

NATURE OF THE MOLECULAR INTERACTION IN DNA-PROTEIN COMPLEXES



**A THESIS SUBMITTED TO THE
CENTRAL DEPARTMENT OF PHYSICS
INSTITUTE OF SCIENCE AND TECHNOLOGY
TRIBHUVAN UNIVERSITY
NEPAL**

**FOR THE AWARD OF
DOCTOR OF PHILOSOPHY
IN PHYSICS**

**BY
RAJENDRA PRASAD KOIRALA**

APRIL 2021

NATURE OF THE MOLECULAR INTERACTION IN DNA-PROTEIN COMPLEXES



**A THESIS SUBMITTED TO THE
CENTRAL DEPARTMENT OF PHYSICS
INSTITUTE OF SCIENCE AND TECHNOLOGY
TRIBHUVAN UNIVERSITY
NEPAL**

**FOR THE AWARD OF
DOCTOR OF PHILOSOPHY
IN PHYSICS**

**BY
RAJENDRA PRASAD KOIRALA**

APRIL 2021

RECOMMENDATION

This is to recommend that **Mr. Rajendra Prasad Koirala** has carried out research entitled “**Nature of the Molecular Interaction in DNA-Protein Complexes**” for the award of Doctor of Philosophy (Ph.D.) in **Physics** under our supervision. To our knowledge, this work has not been submitted for any other degree.

He has fulfilled all the requirements laid down by the Institute of Science and Technology (IoST), Tribhuvan University, Kirtipur for the submission of the thesis for the award of Ph.D. degree.

.....
Dr. Narayan Prasad Adhikari

Supervisor

(Professor)

Central Department of Physics,
Tribhuvan University, Kirtipur,
Kathmandu, Nepal


.....

Dr. Prem Prasad Chapagain

Co-supervisor

(Associate Professor)

Department of Physics,
Florida International University
Miami, United States

April 2021



Ref. No.: (F.No) CDP

Date:

LETTER OF APPROVAL

[Date: 13/04/2021]

On the recommendation of **Prof. Dr. Narayan Prasad Adhikari** and **Dr. Prem Prasad Chapagain**, this Ph.D. thesis submitted by **Mr. Rajendra Prasad Koirala**, entitled “**Nature of the Molecular Interaction in DNA-Protein Complexes**” is forwarded by Central Department of Research Committee (CDRC) to the Dean, IOST, T. U..

.....

Dr. Om Prakash Niraula

Professor

Head

Central Department of Physics,

Tribhuvan University

Kirtipur, Kathmandu

Nepal

DECLARATION

Thesis entitled “**Nature of the Molecular Interaction in DNA-Protein Complexes**” which is being submitted to the Central Department of Physics, Institute of Science and Technology (IOST), Tribhuvan University, Nepal for the award of the degree of Doctor of Philosophy (Ph.D.) is a research work carried out by me under the supervision of Prof. Dr. Narayan Prasad Adhikari of Central Department of Physics, Tribhuvan University and co-supervised by Dr. Prem Prasad Chapagain of Department of Physics, Florida International University, Miami, United States of America.

This research is original and has not been submitted earlier in part or full in this or any other form to any university or institute, here or elsewhere, for the award of any degree.

Rajendra Prasad Koirala

ACKNOWLEDGMENTS

On the behalf of pathways of research works, I could be alone without proper guidance of my supervisors, support of friends and caring of my family members. This work is the outcome of a collaborative efforts and contributions from many helping hands. My deepest gratitude goes to my supervisor Prof. Dr. Narayan Prasad Adhikari, Central Department of Physics, Tribhuvan University (TU), Nepal and co-supervisor Dr. Prem Prasad Chapagain, Department of Physics, Florida International University (FIU), USA, who expertly guided me in every moment of this research work.

My sincere thanks to Prof. Dr. O. P. Niraula, Head, Central Department of Physics, TU and former Head Prof. Dr. B. Aryal for their support. I would like to thank Dr. P. B. Tiwari, Gergetown University, USA for suggestions and supports during my Ph.D. work.

I would offer immense acknowledge to my cousin Dr. G. R. Koirala for his cooperation, fruitful suggestions and caring in every moment of my work. I sincerely thank Prof. A. D. MacKerell Jr., University of Maryland, USA, for his valuable cooperation on input force-field parameters. I would offer my sincere thank to Prof. Dr. S. K. Aryal, Dr. H. P. Lamichhane, Dr. G. C. Kaphle, Dr. N. Pantha, Dr. T. R. Lamichhane and all the faculty members of Central Department of Physics for the suggestions and cooperations. I would be thankful to my lab mates Mr. S. P. Khanal, Mr. J. Powrel, Mr. H. K. Neupane and Mrs B. Thapa for their productive discussion and assistance. The contributions of Dr. R. Pokhrel and Mr. P. Baral, FIU, Dr. S. Lamichhane, Dr. R. P. Adhikari (KU) and Mr. H. P. Bhusal are highly appreciable during my research work. I would be grateful to Mr. R. K. Bachchan and Mr. M. K. Chaudhary for the great help during thesis writing.

I acknowledge the Nepal Academy of Science and Technology (NAST), Khumaltar, Lalitpur, Nepal for three years Ph.D. fellowship. Likewise, I acknowledge the UGC Award No. CRG-73/74-S&T-01 and TWAS research grants RG 20-316.

I would be here from the guidance and support of my family, friends and relatives. My heartfelt gratitude goes to my father Reshmi Raj Koirala and late Fanisara Koirala for loving and caring to make me in this stage. My special thanks goes to my wife Mrs. Laxmi Paudel and my son Mr. Abhinav Koirala for the continuous support, encouragement and patience.

.....
Rajendra Prasad Koirala
April 2021

ABSTRACT

DNA base methylation at the O6 point of guanine is a major cause of cancer. This methylation transits from G:C to A:T mutation pairs during the DNA replication and translation process. The O6-alkylguanine-DNA alkylguanine (AGT) serves as a non-methylating agent, which repairs methylation damage at the O6 point of guanine and O4 point of cytosine by direct damage reversal mechanism. Despite the identification of its role in the methylation damage repair process in the human body, a detailed study is necessary to unlock the underlying mechanism during this methyl transfer process. The present work is focused on the microscopic investigation of DNA-AGT interaction to explore the more insight on DNA damage repair mechanism.

Molecular dynamics (MD) simulation has been carried out to investigate structural basis of the DNA methylation damage repair mechanism, modeling three basic structures: pre-methyl transfer condition (complex-I), transient intermediate state (complex-II), and post-methyl transfer condition (complex-III). Complex-I represents the formation of the DNA-AGT complex and complex-III represents the deformation of the complex AGT from DNA. Complex-II is a close representation of the transient intermediate state of complex-I and complex-II.

The structural and thermodynamic stability of each complex was examined with several physical aspects. Formation of hydrogen bonds as well as energy contributions due to electrostatic as well as van der Waals interactions were taken into consideration to investigate favorable binding of the molecules in all three complexes, which agree with the findings of Daniels et al., 2004. Steered Molecular Dynamics (SMD) results showed that the force in pre-methyl transfer process was greater than that of the post-methyl transfer condition referring to a more favorable binding between DNA and AGT in complex-I. Our findings on force of binding between DNA and AGT agree with AFM experiment of Tessmer and Fried, 2014; and DFT results of Jena et al., 2009. The study was further extended to investigate the changes in the free energy during the interaction of pre- and post-methyl transfer: the former with methylated GUA7 in DNA and the later with methylated CYS145 in AGT. The umbrella sampling method was utilized to calculate the free energy and the results suggested that the change in free energy during the pre-methyl transfer process is greater than that of the post-methyl transfer by 1.3 kcal/mol thereby demonstrating the stronger binding affinity of methylated GUA with AGT than that of the complex in which methylation lies at AGT. Our free energy result agrees with the outcomes of Hu et al., 2007.

TABLE OF CONTENTS

	Page No.
Declaration	i
Recommendation	ii
Letter of Approval	iii
Acknowledgements	iv
Abstract	v
List of Abbreviations	vi
List of Symbols	viii
List of Tables	ix
List of Figures	x
CHAPTER 1	1
1. INTRODUCTION	1
1.1 General Consideration	1
1.2 Proteins and Nucleic Acids	2
1.3 DNA-Protein Interaction	5
1.4 Methylation Damage in DNA	8
1.5 DNA Repair Protein and Repair Mechanisms	12
1.6 Rationale of the Study	15
1.7 Objectives of the Study	16
1.8 Organization of the Thesis	17
CHAPTER 2	19
2. LITERATURE REVIEW	19
2.1 Historical Development of DNA Discovery	19
2.2 DNA Damage Detection	20
2.3 DNA Damage and Repair Mechanism	22
2.3.1 Direct Damage Reversal (DDR)	23
2.3.2 Base Excision Repair (BER)	24

2.3.3	Mismatch Repair (MMR)	24
2.3.4	Nucleotide Excision Repair (NER)	25
2.3.5	Double-strand Break (DSB) and Interstrand Crosslink Repair	25
2.4	DNA Methylation at Guanine	26
CHAPTER 3		30
3.	MATERIALS AND METHODS	30
3.1	Molecular Dynamics	30
3.1.1	Force Calculation	32
3.1.2	Periodic Boundary Condition (PBC)	32
3.1.3	Particle Mesh Ewald (PME)	33
3.1.4	Temperature and Pressure Control	34
3.1.5	Force Fields	34
3.1.6	Modeling of the System	36
3.2	Simulation Details	40
3.2.1	Force Field Parameters for Long-range Interactions	42
3.3	Main Steps of Molecular Dynamics Simulations	43
3.3.1	System Preparation	44
3.3.2	Energy Minimization Run	45
3.3.3	Equilibration Run	45
3.3.4	Production Run	46
3.4	System Setup and MD Simulations in DNA-AGT Systems	46
3.4.1	Steered Molecular Dynamics (SMD)	48
3.4.2	Umbrella Sampling	49
3.4.3	Data Analysis	49
3.4.4	Newly Designed Force Fields	50
3.5	Formulation on MD Analysis	61
3.5.1	Root Mean Square Deviation (RMSD)	61
3.5.2	Root Mean Square Fluctuation (RMSF)	62
3.5.3	Hydrogen Bonding	62
3.5.4	Solvent Accessible Surface Area (SASA)	62
3.5.5	SMD and Pulling Force	63
3.5.6	Umbrella Sampling and Free Energy	64
3.5.7	Weighted Histogram Analysis Method (WHAM)	66
CHAPTER 4		68
4.	RESULTS AND DISCUSSION	68
4.1	General Consideration	68

4.2	DNA-AGT Complexes and Reaction Mechanisms	69
4.3	Stability of Three Complexes	74
4.3.1	Root Mean Square Deviation (RMSD)	74
4.3.2	Root Mean Square Fluctuation (RMSF)	76
4.3.3	Interaction Energy	78
4.3.4	Hydrogen Bonding and Structural Stability	79
4.3.5	Estimation of Distance of Strong Binding Pairs	80
4.4	Electrostatics and van der Waals Interactions	81
4.5	Hydrogen Bonding Analysis	84
4.6	Contact Surface Area	89
4.7	Possible Guanine Base Flipping Mechanism	91
4.7.1	Bond Breaking due to Methylation	92
4.7.2	Stability of CYT20	95
4.8	Steered Molecular Dynamics (SMD) on AGT	97
4.8.1	Pulling of AGT from DNA	98
4.8.2	Solvent Accessible Surface Area (SASA) and Variation of Contact Surface Area during SMD Pulling	102
4.8.3	Maximum Decoupling Force	103
4.9	Estimation of Free Energy	105
4.9.1	Electrostatics and van der Waals Interactions during the Translation of DNA	113
4.10	Intrahelical Flipping of GUA7 after Methyl Transfer	115
CHAPTER 5		118
5. CONCLUSIONS AND RECOMMENDATIONS		118
5.1	Conclusions	118
5.2	Recommendations	121
CHAPTER 6		123
6. SUMMARY		123
6.1	Summary	123
REFERENCES		125
APPENDIX		144
A.	Articles published in International Journals	144
B.	Articles published in National Journals	144
C.	Participation in Conferences	145

LIST OF TABLES

	Page No.
Table 1: Root mean square deviation (RMSD) of three complexes are presented. First column shows the type of complex, other columns show the values of RMSD at the indicated conditions.	76
Table 2: Minimum, maximum and average values of number of hydrogen bonds formed during the 200 ns simulations. First column shows the type of complex, other columns show the values of H-bonds at the indicated conditions.	80
Table 3: Hydrogen bond occupancy of all three complexes during 200 ns NPT simulation. First column shows the type of complex, second and third column shows the residue-nucleotide pairs with specific atoms that were involved in hydrogen bonding, and the fourth column shows the occupancy percentage of residue-nucleotide pairs from corresponding column in second and third.	86
Table 4: Average contact surface area between DNA and AGT that are obtained from 200 ns simulations for all three systems. The net surface area was found maximum in complex-II and minimum in complex-III.	90
Table 5: Variation of electrostatics and van der Waals energy in complex-I and complex-III during the translation of DNA from AGT. Center of mass (COM) distance is the reaction coordinate. The first column contains the COM distance for every successive windows, second and third column are the electrostatics interaction energy in complex-I and complex-III. Fourth and fifth column are the van der waals interaction energy for corresponding complexes.	115

LIST OF FIGURES

	Page No.
Figure 1: Four different structures of Protein.	3
Figure 2: Structure of DNA and RNA: a DNA nucleotides contain G, C, A, T. Purine bases possess the double ring (G and A) and Pyrimidines bases contain single ring (C, T).	5
Figure 3: Block diagram for specific sites of DNA and protein interactions. Protein mostly binds to DNA in the specific sites.	7
Figure 4: Atomic configuration of methylated and non-methylated guanine and cysteine. Methylation at O6 point of guanine is a alkylation damage. After the melthyl transfer to active site cysteine of AGT, DNA is restored into the normal form.	11
Figure 5: Schematic diagram showing DNA damage, possible major effects of corresponding damage and repair mechanisms. First row shows the damaging agents of DNA, second row shows the possible effects of such damaging agents and the third row indicates the corresponding repairs pathways.	23
Figure 6: Periodic Boundary Condition (PBC) showing the original simulation box and some replicated images. The replicated images are assumed as infinitely large numbers so that no boundary effect can be observed.	33
Figure 7: Schematic diagram of bonded stretching and corresponding nature of graph.	37
Figure 8: Schematic diagram of harmonic angle variation and corresponding nature of graph.	38
Figure 9: Schematic diagram of dihedral angle and corresponding nature of graph.	38
Figure 10: Schematic diagram of bonded stretching and corresponding nature of graph.	39
Figure 11: Graphical interpretation of non-bonded interaction. (a) represents the nature of electrostatic potential and (b) represents the nature of lennard Jones potential.	40

Figure 12: Schematic diagram of (a) cutoff and switching points (b) pair-list distance and cutoff distance.	43
Figure 13: The modified structures of GUA7-CYS145 pairs (a) methylated GUA7 with its interacting partner CYS145, (b) covalent linkage of methylated GUA7 and CYS145 and (c) methylated CYS145 with its interacting partner GUA7.	47
Figure 14: Illustration for distributions of overlapping windows during the umbrella sampling simulations. Overlapping windows ensures the representation of entire phase space.	66
Figure 15: Three molecular complexes incorporating targeted parts for methyl transfer mechanism.	73
Figure 16: RMSD of all three complexes. RMSD of complex-II and complex-III were relatively greater fluctuation upto the 100 ns simulation than complex-I, after this time scale all three complexes have the similar fluctuations.	75
Figure 17: Root mean square fluctuation (RMSF) of AGT residues for complex-I complex-II and complex-III. Remote side residues were found more flexible than the interfacial residues indicating the good binding between DNA-AGT molecules in their corresponding complexes.	76
Figure 18: The regions showing the greater fluctuating surfaces. Targeted regions are encircled.	77
Figure 19: The plot of the interaction energy of three complexes as a function of time.	78
Figure 20: The overall hydrogen bonding patterns in complex-I, complex-II and complex-III.	79
Figure 21: Distance between the strongly bounded residue to the nucleotide pairs for (a) complex-I (b) complex-II and (c) complex-III.	80
Figure 22: Distance of SER151(OG) from THY9(O1P) and ADE10(O2P). This shows the shifting of hydrogen bonds of SER151(OG) from THY9(O1P) to ADE10(O2P).	81
Figure 23: Bar diagram representation of the electrostatic, van der Waals, and the sum of these energies for complex- I, complex-II and complex-III indicating the dominant effect of electrostatic energy in overall non-bonded interactions. Electrostatics potential energy is indicated by yellow, vdW potential energy is indicated by red and sum of these potential energies are indicated by blue colour.	83
Figure 24: Comparison of non-bonded (a. electrostatics and b. vdW) energies variations in all complexes.	84

Figure 25: Major interacting sites of DNA and AGT via hydrogen bonding. Site 2 contains many nucleotides that contribute in binding the AGT at the minor groove of DNA. Site 1 also contributes a noticeable hydrogen bonding.	87
Figure 26: Major interacting sites of DNA and AGT showing atomic level hydrogen bonding in major binding sites. Two major regions of binding are shown in the rectangular box in magnified form. . . .	88
Figure 27: Hydrogen bond contributions in nucleotides 6OG7(GUA7), THY9 and THY23 with the protein residues for (a) complex-I (b) complex-II and (c) complex-III. 6OG7(GUA7) has the stable hydrogen bonding in all three complexes, however THY9 and THY23 has the relatively fluctuating in complex-II and complex-III.	89
Figure 28: Contact surface area of DNA and AGT in all three complexes. The area is highly consistent in complex-I but fluctuating in other two complexes.	91
Figure 29: Only DNA system (a) structure of non-methyl GUA7 in dsDNA (b) structure of methyl GUA7 in dsDNA.	93
Figure 30: The RMSD of the DNA for methyl guanine and non-methyl guanine system. More fluctuations is found in methyl condition than that of non-methyl condition.	94
Figure 31: The scheme of formation of hydrogen bonds in non-methylated guanine (GUA7) and CYT20, and methyl guanine (6OG7) and CYT20. All three hydrogen bonds exist in non-methylated guanine condition as the ordinary pairs, however only two hydrogen bonds sustain in methylated guanine.	94
Figure 32: Comparison of the number of hydrogen bonds formed by non-methylated GUA7 and CYT20, and methylated GUA7 and CYT20 pairs. Number of hydrogen bonds are clearly distinguishable in two cases.	95
Figure 33: (a) Hydrogen bonding pattern in ARG128-CYT20 pair showing the strongest hydrogen bond between NH2-N3, (b) Radial distribution for NH2-O2, NH2-N3, and NH2-N4 pairs. NH2-N3 pair has the highest distribution and lie within 3.5 Å, but other two pairs occasionally comes within this range. (c) Distance of N3, N4, and O2 atoms of CYT20 from NH2 atom of ARG128.	97
Figure 34: Pulling of AGT along positive x-direction taking DNA as the reference molecule. In this condition, AGT is pulled in both systems, system-I and system-II, along the same direction (positive x direction) in identical condition.	99

Figure 35: Binding sites of AGT and DNA. Two binding regions (site 1 and site 2) retains upto entire 30 ns simulation in system-I and only one binding region (site 2) exist till the end of the simulation, however, the site 1 was decoupled during the simulation in system-II. . . .	100
Figure 36: The variation of hydrogen bonds in system-I and system-II. The sharp decrease in the number of hydrogen bonds in system-II occurred due to the decoupling of hydrogen bonds from site 1. . .	101
Figure 37: The intramolecular decoupling region of AGT. Breaking of hydrogen bonds in region 3 facilitates the AGT to stretch with the certain part.	102
Figure 38: Contact surface area between DNA and AGT for system-I and system-II. pq and p'q' regions of the graph represent the stable binding, qr and q'r' show sharp down jump due to the intramolecular breaking region in AGT and the fast descending of the contact surface area. rs and r's' regions are constantly running parts that refer to the constant contact surface area between the molecules due to the strong binding in site 2 as shown in molecular structure accompanied with the plot.	103
Figure 39: Decoupling force versus COM displacement of AGT by taking DNA as the reference molecule. Figures (a) and (b) are the force versus displacement curves for five identical simulations of system-I and system-II, respectively. (c) the average values of force versus displacement curve obtained from averaging (a) and (b) separately. The greater value of force in system-I than that of system-II shows more favorable binding of methylation damage DNA with AGT than the binding after methyl transfer.	104
Figure 40: (a) Direction of DNA translation and representative snapshots of displacement of methylated DNA from AGT in the complex-I at COM distances of (b) 24.8 Å (c) 30.8 Å (d) 36.8 Å and (e) 42.8 Å.	108
Figure 41: Plots for distribution of DNA from AGT during the free energy estimation for all 19 windows of each of complex-I and complex-III, respectively. The distribution ensures the sufficient overlapping of windows and are suitable for the free energy calculation. (a) distributions in complex-I and (b) distributions in complex-III . .	109
Figure 42: RMSD measurements for 19 windows obtained from the umbrella sampling for complex-I and complex-III. The RMSD measurements are fairly stable, especially for an essential interaction such as when AGT and DNA were in close contact.	110

Figure 43: Plots for free energy estimation: Change in free energy for complex-I and complex-III taking COM distance as the reaction coordinates.	112
Figure 44: Free energy comparison in 10 ns and 20 ns for complex-I. The free energy change is almost the same showing convergence of 10 ns simulation run.	113
Figure 45: Variation of non-bonded interaction energies with the displacement of DNA from AGT; electrostatic interaction energy (E, red color), van der Waal interaction energy (VDW, orange color), and total non-bonded energy (E+VDW, green color) calculated by using NAMD energy plugin within VMD: (a) for complex-I and (b) for complex-III.	114
Figure 46: (a) Intrahelical returning of flipped guanine after the methyl adduct removal condition. Five snapshots are taken from the trajectories at 0.0 ns, 2.5 ns, 5.0 ns, 7.5 ns and 10.0 ns. (b) the hydrogen bond pattern between GUA7 and CYT20, hydrogen bond formation occurred only after the flipping back of GUA7 into intrahelical region.	116
Figure 47: Intrahelical flipping analysis (a) distance between GUA7 and CYT20 (b) the distribution of GUA7 from CYT20, during the returning back of the distribution is significantly varying.	117

LIST OF SYMBOLS

k_B	Boltzmann constant
k_{ij}^b	Force constant for bond stretching potential
b_{ij}^0	Equilibrium bond length
k_{ijk}^θ	Force constant for bond angle potential
θ_{ijk}^0	Equilibrium bond angle
k_ϕ	Force constant for dihedral angle potential
δ	Equilibrium dihedral angle in degree
k_ξ	Force constant for improper dihedral angle potential
ξ_0	Equilibrium improper dihedral angle in degree
ϵ_0	Permittivity of free space
ϵ_r	Dielectric constant
ϵ	Strength of potential in L-J potential
σ	Length scale in L-J potential
k_i	Force constant for i^{th} window in umbrella sampling

LIST OF ACRONYMS AND ABBREVIATIONS

ABH2	Human AlkB Homologous 2
ABH3	Human AlkB Homologous 3
AFM	Atomic Force Microscopy
AGT	O6-alkylguanine-DNA alkyltransferase
ARG	Arginine
ASN	Asparagine
BER	Base Excision Repair
CHARMM	Chemistry at Harvard Macromolecular Mechanics
COM	Center of Mass
CPD	Cyclobutane Pyrimidine Dimers
CYS	Cysteine
CYT	Cytosine
DDR	Direct Damage Repair
DFT	Density Functional Theory
DNA	Deoxyribo-Nucleic Acid
DSB	Double-Strand Break
dsDNA	double stranded Deoxyribo-Nucleic Acid
EM	Energy Minimization
FFT	Fast Fourier Transform
GROMACS	Groningen Machine for Chemical Simulation
GUA	Guanine
hAGT	human O6-alkylguanine-DNA alkyltransferase
HhH	Helix-hairpin-Helix
HLH	Helix-Loop-Helix
HPLC-MS	High performance Chromatography-Mass Spectroscopy
HR	Homologous Recombination
HTH	Helix-Turn-Helix
IUB	International Union of Biochemistry
IUPAC	International Union of Pure and Applied Chemistry
LEU	Leucine
LMPCR	Ligand-mediated Polymerase Chain Reaction
LYS	Lysine
MD	Molecular Dynamics
MeA	Methylated Adenine
MeC	Methylated Cytosine

MeG	Methylated Guanine
MeT	Methylated Thymine
MGMT	O6-methylguanine methyltransferase
MMR	Mismatch Repair
mRNA	messenger Ribonucleic acid
NAMD	Nanoscale Molecular Dynamics
NER	Nucleotide Excision Repair
NGS	Next-Generation Sequencing
NHEJ	Non-Homologous End-Joining
NMR	Nuclear Magnetic Resonance
NPT	Constant Number, Pressure and Temperature
NTP	Normal Temperature Pressure
NVE	Constant Number, Volume and Energy
NVT	Constant Number, Volume and Temperature
PBC	Periodic Boundary Condition
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PME	Particle Mesh Elward
PSF	Protein Structure File
qPCR	quantitative Polymerase Chain Reaction
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
SAM	S-adenosyl-L-methionine
SER	Serine
SMD	Steered Molecular Dynamics
TIP3P	Transferable Intermolecular Potential with 3 Points
TUNEL	TdT-mediated dUTP-biotin nick end labeling
UV	Ultra-violet
VMD	Visual Molecular Dynamics
WHAM	Weighted Histogram Analysis Method
XRD	X-ray Diffraction
ZNF	Zinc Finger

CHAPTER 1

1. INTRODUCTION

1.1 General Consideration

Biomolecules perform several works to keep the organisms in survival state (Banani et al., 2017; Karplus & McCammon, 2002). They are essential for many biological processes such as physical and mental growth, cell division, self-healing, signal transmission, defensive mechanisms against external pathogens (Cotterill, 2003). Also, many non-biological molecules including water, mineral, and air are integral entities of the overall life processes. Despite the presence of all the biomolecules in the body, the life process is impossible in the absence of these external non-biological molecules. Indeed, non-biological molecules create a suitable environment and involve themselves in life processes like food transportation, absorption, excretion, functionalization of the sensory organs, neurotransmission, etc. (Cleaves II et al., 2012; Lowe et al., 1998). Thus, the living body is the coalescence of such macro-molecules and non-biological molecules under an appropriate cellular environment provided with a certain pressure, temperature, and volume. If there is any complication in a biomolecule, it is propagated to other molecule.

Out of many biological molecules, nucleic acids (Deoxyribose nucleic acid (DNA) and Ribonucleic acid (RNA)) are the genetic information carrying molecules. Basically, DNA expresses the protein and carries the hereditary characters from parents to offspring (Fleck & Munro, 1966). Aberration in DNA may cause the genomics damages, which ultimately produces the genetic effects. Methylation damage is one of the common type of DNA damages that can mutate cytosine-guanine pair into adenine-thymine pair in DNA (Tessmer & Fried, 2014). A protein O6-alkylguanine-DNA alkyltransferase (AGT) has been proposed as methyl damage repair protein. It is still in clinical trial (McKeague et al., 2018). Motivated by DNA repair characteristics of AGT, this work is focused on the investigation of binding mechanism of DNA and AGT, and methyl damage repair process in DNA by utilizing AGT. To build the general background of the problem regarding the DNA-AGT interactions, we begin discuss some general backgrounds of

conformation and important functions of protein and DNA.

1.2 Proteins and Nucleic Acids

Twenty amino acids are the basic building blocks of naturally occurring proteins (Vickery & Schmidt, 1931; Wagner & Musso, 1983). Each amino acid contains an amine-head (NH_2 -) and a carboxyl tail (-COOH) (Hovmöller et al., 2002). When the carboxyl tail of an amino acid covalently links to the amine head of another amino acid by eliminating a water molecule, a dipeptide molecule is formed (Ramachandran & Sasisekharan, 1968; Lumry & Eyring, 1954). The covalent bond between amine and carboxyl groups (CO-NH) of two amino acids is called a peptide bond and it is responsible to unite these amino acids as a single molecule. Likewise, amino acids link together one after another with peptide bonds and thus, forms a long chain like beads in a necklace, called the polypeptide. The biologically functional polypeptide is called a protein (Bulaj, 2005; Dyson & Wright, 1991). Being a polycondensation of the backbone of amino acids (-NH-RCH-CO-), in all proteins are similar, nevertheless, the side chain (-R) is different. The side chain is different in every amino acid, resulting in the structural and functional variation in proteins (Lumry & Eyring, 1954; Krimm & Bandekar, 1986; Khanal, Koirala, et al., 2021). As a long polypeptide cannot form a linear structure, it twists into various shapes thereby seeking stability to minimize the total potential energy of the molecule (Dyson & Wright, 1993; Myers et al., 1997). The structural variation in a protein molecule not only gives the conformation stability but also contributes to the change in the chemical and biological properties of that particular molecule (Hendrix et al., 2005). There are several interactions among the amino acid residues to establish a protein structure. Hydrogen bonds, disulfide bonds, salt bridges, hydrophobic interactions, van der Waals, and electrostatic interactions are the key interaction mechanisms to form a stable and functional protein (Emamjomeh et al., 2019).

