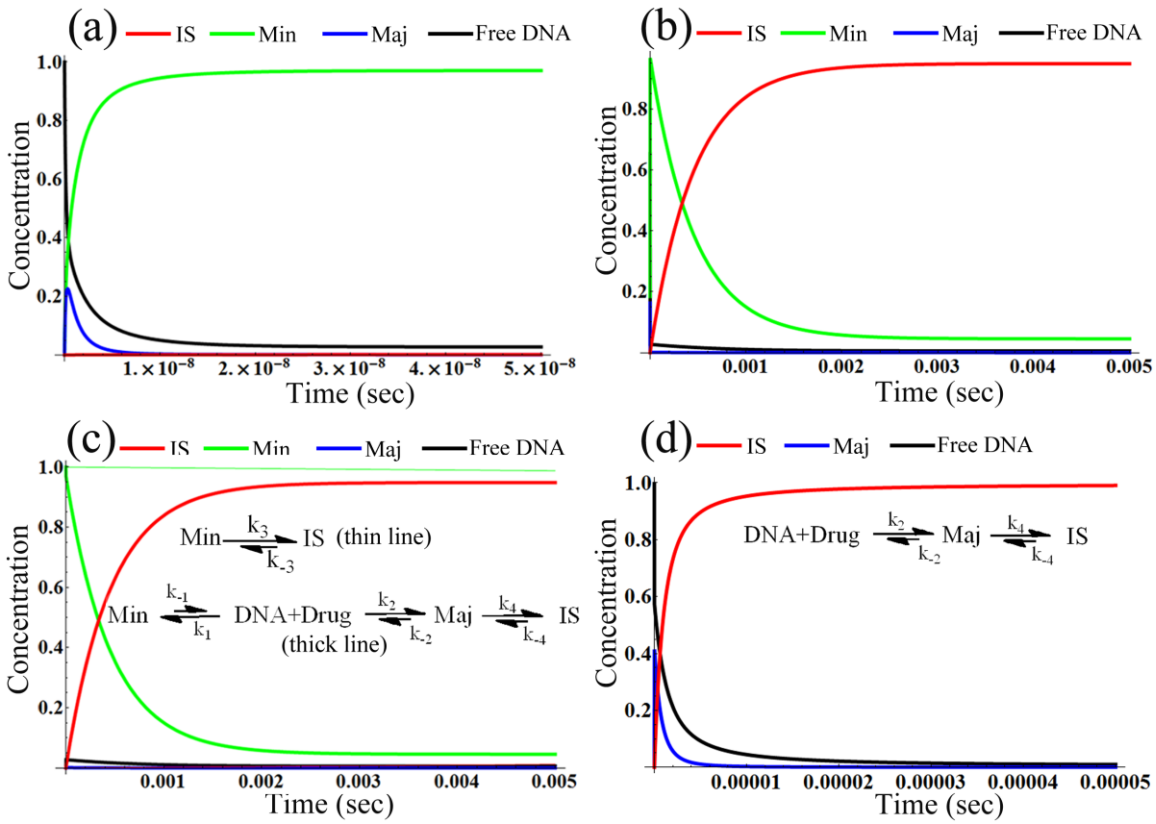


Fig. 3.12. *Kinetics of the intercalation process following scheme 1. (a) short time range to see the fast processes (Scheme 1), (b) entire time range (Scheme 1). (c) kinetic profile for the two competing intercalation pathways starting with the minor groove-bound state. (d) kinetic profile of intercalation through the major groove side in the absence of the minor groove-bound state.*

3.4. Conclusion

We captured in this study the molecular thermodynamics and kinetics of complete intercalation process of an anticancer agent proflavine into DNA through extensive computational methods. The results indicate that intercalation and deintercalation pathways of proflavine proceed through major groove side. However, intercalation is mostly dominated by the stable pre-intercalative minor groove-bound state, which builds up fast (nanoseconds) and slowly leaks to the intercalated state over a timescale of millisecond through major groove side. This long timescale is in excellent agreement with experiment. The free energetic components for the stable states and barrier heights are also in good agreement with experimental findings. However, till date, experimental studies only focused on an ensemble picture, and therefore, the molecular information was unavailable. This study predicts that the intercalated structure may not be as static as appears in the crystal structure and proflavine would likely change its orientation frequently (smaller barrier) even in the intercalated state. The importance of the entropy and desolvation energy is shown to play a major role in the kinetic process. Both the deintercalation and dissociation are energy driven processes and groove-bound states are found to be entropically unfavorable. Therefore, designing of a drug would eventually encounter the seemingly opposite effect that while an increased drug-DNA stacking interaction is required to stabilize the intercalated state, a less stable pre-intercalated state needs to be designed to make the kinetics faster.

3.5. References

1. Lerman, L. S., *J. Mol. Biol.* **1961,** *3*, 18-30.
2. E. F. Gale, E. C., P. E. Reynolds, M. H. Richmond, and M. J. Waring, The Molecular Basis of Antibiotic Action. Wiley: New York, 1972; pp 173-277.
3. Neidle, S., *Prog. Med. Chem.* **1979,** *16*, 151-221.
4. Mukherjee, A.; Lavery, R.; Bagchi, B.; Hynes, J. T., *J. Am. Chem. Soc.* **2008,** *130* (30), 9747-9755.
5. W. D. Sasikala and A. Mukherjee, *J. Phys. Chem. B*, 2012, **116**, 12208-12212.
6. Li, H. J.; Crothers, D. M., *J. Mol. Biol.* **1969,** *39* (3), 461-477.
7. Ramstein, J.; Ehrenberg, M.; Rigler, R., *Biochemistry* **1980,** *19* (17), 3938-3948.

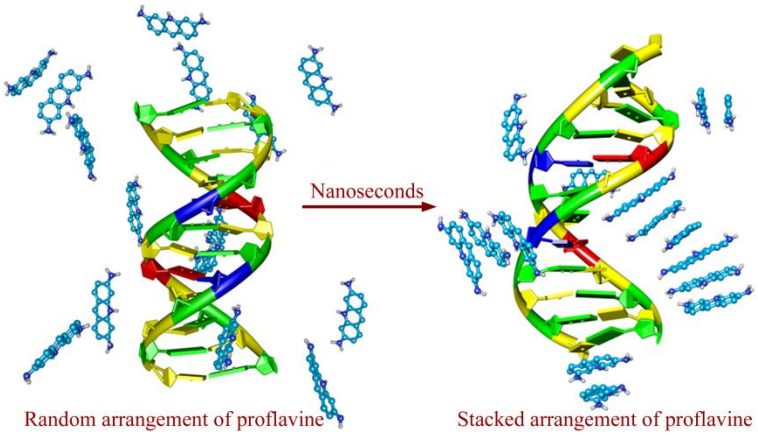
8. C Ciatto, M. D. A., G Natile, F Secco, M Venturini, *Biophys J.* **1999,** *77* (5), 2717-2724.
9. Corin, A. F.; Jovin, T. M., *Biochemistry* **1986,** *25* (14), 3995-4007.
10. Bereznyak, E.; Gladkovskaya, N.; Khrebtova, A.; Dukhopelnikov, E.; Zinchenko, A., *Biophysics* **2009,** *54* (5), 574-580.
11. Biver, T.; Secco, F.; Tinè, M. R.; Venturini, M., *Arch. Biochem. Biophys.* **2003,** *418* (1), 63-70.
12. Holman, G. G.; Zewail-Foote, M.; Smith, A. R.; Johnson, K. A.; Iverson, B. L., *Nat. Chem.* **2011,** *3* (11), 875-881.
13. Ramstein, J.; Leng, M., *Biophys. Chem.* **1975,** *3* (3), 234-240.
14. Macgregor, R. B.; Clegg, R. M.; Jovin, T. M., *Biochemistry* **1987,** *26* (13), 4008-4016.
15. Schwarz, G.; Klose, S., *Eur. J. Biochem.* **1972,** *29* (2), 249-256.
16. Laio, A.; Parrinello, M., *Proc. Natl. Acad. Sci. USA* **2002,** *99* (20), 12562-12566.
17. Barducci, A.; Bussi, G.; Parrinello, M., *Phys. Rev. Lett.* **2008,** *100* (2), 020603.
18. *http://www.rcsb.org/pdb/home/*, DOI:10.2210/pdb3ft6/pdb.
19. Macke, T.; Case., D. A., *Modeling unusual nucleic acid structures*. N.B. Leontes J. SantaLu- cia, Jr. ed.; American Chemical Society: Washington, DC, 1998.
20. Wang, J.; Cieplak, P.; Kollman, P. A., *J. Comput. Chem.* **2000,** *21*, 1049-1074.
21. Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M., *Biophys. J.* **2007,** *92* (11), 3817-3829.
22. Sorin, E. J.; Pande, V. S., *Biophys. J.* **2005,** *88* (4), 2472-93.
23. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., *J. Comput. Chem.* **2009,** *30* (16), 2785-2791.
24. Schlick, T., *Molecular Modeling and Simulation: An Interdisciplinary Guide*. 2nd ed.; Springer: New York, 2010.
25. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., *J. Chem. Theor. Comput.* **2008,** *4* (3), 435-447.
26. Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R., *Comput. Phys. Commun.* **2009,** *180* (10), 1961-1972.

27. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., *J. Comput. Chem.* **2004,** *25*, 1605-1612.
28. Dearing, A.; Weiner, P.; Kollman, P. A., *Nucl. Acids Res.* **1981,** *9* (6), 1483-1498.
29. Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K., *Nucl. Acids Res.* **2009,** *37* (17), 5917-5929.
30. Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T.; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; Maddocks, J. H.; Michon, A.; Osman, R.; Orozco, M.; Perez, A.; Singh, T.; Spackova, N.; Sponer, J., *Nucl. Acids Res.* **2010,** *38* (1), 299-313.
31. S. Li, V. R. C., T. Thonhauser, B. I. Lundqvist and D. C. Langreth, *J. Phys. Chem. B* **2009,** *113*, 11166-11172.
32. Mukherjee, A., *J. Phys. Chem. Lett.* **2011,** *2* (24), 3021-3026.
33. Kolář, M.; Kubař, T. s.; Hobza, P., *J. Phys. Chem. B* **2010,** *114* (42), 13446-13454.
34. Chaires, J. B., *Biopolymers* **1985,** *24* (2), 403-19.
35. Chaires, J. B.; Dattagupta, N.; Crothers, D. M., *Biochemistry* **1982,** *21* (17), 3933-40.
36. Alberty, R. A.; Hammes, G. G., *J. Phys. Chem.* **1958,** *62* (2), 154-159.
37. Malkia, A.; Liljeroth, P.; Kontturi, K., *Electrochem. Commun.* **2003,** *5* (6), 473-479.
38. Onsager, L., *Phys. Rev.* **1931,** *37*, 405-426.

Chapter 4

How Fast is the First Step of Intercalation? Structure and Dynamics of the pre-intercalative Bound State of an Anti-Cancer Agent Proflavine around DNA

Chapter 4: How Fast is the First Step of Intercalation? Structure and Dynamics of the pre-intercalative Bound State of an Anti-Cancer Agent Proflavine around DNA



DNA intercalation by small molecules often leads to arrest of replication, and thereby provides a mechanism for anti-cancer activity. As indicated by experiments, our previous calculations also suggest the formation of a metastable outside bound state or pre-intercalative bound state as the first step of intercalation process. This is also a faster process compared to the insertion into DNA, i.e., intercalation. However, experiments so far were unable to capture the nature of the outside bound state. The experimentally reported timescale of this association, observed using fluorescence kinetics, is in the range of sub-milliseconds. However, the exploration of pathway of intercalation using kinetic study as discussed in third chapter using proflavine as DNA intercalator has shown that the formation of pre-intercalative state happens in nanoseconds at high P/D ratio. Therefore, we have performed molecular dynamics simulations with multiple proflavine molecules to study the structure and dynamics of the formation of pre-intercalative outside bound state around DNA at different ion concentrations. We show that the timescale of outside bound state formation is at least five orders of magnitude faster (in nanoseconds) than the experimentally reported timescale (sub-milliseconds). Moreover, we also observed the stacked arrangement of proflavine all around DNA, which is different from the experimentally predicted stacking arrangement perpendicular to the helical axis of DNA in the close vicinity of the phosphate groups. This study gives an insight into the molecular structure and dynamics of the pre-intercalative bound state – the crucial step preceding intercalation.

4.1. Introduction

In chapter 3, we illustrated the mechanism of two-state binding of proflavine where it initially forms a minor groove-bound state in a diffusion limited free energy downhill process, followed by intercalation through major groove.[1] The calculated free energy stability of this outside-bound minor groove-bound state (-8.8 kcal/mol)[2] was found be closely related to experimental value (-6.8 kcal/mol)[3]. A Smoluchowski equation[4] of downhill diffusion process at P/D ratio 22 (one drug in twelve base pair DNA) showed that the timescale for this association should be in the order of nanosecond[1] which is contrary to several experimental results on the time scale of the association of proflavine to DNA. Fluorescence kinetic studies on proflavine dates four decades back with the pioneering work of Crothers, et al.[5] who suggested that proflavine intercalates into DNA in two steps, subsequently verified by a number of other studies over the last four decades. The kinetic study indicated that the first step is a diffusion limited fast process (0.15 ms) where proflavine binds outside DNA and the second step is a slow process (1.35 ms) characterized by the formation of the intercalated state.[5] Similarly, the temperature jump relaxation studies on the intercalation of proflavine on double helical poly(A) also suggests a two step process where the first step occurs in 0.35 ms and the second step completes in 4.0 ms.[6] To understand the effect of different base composition of DNA on the kinetics of the intercalation process, studies were performed on GC and AT rich DNA at different phosphate concentrations. It has been shown that the formation of the outside bound state, i.e. the first step, is affected by the phosphate concentration while the slow reaction (intercalation) is largely unaffected by the same. At larger phosphate concentration, the time scale of formation of fast outside bound state is 0.4 ms and 0.13 ms for GC and AT rich DNA, respectively. The time scale of formation of slow intercalated state is 5.0 ms and 8.3 ms for GC and AT rich DNA, respectively.[3] Similar trend in time scales is also obtained for a modified DNA (halogenated DNA) where the faster timescale is 0.24 ms and the slower timescale is 1.3 ms.[7] However, Sutin et al. observed a faster timescale, in the range of 30-700 ns, for the dimerization of proflavine in solution using laser temperature jump experiments.[8]

Of the two states, the intercalated state is more stable. Thus, the nature and structure of it was captured by several experimental studies including spectroscopic

methods,[3, 5, 9] NMR[10], and crystallographic[11] studies etc. Outside bound state being a metastable state remained elusive and received relatively less attention. Several studies, however, demonstrated a concentration dependence of proflavine on the formation of the two states. Peacocke and Skerret[12] have shown that the increase in the concentration of proflavine leads to the beginning of a weak binding process characteristic of the interaction among proflavine molecules. The weak binding was ascribed to the formation of proflavine aggregates or micelles on DNA. This was evidenced by the fact that proflavine molecules can interact mutually at concentrations greater than 2.5×10^{-5} M and undergo formation of aggregates or micelles in solution[13]. The optical rotatory dispersion studies of proflavine-DNA complex also suggested the presence of close association of proflavine molecules (stacking) to be the reason for the optical activity of proflavine in presence of DNA[14]. The formation of the intercalated complex and the external bound complex of proflavine to DNA depends on the ratio of phosphate (P) to dye (D) ratio (P/D). P/D is the ratio of the number of nucleotides (base+sugar+phosphate)[15] or phosphates[16] per a bound dye. At higher concentration of the dye, P/D $\leq$1, the visible absorption spectrum shows a blue shift. This indicates that the externally bound structure is stabilized by proflavine-proflavine stacking interaction. At low concentration of proflavine, P/D~10, there is a red shift in the visible absorption spectrum characteristic of the intercalated complex.[15, 17] Similar type of association was predicted for acridine orange, a DNA intercalating agent and belonging to the same chemical family of proflavine, where acridine orange undergoes stacking and aggregation on the double helical DNA.[17-18] This is further confirmed by NMR experiments[19]. Even though crystal structure of proflavine bound to modified Z-DNA complex is resolved where stacked proflavine molecules are bound to the minor groove of the DNA[20], the nature and structure of outside bound state of proflavine to a regular B-DNA remains elusive.

Since the timescale of formation of outside bound state of proflavine to DNA is very fast to be captured by any experimental methods, several experiments were done on proflavine in solution(in the absence of DNA) to understand the association behavior of proflavine at different concentrations. In very dilute solutions (<0.1 nM), dimer formation is absent. However, at high concentrations dimer is formed and it is attributed to the stacking of the neighboring bound molecules.[21] The formation of larger aggregates were

seen at concentrations greater than 10 mM.[22] Further, the dimer[11] and trimer[23] structures of proflavine in solution were confirmed by X-ray crystallography. The thermodynamic studies suggested the dimer formation to be a diffusion controlled process as the change in enthalpy for the process is only 0.5 kcal/mol whereas the dissociation process happens by an enthalpy change of 14.8 kcal/mol.[8] The calculation of thermodynamic parameters of proflavine aggregation in solution revealed that the decrease in free energy (ΔG=-4.0 kcal/mol)of the polymerization reaction of proflavine is governed by favourable enthalpic contributions (ΔH=-7.13 kcal/mol), and opposing entropy contribution (ΔS=-10.5caldeg^{-1}mol^{-1}) for each step in polymerization of proflavine.[24]

It is found that ion concentration also affect the outside bound state of proflavine. The effect of varying concentrations of NaCl on binding of proflavine to DNA shows that there is a reduction in outside binding as the concentration of NaCl is increased. Moreover, the addition of magnesium ions reduced the binding to a greater extent.[12] Also, increase in NaCl concentration is shown to decrease the lifetime of the associated proflavine molecules at the end of the polymeric aggregates.[25]

Even though the structure of the outside bound state at the pre-intercalative stage is not known, it is believed that the stacking between the dye molecules bound externally to the polymer chain plays an important role in the dynamics of the intercalation process.[6] Moreover, in chapter 3, we have shown that how a pre-intercalative state affects the kinetics of the subsequent intercalation kinetics using computational methods[1]. A direct evidence of the timescale of binding of proflavine to DNA however was not done in the previous study. Also, the previous study used only one drug to study the second step (intercalation) and therefore could not provide the structural and dynamical aspects of the binding event outside the DNA. So, the aim of the present study was to investigate how fast is the timescale for the formation of the pre-intercalative outside-bound state during an intercalation process. Also, we wanted to address the structural details for the association of proflavine outside DNA. We addressed these questions using all-atom molecular dynamics simulations involving proflavine molecules and DNA at low P/D ratio and different ion concentrations. The outcome of the observation indicated the re-interpretation of multiple timescale observed in kinetic experiments involving proflavine and other intercalators.

4.2. Experimental Section

4.2.1 Proflavine-DNA System and Equilibration

A 12 base pair DNA, $d(GCGCTCGAGCGC)_2$[2] having the same sequence as discussed in previous chapters, was created using Nucleic Acid Builder (NAB)[26] program. We took 22 proflavine (cationic) molecules ($3.2*10^{-5}$ M) to neutralize the negatively charged DNA and to maintain the P/D ratio as 1. Proflavines were randomly distributed around DNA in a cubic box of length 72 Å and the system was solvated by 12222 TIP3P[27] water molecules. Four such systems were made by varying NaCl concentration such as 0 M, 0.15 M(physiological ion concentration), 0.5 M and 1 M. GAFF(General Amber Force Filed)[28] was used for proflavine whereas, AMBER99/parmbsc0[29] force field was used for DNA. The simulations are performed using GROMACS[30] software. After the initial energy minimization using the steepest descent method[31] each system was heated up to 300 K using a Berendsen thermostat[32] with a coupling constant of 0.2 ps at a harmonic restraint of 25 kcal/mol/Å^2 on heavy atoms of proflavine and DNA. Further, a series of simulation(50ps) and energy minimizations(1000 steps) were done on each system while reducing the restraints from of 25 kcal/mol/Å^2 to 5 kcal/mol/Å^2 at constant temperature 300 K and pressure 1 bar using Berendsen thermostat and barostat. A final equilibration for 2 ns was done with a harmonic restraint of 5 kcal/mol/Å^2 at constant temperature 300 K and 1 bar pressure using a Nose-Hoover thermostat[33] and a Parrinello−Rahman barostat[34] respectively. The time step simulation was kept at 2 fs. Particle mesh Ewald (PME)[35] (cutoff at 10 Å) was used to treat electrostatic interaction, whereas, the van der Waals (vdW) cutoff was taken to be 10 Å. Finally, the constraints on both DNA and proflavine were removed and three 30ns simulations were carried out with different initial velocity distribution to study the time dependent association process. Therefore, overall simulation length was 360 ns.

4.2.2. Proflavine in Water

A set of separate molecular dynamics (MD) simulations was carried out with only 22 proflavine molecules ($3.2*10^{-5}$ M) in solvent containing NaCl at physiological ion concentration to understand the stacking behavior of proflavine in solution. Total simulation length was 90 ns.

4.2.3. Dimer Free Energy Calculation

To calculate the free energetic stability of proflavine dimer, we first created a dimer by docking using AutoDock[36] and solvated in a cubic box containing TIP3P[27] water molecules and physiological ion concentration. Further equilibration and energy minimizations were followed as described above. Finally, umbrella sampling[37] simulations (4 ns) were performed along the distance between the center of mass (COM) of two proflavine molecules in 20 different simulations each window separated by 0.35 Å. The free energy profile was generated by weighted histogram analysis method (WHAM)[38]. Total simulation length was 80 ns.

4.2.4. Stacking Criteria

We have used the following criteria to identify a stacked pair of proflavine molecules: (i) distance between the COM of two molecules should be less than 4 Å and (ii) the angle between the normals should be either less than 45° (parallel stack) or more than 135° (anti-parallel stack). Normals are calculated using same set of atoms for each proflavine molecule. Fig. 4.1.a and 4.1.b show representative stacked pair in parallel orientation and anti-parallel orientation.

To find the number of stacked proflavine around DNA, an additional condition is applied to omit the counting of stacked proflavine in the bulk. Since proflavine can stack around another proflavine which is in turn stacked with a third one, all stacked proflavine need not be close to the DNA. However, the cluster of stacked proflavine molecules may still close to the DNA. Therefore, we put the condition that at least one of a larger cluster of stacked proflavines needs to be close to DNA to be counted as stacking around the DNA.

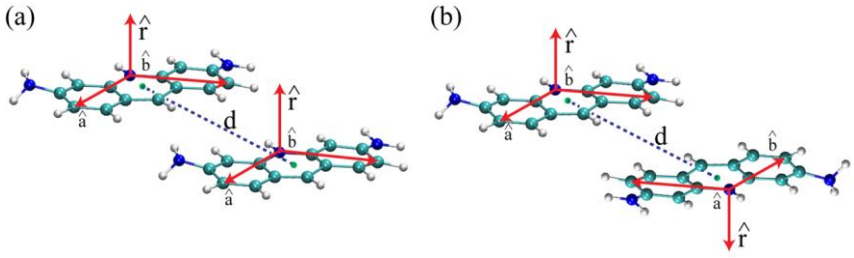


Figure 4.1. *Two proflavine molecule in stacked configuration. Distance (d) between COM is shown by the dashed line. Vectors* $\hat{r_1}$ *and* $\hat{r_2}$ *are the normal vectors obtained from the cross product of* $\hat{a}$ *and* $\hat{b}$ *unit vectors as shown. Representative (a) parallel and (b) anti-parallel orientation is shown.*

4.3. Results and Discussions

4.3.1. Proflavine Aggregation in the Absence of DNA

Before exploring the association of proflavine molecules around the DNA, we studied the aggregation behavior of proflavine in solution itself. We performed MD simulations of 22 proflavine molecules in water. This corresponds to a concentration ($3.2*10^{-5}$ M) higher than that reported for aggregation ($2.5x10^{-5}$M). As the simulation progressed, the initial random arrangement, representatively shown in Fig. 4.2.a, changed to clusters of proflavine through stacking. Figure 4.2.b shows a representative image of the stacked configuration. Formation of stack is consistent with the suggestion made through several experimental studies[21, 24] and crystal structure[23].

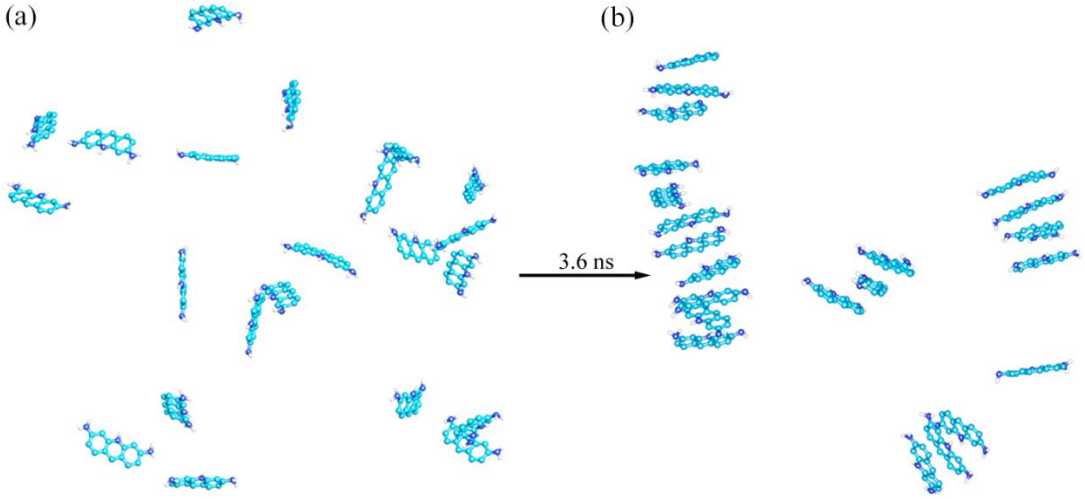


Figure 4.2. *The cluster formation of proflavine molecules by stacking. (a) Initial random distribution of proflavine molecules and (b) the arrangement of proflavine molecules to clusters stabilized by stacking interaction.*

To quantify the amount of stacking and also to understand the timescale of formation of stacked complexes, we calculated number of proflavine molecules in stacked configuration with at least one other proflavine molecule. Figure 4.3.a shows the growth in the number of stacked proflavine molecules with time, averaged over three independent simulations. The curve was fitted with an exponential equation of the form, $\boldsymbol{a_0 - a_0 e^{-(t/\tau)}}$, where, τ is the time constant for stacking, $\boldsymbol{a_0}$ represents the total number of stacked molecule in equilibrium, and t represents the simulation time. Initially, at t=0, all the proflavines are separated from one another, thus making $\boldsymbol{a_0}$=0. The number of

proflavines in stacked configuration increases with time, i.e., proflavine molecules starts stacking with each other. Finally, after a certain time, the number reaches a saturation where $\boldsymbol{a_0}$=19. The average value at equilibrium after 15 ns trajectory is 19.1 ±0.8.

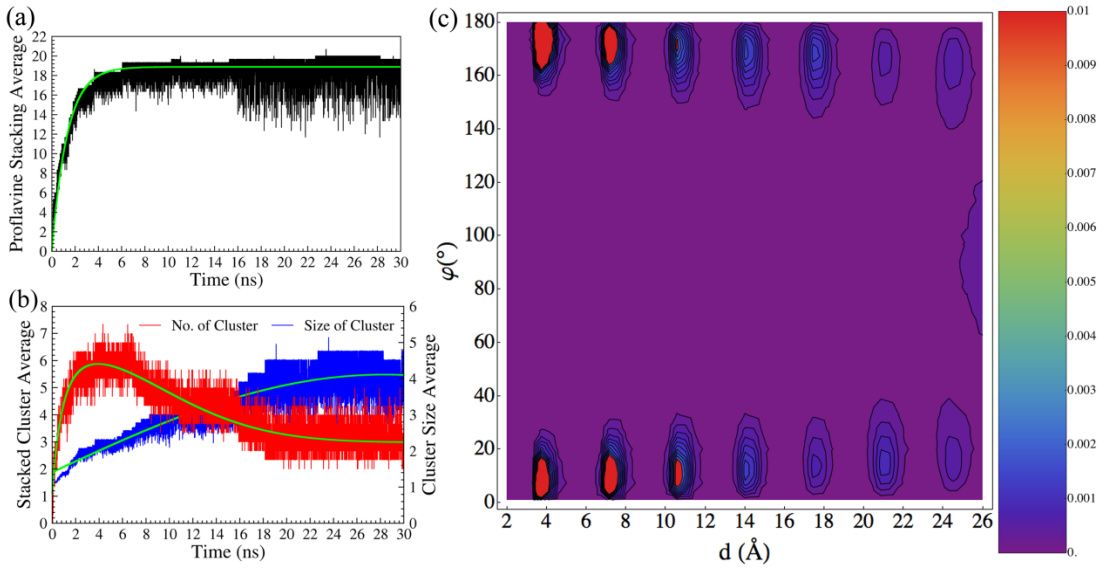


Figure 4.3. *Stacking of proflavine molecules in solution. (a) Average number of stacked proflavine molecules in water along the simulation time, (b) Average number stacked cluster and cluster size of proflavines along simulation time and (c) Probability distribution of distance and orientational distribution of proflavine molecules with one another. The green curves in (a) and size of cluster in (b) denotes fit using the equation,* $a_0 - a_0 e^{-(t/\tau)}$. *The green curve in number of cluster in (b) denotes fit using the equation,* $a_0 + a_3 * exp(-(t - a_1)/a_2) + a_6 * exp(-(t - a_4)/a_5)$.

All the stacked proflavines do not belong to a single cluster, as it would be opposed by entropy[24]. A cluster is determined by the criteria that every proflavine must be in stacking interaction with at least one member of the same cluster. The average number of clusters of stacked proflavine molecules is shown in Fig. 4.3.b. A dimer is denoted by a cluster of two members, whereas a trimer has three proflavine molecules in a cluster. In agreement with experimental studies[11, 21-23], we observed the formation of dimers, trimers, and polymers of proflavine molecules. Initially, the number of stacked clusters is zero, i.e., all the proflavine molecules are separated from each other. With the progress in time, small sized clusters forms and the total number of cluster goes to a maximum of 7.3. Finally the some small clusters join to form bigger clusters and cluster number comes down to an equilibrium value of 3.2. This can be considered as the reflection of cooperativity described experimentally for proflavine and acridine orange around

biopolymers in solution.[21, 25] Cooperativity is again substantiated by the estimation of average number of members of the stacked clusters (cluster size) as shown in Fig. 4.3.b. Initially, at time t=0, none of the proflavine molecules are in stacked configuration indicating that proflavines exist as monomers giving rise to an average cluster size 1. With time, proflavine molecules start stacking as seen in Fig. 4.3.a along with an increase in the average cluster size shown in Fig. 4.3.b. However, almost all proflavines are in stacked configuration within 15 ns where average cluster size is still low (~2.8). Beyond 15 ns, the average cluster size increases reaching a uniform average value, 3.9±0.4. This indicates that in the last phase, smaller clusters coalesce to form bigger cluster as simulation progresses. This is also supported by the number of stacked clusters of proflavines which reaches an average value of 7.3, but decreases thereafter reaching a final average value of 3.2±0.5. The distribution of stacking aggregates is clear in Fig. 4.3.c where the mutual distance and orientation between two proflavine molecules after equilibrium (15 ns to 30 ns) is shown as a contour plot. The distributions are separated by an approximate discrete distance of 3.4 Å and the angle between the proflavine molecules is either 10° (parallel stack) or 170° (antiparallel stack). Figure 4.3.c shows that both parallel and anti-parallel stacking are possible. Also, the probability distribution for proflavines separated at a large distance 25 Å is also dominated by stacking interaction indicating that a large cluster of mutually stacked proflavine molecules exists in solution.

The time constant, τ for stacking of proflavine molecules is 3.6 ns. Although this is much smaller than that (30 ns to 700 ns) reported for proflavine dimer formation using laser temperature jump experiment[8], the difference is only an order of magnitude from the fastest reported timescale. We will subsequently see that for proflavine association around DNA, a seemingly similar phenomenon, the difference is much bigger.

4.3.2. Free Energy of Proflavine Dimer

The formation of multiple stacked proflavine molecules indicates the stability of the chain of proflavine molecules. To understand the stability of the basic building block of a chain of proflavine molecules, we calculated the free energetic stability of dimer formation using umbrella sampling method[37]. Starting with the equilibrated and stable anti-parallel orientation, the dimer was separated into two monomers along the distance

between the COMs of the two. The free energy profile for the dissociation is generated using WHAM[38] as shown in Fig. 4.4.a. The convergence of free energy profile is shown in Fig. 4.4.b. Figure 4.4.a shows that interaction between proflavines extends up to 8 Å beyond which free energy profile becomes flat. Also, there is no barrier for association, indicating a diffusion limited downhill process. The overall free energetic stability for dimer formation is thus found to be -4.9 kcal/mol, in good agreement with experimental value of -4.0 kcal/mol[24]. This interaction can be attributed to the stacking energy between the two. This agreement indicates that the force-field for proflavine is reasonable. The small discrepancy may be due to the dilute dimer concentration considered here compared to an average of multiple varieties of polymeric structures.