Protein structures are classified into four types: primary structure, secondary structure, tertiary structure, and quaternary structure (Pauling et al., 1951; Moutevelis & Woolfson, 2009). The primary structure is a linear chain of amino acids. A very light polypeptide can remain in the primary structure; however, a heavy chain of protein does not stay anymore in this structure. The local organization of the polypeptides via hydrogen bonds generally forms the secondary structures. They are formed in a specific order like alpha-helix, beta sheets, turns, and loops (Harteis & Schneider, 2014; Blake et al., 1978). Three-dimensional folding of the polypeptide comprises the tertiary structure of the protein. Disulfide bonds, hydrogen bonds, salt bridges, hydrophobic interactions, and van der Waals interactions are employed to the tertiary structure of the protein. Multiple

chain involvement to form a single protein molecule establishes the quaternary structure of the protein which remains in a globular form (Serdyuk et al., 2017; Luscombe et al., 2000). Different conformational structures of protein are shown in Figure 1 (Sun et al., 2004).

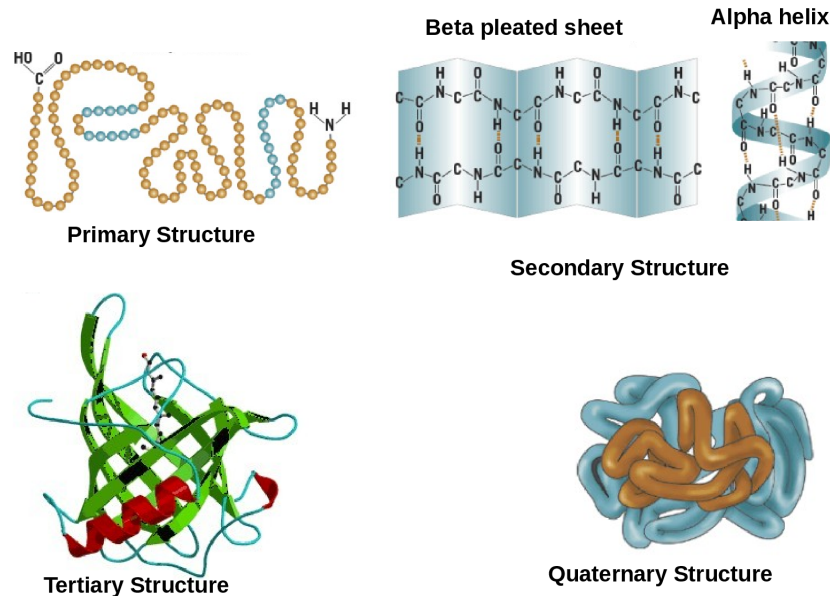


Figure 1: Four different structures of Protein.

There are numerous of proteins in the human body to perform varieties of functions. These functions include serving as carriers of vitamins, oxygen, and carbon dioxide. Similarly, proteins perform a vast array of functions within organisms, including catalyzing metabolic reactions, DNA replications, responding to stimuli, and transporting molecules from one location to another (Chen et al., 2009). Proteins are the invariable participants in all life processes. Protein-enzymes catalyze all chemical, electrochemical, and mechanochemical processes taking place in cells and organisms (Waterlow et al., 1978). The most important function of proteins is that of an enzyme. Specialized enzymes serve as catalysts for all metabolic reactions, DNA replications, the transition of genetic information from DNA to mRNA, and the transfer of the information encoded in the mRNA to the ultimate structure of the protein molecule (Watson, 2004).

The hydrophobic and hydrophilic nature of amino acids plays a crucial role in the interaction of proteins with the surrounding molecules (Bigelow, 1967). Hydrophobicity in protein always tends to minimize the surface area owing to the reduced contact area between the nearby molecules, whereas the hydrophilic amino acids on the surface of protein molecule allow more area to interact with the nearby molecules due to an enhanced surface contact area (Aftabuddin & Kundu, 2007).

Nucleic acids are thread-like macro-molecules that play a central role in hereditary processes: storage of hereditary information, replication, transcription, and translation.

They are categorized into two types: deoxyribose nucleic acid (DNA) and ribose nucleic acid (RNA) (Fleck & Munro, 1966; Watson & Crick, 1953). The building blocks of these nucleic acids are called nucleotides. Each nucleotide contains three subunits: a part of phosphate, a pentose sugar, and a nitrogen base (Watson & Crick, 2010). DNA and RNA differ from each other from the different number of the hydroxyl group in C2' carbon of pentose sugar. While RNA contains two hydroxyl groups bonded on C2', DNA contains a hydroxyl group and a hydrogen atom at the same position as RNA, and hence it is named deoxyribose referring to the deficiency of oxygen (Leontis & Westhof, 2001). DNA contains thymine (T), adenine (A), cytosine (C), and guanine (G). During the synthesis of double strands, T links to A with two hydrogen bonds and C links to G with three hydrogen bonds. RNA also contains four nitrogen bases: adenine (A), cytosine (C), guanine (G), and uracil (U) (Singer, 1968; Watson & Crick, 1953; Avery et al., 1944). Furthermore, the nitrogen bases are categorized into two types depending upon the number of rings. Single ring bases are called pyrimidines and the double ring bases are called purines. Cytosine, thymine, and uracil are pyrimidines; whereas; adenine and guanine are purines. One nucleotide is covalently connected to another nucleotide from C3' to C5' position through phosphodiester bonds. The polycondensation of nucleotides via phosphodiester bonds form a long chain of single-stranded DNA or RNA. When two strands are linked at nitrogen bases through hydrogen bonds, double-stranded DNA (dsDNA) is formed. In such conformation of dsDNA, two strands run oppositely containing 5' to 3' carbon in one strand and 3' to 5' carbon in the complementary strand (Lucas-Lenard & Lipmann, 1971). the structures of DNA, RNA and their nucleotides are shown in Figure 2.

Watson-Crick model, designed in 1953, is the first discovered model of double-stranded DNA structure (Watson & Crick, 1953). They successfully modeled the DNA structure, a right-handed double-helical form, formed by two antiparallel DNA helical strands link via the sequence-specific base pairing of two strands (Wilkins & Randall, 1953). This model has been explained as a "ladder polymer", the rungs being of equal length because of the equality of these AT and GC pairing. The sides of the ladder are composed of regularly alternating sugar and phosphate units. The steric hindrance between sugar and phosphate causes the double helix structure into the twisting form. The twist angle is roughly 30° in two successive base pairs so that one pitch is completed at about 10.6 successive base pairs on average. The distance corresponds to each pitch is about 34 Å (Cotterill, 2003). During the interaction of DNA with other molecules, the diameter and the twisting angle of the double helix are changed. This structural change essentially alters the biochemical function of the DNA molecule (Lucas-Lenard & Lipmann, 1971; Watson & Crick, 1953).

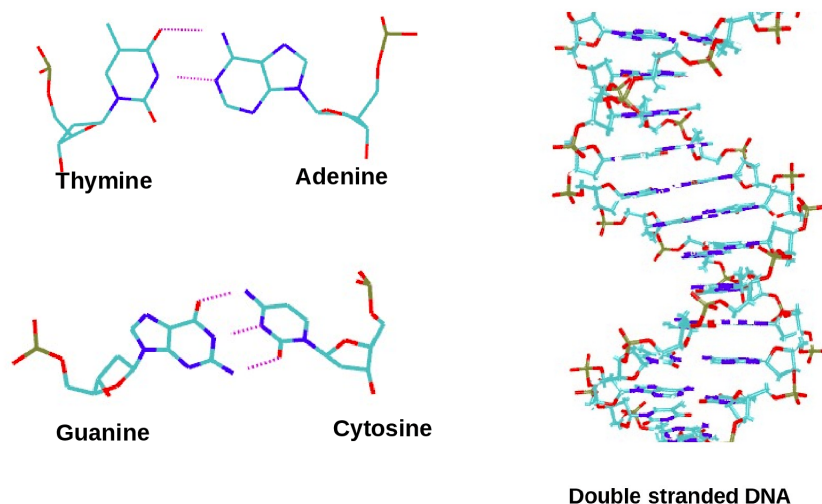


Figure 2: Structure of DNA and RNA: a DNA nucleotides contain G, C, A, T. Purine bases possess the double ring (G and A) and Pyrimidines bases contain single ring (C, T).

Protein interacts with DNA to perform several works. Aside of several functions of protein, DNA damage repair mechanism is one of the important mechanisms in living cells. The entry of protein near the damaged part of DNA is the fundamental requirement to initiate the DNA damage recovery. The protein should approach at a favorable distance for their mutual interaction. In body mechanisms, as soon as the DNA repair protein searches the appropriate position near the damaged site of DNA, they form the biomolecular complex (Ames et al., 1993). To provide the fundamental idea of binding mechanism in DNA and protein, the following section explains the basic concepts of DNA-protein interactions.

1.3 DNA-Protein Interaction

Proteins and nucleic acids are two types of basic alphabets of life. The protein synthesis process begins after the synthesis of messenger RNA (mRNA)(Lengyel & Söll, 1969). In the beginning, mRNA is synthesized into the nucleus as the complementary strands of DNA by the translation process and comes out to the cytoplasm through nuclear pores. The mRNA binds to the ribosome. The ribosome catalyze to amino acids to form the protein molecule. Amino acids stitch over the ribosome and synthesize a long polypeptide. Proteins are synthesized by ribosomes in the living cells in accordance with the genetic code in DNA (Lucas-Lenard & Lipmann, 1971). The synthesized protein can be involved in several mechanisms of a living body. Several proteins are involved in regulating the gene expressions that make code for proteins (Lucas-Lenard & Lipmann, 1971; Lengyel & Söll, 1969). Proteins can interact with DNA and the interaction between them is responsible for the different functions of life activities.

Many crucial biological processes in our body such as the transport and translation of RNA, packaging of DNA, genetic recombination, replication, and DNA repair are controlled by the interaction of proteins and DNA (Dey et al., 2012). The bio-technical manipulation of protein-DNA interaction can regulate the gene expression in some virulence genes, which can be used for the treatment of many diseases (Emamjomeh et al., 2019). Therefore, protein-DNA interaction is very important in the growth, development, differentiation, and evolution of living beings. Proteins bind to DNA with several distinct structural motifs such as the helix-turn-helix (HTH) motif, helix-loop-helix (HLH) motif, helix-hairpin-helix (HhH) motif, zinc finger, and leucine-zipper. HTH motif recognizes the protein entry location in DNA and builds the space to interact with DNA (Rosinski & Atchley, 1999). HLH motif is characterized by two α -helices connected by a loop, generally found in protein dimers, in which one of the helices are active to facilitate DNA binding (Chowdhury & Bagchi, 2015). Zinc finger (ZNF) domains are usually the small protein motifs that contain multiple fingers like protrusions that makes a contact with nucleic acids as well as the other proteins (Ebert & Altman, 2008). It has a key role in the development and differentiation of several tissues. Similarly, the HhH motif plays a very important role in non-sequence-specific DNA bindings, and also mediate in protein-protein interactions (Jones et al., 2003).

DNA-protein binding is mostly sequence-specific in which protein binds to DNA at a site having a specific nucleotide sequence. CpG (cytosine followed by guanine) sequence is usually observed in a sequence-specific manner (Esteller, 2007; Koirala et al., 2020a). Sequence non-specific binding between protein and DNA is also found in many complexes. In sequence non-specific binding, amino acid residues in protein randomly interact with DNA nucleotides. In addition, the interaction between protein and DNA is direct or indirect (Steffen et al., 2002). In the direct interaction, amino acid residues interact directly to nucleotides at the interfacial region without any external mediator, whereas solvent molecules, usually water, mediate the interaction between the molecules in case of the indirect interactions (Emamjomeh et al., 2019; Harris et al., 2014; Coulocheri et al., 2007).

Binding sites of protein with DNA rely greatly on the basic structure of protein like secondary or tertiary. In secondary structure, intra-atomic binding is contributed by the hydrogen bonds, whereas in the tertiary structure, many other components have vital roles like formation of disulfide bond, hydrophobic interactions, salt bridges, non-covalent electron cloud interaction etc (Chou & Fasman, 1978). Similarly, binding sites of DNA are contributed by region specific like major groove or minor groove or base modification. Moreover, the binding is also influenced by the global or local shape of the DNA (Murugan, 2010).

The binding sites of DNA and functional protein are investigated on the basis of energy

level during the formation of complex. The binding is favourable only when the complex structure guarantees the energy level. The binding energy of the complex structure should be sufficiently reduced to its lowest level. If the binding occurs at inappropriate site, the free energy rises up and the complex becomes unstable. The shifting of free energy not only depends on the variation in the active configuration of the local bases of DNA but also in the dimensional change like length and precise angle of local bonds (Emamjomeh et al., 2019). Local environment is also the key factor to create the suitable environment in binding DNA and protein. Water molecules and ions around the binding site also play key role in reducing the free energy of the entire complex. Figure 3, summarizes overall binding patterns of DNA and protein and provides schematic picture of major binding sites of DNA-protein complex and the thermodynamically favourable interactions in site specific binding position (Gapsys et al., 2018).

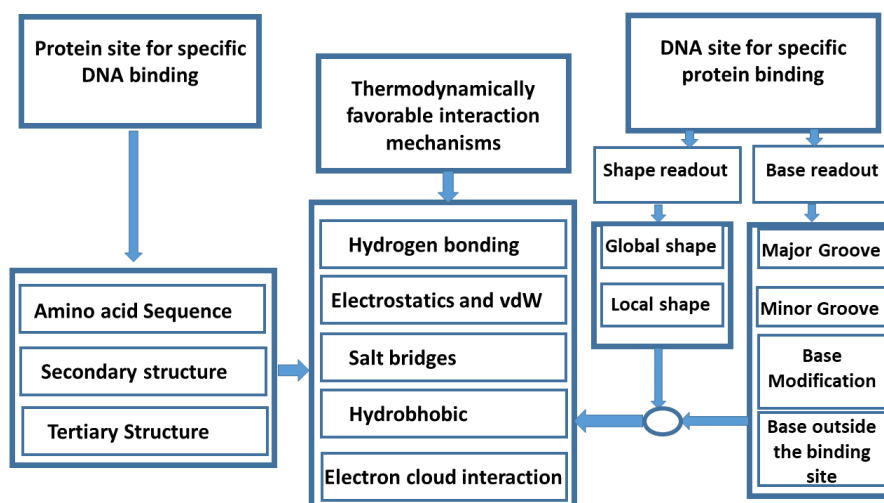


Figure 3: Block diagram for specific sites of DNA and protein interactions. Protein mostly binds to DNA in the specific sites.

In the interfacial region, several types of non-covalent interactions take place like hydrogen bonding, salt bridges, electrostatics, and van der Waals interactions. Since the amino groups act as the hydrogen bond donors in protein main chains, side chains, and DNA bases, they play a decisive role in forming hydrogen bonds in nucleotide-residue pairs. Hydrogen bonds not only play role in intermolecular interactions, but they also contribute to form the three-dimensional structures in the complex (Pace et al., 2014). Similarly, some amino acid residues possess a positive charge at the side chains and the phosphate part of nucleotide has a negative charge to the nucleotide. This condition allows for direct electrostatic interaction. Acidic and neutral residues are poor electrostatic contributors. Van der Waals interaction is pervasive and plays a significant role in binding protein and DNA (Nimrod et al., 2009; Mikles et al., 2013).

Protein usually binds to DNA through the major groove. In the present study, the structure

of the DNA-protein complex is a novel architecture in the sense that the protein binds to DNA through the minor groove. The protein interacts with the minor groove via the helix turn helix motif (Daniels et al., 2004). In the present work, the methyl adducts at the O6-position of guanine, the seventh residue of DNA, GUA7, detaches the hydrogen bonding with its pair partner cytosine and hence facilitates to rotate from the backbone of the corresponding chain. The rotated methylated guanine when approaches to the binding protein, several protein residues come to interact, and the complex becomes more stable (Hu et al., 2008). Many of the other nucleotides and amino acid residues in AGT protein enhances binding DNA and protein molecules, which ultimately provide sufficient time to transfer the methylation damage to the specific site of the protein. Hydrogen bonds and non-bonded electrostatics and van der Waals interactions play pivotal roles in binding these structures. The covalent bonding between DNA and protein is transient (Jones et al., 2003; DiStasio et al., 2014). Now, the major types of DNA damages, basically the cause and effect of methylation damage in DNA will be explained immediately after this section. This work is specifically focus on the methylation damage in guanine base of DNA and its repair mechanisms.

1.4 Methylation Damage in DNA

There are several factors associated with DNA damages. The damage can occur due to internal metabolism or toxic agents from the environment. The internal sources of damage include alkylation, oxidation, hydrolysis, and mismatch of DNA bases (Jackson & Bartek, 2009). Likewise, the ionizing and ultraviolet radiation, drugs for anticancer chemotherapies are external sources of DNA insult (Lindahl & Wood, 1999; Chatterjee & Walker, 2017). Moreover, there are various types of nitrogenous base damages in DNA like chemical modifications of bases, missing bases, crossing the bases, base methylations, etc. Among these, DNA base methylation, i.e., alkylation of DNA base, is one of the major causes for the cell to be carcinogenic (Jackson & Bartek, 2009).

Basically, the damaged portion in DNA blocks the DNA replication and transcription and then, causes the mutations in genomes. Physical structures and chemical properties are altered in such mutated genomes which can also change the biochemical and biophysical characters in the genetic materials (Drabløs et al., 2004). The mutations in the gene, an aberration in DNA, causes a mutation in newly synthesized proteins during the protein biosynthesis. If such types of aberrations in DNA are not repaired or wrongly repaired, the corresponding cell may be mutagenic and can cause tumorigenesis (Tessmer & Fried, 2014). The physiological processes in the body may also cause the DNA damage (Soll et al., 2017). During the DNA replications, the mismatch may take place. Likewise, abortive topoisomerase I and topoisomerase II activity cause the double strands to break

in DNA. Sometimes two partner strands in DNA come at closer proximity and strands may be broken (Tessmer & Fried, 2014).

Chemical attack is also a potent cause of damage in DNA. Oxidation, hydration, and alkylation are the major chemical modifications in DNA (De Bont & Van Larebeke, 2004). The by-products from reactive oxygen compounds may generate the DNA lesions at nitrogen bases. Many other types of lesions are caused by hydrolysis and non-enzyme methylations (Halliwell & Aruoma, 1991). The oxygen and nitrogen compounds produced from macrophages and neutrophils damage the DNA at the site of inflammation and infections. The ultimate effect of chemical attack is also the obstruction in DNA replications and transcriptions. The chemo-toxic materials primarily attack DNA strands and break them (Torgovnick & Schumacher, 2015; De Bont & Van Larebeke, 2004).

The most pervasive cause of exogenous damage in DNA is the impinging of ultraviolet rays on the cell nucleus and hampering on DNA chains (Ames, 1989). Although the ozone layer on the top of the atmosphere absorbs the ultraviolet spectrum, especially the ultraviolet C, the residual ultraviolet A and ultraviolet B can penetrate the atmosphere and reach the earth's surface (Jackson & Bartek, 2009). The exposure of this ultraviolet spectrum violently induces the DNA strands producing about 1,00,000 lesions per second (Ames, 1989; Marnett, 2000). The ionizing radiations produced from artificial and natural sources of radiation greatly infect DNA, both in nitrogen bases and backbones. Uranium, thorium on the earth's crust constantly produces damaging rays like alpha rays, beta rays, and gamma rays which generate damage in DNA and make them carcinogenic (Jackson & Bartek, 2009). The naturally occurring radon gas accumulates at the corners of the room and causes harm to the genetic material of people living at that home (Reuter et al., 2010). Since radon is heavier than that of natural gases in the atmosphere, it can deposit at the corners of home as the dust particles. Besides, many radioactive isotopes, cobalt-60, iodine-131, and technetium-99m are used in cancer therapy. The radiation produced from such artificial sources can have a reversible effect on cancer (Kaina et al., 2007).

DNA base methylation is one of the chemical modifications in DNA in which methyl adduct CH_3 attaches to one of the atom positions of DNA nucleotides. It is the base-pairing damage and is the major cause of cancer (Roos et al., 2004). The methylation at the C5 position of cytosine is ubiquitous. Besides this methylation damage in DNA nucleotide, a similar type of damage can occur in other bases. Methylation occurs not only in the nitrogenous bases but also in the pentose sugar. If the nucleobase is modified due to methylation, it becomes to tumorigenic Wyatt & Pittman (2006). It can block or wrongly encode the newly generated nucleotide that seriously hampers the replication and transcription leading to cellular death (Kaina et al., 2007; Christmann et al., 2011).

During the protein biosynthesis, the selection of amino acids in the mRNA template is not random rather the selection occurs how well it can fit in the three-dimensional structure of a protein and its functional fold (Lucas-Lenard & Lipmann, 1971). Since the mRNA template encoded from the original DNA strands affects the amino acid sequence of newly synthesized proteins, it ultimately results in protein mutation. Though all types of mutations are not harmful, the mutation by methylation is carcinogenic (Tessmer & Fried, 2014).

As soon as the carbon atom of methyl fragment attaches to the O6 position of the guanine base, the double bond between O6 and C6 reduces to a single bond and then N1 and C6 become double-bonded by deprotonation at N1 (Daniels et al., 2004). Thus, the methylated guanine can be structurally conserved as shown in Figure 4. In ordinary base pairing, guanine pairs to cytosine via three hydrogen bonds. However, if the O6 point of guanine is methylated, the guanine impairs to thymine during the DNA replication so that G:C to A:T transition occurs. This transition may cause cell to be carcinogenic and mutagenic. In the next section, will discuss the activity of AGT that is crucial in methylation damage repair process in DNA.

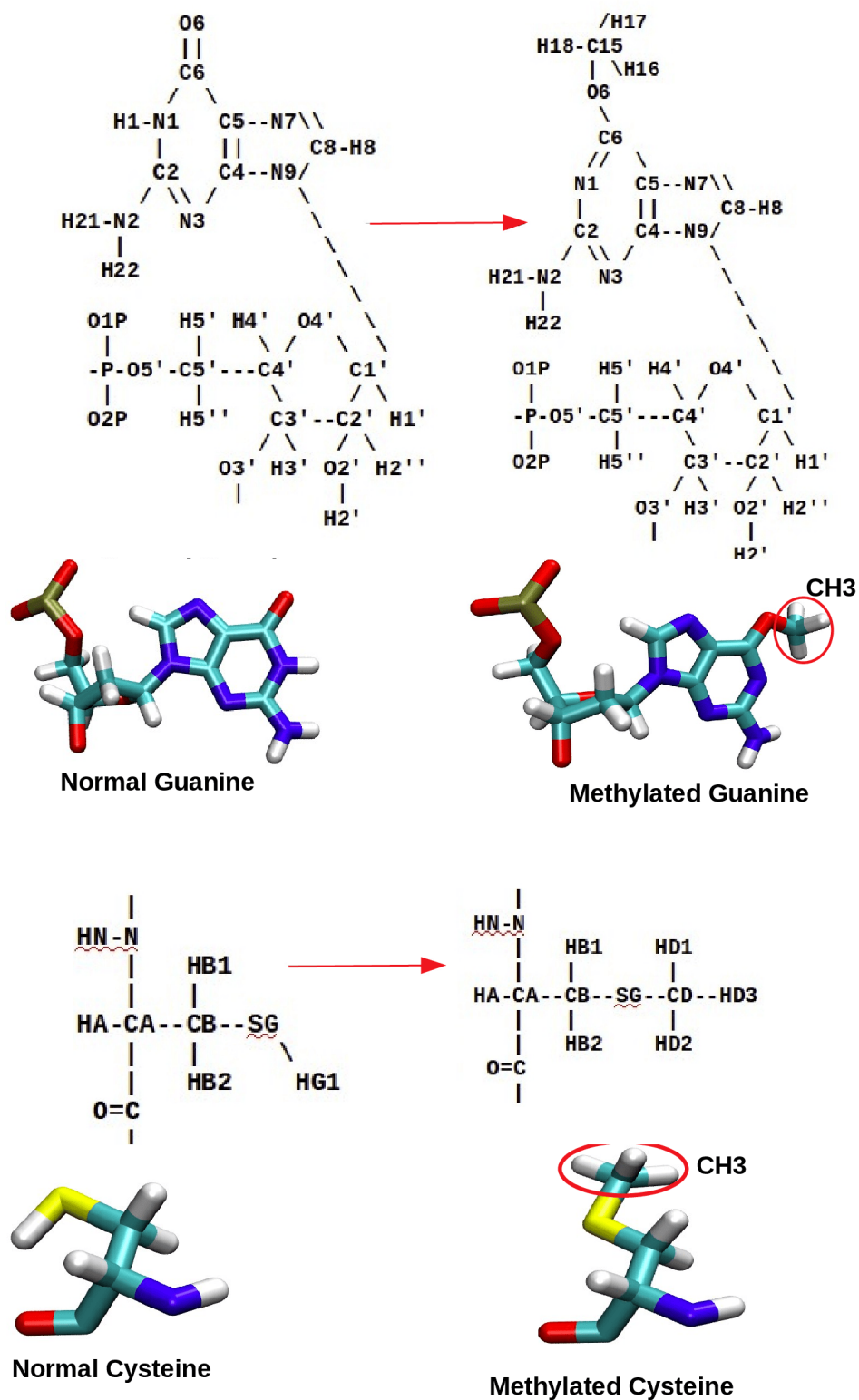


Figure 4: Atomic configuration of methylated and non-methylated guanine and cysteine. Methylation at O6 point of guanine is an alkylation damage. After the methyl transfer to active site cysteine of AGT, DNA is restored into the normal form.

1.5 DNA Repair Protein and Repair Mechanisms

There are approximately 10^{13} cells in the human body, each receives tens of thousands of DNA lesions per day (Jackson & Bartek, 2009). Organisms have several defensive mechanisms against such deleterious effects on DNA (De Bont & Van Larebeke, 2004). DNA glycosylases, endonucleases, DNA polymerases, polynucleotide kinase, and ligase enzymes take part in the direct damage reversal process. If the body enzymes are not enough to repair the damage, this causes cell death by apoptosis or cellular senescence, i.e., permanent cell-cycle withdrawal (Ali et al., 1998). Investigations to date have not been able to discover the specific drug for the removal of methylation damage and the ejection of covalently linked CH_3 from the nitrogen base. A protein, O6-alkylguanine-DNA alkyltransferase (AGT), had been purposed as the demethylation agent for the damaged DNA nucleobase. The investigations targeting development of potential drugs for new cancer types are challenging. The presumption of the appropriate drug from theoretical designing is very helpful in developing new chemotherapeutics to treat such disease (McKeague et al., 2018; Kelley & Fishel, 2008).

DNA regulatory proteins and transcription factors, restriction endonucleases, DNA polymers, and nucleosomes are some important examples of DNA binding proteins that play a fundamental role in the structural and functional features of DNA-protein interaction (Luscombe et al., 2000). Transcription factors control the rate of transcription of genetic information from DNA to mRNA by binding to a specific DNA sequence. DNA and RNA polymerases are used to assemble DNA and RNA molecules, respectively by copying DNA or RNA template strands using base-pairing interactions (Alberts et al., 2002). Nuclease is an enzyme that cleaves the phosphodiester bond between the nucleotide subunits of nucleic acids. Histone proteins are highly alkaline proteins that are found in eukaryotic cell nuclei that package and order the DNA into structural units called nucleosomes (Kauffman & Karypis, 2012; Rohs et al., 2010). Like many DNA binding proteins, AGT has been investigated as a repair agent of the methylation damage in the nitrogen base of DNA.

AGT protein is found in both the cytoplasm and the nucleus of a cell (Arsham et al., 2017). Although phosphorylation can facilitate the transport mechanism, its transport mechanism within the cell is still unclear. It is highly sensitive to the O6-alkylguanine lesion in DNA. When it recognizes the alkyl O6-position of guanine, it shifts towards the lesion (Margison et al., 2003; Acharya et al., 2021). The repair process is quite fast, within a few minutes. It has a remarkably conserved active site in all species from bacteria to human (Arsham et al., 2017). This protein is non-toxic, even though it is highly sensitive to methylating and chloroethylating agents. One interesting mechanism is that it accepts the alkyl group like O6-alkylguanine lesions (Srivenuopal et al., 1996).

AGT is a protein that is encoded by the O6-methylguanine-DNA methyltransferase (MGMT) gene. It repairs the DNA base methylation by a direct damage repair process. The repair process involves the catalysis of the alkylated guanines stoichiometrically and irreversibly by transferring O6-alkyl adduct to active site cysteine (cys145 in human AGT) without damaging the remaining nucleotides in DNA (Tubbs et al., 2007). AGT acts alone to repair methyl lesions, which prevents the unnecessary stress of many molecules during the repairing process. Likewise, this protein performs three different tasks; searches the methyl adduct, binds, and receives the methyl adduct itself. Then, it becomes inactive after accepting the alkyl lesion, so that stoichiometric repair is possible (Liu & Gerson, 2006).

AGT binds to DNA at minor groove with helix turn helix (HTH) motif, identifies the DNA damage, and searches the damaged nucleotide (Daniels et al., 2000). Then, the side chain of arginine residue at 128th position in amino acid sequence in AGT slowly intercalates between the backbones at the damaged region and flips out the damaged nucleotide from the DNA base stack so that the methyl adduct could approach near to the enzyme active pocket. The methyl adduct is then transferred to cysteine at 145th amino acid sequence in AGT forming S-alkyl cysteine (Jena et al., 2009). During the process, the sulfur atom is deprotonated and accepted the methyl adduct. This reaction is called the suicidal reaction of cysteine because of the acceptance of unwanted fragments. After the methyl transformation, the damaged DNA gets free from the methyl lesion (Hu et al., 2008). The methylated cysteine then becomes inactive and does not harm the body cell (Daniels et al., 2000; Falnes et al., 2004).

There are several repair mechanisms of alkylation lesions in DNA. Direct damage reversal and base excision repair, nucleotide excision repair are the major alkylation damage repair processes. Many DNA repair proteins take part in these mechanisms (Dexheimer, 2013). In single step damage reversal, cysteine at the active pocket region of AGT receives the alkylation adduct and let the DNA free from the methylation damage. Moreover, AGT removes the ethyl group from O6-alkylguanine, however, the process is much slower than that in the case of the methylation repair process (Kaina et al., 2007). Human AlkB homologues 2 and 3 (i.e., ABH2 and ABH3) repairs the alkyl damage in diverse points of DNA base like N1MeA, N3MeC, N3MeT, and N1MeG. Both ABH2 and ABH3 repair the same lesion, but they are different in their template specificity (Falnes et al., 2004). ABH2 repairs the alkylating agent at double-stranded DNA and ABH3 repair the alkyl damage in single-stranded RNA substrate. Base excision repair (BER) performs to repair against the alkylation damage like N7MeG, N3MeA, and N3MeG. The glycosylase enzyme is responsible to repair such alkylations. In the beginning, glycosylase recognizes the DNA damage and then proceed with the excision of modified bases (Kaina et al., 2007; Ringvoll et al., 2006).

AGT prevents N1-guanine-N3-cytosine interstrand cross-link damage in the DNA. When chloroethylguanine (-CH₂-CH₂-Cl) is linked at O6 of guanine base, it may have a high probability of cross-linking with N1 intrastrand rearrangement to form the N1-O6-ethanoguanine. This leads to N1-guanine-N3-cytosine interstrand cross-link forming cross-link damage in the DNA (Jaeckle et al., 1998). If AGT can be approached before the cross-linking, it accepts the chloroethyl group, preventing the formation of O6-chloroethylguanine cross-link damage. Besides, the base methylation, MMR deals with mismatches caused by spontaneous or induced base deamination, oxidation, and replication error (Jaeckle et al., 1998; Passagne et al., 2003).

DNA repair protein takes a few minutes to complete the task of the methyl transfer mechanism. Methylated DNA and AGT should form the stable complex as the prerequisite of the transfer mechanism because AGT allows sufficient duration to recognize the alkyl lesion in DNA. Immediately after AGT finds the location of alkyl adduct, one step S_N2 reaction takes place and methyl adduct is transferred to the sulfur atom of the side chain of cysteine residue leaving the DNA non-methylated (Mattosovich et al., 2020). For the quantitative measurement of binding affinity, we calculated binding free energy between methylated DNA and AGT. The change in free energy during the translation of one molecule from another gives the binding affinity between the molecules in the complex. It provides the information regarding the strength of binding as well as the possibility of methyl transfer to AGT active site (Hu et al., 2008). To calculate the binding free energy, we used the umbrella sampling technique to calculate the free energy within the binding site. Since there are several energy barriers present within the configuration states of the complex, it may take several thousand hours to evaluate the free energy to represent all possible configuration states, which is computationally and physically impossible. To eliminate this difficulty, some representative states can be taken so that the free energy change along the path can be calculated (Sugita et al., 2000; Virnau & Müller, 2004; Zang et al., 2005).

The comparison of free energy pre- and post-methyl transfer provides important insight into the methyl transfer process. We calculated the free energy for this methyl transfer process taking pre- and post-transfer complexes. Our presumed that that the free energy change for post-methyl transfer should be smaller than that for pre-transfer (Virnau & Müller, 2004; Sugita et al., 2000; Jayaram et al., 1999). To support our claim, steered molecular dynamics (SMD) (Phillips et al., 2005) was performed to evaluate the force for unbinding of constituent molecules within the complex. This method can reveal the binding strength of AGT and DNA in terms of force. The decoupling force of AGT from DNA in turn gives how strongly the protein binds to DNA.

We also proposed the molecular dynamics study for the stability of DNA-protein covalent linkage. This condition reveals the structural variation of DNA and protein during the

transition state from methylated to non-methylated state in the complex. As the pre-methyl transfer condition gives complex structure before the DNA repair and post-methyl transfer condition conveys the information after methyl transfer to AGT, we designed the intermediate state of these two states of molecule. This intermediate state is transient in nature.

1.6 Rationale of the Study

DNA carries the genetic information for the development, functioning, growth, and reproduction of all kinds of organisms. If some damage occurs in DNA, the corresponding cell does not function properly. The aberrant DNA directly hampers the protein biosynthesis through which the synthesized protein can be mutated (Tubbs et al., 2007). The mutated protein may be carcinogenic. The detection and repairing of such methylation damage in DNA are essential to protect the cell from cancer (Hu et al., 2008).

Previous studies proposed that a protein named O6-alkylguanine-DNA alkyltransferase (AGT) can be a possible candidate to repair methylation damage in DNA, however, its applications against the methylation damage have not been proven yet (McKeague et al., 2018). As such, we carried out a comprehensive study to examine the methyl transfer mechanisms by considering different interaction steps using molecular dynamics (MD) simulations. To perform the work, we prepared the molecular complex taking the initial structure from the protein data bank (PDB) and designed force fields for the methylated segment of both DNA and protein.

Computational investigations, including molecular dynamics (MD) study are attractive complement to experimental investigations. Prediction of molecular behavior during the dynamics is highly beneficial to verify through a wet lab experiment (Banavali & MacKerell Jr, 2002). Therefore, we intended to investigate some important thermodynamic and mechanical characters of molecules in dynamical conditions so that the proposed protein would be suitable as a drug target. This work would be very helpful in designing the drug against cancer caused by methylation at a specific position of guanine.

The present work is focused on the study of methylated guanine-based activity and the mechanism of removing methyl moiety by interacting with the direct damage reversal protein, AGT. Hence, the problem was handled in three different steps to a comprehensive understanding of the methylation transfer process in the targeted molecular complex: entry of methylation at guanine base, intermediate state before and after methyl transfer mechanism the covalent link of the methylated guanine base, and DNA repair protein AGT, and finally the methyl transfer to AGT from DNA. Besides the study of molecular interaction of protein and DNA, molecular dynamics (MD) simulation has been per-

formed to determine the free energy by translating the DNA from binding protein AGT. The free energy approach is a powerful tool to estimate the binding mechanism for large biomolecular complexes (Kästner, 2011). The energy barriers in the configuration states of molecular interaction can be overcome by applying biasing potential for different sample windows of the complex. Moreover, a free energy map of interaction between amino acids and nucleobase can provide the physical basis of discrimination of active residues that take part in the interaction (Virnau & Müller, 2004). The activities of AGT for DNA lesion recognition and repair have not been studied extensively but covers a wide area of research. The molecular dynamics study of such three-step processes of DNA and protein interaction provides the atomic-level study of DNA protein cooperative mechanisms. The accurately designed force fields allow us to obtain a more precise calculation of free energy.

As a close approximation of the transient complex, we designed a covalent linkage between the protein and DNA and performed MD simulations. This work not only investigates the specific problem for the present study but also provides a several avenues for the studies of such an important DNA-protein interaction, including structural stability, and DNA damage-repair process. Moreover, SMD is often executed in the binding mechanisms of protein-ligand molecules. Besides this common application of SMD, we applied this technique in the study of protein-DNA binding mechanism, which is expected as a novel application for the molecular system.

1.7 Objectives of the Study

This study is focused on the interaction between DNA and the DNA binding proteins. Proteins basically interact with DNA by non-covalent bonding methods, nevertheless, in some special conditions, it transiently links to the protein via covalent bonds. During this process, the structure, as well as the physical and chemical properties of both molecules, are altered. In this study, our basic purpose is to investigate the structural and physico-chemical changes in non-covalent and covalent conditions in DNA and protein. In the present study, the methylation damage in DNA was aimed to be non-methylated and hence became free from the cause of cancer.

The study of some basic thermodynamic properties during the interaction of protein and DNA is very important to understand the binding affinity of the corresponding molecules. In the present study, a complex of methylation damaged DNA at the O6 position of guanine and a DNA repair protein O6-alkylguanine-DNA alkyltransferase (AGT) was considered. We purpose the intermediate transient state during the methyl transfer mechanism. This would be the novel structure in the study of DNA and protein interactions. We aimed to form the covalent link of methylated carbon of DNA and

sulfur at the sidechain of cysteine in AGT.

General objective:

Protein repairs several types of DNA damages. The fundamental goal of this work is to study the DNA and protein interaction in covalent and non-covalent linkage to remove the methyl damage from DNA. To fulfill the general objective, the following specific objectives have been setup.

Specific objectives:

- Study the structural features of the DNA-protein complexes.
- Study the interaction sites of the DNA-protein complexes.
- The structural variation of the complex during non-covalent and covalent linkage of the molecules.
- The study of the free energy profile of the DNA-protein complex during non-covalent bonding and analyze energy variations in pre-and post-covalent bonding of interactions.
- Study the effect of DNA base methylation and interaction of such DNA with DNA repair protein.

1.8 Organization of the Thesis

The structure of this thesis is organized as follows:

- Chapter 1 contains a brief introduction of DNA and protein interaction and the repair mechanism of methylation damage in DNA.
- Chapter 2 contains a brief explanation of previous literature in a chronological form which are related to our research field and is named "Literature Review". This section includes the development summary of the works on the different types of DNA discovery, different types of DNA damage, and the repairing techniques. This explanation aims to verify the fundamental basis of our work.
- The theoretical background, necessary formulas, and algorithm that we have used during this research work is covered in Chapter 3 and is named "Materials and Methods". The chapter incorporates the basic concepts of molecular dynamics, its applications, and the necessary mathematical relations to evaluate the thermodynamic parameters in the research work.
- Major outcomes of the entire work of this research are documented and discussed in Chapter 4 "Results and Discussion". In the beginning, the detailed explanation

of the molecular system (DNA-Protein complex) that we used during the work is explained. Then, the binding affinity of DNA with protein is described by different aspects; hydrogen bonding, electrostatics, and van der Waals interactions. The quantitative analysis on decoupling force and free energy of binding of the complex is highlighted for pre-and post-methyl transfer from the guanine base of DNA to cysteine of AGT in the subsequent section. Furthermore, the transient reaction during the methyl transfer mechanism is explained in order to investigate the stability of the covalent complex between the DNA and protein.

- Chapter 5 "Conclusions and Recommendations" expresses the thematic aspects of finding of the entire work and proposes the possible extension of the present work.
- The research work is summarized in the Chapter 6 "Summary".

Finally, the references are listed before closing this document. and published articles, information of participation in scientific conferences and conference certificates are incorporated.

CHAPTER 2

2. LITERATURE REVIEW

Since its identification in 1869 by Swiss chemist Friedrich Miescher, significant advances have been made in the understanding of DNA through experimental and theoretical studies, typically in interpreting the underlying mechanism of DNA and protein interactions (Dahm, 2005). Such interactions in the DNA-protein complex have been analyzed through various aspects including structural modifications, binding affinity, energy variations, etc. For the time being, sophisticated techniques have been developed that could potentially provide a comprehensive study of DNA, its functions, and its aberration close to reality. To elucidate the research progress in DNA-protein interactions, especially the interaction of methylated DNA at the O6 point of guanine and DNA repair protein AGT, we have reviewed and thematically summarized the historical development of DNA discovery, the fundamental mechanism for DNA damage, and its repair.

2.1 Historical Development of DNA Discovery

The major credit for the discovery of DNA double-helix structure was gone to J. Watson and F. Crick, however, its composition and hereditary carrying property had already been studied for decades. For the first time, Swiss physiological chemist Friedrich Miescher, in 1869, proposed the extremely new phosphorus-rich biomolecule into the cell nucleus and he named it 'nuclein' (Miescher, 1869; Dahm, 2005; Avery et al., 1944). Before he discovers nuclein, scientists believed that complex proteins must be the hereditary carrying molecules. On contrary to the existing belief of that period, Miescher experimented and discovered that the cell nucleus contains different material than that of protein. He pointed out that nuclein contains a large amount of phosphorus, which is missing in the composition of protein however, his findings could not be published until his professor Felix Koppe-Seyler repeated this experiment (Dahm, 2005). It took about one year to repeat this experiment and was finally published in 1871 after the verification by Prof. Seyler. After the publication of this work, many other biologists were tempted towards this enticing field of genetic study. About two decades later,

Rechard Altmann renamed 'nuclein' to 'nucleic acid' (Miescher, 1897). Despite many works being carried out at the beginning of the twentieth century, very few noticeable works were published in this field. In the 1940s, two independent works by Oswald T. Avery, Colin MacLeod, and Maclyn McCarthy on the one hand and by Al Hershey and Marta Chase on the other experimentally demonstrated the genetic information-carrying character of DNA. Similarly, in 1953, Pauling and Corey proposed three intertwined chains of DNA structure, having a phosphate near the fiber axis and the bases on the outside, which was later modified by the Watson-Crick model in the same year (Pauling & Corey, 1953).

In 1953, James Dewey Watson and Francis Harry Compton Crick revealed the first insight into the structure of double-helix DNA and its working mechanisms (Watson & Crick, 1953). Their two pages long article "Molecular Structure of Nucleic Acid" in nature publication re-established previous findings, specifically the Meischer discovery of 'nucleic'. In that sense, the Watson-Crick model ultimately revolutionized the biological world. In parallel, Maurice Wilkins had been working on the experimental validation of the DNA double-helix structure from the x-ray diffraction method (Wilkins & Randall, 1953). He also succeeded in presenting the double-helix structure of DNA through x-ray images. For these groundbreaking successes of the discovery of DNA structure, three scientists J. Watson, F. Crick, and M. Wilkins were jointly awarded by Nobel Prize in physiology or medicine in 1962. After the discovery of the Watson-Crick model, the next breakthrough regarding the DNA structure was the discovery of genetic codes provided by the DNA bases. Robert W. Holley and his co-workers cracked out the genetic codes and clearly showed how genetic codes are responsible for the creation of a variety of organisms (Singer, 1968).

2.2 DNA Damage Detection

A large number of DNA damage detection methodologies have been developed following the discovery of the double-helix structure of DNA. Notably, after the advent of DNA sequencing technology in the twentieth century, damage detection became easier and accurate. This method also disclosed the pathway for genetic engineering (W. Li & Sancar, 2020). The research on DNA damage detection was initiated in the 1930s, before the discovery of DNA double-helix structure, however, these studies were not focused on the specific type of damage in DNA. In those periods, some effects on DNA due to the exposure of ultraviolet (UV) rays were reported. Though some deteriorative effects of UV rays were detected in the 1930s and 1940s, the DNA lesion detection was initiated only in 1958 (Rupert, 1960). Varghese and Wang reported that the detection and desolation of DNA damage induced by ultraviolet rays were first successfully

developed in the 1960s by using paper chromatography. Then, a new technology, the radioactive labeling-based method was discovered in 1964. The labeling-based method could successfully detect the nucleotide excision repair in bacteria and mammalian cells (Varghese & Wang, 1967).

Out of many discoveries at the beginning of the 1970s, the fluorescence-based methods were found to be highly effective. Based on this method, acridine orange staining and hello assay was used to detect the breaking of DNA strands (Gruzdev & Kishchenko, 1978). During the same decade, high-performance liquid chromatography-mass spectroscopy (HPLC-MS) was discovered to detect the damage in DNA caused by mycotoxin AFB1. Although HPLC-MS was a sensitive and specific assessment technique, it was not helpful in the detection of oxidative base damages, mainly of guanine (purine base) which is highly prone to oxidation (Essigmann et al., 1977).

DNA sequencing technology revolutionized contemporary research in 1975. Sanger and Coulson reported that this invention was the first generation of sequencing technology through which gene cloning was made possible with gene sequencing and had been a landmark in genetic engineering research (Sanger & Coulson, 1975). Then, in 1980, single-cell gel electrophoresis (comet assay) was invented as new technology to detect the breaking of the DNA strand for a single cell. It is a highly sensitive and versatile method to detect low levels of DNA damage in various cells (Ostling & Johanson, 1984; Freeman et al., 1986). In the same decade, immunoassay-based methods were also invented and were used to quantify UV-induced DNA damage (Mitchell & Clarkson, 1981). Afterward, different kinds of polymerase chain reaction (PCR) including quantitative PCR (qPCR) and ligand-mediated PCR (LMPCR) were widely used for mapping DNA damage at nucleotide resolution in the 1990s (Pfeifer et al., 1991). The TdT-mediated dUTP-biotin nick end labeling (TUNEL) assay was used to label DNA breaks in situ for investigating apoptosis in 1992 (Gavrieli et al., 1992). No significant invention regarding the damage detection was found between 1992 and 2009.

In the early 2010s, the combined use of immunoprecipitation and microarray was reported for mapping UV-induced DNA damage at the chromosomal scale in humans (Cadet et al., 2005). Since 2014, sophisticated techniques have been developed which revolutionized the invention and detection of genomics damages (Zavala et al., 2013). In the last 5-6 years, several next-generation sequencing (NGS)-based methods have emerged for detecting various types of radiolabeling-based techniques. The latest technology regarding sequencing is also called the third-generation technology which was developed for detecting ribonucleotide incorporation and UV-induced DNA (R. P. Sinha & Häder, 2002).

2.3 DNA Damage and Repair Mechanism

The advanced biochemical and biophysical methods allow an in-depth study of DNA structure and its interaction with proteins and other biomolecules whereas computed radiography (CR) and Nuclear Magnetic Resonance (NMR) technologies have made the atomic level research to make more comprehensive and accurate. Similarly, the integration of wet-lab experiments and high-speed computational platforms significantly improved the reliability and time-constraint in the study. While new technology expedites in searching the damage and its recovery, the protein data bank (PDB) structures modeling from x-ray diffraction (XRD) and NMR technologies have revealed an alternative way of study in cellular mechanism and sequencing in DNA and protein molecules (Berman et al., 2002).

Even though there are myriads of factors that can potentially damage DNA, they are broadly categorized into two major factors and are very critical: endogenous and exogenous (De Bont & Van Larebeke, 2004; Friedberg et al., 2004). Endogenous factors basically incorporate the toxins that we intake from food and water, whereas exogenous factors include the UV radiations received from the Sun and radioactive sources of the earth (Ames, 1989; Tubbs et al., 2007). Wood et al., in 2001, mentioned that the DNA base is highly vulnerable to environmental factors like inhaled cigarette smoke or incompletely defined dietary factors. They also emphasized the endogenous factors in the DNA damage process. It is to be noted that the extent of DNA damage is not uniform to all people, even though they live in similar environmental conditions. The DNA damage response varies considerably from tissue to tissue and person to person, in accordance with the genetic and epigenetic mechanism (Wood et al., 2001; Essigmann et al., 1977).

Besides many causes of DNA damages, there are several pathways of DNA repairs in organisms: Direct damage repair (DDR), Base excision repair (BER), and Mismatch repair (MMR), nucleotide excision repair (NER), homologous recombination (HR) and non-homologous end-joining (NHEJ) as summarized in schematic diagram in Figure 5. In the Figure 5, sources of damaging agents, their corresponding damaging effects and repairing mechanisms are summarized.

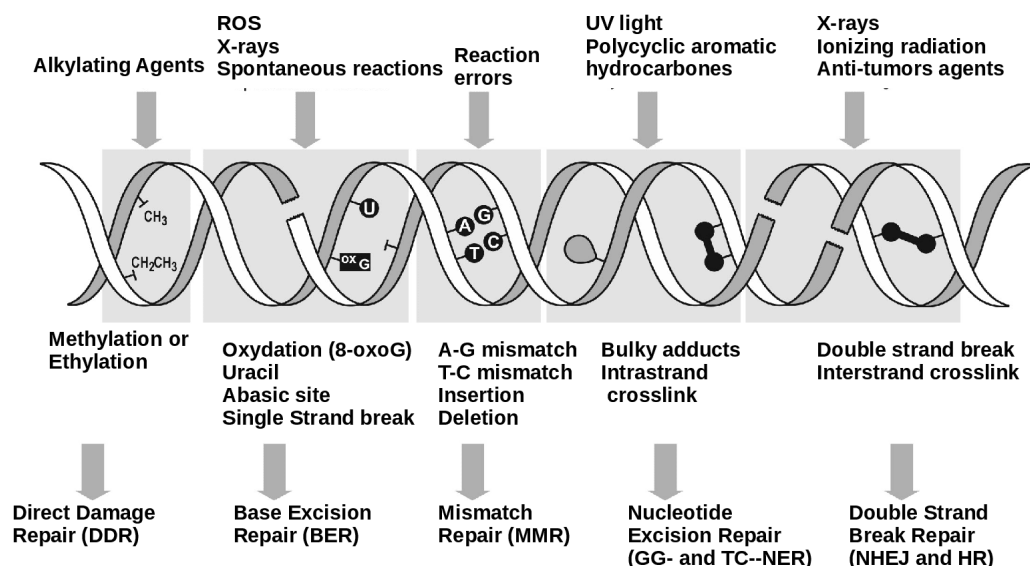


Figure 5: Schematic diagram showing DNA damage, possible major effects of corresponding damage and repair mechanisms. First row shows the damaging agents of DNA, second row shows the possible effects of such damaging agents and the third row indicates the corresponding repairs pathways.

2.3.1 Direct Damage Reversal (DDR)

Methylation and ethylation are the basic alkyl damages in DNA. In 1982, Rydberg and Lindahl experimentally determined the methylation at 7-methylguanine (7-meG) and 3-methyladenine (3-MeA), by incubating DNA with S-adenosyl-L-methionine (SAM) in a neutral aqueous solution. In such a process, the non-enzymatic reaction of SAM acts as a methyl donor and provides the background of mutagenesis and carcinogenesis (Rydberg & Lindahl, 1982). This in turn supplies the information for the discovery of methylation damage reversal enzyme site. David T. Beranek reviewed the previous works regarding the environmental toxin that causes both methylation and ethylation in DNA. He categorized the alkylating agents into three groups: alkyl sulfate, alkyl alkanesulfonates, and nitrosamides (Beranek, 1990).

The DDR mechanism is the one step reaction between DNA and protein. O6-alkylguanine-DNA alkyltransferase (AGT) which is also known as methyl guanine-DNA methyltransferase (MGMT) is considered as the methylation removing agent in DNA. In such a process, the cysteine in AGT accepts the methyl adduct in guanine base by the suicidal reaction. AGT protein is inactivated after receiving the methyl adduct resulting in the DNA which is free from methylation damage. This single protein performs all steps along the methylation elimination mechanism; recognizing the damage, interaction with DNA at the damaged site and receiving the methyl adduct (Pegg et al., 1988).

2.3.2 Base Excision Repair (BER)

The simplest form of endogenous DNA damage is hydrolysis. In DNA, usually, hydrolysis acts in the bond between C1 of deoxyribose sugar and N3 of the base known as N-glycosidic bond, which is prone to acid-catalyzed hydrolysis (Lindahl, 1993). This results in the abasic lesion owing to the loss of purine and pyrimidine bases. This type of base loss may occur due to the error in the base excision repair mechanism. The hydrolysis process may convert the cytosine into uracil due to the spontaneous deamination process which may occur in guanine and adenine. Moreover, the reactive oxygen species (ROS) like O_2^- , H_2O_2 , and $*OH$ can generate several different oxidative DNA adducts, such as base modification, deoxyribose oxidation, single or double-strand breakage, and DNA-protein crosslink (Apel & Hirt, 2004; Marnett, 2000).

DNA base excision repair (BER) is a technique of maintaining generic integrity by removing the damaged part caused by hydrolysis and oxidative reaction. BER mechanism is completed in four steps: incision, end processing, repair synthesis, and ligation. This technique plays an important role in dictating the cellular response to various chemotherapeutic agents and radiotherapy. Indeed, the excision repair mechanism protects DNA against cancer, aging, and neurodegradation (Dexheimer, 2013).

2.3.3 Mismatch Repair (MMR)

Physiological DNA reactions may cause a mismatch reaction in purine and pyrimidine bases leading to a replication error. The mismatch damage occurs due to the incorrect binding of nucleobases of a DNA forming intra- or inter-strand hydrogen bonding. During the replication, the replicated DNA polymerase incorporates the specific base pairs in accordance with the original template of the DNA strand. However, the base pair insertion is not always correct. Sometimes the polymerase incorporates the erroneous pairing; A to G or C. Likewise, deletion or insertion of an unnecessary nucleotide may occur during the replication. The mismatch error can be raised in the base excision repair mechanism (McCulloch & Kunkel, 2008; Modrich, 2006).

Modrich, in 2006, described the mismatch repair process in his review article suggesting that this type of the repair process is mainly caused during DNA replication. MMR is ubiquitous in almost all organisms ranging from bacteria to human, however, its mechanism has been understood in human type and *Escherichia Coli* type. The article highlighted the molecular features of two homologous MutS and MutL as the key enzymes in MMR. Moreover, MutH is a member of the type II family of restriction endonucleases which cleaves the mismatch containing DNA single-strand specifically at hemimethylated (i.e., only one strand methylation) GATC sites. Further, O₆-chloroethylguanine

forms the intramolecular rearrangement to form N1-guanine-N3-cytosine interstrand crosslink (Modrich, 2006).

2.3.4 Nucleotide Excision Repair (NER)

Besides the endogenous factors, many exogenous factors are also highly responsible for DNA damages. Ultraviolet (UV) rays are the major exogenous factor for DNA damage. UV rays induce the physical stress in DNA strands and can form the atypical covalent bond between the adjacent pyrimidine bases via an intrastrand crosslink. Likewise, polycyclic aromatic hydrocarbons, which are commonly found in the diet and harmful compounds produced from cigarette smoke and vehicle exhaust react with DNA backbones or nucleobases can form the bulky adducts. These types of DNA lesions are repaired by the nucleotide excision repair (NER) mechanism (Dexheimer, 2013).

NER is a versatile repair method that can recognize and remove the bulky adducts, intrastrand crosslink, and helix distorting damage in DNA. More significantly, this mechanism is useful in removing the cyclobutane pyrimidine dimers (CPD) and 6-4 photoproducts, generated by UV rays. About 30 different proteins work for the "cut and paste" method to repair such lesions (Shuck et al., 2008). NER process is completed in several steps: DNA damage recognition, local opening of the DNA helix at the vicinity of the lesion, excision of damaged nucleotide, insertion of correctly synthesized nucleotide, and strand ligation. Several DNA repair proteins are employed to perform the entire work. Besides having many advantages, NER may cause DNA damage during nucleotide selection. If the replaced nucleotide is incorrectly inserted into the DNA strands, it can cause genetic disorders (Cleaver et al., 2009; Vermeulen et al., 1997).

2.3.5 Double-strand Break (DSB) and Interstrand Crosslink Repair

Highly energetic electromagnetic radiations like x-rays and gamma rays which are produced from hospitals and natural sources can induce double-strand breaks. Similarly, chemotherapeutic drugs such as topoisomerase I or II inhibitors can also generate the single- or double-strand break by trapping the topoisomerase-DNA covalent complex (B. K. Sinha, 1995). Similarly, some anticancer drugs like methyl methane sulfonate and temozolomide cause the intrastrand and interstrand crosslink. Double-strand break (DSB) is the most complicated as well as a hazardous lesion in DNA. This type of devastating defect can lead the cell apoptosis and the development of cancer cells. These types of complex damages are repaired by two mechanisms in the cell: homologous recombination (HR) and non-homologous end-joining (NHEJ) (X. Li & Heyer, 2008).

HR is usually the error-free mechanism in which the broken region of the DNA strand is joined by protein regarding the genetic information contained in undamaged sister chromatid as a template. It works in three steps: presynapsis (surrounding the DSB with protein), synapsis (conjunction of two broken regions), and postsynapsis (error-free correction of DSB). In contrast to HR, NHEJ eliminates the broken part by direct ligation, which is largely error-prone (McIlwraith et al., 2005; Dexheimer, 2013).

2.4 DNA Methylation at Guanine

Coulondre and Miller identified methyl lesions at the O6 point of guanine base of DNA for the first time in 1977 (Coulondre & Miller, 1977). They reported that the mutagenic character of O6-MeG tends to pair with thymine during replication and to convert the guanine-cytosine pair into adenine-thymine pair (Erickson et al., 1980; Gerson et al., 1986). A similar result had also been detected by Erickson et al. in 1980 by examining the activity of O6-MeG and detected the crosslinking of O6MeG with opposite cytosine. They found that it does not block the DNA replication, rather shifts the G: C pair to the A: T pair in a newly transcribed DNA strand (Pegg, 1990).

In body mechanisms, DNA binding proteins, AGT, serve as the demethyl agent. Cell with deficit AGT is more prone to be methylated. The probability of demethylation of DNA is directly related to the number of AGT in the targeted part of DNA damage (Gerson, 2002). It can remove the light types of alkylating agents like methyl adduct CH_3 , however, the heavy adducts like ethylation (CH_2CH_3) cannot be repaired by AGT. Such heavy adducts are repaired by nucleotide excision repair (NER). As it repairs the alkyl damage in DNA without base excision, genetic information is preserved. Pegg (1990) identified the AGT as the demethylating agent that can transfer the methyl adduct (CH_3) to one's sulfur at the side chain of cysteine by the suicidal reaction. After receiving the methyl adduct, AGT becomes degraded and inactivated, but may not detach from DNA.