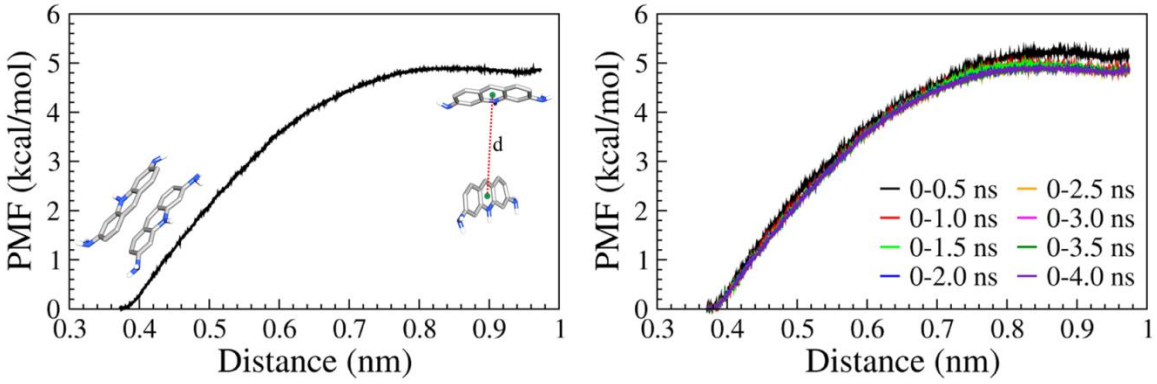


Figure 4.4. *The free energy profile for the dissociation of proflavine dimer showing the initial and final configurations. (a) The red dotted line indicates the reaction co-ordinate, d, denotes the distance between the COMs of two proflavine molecules. (b) Convergence of the Potential Mean Force (PMF) obtained for the dissociation of proflavine dimer along distance between the COMs of two proflavine molecules.*

4.3.3. Proflavine Aggregation in the Presence of DNA

We have seen so far that proflavine molecules tend to stack in solution within a timescale of a few nanoseconds. This study encouraged us further to investigate the structure and timescale of association of these molecules around DNA. Our previous kinetic study on the proflavine intercalation to DNA[1] suggested that the outside bound state formation could happen in nanoseconds, while experimental studies[3, 5-6] argue in favor of sub-millisecond time scale for the same. However, our previous estimate was based on a kinetic analysis from the free energetic stability of proflavine at the outside bound state at high P/D ratio of 22 (single proflavine and a 12 base pair DNA). At this

concentration, no external bound proflavine is seen experimentally.[9, 15] Therefore, we attempt here to directly probe the association using molecular dynamics simulation with a higher concentration of proflavine molecule around DNA (P/D=1) at which outside bound states were predicted experimentally[12]. Moreover, we performed this study at four different ionic concentrations (0.0 M, 0.15 M, 0.5 M and 1.0 M NaCl) to understand the effect of ions in the drug-DNA association.

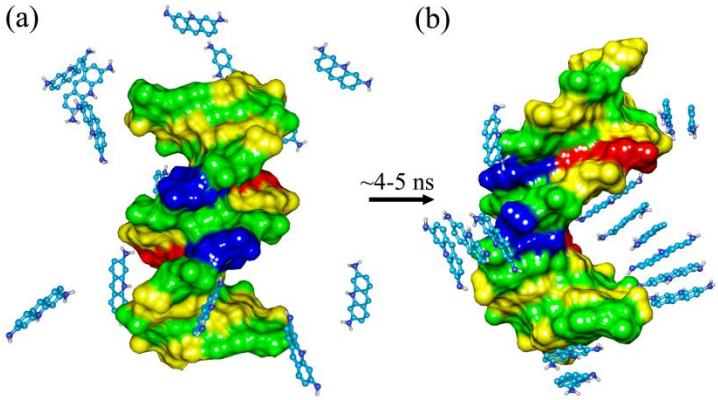


Figure 4.5. *A representative configuration of proflavine around DNA. (a) Initial random arrangement of proflavines around DNA and (b) stacked arrangement of proflavines around DNA.*

As the simulation progressed, we observed that the proflavine molecules, initially randomly distributed, arranges around DNA in stacked orientation. A representative picture of the association is shown in Fig. 4.5. Fig.4.5.a represents the randomly distributed proflavine around DNA and Fig.4.5.b represents proflavine molecules arranged as clusters on the surface of DNA.

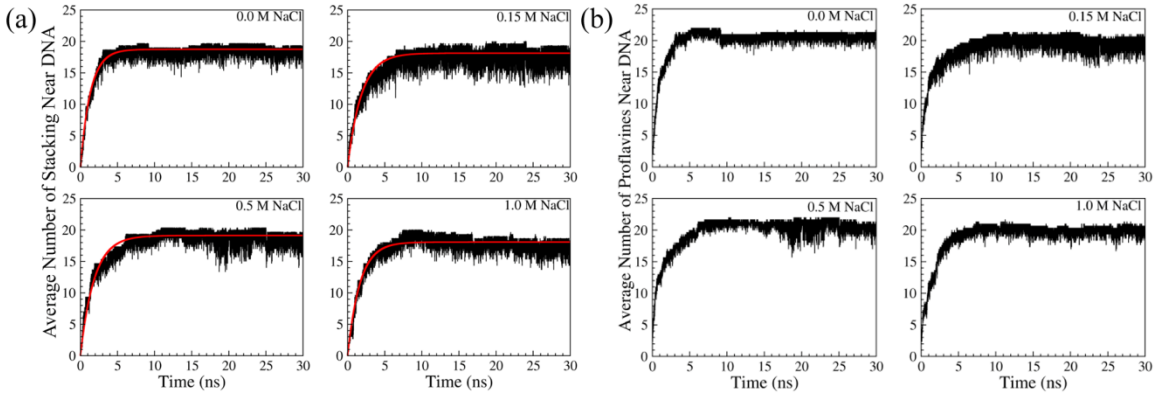


Figure 4.6. *Average number of (a) stacked proflavine and (b) average number of associated proflavine around DNA at four different NaCl concentrations. The red lines show the fit using the equation,* $a_0 - a_0 e^{-(t/\tau)}$.

Figure 4.6 shows the time dependence of the association process where Fig. 4.6.a shows the average number of stacked proflavines around DNA and Fig. 4.6.b shows the average number of associated proflavine including both stacked and unstacked proflavines around DNA at various ion concentrations. Figure 4.6.a shows that number of stacked proflavine molecules increases with time at all ion concentrations. Since proflavine can stack around another proflavine which is in turn stack with a third one, all the stacked proflavine need not be close to the DNA. Therefore, to count the number of stacked proflavine around DNA and to prevent counting the proflavine stacked in the bulk, we used the condition that at least one pair of a larger cluster of stacked proflavine needs to be close to DNA to be counted as stacking around the DNA. Therefore, Fig. 4.6.a does not count the monomeric proflavine around DNA. Total number of proflavine around DNA including unstacked molecules with time is shown in Fig. 4.6.b. From the figure it is clear that for all ion concentrations, after 30 ns simulation, out of 22 proflavine molecules 19 to 21 molecules are closer to DNA.

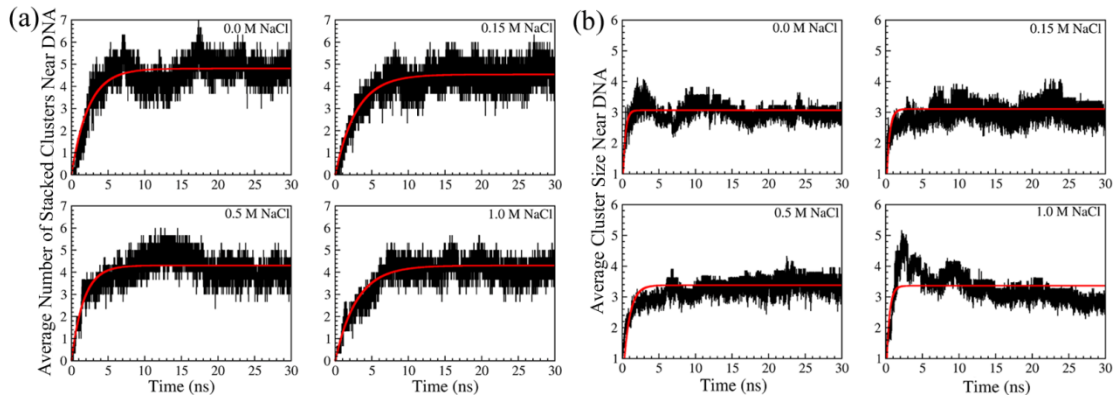


Figure 4.7. *Average number of (a) clusters of stacked proflavine and (b) average cluster size number around DNA at four different NaCl concentrations. The red lines show the fit using the equation,* $\boldsymbol{a_0 - a_0 e^{-(t/\tau)}}$.

The average number of stacked cluster number of proflavines around DNA is shown in Fig. 4.7.a. From the graph, it is clear that as the time progresses, the average number of stacked cluster around DNA increases initially and reaches a plateau region suggesting the increase in the stacked configurations of proflavines are associated with the increase in the cluster formation. So in case of proflavine molecules around DNA, the initial sharp increase followed by the decrease in cluster formation is not seen as we have observed in case of proflavine in solution only. This suggests that the cooperativity seen

in case of proflavine molecules in solution is not observed in case proflavine around DNA. This is again demonstrated by the average number of cluster size of proflavine molecules around DNA as shown in Fig. 4.7.b. We can consider the presence of cooperativity if the plateau region of cluster formation (see Fig. 4.7.a) is characterized by an increase in cluster size at that region. However, from the Fig. 4.6.b, it is evident that there is no further increase in the cluster size after 15 ns. Initially, at time, t=0, the average number is one. With time, the number starts increasing and finally reaches to a uniform value suggesting the absence of cooperativity. Figure 4.7 shows that both the number of cluster of stacked proflavine and also the average number of cluster size of stacked proflavine increases with time for all the different ion concentrations. Here, increase in the number of cluster size around DNA may be just a reflection of the increase in the number of stacked proflavine in the existing clusters.

Table 4.1*: Average number of stacking, time constant (τ_{stack}), number of stacked cluster and stacked cluster size at different ion concentrations.*

Concentration of ions (M)	Ave. No. of Stacked Proflavine near DNA	τ_{stack} (ns)	Ave. No. of Stacked Custer near DNA	Ave. No. of Stacked Cluster Size
0.00	18.3±1.3	1.3	4.9±0.3	3.0±0.2
0.15	18.2±1.8	1.9	4.6±0.4	3.1±0.6
0.50	19.1±1.6	1.8	4.2±0.8	3.5±0.4
1.00	17.7±1.6	1.7	4.2±0.9	3.0±0.2

Table 4.1 shows the number of stacked proflavine around DNA at equilibrium at different ion concentrations. Out of total 22 molecules, 18 to 19 molecules stack around DNA. Also, there are on an average 2 molecules, which are unstacked (separate monomer) around the DNA. This is expected given the free energetic stability of the proflavine molecules around DNA reported earlier by our earlier computational study[1-2]and several experimental studies[24]. However, the important point here is to show that most of the proflavine around DNA (~21) remains in stacked configuration (~19).

The time dependence of proflavine-DNA association in each case was fitted to the similar equation mentioned above. The time constant for stacking around DNA is shown in Table 4.1. At all ion concentrations, the stacking happened within a few nanoseconds, consistent with our earlier kinetic study at higher P/D ratio[1]. This study, therefore, shows

for the first time that the formation of the outside-bound state in the pre-intercalative stage is five orders of magnitude faster when compared to what deduced from experimental observations[3, 7, 9-11, 39-40]. Therefore, the interpretation of the sub-millisecond time scale observed in experiment may be attributed to a different facet of intercalation itself, rather than to the timescale for outside binding as thought earlier.

The timescales do not depend significantly on the ion concentration (Table 4.1), although we observed a small increase in the timescale with ion concentration within the range considered here (<1M). This small increase may be attributed to the competition of NaCl with proflavine for the binding sites on DNA as shown by the experimental studies[25]. This is evidenced by the fact that there is an increase in the NaCl concentration around DNA (at 1.0 M NaCl) as shown in the comparison of radial distribution function of Na+ and proflavine around DNA (See Fig. 4.8). In case of proflavine around

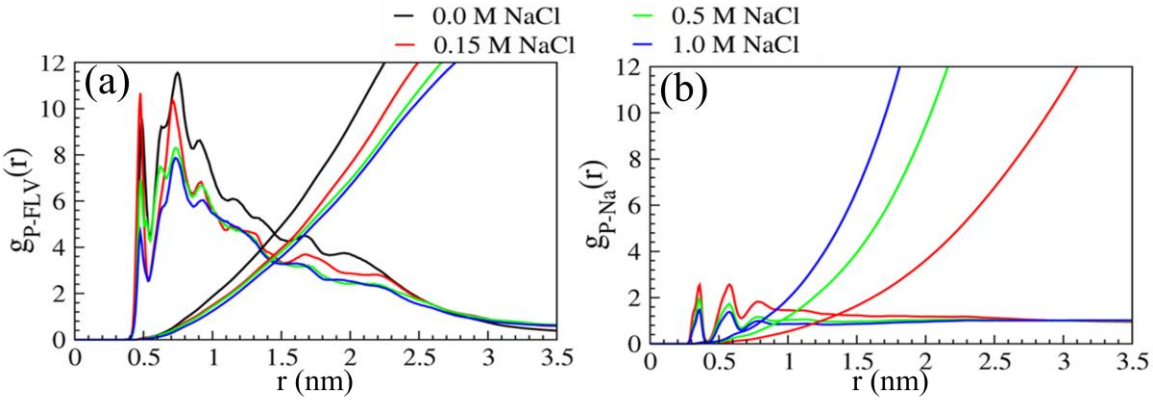


Figure 4.8. *Radial distribution function and co-ordination number of proflavine molecules and Na+ around backbone of DNA. (a) Proflavine around phosphate group of DNA (b) Na+ around phosphate group of DNA.*

phosphate group of DNA (Fig. 4.8.a), there is a slight decrease in the co-ordination number as the concentration of NaCl is increased. This may be due to the competition of Na+ ions with proflavine for the phosphate group which is evidenced by the slight increase in the co-ordination number of Na+ ions around phosphate as shown in Fig. 4.8.b. Since the highest ion concentration used in this study is already much higher compared to that used in experimental kinetic studies[25], we can rule out the effects of ion in slowing down the association kinetics observed in experiments. Experimental study[21, 25] indicates that there is a cooperativity associated with the stacking of proflavine

molecules around linear biopolymers. In cooperative binding, proflavine monomer or dimer binds to the biopolymer first. This is followed by binding of another proflavine monomer or dimer in the immediate neighbor of already stacked proflavine molecules. Finally, these adjacent neighbors stack together to form cluster. The studies on proflavine stacking show that because of the co-operative effect, the members of the cluster can be up to 6[21, 25]. Although the number of cluster is in excellent agreement for all ion concentrations (Table 4.1), we do not observe any signature of cooperativity at all ion concentrations which require an initial increase followed by a decrease in the number of clusters (or increase in average size) as was observed for proflavine in solution. To understand the cooperativity, we monitored the average size of the cluster (including monomer) with time. The average size increases indicating that the proflavines bind with the same cluster. However, eventually, the size goes down to 3.

4.3.4. Structure of the Outside-bound State

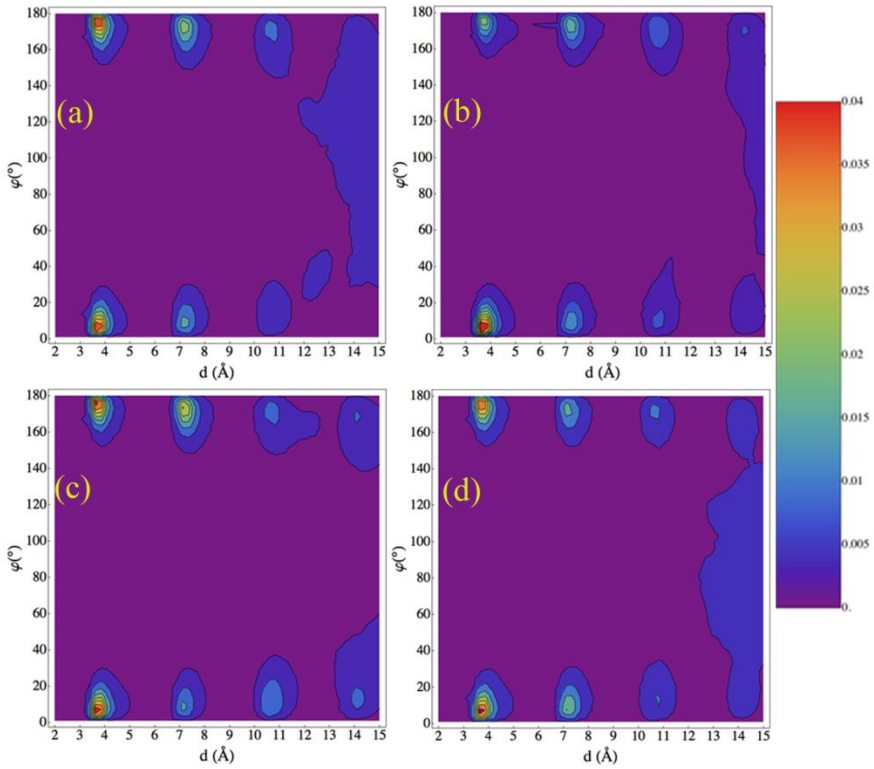


Figure 4.9. *Probability distribution of distance and orientational distribution of proflavine molecules with one another at (a) 0 M, (b) 0.15M (c) 0.5M, and (d) 1M ion concentration. The color bar shows values of the contours. Contours above 0.001 are drawn at 0.005 intervals.*

Once we captured the timescale for proflavine stacking around DNA, we probed further into the detailed structural arrangement of the outside-bound state. This pre-intercalative state remained elusive because of its metastable nature. Only recently, it was shown that low P/D ratio makes the outside-bound state stable.[9]The outside bound state is also crystallized using modified DNA[20]. It was proposed that stacking of proflavine would happen along the DNA axis because of the presence of phosphate groups, which would attract the cationic proflavine molecules.[12] To investigate the arrangement of proflavine molecules, we first calculated the probability distribution of mutual distance and orientation among them and plotted in Fig. 4.9. The signature of parallel and anti-parallel stacking is clear from the distribution at high and low angles at discrete separations of 3.4 Å. However, presence of DNA reduces the range of stacking probability compared to proflavine solution alone (Fig. 4.3.d).

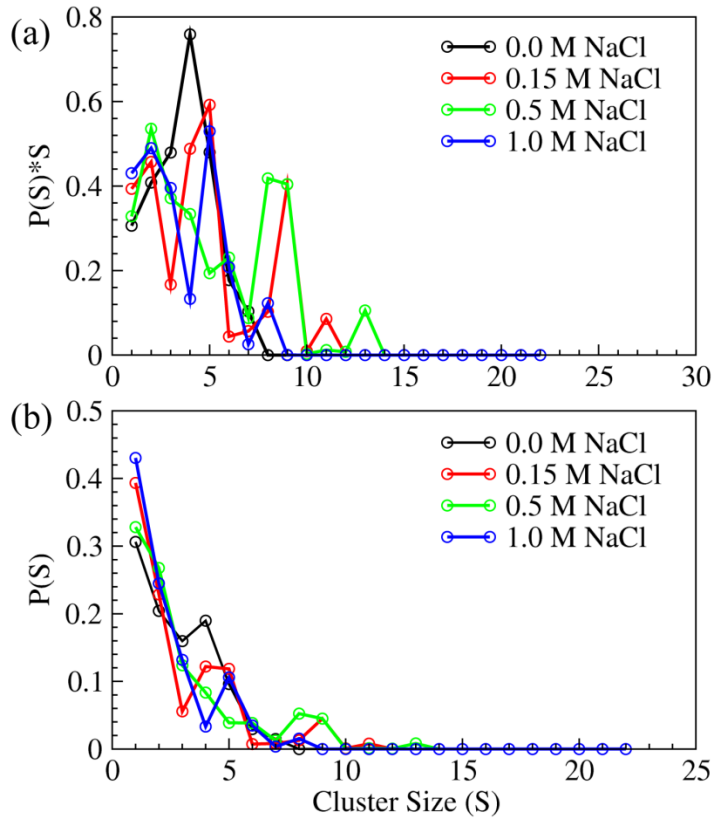


Figure 4.10. *(a) Probability of proflavine molecules to be in a cluster of a particular size. Note that although number of monomer is the highest, the number of molecules in the monomer is less. (b) The distribution of different sizes of clusters of proflavine around DNA.*

This is reinstated in Figure 4.10.a where we have plotted the probability of formation of a particular size of cluster, P(S), multiplied by the members in the cluster (S). This gives us the probability of formation of a stacked cluster of a particular size. This shows that proflavine molecules in presence of DNA typically form a stacked cluster involving either 4-5 molecules or 8-9 molecules. The distribution of different sizes of clusters of proflavine molecules around DNA is shown in Fig. 4.10.b.

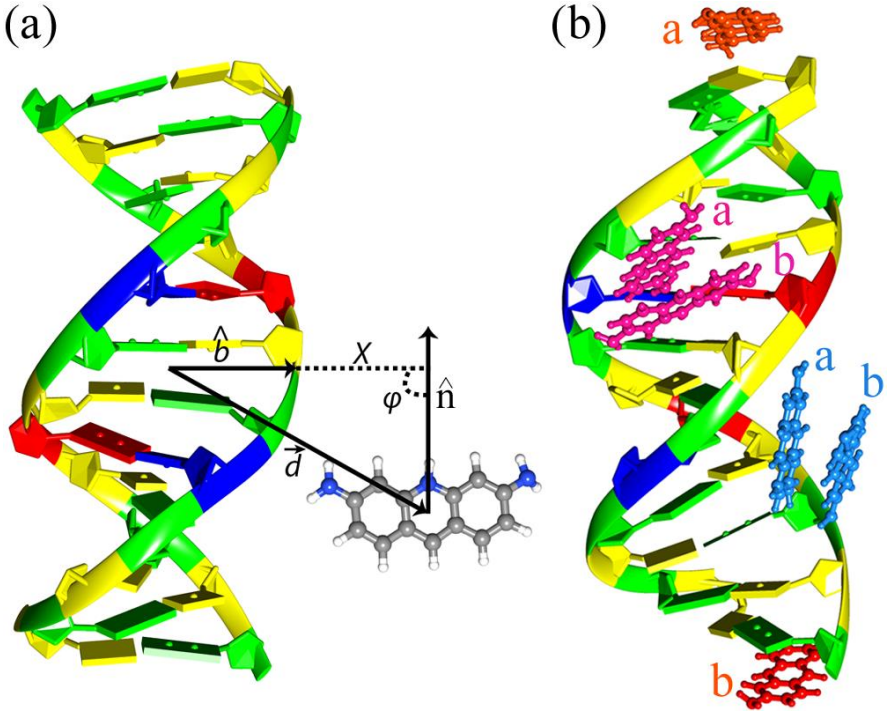


Figure 4.11. *(a) Schematic representation of coordinates X and φ used in proflavine orientation and (b) some representative structures.* $\hat{b}$ *represents the unit vector between COMs of 4 bases of DNA and two sugar units closer to the four bases.* $\vec{d}$ *represents the vector from COM of four bases to COM of proflavine molecule. The unit vector,* $\hat{n}$*, represents the vector normal to the plane of proflavine molecules. Here,* $X = \hat{b}.\vec{d}$ *and* φ *is the angle between the two vectors,* $\hat{b}$ *and* $\hat{n}$*. The values of X and* φ *for the representative structures shown in (b) are: Minor Groove(pink): a: X=5.4 Å,* $\varphi=105.4^{o}$*; b: X=5.5 Å,* $\varphi=100.7^{o}$*; Major Groove (blue): a: X=-3.2 Å,* $\varphi=43.9^{o}$*; b: X=-6.1 Å,* $\varphi=56.8^{o}$*; Ends of DNA (red): a: X=-5.6 Å,* $\varphi=91.4^{o}$*; b: X=-2.6 Å,* $\varphi=82.1^{o}$*.*

To understand the orientation of proflavine with respect to DNA helical axis, we calculated the angle between the body fixed vector $\hat{b}$ and the vector normal to the proflavine, $\hat{n}$ (Fig. 4.11.a). Some of the representative stacked proflavine around DNA along with corresponding values of X and φ are also shown in Fig. 4.11.b. Figure 4.12 shows the distribution of the stacked proflavine molecules around DNA with respect to

axis perpendicular to DNA helical axis at different ionic concentrations. X is positive towards the minor groove and negative towards the major groove. Also, the angle φ is 0° if the molecule parallel with DNA axis and ~90° if they are perpendicular to DNA axis. The assumption earlier was that the because of the interaction between the amino group of proflavine and phosphate groups of DNA, proflavine will stack perpendicular to DNA. Figure 4.12 shows that the proflavine distribution is concentrated within 10 Å from the

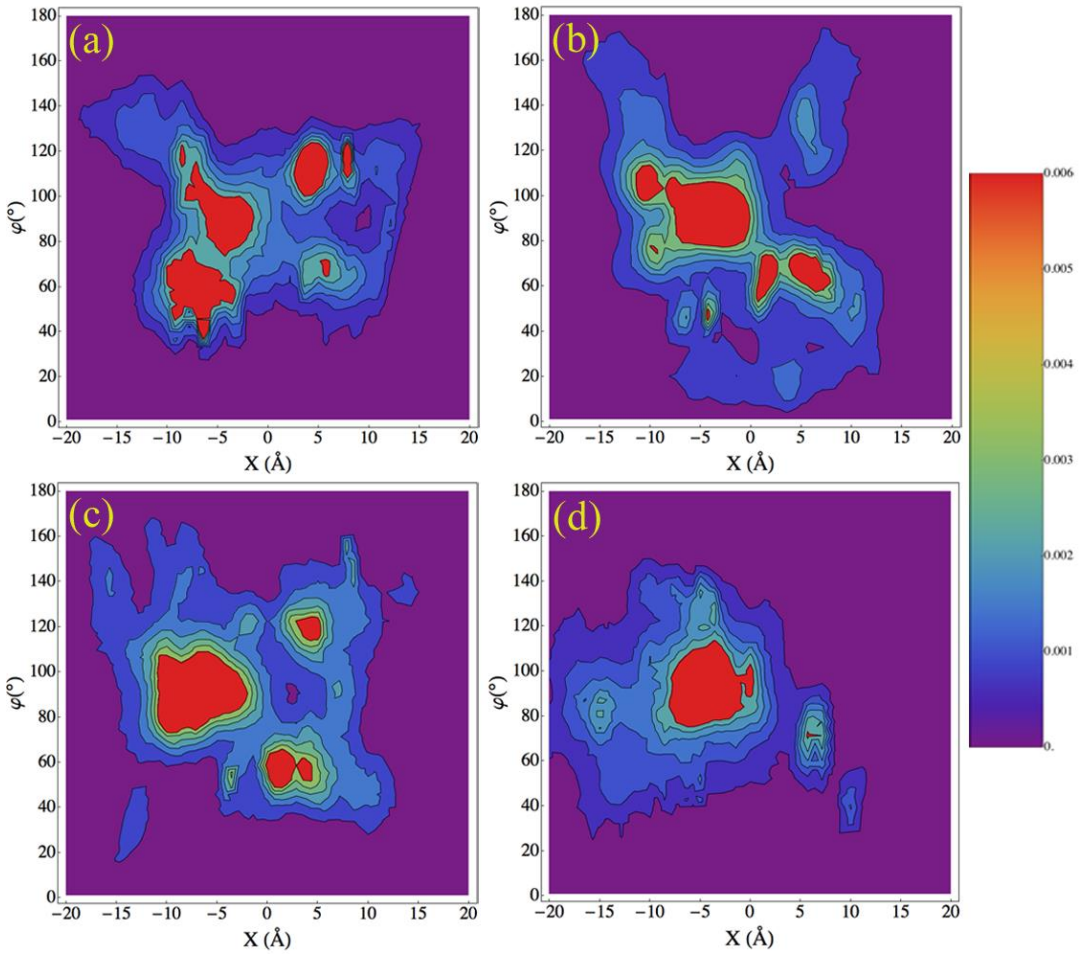


Figure 4.12. *Distribution of proflavine around DNA at a particular distance perpendicular to DNA axis and orientation at (a) 0 M, (b) 0.15M (c) 0.5M, and (d) 1M ion concentration. The color bar shows values of the contours. Contours above 0.001 are drawn at 0.005 intervals.*

DNA irrespective of ion concentrations used here. This indicates that ions do not prevent the association within the timescale (30ns) used in this study. Moreover, in all cases, the distribution is high in the negative values of X (in between -10 Å and 0 Å) for angle between 80° and 110° indicating a tilted association of proflavine in the major groove

side. A distribution to lesser extend is seen in the minor groove side at all ion concentrations represented by the positive X values between ranging 0 Å and 8 Å and angle between 50^0 and 70^0. Figure 4.11.b shows some examples of proflavine arrangement around DNA chosen from the regions with higher distribution as indicated by their φ and X values. Therefore, the driving force of proflavine association with DNA may not be only the electrostatic interaction between the proflavine and the phosphate backbone. The angular arrangement may help the proflavine molecules to gain more interaction with DNA despite getting the stacking interaction with the neighboring proflavine molecules.

4.4. Conclusion

In this study, we investigated the structure and timescale of proflavine binding to DNA, considered as the first step in the intercalation process, using molecular dynamics simulation. The results here indicate that proflavine associates in nanoseconds – six orders of magnitude faster than argued using experimental kinetic studies. Interestingly, the discrepancy is less for proflavine association in solution where experimental studies indicate a timescale between 30 to 700 ns.[8] The nanosecond timescale was also predicted from our previous kinetic study from the free energy of association between proflavine and DNA, albeit at lower drug concentration.[1] Therefore, we strongly believe that the timescale of association of proflavine around DNA also would be similar. The ion concentration can be thought of as a deterrent of association of proflavine around DNA. Therefore, we have performed our studies at four different ion concentrations, the highest being 1.0 M.

To validate our result, we have calculated the dimerization free energy of proflavine and found good agreement with experiment. Moreover, we found that the association of proflavine in solution is cooperative nature in agreement with experiment. The number of clusters matches well with the experimental result. Therefore, we believe that the force-field is reliable and the method adapted is reasonable to study the process.

We have shown that the association of proflavine both in solution and around the DNA happens through parallel or anti-parallel stacking. However, the number of stacks is less

for proflavine around DNA. From the cluster size distribution, we found that proflavine can form 4-5 member stack around DNA and sometime even 8-9 member stack.

Our previous study on the mechanism of proflavine intercalation showed that a direct intercalation from the major groove-bound state would result in microsecond timescale. However, due to the presence of a captive, more stable minor groove-bound state, the overall kinetics appears in the millisecond regime.[1] It is possible that experiments detected a different outside-bound state compared to what is observed here. However, intercalation of proflavine was shown to be a two-step process. [1, 3, 5-6] Even for daunomycine, the first pre-intercalative state was shown to be a minor groove-bound state.[41] Therefore, we believe that the submillisecond timescale observed in the experiment is another facet of intercalation itself, rather than outside binding. The new perspective obtained from the study may be useful to understand drug-DNA intercalation process and help to design new intercalating anti-cancer drugs.

4.5. References

1. Sasikala, W. D.; Mukherjee, A., *Phys. Chem. Chem. Phys.* **2013,** *15* (17), 6446-6455.
2. Sasikala, W. D.; Mukherjee, A., *J. of Phys. Chem. B* **2012,** *116* (40), 12208-12212.
3. Ramstein, J.; Ehrenberg, M.; Rigler, R., *Biochemistry* **1980,** *19* (17), 3938-3948.
4. Alberty, R. A.; Hammes, G. G., *J. of Phys. Chem.* **1958,** *62* (2), 154-159.
5. Li, H. J.; Crothers, D. M., *J. Mol. Biol.* **1969,** *39* (3), 461-&.
6. C. Ciatto, M. L. D. A., G. Natile, F. Secco, M. Venturini, *Biophys J.* **1999,** *77* (5), 2717-2724.
7. Corin, A. F.; Jovin, T. M., *Biochemistry* **1986,** *25* (14), 3995-4007.
8. Turner, D. H.; Flynn, G. W.; Lundberg, S. K.; Faller, L. D.; Sutin, N., *Nature* **1972,** *239* (5369), 215-217.
9. Bereznyak, E.; Gladkovskaya, N.; Khrebtova, A.; Dukhopelnikov, E.; Zinchenko, A., *Biophysics* **2009,** *54* (5), 574-580.
10. Tang, P.; Juang, C. L.; Harbison, G. S., *Science* **1990,** *249* (4964), 70-2.
11. Neidle, S.; Jones, T. A., *Nature* **1975,** *253* (5489), 284-285.

12. Peacocke, A. R.; Skerrett, J. N. H., *Trans. Faraday Soc.* **1956,** *52* (0), 261-279.
13. Albert, A., *The acridines: their preparation, physical, chemical, and biological properties and uses*. E. Arnold: 1951.
14. A. Blake, A. R. P., *Nature* **1965,** *206*, 1009-1011.
15. Dourlent, M.; Hélène, C., *Eur. J. Biochem.* **1971,** *23* (1), 86-95.
16. Araki T, Y. M., *Histochem.* **1985,** *83* (4), 299-301.
17. Bradley, D. F.; Wolf, M. K., *Proc. Natl. Aca. Sci.* **1959,** *45* (7), 944-952.
18. Stone, A. L.; Bradley, D. F., *J. Am. Chem. Soc.* **1961,** *83* (17), 3627-3634.
19. Blears, D. J.; Danyluk, S. S., *J. Am. Chem. Soc.* **1967,** *89* (1), 21-26.
20. Westhof, E.; Hosur, M. V.; Sundaralingam, M., *Biochemistry* **1988,** *27* (15), 5742-5747.
21. Schwarz, G.; Klose, S.; Balthasar, W.,Eur. J. Biochem. **1970,** *12* (3), 454-460.
22. Haugen, G. R.; Melhuish, W. H., *Trans. Faraday Soc.* **1964,** *60* (0), 386-394.
23. Obendorf, S. K.; Carrell, H. L.; Glusker, J. P., *Acta Crystallo. Sec. B* **1974,** *30* (6), 1408-1411.
24. Shiao, D. D. F.; Sturtevant, J. M., *Biochemistry* **1969,** *8* (12), 4910-4917.
25. Schwarz, G.; Klose, S.,Eur. J. Biochem. **1972,** *29* (2), 249-256.
26. Macke, T.; Case., D. A., *Modeling unusual nucleic acid structures*. N.B. Leontes J. SantaLu- cia, Jr. ed.; American Chemical Society: Washington, DC, 1998.
27. Mark, P.; Nilsson, L., *J. Phys. Chem. A* **2001,** *105*, 9954-9960.
28. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., *J. Com. Chem.* **2004,** *25*, 1157-1174.
29. Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M., *Biophys J* **2007,** *92* (11), 3817-29.
30. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., *J. Chem. Theor. Comput.* **2008,** *4* (3), 435-447.
31. Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T., *Numerical Recipes in FORTRAN: The Art of Scientific Computing* 2nd ed ed.; Cambridge University Press: Cambridge, England, 1992.
32. Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; DiNola, A.; Haak, J. R., *J. Chem. Phys.* **1984,** *81* (8), 3684-3690.