Alkylating agents are important drugs in cancer chemotherapy. Both methylating and chloroethylating agents produce O6-alkylguanine damage in DNA. The non-methylating agents like AGT work in the reverse reaction than that of alkylating agents, i.e., it prevents the cell from alkylation so that the cancer cell can be survived. To support this fact, the AGT cDNA transfection experiment demonstrated that AGT provides expression dependent resistance to methylating and chloroethylating drugs. This reveals that it plays a negative role during cancer chemotherapy (Ishiguro et al., 2010).

Daniels et al. (2000), examined the responses of AGT in the alkyl DNA damage detection and binding mechanism with the O6MeG. They observed that the asparagine hinge

together with the helix turn helix (HTH) motif contributes a major role in forming the mDNA-AGT complex. To facilitate methylated guanine rotation from its base stack to approach in the active pocket of AGT, arginine residue intercalates into the DNA backbone and then provides the stability to cytosine which previously pairs to methylated guanine with two hydrogen bonds instead of regular three hydrogen bonds (Daniels et al., 2000). Further, Duguid et al. (2003) experimented to probe the alkyl damage searches mechanism by using two O6-alkylguanine-DNA alkyl transferases, the E Coli CAdA and the human AGT. Their result showed that there is no necessity of flipping back of damaged nucleotide to recognize the lesion, instead, they can locate by sensing the unstable nature of the damaged base pair.

In order to study the methyl transfer mechanism from O6-MeG of DNA to a targeted cysteine residue in AGT, Noll & Clarke (2001), prepared a covalent crosslink between DNA and protein via N1, O6-ethanoxanthosine. The formation of this complex was dependent on both active sites of cysteine and N1, O6-ethanoxanthosine. The outcomes of this complex revealed that the crosslinking between them was detected stable and the resulting complex appeared to be well suited for further biochemical and biophysical characterization in DNA-protein covalent complex.

Rasimas et al., in 2003, examined the structural stability of complex formation of AGT with single stranded DNA (ssDNA) and double stranded DNA (dsDNA) by using the analytical ultracentrifugation and electrophoretic mobility shift assays. The experimental results indicated that the stoichiometry complex formation is 4:1 on ssDNA and dsDNA with AGT. Additionally, there is no significant change in complex formation, despite the mutation of a large proportion of the active sites of DNA. They also argued that the outcomes of the experiments can provide significant implications in the formation of the AGT-DNA complex and its structural stability (Sabharwal & Middleton, 2006).

Daniels et al. (2004), pointed out the methyl transfer mechanism from nucleotide to AGT active region by damage reversal process. AGT was found to be binding at the minor groove of methylated DNA via helix turn helix (HTH) motif which in turn induces methylated nucleotide with arginine finger. Arginine forms a hydrogen bond to cytosine, pair partner of flipped out guanine, and send the methyl adduct to the AGT active pocket. Then, the methyl adduct is transferred by deprotonating the sulfur in cysteine. In such a process, a single molecule of AGT is sufficient to perform this entire mechanism and the process is completed within a few milliseconds (Lindahl et al., 1982). Regarding the AGT expression, the highest AGT expression occurs in the livers, lungs, kidneys, and colons, however, the lowest level of expression takes place in the pancreas, hematopoietic cells, and lymphoid tissue (Zang et al., 2005).

Daniels and coworkers also reported the protein data bank (PDB) structure of human

AGT in a complex having double-stranded DNA. The designed PDB structure from their study could be the landmark in the computational study of the interaction between protein and DNA. The designed structures are two types: one contains the non-covalent bonding on AGT and mDNA at O6-position of guanine and the other contains the covalent cross-linking of sulfur of cys145 to N1-O6-ethnoanthosine. First architecture facilitates the study of the DNA lesion recognition and protein binding mechanism in the minor groove of DNA, which is considered as the rare binding location of the protein in DNA. Furthermore, this structure is useful in the understanding of nucleotide flipping and methyl transfer process from O6-position of guanine to AGT active pocket cysteine. Similarly, the second structure is designed to investigate the preventive measure of DNA from methylation and ethylation (Daniels et al., 2004).

Zang et al., in 2005, estimated that the rate of individual steps in the removal of alkyl groups from O6-methyl (O6-MeG) and -benzyl (O6-BeG) guanine in oligonucleotides with human AGT by using rapid reaction kinetic methods. This experiment describes the overall pattern of the rate of methyl group transfer versus AGT concentration. Besides this, they measured the effects of mutations and the affinity of recognition of O6-alkylguanine lesion. The overall reaction showed that the reaction rate versus concentration is 100-folds faster in O6-BeG than that of O6-MeG. The dissociation of AGT from alkylated and normal DNA were observed with different rates. The rate measurement experiment determined that the dissociation rate is higher in normal DNA than that of alkylated DNA showing the favorable binding of AGT with methylation damaged DNA (Rasimas et al., 2003).

Hu et al. (2008) observed two steps of nucleotide flipping processes that enable to recognize the lesion in DNA, rather than a thermodynamic, gate-keeping strategy for lesion discrimination. They made the connection with contemporary research on single-molecule studies of DNA-repair proteins sliding on DNA to understand how they sense minute chemical differences between bases efficiently. Afterward, Jena and coworkers in 2009, studied the major interacting amino acids of AGT to mDNA by using density functional theory (DFT). They proposed three steps reaction processes during the elimination of methyl lesion at the O6 position of guanine. In their study, they took six amino acids Cys145, His146, GLU172, Tyr114, Lys165, and Ser159 of AGT as the catalytic amino acid residues and investigated their role during the stable complex formation of AGT and mDNA and the mechanism of methyl transfer strategies. In such a process, Cys145-water-His146-Glu172 tetrad is converted into cysteine thiolate anion followed by the deprotonation of Tyr114 by N3 site of O6-MeG and finally, methyl group transfer would be possible to Cys145 of AGT (Jena et al., 2009).

McKeague et al. (2018) experimented to draw the relationship of methylated guanine at O6 (O6-MeG) with all canonical nucleotides and then tested the possibility of binding

AGT with various possible base pairs of nucleotides taking the frequently mutated 12 of the K-RAS gene. They observed that the rate of repair mechanism of O6-MeG was lowered by 2-fold when O4-MeG pairs with guanine (G), whereas all other canonical bases have no impact on the binding rate. Resulting 2-folds of decrease in repairing is due to the syn conformation of the glycosidic bond precluding the formation of a repair active complex. This experiment not only studies the possibility of the binding rate of hAGT with possible nucleotide pairs, but it also reveals the biochemistry of damage repair processes.

CHAPTER 3

3. MATERIALS AND METHODS

Present work is based on the molecular dynamics simulations to investigate the protein and DNA interactions. A simulation package NAnoscale Molecular Dynamics (NAMD) (Phillips et al., 2005) has been used for MD simulations. Visual Molecular Dynamics (VMD) (Humphrey et al., 1996), pymol (DeLano, 2002) and Microsoft excel were used to visualize the molecular structures as well as analyze the outcomes obtained from the simulations. Several analysis tools were also exploited to estimate the computational results, and the results were compared with previously published research articles.

In this research work, DNA and protein structures were modified by adding methyl adduct at certain points, and they were also linked with covalent bond to study the transient state of methyl transfer. To perform MD simulation, the ordinary force fields were insufficient and hence new force fields were designed and implemented during the system preparation and MD simulations. The newly designed force fields are incorporated in the later part of this section. This section begins with the concepts and applications of molecular dynamics methods that we have used in the following sections. Then, the technical details of system generation, steps of MD simulation run are presented in detail. All associated methods of result analysis and related mathematical relations will be explained to evaluate the thermodynamic parameters like RMSD, RMSF, SASA, free energy, rupture forces, etc. Bonded and non-bonded energy functions are also discussed which helps to understand the biochemical properties of the molecular systems.

3.1 Molecular Dynamics

Understanding the biomolecular activities is crucial on analyzing the life process. The phenomenon of jiggling and wiggling of atoms and molecules governs the mechanism of living systems and thus applying the atomic/molecular approaches helps in reducing the complex biological functionalities in the regime of chemistry and physics (Rapaport, 2004). The interactions of biomolecules can be modeled by considering many body problems from the aspect of classical physics, for example a bond between two atoms

can be considered as two bodies connected with virtual spring that is oscillated across a certain amplitude. The molecular dynamics simulations utilize the laws of classical mechanics to the microscopic particles even for the nuclear motion in an atom. Classical concept can be the appropriate approach to analyze the many body systems for the prediction and estimation of various mechanical and thermodynamic problems associated with biopolymers (Allen & Tildesley, 1991; Frenkel & Smit, 2002).

Molecular dynamics simulation techniques mimics the real experiments and extracts the information from the simulating systems by employing the mathematical relationships from statistical mechanics. It is the computational technique that describes the various aspects of molecular mechanism in the suitable environment (Rapaport, 2004). This computational method calculates the time-dependent behavior of a molecular system. Bulk properties/macrosopic properties of a system is related to the molecular level. The connection between microscopic simulations and macroscopic properties could be connected via statistical mechanics which provides the rigorous mathematical expressions that relate macroscopic properties to the distribution and motion of the atoms and molecules of the N-body system (Dill & Bromberg, 2010; Chandler, 2005). Thus, the MD simulation method can compute the mechanical, thermodynamic, electronic, magnetic properties etc., of the many body system.

In classical molecular dynamics, Newton's equation of motion is solved for a system of N number of particles interacting through a potential energy $V(r_i)$ that is assigned by the appropriate parameters for bonded and non-bonded interactions (Karplus & Petsko, 1990; Lemkul, 2020). Newton's equation of motion for i^{th} particle in a system is stated as,

$$m_i \frac{\partial^2 r_i}{\partial t^2} = -\nabla V(r_i) = F_i \quad (3.1)$$

In the Equation 3.1, m_i is the mass and r_i is the position of the i^{th} particle. The negative gradient of potential energy in the right hand side (RHS) gives the force on the corresponding particle of the system. MD simulations have provided detailed information on the fluctuation and conformation changes of proteins and nucleic acids. For N number of particles in a system, there should be N number of Equations. Moreover, all such N number of equations should be solved in each time steps in the MD simulations. Solving these equations manually for each time step of simulation is impossible. So, special computational programming called the computational simulation is used. The simulation programs are able to solve these Equations for each time step $\sim 10^{-15}$ s. The calculated values are printed in output files describing the trajectory of the molecules (Allen & Tildesley, 2012).

3.1.1 Force Calculation

Calculation of force on each particle within the periodic box is one of the challenging as well as the time consuming procedure in molecular dynamics simulations. The force on a particle is evaluated from the negative gradient of potential $V(r_i)$ and is determined by pairwise addition of particles. $V(r_i)$ is the pairwise additive interaction potential which is determined by the contributions to the force on i^{th} particle due to all its neighbours. This implies that there are $\frac{N \times (N-1)}{2}$ pair distances for a particle to interact with remaining particles in the system having N particles (Makov & Payne, 1995). This expresses that the time required for the evaluation of the force on a particle by the remaining particles scales as N^2 . More challenging part in molecular dynamics is the calculation of each of the particle with every neighbouring particles is to be calculated. This creates the serious problem in the computational work. To resolve such problem, periodic boundary condition is used and the cutoff distance is specified in such a way that multiple calculations of the force between same particle pairs are avoided (Shimada et al., 1994; Thompson et al., 2009).

3.1.2 Periodic Boundary Condition (PBC)

To perform the MD simulation, a molecule or a complex of molecules is to be kept into simulation box by solvating in water and ion environment. During the simulations, the contents inside the box move randomly so that large number of molecules collide on the inner wall of the box. The molecules on the boundary experience different force than the molecules well within the box. These types of edge effects in MD simulation are tackled by implementing the periodic boundary condition (PBC). In PBC, all the molecules are enclosed in a box and is replicated in all the dimensions to form an infinite lattice so that each particle interacts with periodic images of particles in the same system. When a molecule leaves the original central box, one of its periodic image enters through the opposite face such that the number density in the original simulation box remains conserved (Friedman et al., 2013; Braun et al., 2019).

One of the major problem in PBC is the possibility of interaction of the actual molecule with its image, which is undesirable. To prevent the system from this situation, a cutoff distance for the non-bonded interactions is so chosen that it is less than half the length of the simulation box. The schematic diagram on the implementation of PBC is shown in Figure 6.

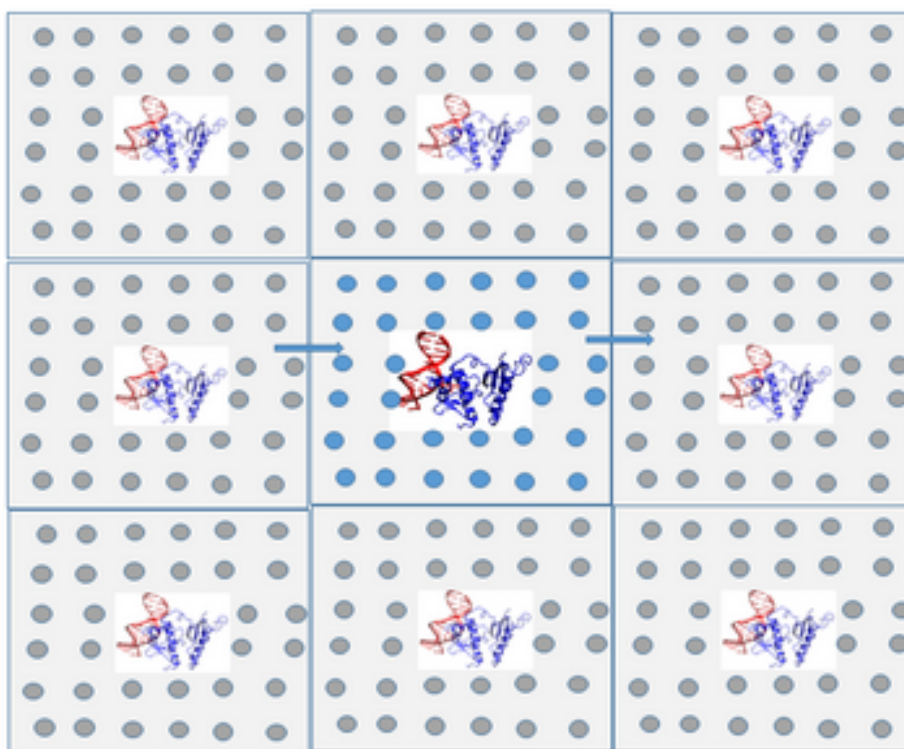


Figure 6: Periodic Boundary Condition (PBC) showing the original simulation box and some replicated images. The replicated images are assumed as infinitely large numbers so that no boundary effect can be observed.

3.1.3 Particle Mesh Ewald (PME)

To handle the long-range interactions, especially the coulomb interaction between the charge particles is a complicated task in computational simulations. To resolve such a problem, Particle Mesh Ewald (PME) summation is commonly used. This summation technique is based on replacing the point charge distribution by a continuous charge distribution with appropriate charge density. Furthermore, it is a grid based technique in which the charge density is calculated for each grid and potential is calculated taking the charge density of the grids (Darden et al., 1993). Overall, there are 5 steps to deduce the potential using the PME technique:

- First, charge is assigned to each of the grid point.
- Then, the grids are transformed into the reciprocal space. A Fast Fourier Transform (FFT) is used to convert the charges on the grid to their equivalent in the Fourier space.
- Poisson's equation in Fourier space is employed to calculate the reciprocal space potential which is stored on the grid.
- Grid is then transformed to the real space by using inverse FFT.

- Finally, force is calculated from the gradient of potential.

3.1.4 Temperature and Pressure Control

Thermostats

Molecular dynamics simulations are performed under a specified thermodynamic ensemble in order to mimic the experiments that are performed in laboratory. For this, the temperature of a system should be controlled at a specific value. Thermostats are used to maintain the temperature of the system at that desired value. In MD simulations, thermostats are set by developing the algorithms. The temperature that is set for a molecular dynamics simulations is measured by using the kinetic energies from the equipartition theorem,

$$N \left(\frac{3}{2} k_B T \right) = \left\langle \sum_{i=1}^N \frac{1}{2} m_i v_i^2 \right\rangle \quad (3.2)$$

The time average of kinetic energies of N-particles in the system gives the instantaneous temperature. The temperature calculated from this method may not be equal to the targeted temperature rather it fluctuates around that point (Braun et al., 2019; Zuckerman, 2010). Some thermostats which are in practice are Gaussian, Simple velocity rescaling, Berendensen, Bussi-Donadio-Parrinello, Ansersen, Langevin, and Nose-Hoover.

Barostats

As the thermostats control the temperature of a molecular system, barostats are designed to control the pressure. The algorithms prepared for the barostat control the pressure. This works together with thermostats and hence both temperature and pressure are controlled (Tuckerman, 2010). There are some popular barostats in practice: simple velocity rescaling, Berendensen, Andersen, Parrinello-Rahman, Martyan-Tuckerman-Tobias-Klein and Monte Carlo.

3.1.5 Force Fields

Force fields for MD simulation comprises topology information and empirical potential energy functions that are effective among the atoms in a molecular system and its surroundings. The common potential function includes the bonded and non-bonded terms. Bond, angle, proper dihedral and improper dihedral belong to the bonded terms, whereas Coulomb and Lenard Jones terms lie within the non-bonded potential function.

Improper dihedral is included in the potential function to enforce planarity of cyclic portions of the molecules (Leach & Leach, 2001). The result obtained from the same force field may be influenced by many other factors: cutoff parameters of long range forces, selection of water models, and the type of statistical ensemble under which the MD simulation is carried out. The topology and parameters for ordinary amino-acids and DNA nucleotides are available in the existing force fields like CHARMM, GROMOS, OPLS, etc; however, topology and parameters should be designed for the newly discovered ligands and mutations in ordinary molecules (Best et al., 2012). New force field designing might be challenging: special care should be given during the designing of parameters for such molecules.

Before starting computational simulation, we should prepare the molecular system that resembles the real environment of our body cell containing the targeted molecule surrounded with water and ions (Klauda et al., 2010). To do so, we need force field information which are categorized into topology and parameters. In entire simulations, we have used the CHARMM36m force fields.

Topology Files

Topology file contains the fundamental information for the computational construction of the molecular structure (Michaud-Agrawal et al., 2011). Protein data bank (pdb) provides only the limited information for the protein structure (Berman et al., 2000). The topology file defines partial charges, assigns the atoms that are connected to one another through chemical bonds, and incorporates group of atoms that form angles and dihedrals. Furthermore, the internal coordinates are also assigned with the coordinates of hydrogen atoms and other atoms missing from a crystal PDB file (Wang et al., 2004).

Aside the ordinary force fields, we have designed additional topology files for the modified structures. The additional topology files were designed for methylated guanine at O6 point, methylated cysteine at sulphur atom, and the covalent linkage between sulfur atom of cysteine and methylated carbon at guanine atom. After implementing all necessary requirements during the system preparation, the appropriate systems could be constructed.

Parameter Files

Parameter files contain all of the numerical constants that are needed to evaluate the energy and forces of the atoms involved in the interactions. psf file provides the molecular structure information and the pdb file provides the atomic coordinates. In every step of system preparation, corresponding psf file is generated along with the pdb file. Parameter

files are inter-related to topology files that provide bonded and non-bonded information and are implemented during the energy calculations. Parameter files contain basically the constants for both bonded and non-bonded terms (Lee et al., 2016; Feller & MacKerell, 2000).

Bonded terms comprises the bond stretching, bond angle, dihedral angle and a out of plane angle (called the "improper" dihedral). The assigned values for the bond stretching (entry parameter "BONDS") are two atom types, spring constant and equilibrium length. Each type of entry is available only in nearest covalent bond of consequent two atoms. Similarly, the entry parameters in "ANGLES" keyword consists of three atom types, a spring constant and an equilibrium angle. There is a minority like term between the first and third atoms shows the bonded stretching term during the formation of angle and is called the Urey-Bradly term. The constants associated in these conditions are calculated with the relation $V_{(U-B)} = k_{ub}(s - s_0)^2$ (Phillips et al., 2005).

In the next section, the parameters are assigned for every type of dihedral present in the topology file. Actually, dihedral represents the energy of rotation around a covalent bond. The entry parameters of dihedral terms are four atom types, periodicity and angle between the planes of first three and last three atoms. Dihedral angle is the source of flexibility in biomolecules. The final bond like term in the parameter file are improper. It is used exclusively and explicitly in the molecular biology to maintain the planarity. The backbone dihedral energy has been corrected by CMAP (Phillips et al., 2005; Jorgensen et al., 1983; MacKerell Jr et al., 1998).

3.1.6 Modeling of the System

In classical force fields, the potential functions are derived empirically to describe atomic interactions. The atoms are treated as spherically symmetric particles that are connected through covalent bonds to form molecules. Each atom experiences a force resulting from its pairwise interaction with the rest of the atoms of the system. The total potential V_{total} is calculated by summing over all bonded and non-bonded interaction potential terms (Jorgensen et al., 1983; MacKerell Jr et al., 1998) and are given by,

$$V_{total} = V_{bond} + V_{angle} + V_{dihed} + V_{impr} + V_{vdw} + V_{coulomb} \quad (3.3)$$

In Equation 3.3, V_{bond} representing bond stretching potential, V_{angle} representing bond angle potential, V_{dihed} representing proper dihedral potential and V_{impr} representing improper dihedral potential are included in the bonded potentials. Similarly, the non-bonded terms are V_{vdw} representing van der Waals potential and $V_{coulomb}$ representing coulomb potential. The bonded terms- V_{bond} , V_{angle} , and V_{impr} are harmonic potentials,

but the term V_{dihed} is a periodic potential as shown in Figure 7.

Bonded Interactions

Bonded interactions are based on a fixed list of atoms. They are not exclusively pair interaction, but 3- and 4-body interactions as well. Bond stretching (2-body), bond angle (3-body), and dihedral angle (4-body) interactions are included in these categories.

Bond Stretching potential

Any two covalently bonded atoms are physically modeled by providing a harmonic potential. During the modeling, two such atoms are assumed to be connected by a spring with certain force constant k . The value of k is specific for each pair of bonded atoms and are taken from the experiment. The energy of a bond is approximated from the function of displacement from the ideal bond length b . The bond stretching between two covalently bonded atoms i and j is represented by harmonic potential (Phillips et al., 2005) as,

$$V_{\text{bond}}(r_{ij}) = \frac{1}{2}k_{ij}^b (r_{ij} - b_{ij})^2 \quad (3.4)$$

where, b_{ij} is the equilibrium bond length and k_{ij}^b is the force constant.

Also, the corresponding force is given by,

$$F_i(r_{ij}) = \frac{1}{2}k_{ij}^b(r_{ij} - b_{ij})\frac{\mathbf{r}_{ij}}{r_{ij}} \quad (3.5)$$

Figure 7 shows the potential due to bond stretching.

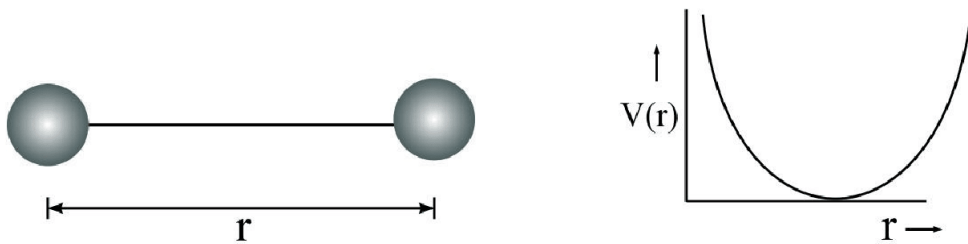


Figure 7: Schematic diagram of bonded stretching and corresponding nature of graph.

Bond angle potential

The bond-angle vibration between a triplet of atoms $i - j - k$ is also represented by a harmonic potential. Its nature is similar to the bond stretching, but the force constant is assigned for a variation of angle formed by three consecutive atoms. The harmonic

potential representing bond-angle vibration (Jorgensen et al., 1983) is given by,

$$V_{\text{angle}}(\theta_{ij}) = \frac{1}{2}k_{ijk}^{\theta} (\theta_{ijk} - \theta_{ijk}^0)^2 \quad (3.6)$$

where, θ_{ijk}^0 is the equilibrium bond angle and k_{ijk}^{θ} is the force constant. Figure 8 shows the potential due to bond-angle oscillation.

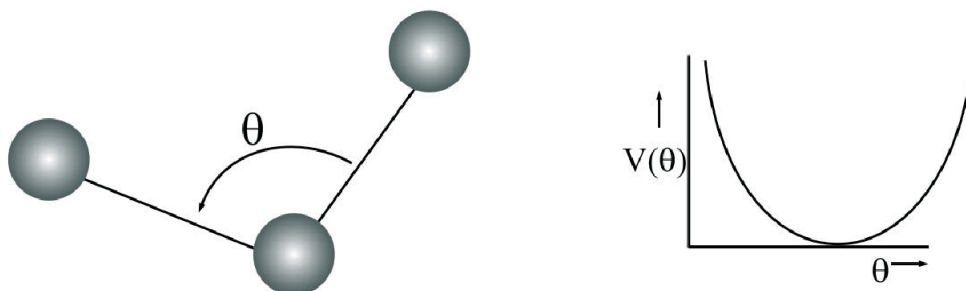


Figure 8: Schematic diagram of harmonic angle variation and corresponding nature of graph.

Proper Dihedrals Potential

This is the torsion angle potential function and is defined with proper dihedral angles in accordance to the IUPAC-IUB convention, where ϕ is the angle between the ijk and jkl planes as shown in Figure 7(c). The periodic potential (Jorgensen et al., 1983) is defined as,

$$V_{\text{dihed}}(\phi_{ijkl}) = k_{\phi}(1 + \cos(n\phi - \phi_s)) \quad (3.7)$$

where, ϕ is the dihedral angle, n is the integer which determines the periodicity of the potential (number of minima/maxima) in the interval $[0, 2\pi]$. Likewise, ϕ is the angle between the plane containing the first three atoms in the dihedral and the plane containing the last three and ϕ_s is the equilibrium angle in degree. The nature of proper dihedral potential is shown in Figure 9.

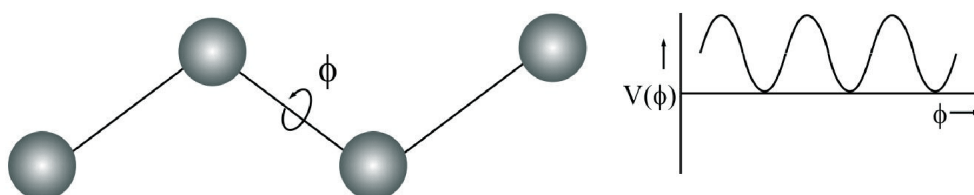


Figure 9: Schematic diagram of dihedral angle and corresponding nature of graph.

Improper dihedral potential

The improper dihedral angle ξ is defined as the angle between planes (i, j, k) and (j, k, l) in all cases as shown in Figure 10. The simplest improper dihedral potential is a harmonic

potential (Phillips et al., 2005) which is given by,

$$V_{\text{improper}}(\xi_{ijkl}) = \frac{1}{2}k_{\xi} (\xi_{ijkl} - \xi_0)^2 \quad (3.8)$$

The variation of improper dihedral potential with dihedral angle is shown in Figure 10.

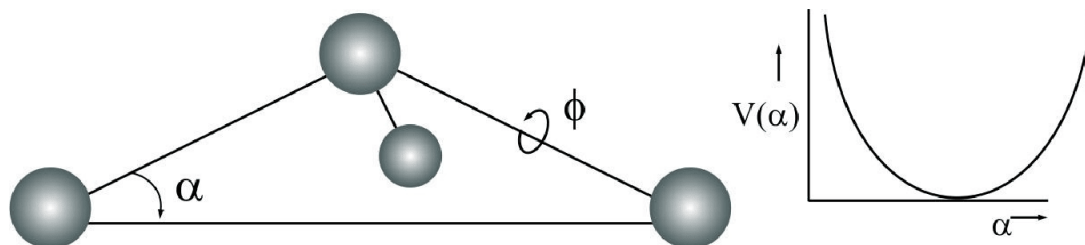


Figure 10: Schematic diagram of bonded stretching and corresponding nature of graph.

Non-bonded Potential Functions

Non-bonded interactions act between atoms which are not linked with covalent bonds. The potential function for such non-bonded interactions have two components: electrostatics and Lennard-Jones (L-J).

Electrostatic Potential

The atoms in a molecule possess partial charges despite the entire system is chargeless. The interaction between atoms with their partial charges has the significant effect in intra- and inter-molecular binding. The interaction between two point charges q_i and q_j is defined by a electrostatic potential (Phillips et al., 2005; Leach & Leach, 2001) as,

$$V_{\text{coulomb}} = \frac{q_i q_j}{4\pi\epsilon_r\epsilon_0 r_{ij}} \quad (3.9)$$

In Equation 3.9, the ϵ_r denotes the dielectric constant of the medium, ϵ_0 is the permittivity of free space, and r_{ij} is the distance between the charges. As atoms and molecules are not charged unless they are ions, the atomic charges in this expression are an artificial construct. However in molecules, the atoms share their valence electrons and the electron density may be shifted due to different electronegativity of the atoms. These different charge densities can be mapped into partial point charges which are represented by q_i and q_j as indicated in the Equation 3.9. The nature of variation of electrostatic potential between two charged atoms are shown in Figure 11a).

Lennard-Jones (L-J) Potential

Lenard-Jones terms well describe the short range repulsion and long range attraction. Short range repulsion scales as $\frac{1}{r^{12}}$ and the long range attraction scales as $\frac{1}{r^6}$. Although

it is not suitable for systems where strong localized bonds may form (as in covalent system) or there is a delocalized "electron sea" where the ions stay (as in metals), the L-J potential helps us to understand basic points in many areas of condensed matter physics (Braun et al., 2019). The L-J potential for interaction between a pairs of atoms (Leach & Leach, 2001) is given by,

$$V_{LJ}(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (3.10)$$

where, ϵ defines the strength of the interaction and σ defines length scale. The $\frac{1}{r^{12}}$ term dominates at short distances and the potential is strongly repulsive due to non-bonded overlap of electronic orbitals (according to Pauli Exclusion Principle). The $\frac{1}{r^6}$ term dominates at large distances and the potential is attractive due to van der Waals dispersion forces caused by the dipole-dipole interactions. The variation of L-J potential is shown in Figure 11(b).

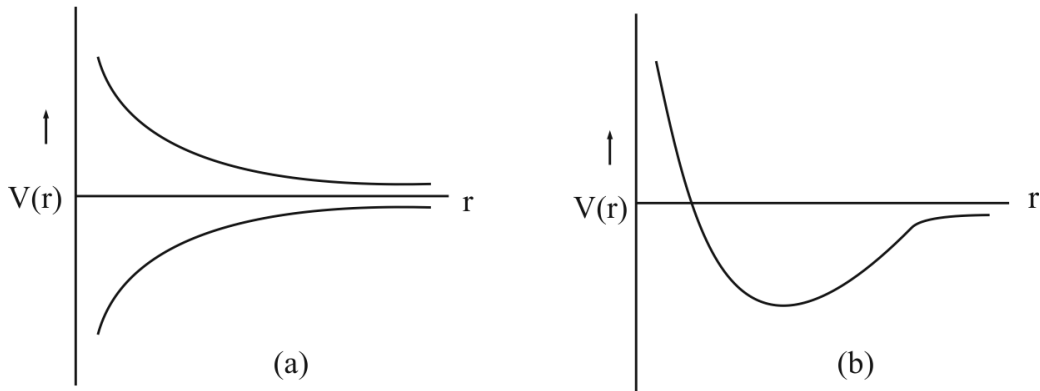


Figure 11: Graphical interpretation of non-bonded interaction. (a) represents the nature of electrostatic potential and (b) represents the nature of lennard Jones potential.

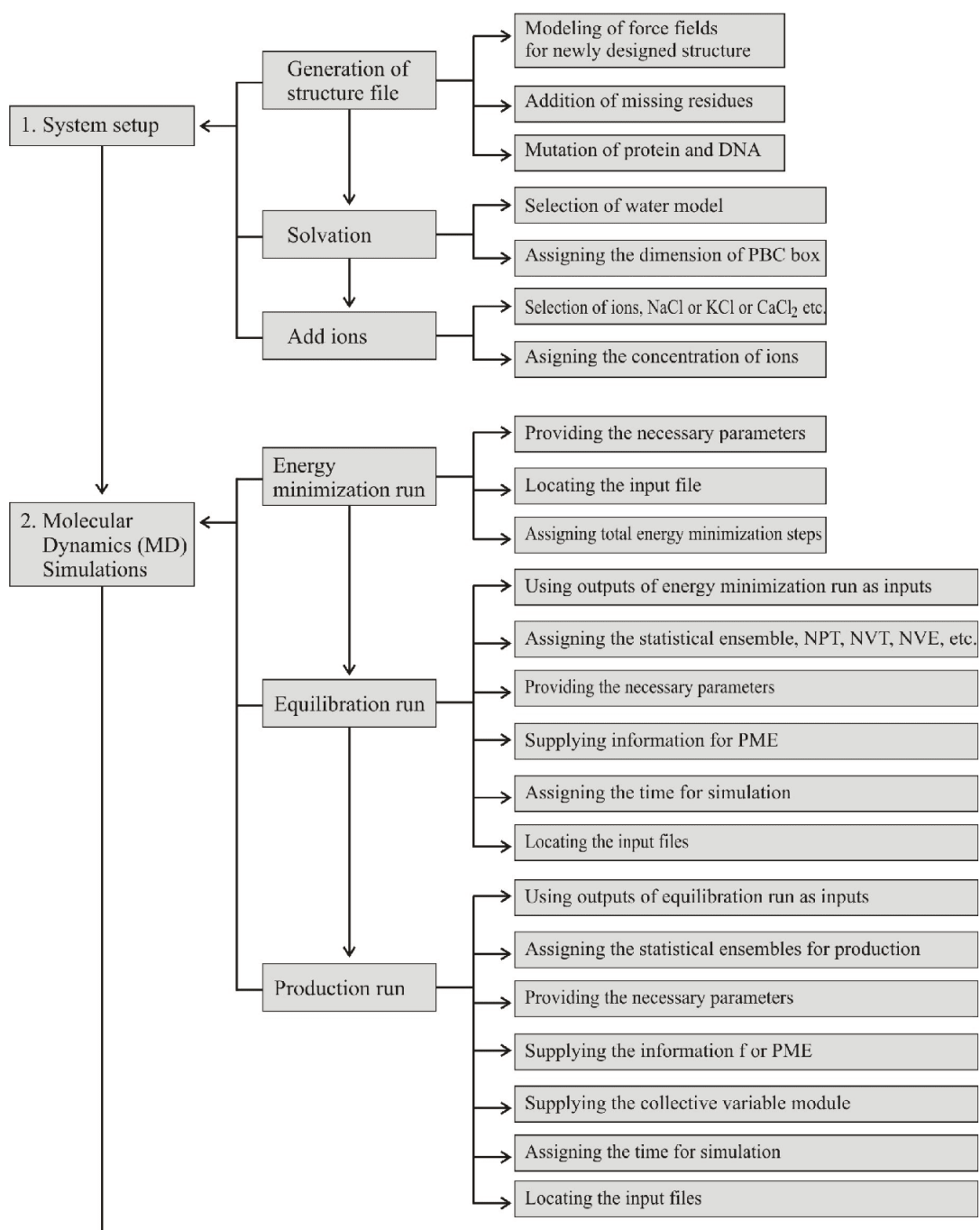
3.2 Simulation Details

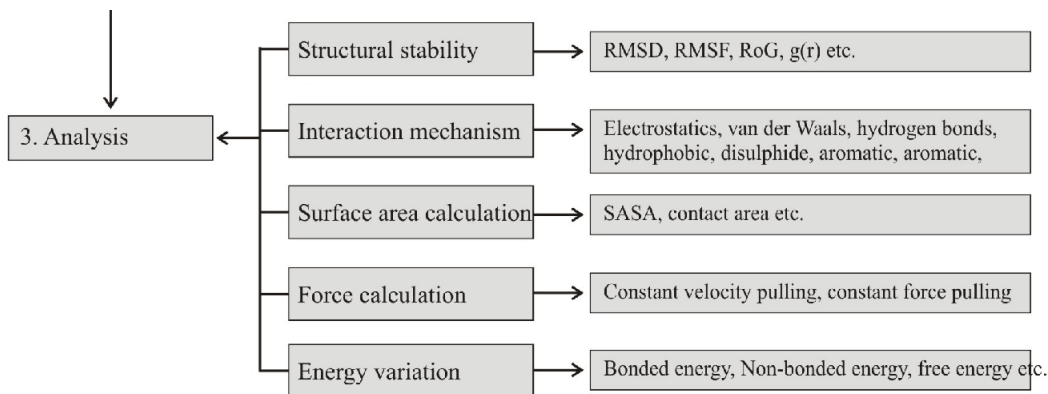
There are three basic steps in molecular dynamics study. These steps comprise the molecular system design, MD simulations, and analysis. These steps and their supporting points are presented in Flow Chart.

As shown in the Flow Chart, the molecular dynamics study comprises the different steps to estimate the meaningful outcomes. When molecule is selected from a source files, it should be set into PBC box generating the protein structure file (PSF), then the molecule is dissolved into water and electrically neutralized by adding Na^+ and Cl^- ions. This setup resemble the system as the cellular environment. The prepared systems are propagated into three MD simulations: energy minimization, system equilibration and production run. In molecular dynamics simulations the outputs of first step (i.e., energy

minimization run) is taken as the input for the equilibration run and consequently, the output of equilibration run is used as the input for the production run. The outputs of production run are analyzed to study the properties of the molecule in the aqueous environment.

Outputs of simulation run are analyzed several mathematical tools. The simulation outputs are achieved in the trajectory files with different file formats, like collective variables outputs, dcd files, log files etc. These files are utilized to obtain the necessary analysis parameters.





Flow Chart: NAMD system setup, simulations and the analysis.

3.2.1 Force Field Parameters for Long-range Interactions

Several adjustments need to be done to specify the long-range interactions, electrostatics and van der Waals, in the PBC box. These adjustments commands are written in configuration files so that the simulation run proceeds in accordance with the input information provided in these files.

Exclude: This term indicates the atomic interactions that are to be excluded during the simulation. The excluded value indicates the omission of atomic interaction of any three consequent atoms for van der Waals interaction. With the value of none, no bonded pairs of atoms will be excluded. With the value of 1-4, all 1-3 pairs will be excluded along with all pairs connected by a set of two bonds (i.e., if atom A is bonded to atom B, and atom B is bonded to atom C, and atom C is bonded to atom D, then the atom pair A-D would be excluded). The electrostatic interactions for such pairs are modified by the constant factor defined by 1-4-scaling. The van der Waals interactions are modified by using the special 1-4 parameters defined in the parameter files (Phillips et al., 2005).

Cutoff: The calculation of interaction between atoms of the entire volume via electrostatics and van der Waals contact is computationally too costly. To avoid such difficulty cutoff distance is set in the simulations so that the interaction beyond the cutoff are not taken into account. The schematic diagram showing the cutoff distance is shown in Figures 12(a) and 12(b). In NAMD software package, it is assigned in Å. Within the cutoff region, the short-ranged part is evaluated correctly and long range part is approximated by summation techniques like Particle Mesh Ewald (PME) summation (Phillips et al., 2005). This parameter is used when switching is set to ‘on’ to specify the allowable distance between atoms for inclusion in the pair list.

Switch distance: If the cutoff is assigned alone, the electrostatic and van der Waals interactions do not terminate smoothly so that the calculation diverges. In order to make the smooth termination, a switching function is employed by assigning the switch

distance. This function allows smoothing such interaction to zero at the cutoff distance (Phillips et al., 2005). The value of switchdist must be less than or equal to the value of cutoff, since the switching function is only applied on the range from switchdist to cutoff as shown in Figure 12(a).

Pairlist distance: This is very important to identify the atoms whose interactions are to be calculated for long-range forces. During the MD simulations, atoms are wiggling and jiggling within a certain volume. Several atoms enter and leave the cutoff boundary within the simulation time step as shown in Figure 12(b). To address the solution to such a difficulty during the force calculation, a boundary surface is set that selects the atom pairs within an arena and beyond this region are excluded. The maximum distance between any two atoms within which long-range interactions are taken is called the pairlist distance. Pairlist distance must be assigned greater than the cutoff distance (Phillips et al., 2005).

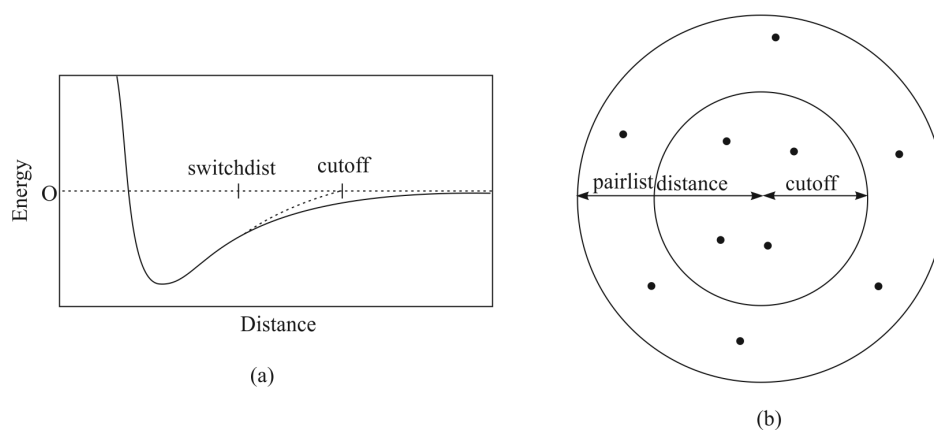


Figure 12: Schematic diagram of (a) cutoff and switching points (b) pair-list distance and cutoff distance.

3.3 Main Steps of Molecular Dynamics Simulations

Molecular dynamics simulation is the best alternative to predict the molecular behaviour in the aqueous environment. Several physical and chemical properties of a molecule or a molecular complex can be evaluated by completing some essential steps in MD simulations among which preparation of molecular system is the foremost one. Then, three steps of simulation run are propagated: energy minimization run, equilibration run, and production run (Braun et al., 2019). The outcomes of production run are applied to estimate the several thermodynamic and mechanical properties of the corresponding system.

The details of the simulation procedures are shown in chart diagram below. The leftmost column shows the main steps of the MD simulations. The middle column shows the necessary procedures for corresponding steps of simulations and the rightmost column

shows the detail processes that should be done to carry out the simulations. This chart is basically constructed targeting the simulations for the present work. The analysis techniques can be modified in accordance with the system of molecules and aim of investigation of chemical and physical parameters.

3.3.1 System Preparation

The initial structure of a molecule is either taken from protein data bank or designed by allocating the appropriate coordinates for atoms in a phase space. Basically, the coordinates assigned in accordance with the image taken from x-ray diffraction (XRD) or nuclear magnetic resonance (NMR) technique are used in order to design the pdb structure. Then, protein structure file (psf) is generated. In every psf file, an associated pdb is also generated. In original PDB structure, hydrogen atoms are not included. During the generation of psf file, hydrogen atoms are added in the appropriate positions so that the molecular structure should be chemically accurate (Van Gunsteren & Mark, 1992; Lee et al., 2016). Also, in many pdb structures, some essential amino acid residues may be missing. These incomplete structures never give the correct result. Therefore, these missing residues must be modelled with appropriate software, usually called modeler. In the present study, visual molecular dynamics (VMD) and online web server CHARMM-GUI have been employed as a modeler. In the formation of psf and pdb structures, a topology file is accompanied with the original structure. The software program constructs the structure in accordance with the bonds, angles, dihedrals, mass, and charge provided by the topology files. The newly generated PSF and PDB files are then solvated in water to create the cellular environment. Further, the solvated system is to be electrically neutralized by adding necessary ions. NaCl, KCl, CaCl₂ and MgCl₂ are often used to neutralize the solvated system. In this work, NaCl has been used for the entire simulation work considering the maximum concentration of these ions in the body cell.

System preparation is the most critical stage of the simulations. If the atoms are wrongly allocated or missed, the result would be flawed. Therefore, one should be careful in whether the molecule contains suitable bond, protonation/deprotonation stage, and chemically valid structure (Braun et al., 2019). The information provided by the topology file is the key factor to construct the psf file. If the molecular system is wrongly prepared, further steps of molecular dynamics never estimate the accurate result.

3.3.2 Energy Minimization Run

This is the first step of molecular dynamics simulations. Lack of accuracy of allocation of positions of atoms taken from XRD or NMR might lose the stability of molecule (Gallo et al., 2009). Also, the hydrogen atoms are modelled in a molecule with the help of computer software which are sole guess coordinates. These limitations of the modeling a system leads the unstable state of potential energy. Therefore, the prerequisite of MD simulation is to bring the structure in minimum potential energy state. This can be done by energy minimization run. If the system is simulated without maintaining at the lowest energy state, the system may be blown up and the simulation terminates abruptly so that no further simulation would be possible. Therefore, the structure should be relaxed to the local minimum energy state.

3.3.3 Equilibration Run

Energy minimization (EM) run removes the errors associated during the design of protein data bank files and in modeling of hydrogen atoms. This run also sets the system in minimum and stable energy state. Without EM simulation run, the simulation outcomes may not be converged. In the next step, the system is prepared for resembling to living body environment. Temperature, volume, and pressure of the system should be set for the realistic condition (for example room temperature or human body temperature with suitable pressure). This environment can be created from the equilibration run in molecular dynamics simulations (Gallo et al., 2009; Braun et al., 2019).

The main aim of the equilibration is to set the system into a targeted state point, i.e., appropriate thermodynamic conditions. During simulations, thermodynamic variables like temperature and pressure may change due to the motion and interaction of atoms, so they should be controlled to maintain the system at certain thermodynamic condition. This can be done by adding or removing the heat in the system as it approaches the correct partitioning of kinetic and potential energies. Once the kinetic and potential energies of the system are reached at almost the constant values and other key properties like temperature and pressure are no longer changing with time, the system is said to be well-equilibrated. Equilibration of system ensures the convergence of simulation in accordance with assigned thermodynamic state.

The coordinates and velocities achieved at the end of equilibration run are used to initiate the production run. Even though many ensembles pairs (NPT-NVT, NVT-NPT, NPT-NVE, etc.) are suggested for equilibration to production run, the choice of ensembles depend on what properties we want to estimate from the simulation. In general, two different ensembles are chosen for equilibration to production run. For example, if NVT

ensemble is selected for the production run, corresponding equilibration run should be NPT. Sometimes an additional simulation run may be suitable as an extension of the equilibration run.

3.3.4 Production Run

This is the final step of MD simulations. When the system is well equilibrated, it is then propagated for the production run. The data obtained from the production run are employed to evaluate the physical parameters like heat capacity, heat diffusion, elastic and viscous property, transport phenomena, potential mean force, free energy, coupling and decoupling forces, etc. The data obtained in the outputs rely on the input command provided by the configuration file. In addition to the ordinary information in configuration file, some special physical parameters are to be calculated by assigning the collective variables accompanied with the other outputs in the simulations (Braun et al., 2019). For example, reaction coordinates should be assigned in order to calculate free energy. Similarly, the speed and direction should be assigned in SMD pulling.

The initial steps of production run are inherent from equilibration run. The initial data obtained from production run are sometimes discarded by many researchers to eliminate such inherent effect of equilibration run. Also, simulation can be considered as the randomized process, so simulation of a molecular system is repeated many times to validate the result and to make the result more reliable. In general, production run is performed for a relatively longer time than that of equilibration. Longer simulation stabilize the system, reduces the simulation error, and estimate the desired parameter closer to the reality (Braun et al., 2019).

3.4 System Setup and MD Simulations in DNA-AGT Systems

The present work has been carried out to analyze the interaction mechanism of damaged DNA and DNA repair protein AGT. Originally, DNA contains methylation damage at the O6 position of GUA7. A DNA repair protein, AGT, also named O6-methylguanine-DNA methyltransferase (MGMT), was used to remove methylation damage from DNA. The interaction between AGT with methylated DNA (mDNA) was assumed as the methyl transfer reaction in which the methyl adduct could be transferred to the specific cysteine residue of AGT. In this research work, we divided our investigations into three steps: system with methylated DNA (complex-I) represents the pre-methyl transfer condition, system with methylated Cys145 of AGT (complex-III) represents the post-methyl transfer condition, and covalent linkage between methylated DNA and AGT (complex-II), which represents the transient intermediate state of pre- and post-methyl transfer state.

The structural modification of methylated guanine, methylated cysteine with their pair partners, and the covalent links between GUA7 and CYS145 are shown in Figure 13.

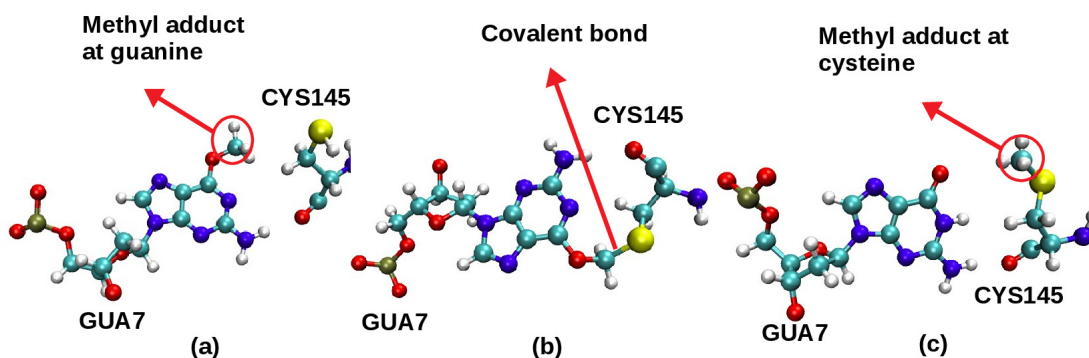


Figure 13: The modified structures of GUA7-CYS145 pairs (a) methylated GUA7 with its interacting partner CYS145, (b) covalent linkage of methylated GUA7 and CYS145 and (c) methylated CYS145 with its interacting partner GUA7.

200 ns NPT Simulations

Methylated GUA System (complex-I): The X-ray crystal structures of two molecular systems were obtained from the protein data bank (PDB) 1T38.pdb. The PDB structure of this structure contains the methylated guanine at DNA and AGT complex with non-covalent bonding between them. In this structure, seventh residue of dsDNA has originally been methylated at O6 of guanine. The existing double bond between O6 and C6 was converted into single bond and C6 to N1 became double bonded. Finally, N1 was deprotonated. This type of modifications in guanine base causes the cell mutagenesis and carcinogenic (Kyrtopoulos et al., 1997; Bruner et al., 2000). The missing residues were modelled by using CHARMM-GUI online web server. The protein was mutated at SER145 by another amino acid CYS145 (i.e., S145C), which is necessary for methylated base flipping and methyl-transfer mechanism. Cysteine is only that amino acid which accepts the methyl adduct in its sidechain by deprotonating at sulphur. Newly designed force fields 6og.rtf and 6og.prm were used to generate the system and the simulations.

Methylated CYS System (complex-III): It was also modelled from PDBID 1T38.pdb. For this structure, the methylated guanine (6OG7) was taken as the normal guanine (GUA7) and cysteine at 145 residue of AGT (CYS145) was modelled with methylated at SG point. Remaining part of the modeling was similar as explained in the complex-I. Newly designed force fields ort.rtf and ort.prm were implemented for the system generation and the simulations.

Covalent system (complex-II): The PDB structure in system II contains the DNA-protein complex with covalent bonding between them via GUA7 and CYS145. CHARMM-GUI was used to fulfill the missing residues. In order to generate the AGT-DNA covalent

system, we first converted the residue E1X7 in the PDB ID 1T39 to methylated GUA7 and then applied a patch (PRES: C6OG) to connect the methylated GUA7 and CYS145, resulting in a covalent bond between the protein and DNA. During the process, sulfur at CYS145 and CH₃ at guanine were deprotonated. In such process, we have considered the sp³ hybridization scheme.

We, thus, prepared three molecular systems to investigate the non-covalent and covalent mechanisms between targeted DNA and AGT. Then, all the systems were solvated with transferable intermolecular potential with 3 points (TIP3P) water in periodic boundary condition (PBC) simulation box (PBC box size (85 × 85 × 85) Å³ for system-I and system-III, and (90 × 90 × 90) Å³ for system-II) and were electrically neutralized in 0.15 M salt concentration of NaCl.

All-atom molecular dynamics (MD) simulations were performed by using computer software package NAMD. The CHARMM36m (Huang et al., 2017) force fields were employed for all simulation systems. Some force fields were designed for modified structures; methylated GUA7, methylated CYS145 and covalently linkage of DNA and AGT. The newly designed force fields have been incorporated in the later part in this section. The Particle Mesh Ewald (PME) was used to treat the long-range interactions with a 12.0 Å non-bonded cutoff. The energy minimization was performed for 10,000 steps by using the conjugate gradient algorithm. Then, the system was equilibrated at 300 K (27 °C) with harmonically restrained alpha carbon atoms of AGT and backbone atoms of DNA with 1 fs time step. Finally, the production run was propagated by using Langevin dynamics with a damping constant of 1 ps⁻¹ under NPT conditions with time step of 2 fs for 200 ns.

To study the nucleotide flipping mechanism, we have prepared two systems picking only the DNA from PDB ID 1T38.pdb: a DNA containing non-methylated guanine and another the methylated guanine. Then, MD simulations were run for 100 ns each. The system preparation is similar to the above method.

3.4.1 Steered Molecular Dynamics (SMD)

To investigate the decoupling force between DNA and AGT, we have carried out the SMD simulations. Two different systems have been prepared similar to complex-I and complex-III as explained above in 30 ns NPT simulation. Molecular systems for SMD were also prepared from CHARMM-GUI. The missing residues were modelled and S145C mutation was done by the same server. The system was solvated in orthorhombic box of dimensions (128 × 64 × 72) Å³ size with TIP3P water and electrically neutralized by adding NaCl. The molecular systems were named system-I and system-II taken from complex-I and complex-II respectively.

To perform the SMD simulations, the AGT protein was pulled along positive x direction under constant velocity pulling method with 1 Å/ns taking the DNA as the reference molecule under NPT condition. The total trajectory length was 30 ns for each, which was sufficient to estimate the maximum pulling force to decouple AGT and mDNA for the supplied uniform velocity. The spring constant was specified to be 1 kcal/(mol·Å²). This procedure takes the force as the function of displacement. Same procedure has been performed for both the systems. Each of the system is repeated five times and force-displacement curves were plotted for each of corresponding system. Then, the average value obtained for each of the system after the SMD run was also plotted and compared the result to identify the binding strength of DNA and AGT in two situations.

3.4.2 Umbrella Sampling

In the present work, we focused on the estimation of free energy of binding of DNA and AGT by using umbrella sampling method. Umbrella sampling was performed in two different systems for complex-I and complex-III. We prepared 19 different windows for each complex. To prepare the windows AGT protein was taken as reference molecule and DNA was translated along negative x-direction. Windows for both systems were prepared manually displacing the DNA by using VMD. We used 19 windows of 1 Å each along the negative x-axis to ensure the sufficient overlapping of windows to cover the entire reaction coordinate space. Originally, in PDB 1T38.pdb, DNA is about 24.8 Å center of mass (COM) distance far from the protein. The system was solvated in orthorhombic box of dimensions (128 × 64 × 72) Å³ size with TIP3P water and electrically neutralized by adding NaCl.

Energy minimization has been performed each of the window equally for 10,000 steps. Each of these systems is equilibrated at 300 K for 1 ns. Then, the production run was performed for 10 ns for each window of both systems. The reaction coordinate was chosen as the distance between the center of mass (COM) of AGT and DNA along the negative x-axis. We used the force constant of 1.5 kcal/(mol·Å²). Then, the free energy was estimated by the Weighted Histogram Analysis Method (WHAM) (Kumar et al., 1992) program.

3.4.3 Data Analysis

VMD was used to analyze the simulation trajectories and visualize the structures. The NAMD energy plugin, available in VMD, was used to calculate the non-bonded interaction energies; electrostatics (E), van der Waals (VDW) contributions and hydrogen bonding between AGT residues and DNA nucleotides. The free energy was estimated by

the Weighted Histogram Analysis Method (WHAM) program. Several tool command language (tcl) scripts and fortran (FORMula TRANslateion) codes have been used to estimate the physical parameters during the interactions of molecules. GaphPad prism (San Diego, CA) and xmgrace and Microsoft excel softwares were used to plot the graphs. Some tools of microsoft excel were also used in the calculation of physical parameters.

3.4.4 Newly Designed Force Fields

In this research work, the molecular complexes were not the complex of ordinary DNA and protein, rather they were modified.

In original protein data bank structure, seventh nucleotide, GUA7, is methylated at O6 point forming single bond between O6 and C6. This condition sets the double bonds between C6 and N1 by deprotonating at N1. These modifications inherent in the partial charges as well as the internal coordinates. Therefore, a new topology file had to be designed such that the protein structure file would be correct. We have designed the topology file for entire structure of methylated guanine base in seventh residue of the DNA. For the modified structure (i.e., methylated guanine base), we have designed the topology and parameter files.

The topology and parameter information of all three systems are as follows:

- Complex-I: Topology: This file contains the topology information for complex-I in which the guanine is methylated at O6 point.
- Complex-I: Parameters: This file contains the corresponding parameters for the complex-I.
- Complex-II: Topology: This file contains the topology information of complex-II. PRES gives the topology to patch the guanine and cysteine.
- Complex-II: Parameters: This file contains the parameters for the covalent linkage of guanine and cysteine.
- Complex-III: Topology: This file contains the topology for the methylated cysteine at sulfur atom at its side chain.
- Complex-III: Parameters: This file contains the parameters for the methylated cysteine in which cysteine is methylated at sulphur.