33. Evans, D. J.; Holian, B. L., *J. Chem. Phys.* **1985,** *83* (8), 4069-4074.

34. Parrinello, M.; Rahman, A., *J. Appl. Phys.***1981,** *52* (12), 7182-7190.

35. Tom Darden, D. Y., Lee Pedersen, *J. Chem. Phys.* **1993,** *98* (12), 10089-10092.

36. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., *J. Comput. Chem.* **2009,** *30* (16), 2785-2791.

37. Torrie, G. M.; Valleau, J. P., *J. Comp. Phys.* **1977,** *23*, 187-199.

38. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M., *J. Comp. Chem.* **1992,** *13*, 1011-1021.

39. Li, H. J.; Crothers, D. M., *J. Mol. Biol.* **1969,** *39* (3), 461-477.

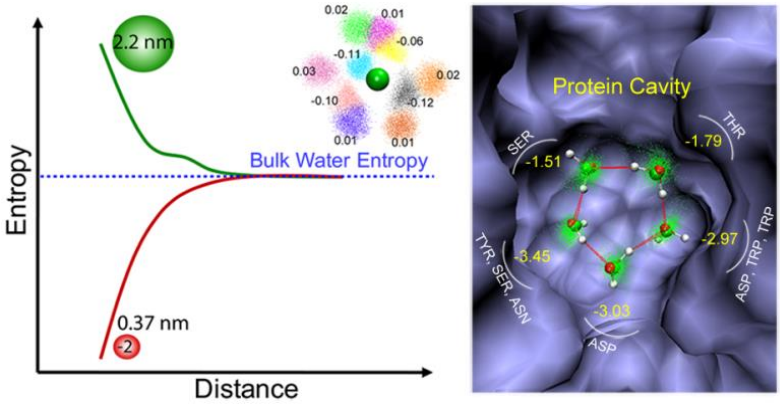
40. Ramstein, J.; Leng, M., *Biophys. Chem.* **1975,** *3* (3), 234-240.

41. Wilhelm, M.; Mukherjee, A.; Bouvier, B.; Zakrzewska, K.; Hynes, J. T.; Lavery, R., *J. Am. Chem. Soc.* **2012,** *134* (20), 8588-8596.

Chapter 5

Water Entropy: Hydrophobic Crossover and Application to Drug Binding

Chapter 5: Single Water Entropy: Hydrophobic Crossover and Application to Drug Binding



Entropy of water plays an important role in both chemical and biological processes e.g. hydrophobic effect, molecular recognition etc. Here we used a new approach to calculate translational and rotational entropy of the individual water molecules around different hydrophobic and charged solutes. We showed that for small hydrophobic solutes, the translational and rotational entropies of each water molecule increase as a function of its distance from the solute reaching finally to a constant bulk value. As the size of the solute increases (0.746 nm), the behaviour of the translational entropy is opposite; water molecules closest to the solute have higher entropy that reduces with distance from the solute. This indicates that there is a crossover in translational entropy of water molecules around hydrophobic solutes from negative to positive values as the size of the solute is increased. Rotational entropy of water molecules around hydrophobic solutes for all sizes increases with distance from the solute, indicating the absence of crossover in rotational entropy. This makes the crossover in total entropy (translation + rotation) of water molecule happen at much larger size (>1.5 nm) for hydrophobic solutes. Translational entropy of single water molecule scales logarithmically ($S_{tr}^{QH} = C + k_B \ln V$), with the volume V obtained from the ellipsoid of inertia. We further discuss the origin of higher entropy of water around water and show the possibility of recovering the entropy loss of some hypothetical solutes. The results obtained are helpful to understand water entropy behavior around various hydrophobic and charged environments within biomolecules. Finally, we show how our approach can be used to calculate the entropy of the individual water molecules in a protein cavity that may be replaced during ligand binding. This method can also be used to study the role of water molecule in DNA-drug intercalation process.

5.1. Introduction

5.1.1. Role of Water in Molecular Recognition Process

Water is considered to be a 'biomolecule' and it plays a pivotal role in life as it constitutes 70% of human body[1]. Water present around different bio-macromolecules modulates its structure and function during molecular recognition processes[1-3]. Water surrounding different bio-macromolecule can be classified into three categories: (i) tightly bound water (ii) surface water and, (iii) bulk water (See Figure 5.1).

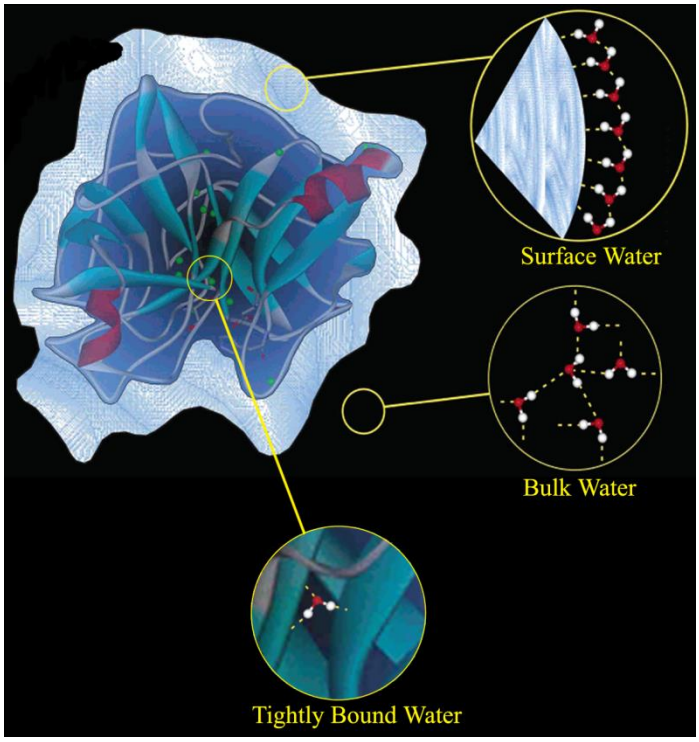


Figure 5.1. *Different kinds of water molecule around a bio-macromolecule. Tightly bound water molecule forms multiple hydrogen bonding interactions with groups present in the core of the molecule. The surface water molecules form the network of water molecules through hydrogen bonding interactions among itself and with the surface of the bio-macromolecule. The bulk water molecules form loose network of water molecules through intermolecular hydrogen bonding. (Figure adapted and modified from Chem. Rev.* ***2004,*** *104, 2099-2123)*

Though most water molecules present in and outside of bio-macromolecules are movable and displaceable, a few of them are tightly bound. These tightly bound water molecules play an imperative role in modulating the conformation and biological activity of bio-macromolecule. Hence, they are considered to be the integral part of the structure

of the bio-macromolecule.[4-5] One such example is the presence of tightly bound water molecule in the form of 'spine of hydration' (as shown in Fig. 5.2) in the minor groove DNA which give shape and stability to B-form of the duplex[6]. Likewise, several enzymes and proteins also have deeply bound water molecules in the core assisting in the stability of the enzymes[7]. For example, the enzyme pancreatic trypsin inhibitor (PTI), which inhibits trypsin holds four such internal water molecules.[8] Also human lysozyme contains deeply buried internal water molecules arranged in the form of semicircle through hydrogen bonding interactions.[9] Moreover, in many cases, tightly bound water molecule conserved in the core of the protein act as a bridge between protein and ligand, thereby facilitating the molecular recognition process. One such example is the presence of bridging water molecule between the protein, concanavalin A and the ligand, trimannoside with several hydrogen bonds.[10] Figure 1.2 shows some examples of tightly bound water molecules discussed above.

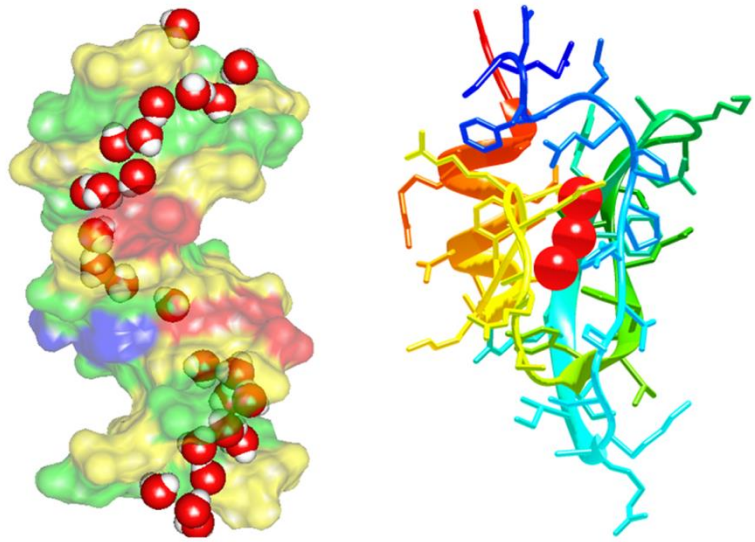


Figure 5.2. *Tightly bound water molecules in the minor groove of DNA and the core of the protein. The red spheres represent water molecules.*

A critical peripheral hydration level is necessary to establish the molecular conformation and dynamics of bio-macromolecules. These water molecules are referred to as peripheral water molecules and usually form hydrogen bonding interactions with bio-macromolecules.[8] Thus, they form network of water molecules which helps in linking the polar groups present on the surface of the bio-macromolecule. In case of proteins, it is

proposed that these peripheral water molecules can form clathrate-type structures at surfaces[11]. For example, pentagonal arrays of water clusters are seen on the surface of a protein, crambin, covering the hydrophobic region of the protein[12]. Moreover, the peripheral water molecules also assist in molecular recognition between two interfaces such as protein-protein or protein-nucleic acid interaction. These interfacial water molecules also contribute to the stability and specificity to such type of interactions. For example, three interfacial water molecules present in the *trp* repressor-DNA complex handle the specific affinity of repressor to the operator sequence of the DNA[6]. Similarly, an appropriate example for the protein-protein interaction is barnase-barstar complex. Here, the interfacial water molecules present between two proteins stabilize the interaction by forming adhesive hydrogen-bond network between the interfaces.[13] Figure 5.3 shows different systems having surface and interfacial water molecules.

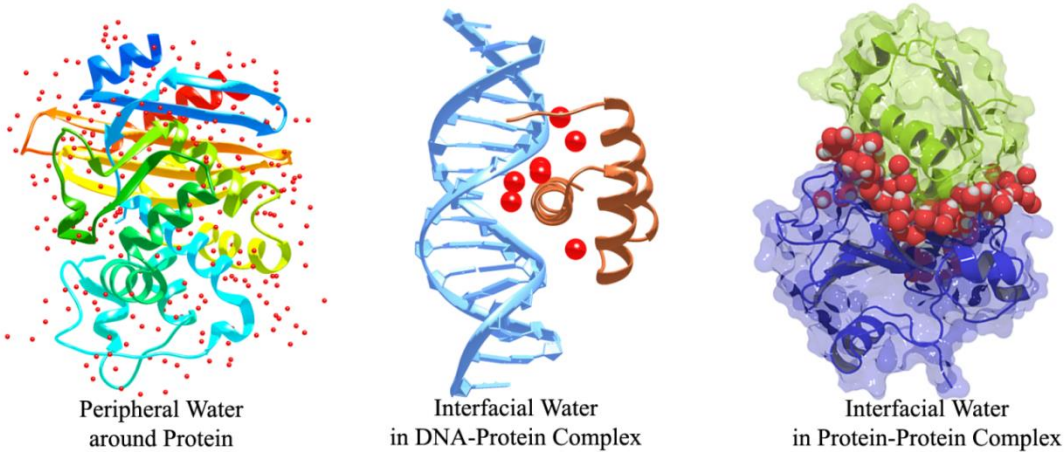


Figure 5.3. *Shows different molecular recognition processes with surface and interfacial water molecules. Red spheres represent water molecules.*

5.1.2. Hydrophobic Effect and Importance of Water Entropy

In biology, during molecular recognition process, usually small ligands interact with bio-macromolecular cavities or interfaces constituting ~75% non-polar regions[14]. These non-polar regions are occupied by several water molecules which could not form hydrogen bonding interactions with the hydrophobic cavities. Thus, the water molecules re-order themselves to form intermolecular hydrogen bonding and form highly ordered hydrogen bonded networks as shown in Fig. 5.4[14-15]. This leads to reduction in the mobility of water molecules in the cages leading to free energetically less stable state due

to the reduction in translational and rotational entropy of water molecules[16]. But during the molecular recognition process, most of the water molecules present in and around the bio-macromolecule is displaced to the bulk[16]. A positive entropy change and, hence favourable free energy contribution, can arise from highly structured water molecules in the binding pocket that are released to a less ordered state in the bulk solvent upon binding known as 'hydrophobic effect' (HE).[3, 17]

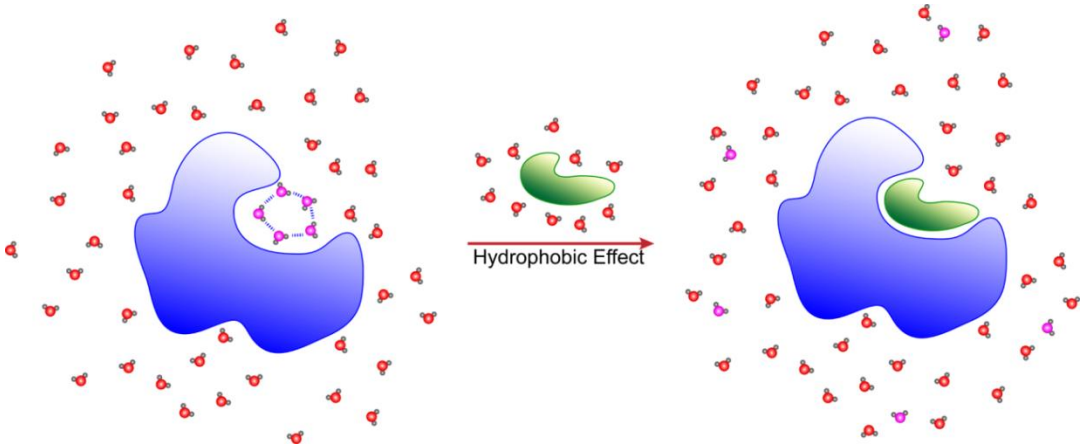


Figure 5.4. *Schematic representation showing hydrophobic effect. Blue colored shape represents protein and green colored shape indicates ligand. The highly structured water network in the protein cavity is shown as water molecules (magenta colored oxygen) arranged in pentagonal shape formed by intermolecular hydrogen bonding. During protein-ligand interaction, these water molecules are displaced into bulk (surface water and bulk water are represented by red oxygen atoms).*

The binding between molecules is usually accompanied by both the displacement of bound water molecules from the binding sites and the formation of direct interactions between the molecules. Whether such interactions are favourable, neutral, or unfavourable to binding affinity is determined by the balance between the direct interactions gained and the solvation interactions lost.[18] This clearly suggests that not only the binding interactions between the macromolecule-ligand or between interfaces are important, but the role of water is very crucial and hence the entropy of water molecule because the net contribution of the entropy of expelled highly ordered water molecules in the overall free energy of binding process is higher as compared to less ordered water molecules.[19] This hydrophobic effect imparted by water molecules play a crucial role in many molecular recognition processes and considered to be one of the driving forces for protein folding. Proteins are made up of both polar and non-polar

amino acids. In case of un-folded proteins, both polar and non-polar amino acids are exposed to water (Fig.5.5). The water molecules closer to non-polar amino acids are not able to form hydrogen bonding interactions with non-polar amino acids. But these water molecules maintain the number of hydrogen bonds by reordering to form intermolecular hydrogen bonding among water molecules leading to cage like water structures. These water molecules have limited movement leading to decrease in entropy of water molecules. This leads to folding of proteins in such a way that the core of the protein constitutes non-polar residues which are not exposed to water molecules and the polar amino acids are exposed to water molecules (Fig.5.5). During the folding process, water molecules having low entropy are displaced to the bulk where the entropy of water molecules is increased as compared to the cage-like structures.[20]

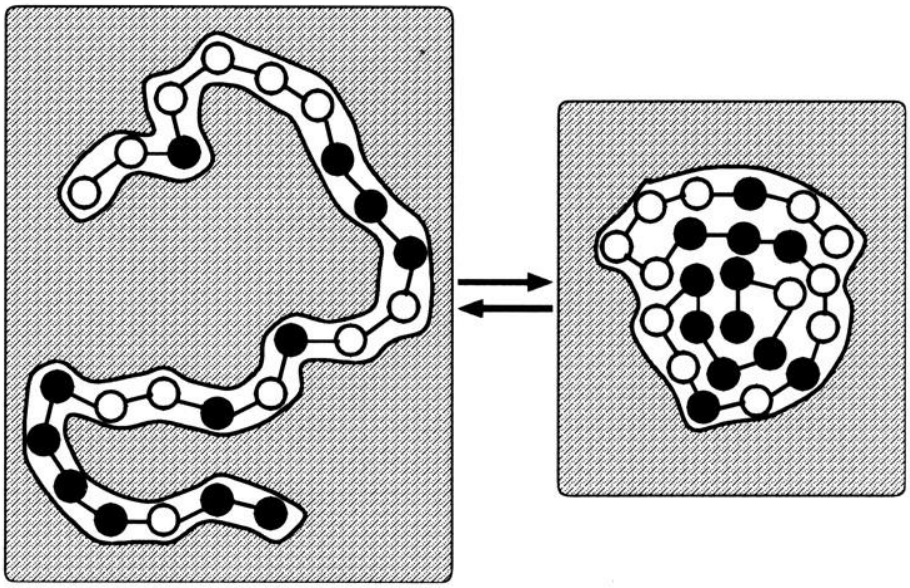


Figure 5.5. *Schematic diagram showing protein folding. White and black spheres represent polar and non-polar amino acids of the polypeptide chain respectively. The dashed region represents the solvent. Left panel represents un-folded protein and right panel represents folded protein (Figure adapted from PNAS, 97, 15, 2000, 8324-8327).*

Similarly, entropy of water molecules plays a crucial role in many biological processes involving small molecules. Here, hydrophobic effect depends on the size and nature of the solute. For small solutes like methane at room temperature, larger negative entropy change over smaller favourable enthalpy leads to this phenomenon.[21] This entropy loss for water around a small hydrophobic solute like methane is argued to

originate from the tendency of the solvent to maintain necessary hydrogen bonds around the solute[22-23], leading to enhanced order[24-25].

Clathrate-like water arrangement[26] around small solutes interestingly changes to loose dangling hydrogen-bonded arrangement around bigger hydrophobic particles where water regains entropy at the cost of losing enthalpic stability.[22-23] Recently, the existence of dangling hydrogen-bonds around bigger hydrophobic solutes was shown experimentally[27] and was subsequently shown to be cooperative.[28] This "two faces of water"[29] has been the topic of intense research for several decades, primarily because of the fundamental nature of its application in protein folding[1], aggregation[30] etc.

The origin of entropy loss of water molecules around small solutes has often been debated.[31-32] It is argued that entropy loss is caused by the cohesive forces between solvent particles that get perturbed by small solutes[26, 33]. Some other theoretical studies argue that the small size of the solvent (water) is responsible for the HE because smaller size requires higher entropy cost to open the cavity for a solute insertion[34], or it could be a competition between both the above[35]. The iceberg model[26] hypothesized an increase in tetrahedral order around methane-like small molecules, substantiated recently using simulation[24] and experiment[25]. Binding energy of water and water-water hydrogen-bond (H-bond) data showed, on the other hand, that the structure of the H-bond is maintained, rather than enhanced[36-38]. Recently, long range (up to the second hydration shell) correlation, rather than the accessible volume[9,21] was argued to be responsible for the entropy loss of water around hydrophobic solutes.[39]

The change of "face" from entropy to enthalpy dominated HE is predicted to happen around solute radius 1 nm.[23, 40-41] This theoretical prediction based on information theory[42] has been captured through structural[25, 41, 43]and dynamical[44-45] aspects. However, demonstration of this crossover directly through water entropy calculations has not been done. Only recently, Goddard and co-workers calculated water entropy using 2PT (dividing the liquid density of states into solid and gas using a fluidicity factor) decomposition[46] to show that for a hydrophobic particle and mildly attractive one as methane, the crossover from negative to positive entropy occurs around 0.6 nm and 0.9 nm, respectively.[47] However, the above study focussed only on the translational contribution of water molecules in the hydration shell with the objective of showing the

transition from a hydrophobic to hydrophilic character of a solute at a small charge 0.4e.[47]

5.1.3 Calculation of Entropy

The entropy change of water around a solute has its contribution from every single water molecule possessing both translational and rotational entropy (with rigid approximation). The vibrational entropy of water molecule is considered to be negligible compared to other two components of entropy. Estimations of both translational and rotational entropies are widely debated.[48] In computational methods, for studying molecular recognition processes, either this entropy contribution is neglected[2, 49] or the expression for an ideal gas[50-53] is used. Translational entropy of a monatomic ideal gas, known as the Sackur-Tetrode (ST) equation[54], is expressed as $S_{tr}^{id} = C(T) + k_B \ln V$, where C is the momentum contribution and V is the configurational volume. The momentum contribution is integrable and unambiguous. However, the configurational volume has been a source of debate. One often uses the volume of a gas (22.4 litres at standard temperature and pressure) as the configurational volume. While this is applicable for molecules in the gas phase, this volume is a gross overestimation in the liquid environment. Therefore, the ST approach estimates very high translational entropy (12-15 kcal/mol)[55-56]. Alternate force-based method known as cratic entropy[57] uses an equation similar to ST, where configurational volume is based on concentration. However, cratic entropy is highly debated[58]. Similarly, rotational entropy for an ideal gas depends only on the moment of inertia of the molecule. However, in an interacting system, both translational and rotational entropy should depend on the environment. Water entropy in a realistic system, therefore, is difficult to compute. Amongst the few existing approaches, Lazaridis and Paulaitis proposed the inhomogeneous solvation theory (IST) providing reasonable estimates for the change in the translational and rotational entropy of water in the solute hydration shell.[59] An alternate approach of water entropy calculation, 2PT decomposition mentioned above, was subsequently proposed by Goddard *et al.*[60] Recently, Stanley *et al*. proposed tetrahedral entropy of water[61].

Since the overall water entropy must have contributions from the individual water molecules constituting the solvation shell, it is important to address the entropy of every

single water molecule in the solvation shell, or in a biomolecular cavity. It is of particular interest to estimate entropy of each water molecule for drug design[62] as even a single water molecule can change the overall free energy of a molecular recognition process. Studies on various protein-ligand complexes show the presence of highly localized anchoring water molecules between the protein surface and ligands.[18, 63-65] Ligand modifications are done to displace the low entropy water molecules from the binding sites as it favours the binding process of ligand[3, 66]. Entropic behavior of an individual water molecule, however, was not addressed until recently where Friesner *et al.* calculated translational and rotational entropy of single water molecules in protein cavities and interfaces by applying IST[59] on configurations obtained from molecular dynamic simulations.[19] This method has been incorporated as the "WaterMap" in the docking program called GlideXP1 and is used for calculating water molecule entropy in various systems[67-70]. This entropy correction was shown to improve binding energy prediction for various systems[19].

5.1.4. Outline of Work

Here we propose an alternate method to calculate single water translational and rotational entropy, and thereby probe into the general behavior of the individual water molecules around different types of solute environments (charged, hydrophobic, attractive etc.) found in complicated non-polar and polar protein cavities or interfaces. We calculated translational and rotational entropy of every water molecule as a function of its distance from hydrophobic solutes of different size and various charged particles. Finally, we took an example of ligand binding to a protein cavity to show the extent of entropy loss by the water molecules that occupy the particular cavity in the absence of ligand. This in turn indicates the favorable free energy contribution from the water entropy for this particular ligand binding process.

5.2. Experimental Section

5.2.1. Simulation Details

United atom model of methane was selected for calculating water entropy around methane using molecular dynamics simulations. All the simulations were carried out in

GROMACS[71]. Methane was considered as Lennard-Jones particle with σ = 0.373 nm and ε = 0.294 kcal/mol[72]. TIP4P water model was selected with OPLS force field[73]. Six different sets of simulations of hydrophobic methane particles were carried out by changing the diameter of the particles from σ to 6σwhile keeping ε at 0.294 kcal/mol. For the hydrophobic solutes, attractive potential is not used in the simulations. Furthermore, simulations were performed involving methane with +1, -1, +2, -2 charges and hydrophobic methane with charge -0.5. For each simulation, the solute is solvated with TIP4P water molecules in a cubic box where 1000 to 4000 water molecules were added to solvate the solute depending on the size of the solute to ensure that at least three layers of water molecules are present around the solute. For the simulations with charge on methane, counter ions (Na+ or Cl-) were added in the system to neutralize the system charge. Each system was energy minimized using steepest descent algorithm[74]. Further equilibration was carried out for 2 ns at constant 25°C temperature and constant 1 bar pressure (NPT) using Nose-Hoover thermostat[74] and Parrinello-Rahman barostat[74], respectively. After NPT equilibration, final 100 ns simulation was performed at constant 25°C temperature and 1 bar pressure using Nose-Hoover thermostat and Parrinello-Rahman barostat, respectively using0.4 ps coupling constants for each. PME (Particle Mesh Ewald)[74] was used for electrostatic interaction and a cut off not less than 1.2 nm was used for both electrostatic and van der Waals interactions. From each trajectory obtained after simulation, a sphere of 500 to 1000 water molecules around the solute was cut out depending upon the diameter of the solute. Subsequently, Permutation Reduction (PR)[75] was applied on each system prior to the calculation of translational and rotational entropies of the individual water molecules around the solute.

The monomer unit of streptavidin (PDB ID: 1STP[76]) was used for the protein system simulation after removing biotin from the binding site and water molecules present in the crystal structure. Streptavidin was put in a cubic box (7.3 nm x 7.3 nm x 7.3 nm) and solvated with TIP4P water molecules. Two sodium ions were added to neutralize the system. OPLS force field[73] was used for the protein, water and ions. The energy minimization was carried out using steepest descent algorithm[77]. The system was heated up to 298 K using Berendsen thermostat[78] with 25 kcal/mol restraints on the heavy atoms of the protein. Series of equilibration and energy minimization were performed on

the system where the restraint was reduced from 25 kcal/mol/Å^2 to 5 kcal/mol/Å^2. Final equilibration for 2 ns followed by 100ns production run was performed with restraints on the heavy atoms of the protein at constant 1 bar pressure and 298 K temperature using Parrinello−Rahman barostat[79] and Nose-Hoover thermostat[80-81], respectively. All the simulations were carried out using GROMACS[71] molecular dynamic simulation software. We have used ANN program of Mount et al[82] to calculate the k^{th} nearest neighbor entropy and thereby estimated the anharmonic contributions.

The stepwise protocol is mentioned below:

(i) We chose a reference configuration. In this study, it was the first frame after a 2ns equilibration.

(ii) This was followed by centering the solute and removing its translational and rotational motion. Later is required only for non-spherical solute. This will ensure a correct configuration of water around a particular site of a solute.

(iii) Water molecules were ordered with respect to the distance from the solute. This step helps in reducing the number of water molecules to be permuted.

(iv) Only the required number of water molecules around the solute is retained. We kept around 500 to 1000 water molecules depending on the RDF.

(v) Permutation reduction was performed for individual water molecules separately using only oxygen-oxygen distance.

(vi) RMSDs of permuted water molecules were calculated to check consistency of permutation over the whole trajectory.

(vii) Entropy calculations were performed using Eq. 2 and 3.

Convergence of entropy with time was checked for each water molecule.

5.2.2. Methods

Difficulty in water entropy calculations arises from the fact that water is diffusive. Therefore, standard methods based on quasi-harmonic approach[83] used in biomolecular entropy calculation cannot be applied here. A recent approach of Grubmuller and co-workers, known as permutation reduction (PR)[75, 84], overcomes this difficulty. PR uses linear assignment approach to keep the identity of each water molecule based on a reference configuration (or a particular site). Fig. 5.6 represents the positions of single

water oxygen before and after permutation reduction. Permutation reduction makes the permutated water molecules move only in a localized region as shown in Fig.5.6b. Fig. 5.7 represents the positions of several water oxygens around methane after permutation reduction. As apparent from Fig.5.7.a, each permuted water molecule occupies different configurational volume with minimal overlap (Fig. 5.7.b), and, therefore, may possess different entropy.

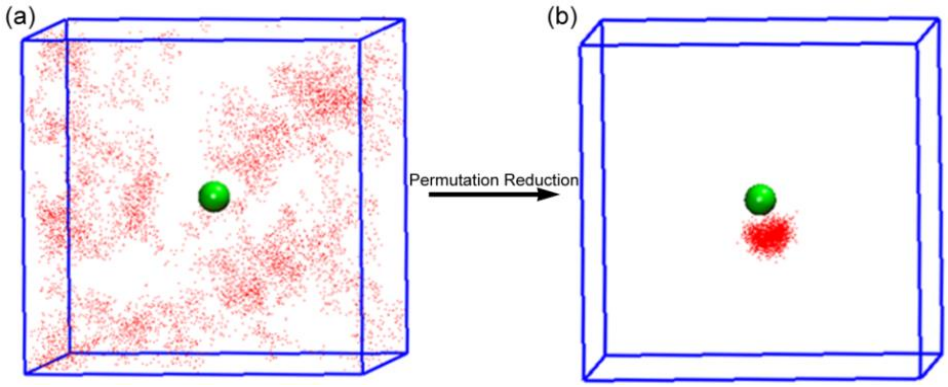


Figure 5.6. *Positions of water oxygen around methane (a) before permutation reduction and (b) after permutation reduction.*

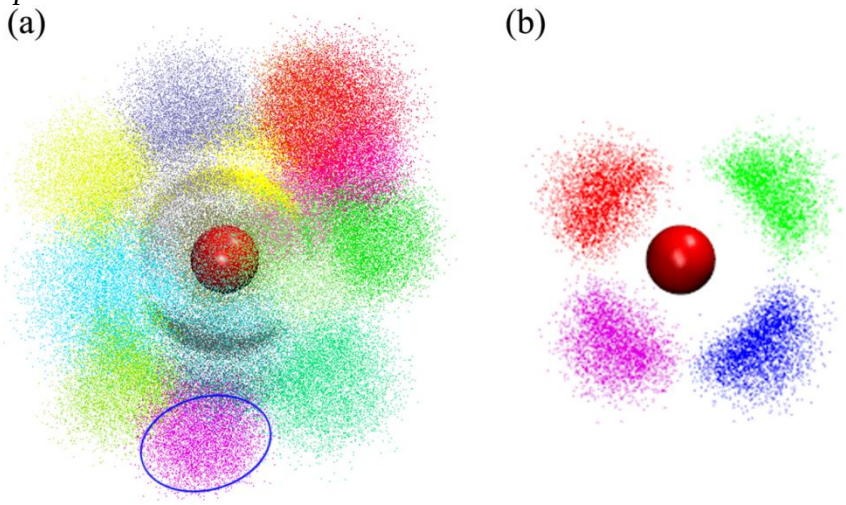


Figure 5.7. (a) *Configurations of water oxygens around methane (10000 frames) after permutation reduction. Each water molecule is colored differently. Ellipsoid is a schematic guide to the region occupied by one particular water molecule. (b) Configurations of four closest water oxygen atoms around methane (3333 frames) after permutation reduction. Fifth water molecule is behind methane and not shown for clarity. Each water molecule is colored differently. Each ellipsoid shows the region occupied by one particular water molecule indicating minimal overlap of ellipsoids formed by adjacent water molecules.*

We applied quasi-harmonic method on the translational motions of these permuted water molecules to calculate their translational entropy individually. For that, we created the covariance matrix of translational fluctuation along X, Y, and Z direction giving rise to 3X3 matrix. For an individual molecule, diagonalization of the above matrix provides three eigenvalues $(\lambda_i, i = 1,2,3)$ and thereby three frequencies, $\omega_i = \sqrt{k_B T/\lambda_i}$ $(i = 1,2,3)$. With these frequencies, we estimated the translational entropy using the following expression of entropy of solid state quantum harmonic oscillator[85].

$$S_{tr}^{QH} = k_B \sum_{i=1}^{3} \frac{\hbar\omega_i/k_B T}{e^{\hbar\omega_i/k_B T} - 1} - \ln\left(1 - e^{-\hbar\omega_i/k_B T}\right), \qquad (1)$$

where $\hbar = h/2\pi$, h is Planck's constant, T is temperature, and k_B is the Boltzmann's constant. Note that, calculation of the above frequencies depends on the simulation time. Therefore, we simulated each system long enough (100 ns) to ensure convergence in the entropy estimate (see Fig. 5.8) for each water molecule. Fig. 5.8 shows the convergence of translational entropy for first five closest permutated water molecules around methane. The convergence is obtained at around 75 ns for all water molecules. Permutation of each water molecule was carried out separately using our asymmetric permutation[53] approach.