ATOM C2 CN2 0.83
 ATOM N1 NN3G -0.84
 ATOM C6 CN1 0.80
 ATOM O6 ON5 -0.48
 ATOM C5 CN5G 0.00
 ATOM N7 NN4 -0.60
 ATOM C8 CN4 0.25
 ATOM H8 HN3 0.16
 ATOM C15 CN9 0.12
 ATOM H16 HN9 0.09
 ATOM H17 HN9 0.09
 ATOM H18 HN9 0.09
 GROUP
 ATOM C2' CN7B 0.14
 ATOM H2'' HN7 0.09
 ATOM O2' ON5 -0.66
 ATOM H2' HN5 0.43
 GROUP
 ATOM C3' CN7 0.01
 ATOM H3' HN7 0.09
 ATOM O3' ON2 -0.57
 BOND P O1P P O2P P O5'
 BOND O5' C5' C5' C4' C4' O4' C4' C3'
 O4' C1'
 BOND C1' N9 C1' C2' N9 C4 N9 C8
 C4 N3
 BOND C2 N2 C2 N1 N2 H21
 BOND N2 H22 C6 C5 C6 O6
 BOND O6 C15 C15 H16 C15 H17 C15 H18
 BOND C5 N7 C2' C3' C3' O3' O3' +P
 BOND C2' O2' O2' H2'
 BOND C1' H1' C2' H2'' C3' H3' C4' H4'
 C5' H5'
 BOND C5' H5'' C8 H8
 DOUBLE C2 N3 C4 C5 N7 C8 N1 C6
 IMPR C2 N3 N1 N2 C6 N1 C5 O6 N2 H21
 C2 H22
 DONO H21 N2
 DONO H22 N2
 DONO H2' O2'
 ACCE N3
 ACCE N7
 ACCE O1P P
 ACCE O2P P

ACCE O2'
 ACCE O3'
 ACCE O4'
 ACCE O5'

IC C5 C6 O6 C15 1.3600 120.00 180.00 108.00
 1.2340
 IC C6 O6 C15 H16 1.2340 108.00 180.00 108.89
 1.1110
 IC H16 O6 *C15 H17 1.1110 108.89 120.00 108.89
 1.1110
 IC H16 O6 *C15 H18 1.1110 108.89 -120.00 108.89
 1.1110

END


```

ATOM C5   CN5G   0.00   !
ATOM N7   NN4   -0.60   !
ATOM C8   CN4    0.25   !
ATOM H8   HN3    0.16   !
ATOM C15  CN9    0.12   !
ATOM H16  HN9    0.09   !
ATOM H17  HN9    0.09   !
ATOM H18  HN9    0.09   !
GROUP
ATOM C2'  CN7B   0.14
ATOM H2'  HN7    0.09
ATOM O2'  ON5   -0.66
ATOM H2'  HN5    0.43
GROUP
ATOM C3'  CN7    0.01
ATOM H3'  HN7    0.09
ATOM O3'  ON2   -0.57
BOND P    O1P    P    O2P    P    O5'
BOND O5'  C5'    C5'  C4'    C4'  O4'    C4'  C3'
O4'  C1'
BOND C1'  N9     C1'  C2'    N9  C4     N9  C8
C4  N3
BOND C2  N2     C2  N1     N2  H21
BOND N2  H22    C6  C5     C6  O6
BOND O6  C15    C15 H16    C15 H17    C15 H18
BOND C5  N7     C2' C3'    C3' O3'    O3' +P
BOND C2' O2'    O2' H2'
BOND C1' H1'    C2' H2' '  C3' H3'    C4' H4'
C5' H5'
BOND C5' H5' '  C8  H8
DOUBLE C2  N3    C4  C5     N7  C8     N1  C6
IMPR C2  N3    N1  N2     C6  N1  C5  O6     N2  H21
C2  H22
DONO H21  N2
DONO H22  N2
DONO H2'  O2'
ACCE N3
ACCE N7
ACCE O1P  P
ACCE O2P  P
ACCE O2'
ACCE O3'
ACCE O4'
ACCE O5'

```

IC C5	C6	O6	C15	1.3600	120.00	180.00	108.00
1.2340							
IC C6	O6	C15	H16	1.2340	108.00	180.00	108.89
1.1110							
IC H16	O6	*C15	H17	1.1110	108.89	120.00	108.89
1.1110							
IC H16	O6	*C15	H18	1.1110	108.89	-120.00	108.89
1.1110							

PRES C6OG 0.21

DELETE ATOM 1HG1

DELETE ATOM 2H18

GROUP

ATOM 1CB CT2 -0.10

ATOM 1SG S -0.08

GROUP

ATOM 2C15 CN9 0.21

ATOM 2H16 HN9 0.09

ATOM 2H17 HN9 0.09

BOND 1SG 2C15

END

IC -C	N	CA	C	1.3478	124.2100	180.0000	106.3100
1.5195							
IC N	CA	C	+N	1.4510	106.3100	180.0000	117.7400
1.3471							
IC +N	CA	*C	O	1.3471	117.7400	180.0000	120.6400
1.2288							
IC CA	C	+N	+CA	1.5195	117.7400	180.0000	124.5200
1.4471							
IC N	C	*CA	CB	1.4510	106.3100	121.6200	111.8800
1.5546							
IC N	C	*CA	HA	1.4510	106.3100	-116.9800	107.5700
1.0832							
IC N	CA	CB	SG	1.4510	111.2500	180.0000	115.9200
1.5460							
IC SG	CA	*CB	HB1	1.5460	115.9200	120.5600	106.9000
1.1153							
IC SG	CA	*CB	HB2	1.5460	115.9200	-124.8000	109.3800
1.1129							
IC CA	CB	SG	CD	1.5546	115.9200	180.0000	110.2800
1.8219							
IC CB	SG	CD	HD1	1.8219	98.9400	-179.4200	110.9100
1.1111							
IC HD1	SG	*CD	HD2	1.1111	110.9100	119.9500	111.0300
1.1115							
IC HD1	SG	*CD	HD3	1.1111	110.9100	-119.9500	111.0900
1.1112							

END

3.5 Formulation on MD Analysis

There are several formulae associated with the MD analysis. In this part, we have included those mathematical relations which are taken into our consideration for analysis.

3.5.1 Root Mean Square Deviation (RMSD)

Root mean square deviation is a thermodynamic variable that estimates the structural stability of entire molecular system in aqueous environment. RMSD of backbone of the system is usually calculated to evaluate the stability. For the structural stable state, RMSD of the system should be almost constant (Martínez, 2015). If the RMSD curve shows the increasing pattern till the end of the simulation run, the system is considered not properly equilibrated. If the system is not brought properly equilibrated, the calculated parameters taken after the production run will be erroneous.

RMSD is plotted over the time function during the simulation run (Kuzmanic & Zagrovic, 2010; Carugo & Pongor, 2001). It is the numerical measure of structural variation of molecule during the MD simulations and estimate whether the molecule is thermodynamic stability or not. To find the RMSD, a reference structure is taken as standard and the variation of structure under each frame of simulation is calculated. The difference of structure for two different frames is calculated as,

$$\text{RMSD}(t) = \sqrt{\frac{\sum_{\alpha=1}^N (\vec{r}_{\alpha}(t_j) - \langle \vec{r}_{\alpha} \rangle)^2}{N_{\alpha}}} \quad (3.11)$$

where, N_{α} is the number of atoms whose positions are being compared, N_t is the number of time steps over which atomic positions are being compared, $\vec{r}_{\alpha}(t_j)$ is the position of atom α at time t_j , and $\langle \vec{r}_{\alpha} \rangle$ is the average value of the position of atom α to which the position $\vec{r}_{\alpha}(t_j)$ are being compared. It is defined as (Bornot et al., 2011),

$$\langle \vec{r}_{\alpha} \rangle = \frac{1}{N_t} \sum_{\alpha=1}^N \vec{r}_{\alpha}(t_j) \quad (3.12)$$

The RMSD value of molecule is being changed over time. Its value in each time step is calculated taking the reference coordinate of first time step of the simulation.

3.5.2 Root Mean Square Fluctuation (RMSF)

Root mean square fluctuation of a protein molecule has been calculated to find the relative stability of amino acid residues that fluctuate over the simulation run. The comparison of structural fluctuation helps to find the relatively stable and flexible residue in the protein chain. High RMSF value gives the more flexible residue and vice versa (Fuglebakk et al., 2012).

Each amino acid necessarily contains one alpha carbon at its backbone. So, the flexibility of alpha carbon is considered as the flexibility of corresponding amino acid in a protein chain. The mathematical formula to find the RMSF (Bornot et al., 2011) is,

$$\text{RMSF} = \sqrt{\langle r_i - \langle r_i \rangle \rangle^2} \quad (3.13)$$

where, r_i and $\langle r_i \rangle$ are the current position and average position of the representative atom to evaluate the RMSF.

3.5.3 Hydrogen Bonding

Hydrogen bonding is a special type of non-covalent interaction which is responsible in protein folding and existence of DNA and RNA structures. This bonding has the crucial role not only in giving the secondary structure of bio-molecules but also in the inter-molecular interaction in biological processes, like protein-ligand interaction, DNA-protein complex formation, protein-drug bindings etc (Bornot et al., 2011). It has the pivotal role in governing the shape, properties and functions of bio-molecules. Understanding the role of hydrogen bonding in inter-molecular binding is really exciting in drug designing (Thakuria et al., 2017; Pace et al., 2014). In this work, the role of hydrogen bonding between AGT and mDNA is extremely useful in understanding the stability and binding affinity of the complex.

3.5.4 Solvent Accessible Surface Area (SASA)

Solvent accessible surface area of a molecule is the measure of its available exposed surface into the solvent. If SASA increases during the molecular dynamics simulation, the amino acids lie on the surface of the molecule are hydrophilic in nature, whereas the decrease in SASA depicts the hydrophobic nature of surface amino acid residues. Besides this application, SASA is usually used to evaluate the contact area of two or more molecules in a complex (Yunta, 2017). The contact surface area is the surface buried at the interface between two molecules which contribute in binding to stabilize the molecular complex. Larger the contact surface area can have the greater binding

affinity. The contact area also influences the binding free energy and hence contributes in the formation of stable molecular complex. In this work, we have calculated the stability of protein and DNA employing this tool.

Let $S_{\text{DNA}}(t)$, $S_{\text{prot}}(t)$, $S_{\text{DNA+prot}}(t)$ the the time dependent SASA of DNA only, protein only and the DNA-protein complex respectively. Now, the contact surface area between DNA and protein has been evaluated from the relation (Yunta, 2017),

$$\sigma(t) = \frac{S_{\text{DNA}}(t) + S_{\text{prot}}(t) - S_{\text{DNA+prot}}(t)}{2} \quad (3.14)$$

The increase in surface area indicates the increase in stability of the complex and vice versa.

3.5.5 SMD and Pulling Force

Steered Molecular Dynamics (SMD) mimics the atomic force microscopy (AFM) experiments. This technique explores the elastic properties and unfolding nature in biomolecules (Zou et al., 2012; Phillips et al., 2005). This can be performed at constant velocity or constant force. In this technique, external force is applied to one or more atoms so that the binding strength or folding strength of a molecule can be estimated. The atom that is fixed at a specific coordinate is called SMD atom and the atom which is pulled with constant velocity or the constant force is called dummy atom. During the modeling, SMD atom and dummy atom are attached with a virtual spring having a certain force constant ' k '. The spring constant is restoring in nature and the external force tends to overcome the restoring force. To estimate the correct prediction of elastic properties, spring constant should be chosen correctly. It is usually chosen from verified experiments.

In this process, the backbone atoms of DNA were taken as the fixed atoms and alpha carbons in AGT were taken as the dummy atoms; and were pulled from their center of mass (COM) along the positive x-direction with constant velocity ($\vec{v} = d\vec{x}/dt$) in water and ions environment so that the SMD atom experiences the force $\vec{F}(t) = k(\vec{v} t - \Delta\vec{x})$ (Isralewitz et al., 2001), providing the external potential energy (Martínez et al., 2008),

$$U(x, t) = \frac{1}{2}k(\vec{v} t - \Delta\vec{x} \cdot \hat{n})^2 \quad (3.15)$$

where, k ($= 5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$) is the spring constant and gives the stiffness of the applied harmonic restraining force, and $\Delta\vec{x}(t) = \vec{x}(t) - \vec{x}_0$, is the displacement between the molecules being actual position $\vec{x}(t)$ and initial position \vec{x}_0 and \hat{n} is the unit vector along the direction of pulling.

3.5.6 Umbrella Sampling and Free Energy

Several energy barriers encounter during the chemical reaction, which should be overcome to proceed the reaction. Thermal energy and the enzyme activity make possible to cross such barriers in the cellular mechanism. In molecular dynamics simulations, it is really a challenging task to develop the technique of mimicking such situations. If there is a large energy barrier along the reaction pathway, it is almost impossible to jump the molecular system from one energy state to another energy state (Gullingsrud et al., 1999; Chandler, 1978). As a result, molecule at the existing state is trapped into a certain phase space. In such complicated situations, umbrella sampling can be a best sampling technique to find the free energy change along the reaction coordinates.

In umbrella sampling, entire reaction coordinates are divided into many windows and corresponding configuration states of the system are taken as the sample structure of the molecule. The reaction coordinates are so chosen that their molecular distribution overlaps in distributions obtained from every preceding and succeeding windows. In addition to this, a bias, an additional energy term, is applied in each window to ensure the sufficient overlapping of distributions in the calculation of free energy. The bias potential provides the harmonic potential, which connects the energetically separated regions in the phase space (Chandler, 1978; Kästner, 2011; McDonald & Singer, 1967).

An artificial biased potential is added to the unbiased potential to execute the umbrella sampling. The sum of these two potentials is called the biased potential and is written as,

$$V^b(r) = V^u(r) + \omega_i(\xi) \quad (3.16)$$

The superscript 'b' represents the biased quantity and the superscript 'u' represents the unbiased quantity. Also, $\omega_i(\xi)$ denotes the artificial biased potential for i^{th} window. The artificial harmonic potential $\omega_i(\xi)$ is calculated from,

$$\omega_i(\xi) = \frac{1}{2}k_i(\xi_i - \xi_i^{\text{ref}})^2 \quad (3.17)$$

where, ξ is the reaction coordinate along the path of reaction progress and ξ_i is the center of reaction coordinate (mean position of harmonic motion) for i^{th} window. In each window, the biased function attempts to keep the system close to the ξ_i^{ref} . k is the force constant of the virtual spring that oscillates the molecule within a certain phase space. Therefore, the Equation 3.16 is written as,

$$V^b(r) = V^u(r) + \frac{1}{2}k_i(\xi_i - \xi_i^{\text{ref}})^2 \quad (3.18)$$

Supplying these potentials and other usual parameters, MD simulation should be run for a specific window w_i , then the same process is repeated for window w_{i+1} and so on. Thus, umbrella sampling simulations are run for all necessary windows. To obtain unbiased free energy, the necessary unbiased probability distribution for the i^{th} window,

$$P_i^u(\xi) = \frac{\int \exp[-\beta E(r)] \delta[\xi'(r) - \xi] d^N r}{\int \exp[-\beta E(r)] d^N r} \quad (3.19)$$

But, we need the biased probability distribution to execute the umbrella sampling, since artificial potential is applied for the MD simulation. The biased probability distributions along the reaction coordinate is,

$$P_i^b(\xi) = \frac{\int \exp\{-\beta E(r) + \omega_i(\xi'(r))\} \delta[\xi'(r) - \xi] d^N r}{\int \exp\{-\beta [E(r) + \omega_i(\xi'(r))]\} d^N r} \quad (3.20)$$

We need the unbiased probability distribution in terms of biased probability distribution to evaluate the free energy of the system. After calculation, we get the relation between $P_i^u(\xi)$ and $P_i^b(\xi)$ is,

$$P_i^u(\xi) = P_i^b(\xi) \exp[\beta \omega_i(\xi)] \langle \exp[\beta \omega_i(\xi)] \rangle \quad (3.21)$$

Now, the unbiased free energy $A_i(\xi)$ can be calculated as,

$$A_i(\xi) = -\frac{1}{\beta} \ln(P_i^b(\xi)) - \omega_i(\xi) + F_i \quad (3.22)$$

And, the value of F_i (McDonald & Singer, 1969) is written as,

$$F_i = -\frac{1}{\beta} \ln \langle \exp[-\beta \omega_i(\xi)] \rangle \quad (3.23)$$

which is independent of ξ .

Then, the free energy obtained from each window has to be added to find the global free energy.

Selection of value of k : The choice of k is crucial in umbrella sampling. It has the key role in selecting the number of windows as well as centers of reaction coordinates. If the value of k_i is taken very small, $\omega(\xi)$ is negligible with respect to V^b . So,

$$V^b \approx V^u \quad (3.24)$$

Also, very large k_i will provide the very narrow distribution that may have difficulty of window overlapping. Increasing the value of k , may lead to increase in errors in the

numerical integration.

Selection of centers of reaction coordinate: There is also an important fact that how far the center of consecutive windows are taken. The Y-axis in Figure 14 is the pair distributions of center of mass distance of two molecules. If the distribution peak is high and width is low, the molecule oscillate in narrow region, but the large distribution width shows the larger fluctuation of the molecules. For the reliable result, the probability distribution should be sufficiently overlapped as shown in Figure 14. If the consecutive centers would chosen larger than that of overlapping condition, the free energy so obtained gives erroneous. Hence, the centers of reaction coordinates are appropriately chosen.

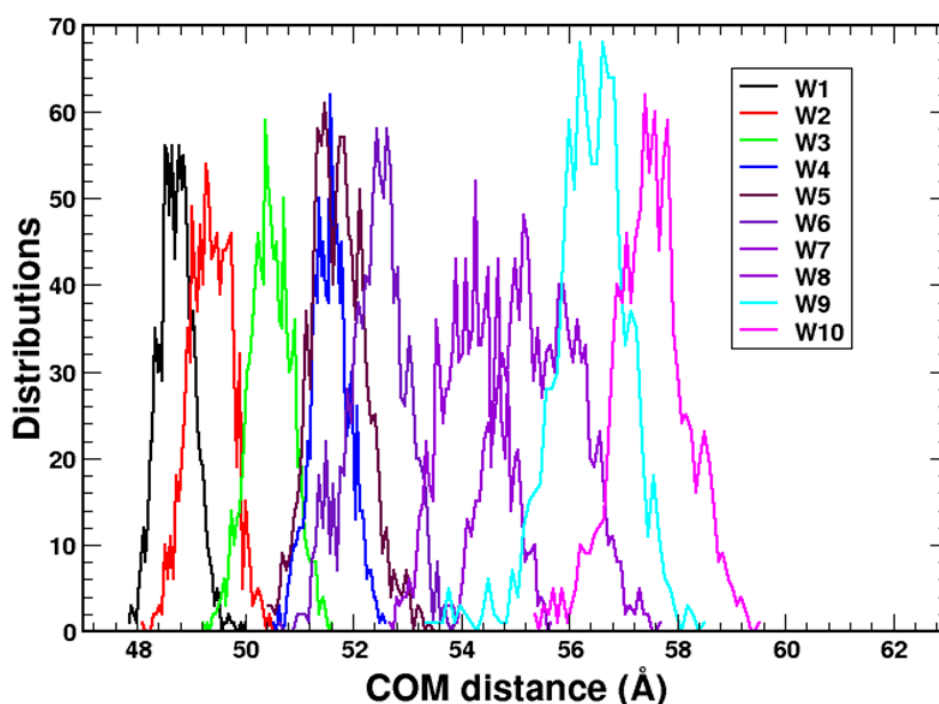


Figure 14: Illustration for distributions of overlapping windows during the umbrella sampling simulations. Overlapping windows ensures the representation of entire phase space.

3.5.7 Weighted Histogram Analysis Method (WHAM)

WHAM (Curuksu & Zacharias, 2009; Souaille & Roux, 2001) is a numerical method for the estimation of free energy. This method minimizes the statistical error of $P^u(\xi)$. It estimates the free energy of individual window and connects the global distribution by a weighted average of the distributions derived from the individual windows. So,

$$P^u \xi = \sum_{i=1}^N p_i(\xi) P_i^u(\xi) \quad (3.25)$$

The weights p_i are chosen in order to minimize the statistical error of P^u , i.e.,

$$\frac{\delta\sigma^2(P^u)}{\delta p_i} = 0 \quad (3.26)$$

and p_i should verify the condition as, $\sum p_i = 1$ and is calculated as,

$$p_i = \frac{a_j}{\sum_j a_j} \quad (3.27)$$

where,

$$a_i\xi = N_i \exp[-\beta\omega(\xi) + \beta F_i] \quad (3.28)$$

With N_i being the total number of steps sampled for window i . The F_i are calculated by

$$\exp(-\beta F_i) = \int P^u(\xi) \exp[-\beta\omega(\xi)] d\xi \quad (3.29)$$

CHAPTER 4

4. RESULTS AND DISCUSSION

4.1 General Consideration

In this chapter, we present, analyze and discuss the main outcomes of our work. The work is basically related to the methylation damage repair process in DNA by interacting with a DNA repair protein AGT. It incorporates the atomic level investigation on molecular complex formation between DNA and AGT, their reaction mechanism, and transfer of methyl adduct to AGT from DNA. To examine such reaction pathways, we have carried out the molecular dynamics (MD) simulations under different thermodynamic conditions.

We utilized simulation outcomes to estimate various mechanical and thermodynamic parameters like molecular stability, maximum rupture force, free energy etc. The atomic-level investigations in the contact region of the methylated DNA and AGT complex can provide an in-depth understanding of the binding between these molecules. The results allow to evaluate of silico-drug nature of AGT and is useful in removing the methylation damage from DNA. The main findings of the work are organized as follows:

- In Section 4.2, we discuss the molecular conformation of methylated DNA and AGT, structural variation in different stages of interactions in DNA-AGT complex, and the chemical reaction schemes. It provides the brief overview of molecular complex formation that will be useful to investigate the further sections of this chapter.
- The structural and energetic stability of the molecular complexes are presented and discussed in Section 4.3. This section provides the information about the comparative aspects of formation and dissociation of complexes.
- Section 4.4 presents the coulomb and van der Waals contributions during the complex formation. It provides the quantitative estimation of non-bonded interactions between DNA and AGT.

- In Section 4.5, the atomic level contributions of hydrogen bonding are explained. The results obtained from hydrogen bond analysis are compared for three complexes. Main interacting nucleotide-residue pairs are detected and interpreted in terms of occupancy percentages.
- The contact surface area between DNA and AGT are estimated and are compared to identify the stability of the of the complexes in Section 4.6. Solvent accessible surface area (SASA) is employed to estimate the contact surface area.
- In Section 4.7, atomic level investigations are carried out for the methylated guanine rotation from its backbone and discussed the stability of unpaired nucleotide.
- The structural dissociation of the complexes during the translation of AGT from DNA is discussed in Section 4.8. The steered molecular dynamics (SMD) has been utilized to estimate the decreasing pattern of hydrogen bonding and contact surface area. Further, maximum decoupling forces between AGT and DNA are estimated and discussed.
- In Section 4.9, free energy is estimated by translating DNA from AGT by using umbrella sampling technique. In addition, the variation of electrostatics and van der Waals energies are estimated at different windows designed for umbrella sampling.
- Section 4.10 describes the possibility of restoring mechanism of flipped out guanine into the base stack. This process is analyzed in terms of hydrogen bond formation and radial distribution of targeted guanine-cytosine pair partners.

4.2 DNA-AGT Complexes and Reaction Mechanisms

Living organisms are continuously exposed to many DNA damage agents (Chatterjee & Walker, 2017). As the DNA is a hereditary material in organisms, the preservation of information in genomics sequence in DNA is very important. Loss of genetic information may cause several health complications like cell mutation, cell cancer etc., which eventually leads the cell apoptosis and cell death (Kaina et al., 2007). The damaged DNA base may provide the wrong information to newly synthesized DNA strands or miss the ordinary coding. All the above difficulties that come up into the genetic materials, lead towards cell abnormalities and eventually affect in the entire body mechanism. Body cancer is one of the worst condition of genomics mutation (Robertson, 2005). Aiming as such AGT as the non-methylating agent in DNA, we have prepared the three molecular complexes that represent complete reaction path ways of methylated DNA repair process.

In this work, a molecular complex containing methylated DNA and AGT was taken into account. It repairs the methylation damage at the O6 point of guanine and O4 point of thymine bases of DNA (Zak et al., 1994). This protein contains a total of 188 amino acid residues, however, we have taken 170 residues and the remaining 18 residues towards the N-terminal domain were not considered because they contribute neither in binding nor in methyl transfer mechanism (Daniels et al., 2004). Also, the methylated DNA in the complex contains 26 nucleotides. The amino acid sequence in AGT and nucleotide sequence of double-stranded DNA (dsDNA) are shown in sequence chart below. The CYS145 in AGT is the targeted amino acid which is considered as the methyl acceptor and GUA7 in DNA is the methyl guanine that is needed to be repaired. The position of methyl guanine in DNA and active cysteine residue in AGT are indicated with the arrow as shown in sequence chart below. The sequences of amino acids and nucleotides in AGT and DNA are represented by one letter symbol. The digit just above the one letter symbol indicates the residue identity (RES ID). RES ID is given for first and last residue in each line, others are identified counting from left to right in regular counting order. Understanding the amino acid sequences in AGT and nucleotide sequences in DNA is useful in the molecular level study of the coming sections. These sequences are also important to identify the major interacting residue-nucleotide pairs in both non-covalent and covalent conditions.

Amino acid sequence in AGT

6	E M K R T T L D S P L G K L E L S G C E	25
26	Q G L H E I K L L G K G T S A A D A V E	45
46	V P A P A A V L G G P E P L M Q C T A W	65
66	L N A Y F H Q P E A I E E F P V P A L H	85
86	H P V F Q Q E S F T R Q V L W K L L K V	105
106	V K F G E V I S Y Q Q L A A L A G N P K	125
126	A A R A V G G A M R G N P V P I L I P C	145
146	H R C V C S S G A V G N Y S G G L A V K	165
166	E W L L A H E G H R L	176

Nucleotide sequence in DNA

1	5'- G C C A T G G C T A G T A -3'	13
26	3'- C G G T A C C G A T C A T -5'	14

Residues-nucleotides sequences: Amino acid sequence in AGT and the nucleotide sequence in DNA. The targeted residues CYS145 and 6OG7 are indicated with arrow. AGT contains 170 amino acid residues and methylated DNA contains 26 nucleotides.

The complex which we have taken into consideration contains chemically modified structures. This chemical modification in DNA causes the structural modification which ultimately obeys the induced fit rule during the binding of DNA and AGT. We basically focused on the study of methyl guanine base activity and the mechanism of transfer of methyl moiety to AGT by interacting with the direct damage reversal mechanism. We handled the problem in three different steps:

- Entry of AGT to methylated DNA: This is the first step of reaction process in

which methylated DNA and its repair protein AGT form a non-covalent complex. It represents the pre-methyl transfer condition (complex-I).

- Covalent linkage between methylated DNA and AGT: This step represents the transient intermediate state of the reaction mechanism (complex-II). The covalent bonding was modelled for the investigation of special purpose to represent the methyl transfer event from DNA to AGT. It links the pre- and post-methyl transfer conditions.
- Dissociation of AGT from DNA: This is the final step of the reaction. The methyl adduct is supposed to be attached at SG point of cysteine at 145th residue of AGT, which is considered as the repaired DNA after transferring the methylation damage to cysteine. It represents the post-methyl transfer condition (complex-III).

To represent the above three conditions of reaction steps, we generated three molecular complexes taken from two protein data bank structures (PDB ID 1T38 and 1T39) (Daniels et al., 2004). PDB ID 1T38 was modified to prepare the structures for complex-I and complex-III as explained in Section 3.4 of Chapter ‘Materials and Methods’. Complex-I has methylated GUA7 (6OG7) in DNA and represents the molecular structure for pre-methyl repair conditions. Similarly, complex-III has methylated cysteine (ORT145) in AGT and represents molecular structure for post-methyl repair conditions. Complex-II, originally taken from PDB ID 1T39, has the covalent complex between methylated GUA7 at DNA and CYS145 at AGT that represents a possible model of a transient, intermediate state. During the methyl transfer process, the methyl adduct at GUA7 is supposed to be transferred to CYS145 by deprotonating at sulfur. The molecular structures of all three complexes are shown in Figure 15. In Figure 15, the targeted residue-nucleotide pairs during the methyl transfer process are enclosed with a rectangular box and are magnified on the right side of the corresponding complexes.

Since methylation damage is a modified structure in DNA, the existing ordinary force fields cannot be used in the computational simulations of such reformed structures. Therefore, we designed force fields for all three complexes as explained in the Section 3.3 of Chapter ‘Materials and Methods’ and the molecular dynamics simulation was executed. This work was primarily aimed to model the AGT-DNA complexes for computational investigations with molecular dynamics simulations. The topology and parameter information that we have presented in Chapter ‘Materials and Methods’, especially for methylated CYS and methylated GUA, are valuable for future investigations of systems with these modifications, not just for the AGT-DNA complex.

To generate the correct structure in accordance to our objective, we have mutated the SER145 to CYS145 to investigate as suggested by Daniels et al. (2004). Cysteine is a actively reactive amino acid in the intermolecular interactions. Besides the reactive

behaviours, it also forms the disulphide bonds that is essential in the formation of tertiary structure of protein. Study of its transport properties is also very important (Koirala et al., 2020).

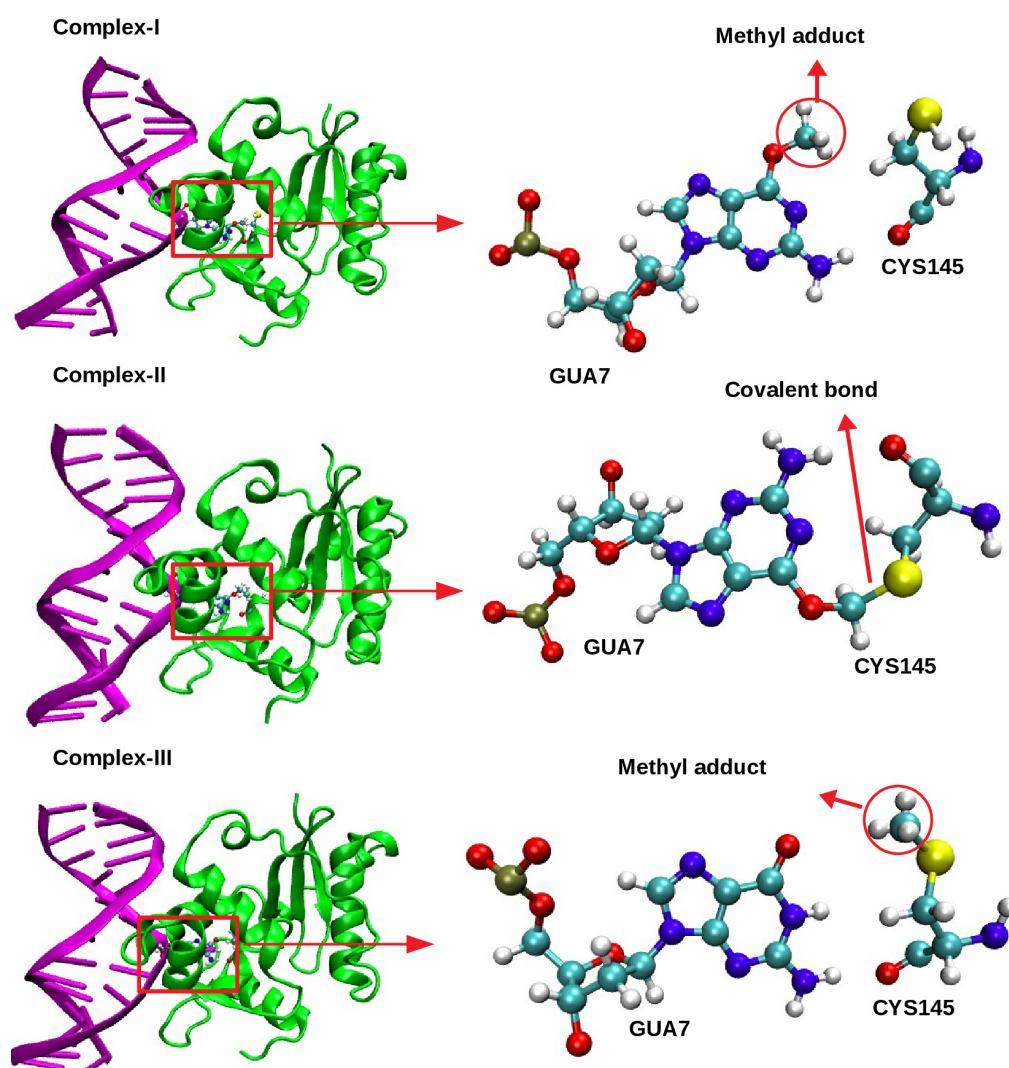


Figure 15: Three molecular complexes incorporating targeted parts for methyl transfer mechanism.

Out of above three complexes, complex-II is the transient intermediate state. Our transient intermediate model used covalently linked sp^3 hybridized carbon between S and O (i.e., $-S-CH_2-O-$). However, the transfer of methyl group to sulfur from methyl guanine was considered to occur via an S_N2 -like mechanism (Mattosovich et al., 2020; Y. Fu et al., 2021) and is very short likely occur in the order of 100 ns time scales (Patra et al., 2016), which involves a planar, penta-coordinated carbon ($-S-CH_3-O-$) (Fernández et al., 2007) as the transient intermediate. The orientations of the groups and the nature of the cavity might be sufficiently represented by our model to stabilize the complex and the further work with appropriate force-field modifications for pentacoordinate carbon is needed for a more detailed analysis of the geometry and energetic of the transition state.

4.3 Stability of Three Complexes

In the investigation of intra- and inter-molecular interactions, inspection of structural and thermodynamic stability are essential. There are several ways of estimation of structural and thermodynamic stability of a molecule in molecular dynamics simulation. The structural stability are usually examined by estimating root mean square deviation (RMSD), root mean square fluctuation (RMSF) etc. and the thermodynamic equilibrium condition comprises the stabilization of pressure, temperature, volume in accordance with the ensembles under which the simulation is to be performed.

We have used the original molecular structures designed from protein data bank (PDB). These initial structures of molecules taken from protein data bank may not be structurally and thermodynamically stable. There are several limitations in the accurate prediction of the coordinates assigned from x-ray diffraction (XRD) and nuclear magnetic resonance imaging (NMR). Many errors inherent during the assigning of coordinates in the PDB structure hold on the molecule in energetically unstable state. Also, the guessed coordinates of hydrogen atoms during the formation of protein structure file also produces errors. To resolve such technical complications, we performed energy minimization run. This run minimizes the potential energy, thereby removing any steric hindrance in the atoms of the molecules. In this work, pressure and temperature were equilibrated at 1 bar pressure and 300 K temperature. The energy minimization run and the equilibration run thus created the suitable environment of the molecular system and could mimic the real body cell environment. The outputs of equilibration run are assumed appropriate to execute the production run. Finally, the trajectory data obtained from production run was used to study the specific properties of equilibrated system.

Considering the structural stability as the foremost requisite of molecular dynamics simulation, we have estimated some basic structural and thermodynamic parameters like root mean square deviation (RMSD), root mean square fluctuation (RMSF), total energy, consistency of hydrogen bond, and confinement of residue within a suitable distance. In first step of the investigation, we probed from 200 ns MD simulation for each of the molecular complexes to ensure the stability of the complexes and to investigate the structural dynamics.

4.3.1 Root Mean Square Deviation (RMSD)

Root mean square deviation is the measure of structural stability of the molecules. Consistent as well as lower value of RMSD confirm the stable structure in the complex. A consistent RMSD value ensures the structural stability and that of lower value depicts the best orientation of the molecule in three dimensional structure (Neveu et al., 2018).

In all complexes, the RMSD was estimated for all atoms of DNA and AGT backbones without taking hydrogen atoms. The atomic coordinates of the first step of the simulation were taken as the reference coordinates. Then, the RMSD of each complex is plotted as the time-dependent function of the MD simulation. The RMSD for all three complexes is shown in Figure 16.

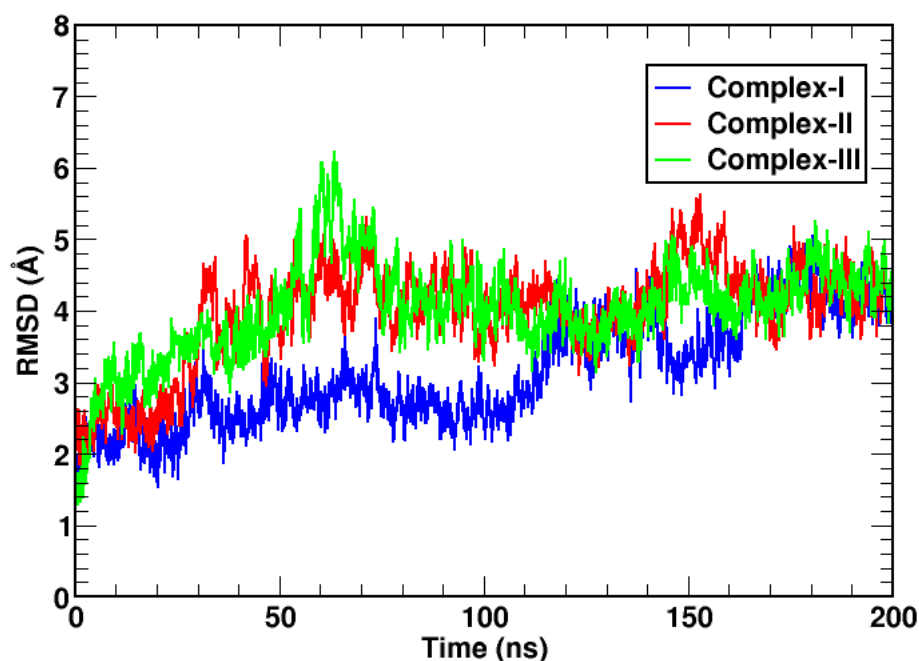


Figure 16: RMSD of all three complexes. RMSD of complex-II and complex-III were relatively greater fluctuation upto the 100 ns simulation than complex-I, after this time scale all three complexes have the similar fluctuations.

The RMSD value of complex-I was found fluctuating within the range of 1.4 Å to 5.3 Å with average value of 3.5 Å. Likewise, the RMSD values of complex-II and complex-III were detected ranging 1.7 Å to 6.3 Å and 1.5 Å to 6.8 Å respectively with corresponding average values of 4.3 Å and 4.4 Å. The RMSD values of complex-II and complex-III were observed to be slightly larger in comparison to complex-I. The value of complex-I was well below 5.5 Å, however, in case of complex-II and complex-III, they were sometimes found to be above 6 Å. After 120 ns time, all three complexes have almost the equal RMSD as shown in Figure 16. The consistent values of RMSD ensured the structural stability of all three complexes. To interpret the easy comparison, the minimum, maximum and the average values of RMSD of each complex summarized in Table 1. This shows that there is no significant fluctuation of RMSD in all three complexes, which could produce a stable trajectory for further analysis (Koirala et al., 2021).

Table 1: Root mean square deviation (RMSD) of three complexes are presented. First column shows the type of complex, other columns show the values of RMSD at the indicated conditions.

Complexes	RMSD minimum (Å)	RMSD maximum (Å)	RMSD average (Å)
I	1.4	5.3	3.5
II	1.7	6.3	4.3
III	1.5	6.8	4.4

4.3.2 Root Mean Square Fluctuation (RMSF)

RMSF of amino acid residues in a molecule distinguishes the mobile part and the stable part of the targeted molecule (Koirala et al., 2020b; Dadou et al., 2017). The stable part contains the region of bounded residues and fluctuating part contains the residues of relatively flexible region. In contrast to RMSD calculation taking first step coordinates as the reference coordinates, RMSF is calculated from an average structure of the entire simulation for each residue. We estimated the RMSF of alpha carbon of AGT in all three complexes, complex-I, complex-II and complex-III as shown in Figure 17. RMSF of alpha carbon best represents the fluctuations of the corresponding residues (Grasso et al., 2018).

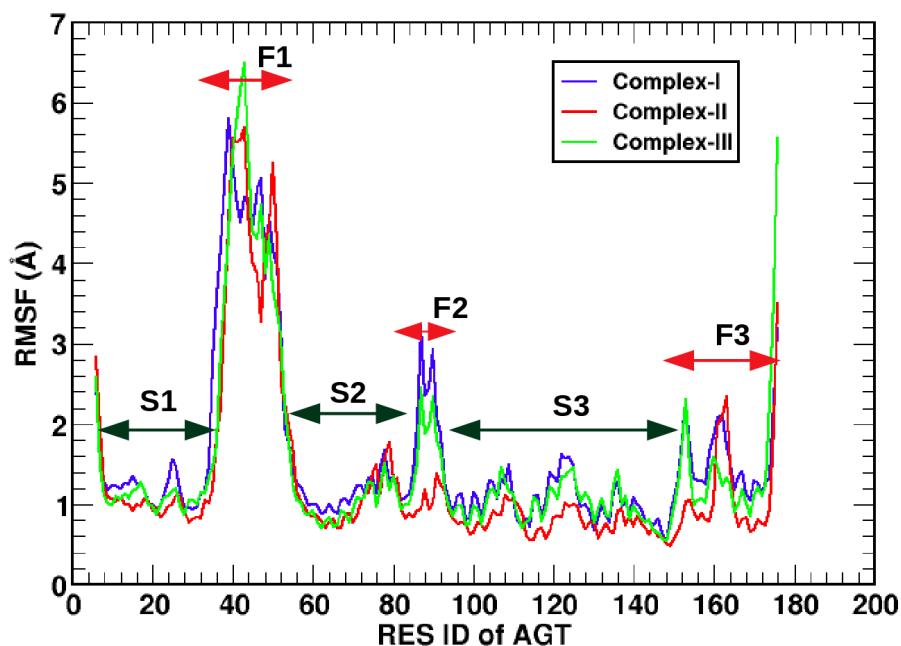


Figure 17: Root mean square fluctuation (RMSF) of AGT residues for complex-I complex-II and complex-III. Remote side residues were found more flexible than the interfacial residues indicating the good binding between DNA-AGT molecules in their corresponding complexes.

The fluctuating and stable residues are shown in Figure 17. During the 200 ns NPT simulations, we have represented stable regions by S_i and the fluctuating regions by F_i , where, $i = 1, 2, 3$. $RMSF > 2.0 \text{ \AA}$ are taken as the fluctuating region and that of $RMSF$

$< 2.0 \text{ \AA}$ are taken as the stable region. This is not the standard value to distinguish fluctuating and stable region, rather it is taken as the reference to explain the present situation. S1 (RES ID 7-34), S2 (RES ID 53-86) and S3 (RES ID 92-160) represent the stable regions in all three complexes, whereas the residues F1 (RES ID 35-52), F2 (RES ID 87-91) and F3 (RES ID 161-176) represent the relatively fluctuating regions in complex-I and complex-III. But, the residues within the F2 region in complex-II are stable as shown in Figure 18 in comparison to complex-I and complex-III. Since the complex-II has the covalent link between DNA and AGT, the region F2 nearer to that link is relatively more stable. In addition, the residues in complex-II are also less fluctuating in region F3. In the molecular complexes, these highly fluctuating residues lie in the remote side of the DNA. Moreover, the residues in the N-terminal and C-terminal regions have the relatively greater fluctuation. In both regions, the corresponding residues are resided at the surface of molecule and are relatively free from intra-molecular as well as inter-molecular interactions. The residues fluctuation in complex-I and complex-III have almost the same pattern.

Furthermore, the amino acid residues in AGT, which lie at the vicinity of the interfacial region of DNA, were found less mobile and bounded strongly to DNA nucleotides than that of remote sides. The interfacial residues were stabilized and contributed to the formation of a stable complex. Some targeted amino acids like CYS145, ARG128, THY114, SER151, were found highly stable and strongly interacting with DNA nucleotides. These residues might play important role in the complex formation and methyl transfer mechanism.

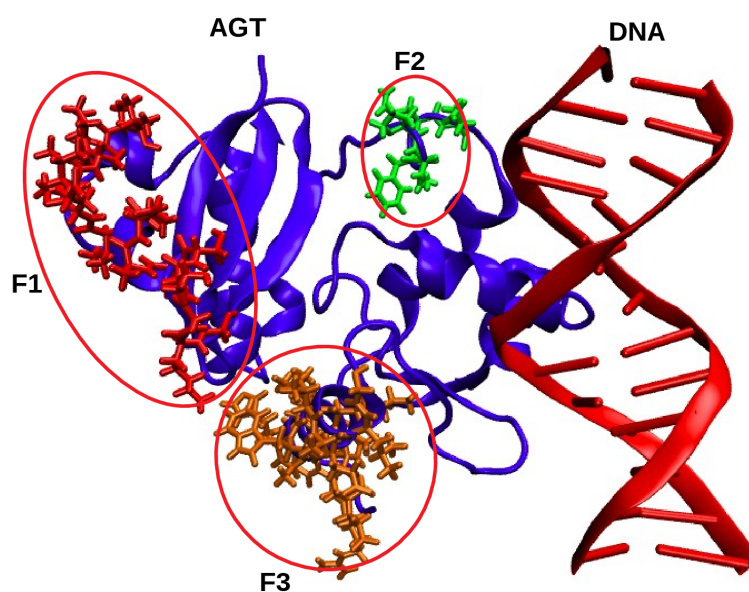


Figure 18: The regions showing the greater fluctuating surfaces. Targeted regions are encircled.

4.3.3 Interaction Energy

Considering the interaction energy between DNA and AGT as a key determining factor of stability of complexes, we have estimated the total interaction energy of all three complexes. 200 ns MD simulation trajectories were taken to estimate the total interaction energy of the protein and DNA molecules. The interaction energy of complex-I was obtained fluctuating from (-772.00 to -432.01) kcal/mol with average of (-599.81 ± 45.03) kcal/mol. Likewise, complex-II possesses the interaction energy ranging from (-760 to -430.12) cal/mol with average value of (-611.68 ± 59.16) kcal/mol and that of complex-III ranging from (-760.34 to -427.37) kcal/mol with average value of (-582.30 ± 46.40) kcal/mol. The total interaction energy of complex-II is the maximum negative value. It is probably due to the covalent linkage between the AGT and DNA that tends to keep the nucleotide-residue pairs in the close proximity. In comparison of other two complexes, complex-I has slightly greater negative value predicting the more stable structure. The plot of total energy versus time of simulation is shown in Figure 19.

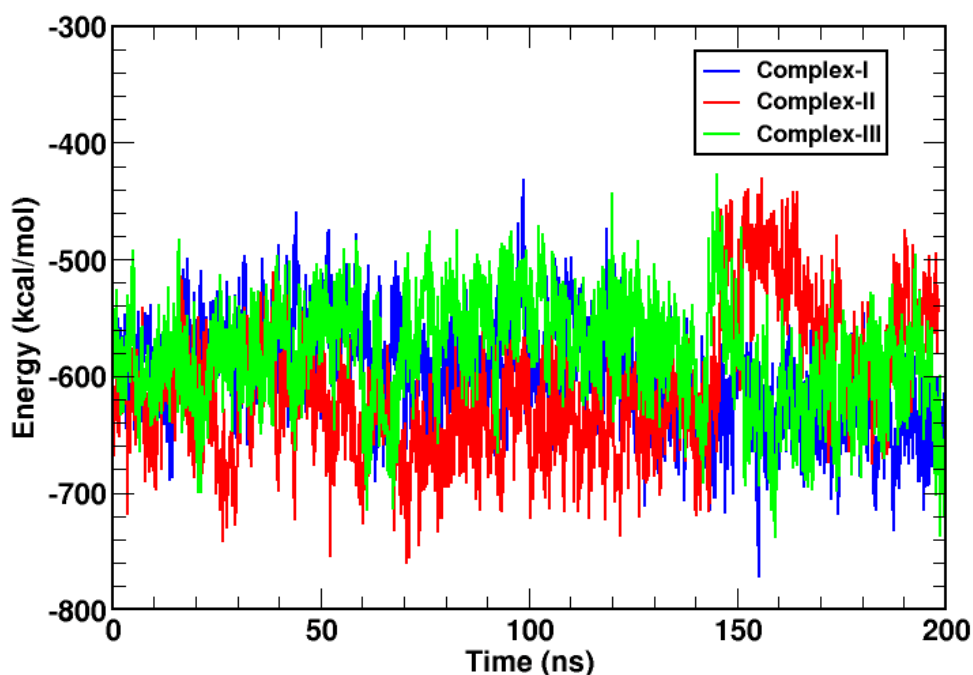


Figure 19: The plot of the interaction energy of three complexes as a function of time.

The interaction energy of all three complexes were obtained fluctuating nearer to their average values and remains almost the constant. The interaction energy in complex-II was found increasing after the time 150 ns, this may be reason of transient state of the interaction. Till the end of the simulations, complex-I has the maximum negative value, showing the most favourable structure.

4.3.4 Hydrogen Bonding and Structural Stability

The hydrogen bonding play important role in binding DNA and protein. It is pervasive in nature during the intermolecular interactions. This bonding contributes vital role in all chemical, physical and biological systems. The importance of hydrogen bonding in drug designing is well established. It dictates in the conformation of molecular structure and binding affinity (Tiwari et al., 2016). Drug binding on the targeted molecule is mediated by the number of potential hydrogen bonds. The breakage of hydrogen bonds illustrates the unfavorable binding of two molecules in a complex, whereas the increasing pattern of hydrogen bonds shows that the molecules in the complex still need to rearrange to form the stable structure. The constant number of hydrogen bonds during the simulation portrays the stable structure and the complex formation is favourable (Pace et al., 2014).

Assuming the hydrogen bonding as the pivotal role in intermolecular binding, we employed 200 ns MD simulation of all three complexes to analyze and compare the pattern of variation in hydrogen bonds during the dynamics as shown in Figure 20. Hydrogen bonds (H-bonds) were measured quantitatively in terms of number of hydrogen bonds.

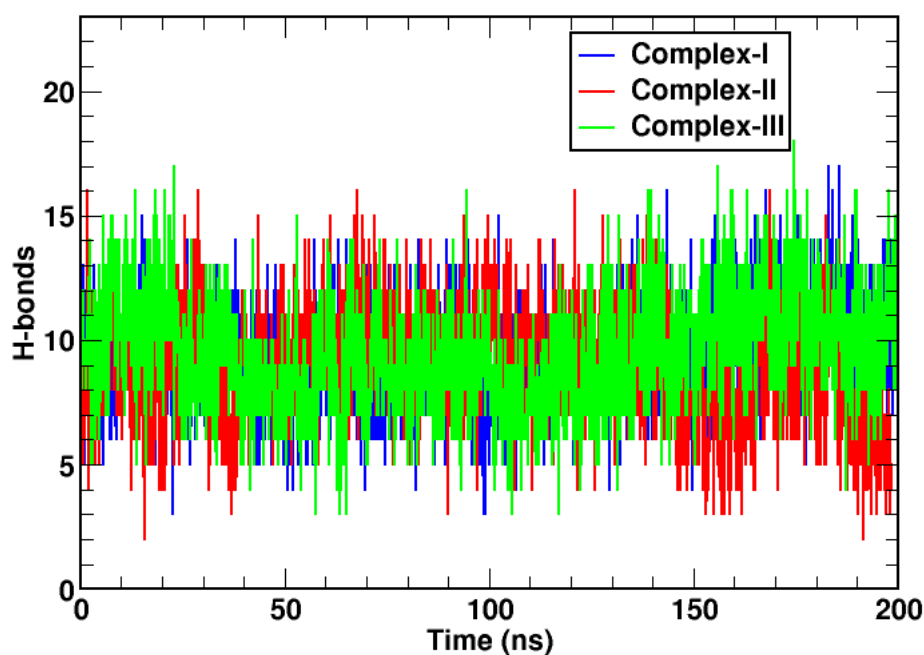


Figure 20: The overall hydrogen bonding patterns in complex-I, complex-II and complex-III.

The minimum, maximum and average values of number of hydrogen bonds are shown in Table 2. The average values of hydrogen bonds are shown in inequality sign (< or >), as these average values were obtained in decimal value. The expression in inequality sign makes it easier to compare H-bonds for different complexes. The average value of H-bonds in complex-II was obtained in decreasing pattern towards the end of simulations. In overall analysis, we have detected the strong and highly consistent hydrogen bonds

in SER151-THY9, CYS145-6OG7(GUA7) and TYR114-GUA7(GUA7) pairs. These pairs may be important in the formation of the DNA-AGT complex. The major atoms that contribute in the hydrogen bonds will be analyzed in the later part of this section.

Table 2: Minimum, maximum and average values of number of hydrogen bonds formed during the 200 ns simulations. First column shows the type of complex, other columns show the values of H-bonds at the indicated conditions.

Complexes	H-bonds minimum	H-bonds maximum	H-bonds average
I	3	17	> 9
II	1	17	< 9
III	1	16	> 9

In addition to the DNA-AGT intermolecular hydrogen bonding, we also detected three water mediated hydrogen bonds. Among three water-mediated hydrogen bonds, two were observed between ARG128 and CYT8 and one was observed between LYS125 and CYT21. They were detected only in the complex-I. These three bonds work as the additional contributions to stabilize this complex. Such type of hydrogen bonds were not detected in complex-II and complex-III within our 200 ns time scales.

4.3.5 Estimation of Distance of Strong Binding Pairs

As the potential energy is the function of distance between two atoms, it is useful to analyse the variation of distance of key binding atoms. We plotted the distances between the atoms that are responsible in establishing the hydrogen bonds in the three major residue-nucleotide pairs: 6OG7(GUA7)-TYR114, 6OG7(GUA7)-CYS145 and THY9-SER151 as shown in Figure 21.

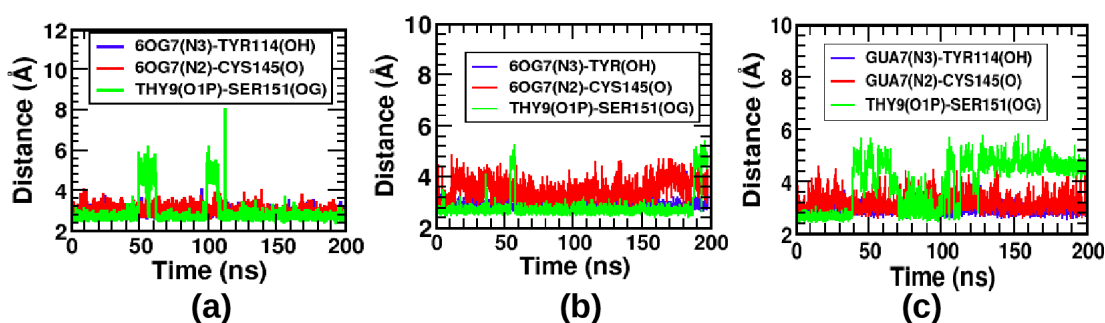


Figure 21: Distance between the strongly bounded residue to the nucleotide pairs for (a) complex-I (b) complex-II and (c) complex-III.

During the 200 ns simulations, we detected stable hydrogen bond between 6OG7(N3) (i.e., GUA7(N3)) and TYR114(OH) atoms interacting with close proximity in all complexes as shown in Figure 21(a). The distance of these atoms are almost the equal

throughout the simulations. Similarly, another nucleotide pairs in 6OG7(N2) and CYS145(O) has also the strong contributions in the formation of DNA-AGT complex. Hydrogen bonds between these atoms is most consistent in complex-I as shown in Figure 21. In complex-I and complex-II, the distance between THY9(O1P) and SER151(OG) were also almost constant, despite some fluctuations in some trajectory frames. However, there were some more fluctuations observed in THY9-SER151 in complex-III as shown in Figure 21(c). After 100 ns simulation run complex-III, the distance between THY9(O1P) and SER151(OG) was found increasing significantly. To examine the cause of such significant increment of distance, we searched the alternative residue that was activated to stabilize SER151(OG). We found that SER151(OG) rotated towards the ADE10(O2P) atom and formed the stable bond. We further estimated the distance between SER151(OG) and ADE10(O2P) during the entire simulation as shown in Figure 22. Distance between these two atoms were found decreasing after 100 ns time. The overall graph shows that when SER151(OG) comes closer to ADE10(O2P), THY9(O1P) moves farther from SER151(OG) and vice versa.

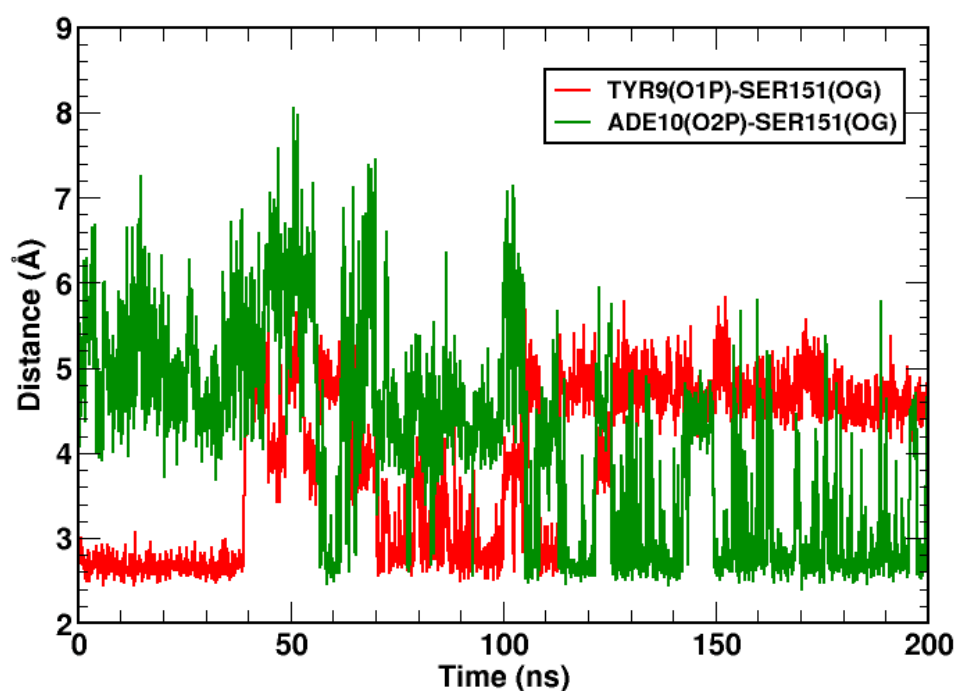


Figure 22: Distance of SER151(OG) from THY9(O1P) and ADE10(O2P). This shows the shifting of hydrogen bonds of SER151(OG) from THY9(O1P) to ADE10(O2P).