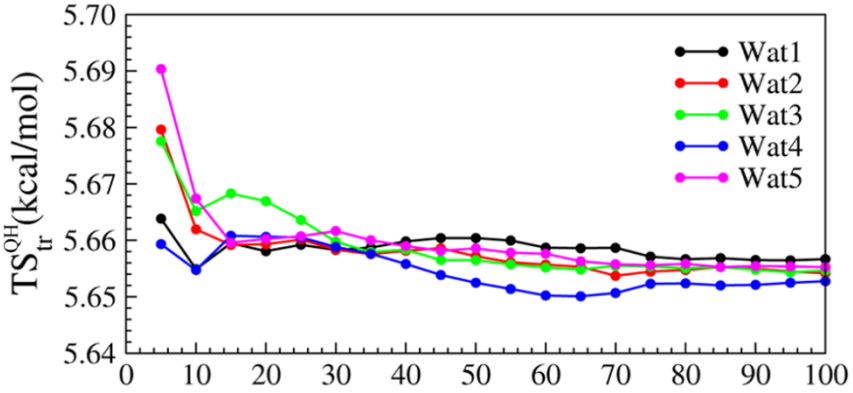


Figure 5.8. *Convergence of translational entropies of five nearest water molecules around the solute methane (σ) with respect to the first frame. The average distance of each water molecule (from Wat1 to Wat5) from COM of solute is 0.410±0.060 nm, 0.419±0.061 nm, 0.413±0.061 nm, 0.417±0.059 nm, and 0.411±0.061 nm respectively.*

Although vibrational formula (Eq. 1) was used to estimate the individual water entropy, it nonetheless provides the translational entropy in the present approach since the frequency here corresponds to the center of mass displacement of the entire molecule.

The above formula in the classical limit $\hbar\omega/k_BT < 1$ can be written as (after series expansion),

$$S_{tr}^{QH} = k_B \sum_{i=1}^{3} 1 - \ln(\hbar\omega_i/k_BT) = k_B[3 - \ln(C(T)/V)], \qquad (2)$$

, where,

$C(T) = \left(\sqrt{2}\pi/3\right)(\hbar^2/\mathrm{k_B T})^{3/2}$ and $V = 4\pi/3\left(1/2\sqrt{2}\right)\sqrt{\lambda_1\lambda_2\lambda_3}$. $\sqrt{\lambda_\mathrm{i}}$ is the length of the principle axes of the ellipsoid. C(T) is a constant at a given temperature and found to be 5.79 kcal/mol(multiplied by T=298 K). Volume, however, varies for every water molecule in every system. Eq. 2 resembles the ST equation for translational entropy of monatomic ideal gases. Depending on the strength of interaction of water molecules with different solutes, the contribution of entropy due to anharmonicity increases[75]. To calculate the anharmonic (non-Gaussian) contribution, we have used a method based on k^{th} nearest neighbor (NN) distances, where k<(n-1)(n=sample points)[86], following Numata *et al.*[87] Similar anharmonic corrections were incorporated recently in the quasi-harmonic entropy calculations of drug, DNA, and water in the intercalation process[53].

However, we used the angular distribution of the individual permuted water molecules to calculate the rotational entropy of a single water molecule as shown below.

$$S_{rot} = -k_B \int p(\theta,\chi)c \ln\{p(\theta,\chi)c\} \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\chi, \qquad (3)$$

where, $p(\theta,\chi)$is the angular distribution of a single water molecule and $c$is a normalization factor. θ is the angle between the dipole vector and the water oxygen to solute vector. χ is the angle between H-H vector of the water and the normal to the plane defined by the solute-oxygen vector and the dipole vector (plotted in Fig. 5.9.a). The above method is similar to that of IST. However, the primary difference is that it is applied here to the individual water molecules over the whole trajectory compared to the solvation shell as practised in IST. As discussed by Lazaridis *et al.*, the third angle is not important.[88] We calculated the angular distribution function $p(\theta,\chi)$ for every permuted water molecule and thereby calculated the rotational entropy of them as discussed later. The convergence for rotational entropy is shown in Fig. 5.9.b for the first five closest permuted water molecules. Here, convergence happened around 40 ns.

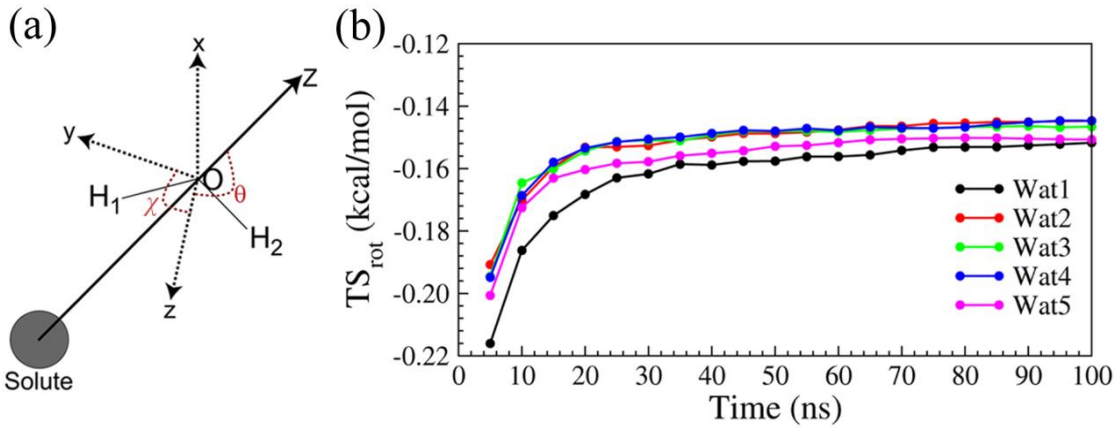


Figure 5.9. *(a) Schematic diagram showing the two angles θ and χ. θ is the angle between the dipole vector and the vector from water's oxygen to the solute. χ is the angle between the H-H vector of the water and the normal to the plane defined by the solute-oxygen vector and the dipole vector. (b) Convergence of rotational entropies of five nearest water molecules around the solute methane (σ) with respect to the first frame. The average distance of each water molecule (from Wat1 to Wat5) from COM of solute is 0.410±0.060 nm, 0.419±0.061 nm, 0.413±0.061 nm, 0.417±0.059 nm, and 0.411±0.061 nm respectively.*

Reason for not using quasi-harmonic entropy for rotational motion is that the moment of inertia associated with the angular motion of the water molecule relative to a solute is ambiguous. Moreover, it is inappropriate to use harmonic approximation for a freely rotating molecule.[89]

5.3. Results and Discussions

We have applied our method to study the water entropy and associated properties on two sets of systems. The first set consists of hydrophobic methane-like molecules (MHF) that vary only in size, but contain the same interaction potential. MHF contains only the repulsive part of the van der Waals (vdw) potential (see Simulation Detail). The second set of systems contains methane-like particles of the same size, however, containing different amount of charges. We will describe the water structure around these systems first before going to their role in the water entropy.

5.3.1. Water Structure and Inhomogeneous Solvation Theory (IST)

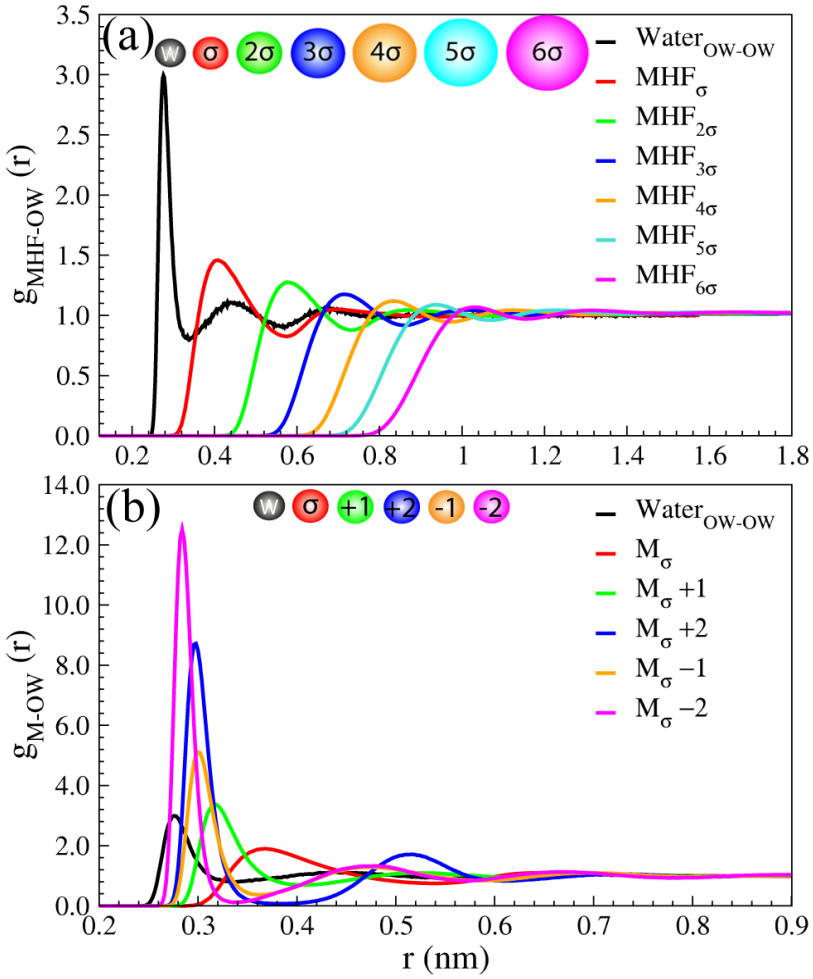


Figure 5.10. *Radial Distribution Function of water oxygen around (a) hydrophobic particles (MHF) of different sizes (σ to 6σ) (b) methane-like particles with an additional charge. For comparison, results for water and methane (M_σ) solutes are also shown. Inset shows the solutes schematically.*

Fig. 5.10 shows the radial distribution function (RDF) of water's oxygen atom around the hydrophobic particles (MHF) of different sizes (Fig. 5.10.a) and around methane-like particles (M_σ) with different charges (Fig. 5.10.b). For comparison, we have also taken a water (w) molecule as a solute (water-solute). RDF around its oxygen atom is also shown. As expected, distribution functions around hydrophobic solutes are broader compared to pure water RDF, indicating less compact water structure around the former. As the size of the hydrophobic solute increases, the peak of the RDF decreases further. We have chosen sizes in the multiples of methane diameter (0.373 nm denoted as

σ). Largest size considered here is 6σ, where RDF still retains some structure, i.e., the peak of the RDF is close to 1.0. Inhomogeneous solvation theory (IST) estimates translational entropy difference based on the RDF between the solute and the solvent, $S_{tr}^{IST} = -4\pi k_B \int dr r^2 g(r) \ln g(r)$. Entropy goes to zero when $g(r)$ either goes to 0 (short distance) or to 1 (long distance). However, $g(r)$ is normalized to 1 by the probability of ideal (non-interacting) gas. Therefore, the reduction in the water entropy captured by IST is the difference between the bulk water and the water molecules around a solute.

Table 1: *Translational entropy reduction calculated using IST approach for the systems considered here. Each value is multiplied by T=298K***.**

System	Translational Entropy (IST) (kcal/mol)
Water	-1.6
M_σ	-2.7
M_σ+1	-3.9
M_σ+2	-11.5
M_σ-1	-4.6
M_σ-2	-10.9
MHF_σ	-2.1
$MHF_{2\sigma}$	-4.6
$MHF_{3\sigma}$	-7.2
$MHF_{4\sigma}$	-10.8
$MHF_{5\sigma}$	-15.2
$MHF_{6\sigma}$	-20.0
MHF-0.5	-1.6

Table 1 shows the IST translational entropy difference (S_{tr}^{IST}), multiplied by 298 K, of all systems studied here. As shown before by Lazaridis *et al.*[88], reduction in the water entropy around methane is rightly captured here. Interestingly, if we consider a particular water molecule as a solute, we get a reduction in entropy of water around this water-solute itself. However, this contribution is constant and will remain even in pure water.

However, IST translational entropy is negative even for the biggest particle (6σ) considered here. This is understandable as the peak of the RDF around this particle is close to 1 even at this size. Moreover, larger particle has higher coordination number. Therefore, the entropy attains a more negative value with the size of the particle. Reduction in the water entropy is more around similar sized particle with increasing the magnitude of charge as expected. However, negative charges create enhanced structure compared to the same amount of positive charges (Fig. 10.b), reflecting in entropy estimates as discussed later.

5.3.2. *Single Water Translational Entropy*

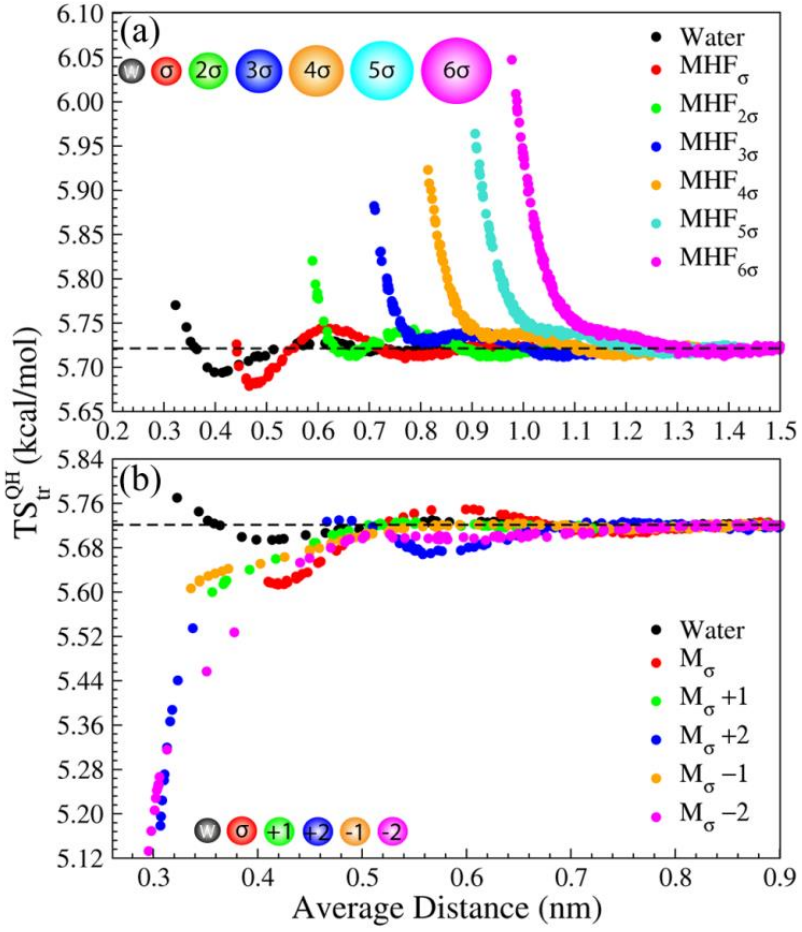


Figure 5.11. *Translational entropy (TS_{tr}^{QH}, T=298K) of individual water molecules against their average distance from the (a) hydrophobic solutes of different sizes (σ to 6σ, σ=0.373nm), (b) methane-like solutes with additional charges. For comparison, water-solute is also shown. Dashed lines denote the average bulk value (far away from solute).*

Fig. 5.11 shows the translational entropy (multiplied by 298 K) calculated using the present method for individual water molecules as a function of its average distance from the solute. Each point in the graph represents the translational entropy value of a single water molecule. Fig. 5.11.a shows the entropy variation of water molecules around hydrophobic solutes of varying sizes, while Fig. 5.11.b shows the same for methane-like solutes with varying charges. For comparison, water is also used as a central solute around which entropy of other water molecules is calculated. The translational entropy of all the water molecules around every solute (Figs 5.11a,b) converges, as expected, to the same value 5.72 kcal/mol far away(where RDF goes to 1 in Fig. 5.10) from the solute. Note that each water molecule has some distinct value. This indicates that the configurational space sampled by each water molecule is inherently different. Fig. 5.11 indicates that even for spherical particles, each water molecule around it has different translational entropy depending on its average proximity to the solute. Here, the translational entropy values were calculated from the trajectories that are permuted using initial reference configuration (at t=0 ns).

It was recently suggested that many water molecules slow down near a hydrophobic solute.[90] Fig. 5.11 here shows the number of water molecules whose translational entropy is affected by the hydrophobic solutes of varying sizes. It also depicts the spatial extent of this effect. In case of MHF_{σ}, the translational entropy of the closer water molecules decreases initially and then gradually increases to the bulk water entropy (5.72 kcal/mol). For this system, roughly 19 out of 20 water molecules in the solvation shell loose translational entropy giving rise to a total -0.38 kcal/mol reduction (Table 2). When the size becomes double ($MHF_{2\sigma}$), water molecules closer to the solute have higher entropy than that of the bulk water. Therefore, the crossover in translational entropy from negative to a positive value is captured at single water level itself. Translational entropy of the closest water molecule becomes larger with the size of the hydrophobic particle. For these bigger hydrophobic particles, translational entropy of the water molecules decreases with distance from the solute and finally matches the bulk value at large separation. Together with the increased number in the coordination shell, total translational entropy of the water molecules increases non-linearly with the size of the hydrophobic solutes for the set of systems considered here. Thus, the total change in

the translational entropy around the hydrophobic particle of size 2σ is 0.69 kcal/mol. For hydrophobic solutes of bigger size, this trend persists to make a difference of 10.64 kcal/mol for the biggest hydrophobic particle considered here ($MHF_{6\sigma}$).

It is also interesting to observe the behavior of the water entropy around a central water molecule acting as a solute. A few (five) water molecules closer to the water-solute have slightly higher entropy than the bulk water. With the increase in distance, the translational entropy first decreases and then increases before reaching the bulk value. The overall reduction obtained in the translational entropy for water molecules around a particular "water-solute" is -0.11 kcal/mol (Table 2).

To understand water entropy behavior around the attractive solutes, we used a set of systems containing methane-like solutes with varying charges. Fig. 5.11.b shows the translational entropy of the individual water molecules as a function of their average distances from these solutes. As expected, the water molecules closer to these solutes have low entropy that increases with distance reaching finally to the bulk value. Note that, even the weakly attractive methane lowers water entropy more than the hydrophobic solute of the same size. With the incorporation of charges, entropy loss increases. However, entropy loss is more for the negative charge compared to the equivalent positive one (Table 2). Methane-like particle with -2 charge reduces the translational water entropy by -6.54 kcal/mol, while reduction around the same with +2 charge is -5.16 kcal/mol.

To see the range of the anharmonic contribution, we chose three different types of solutes -- methane (M_σ), biggest hydrophobic solute ($MHF_{6\sigma}$), and highly charged solute (M_σ-2) -- that have varying interactions, and calculated anharmonic contributions of water molecules present in the first solvation shell of these solutes. The average anharmonic contribution is found to be 1.0±0.21% for M_σ, 0.18±0.04% for $MHF_{6\sigma}$, and 5.6±0.6% for M_σ-2.The anharmonic contribution of the bulk water entropy is found to be 0.49±0.02%. Therefore, the anharmonic contribution seems to depend on the interaction strength of the solute. Table A-5.1 in Annexure 5 shows the percentage of the anharmonic contribution to the translational entropy of the closest water molecule around different solutes. Since these anharmonic contributions are relatively small and will not change the qualitative picture of entropy variation with distance from the solute, we have

not incorporated anharmonic corrections to the translational entropy of the water molecules for the spherical solute systems.

We now compare the translational entropy loss from both the above approaches: IST (Table 1) and the present method based on permutation reduction

Table 2: Sum of change in rotational entropy and translational entropy of water molecules around different types of solute.

System	Total $T\Delta S_{tr}^{QH}$ (kcal/mol)*	Total $T\Delta S_{rot}$ (kcal/mol)*	Total $T\Delta S_{tr}^{QH} + T\Delta S_{rot}$ (kcal/mol)
Water	-0.11	-0.48	-0.59±0.25
MHF_{σ}	-0.38	-1.58	-1.96±0.31
$MHF_{2\sigma}$	0.69	-2.94	-2.25±0.54
$MHF_{3\sigma}$	2.06	-3.92	-1.86±0.70
$MHF_{4\sigma}$	4.30	-4.76	-0.46±0.91
$MHF_{5\sigma}$	6.63	-5.96	0.67±1.11
$MHF_{6\sigma}$	10.64	-6.93	3.71±1.40
M_{σ}	-1.79	-1.84	-3.63±0.25
M_{σ}+1	-1.02	-2.72	-3.74±0.19
M_{σ}+2	-5.16	-11.55	-16.71±0.34
M_{σ}-1	-1.41	-7.22	-8.63±0.21
M_{σ}-2	-6.54	-25.64	-32.18±0.45
MHF_{σ}-0.5	0.16	-1.04	-0.88±0.15

**The error in each water entropy ranges between 0.001 to 0.007 kcal/mol. Here, we have taken 0.007 kcal/mol as the upper bound of error and calculated the total error by multiplying with total number of water molecules contributing to the entropy change.*

(Table 2). For methane particle, the water entropy reduction differs between IST (-2.7 kcal/mol) and the present (-1.79 kcal/mol) method, although the qualitative picture remains the same. However, for big hydrophobic solutes, even the qualitative trend is different. While the present approach captures the crossover in the translational entropy for the hydrophobic particles beyond 2σ size, IST fails to do so at least up to the largest

size (6σ) considered here. Note that, IST takes into account only the solute-solvent pair correlation. Therefore, the absence of crossover in IST may result from the missing solvent-solvent correlation, or a long-range correlation in water.

To check the sensitivity of the results with the reference frame, we carried out permutation reduction for the water molecules close to methane using different frames, at 0.0 ns, 50 ns, and 100 ns. The comparison of the translational entropy of the permuted water molecules around methane (Fig. 5.12) using different reference configurations (at t=0, 50, and 100 ns) shows that the translational entropy follows the same trend with the increase in distance from the solute.

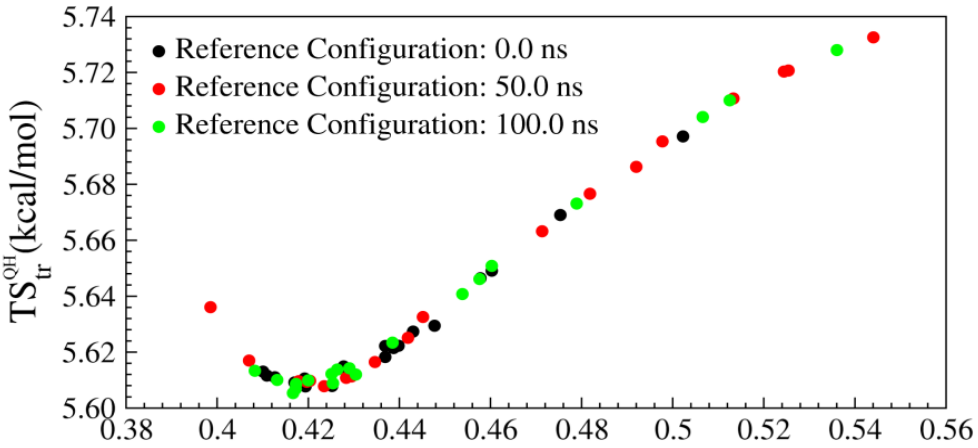


Figure 5.12. *The translational entropy of water molecules against the average distance from methane calculated using different reference configurations.*

5.3.3. *Single Water Rotational Entropy*

The rotational entropy of each permuted water molecule was calculated from its orientational distribution using Eq. 3. Fig. 5.13 shows the rotational entropy of the individual water molecules around hydrophobic solutes of various sizes (Fig. 5.13.a) and methane like particles with various charges (Fig. 5.13.b) against their average distance from the solute.

The behavior of rotational entropy is similar to that of the translational entropy with a few exceptions. The water molecules closer to a solute have lowest rotational entropy that increases with distance from the solute, finally reaching the bulk value close to zero (-0.03kcal/mol). Therefore, far away from the solute, water's orientational distribution is independent of the solute, as expected. The small negative rotational

entropy of bulk water may be attributed to numerical inaccuracy. As the size of the hydrophobic solute increases, there is a small increase in the rotational entropy of water molecules close to the solute. However, total loss in the rotational entropy is much higher for a bigger solute due to the increased coordination number. The total rotational entropy is still negative even for the largest size considered here (6σ), (Table 2). The reason for the absence of crossover lies in the fact that total rotational volume is constant which constrains to maximum value of the entropy to zero (Eq. 3).

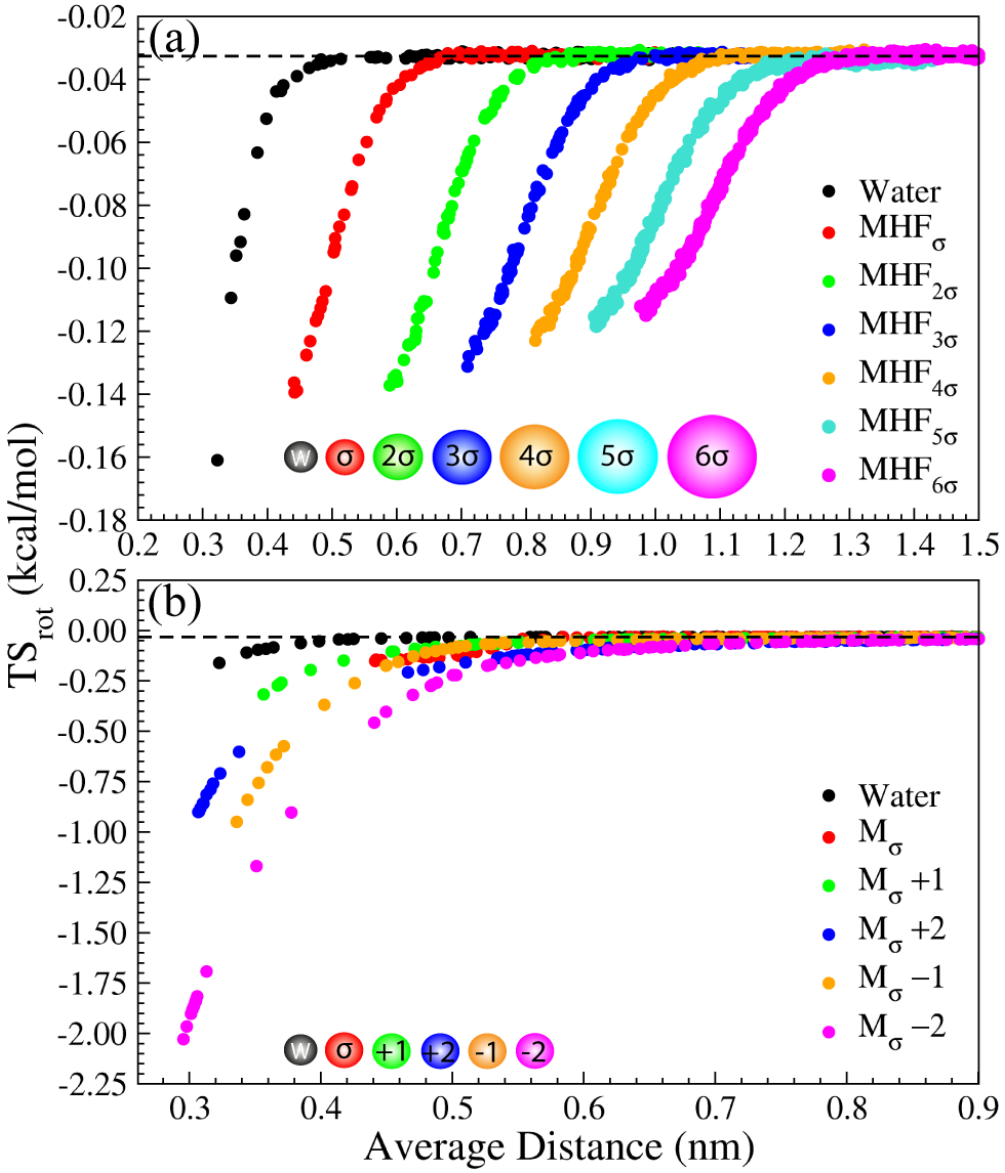


Figure 5.13. *Rotational entropy (TS_{rot}, T=298K) of individual water molecules against their average distance from the (a) hydrophobic solutes of different sizes (σ to 6σ, σ=0.373nm), (b) methane-like solutes with additional charges. For comparison, water-solute is also shown. Dashed lines denote the average bulk value (far away from solute).*

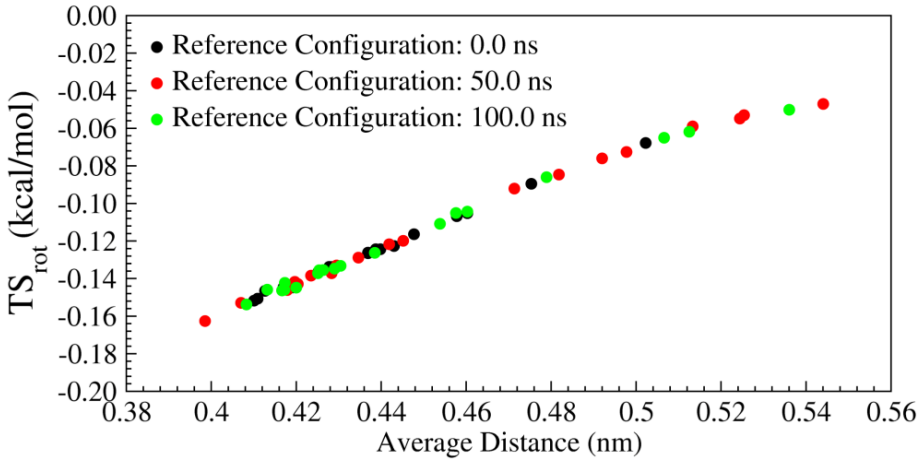


Figure 5.14. *The rotational entropy of water molecules against the average distance from methane calculated using different reference configurations*

In the case of water acting as a solute, we observed a smaller reduction in rotational entropy (less than that of methane) for the water molecules surrounding it. Reduction in rotational entropy around charged solutes is more than that of the hydrophobic solutes (Fig. 5.13.b, Table 2). As in the case of translational entropy, reduction in rotational entropy of water is also more for the negatively charged solutes compared to that with similar magnitude of positive charge.

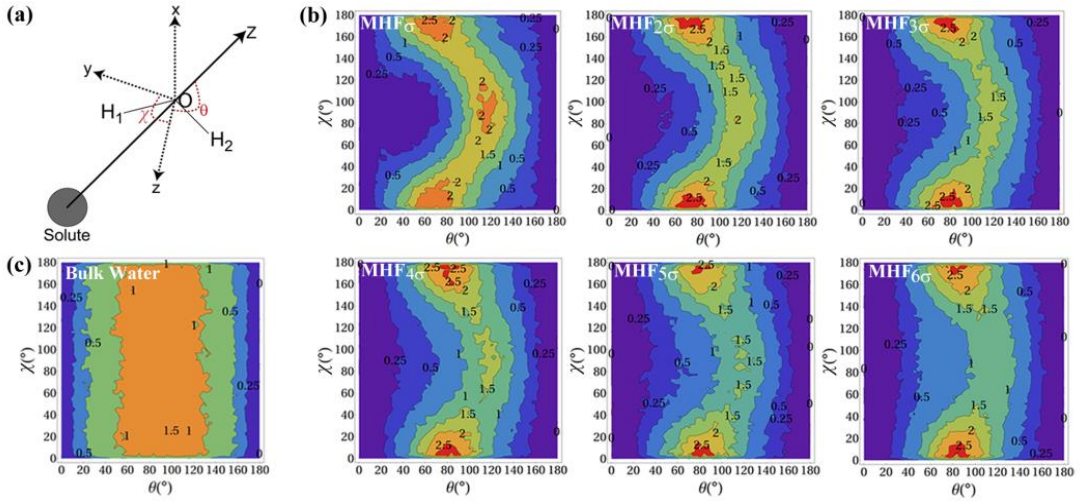


Figure 5.15. *(a) Schematic diagram showing the two angles θ and χ. θ is the angle between the dipole vector and the vector from water's oxygen to the solute. χ is the angle between the H-H vector of the water and the normal to the plane defined by the solute-oxygen vector and the dipole vector. (b) Orientational distribution of the water molecule closest to the hydrophobic solutes of different sizes. Note that, although the peak of the distribution shifts to a higher value of θ with an increase in size, distribution does not become uniform. (c) Orientational distribution of a permuted water molecule in the bulk*

Similar to the translational entropy, choice of the reference configuration is not important in the variation in rotational entropy (Fig. 5.14). Fig. 5.14 shows that the reference frame selection does not affect the entropy as a function of its average distance from the solute.

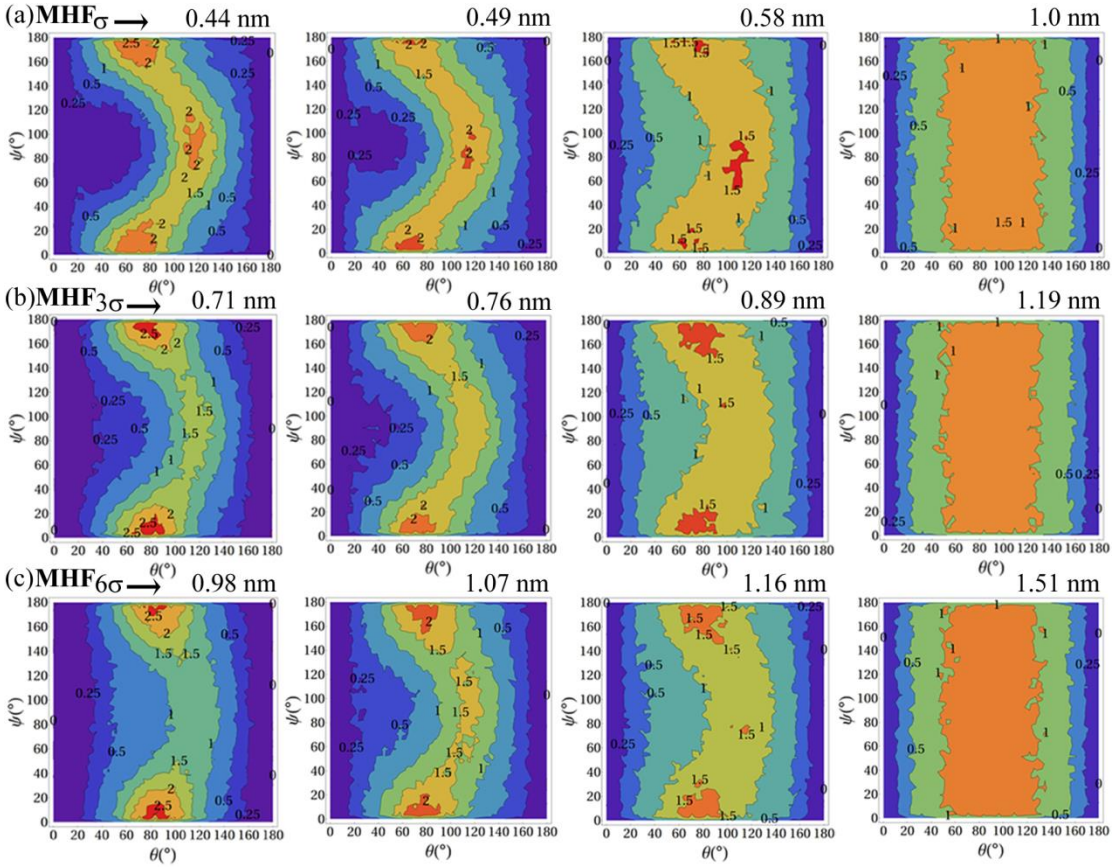


Figure 5.16. *Representative two-dimensional contour plots for orientational distribution of certain water molecules present around hydrophobic solutes of different sizes and at different distances. Direction of the arrow (along the row) indicates increase in distance with respect to solute.*

Absence of crossover in rotational entropy is intriguing. Rotational entropy of a water molecule depends on its orientational distribution (Fig. 5.15) with respect to the solute. Fig. 5.15.a schematically represents the orientation of a water molecule with respect to a solute using two angles θ and χ. Fig. 5.15.b shows orientational distributions of the closest water molecule around hydrophobic solutes of different sizes. Note that in Fig. 5.15.b the peak of the orientational distribution changes to different locations (higher θ value) with the increase in the size of the hydrophobic particle. However, there is not much reduction in the peak height. Thus, the distribution does not become uniform as

seen in the case of bulk water (Fig. 5.15.c). Therefore, with an increase in the size of the hydrophobic solute, we do not observe any major increase in the rotational entropy of the water closer to a solute causing a crossover from negative to positive entropy change. Orientational distributions of other water molecules at various distances are shown in Fig. 5.16. The average distance of water molecules from the COM of corresponding solute is shown on top of each panel. The 4th column represents water in the bulk for each system. As the size increases, peaks of the distribution for the water molecules closest to the solute shift to higher value of θ. However, it does not become uniform. With increasing distance, however, the distribution becomes uniform. It was observed earlier that even around LJ wall (~infinite radius) orientational distribution peaks at different location.[91]

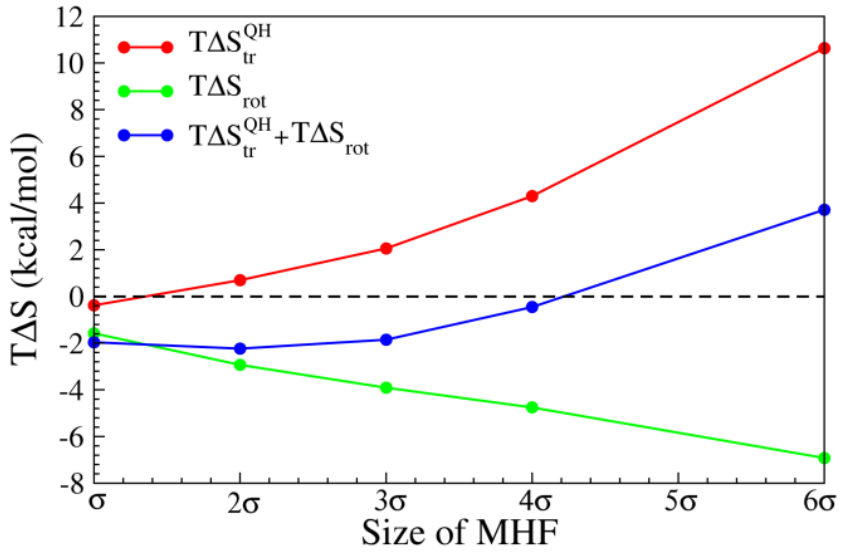


Figure 5.17. *Total change in the water entropy around the hydrophobic solutes of different sizes. As translational entropy increases with size of the hydrophobic particles, rotational entropy decreases, causing the crossover to happen at a larger size.*

The above result signifies that the observed entropic crossover[25] from smaller to bigger hydrophobic particle originates from the translational entropy alone. The rotational entropy opposes the crossover and forces the crossover in the overall entropy to happen at a larger length scale. Fig. 5.17 shows the total translational, rotational, and combined entropy reduction of solvation water as a function of the size of the hydrophobic particle. It shows that while the crossover in the translational entropy occurs at around 1.5σ, the crossover in the total entropy happens at around 4.5σ. Our result is

similar to that of Goddard *et al.*[47], where the crossover for hydrophobic particles at around 0.6 nm (1.6σ) was obtained through translational entropy alone.

5.3.4. Hydrogen Bond (H-bond) Variation

We have discussed above the contribution of each water molecule to the overall change in the water entropy around different solutes. Reduction in the water entropy around small hydrophobic solutes has often been attributed to an enhanced structure formation[23], whereas the entropy increase around bigger hydrophobic solutes is attributed to the formation of a loosely bound structure[23].

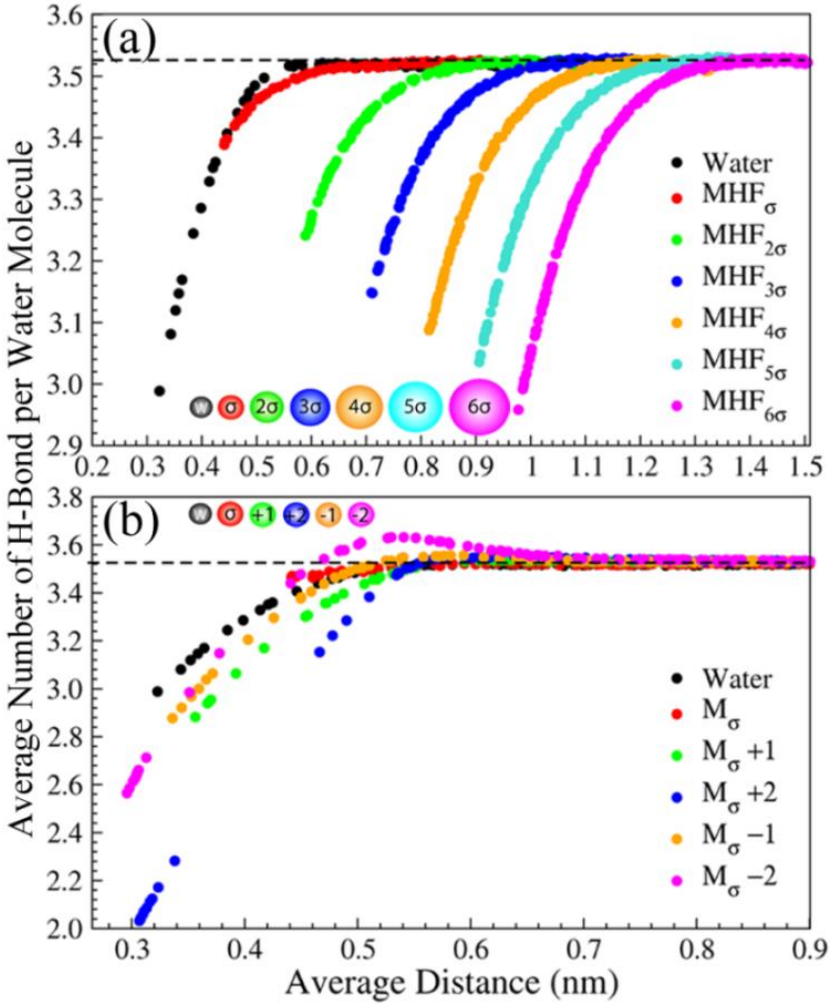


Figure 5.18. *Average number of H-bond formed by a single permuted water molecule (without considering H-bond to the solute) as a function of their average distance from the particle for the (a) hydrophobic particles of different sizes (σ to 6σ, σ=0.373nm) and (b) methane-like particles with additional charges. For comparison, water-solute is also shown.*

To understand the role of hydrogen bonding in the single water entropy, we have calculated the average number of H-bonds formed by each permuted water molecule with the rest of the water molecules. H-bond is defined when donor-acceptor distance is less than 0.35 nm and acceptor-donor-hydrogen angle is less than 30°. In an ideal tetrahedral geometry, a water molecule participates in four H-bonds, two by donating and two by accepting. However, the average number of H-bonds formed by a TIP4P water molecule is around 3.57 in bulk water[92]. Figure 5.18 shows the average number of H-bond donated and received by an individual (permutated) water molecule against their average distance from the hydrophobic (Fig. 5.18.a) and charged solutes (Fig. 5.18.b). Since hydrophobic molecules do not participate in H-bond, we did not consider H-bonds between the solute and the surrounding water molecules for any system so that we can compare the effect of the solute on H-bond network among solvent water only.

Water molecules close to any solute loose H-bond. However, as the distance increases, the average number of H-bond increases gradually reaching finally the average bulk value 3.52 at a larger distance from the solute. It is interesting to note that the loss in H-bond for the closest water molecule around a water-solute (26.3%) is more compared to the same around a hydrophobic solute of the size of methane (13.0%), and even around the hydrophobic solute of size 5σ (24.0%). Therefore, there is an enhancement in the H-bond among solvent water molecules around small hydrophobic solutes compared to water-solutes, similar to what observed in tetrahedral order parameter[93-94]. Decrease in the H-bond among solvent water is more around charged particles (Fig. 5.18.b). When the solute particle is considered as an H-bond partner, the average number of H-bond formed by the closest water molecule around water-solute attains bulk value. Therefore, the apparent reduction in the H-bond around water-solute and charged particle is due to participation of the solute itself in H-bond with the solvent. Since hydrophobic solutes do not form H-bond with the water, the average number of H-bonds among solvent molecules alone appears high.

As the size of the hydrophobic solute increases, the average number of H-bond decreases. This is attributed to dangling H-bond formation[27], corroborated recently using experiment[28]. To see this effect, we calculated the percentage of dangling H-bond with and without the solute being an H-bond partner (Fig. 5.19). This analysis clearly shows

that with an increase in the size of the hydrophobic solute, the percentage of dangling H-bond increases (blue line). For these solutes, the increase in dangling H-bond is the same with and without solute as H-bond partner, indicating that water does not take part in H-bond with the solute. For the biggest solute considered here (6σ), the dangling hydrogen bond is 26%.

For charged solutes, the behavior is different. When water acts both as a solute and an acceptor, the number of dangling H-bond becomes similar to the bulk value (11.5%; dotted black line in Fig. 5.19). However, if it is not considered as an acceptor, the same increases to 23.5% (solid black line in Fig. 5.19). Therefore, the water-solute

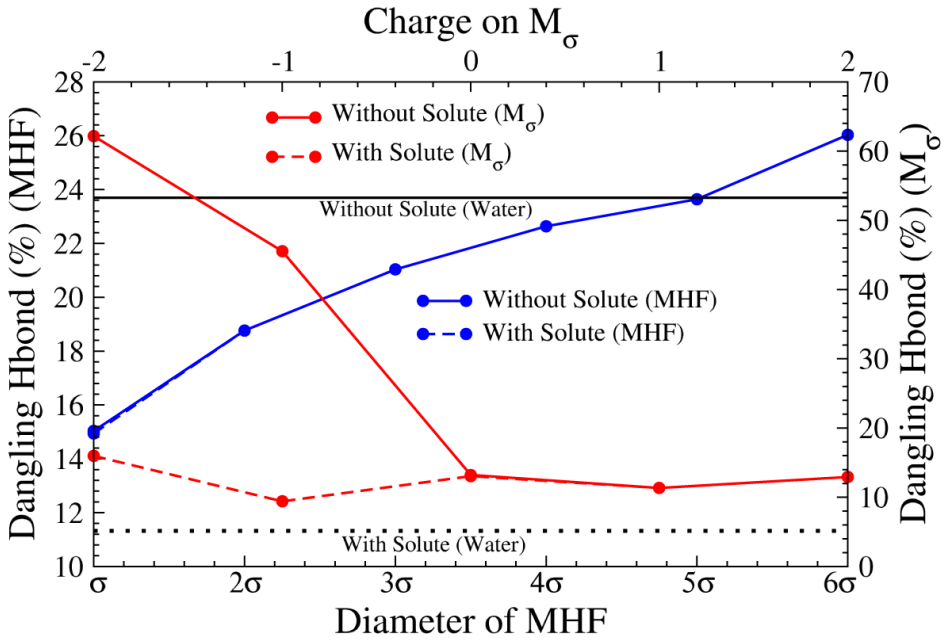


Figure 5.19. *Dangling H-bond for the water closest to hydrophobic solutes of different sizes (left-bottom) and charged solutes (right-top) with and without solute as acceptor.When both the hydrogen atoms of a water molecule are not involved in H-bond, the water is considered to form 100% dangling H-bond. Black solid and dashed lines show the values for water-solute.*

contributes 12% of H-bonds formed by the closest water molecule. For negatively charged particles, on the other hand, a significant difference is observed. The number of dangling H-bond is very high when the solute is not considered as an acceptor (62% for -2 charge). It is close to the bulk otherwise (16% for -2 charge). Since solutes with zero

and positive charges do not act as acceptors, the number of dangling H-bond remains same with and without solute's involvement in the H-bond calculation.

Therefore, dangling H-bond (with or without solute's participation) is not a good indicator for entropy loss for a charged particle. However, entropy gain for bigger hydrophobic solutes can loosely be correlated with the number of dangling H-bonds.

5.3.5. Recovery of the Entropy Loss

The previous results show that both the small hydrophobic and the attractive solutes reduce the water entropy. However, it is interesting why water, with both partial charges and attractive vdw interaction, induces lesser entropy reduction. To understand the origin of this, we created a hydrophobic solute of the size of methane with charge -0.5 (MHF_{σ}-0.5). This was done to induce energetic frustration in the surrounding water. Being hydrophobic in nature, it should force the surrounding water to form H-bond with other water molecules, whereas its negative charge would drive the H-bond formation of the surrounding water with itself. Table 1 shows that the translational entropy loss, calculated from IST, around this particle (-1.6 kcal/mol) is same as that around the water-solute (-1.6 kcal/mol). Table 2 shows that this particle has higher translational entropy than water-solute, compensated by lower rotational entropy, giving rise to a comparable total entropy loss of surround water molecules. Therefore, water being dipolar may create similar energetic frustration to reduce the entropy loss of water surrounding it.

5.3.6. Application to Drug Design

Our method of calculation of the single water entropy can be applied to drug design. When a drug binds to an active site of a biomolecule (protein/DNA), water is replaced to the bulk increasing the entropy of the overall system. Many docking methods that predict the binding energy do not account for this extra entropy.[95] However, recently Friesner and coworkers[19] developed the "water map" from the inhomogeneous solvation theory, which accounts for this change in the water entropy[2, 49]. Our method provides an alternate route to calculate the entropy of the individual water molecules that may be replaced upon drug binding.

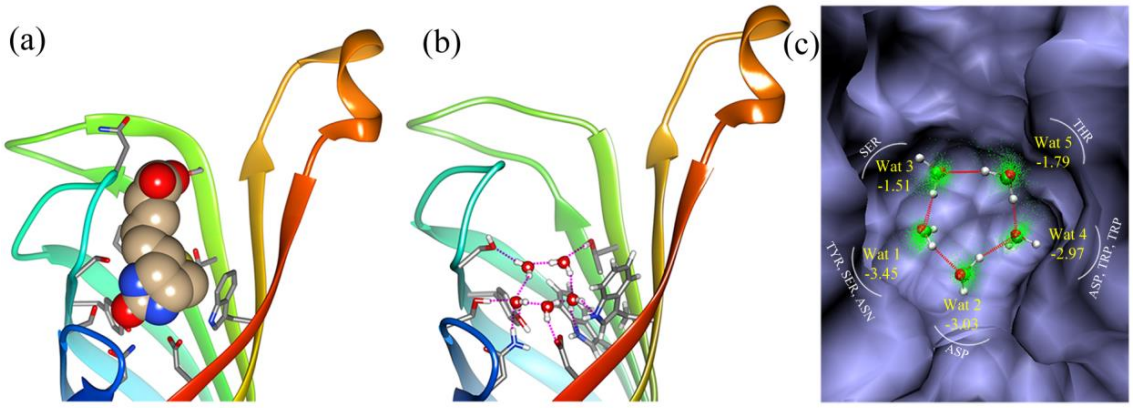


Figure 5.20. *The biotin binding cavity of streptavidin. (a) biotin-bound streptavidin cavity. Biotin is shown as VDW model and the protein as ribbon structure with stick model of amino acids involved in the binding process (b) pentameric arrangement of water molecules in the binding cavity in the absence of biotin. (c) Same as (b) labeled with total entropy of the individual water molecules. Protein cavity is shown with surface representation and water as CPK model. Green dots indicate positions of water oxygen atoms in each frame combine for several frames (~2000) after permutation reduction. The amino acids, which form hydrogen bonding with individual water molecules, are also labeled.*

Results from the above systems of hydrophobic and charged solutes provide a qualitative idea about how a particular solute may modify the entropy of surrounding water molecules. Here, we use a more realistic example, one studied by Friesner[19], of biotin binding to streptavidin. Friesner and co-workers have shown that the cavity in which biotin binds has five low-entropy water molecules in pentameric H-bonded arrangement. These water molecules are replaced to the bulk upon biotin binding, increasing the entropy component by 7 kcal/mol[19]. Figure 5.20.a shows the streptavidin-biotin complex where the head portion of the molecule sits in the deeper portion of the cavity stabilized by hydrogen bonding interactions with the surrounding amino acids such as SER, TYR, ASN, ASP, THR and TRP. In the absence of biotin, the structural arrangement of these amino acids triggers five water molecules to come together and form a pentagonal arrangement through inter-water hydrogen bonding.

Here, we performed molecular dynamics simulation of streptavidin in water after removing biotin from the cavity. We also find the similar pentameric arrangement of five water molecules in the cavity as shown in Fig. 5.20.b. Moreover, there are three other

water molecules (not shown in the figure) present in the mouth of the cavity. All these eight water molecules are replaced upon biotin binding.

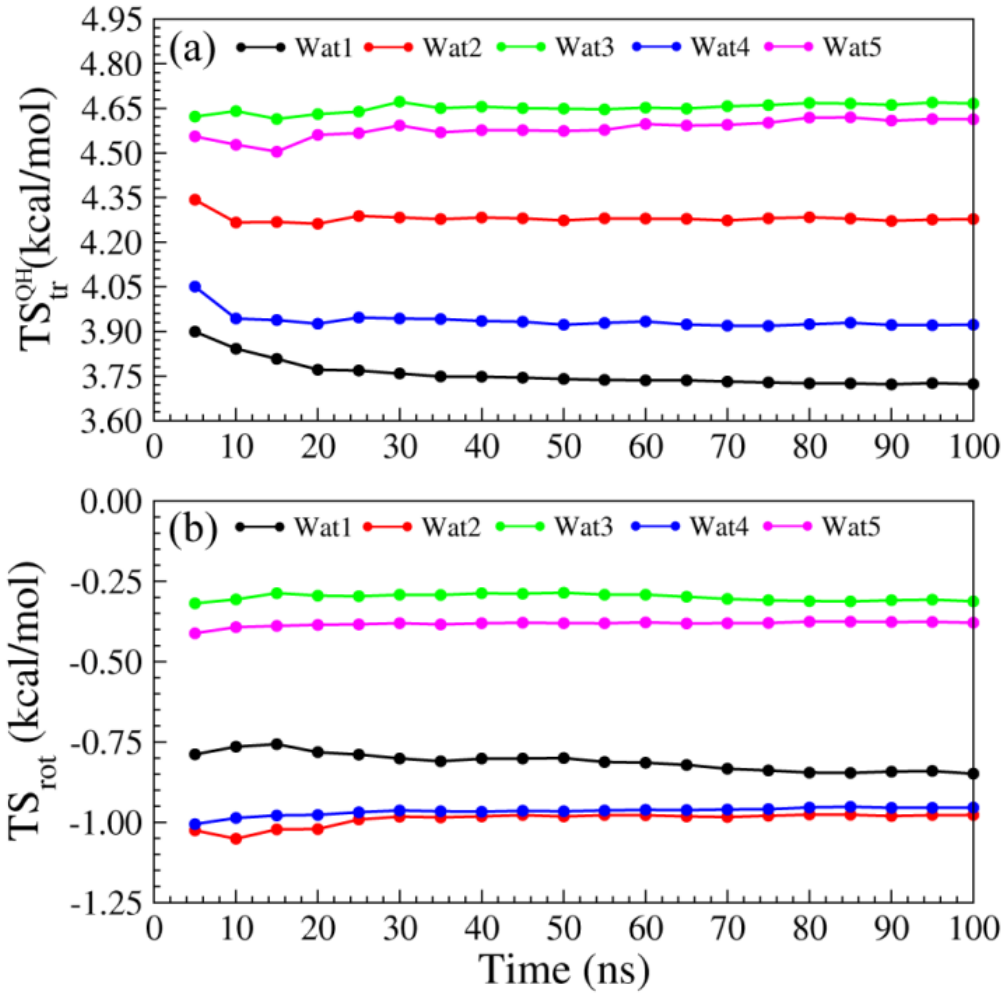


Figure 5.21. *Convergence of (a) translational and (b) rotational entropy of individual water molecules (shown in Fig. 5.20.c) in the streptavidin cavity.*

Since entropy of these water molecules contributes to the free energy of the biotin binding process, we calculated entropy of the individual water molecules in the cavity. Figure 5.20.c shows a representative structure of the five water molecules in the binding cavity with their entropy values calculated using the present method. The dots show several frames (~2000) of each permuted water molecule. Fig. 21a and b respectively shows the convergence of translational and rotational entropies of five water molecules found in the streptavidin cavity as a pentamer. In both cases, the convergence has reached around 40 ns.

Table 3: *Decomposition of the translational entropy (multiplied by T=298K) for the water molecules present in the streptavidin cavity.*

Water number	TS_{tr}^{QH} (kcal/mol)	TS_{tr}^{AH} (kcal/mol)
Wat 1	3.74	0.68
Wat 2	4.26	0.66
Wat 3	4.64	0.18
Wat 4	3.94	0.29
Wat 5	4.59	0.34
Wat 6	5.42	0.09
Wat 7	5.32	0.31
Wat 8	4.48	0.15

Table 3 shows the quasiharmonic and anharmonic translational entropy of water molecules present in the streptavidin cavity. The corrected translational entropy loss of water molecules in the cavity is shown in Table 4. It also shows the change in the rotational entropy and total change in entropy of every water molecule in the cavity. The total change in entropy for the five water molecules is -12.75 kcal/mol, comparable to -7.0 kcal/mol obtained by Friesner *et al*.[19] The total loss in entropy for all eight water molecules in the cavity is -15.50 kcal/mol (Table 4). However, we have observed that a 10 ns long simulation, performed by Friesner *et al*.,[19] may not be enough for the convergence of entropy values. In our study, it requires at least 40 ns to get the convergence (Fig. 5.21). Further, we performed H-bond analysis to understand the reduction in entropy of water molecules in the cavity. The analysis (green line in Fig. 5.22.a) shows that there is a negative correlation (correlation coefficient -0.96) between the entropy and the total number of H-bonds formed by a single water molecule with rest of the water molecules and surrounding amino acids.

Table 4: Change in the translational and rotational entropy (multiplied by T=298K) for water molecules present in the streptavidin cavity. $T\Delta S_{tr}^{QH-AH} = TS_{tr}^{QH-AH} - TS_{tr(Bulk)}^{QH-AH}$, *where,* $TS_{tr(Bulk)}^{QH-AH}$ *= 5.69 kcal/mol.*

Water number	$T\Delta S_{tr}^{QH-AH}$ (kcal/mol)	$T\Delta S_{rot}$ (kcal/mol)	$T\Delta S_{tr}^{QH-AH} + T\Delta S_{rot}$(kcal/mol)
Wat 1	-2.64	-0.81	-3.45
Wat 2	-2.09	-0.94	-3.03
Wat 3	-1.23	-0.28	-1.51
Wat 4	-2.05	-0.92	-2.97
Wat 5	-1.44	-0.35	-1.79
Wat 6	-0.37	-0.10	-0.47
Wat 7	-0.68	-0.07	-0.75
Wat 8	-1.32	-0.21	-1.53
Total	-11.82	-3.68	-15.50

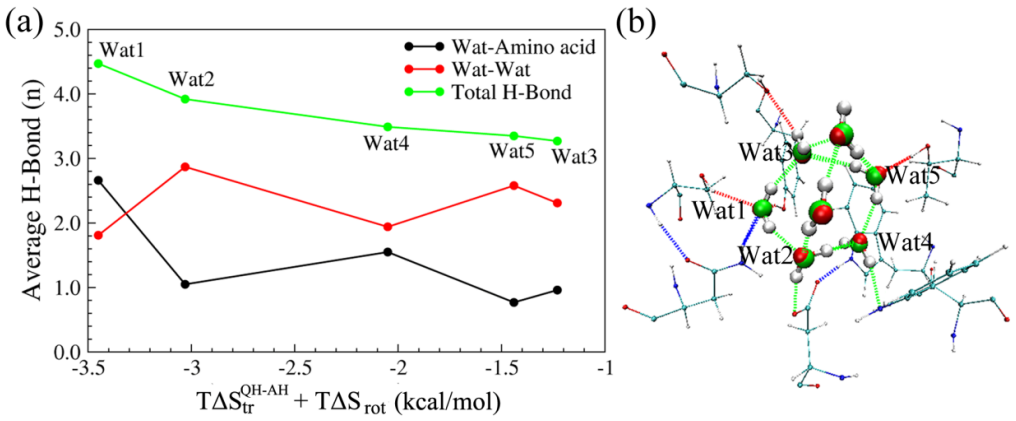


Figure 5.22. *Analysis of hydrogen bonding of individual water molecules in the streptavidin cavity against their total entropy. a) Average hydrogen bonding of the individual water molecules as a function of the total entropy. Here, hydrogen bond (H-bond) analysis (green line) shows a negative correlation (correlation coefficient -0.96), which indicates that entropy decreases with more hydrogen bonding. (b) Representation of hydrogen bonding formed by five water molecules with surrounding amino acids and other water molecules. Wat1 and Wat2 are the slowest water molecules with the maximum total number of H-bonds. Wat3, Wat4, and Wat5 are the faster water molecules compared to Wat1 and Wat2, but have lower number of H-bonds. Anharmonic contributions are removed from the translational entropy values of these water molecules.*

Although, in this test case, we kept the protein's heavy atoms restrained to compare with the result of Friesner and coworkers[19], the present method does not require any constraint. Therefore, using this method, we can estimate translational and rotational entropy of water around any system of choice at a single molecule level.

5.4. Conclusion

Here we proposed a method based on permutation reduction method to calculate the translational and rotational entropy of individual water molecules. Using the permuted water molecules, we adopted quasi-harmonic approach to calculate translational entropy and angular distribution to calculate rotational entropy individually. We applied this method for the water entropy calculations around hydrophobic solutes of different sizes and methane like particle with different charges. Results showed that attractive and small hydrophobic solutes (size < 0.56 nm) decrease the translational entropy of water. However, water molecules surrounding bigger hydrophobic solutes of size greater than 0.56 nm (1.5σ) has more translational entropy than bulk water. This crossover was observed even at single water molecule level depending on its position with respect to the solute. We showed that there is no crossover in rotational entropy, pushing the crossover in total entropy (translational + rotational) to occur at a larger size (> 4.5 σ).

We also showed that there is a loss of entropy around water itself when it is considered as a solute. However, this loss is less compared to the small hydrophobic solute of the size of methane (MHF). Our analysis demonstrated that the structural enhancement of water is more around MHF compared to another water due to the solute's inability to take part in hydrogen bonding with water. Increase in water entropy around bigger hydrophobic solutes was found to be correlated with the dangling H-bonds.

We showed that translational entropy depends on the accessible volume similar to that in the case of an ideal gas. However, the accessible volume here is the volume of the ellipsoid of inertia. Our results showed that the entropy of water depends on the nature of the configurational space itself, governed by the environment (i.e., proximity to the solute, for example).

Since any molecular recognition site such as proteins or nucleic acids consists of hydrophobic, positive or negative residues, objective of this work was to take different types of coarse-grained solutes and understand their effect on the entropy of the individual water molecules. We showed that there can be significant differences in the entropy loss between water molecules in the same protein cavity. Therefore, this method will help to design better drug aiming at specific water replacement. It will also help to understand the role of the water entropy in binding processes itself.

5.5. References

1. Levy, Y.; Onuchic, J. N., *Annu. Rev. Biophys. Biomol. Struct.* **2006,** *35* (1), 389-415.
2. Gilson, M. K.; Zhou, H.-X., *Annu. Rev. Biophys. Biomol. Struct.* **2007,** *36* (1), 21-42.
3. Hamelberg, D.; McCammon, J. A., *J. Am. Chem. Soc.* **2004,** *126* (24), 7683-9.
4. Agnieszka, K. B., Thermodynamics of Ligand-Protein Interactions: Implications for Molecular Design. 2011.
5. Li, Z.; Lazaridis, T., *J. Phys. Chem. B* **2004,** *109* (1), 662-670.
6. Jayaram, B.; Jain, T., *Annu. Rev. Biophys. Biomol. Struct.* **2004,** *33*, 343-61.
7. Kazufumi Takano, Y. Y. a. K. Y., *Protein Engg.* **2003,** *16* (1), 5-9.
8. Pethig, R., *Annu. Rev. Phys. Chem.* **1992,** *43* (1), 177-205.
9. Blake, C. C. F.; Pulford, W. C. A.; Artymiuk, P. J.; Huber, R., *J. Mol. Biol.* **1983,** *167* (3), 693-723.
10. Poortmans, L. W. a. J. B., Remy Loris, Dominique Maes, Freddy, *J. Biol. Chem.* **1996,** *271*, 30614-30618.
11. Klotz, I. M., *Science* **1958,** *128*, 815-822.
12. Teeter, M. M., *Proc. Natl. Acad. Sci. USA* **1984,** *81*, 6014-6018.
13. Ahmad, M.; Gu, W.; Geyer, T.; Helms, V., *Nat. Commun.* **2011,** *2*, 261.
14. Houk, K. N.; Leach, A. G.; Kim, S. P.; Zhang, X., *Ang. Chem. Int. Ed.* **2003,** *42* (40), 4872-4897.
15. Matthews, B. W.; Liu, L., *Protein Science* **2009,** *18* (3), 494-502.

16. Snyder, P.; Lockett, M.; Moustakas, D.; Whitesides, G., *Eur. Phys. J. Special Topics* **2013**, 1-39.
17. Chodera, J. D.; Mobley, D. L., *Ann. Rev. Biophys.* **2013,** *42* (1), 121-142.
18. Li, Z.; Lazaridis, T., *J. Phys. Chem. B* **2004,** *109* (1), 662-670.
19. Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A., *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104* (3), 808-813.
20. Huang, D. M.; Chandler, D., *Proc. Natl. Acad. Sci. U. S. A.* **2000,** *97* (15), 8324-8327.
21. Ben-Naim, A.; Marcus, Y., *J. Chem. Phys.* **1984,** *81* (4), 2016-2027.
22. Stillinger, F. H., *J. Sol. Chem.* **1973,** *2* (2-3), 141-158.
23. Chandler, D., *Nature* **2005,** *437* (7059), 640-647.
24. Galamba, N., *J. Phys. Chem. B* **2013,** *117* (7), 2153-2159.
25. Davis, J. G.; Gierszal, K. P.; Wang, P.; Ben-Amotz, D., *Nature* **2012,** *491* (7425), 582-585.
26. Frank, H. S.; Evans, M. W., *J. Chem. Phys.* **1945,** *13* (11), 507-532.
27. Perera, P. N.; Fega, K. R.; Lawrence, C.; Sundstrom, E. J.; Tomlinson-Phillips, J.; Ben-Amotz, D., *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (30), 12230-12234.
28. Davis, J. G.; Rankin, B. M.; Gierszal, K. P.; Ben-Amotz, D., *Nat Chem* **2013,** *5* (9), 796-802.
29. Chandler, D., *Nature* **2002,** *417* (6888), 491-491.
30. Thirumalai, D.; Reddy, G.; Straub, J. E., *Acc. Chem. res.***2011,** *45* (1), 83-92.
31. Lazaridis, T., *Acc. Chem. res.***2001,** *34* (12), 931-937.
32. Baldwin, R. L., *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (19), 7310-7313.
33. Kauzmann, W., *Nature* **1987,** *325* (6107), 763-764.
34. Lee, B., *Biopolymers* **1985,** *24* (5), 813-823.
35. Lazaridis, T., *J. Phys. Chem. B* **2000,** *104* (20), 4964-4979.
36. Blokzijl, W.; Engberts, J. B. F. N., *Ang. Chem. Int. Ed.* **1993,** *32* (11), 1545-1579.
37. Meng, E. C.; Kollman, P. A., *J. Phys. Chem.* **1996,** *100* (27), 11460-11470.
38. Jorgensen, W. L., *J. Chem. Phys.* **1982,** *77* (11), 5757-5765.
39. Godec, A.; Merzel, F., *J. Am. Chem. Soc.* **2012,** *134* (42), 17574-17581.

40. Lum, K.; Chandler, D.; Weeks, J. D., *J. Phys. Chem. B* **1999,** *103* (22), 4570-4577.
41. Garde, S.; Patel, A. J., *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (40), 16491-16492.
42. Hummer, G.; Garde, S.; Garcia, A. E.; Pohorille, A.; Pratt, L. R., *Proc. Natl. Acad. Sci. U. S. A.* **1996,** *93* (17), 8951-8955.
43. Li, I. T. S.; Walker, G. C., *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (40), 16527-16532.
44. Laage, D.; Stirnemann, G.; Hynes, J. T., *J. Phys. Chem. B* **2009,** *113* (8), 2428-2435.
45. Berne, B. J.; Weeks, J. D.; Zhou, R., Dewetting and Hydrophobic Interaction in Physical and Biological Systems. In *Annu. Rev. Phys. Chem.*, 2009; Vol. 60, pp 85-103.
46. Lin, S. T.; Blanco, M.; Goddard, W. A., *J. Chem. Phys.* **2003,** *119* (22), 11792-11805.
47. Pascal, T. A.; Lin, S. T.; Goddard, W. A.; Jung, Y., *J. Phys. Chem. Letters* **2012,** *3* (3), 294-298.
48. Pascal, T. A.; Lin, S. T.; Goddard, W. A., 3rd, *Phys. Chem. Chem. Phys.* **2011,** *13* (1), 169-81.
49. Freire, E., *Drug Discovery Today* **2008,** *13* (19–20), 869-874.
50. Page, M. I.; Jencks, W. P., *Proc Natl Acad Sci U S A* **1971,** *68* (8), 1678-83.
51. Finkelstein, A. V.; Janin, J., *Protein Eng* **1989,** *3* (1), 1-3.
52. Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A., *Biophys. J.* **1997,** *72* (3), 1047-1069.
53. Mukherjee, A., *J. Phys. Chem. Letters* **2011,** *2* (24), 3021-3026.
54. Mcquarrie, D. A., *Statistical Mechanics*. University Science Books: 2000.
55. Pollack, G. L., *Science* **1991,** *251* (4999), 1323-1330.
56. Murphy, K. P.; Xie, D.; Thompson, K. S.; Amzel, L. M.; Freire, E., *Proteins* **1994,** *18* (1), 63-7.
57. Irudayam, S. J.; Henchman, R. H., *J Phys Chem B* **2009,** *113* (17), 5871-84.
58. Holtzer, A., *Biopolymers* **1995,** *35* (6), 595-602.

59. Lazaridis, T., *J. Phys. Chem. B* **1998,** *102* (18), 3531-3541.

60. Lin, S.-T.; Maiti, P. K.; Goddard, W. A., *J. Phys. Chem. B* **2010,** *114* (24), 8191-8198.

61. Kumar, P.; Buldyrev, S. V.; Stanley, H. E., *Proc. Natl. Acad. Sci.* **2009, 106, 22130-34**.

62. Baron, R.; McCammon, J. A., *Annu. Rev. Phys. Chem.* **2013,** *64* (1), 151-175.

63. McGrath, M. E.; Vasquez, J. R.; Craik, C. S.; Yang, A. S.; Honig, B.; Fletterick, R. J., *Biochemistry* **1992,** *31* (12), 3059-3064.

64. Kuroda, D. G.; Bauman, J. D.; Challa, J. R.; Patel, D.; Troxler, T.; Das, K.; Arnold, E.; Hochstrasser, R. M., *Nat Chem* **2013,** *5* (3), 174-181.

65. Baldwin, E. T.; Bhat, T. N.; Gulnik, S.; Liu, B.; Topol, I. A.; Kiso, Y.; Mimoto, T.; Mitsuya, H.; Erickson, J. W., *Structure* **1995,** *3* (6), 581-590.

66. Lam, P. Y.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; et, a., *Science* **1994,** *263* (5145), 380-384.

67. Breiten, B.; Lockett, M. R.; Sherman, W.; Fujita, S.; Al-Sayah, M.; Lange, H.; Bowers, C. M.; Heroux, A.; Krilov, G.; Whitesides, G. M., *J. Am. Chem. Soc.* **2013,** *135* (41), 15579-15584.

68. Brodney, M. A.; Barreiro, G.; Ogilvie, K.; Hajos-Korcsok, E.; Murray, J.; Vajdos, F.; Ambroise, C.; Christoffersen, C.; Fisher, K.; Lanyon, L.; Liu, J.; Nolan, C. E.; Withka, J. M.; Borzilleri, K. A.; Efremov, I.; Oborski, C. E.; Varghese, A.; O'Neill, B. T., *J. Med. Chem.* **2012,** *55* (21), 9224-9239.

69. Myrianthopoulos, V.; Kritsanida, M.; Gaboriaud-Kolar, N.; Magiatis, P.; Ferandin, Y.; Durieu, E.; Lozach, O.; Cappel, D.; Soundararajan, M.; Filippakopoulos, P.; Sherman, W.; Knapp, S.; Meijer, L.; Mikros, E.; Skaltsounis, A.-L., *ACS Med. Chem. Lett.* **2012,** *4* (1), 22-26.

70. Tran, Q.-T.; Williams, S.; Farid, R.; Erdemli, G.; Pearlstein, R., *Proteins: Structure, Function, and Bioinformatics* **2013,** *81* (2), 291-299.

71. Lindahl, E.; Hess, B.; Spoel, D. v. d., *J. Mol. Model* **2001,** *7*, 306-317.

72. Jorgensen, W. L.; Madura, J. D.; Swenson, C. J., *J. Am. Chem. Soc.* **1984,** *106* (22), 6638-6646.

73. Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J., *J. Am. Chem. Soc.* **1996,** *118* (45), 11225-11236.
74. Schlick, T., *Molecular Modeling and Simulation: An Interdisciplinary Guide: An Interdisciplinary Guide*. Springer: 2010.
75. Reinhard, F.; Grubmuller, H., *J Chem Phys* **2007,** *126* (1), 014102.
76. Weber, P. C.; Ohlendorf, D. H.; Wendoloski, J. J.; Salemme, F. R., *Science* **1989,** *243* (4887), 85-88.
77. Press, W. H., *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press: 2007.
78. Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; DiNola, A.; Haak, J. R., *J. Chem. Phys.* **1984,** *81* (8), 3684-3690.
79. Parrinello, M.; Rahman, A., *J. Appl. Phys.* **1981,** *52* (12), 7182-7190.
80. Nosé, S., *J. Chem. Phys.* **1984,** *81* (1), 511-519.
81. Hoover, W. G., *Physical Review A* **1985,** *31* (3), 1695-1697.
82. Arya, S.; Mount, D. M.; Netanyahu, N. S.; Silverman, R.; Wu, A. Y., *J. ACM* **1998,** *45* (6), 891-923.
83. Karplus, M.; Kushick, J. N., *Macromolecules* **1981,** *14* (2), 325 - 332.
84. Reinhard, F.; Lange, O. F.; Hub, J. S.; Haas, J.; Grubmuller, H., *Comput. Phys. Commun.* **2009,** *180* (3), 455-458.
85. Andricioaei, I.; Karplus, M., *J Chem Phys* **2001,** *115* (14), 6289 - 6292.
86. Singh, H.; Misra, N.; Hnizdo, V.; Fedorowicz, A.; Demchuk, F., *Amer. J. Math. and Management Sci.* **2003,** *23*, 301-321.
87. Numata, J.; Wan, M.; Knapp, E. W., *Genome Inform* **2007,** *18*, 192-205.
88. Lazaridis, T.; Paulaitis, M. E., *J. Phys. Chem.* **1992,** *96* (9), 3847-3855.
89. Carlsson, J.; Åqvist, J., *J. Phys. Chem. B* **2005,** *109* (13), 6448-6456.
90. Rossato, L.; Rossetto, F.; Silvestrelli, P. L., *J. Phys. Chem. B* **2012,** *116* (15), 4552-4560.
91. Lee, C. Y.; McCammon, J. A.; Rossky, P. J., *J. Chem. Phys.* **1984,** *80* (9), 4448-4455.
92. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., *J Chem Phys* **1983,** *79*, 926-935.

93. Raschke, T. M.; Levitt, M., *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102* (19), 6777-6782.

94. Chau, P. L.; Hardwick, A. J., *Mol. Phys.* **1998,** *93* (3), 511-518.

95. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., *J. Comput. Chem.* **2009,** *30* (16), 2785-2791.

Chapter 6

Summary and Future Directions

Chapter 6: Summary and Future Directions

6.1. Summary

The thesis entitled "***Molecular Mechanism of Drug-DNA Intercalation and the Role of Water Entropy in Molecular Recognition"*** provides the molecular details of the mechanism, kinetics and thermodynamics of a biomolecular recognition process DNA-intercalation by small molecule, proflavine. It also give insight into the importance of water entropy in molecular recognition processes.

DNA intercalation is one of the methods by which anti-cancer drugs function. Despite extensive experimental studies, molecular level understanding of the process is minimal. Therefore, we used computational methods to probe into the molecular mechanism and the pathway intercalation of an anticancer agent proflavine to DNA.[1]

By using well-tempered metadynamics simulations along with careful incorporation of configurational restraints, we captured direct intercalation of proflavine into DNA. Our study revealed that the intercalation happens by drug-induced cavity formation that defies natural fluctuation hypothesis. Comparison of the free energy, various energy and DNA structural parameters indicated that intercalation of proflavine adopts a minimal base stacking penalty pathway characterized by the sequence Twist➔Shift/Slide➔Buckle/Roll➔Rise. The general principles understood from the study of proflavine may apply to various other simple, flat aromatic molecules. We also showed that for intercalation from the minor groove, both entropy and energy contribute to the barrier of intercalation process. Here, water plays an important role in modulating the energy components of the barrier. The entropy component is mostly affected by the drug.[2]

Further, to understand the complete intercalation process and its associated thermodynamic and kinetic role, we performed intercalation and dissociation from both the grooves of DNA and deintercalation through both grooves. The results indicated that intercalation and deintercalation pathways of proflavine proceed through major groove side. However, intercalation is mostly dominated by the stable pre-intercalative minor groove-bound state, which builds up fast (nanoseconds) and slowly leaks to the intercalated state over a timescale of millisecond through major groove side. This long

timescale is in excellent agreement with experiment. This study predicted that the intercalated structure may not be as static as appears in the crystal structure and proflavine would likely change its orientation frequently (smaller barrier) even in the intercalated state. From this study it is evident that the design of a drug would eventually encounter the seemingly opposite effect that while an increased drug-DNA stacking interaction is required to stabilize the intercalated state, a less stable pre-intercalated state needs to be designed to make the kinetics faster.[3]

Similar to experimental methods, the computational methods adapted here to understand the intercalation process illustrated a two-step binding process: faster outside bound state formation and a slower intercalation process. Our previous calculation suggested the formation of outside bound state happens in nanoseconds. This observation is contrary to the experimental results which predicted the formation of outside bound state in sub-milliseconds time interval. So to understand this discrepancy, we used simple molecular dynamics simulations to carry out the association of proflavine around DNA at higher dye concentrations (P/D=1). Vigorous molecular dynamic simulations had shown that the association of proflavine around DNA happens in the time scale of nanoseconds matching with our previous prediction as well as the time scale obtained experimentally for proflavine dimer formation in solution (30 to 700 ns). Structural analysis had shown that proflavine molecules arrange as clusters of stacked molecules all around DNA compared to the experimental predictions that the association will take place close to the phosphate group.[4]

Water plays an important role in various molecular recognition processes through both entropic and enthalpic contributions of which the first one, entropy, is difficult to estimate. So we proposed a new method based on permutation reduction to calculate single water translational and rotational entropy to understand the general role of water present in different solute environments or in the bio-molecular cavities, surfaces or interfaces etc. Initially, we applied this method for the water molecules present around hypothetical solutes having hydrophobic nature of different sizes and methane-like particles with varying charges. Smaller hydrophobic solutes (size < 0.56 nm) and charged solutes decreased the translational entropy of surrounding water molecules. But, larger hydrophobic solutes (size > 0.56 nm) increased translational entropy of surrounding

water molecules. However, in case of rotational entropy, we did not see the crossover of entropy from negative to positive seen in case of translational entropy with respect to the size of the solutes. This led to the overall (translational + rotational entropy) cross over in entropy to happen at a larger size (> 1.68 nm). We applied this method on the water molecules present in the biotin binding cavity of streptavidin. Using our method, the total entropy of water molecules present in the cavity was found to be -12.75 kcal/mol. This new method will help in drug design based on specific water displacement and also will help to understand the role of water molecules in various molecular recognition processes.[5]

6.2. Future Directions

6.2.1. General Mechanism of flat intercalators and effect of substituents

By using proflavine as a prototype molecule to study the molecular mechanism of intercalation, we were able to capture the pathway of intercalation process where, proflavine intercalates through major groove but limited by a pre-intercalative minor groove bound state. The intercalation process was found to be a millisecond process which was attributed to the formation of initial minor groove bound state followed by intercalation through major groove. So, the time scale of formation of intercalated state depends on how strongly the pre-intercalative bound state is formed. But the stability of pre-intercalative bound state of the ligand depends on the nature of substituents. Hence, to understand the effect of substituents on the stability of pre-intercalative bound state and its influence of the overall time scale of intercalation, we would like extend the study using different proflavine derivatives using the same methodology what we have used for proflavine. Various proflavine derivatives chosen are shown in Fig. 6.1.

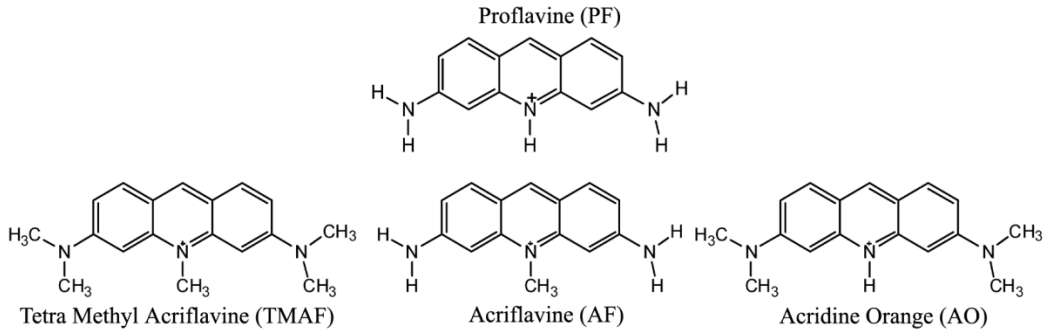


Figure 6.1. *Structure of proflavine and different proflavine derivatives.*

❖ ***Preliminary Results***

Initially, we performed the intercalation of proflavine derivatives, tetra methyl acriflavine (TMAF), acriflavine (AF), and acridine orange (AO) through minor groove using the same methods and methodology as described in the chapters 2 and 3. TMAF intercalated inside the DNA, but it broke the DNA characterized by the complete flipping of the fifth base pair of DNA. As a result, there was no increase in the length of 12 base pair DNA. The barrier was calculated to be 15.9 kcal/mol. The intercalation of acriflavine (AF) happened after crossing the energy barrier 19.5 kcal/mol with the intercalating base flipping with an increase in the Rise of the intercalating base pair. In case of acridine orange (AO), the intercalation barrier was found to be 15.5 kcal/mol. Here, the process did not cause any base flipping. The free energy surfaces and the intercalation geometry for all proflavine derivatives are shown in Fig. 6.2.

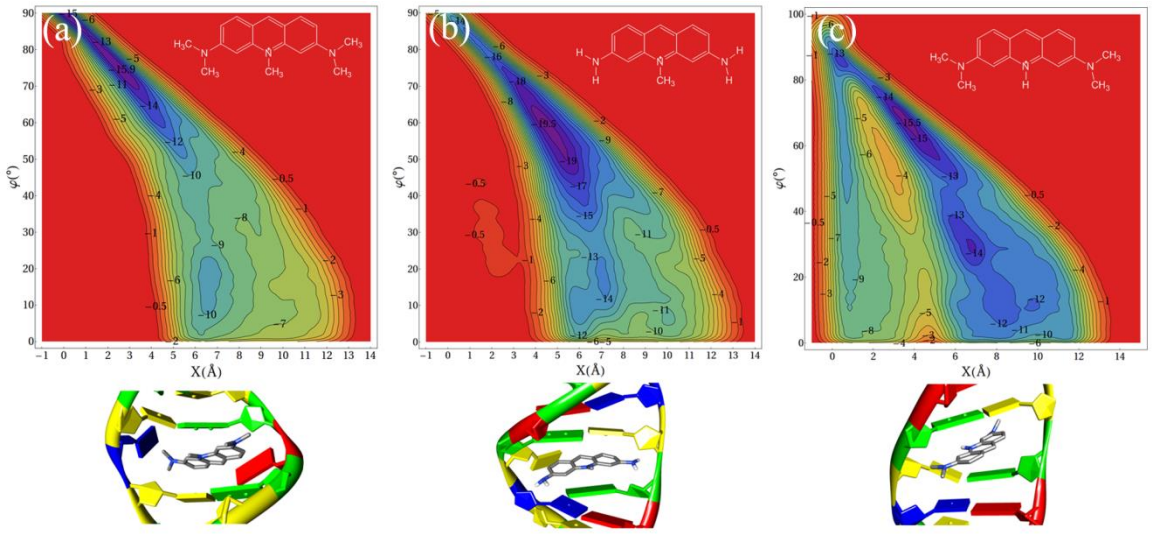


Figure 6.2. *Free energy surfaces against the collective variables (X, φ) and the corresponding intercalation geometry of (a) tetra methyl acriflavine (TMAF), (b) acriflavine (AF) and (c) acridine orange (AO).*

From this initial study, we can conclude that bulkier substituents on all nitrogen atoms in the case of TMAF make the intercalation through minor groove improbable. The intercalation barrier for AF is less as compared to AO suggesting the absence of bulkier group in the middle nitrogen atoms favors the intercalation process. To understand the effect of substituents on the stability of pre-intercalative bound state and the time scale of

overall intercalation process, we need to perform intercalation through major groove, deintercalation through minor and major groove and dissociation from major and minor groove bound states.

6.2.2. Molecular Mechanism of an Anticancer Agent- Ellipticine

We would also like to extend our study to understand the molecular mechanism of a potent DNA intercalating agent, ellipticine, which has been used an important anticancer agent to treat various cancers[6].

❖ ***Preliminary Results***

The preliminary results indicate that the intercalation of ellipticine happens through the major groove. All efforts for the intercalation through the minor groove failed as ellipticine causes DNA strand breakage. The barrier for intercalation through major groove was found to be 7.5 kcal/mol. The free energy surface along with different stages of intercalation process is shown in Fig. 6.3. Further analysis has to be done to understand the molecular mechanism of ellipticine intercalation through major groove. To understand the overall intercalation process in detail including the time scale, pathway etc, we need to perform the deintercalation and dissociation of ellipticine from DNA.

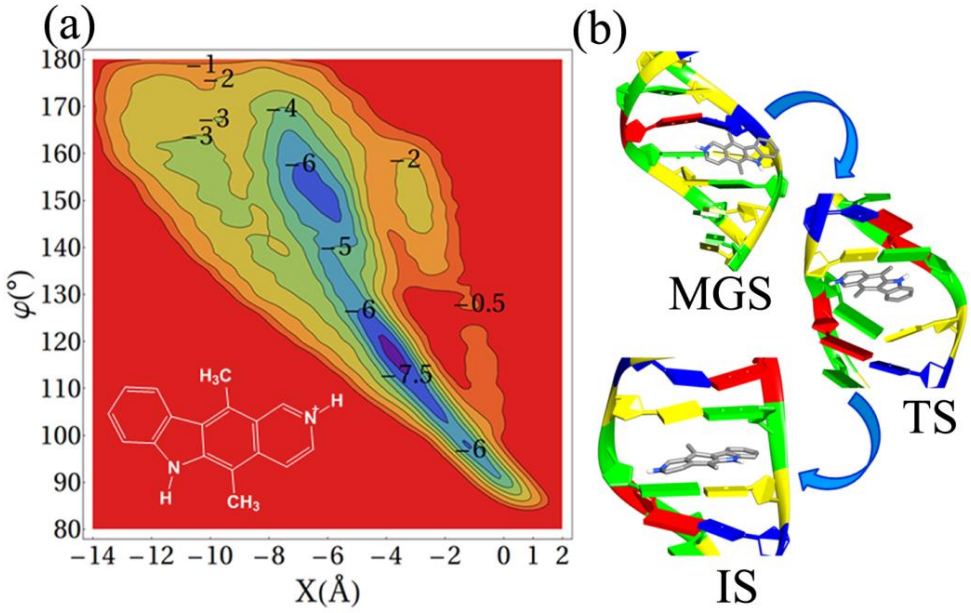


Figure 6.3. *(a) Free energy surface against the collective variables (X, φ) and (b) different stages of intercalation process: initial major groove bound state (MGS), transition state (TS) and final intercalated state (IS).*

6.2.3. Role of Water Entropy in the Difference in Stability between DNA and RNA in water

We would also like to apply the new method of calculation of entropy of water molecules to different biological systems to understand the role of water molecules in various molecular recognition processes. Two such systems are duplex DNA and RNA. We already know that duplex DNA is more stable in water as compared to duplex RNA. Hence, one of our aim is to understand the role of water molecule and its entropy in the difference between the stability of duplex DNA and RNA having the same nucleotide sequences.

6.2.3. Use of Water Entropy in Drug Discovery

Also we would like to extend our studies to design of new ligands to combat various ailments using water entropy calculation. Here, the ligand design is based on the ability of the molecule to displace low entropy water molecules from a protein cavity, surface or biomolecular interface as it will contribute more to the free energy of binding process of the ligand with the receptor.

6.3 References

1. Mukherjee, A.; Sasikala, W. D., Chapter One - Drug–DNA Intercalation: From Discovery to the Molecular Mechanism. In *Advances in Protein Chemistry and Structural Biology*, Tatyana, K.-C., Ed. Academic Press: 2013; Vol. Volume 92, pp 1-62.
2. Sasikala, W. D.; Mukherjee, A., *J. Phys. Chem. B* **2012,** *116* (40), 12208-12212.
3. Sasikala, W. D.; Mukherjee, A., *Phys. Chem. Chem. Phys.* **2013,** *15* (17), 6446-6455.
4. Sasikala, W. D.; Mukherjee, A., *Phys. Chem. Chem. Phys.* **2016,** DOI: 10.1039/c5cp07789c
5. Sasikala, W. D.; Mukherjee, A., *J. Phys. Chem. B* **2014,** *118* (36), 10553-10564.
6. Canals, A.; Purciolas, M.; Aymami, J.; Coll, M., *Acta Crystal. Sec. D* **2005,** *61* (7), 1009-1012.

List of Publications

List of International Publications

1. **Wilbee D. Sasikala** and Arnab Mukherjee*, Structure and dynamics of proflavine association around DNA, *Phys. Chem. Chem. Phys.,* **2016**, DOI: 10.1039/c5cp07789c.

2. Reman Kumar Singh, **Wilbee D. Sasikala** and Arnab Mukherjee*, Molecular origin of DNA kinking by transcription factors; *J. Phys. Chem. B,* **2015,** *119* (35), 11590-11596.

3. **Wilbee D. Sasikala** and Arnab Mukherjee*, Single Water Entropy: Hydrophobic Crossover and Application to Drug Binding, *J. Phys. Chem. B,* **2014,** *118* (36), 10553-10564.

4. **Wilbee D. Sasikala** and Arnab Mukherjee*, Intercalation and Deintercalation Pathway of Proflavine through the Minor and Major Grooves of DNA: Role of Water and Entropy, *Phys. Chem. Chem. Phys.* **2013,** *15* (17), 6446-6455.

5. Arnab Mukherjee* and **Wilbee D. Sasikala**, Chapter One - Drug–DNA Intercalation: From Discovery to the Molecular Mechanism. In *Advances in Protein Chemistry and Structural Biology*, Tatyana, K.-C., Ed. Academic Press: **2013**; Vol. Volume 92, pp 1-62. *(Invited Book Chapter)*

6. Krishna Gavvala, **Wilbee D. Sasikala**, Abhigyan Sengupta, Siddhi A. Dalvi, Arnab Mukherjee and Partha Hazra*, Modulation of excimer formation of 9-(dicyano-vinyl)julolidine by the macrocyclic hosts, *Phys. Chem. Chem. Phys*., **2013**,15**,** 330-340. *(Mentioned in Annexure-Collaborations)*

7. **Wilbee D. Sasikala** and Arnab Mukherjee*, Molecular mechanism of direct proflavine-DNA intercalation: evidence for drug-induced minimum base-stacking penalty pathway, *J. Phys. Chem. B* **2012,** *116* (40), 12208-12212.

8. Abhigyan Sengupta, **Wilbee D. Sasikala**, Arnab Mukherjee*, Partha Hazra*, Comparative study of flavins' binding with human serum albumin: a fluorometric, thermodynamic, and molecular dynamics approach, Chem. Phys. Chem. 13, 2012, 2142. *(Mentioned in Annexure-Collaborations)*

List of Publications in International Conference Proceedings

1. **Wilbee D. Sasikala** and Arnab Mukherjee, Intercalation Mechanism of Proflavine into DNA, Theoretical Chemistry Symposium (TCS), 8-12 December, 2010, IIT Kanpur.

2. **Wilbee D. Sasikala** and Arnab Mukherjee, Study on Molecular Mechanism of Intercalation of Anti-Cancer Drug Proflavine to DNA, Dynamics and Spectroscopy, 20 April, 2012, IISER Pune.

3. **Wilbee D. Sasikala** and Arnab Mukherjee, Study on Molecular Mechanism of Intercalation of Anti-Cancer Drug Proflavine to DNA, Theoretical Chemistry Symposium (TCS), 19-22 December, 2012, IIT Guwahati.

4. **Wilbee D. Sasikala** and Arnab Mukherjee, Molecular mechanism of direct proflavine-DNA intercalation, International Conference on Biomolecular Forms and Functions, 8-10 January 2013, IISc, Bangalore.

5. **Wilbee D. Sasikala** and Arnab Mukherjee, Study on Molecular Mechanism of Intercalation of Anti-Cancer Drug Proflavine to DNA, International Meeting on Chemical Biology (IMCB-2013), 26-28 May 2013, IISER Pune.

6. **Wilbee D. Sasikala** and Arnab Mukherjee, Single Water Entropy: Hydrophobic Crossover and Application to Drug Binding, International Symposium on the Recent Advances in Modeling Rare Events (Rare2014), 29th May – 1st May 2014, Kerala, India.

Annexures

Annexure 3

1. Superposition of the equilibrated crystal structure of the intercalating part of DNA with the two stable intercalated states, corresponding to different orientations of proflavine, obtained from the free energy calculation.

Here non-hydrogen atoms of the base pairs C6-G7 were considered for RMSD calculation. RMSD of stable C6-G7 base pairs where amine groups of intercalated proflavine faces towards major groove is 0.9 Å. RMSD of stable C6-G7 base pairs where amine groups of intercalated proflavine faces towards minor groove is 1.0 Å.

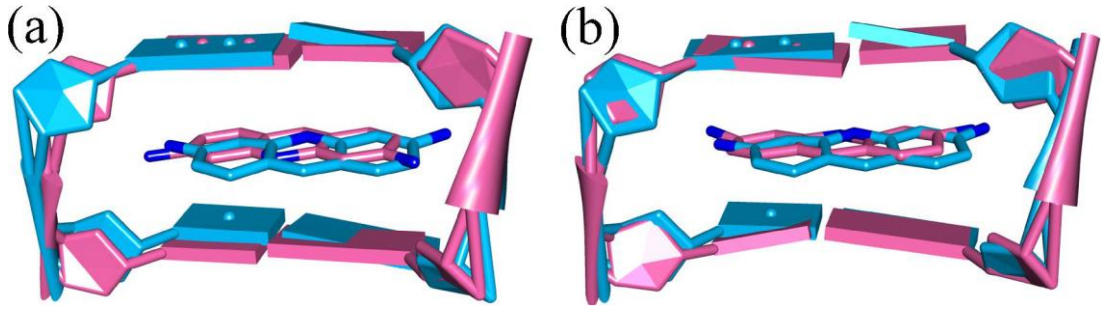


Figure A-3.1. Superposition of the equilibrated crystal structure of the intercalating part of DNA with the two stable intercalated states. The intercalated relaxed crystal structure of base pairs with proflavine is shown as deep cyan colored ribbon structure with amine groups of proflavine (deep cyan colored stick model) facing major groove side. Figure A1a shows C6-proflavine-G7 complex in pink colored ribbon structure with amine groups of proflavine (stick model with pink color) facing minor groove side which is opposite to that of crystal structure orientation. Figure A1b shows C6-proflavine-G7 complex with pink colored ribbon structure with amine groups of proflavine (pink colored stick model) facing the major groove side. The amine groups of proflavine are shown as element based stick model.

2. Analysis of deintercalation from major and minor groove sides: Figures A2 and A3 show the analysis of deintercalation from major groove and minor groove sides respectively.

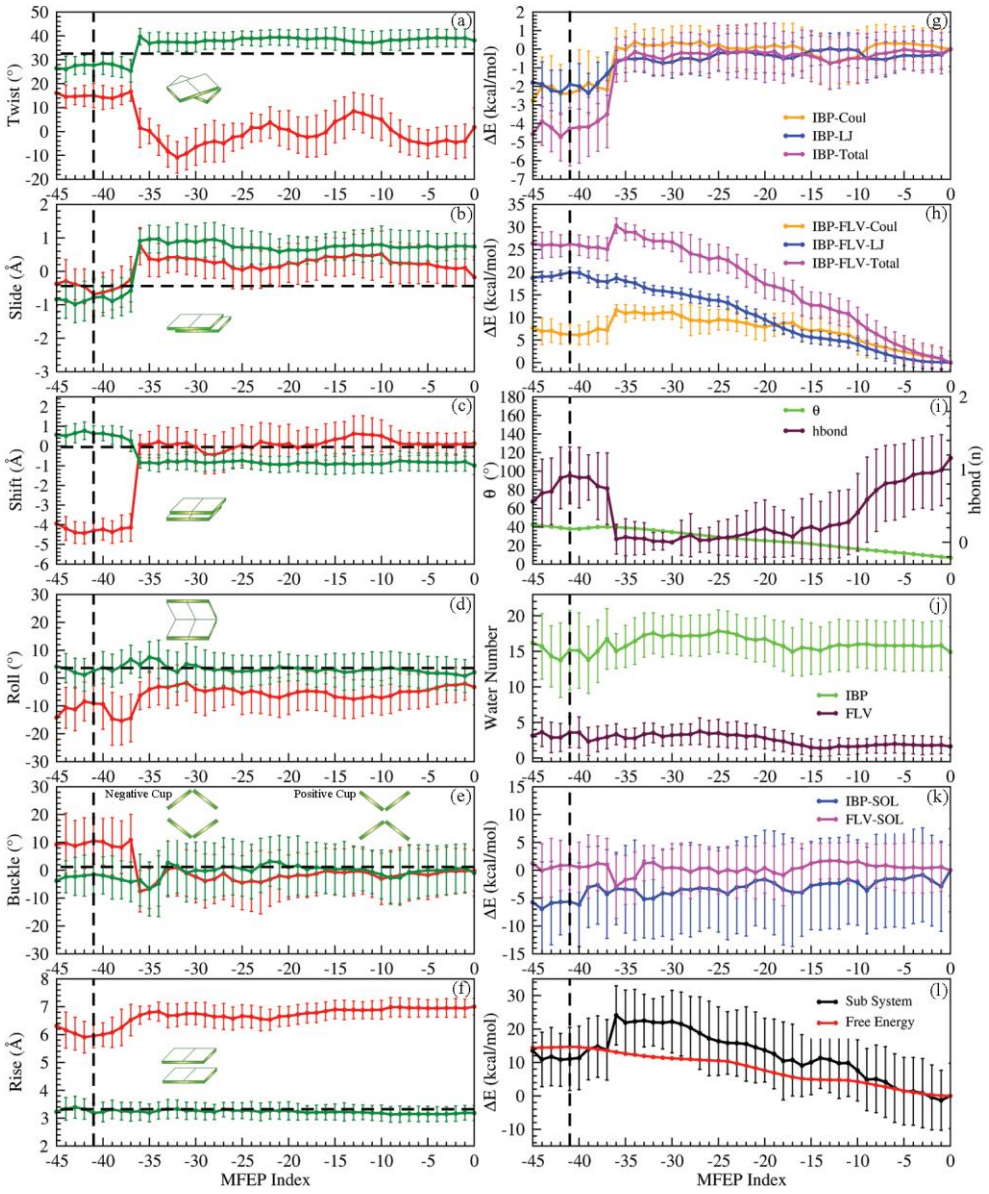


Figure A-3.2: Analysis of deintercalation process through major groove along MFEP. The vertical black dashed line represents the transition state (MFEP index -41). The energy values are taken to be zero at the intercalated state (MFEP index 0). Rest is same as Fig. 3.9. At the TS, base pair Twist increases and the corresponding Shift of the bases leads to opening of the base pairs towards the major groove side represented by negative Roll. Drug's exit is reflected by the decrease in Rise. As the drug starts coming out, (g) the interaction energy between base pairs decreases and (h) the energy between drug and base pairs increases. The low value of θ at TS indicates deintercalation happens with amino groups of the proflavine facing the bases. Main contribution of energy barrier come from the loss of stacking interaction between IBP and FLV where other interactions (IBP stacking interaction or IBP-water interaction) lead to energetic stabilization to reduce the energy barrier.

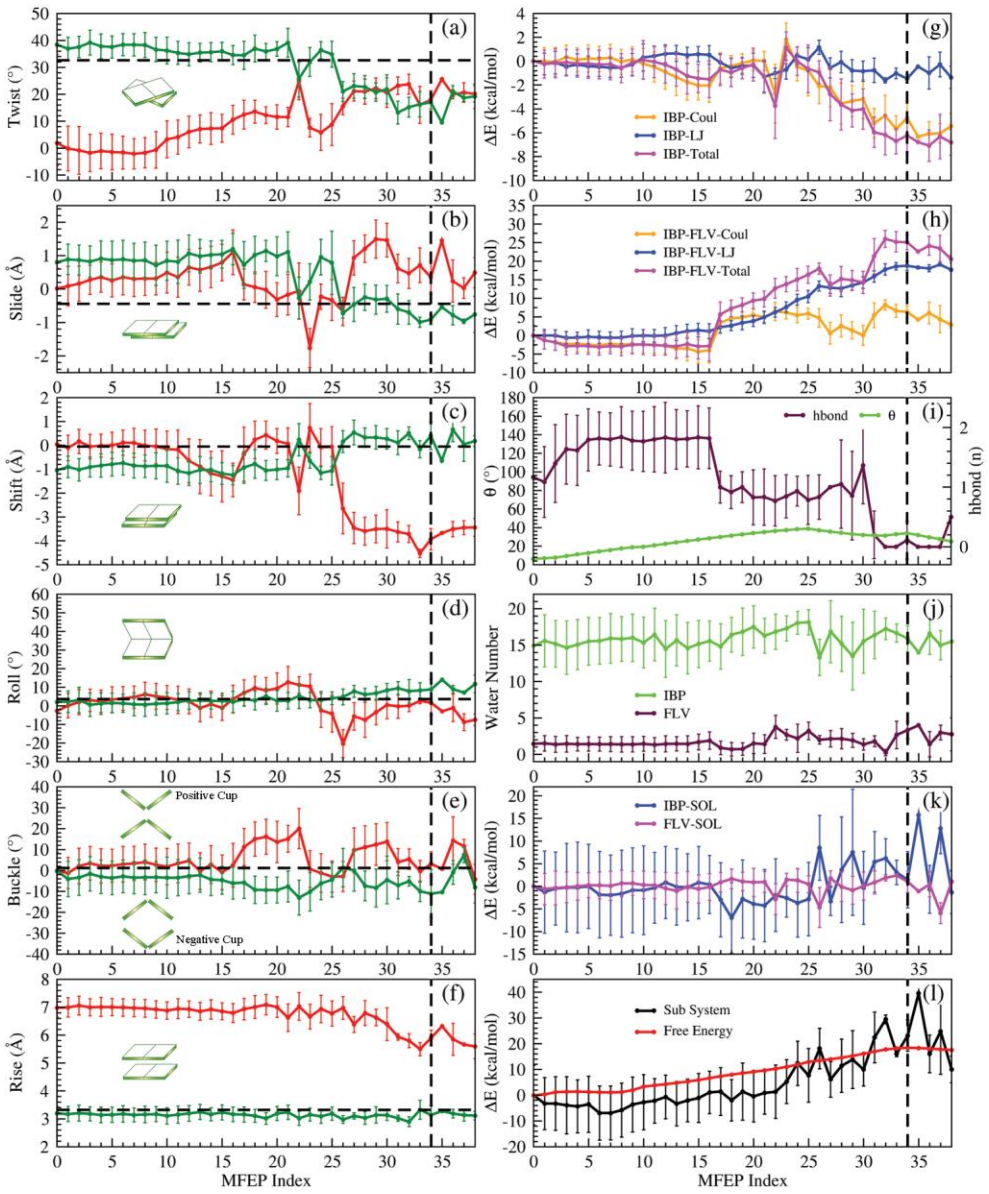


Figure A-3.3: Analysis of deintercalation process through minor groove along MFEP. Transition state is now at MFEP index 34. The energy values are taken to be zero at the intercalated state (MFEP index 0). Rest is same as Fig. 3.9. DNA parameters change similar to deintercalation towards major groove. (g) As the drug goes out, the distance between the two base pairs decreases and this leads to decrease in stacking energy where the main contribution comes from Coulombic interactions. (h) The energy between drug and base pairs increases again due to loss of stacking and decrease in hydrogen bond. (i) The low value of θ represents that the amine groups of the drug faces water during its exit from the base pairs. At TS, energy of IBP-water and drug-water is zero indicating negligible desolvation energy cost for deintercalation through minor groove side. Therefore, the main contribution of the energy cost for deintercalation through minor groove comes from the decrease in the interaction between the drug and DNA.

3. Analysis of dissociation from major and minor groove-bound states: Figures A4 and A5 show the analysis of dissociation of proflavine from major groove and minor groove bound states respectively.

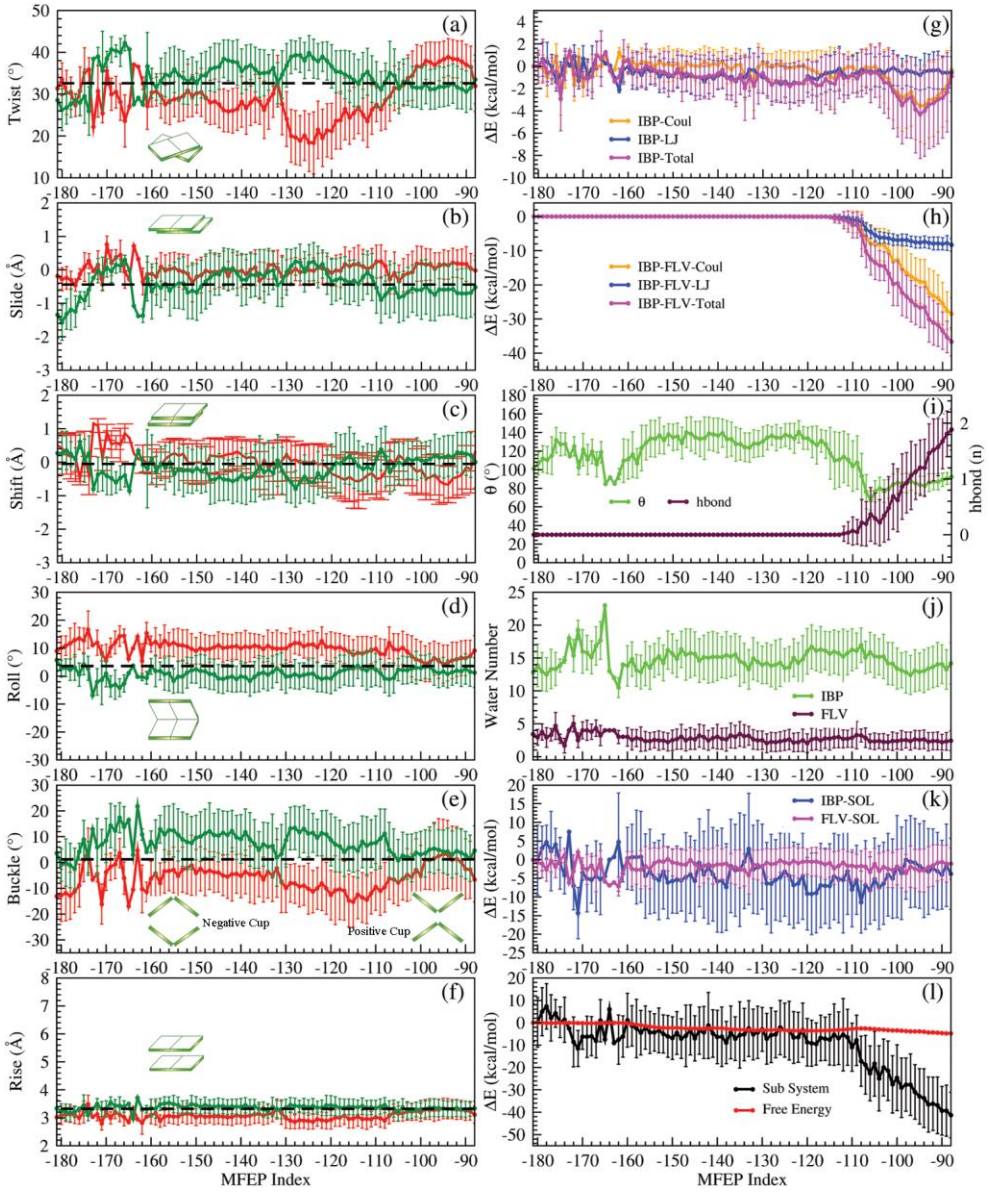


Figure A-3.4: Analysis of dissociation process from major groove-bound state along MFEP. The energy values are taken to be zero at completely dissociated state (MFEP index -181). MFEP index -88 represents major groove bound state. Rest is same as Fig. 3.9. During the dissociation from major groove, there is no significant change in the DNA parameters along MFEP. In the major groove bound state, the energy between drug and base pairs is very low where the main contribution comes from electrostatic interaction through hydrogen bonding. As the drug goes out, it loses hydrogen bonding, thereby increasing the energy between the drug and base pairs.

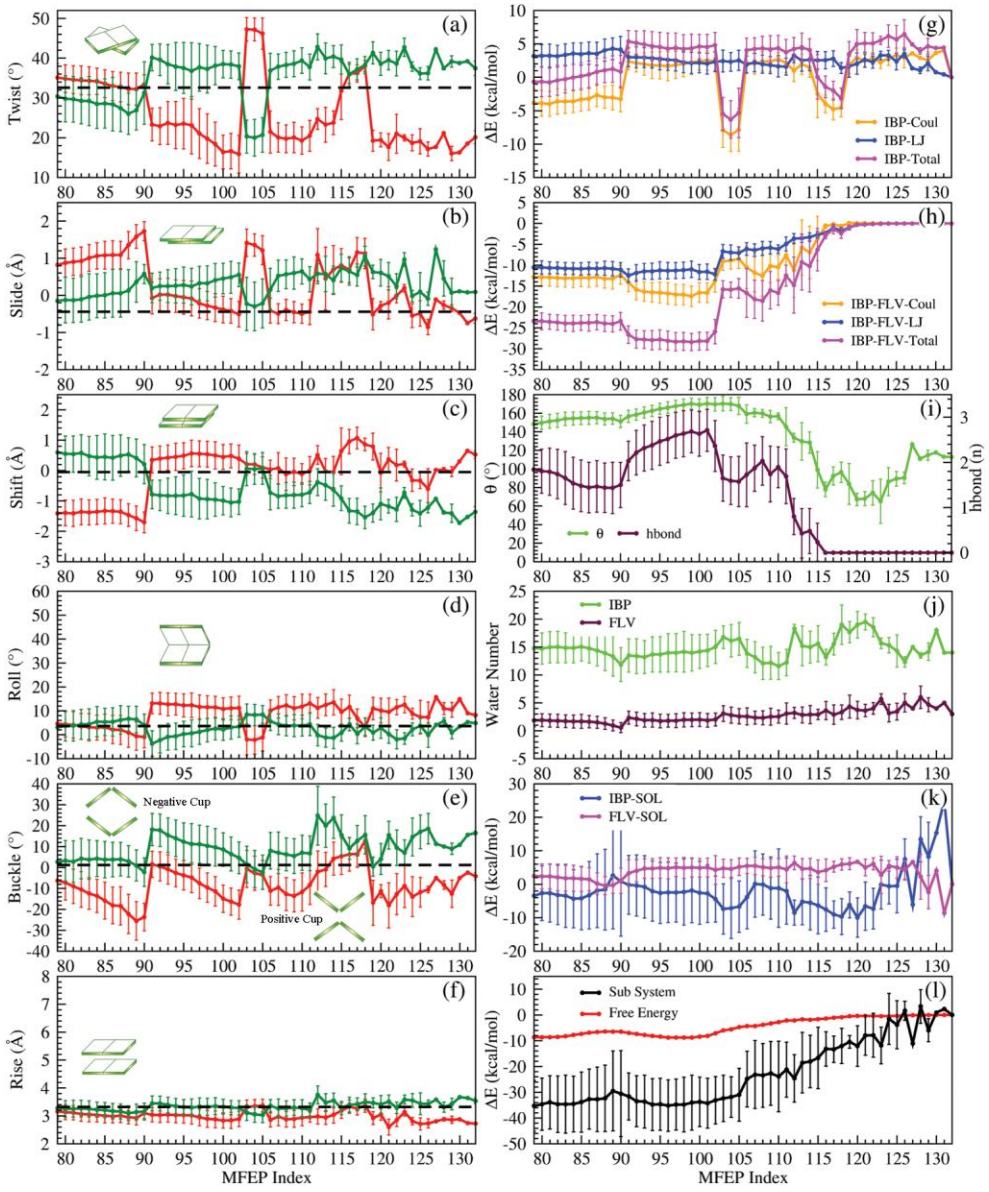


Figure A-3.5: Analysis of dissociation process from minor groove-bound state along MFEP. The energy values are taken to be zero at completely dissociated state (MFEP index 132). MFEP index 79 represents minor groove bound state. Rest is same as Fig.3.9. As in the case of dissociation from major groove-bound state, the dissociation from minor groove-bound state does not affect the DNA parameters significantly. As the drug moves out of the minor groove, the energy between IBP and drug increases, the main contribution of this comes from the Coulombic energy. The increase in Coulombic energy is due to the loss of hydrogen bonding between IBP and drug, which in turn is due to the change in θ from ~160˚ to ~60˚ indicating that the amine groups now face the water.

4. Estimation of error for intercalation from major groove side: To find out error for intercalation from major groove side, we performed multiple well-tempered metadynamics simulations, as direct error estimation is not applicable in well-tempered metadynamics[1]. We started the simulation with different initial velocity distributions. Figure A6 shows the additional free energy surface for intercalation of proflavine from major groove side.

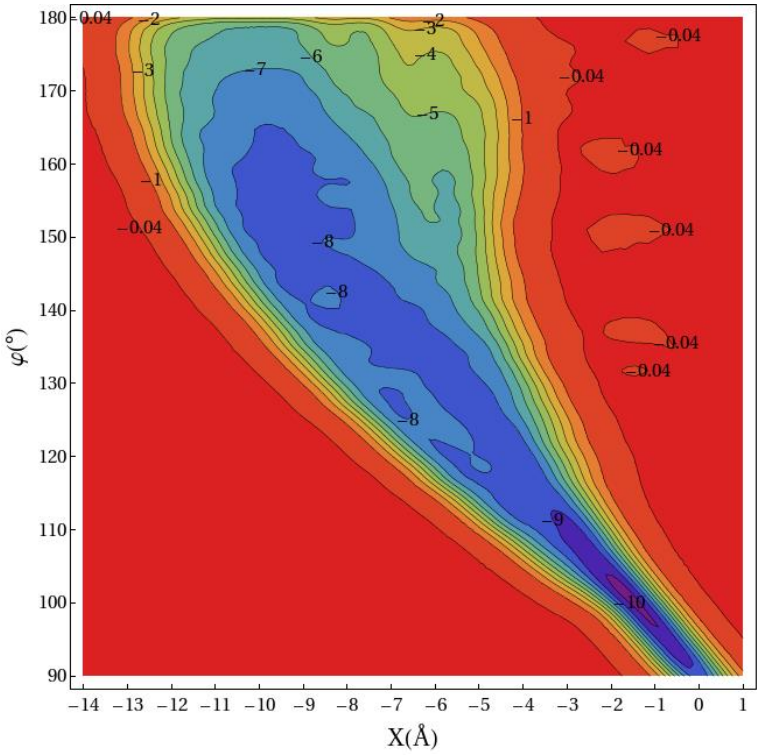


Figure A-3.6. Free energy surface for intercalation from major groove side along X and φ. Free energy values are written on the contour lines in kcal/mol. Height of added Gaussians is 0.048 kcal/mol and width (σ) of Gaussian for X is 0.5 Å and for φ is 0.2 radians with a bias factor 15. The free energy barrier for intercalation is ~ 10 kcal/mol, which is comparable with 8.8 kcal/mol shown in the main article for intercalation barrier from major groove.

5. References

1. W. D. Sasikala and A. Mukherjee, *J. Phys. Chem. B*, 2012, **116**, 12208-12212

Annexure 5

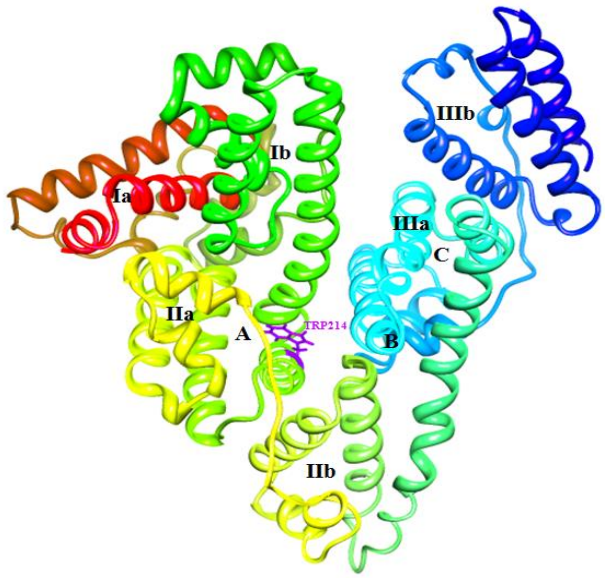
Table A-5.1. Percentage of anharmonic contribution to the translational entropy for the closest water molecule around different solutes. Anharmonic contribution for the bulk water is 0.49%.

System	TS_{tr}^{AH} (%)
Water	0.87
MHF_{σ}	0.51
$MHF_{2\sigma}$	0.33
$MHF_{3\sigma}$	0.28
$MHF_{4\sigma}$	0.20
$MHF_{5\sigma}$	0.21
$MHF_{6\sigma}$	0.18
M_{σ}	1.20
M_{σ}+1	1.80
M_{σ}+2	3.55
M_{σ}-1	2.82
M_{σ}-2	5.77
MHF_{σ}-0.5	0.69

1. Comparative Study of Flavins Binding with Human Serum Albumin: A Fluorometric, Thermodynamic, and Molecular Dynamics Approach

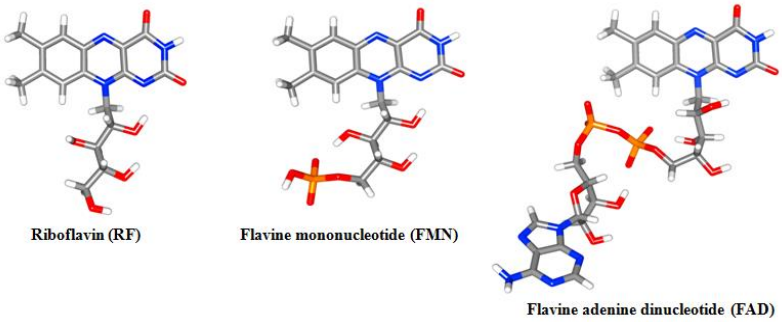
1. Introduction

Flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN) are derivatives of riboflavin (RF), a water-soluble vitamin, more commonly known as vitamin B2. Flavins have attracted special attention in the last few years because of the recent discovery of a large number of flavoproteins. In this work, these flavins are used as extrinsic fluorescence markers for probing the microheterogeneous environment of a well known transport protein, human serum albumin (HSA). Steady state and time-resolved fluorescence experiments confirm that both FMN and FAD bind to the Sudlow's site-1 (SS1) binding pocket of HSA, where Trp214 resides. In the case of RF, a fraction of RF molecules binds at the SS1, whereas the major fraction of RF molecules remains unbound or surface bound to the protein. Different binding sites on human serum albumin are shown in Scheme 1 and the chemical structures of FAD, FMN and RF are shown in Scheme 2.



Scheme 1. *Molecular structure of HSA, obtained after 10 ns of simulation with labels on six subdomains. A: Sudlow's site-1 (SS1), B: site close to Lys444 (SL), C: Sudlow's site-2 (SS2).*

Docking and molecular dynamics simulations are essential tools to understand the molecular interaction and dynamics of two bound species. To verify the experimental observations of FMN and FAD binding to SS1 and only a fraction of RF binding to the same SS1 pocket of HSA, we employed docking studies of flavins with HSA followed by molecular dynamics (MD) simulation study, detailed outcome of which is presented here study. We have first performed 10 ns simulation of the protein alone in water to observe its stability.



Scheme 2. *Quantum chemically optimized structures of the flavins: Riboflavin (RF), Flavin mononucleotide (FMN) and Flavin adenine dinucleotide (FAD).*

2. Experimental Seciton

The computational study comprises molecular dynamics of the protein and the complex created by docking. The crystal structure of HSA (PDB ID 1AO6) was obtained from Protein Data Bank with a resolution of 2.5 Å.[1] Crystal structure of HSA contains A and B polypeptide chains. We selected chain A for the present study. While we used GROMACS[2] to perform the molecular dynamics simulation, we used modified AMBER99SB force-field[3] to generate the topology and parameter of HSA.

We have added the missing hydrogen atoms after removing B chain, and performed vacuum minimization of the protein with no constraints by using steepest descent method for 1000 steps[4] followed by conjugate gradient method[5] for 1000 steps. The optimized protein structure was then solvated with water in a cubic box using tip4pEW water model[6]. We have added 15 sodium ions to neutralize the system. After adding water and ions, the system was again energy minimized using conjugate gradient method and heated up to 300 K during 100 ps of simulation time using V-rescale[7] thermostat at constant pressure of 1 bar using Parrinello-Rahman[8] barostat. Temperature

coupling constant was kept at 0.4 ps, while pressure coupling was kept at 0.6 ps. During the heating, heavy atoms of the protein were restrained by a harmonic force with a force constant 25 kcal/mol. After heating, equilibration was performed at constant temperature and pressure for 100 ps. Finally, molecular dynamic simulation was carried out for 10ns at constant temperature of 300 K using nose-hoover thermostat.[9] The ribbon diagram of optimized structure of HSA after 10 ns of simulation is shown in Scheme 1, where six different sub domains are shown in different colors.

Following the usual method for creating the general amber force field (GAFF),[10] we have optimized the flavins using Hartree Fock theory with 6-31G* basis set using Gaussian03,[11] which is followed by calculation of Merz-Kollmann ESP charges.[12] Finally restrained electrostatic potential (RESP) charges[13] for each atom were calculated using AMBER tools.[14] The total charge of RF, FMN and FAD is 0, -1 and -2 respectively. The quantum chemically optimized structures of the three flavin molecules used in this study are shown in Scheme 2. Preliminary docking studies with predefined charges showed that FAD binds to HSA mainly in the 'unstack' or extended conformation. Hence, we used an extended initial conformation of FAD for optimization and subsequent charge calculation using quantum chemistry calculation.

The final structure of HSA obtained at the end of 10 ns simulation was used as receptor and the three flavin molecules (RF, FMN and FAD) were docked into it using AutoDock[15] software. AMBER99SB charges of the protein and quantum chemically derived charges for the flavin were kept unchanged during docking. Initially, the whole protein was used to search for a binding site with a coarse grid spacing 0.16 Å. Genetic algorithm[16] was used as a search method. Out of the 10 different docked structures for each flavin, the best one in terms of binding free energy was selected for subsequent molecular dynamic simulation.

The best-docked structures for each of the HSA-flavin complexes were subjected to similar initial preparation (vacuum minimization for 10000 steps, followed by solvation withtip4pEW water and neutralization by addition of Na^+ ion), energy minimization, and position restraint. Finally 4 ns simulation was performed at constant temperature 300 K and pressure 1 bar with integration time step 1 fs. Thermostat and barostat at each step were same as described above. During the simulation, FMN was

unstable due to terminal phosphate group. Therefore, following the protocol of Case *et al.*,[17] we joined terminal hydrogen atom with the phosphate atom through a bond such that the terminal P-O-H is confined like rigid TIP3P water model.

3. Results and Discussion

We have first performed 10ns simulation of the protein alone in water to observe its stability. The root mean square deviation (RMSD) of the full protein and its different domains are shown in Fig. 1(a), which suggests that the protein is well equilibrated as the RMSD value fluctuates around 2.5 Å for the last 4ns. Radius of gyration (Rg), which is a measure of the size of the protein, is shown in Fig. 1(b). The radius of gyration of HSA also fluctuates around 26.68Å which is comparable to that of experimental value of 27.4±0.350 Å.[18] Note that neither the RMSD, nor the Rg show large deviation from the crystal structure as opposed to Ref.19 and in agreement with Ref.20.

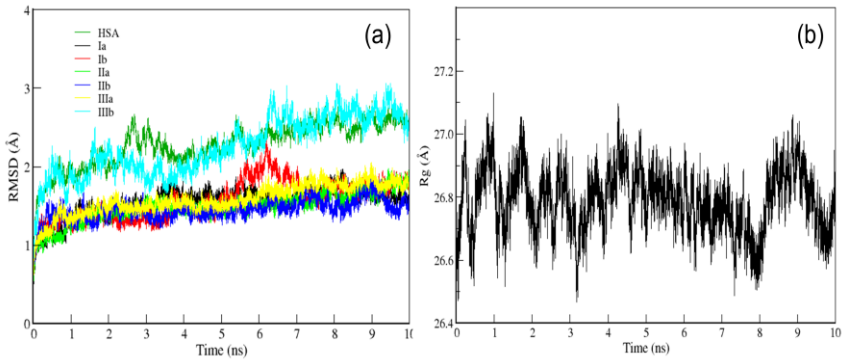


Figure 1: *(a) Time evolution of the RMSD for the whole protein (HSA) and its different domains from the crystal structure, (b) time evolution of the radius of gyration (Rg).*

To obtain preliminary information about the binding site, first we have searched the whole protein using Genetic algorithm in docking software AutoDock.[15] This preliminary docking showed that RF binds to three different binding sites, SS1, SL, and SS2 (A, B and C respectively in Scheme 1), while FAD and FMN show binding affinity only to SS1 site that accommodates the amino acid tryptophan (TRP214, see Scheme 1). To obtain best docked structures for all the flavins, we performed docking of each flavin to their preferred binding sites using more refined grid having 126 number of points in x,

y, and z direction with a spacing of 0.16 Å. Table 1 summarizes the binding energy and other parameters of all the docking studies.

Table 1: *Different interaction energies between flavin derivatives and the protein in the best docked structureobtained from AutoDock. Binding site for each is also mentioned.*

Flavin Derivatives	Binding site	Binding Energy(kcal/mol)	Vdw-hb-desolvation energy (kcal/mol)	Electrostatic interaction energy (kcal/mol)	Tortional Energy(kcal/mol)
RF	SS1	-3.19	-4.32	-1.56	2.68
RF	SL	-2.32	-3.3	-1.7	2.68
RF	SS2	-0.86	-2.09	-1.45	2.68
FMN	SS1	-1.82	-3.3	-2.1	3.28
FAD	SS1	2.74	-0.72	-2.21	5.67

We used rigid docking method, which allows flexibility of the ligands only and not the receptor, to find out the best docked structures for the flavins. Even flexible docking allows only certain torsional degrees to be varied during docking. Therefore, docking method does not allow flexibility of the receptor and ligands in a dynamical way. To obtain more accurate and flexible bound complex, we performed molecular dynamic simulations of HSA-flavin systems for 4ns starting with the best docked structure of each flavin molecule onto previously 10 ns equilibrated protein.

RF shows affinity towards three different sites of the HSA, out of which two are comparable as shown in Fig. 2. The binding energies of RF to the three sites are comparable (Table 1). Figure 2a shows RF bound to the SS1 pocket forming stacking interaction with TRP214. The hydroxyl groups of ribityl side chain of riboflavin forms intermolecular hydrogen bonding with GLY196, LYS199, GLU153 and the carbonyl group of isoalloxazine ring involves hydrogen bonding interaction with ARG218. The stability is also contributed by the electrostatic interaction between the ribityl side chain and basic amino acid residues present in the hydrophilic side of the binding pocket.

Figure 2b shows the same structure obtained at the end of 4 ns molecular dynamics simulation. This structure indicates better stacking interaction of RF with TRP214 compared to the docked complex. Fig. 2c shows the binding of riboflavin to the surface of HSA close to LYS444 at the SL site. This binding site is contributed by subdomains IIIa, IIa and IIb. The isoalloxazine ring is flanked by PRO339 and PRO447. Ribityl side chain forms hydrogen bonding with GLU297, GLU294 and LYS444. Fig. 2d shows the same interaction obtained after 4 ns simulation, and it indicates that the surface binding pocket (SL site) of riboflavin (close to LYS444) has also undergone conformational change to provide better interactions with the side chain of riboflavin. Here, the methyl side chain of isoalloxazine ring forms hydrophobic interaction with LEU407 and ribityl side chain forms hydrogen bonding with ARG410 and LYS414.

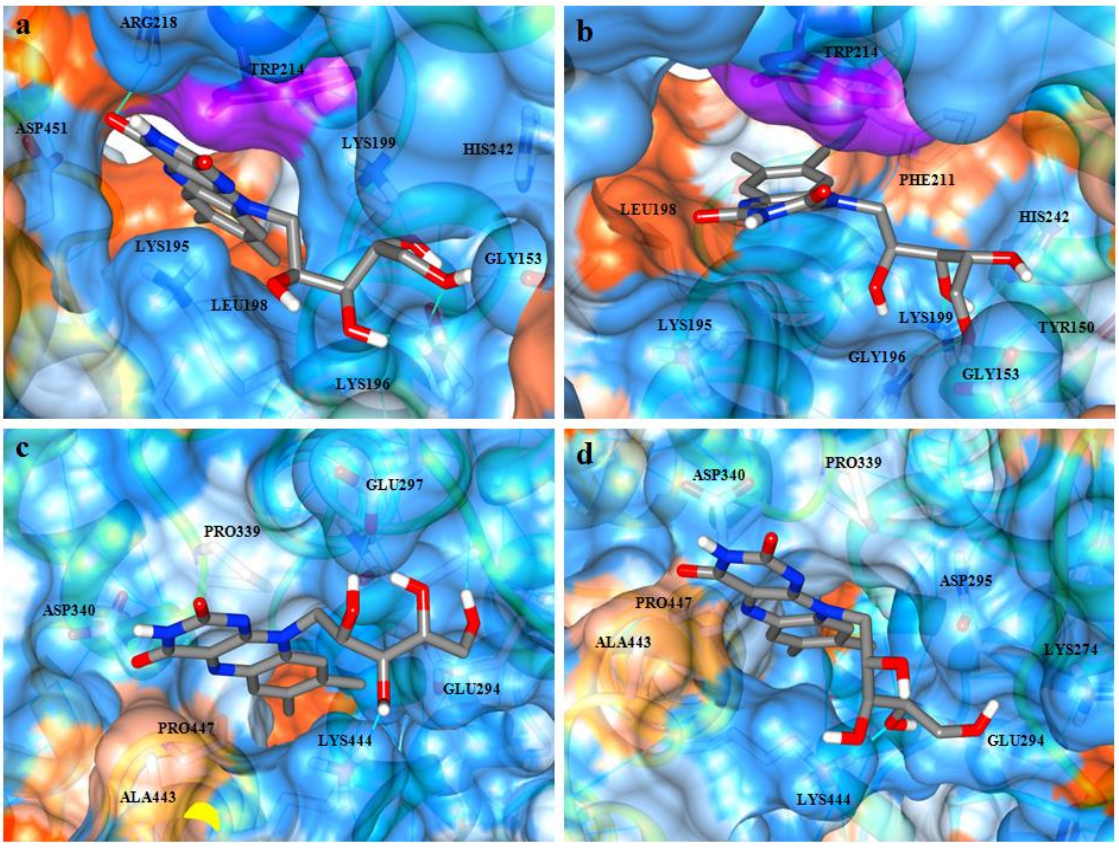


Figure 2: *(a) RF bound to SS1 after docking (b) RF bound to SS1 after 4ns simulation. (c) RF bound to SL after docking (d) RF bound to SL after 4ns simulation. RF is given element-based color. The amino acids are colored according to hydrophobicity except TRP214 which is shown in violet color. The picture is plotted using chimera.*[100]

Binding of RF to SS1 can lead to strong stacking interaction between tryptophan (TRP214) and isoalloxazine ring of RF, and this strong stacking interaction may result in formation of almost dark complex between the two, as they may be involved in intermolecular electron transfer process. However, binding to the other two sites will not affect the fluorescence of the RF (since these sites are away from TRP214). Steady state fluorescence experiment confirms almost 38% reduction in fluorescence of RF on mixing with HSA. Therefore, presence of three comparable binding sites of RF results in binding a fraction of RF molecules to SS1 site, which leads to only a moderate quenching observed in the fluorescence experiment.

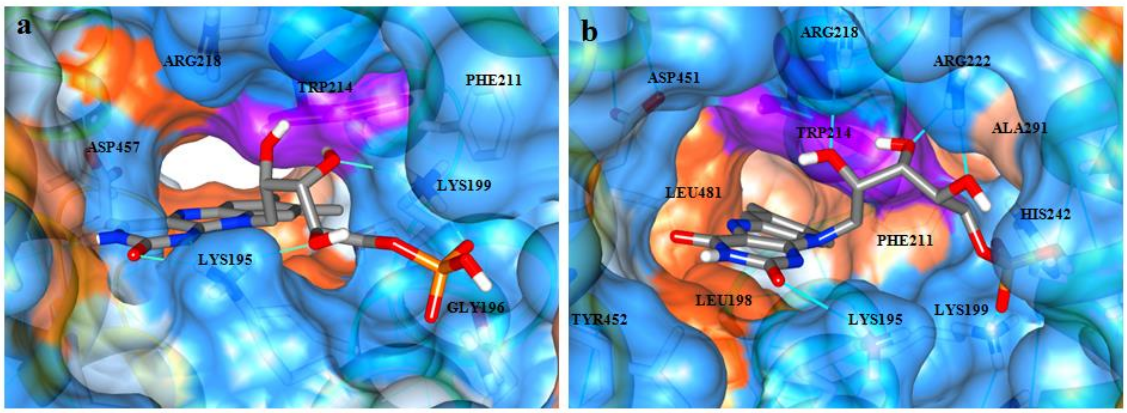


Figure 3: *(a) FMN bound to SS1 after docking (b) FMN bound to SS1 after 4ns simulation. Other details are same as Fig. 2.*

Figure 3a shows the binding of FMN to SS1binding pocket of HSA only. Here, the negatively charged FMN molecule is more towards the hydrophilic region of the binding pocket and makes the interactions with surrounding positively charged amino residues such as ARG218 and LYS199. Like RF, FMN forms stacking interaction with TRP214 which is responsible for the drastic decrease in fluorescence intensity. The methyl groups of isoalloxazine ring face towards hydrophobic part of the SS1. Moreover, the isoalloxazine ring forms hydrogen bonding with ASP451 and LYS195, whereas ribityl side chain involves hydrogen bonding interaction with LYS195. The stability is also increased by electrostatic interaction between terminal phosphate group of FMN and surrounding two positively charged amino acid residues (LYS199 and GLY196). Consequently, the electrostatic interaction energy of FMN is more negative than RF(Table 1). Figure 3b shows the same structure after 4ns of molecular dynamics

simulation. In the case of FMN, SS1 has undergone more conformational change in order to involve in increased stacking interaction with TRP214. This also increases the electrostatic interactions between the basic amino acid residues and negatively charged side chain of FMN as shown in Fig. 3b.

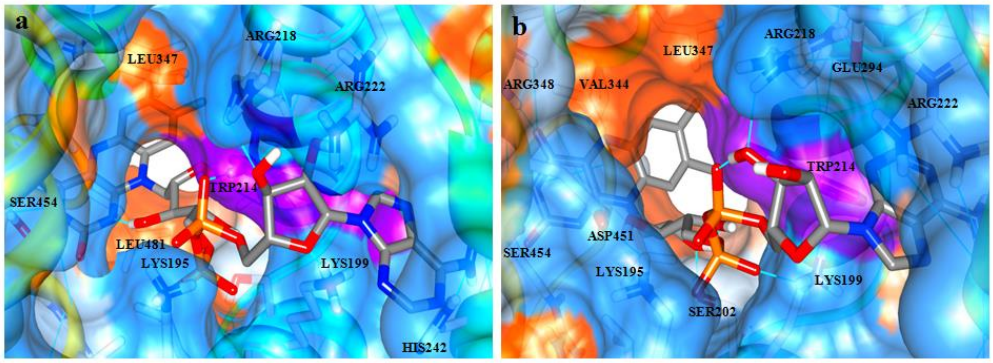


Figure 4: *(a) FAD bound to SS1 after docking (b) FAD bound to SS1 after 4ns simulation. Other details are same as Fig. 2.*

Figure 4a displays the binding of FAD to SS1 site of HSA. FAD binds to this binding pocket only in the extended conformation. Here, the isoalloxazine ring points more towards hydrophobic portion, whereas methyl groups face towards LEU347.The ribityl side chain directs towards the hydrophilic region of the binding pocket. Interestingly, here the isoalloxazine ring is on one side of the tryptophan residue and the adenine ring on other side of tryptophan. The carbonyl groups of the ring make hydrogen bonding with SER454. Moreover, the stability is increased by hydrogen bonding between ribityl side chain and ARG218, ARG222, LYS195 and LYS199. The adenine ring of FAD is stabilized by interaction with HIS242. Furthermore, the stability is gained by the electrostatic interactions between two negatively charged phosphate groups and positively charged surrounding amino acid resides. However, total binding energy obtained for FAD from AutoDock is positive. The reason for the positive binding energy lies in the unusually high torsional energy of (5.7 kcal/mol) of the drug (see Table 1). It should be mentioned here that docking energy is only empirical in nature; moreover, due to absence of dynamics, flexibile optimization is also not possible. Therefore, molecular dynamics simulation is required to properly optimize the bound molecules which has been done here for all the flavins. Fig. 4b shows the FAD bound to SS1 after 4 ns

simulation. Here, conformational change has taken place to form hydrophobic interactions with VAL344 and LEU347 and adenine ring of FAD is shifted more towards TRP214.

It is clear from Fig. 4b that isoalloxazine ring is on one side of tryptophan residue and adenine ring on the other side of tryptophan. Here the ring is close to TRP214 but the distance and orientation of the isoalloxazine ring is not suitable for stacking interactions like RF and FMN. However, the closeness of the flavin ring towards the tryptophan residue may be responsible for fluorescent quenching of FAD, albeit, lesser than FMN. Although RF has better stacking interaction than FMN and FAD, it has other competitive binding sites where quenching is not possible. Hence, overall RF exhibits trifling quenching than that of FMN and FAD.

Table 2: *Short range Coulombic and Lennard-Jones Potential between HSA and derivatives of flavin in different systems.*

System	Coul-SR (kcal/mol)	LJ-SR (kcal/mol)
HSA-RF (SS1)	-37.46±0.105	-31.93±0.050
HSA-RF (SL)	-14.41±0.048	-21.764±2.318
HSA-FMN	-51.77±0.118	-36.921±0.027
HSA-FAD	-77.605±0.085	-77.975±0.0451

Table 2 shows the average non-bonding interaction energy (electrostatic interaction and van der Waals interaction) between HSA and flavins obtained from the last 2 ns of molecular dynamics simulations. It is clear from the results that RF binds to SS1 pocket stronger than SL in terms of both the van der Waals and electrostatic interaction. Both FMN and FAD show more electrostatic interaction energy compared to RF because of the presence of negatively charged terminal phosphate groups in the formers. In case of FMN, electrostatic interaction predominates over van der Waals interaction. Among FMN and FAD, later shows stronger electrostatic interaction with the basic amino acids present in the SS1 because of the presence of two phosphate groups in

FAD. For binding of FAD in SS1 site, both electrostatic and van der Waals interactions contribute similarly.

It should be noted that the binding energy obtained from Docking and molecular dynamics simulation are different due to the different force-field used in both cases. Docking uses empirical energy function based on specific atom types and fixed interaction energy, however, tuned to match qualitative experimental trend. Molecular dynamics, on the other hand, uses quantum chemically derived all atom force-field and uses distance dependent energy functions. Therefore, the energy values obtained from two above methods are not comparable. However, they can be used to understand the qualitative aspects of the interaction between the protein and the ligand (flavins). Moreover, in molecular dynamics simulation, we provided only the average potential energy, not the free energy as the later would require much more extensive set of simulations and may be beyond the scope and the goal of the present article.

Table 3: RMSD and Radius of Gyration of HSA chain in different HSA-flavin systems.

System	RMSD (Å)	Radius of Gyration (Å)
HSA	2.09±0.087	26.89±0.095
HSA-RF (SS1)	2.52±0.169	27.19±0.097
HSA-RF (SL)	2.60±0.119	27.10±0.101
HSA-FMN	2.63±0.120	27.13±0.096
HSA-FAD	2.32±0.128	27.26±0.114

Average RMSD and radius of gyration of HSA in different systems are shown in Table 3. RMSD after 2.5ns simulation fluctuates around 2 Å in all the systems indicating the stability of systems and shows the reliability of the docking. The RMSD average for HSA in simulation of HSA alone is 2.09 Å as shown in Table 6. RMSD's for protein-flavin complexes are also within permissible range of molecular dynamcis simulation.

The conformational fluctuation of each of the amino acids in the binding site (SS1) is analysed using root mean square fluctuation (RMSF) values. Figure 5(A),

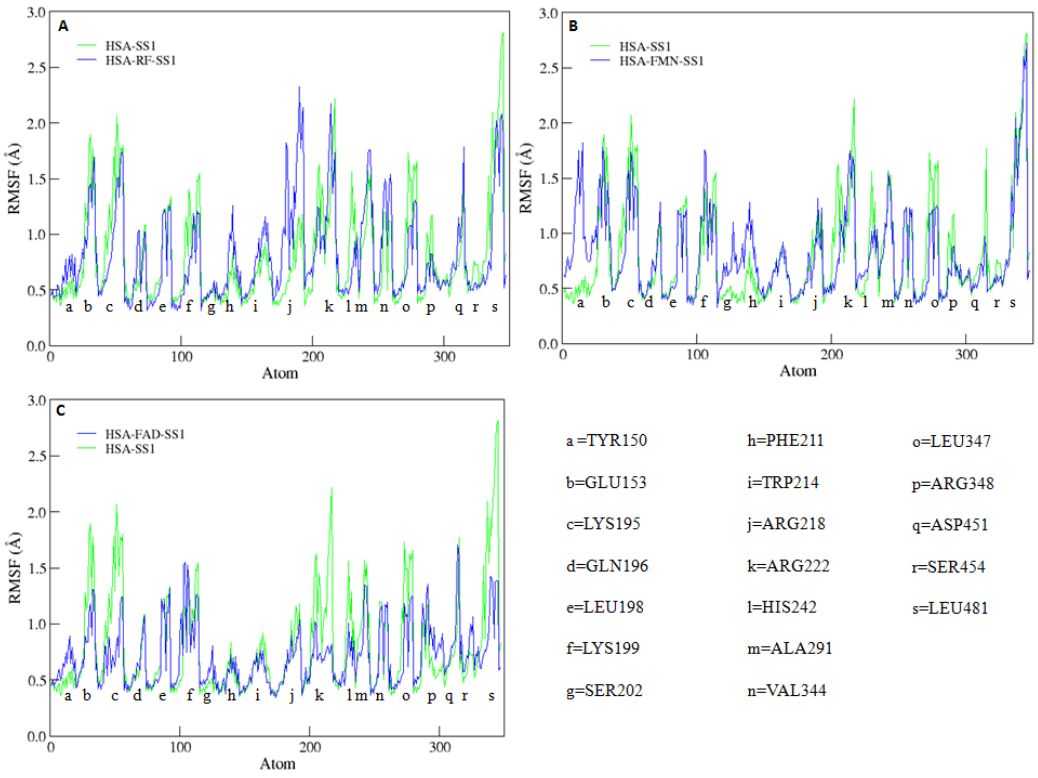


Figure 5: *RMSF of amino acids of active site SS1 in A) HSA-RF, B) HSA-FMN and C) HSA-FAD systems in comparison with RMSF of amino acids of SS1 in HSA only. Amino acids present in the SS1 binding pockets are also mentioned.*

5(B) and 5(C) depicts RMSF of HSA complexed with RF, FMN and FAD, respectively compared to that of the uncomplexed HSA. In RF bound protein, fluctuation of amino acids PHE211 (h), TRP214 (i) and ARG218 (j) is more than the uncomplexed structure. Here, the conformational change of TRP214 leads to better stacking interaction with RF as compared to the docked structure. Figure 5(b), the FMN-HSA complex, shows high RMSF of LYS199(f) because conformational change has taken place to change the position nearer to the phosphate group of FMN to form hydrogen bonding (Initially LYS199(f) was close to PHE211(h) which reflected as blue color in the hydrophobicity surface of PHE211 in Fig. 3(a)). Also, PHE211(h) underwent high flexibility to increase the hydrophobic interaction with isoalloxazine ring of FMN. The strong hydrogen bonding interactions of two hydroxyl groups of side chain of FMN with ARG222(k) decreases the RMS fluctuation. The hydrophobic interactions of LEU347(o) with flavin ring of FMN and PHE211(h) decreases the RMS fluctuation (o). Figure 5(c) shows

overall decrease of RMS fluctuation of amino acids in SS1 bound to FAD. The decreased RMSF values of LYS195(c), ARG218(j), ARG222(k) and HIS242(l) indicate strong electrostatic interaction with FAD. LEU481 shows strong hydrophobic interaction with flavin moiety of FAD, and hence, the RMSF value of LEU347 (o) is decreased.

4. Conclusions

The binding interactions of flavins with HSA have been studied with the help of steady state and time-resolved fluorescence spectroscopy, isothermal titration calorimetry, and circular dichroism studies and supported by docking and molecular dynamics simulations. Flavins have been documented for the first time as an extrinsic fluorescence marker for a model transport protein HSA. Steady state and time-resolved fluorescence studies confirm that major fraction of FMN and FAD molecules bind to the SS1 binding pocket of HSA. For RF, a small fraction binds strongly with TRP214 in the SS1 binding pocket of HSA and forms a dark complex with TRP214. The remaining fraction of RF molecules either reposes on protein surface or remains unbound. The thermodynamic parameters like free energy, enthalpy and entropy change of the binding interactions between flavins and HSA were determined with the help of ITC study. The sites of binding for these yellow pigments are determined with the help of molecular dynamics simulation as well as docking studies. Moreover, the experimental results are supported by docking and molecular dynamics simulation studies.

5. References

1. Sugio, S.; Kashima, A.; Mochizuki, S.; Noda, M.; Kobayashi, K., *Protein Eng.* **1999,** *12*, 439-446.
2. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C., *J. Comput. Chem.* **2005,** *26* (16), 1701-1718.
3. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., *Proteins: Structure, Function, and Bioinformatics* **2006,** *65* (3), 712-725.
4. Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T., *Numerical Recipes in FORTRAN: The Art of Scientific Computing*. 2nd ed. ed.; Cambridge University Press: Cambridge, 1992.

5. Zimmermann, K., *J. Comput. Chem.* **1991,** *12* (3), 310-319.
6. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T., *J. Chem. Phys.* **2004,** *120* (20), 9665-9678.
7. Bussi, G.; Zykova-Timan, T.; Parrinello, M., *J. Chem. Phys.* **2009,** *130* (7), 074101-9.
8. Parrinello, M.; Rahman, A., *J. Appl. Phys.* **1981,** *52* (12), 7182-7190.
9. Nosé, S., *Phys. Rev. E* **1993,** *47* (1), 164.
10. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., *J. Comput. Chem.* **2004,** *25* (9), 1157-1174.
11. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A. J.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson,; G. A., N., H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M.,; Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian,; H. P., C., 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., Ayala, P. Y.,Morokuma,; K., V., G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels,; A. D., S., M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman,; J. B., O., J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu,; G., L., A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham,; M. A., P., C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen,; W., W., M. W., Gonzalez, C., Pople, J. A., *Gaussian Inc.* **2004,** *Revision C.02.*
12. Besler, B. H.; Merz, K. M.; Kollman, P. A., *J. Comput. Chem.* **1990,** *11* (4), 431-439.
13. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollmann, P. A., *J. Am. Chem. Soc.* **1993,** *115* (21), 9620-9631.
14. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., *J. Comput. Chem.* **2005,** *26* (16), 1668-1688.

15. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., *J. Comput. Chem.* **2009,** *30* (16), 2785-2791.
16. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., *J. Comput. Chem.* **1998,** *19* (14), 1639-1662.
17. Case, D. A. terminal phosphate group in RNA. http://ambermd.org/Questions/mail/251.html.
18. Kiselev MA, G. I., Dobretsov GE, Komarova MN, *Biofizika* **2001,** *46* (3), 423-7.
19. Sudhamalla, B.; Gokara, M.; Ahalawat, N.; Amooru, D. G.; Subramanyam, R., *J. Phys. Chem. B* **2010,** *114* (27), 9054-9062.
20. Deeb, O.; Rosales-Hernández, M. C.; Gómez-Castro, C.; Garduño-Juárez, R.; Correa-Basurto, J., *Biopolymers* **2010,** *93* (2), 161-170.

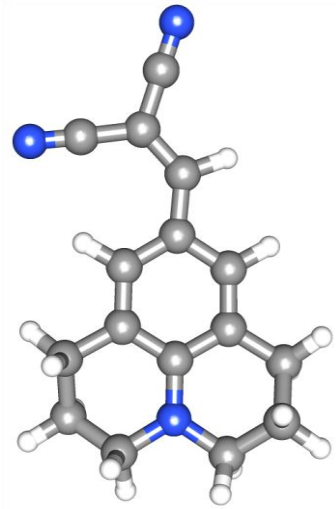
2. Modulation of Excimer Formation of 9-(dicyano-vinyl)julolidine (DCVJ) by the Macrocyclic Hosts

1. Introduction

Supramolecular hosts are good model system for nano-cavities due to its various size, shape and hydrophobicity.[1-7] They have been found particularly useful to monitor the formation of host-guest inclusion complex by incorporating fluorescent dyes inside their cavity.[1-7] The inclusion of the fluorescent guest into their cavity is accompanied by large change of fluorescent properties as a consequence of an altered microenvironment.[1-7] One typical example of such supramolecular host-guest inclusion complexes consists of cyclodextrin (CD) and guest fluorescent molecule.[1-5] Cyclodextrins (CDs) are cyclic oligosaccharides with 6, 7 and 8 D-glucose units, and they have lipophilic cavity with different inner diameters (5.7, 7.0, 8.5 Å for α-, β- and γ-CD, respectively), which provide a relatively hydrophobic cavity for guest binding.[1-3] The most remarkable feature of the CDs is their ability to form inclusion complexes with a variety of organic molecules. Although stoichiometry of the complex is usually 1:1 in aqueous solution, 1:2 host-guest inclusion complex is also feasible for higher cavity size.[8-12] This property of cyclodextrins is used to control the monomer-dimer equilibrium of dyes by monitoring their absorption and emission spectral properties.[8-12]

9-(dicyano-vinyl)julolidine (DCVJ) is a classic example of a molecular rotor (Scheme 1). A molecular rotor is a fluorescent molecule that has ability to undergo an intramolecular twisting motion in the excited state.[13,14] Unlike other molecular motor like, p-(dimethylamino)benzonitrile (DMABN), DCVJ always exhibits only single emission band even in polar solvent as the energy gap between S_1 and S_0 in the TICT state of DCVJ is almost 3 times smaller than the LE energy gap.[13] Hence, non-radiative relaxation from TICT is preferred over radiative relaxation.[13] Very recently, it has been reported that primary mechanism of non-radiative relaxation of DCVJ is isomerisation rather than TICT state formation.[15,16] DCVJ has been extensively used in biological applications. It has been reported that fluorescence quantum yield of DCVJ increases

when it binds to tubilin.[17] Moreover, during tubilin to tubules aggregation, further enhancement of DCVJ quantum yield takes place.[17] The transformation of actin from its monomeric G-state to its polymeric F-state is monitored with the help of fluorescence of DCVJ.[18] DCVJ has also been used as a probe for measurement of micro viscosity and free volume in phospholipids bilayers and micelles.[19,20]



Scheme 1: *Quantum chemically Optimized (B3LYP/6-31G*) structure of DCVJ represented in ball and stick model with element based colors*.

Several studies have been devoted to understand the excited state properties of DCVJ in different solvents,[16] polymers[21] and ionic liquids[22,23]. However, relatively few investigations have dealt with its excited state aggregation behaviour in water despite that fact that dimerization/aggregation has strong effect on emission spectra of DCVJ, which might have significant implications for analyzing protein structure and conformation by DCVJ. Bhattacharyya *et al*. first time noticed that emission spectra of DCVJ in water depend on concentration, and the intensity of longer wavelength emission band increases with increasing DCVJ concentration.[24] They also observed that longer wavelength emission band intensity increases with the increase in γ-CD concentration, and this was attributed to the formation of DCVJ excimer inside the γ-CD cavity.[24] However, the dynamics aspects, which provides deep perception to this excimer formation process, has not been addressed in their study. In this work, we continue our efforts in this direction to get insight about the detailed picture of the excimer formation dynamics of DCVJ in

presence of α-, β- and γ-CD with the help of steady state and time-resolved fluorescence spectroscopic techniques. Our studies reveal the formation of both 1:1 and multiple inclusion complexes depending on the cavity sizes of CDs. Finally, docking as well as some theoretical calculations have been employed to understand the molecular picture of this host-guest interaction. We think that this kind of stoichiometry as well as size based excimer formation drift of a molecule might have some applications towards fluorescence on/off switches.

2. Experimental Section

To obtain the molecular picture of orientation of DCVJ in the inclusion complexes with various CDs as well as to gain insight into the stabilization achieved due to encapsulation, we have docked the ligand (DCVJ) into the CDs, followed by a quantum chemical optimization. Crystal structures of α-CD (PDB ID: 2ZYM), β-CD (PDB ID: ZYN) and γ-CD (PDB ID: 2ZYK) were taken as receptor and optimized geometry of either monomeric or dimeric form of the DCVJ was taken as ligand. In all docking, we have used Gasteiger charges for both the CDs and the ligand. The grid was generated on whole receptor with grid points of 126 on each of the orthogonal directions with a spacing of 0.16 Å. Search was performed using Genetic algorithm using AutoDock (version 1.4) software.[25] During docking, the receptor was kept rigid and the ligand was flexible. Following the experimental evidence, we have also tried to find 1:2 or 2:2 inclusion complex (DCVJ:γ-CD) in case of γ-CD by docking the dimeric form of the ligand on to the host. Although experimental studies have indicated an excimer complex, we have made an approximation that geometry of the ground state dimer would reflect that of the excimer. To create the dimeric form of ligand, we have docked one DCVJ molecule on to the other using a smaller grid spacing 0.08 Å. The most stable DCVJ dimer complex obtained from docking was then optimized using ab initio density functional theory using WB97XD functional to account for the dispersion interaction, which stabilizes the stacked geometry of the DCVJ dimer, with 6-31G* basis set in Gaussian 09.[26]

3. Results and Discussion

For all 1:1 inclusion complexes of three macrocyclic hosts, the best docked structures were then optimized using ab initio density functional theory using B3LYP functional with 6-31G* basis set (Figure 1).[26] The binding energy was calculated by optimizing the

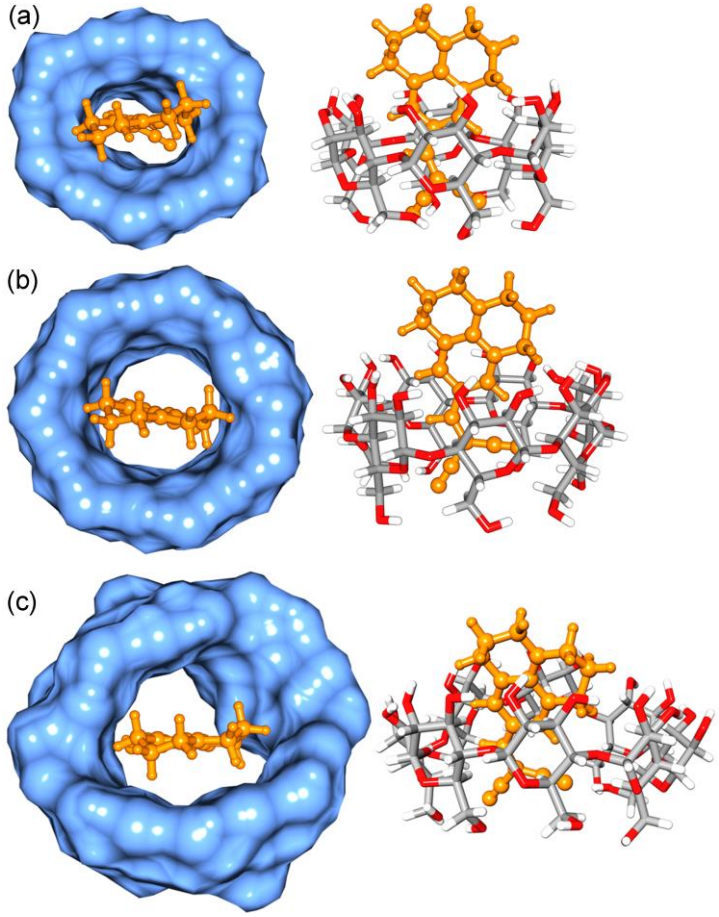


Figure 1: *Optimized structure of (a) docked α-CD:DCVJ complex (1:1), (b) docked β-CD:DCVJ complex (1:1) and (c) docked γ-CD:DCVJ complex (1:1). DCVJ is shown in ball and stick model (orange color) and CDs are shown in both surface representation (blue color) and stick model (element based color).*

complex, individual molecules, and then subtracting energy of the individual molecules from the complex. A full ab initio quantum chemical calculation for higher inclusion complexes (1:2 or 2:2) of γ-CD is computationally prohibitive and we believe that PM6 semi-empirical method would provide at least geometrically correct result for higher order inclusion complexes of γ-CD (Figure 2-3). Unlike, ab initio density functional theory, semi-empirical method does not provide binding energy. However, it provides the heat of formation (ΔH/kcal.mol^{-1}) for the inclusion complexes. Therefore, in case of

multiple inclusion complexes of γ-CD, instead of calculating the binding energy, we have provided the heat of formation of inclusion (ΔH/kcal.mol^{-1}) complex by optimizing the complex, individual molecules by PM6 method in MOPAC,[27] and then subtracting the heat of formation of the individual molecules from the complex. Since the docking as well geometry optimization were done without consideration of any solvent and other parameters, these geometries provide only qualitative picture of the structures in the ground states.

Although docking study indicates the formation of inclusion complex between DCVJ and α-CD (Figure 1a), the binding energy (after BSSE correction) obtained from quantum chemical optimization method is positive (32.27 kcal/mol), and it indicates that energetically DCVJ does not prefer to be included inside the cavity of α-CD probably due to smaller size of the host cavity. Therefore, theoretical calculation supports our experimental findings that DCVJ does not from inclusion complex with α-CD. In case of β-CD, the binding energy (after BSSE correction) calculated from quantum chemical optimization method is negative (-9.70 kcal/mol), inferring that 1:1 inclusion complex formation between DCVJ and β-CD is energetically feasible as we have observed experimentally (Figure 1b).

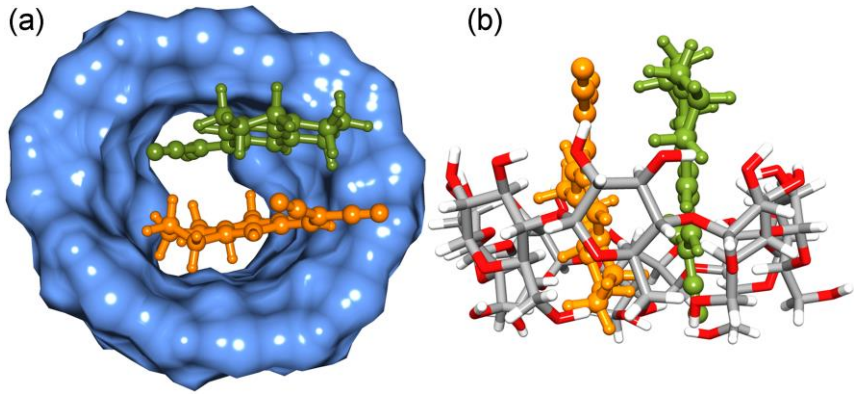


Figure 2: Optimized *structures of γ-CD:DCVJ (1:2) complex. (a) and (b) correspond to the different orientations of the same optimized complex. The DCVJs are represented as ball and stick model (ornage and green colors). Rest is same as that of Figure 1.*

For γ-CD, the 1:1 host-guest inclusion complex exhibits energetically feasible negative binding energy (-4.62 kcal/mol) obtained from ab initio quantum chemical optimization method (Figure 1c). Following the experimental evidence of formation of 1:2 inclusion complex (DCVJ:γ-CD) in presence of γ-CD, we have docked optimized

dimer of DCVJ to γ-CD host, and then optimized the docked structure of 1:2 inclusion complex (γ-CD:DCVJ) by semi-empirical calculation (Figure 2). Considering the experimental proof of formation of 2:2 complex in γ-CD, we have first created a dimeric complex of γ-CD by docking one γ-CD on to the 1:2 complex of γ-CD and DCVJ. This was done because docking the γ-CD on to an empty γ-CD resulted in an unusual interpenetrating ring structure where 2:2 complex won't be feasible. Subsequently, we removed the DCVJ dimer from the host and obtained a dimeric complex of γ-CD. This was followed by docking of DCVJ dimer to obtain the best docked configuration of the 2:2 complex of γ-CD and DCVJ complex. All the allowable torsions for DCVJ dimer were switched off during the docking. Evaluating heat of formation from PM6 semi-empirical calculations, we confirm that 1:2 and 2:2 stoichimetries, as shown in Figure 2-3, are energetically feasible by ΔH_f of -25 kcal.mol^{-1} and -67 kcal.mol^{-1}, respectively. Therefore, it confirms that higher stoichiometric complex is more favourable than the lower stoichiometry. It is also noticeable that inside the cavity two DCVJ molecules are separated by ~3.65 Å, which is in agreement with the literature report, where it was observed that aromatic molecules involves in stacking interaction with approximately parallel molecular plane separated by 3.3-3.6 Å.[28] Moreover, it is also seen from Figure 3 that 2:2 inclusion complex provides better rigidity to the excimer than that of 1:2 complex as free rotation of DCVJ molecules is completely restricted in 2:2 inclusion complex, and therefore, the quantum yield of excimer formation in 2:2 inclusion complex is higher than that of lower stoichiometric complexes as observed in experiment.

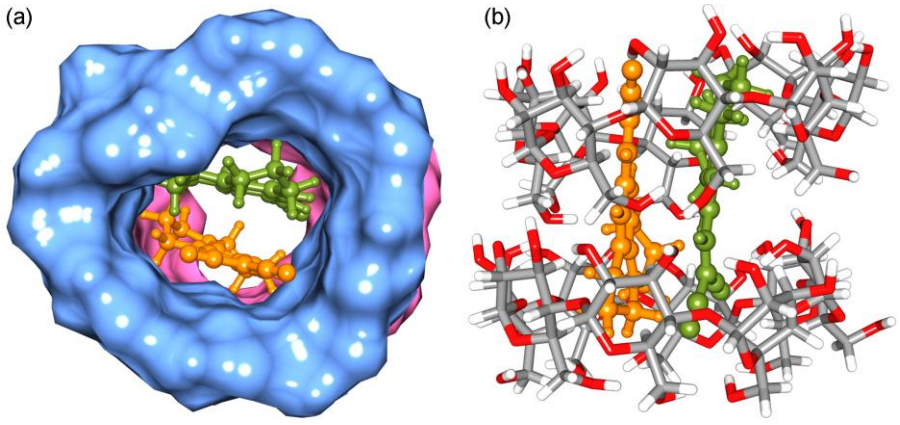


Figure 3: *Optimized structures of γ-CD:DCVJ (2:2) complex. Two cyclodextrins are shown in both surface (blue and pink colors) and stick representations. Rest is same as that of Figure 1.*

4. Conclusions

In this work, we report about the modulation over excimer formation drift of a molecular motor, called DCVJ, by various cavity sizes of cyclodextrin macrocyclic host. The experimental and theoretical studies confirm that DCVJ does not form inclusion complex with α-CD due to the smaller cavity size of the host. However, β-CD, a relatively bigger size of CD, accommodates one DCVJ molecule into its cavity, and therefore excimer population of DCVJ in solution is significantly reduced. Most striking observation is noticed in case of γ -CD having bigger cavity size. Our experimental results indicate that at lower γ-CD concentration, only one DCVJ molecule is being encapsulated by the host cavity. However, at relatively higher γ-CD concentration, two DCVJ molecules are included host, and they form excimer inside the cavity. At a very high concentration of the host, we also got substantial evidence for the formation of 2:2 inclusion complexes between γ-CD and DCVJ. The molecular picture of the interaction between DCVJ and CDs are supported by docking and quantum chemical calculations.

5. References

1. J. Szejtli, *Chem. Rev.*, 1998, **98**, 1743-1754.
2. A. Douhal, *Chem. Rev.*, 2004, **104**, 1955-1976.
3. R. N. Dsouza, U. Pischel and W. M. Nau, *Chem. Rev.*, 2011, **111**, 7941-7980.
4. V. Ramamurthy and D. F. Eaton, *Acc. Chem. Res.*, 1988, **21**, 300-306.
5. V. K. Indirapriyadharshini, P. Karunanithi and P. Ramamurthy, *Langmuir*, 2001, **17**, 4056-4060.
6. R. Kaliappan, Y. Ling, A. E. Kaifer and V. Ramamurthy, *Langmuir*, 2009, **25**, 8982-8992.
7. A. Baldridge, S. R. Samanta, N. Jayaraj, V. Ramamurthy and L. M. Tolbert, *J. Am. Chem. Soc.*, 2010, **132**, 1498-1499.
8. T. Yorozu, M. Hoshino and M. Imamura, *J. Phys. Chem.*, 1982, **86**, 4426-4429.
9. C. Retna Raj and R. Ramaraj, *Chem. Phys. Lett.*, 1997, **273**, 285-290.
10. T. C. Barros, K. Stefaniak, J. F. Holzwarth and C. Bohne, *J. Phys. Chem. A*, 1998, **102**, 5639-5651.

11. C. Lee, Y. W. Sung and J. W. Park, *J. Phys. Chem. B*, 1999, **103**, 893-898.
12. A. S. M. Dyck, U. Kisiel and C. Bohne, *J. Phys. Chem. B*, 2003, **107**, 11652-11659.
13. M. Haidekker and E. Theodorakis, *J. Biol. Eng.*, 2010, **4**, 11.
14. J. A. Levitt, P.-H. Chung, M. K. Kuimova, G. Yahioglu, Y. Wang, J. Qu and K. Suhling, *ChemPhysChem*, 2011, **12**, 662-672.
15. C. Swalina and M. Maroncelli, *J. Phys. Chem. C*, 2009, **114**, 5602-5610.
16. H. Jin, M. Liang, S. Arzhantsev, X. Li and M. Maroncelli, *J. Phys. Chem. B*, 2010, **114**, 7565-7578.
17. C. E. Kung and J. K. Reed, *Biochemistry*, 1989, **28**, 6678-6686.
18. M. Lindgren, K. Sörgjerd and P. Hammarström, *Biophys. J.*, 2005, **88**, 4200-4212.
19. S. Luka, *J. Am. Chem. Soc.*, 1984, **106**, 4386-4392.
20. C. E. Kung and J. K. Reed, *Biochemistry*, 1986, **25**, 6114-6121.
21. A.-Y. Jee, E. Bae and M. Lee, *J. Phys. Chem. B*, 2009, **113**, 16508-16512.
22. H. Jin, X. Li and M. Maroncelli, *J. Phys. Chem. B*, 2007, **111**, 13473-13478.
23. A. Paul and A. Samanta, *J. Phys. Chem. B*, 2008, **112**, 16626-16632.
24. A. Bhattacharya, K. Bhattacharya, B. Bhattacharya and S. Roy, *Indian J. Biochem. Biophys.*, 1995, **32**, 442-446.
25. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **30**, 2785-2791.
26. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. J. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson,, N. G. A., H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M.,, T. Nakajima, Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian,, C. H. P., J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E.,, O. Yazyev, Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J.W., Ayala, P. Y.,Morokuma,, V. K., G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels,, S. A. D., M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman,, O. J. B., J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu,, L.

G., A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham,, P. M. A., C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, and W. W., M. W., Gonzalez, C., Pople, J. A., *Gaussian Inc.*, 2004, **Revision C.02.**

27. J. J. P. Stewart, *MOPAC2009; Stewart Computational Chemistry*: Colorado Springs, CO, 2008.

28. M. O. Sinnokrot, E. F. Valeev and C. D. Sherrill, *J. Am. Chem. Soc.*, 2002, **124**, 10887-10893.

****