4.4 Electrostatics and van der Waals Interactions

Electrostatics and van der Waals interactions play crucial roles in intramolecular as well as intermolecular binding in biomolecules. In this work, we estimated the intermolecular binding of DNA and AGT. Each phosphate group of DNA nucleotides possesses the net negative charge (equal to the charge of an electron). Also, some of the amino acids

have non-zero charges, positive or negative. Besides these, atoms in the molecules have non-zero partial charges. In many situations, they form the dipole moment. Such partial or net charges in the molecules play significant role in binding DNA and protein molecules via electrostatic interaction (Grimme et al., 2006; Jones et al., 2003). On the other hand, van der Waals interaction is ubiquitous in all types of atoms and assists in strengthening the binding of the molecules in a complex (Matthew & Ohlendorf, 1985; DiStasio et al., 2014). The interaction between two point charges q_i and q_j is defined by electrostatic potential as,

$$U_{\text{coulomb}} = \frac{q_i q_j}{4\pi\epsilon_r\epsilon_0 r_{ij}} \quad (4.1)$$

In equation, the ϵ_r denotes the dielectric constant of medium, ϵ_0 is the permittivity of free space and r_{ij} is the distance between the charges. The potential energy provided by the van der Waals interaction between a pairs of atom is given by,

$$U_{\text{vdW}} = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (4.2)$$

where, ϵ defines the strength of the interaction and σ defines length scale.

In this work, we carried out quantitative estimation of the binding mechanism of DNA and AGT in complex-I, complex-II and complex-III due to electrostatic and van der Waals contacts. The averaged values of such interaction energies were estimated from outputs of 200 ns simulations for all complexes. During 200 ns MD run under NPT condition, the contributions of electrostatic interaction were greater than that of the van der Waals interactions as expected. In complex-I, the observed potential energies due to electrostatics, van der Waals, and the total (electrostatics + van der Waals) were (-492.70 ± 44.20) kcal/mol, (-107.10 ± 5.52) kcal/mol, and (-599.81 ± 45.03) kcal/mol, respectively. Similarly, in complex-II, the respected values were (-500.76 ± 54.61) kcal/mol, (-110.92 ± 8.28) kcal/mol, and (-611.68 ± 59.16) kcal/mol, respectively. Also, in complex-III, the potential energies for the respective interactions were (-475.57 ± 45.19) kcal/mol, (-106.73 ± 6.26) kcal/mol, and (-582.30 ± 46.40) kcal/mol. The interaction energy for all three complexes are compared in bar diagrams as shown in Figure 23.

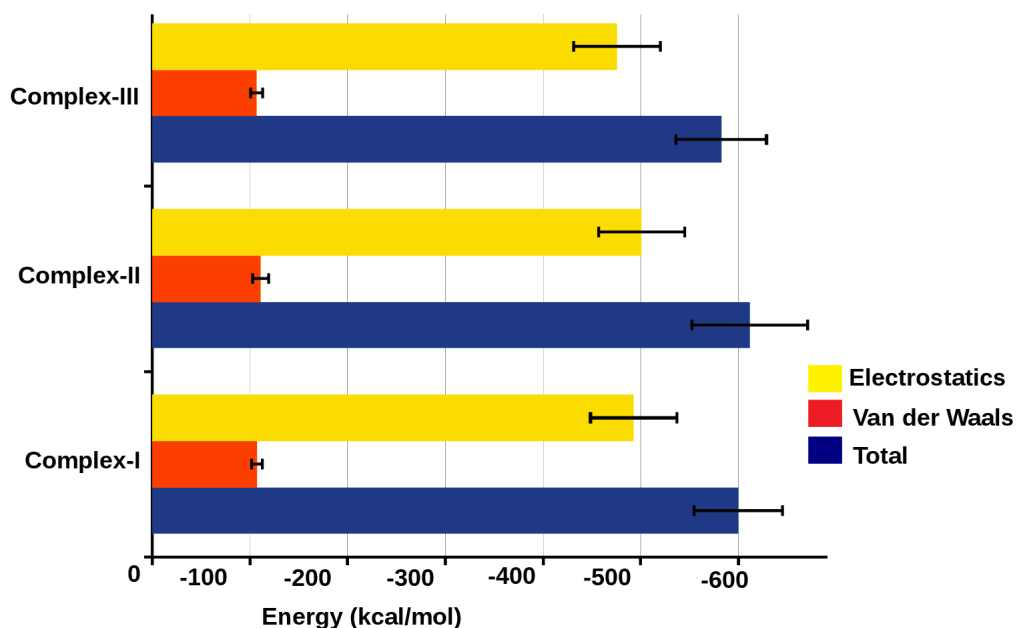


Figure 23: Bar diagram representation of the electrostatic, van der Waals, and the sum of these energies for complex- I, complex-II and complex-III indicating the dominant effect of electrostatic energy in overall non-bonded interactions. Electrostatics potential energy is indicated by yellow, vdW potential energy is indicated by red and sum of these potential energies are indicated by blue colour.

The magnitude of total energies (i.e., the sum of electrostatic and van der Waals) is minimum in complex-II. In this complex, the covalent linkage may keep DNA and AGT molecule in closer proximity. In comparison of other two complexes, complex-I has the lower energy than that of complex-III, suggesting that the possible formation of pre-methyl transfer (i.e., formation of the complex) is more favorable than that of post-methyl transfer (dissociation of the complex). Comparing the energy in these two complexes, the contribution of each energy type (i.e., electrostatic and van der Waals) is smaller in complex-I than in the complex-III. Errors were taken from the standard deviation (S.D.) of the corresponding data. The S.D. on energies indicate that the fluctuation of energy is relatively higher in complex-III than in complex-I. The simulation results show that energy due to electrostatic interaction has the dominating contributions in the binding of methyl-DNA and AGT, but no significant difference was observed while comparing the similar type of energy in these complexes.

The variations in electrostatics and vdW energies for all three complexes are plotted in Figure 24. Up to 30 ns simulation run, the energy variation in all complexes were found almost the same, however the values were significantly fluctuating in complex-II after 30 ns. The pattern of variation was similar for both electrostatics and van der Waals energies. Complex-III has the fairly stable energy, whereas the energies were most stable throughout the simulation in complex-I.

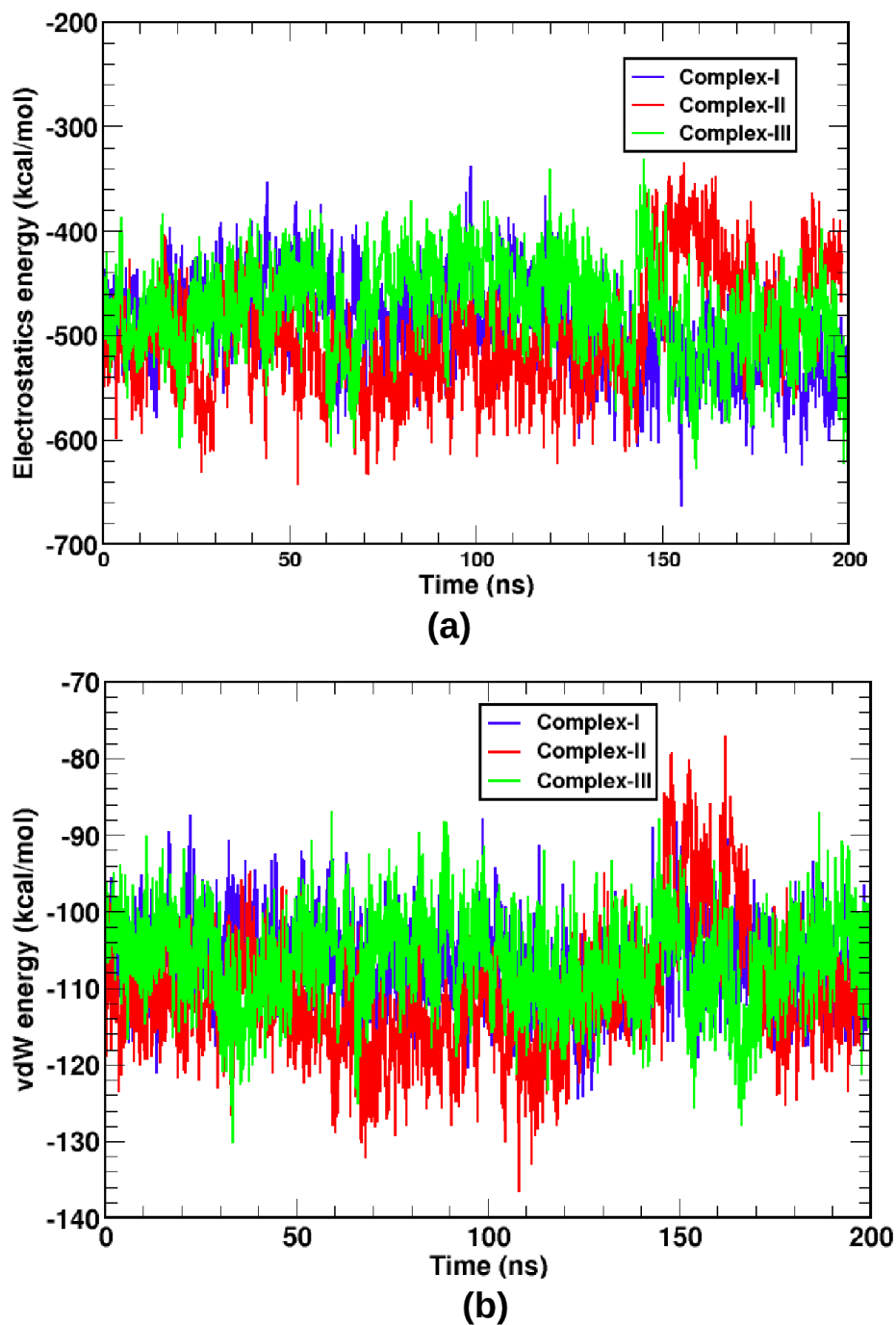


Figure 24: Comparison of non-bonded (a. electrostatics and b. vdW) energies variations in all complexes.

4.5 Hydrogen Bonding Analysis

Hydrogen bonding is a special type of non-covalent interaction that plays important role in protein folding and the existence of a helical structure of DNA and RNA. The double helix structure of DNA is possible only due to the hydrogen bonding among the nucleotide within the same strand or another strand. In addition, they have crucial role in the intermolecular interaction of biomolecules in the biochemical processes,

like protein-ligand interaction, DNA-protein complex formation, protein-drug bindings, etc. (Thakuria et al., 2017). Thus, they have important role in governing the shape, properties and functions of biomolecules. Intermolecular interaction via hydrogen bonding is important in drug designing. In this work, the role of hydrogen bonding between AGT and methylated DNA is extremely useful in understanding the stability and binding affinity in their complex (Tiwari et al., 2018). We have performed the detailed atomic-level analysis of hydrogen bonds in all three complexes. This analysis provides an overview of the interatomic hydrogen bonding mechanisms .

Hydrogen bonding was quantitatively analyzed by estimating the hydrogen bond occupancy. Greater the value of H-bond occupancy, higher the possibility of strong binding of that residue with its pair partner (Kannam et al., 2017). We have analyzed the important interactions between DNA nucleotides and AGT residues in all three complexes, complex-I, complex-II, and complex-III.

There were several nucleotides in DNA that took part in interactions. Out of them, in each of three complexes, 6OG7 (GUA7), THY9, and THY23 were observed as the major binding nucleotides in DNA. Besides these major pairs, many other nucleotides also assist in the formation of the complex. These pairs occasionally came into interactions and were short lived. Most of the time, they lie out of the interacting range of residues in AGT. Majority of the interacting nucleotide-residue partners were detected similar in all three complexes. Active nucleotides lie at the minor groove and hence contributed in binding AGT at this position except THY23.

After recognizing the major interacting nucleotides, we investigated the atomic level interaction by examining the hydrogen bond occupancy. The hydrogen bond occupancy percentage of some residues may be greater than 100%, because of interaction of one nucleotide to two or more atoms of same or other amino acid residues. In this part, we have determined the occupancy percentage by taking unique hydrogen bonds, which are smaller than 100%.

In the case of complex-I, TYR114, CYS145, SER151, THR95 and GLN115 play major role in the formation of hydrogen bonds with the nucleotides. Out of several other significant binding pairs, TYR114-6OG7 has the maximum hydrogen bond occupancy with 87.6%. Likewise, this pair was followed by CYS145-6OG7 (87.1%) and SER-THY9 (85.6%) pairs. Besides these, THR95 and GLN115 were found to be contributed in the formation of hydrogen bonding having > 40% occupancy as summarized in the complex-I part of Table 3. In Table 3, atoms that involved in hydrogen bonding are also included into the brackets together with its corresponding residue.

Table 3: Hydrogen bond occupancy of all three complexes during 200 ns NPT simulation. First column shows the type of complex, second and third column shows the residue-nucleotide pairs with specific atoms that were involved in hydrogen bonding, and the fourth column shows the occupancy percentage of residue-nucleotide pairs from corresponding column in second and third.

AGT-DNA complex	Residues-Nucleotides pairs		Occupancy
	Residues	Nucleotides	
Complex-I	TYR114 (OH)	6OG7 (N3)	87.6%
	CYS145 (O)	6OG7 (N2)	87.1%
	SER151 (OG)	THY9 (O1P)	85.6%
	THR95 (OG1)	THY23 (O1P)	80.3%
	SER151 (N)	THY9 (O2P)	69.6%
	THR95 (N)	THY23 (O1P)	62.8%
	GLN115 (N)	THY9 (O1P)	53.2%
Complex-II	SER151 (OG)	THY9 (O1P)	90.0%
	THR114 (OH)	6OG7 (N3)	71.6%
	THR95 (N)	THY23 (O1P)	69.2%
	SER151 (N)	THY9 (O2P)	68.4%
	THR95 (N)	THY23 (O1P)	56.1%
	CYS145 (O)	6OG7 (N2)	53.8%
	GLN115 (N)	THY9 (O1P)	48.9%
ARG128 (NH2)	CYT20 (N3)	47.8%	
Complex-III	TYR114 (OH)	GUA7 (N3)	59.0%
	SER151 (N)	THY9 (O1P)	55.8%
	PHE94 (N)	THY23 (O1P)	48.4%
	ORT145 (O)	GUA7 (N1)	47.8%
	ARG135 (NH1)	GUA7 (O5')	44.6%
	ASN157 (ND2)	GUA7 (O4')	43.7%
	ORT145 (O)	GUA7 (N2)	40.9%

In complex-II, 6OG7 and CYS145 have covalent bonding. It has a slight variation in hydrogen bond occupancy percentage than that of complex-I. In this complex, SER151-THY9 has the maximum contribution, with the hydrogen bond occupancy percentage of 90.0%. This pair was followed by THR114-6OG7 (71.6%) and THR95-THY23 (69.2%). The occupancy of the remaining pairs is shown in the complex-II part of Table 3. One new pair, ARG128-CYT20 was noticeable having hydrogen bond occupancy percentage > 40% in complex-II. The newly detected pair is ARG128-CYT20 (47.8%) in complex-II. The atomic level investigations hydrogen bonding of all possible pairs that have occupancy percentage > 40% are shown in complex-II part of Table 3.

The hydrogen bond occupancy percentage of residue-nucleotide pairs was significantly lower in complex-III than that of complex-I and complex-II. In complex-III, the pair TYR114-GUA7 possessed the maximum value of hydrogen bond occupancy percentage with 59.0% followed by SER151-THY9 (55.8%) and PHE94-THY23 (48.4%). The remaining pairs with their corresponding occupancy percentage are shown in the complex-III part of Table 3. A new pair ARG135-GUA7 appeared possessing an occupancy percentage > 40% in this complex, but this bond has occupancy percentage < 40% complex-I and complex-II.

The overall pattern of hydrogen bond occupancy percentage revealed that the contribution of hydrogen bonds in complex-III has the lowest value among the three complexes. Complex-II has the relatively greater percentage of hydrogen bond occupancy due to the fact that covalent bond tends to approach the atoms in DNA and AGT nearer to each other. But, this is the short transient state, so these bonds can not sustain for long time (Patra et al., 2016; Y. Fu et al., 2021). In comparison to two stable states, complex-I has a higher binding affinity than complex-III. This depicts that complex-I refers to the complex formation stage and complex-III refers to the dissociation stage of DNA and AGT. Though the binding affinity is low, complex-III was not in the condition of easy dissociation, rather its intermolecular binding is degraded and AGT can be inactivated. The binding regions (site 1 and site 2) and corresponding residue-nucleotide pairs are shown in Figure 25.

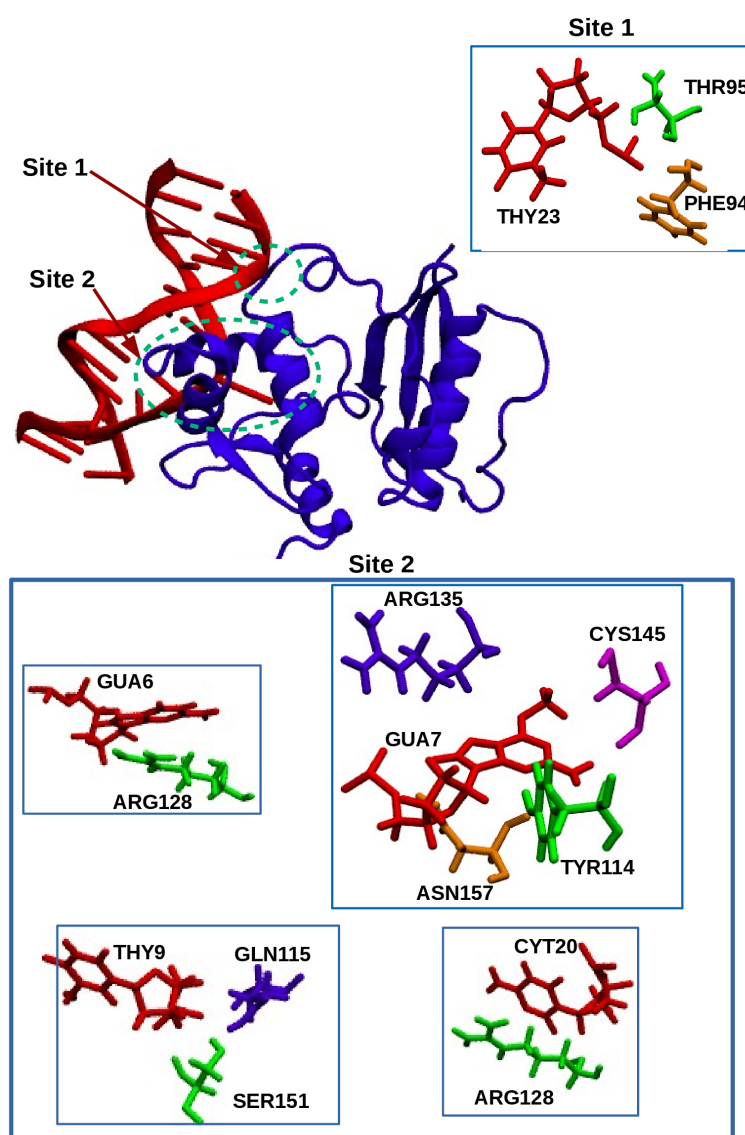


Figure 25: Major interacting sites of DNA and AGT via hydrogen bonding. Site 2 contains many nucleotides that contribute in binding the AGT at the minor groove of DNA. Site 1 also contributes a noticeable hydrogen bonding.

Targeting the atomic level visualization in major binding partners of nucleotides and residues, we represented them forming two sites; site 1 and site 2 as shown in Figure 25. In this figure, noticeable binding partners are shown. Major binding partners are kept in rectangular boxes. In site 1, THY23 has the major contribution in binding with AGT. The AGT residues, THR95 and PHE94, have the leading role in the formation hydrogen bonds at this site. Site 2 is the strongest binding region between DNA and AGT. In this region, several nucleotides actively participated in the formation of hydrogen bonding with AGT residues. These pairs were chosen in accordance of hydrogen bond occupancy percentage in Table 3.

To analyze more specifically, the hydrogen bonding formation patterns are shown in Figure 26. This figure includes the atomic level information for the major bindings partners and are also common in all three complexes. The major interacting nucleotide-residue partners were somehow similar in all three complexes, but the strengths are different.

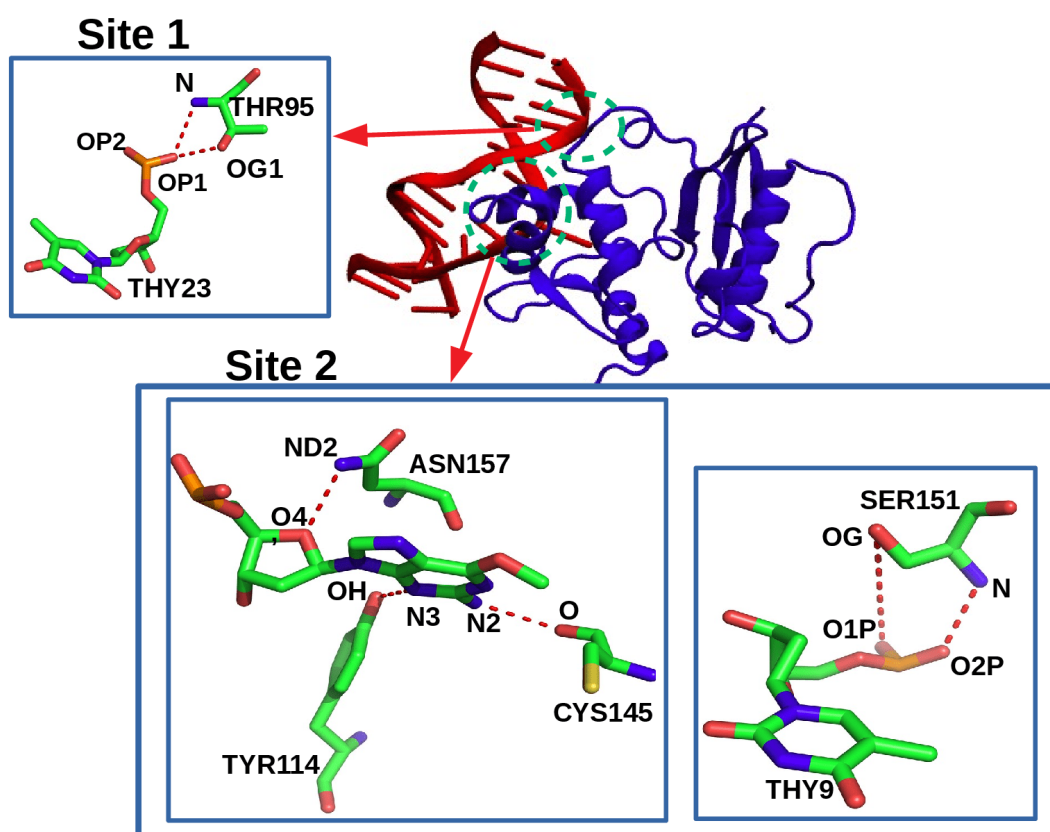


Figure 26: Major interacting sites of DNA and AGT showing atomic level hydrogen bonding in major binding sites. Two major regions of binding are shown in the rectangular box in magnified form.

The atomic level hydrogen bonding analysis revealed that 6OG7(GUA7), THY9 and THY23 have the viral role in formation of stable complex of DNA and AGT. We further investigated the overall hydrogen bonds variation of these nucleotides within the simulation time scales. Number of hydrogen bond versus simulation time are shown

in Figure 27. The hydrogen bond contributions in three nucleotides, 6OG7(GUA7), THY9 and THY23 in all three complexes were compared. We found that all three nucleotides formed stable hydrogen bonding in complex-I as shown in Figure 27(a), but the these bonds were slightly fluctuating in complex-II and complex-III. Moreover, It has been observed that the fluctuations detected in complex-II and complex-III were due to relatively unstable binding of THY9 and THY23. 6OG7 (GUA7) was found to be most consistent in all three complexes.

From Figure 27, we can clearly seen that THY23 has almost zero hydrogen bonding after 150 ns in complex-II as shown in Figure 27(b). This bond is also fluctuating in complex-III as shown in Figure 27(c). In conclusion, GUA7 has significant contribution not only in the methyl transfer process but also in the formation of stable complex. Our findings regarding the structural stability of DNA-AGT complex due to hydrogen bonding, electrostatics and van der Waals contact agree with many previously reported results of (Daniels et al., 2004; Patra et al., 2016; Y. Fu et al., 2021).

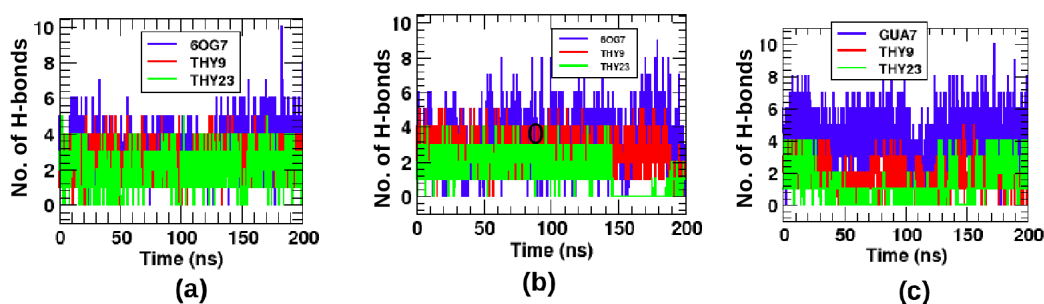


Figure 27: Hydrogen bond contributions in nucleotides 6OG7(GUA7), THY9 and THY23 with the protein residues for (a) complex-I (b) complex-II and (c) complex-III. 6OG7(GUA7) has the stable hydrogen bonding in all three complexes, however THY9 and THY23 has the relatively fluctuating in complex-II and complex-III.

4.6 Contact Surface Area

The contact surface area between DNA and AGT is useful in examining the structural stability of the molecular complex. Solvent accessible surface area (SASA) is used to estimate the contact surface area of two molecules. We estimated the contact surface area of all three complexes during the 200 ns simulation to compare the binding affinity in these systems. For this, an average value of solvent accessible surface area (SASA) of an individual molecule and their complexes were determined. Then, these data were used to determine the contact surface area of AGT and DNA by using the relation (Yunta, 2017),

$$\text{Contact surface area (A)} = \frac{A_1(t) + A_2(t) - A_3(t)}{2} \quad (4.3)$$

where, $A_1(t)$ is the SASA of AGT, $A_2(t)$ is the SASA of DNA, and $A_3(t)$ is the SASA of their complex. In the Table 4, A_1 , A_2 and A_3 are the average solvent accessible surface area (SASA) of AGT, DNA and their complex (complex-I); A'_1 , A'_2 and A'_3 are the SASA of AGT, DNA and their complex (complex-II) and A''_1 , A''_2 and A''_3 are the SASA of AGT, DNA their complex (complex-III).

Table 4: Average contact surface area between DNA and AGT that are obtained from 200 ns simulations for all three systems. The net surface area was found maximum in complex-II and minimum in complex-III.

Complex	SASA (\AA^2) for			Double contacts $A_1 + A_2 - A_3$	Net contact surface area (A) $(A_1 + A_2 - A_3)/2$
	AGT (A_1)	DNA (A_2)	complex(A_3)		
I	10055.72	5531.20	13798.48	1788.44	$(894.22 \pm 29.57) \text{\AA}^2$
	AGT (A'_1)	DNA (A'_2)	complex(A'_3)	$A'_1 + A'_2 - A'_3$	Net contact surface area (A')
II	10236.34	5679.28	14095.76	1819.86	$(909.93 \pm 68.69) \text{\AA}^2$
	AGT (A''_1)	DNA (A''_2)	complex(A''_3)	$A''_1 + A''_2 - A''_3$	Net contact surface area (A'')
III	10214.99	5669.13	14127.16	1756.96	$(878.48 \pm 40.19) \text{\AA}^2$

Table 4 shows the quantitative estimation of mean contact surface area between DNA and AGT in all three complexes. Among three complexes, complex-II has the largest average contact surface area $(909.93 \pm 68.69) \text{\AA}^2$ with greatest fluctuation $(653.54 \text{ to } 1086.78) \text{\AA}^2$, complex-III has the smallest average contact surface area $(878.48 \pm 40.19) \text{\AA}^2$ with fluctuation from $(711.88 \text{ to } 1037.62) \text{\AA}^2$. Complex-I has the intermediate value $(894.22 \pm 29.57) \text{\AA}^2$ ranging from $(761.53 \text{ to } 1019.25) \text{\AA}^2$ showing the least fluctuation having consistent contact surface area as shown in Figure 28. Supporting the hydrogen bond occupancy percentage in Table 3, complex-II has the greatest contact surface area in this complex.

The time evolution of contact surface area is shown in Figure 28. The contact surface area is almost the same in complex-I throughout the simulation predicting the most favorable complex in comparison to other two complexes. The surface area was detected fluctuating in complex-II and complex-III between 140 ns to 170 ns. The fluctuation in surface area gives the evidence of unstable binding between DNA and AGT. Comparison of surface area of contact for three complexes during the 200 ns simulation is shown in Figure 28.

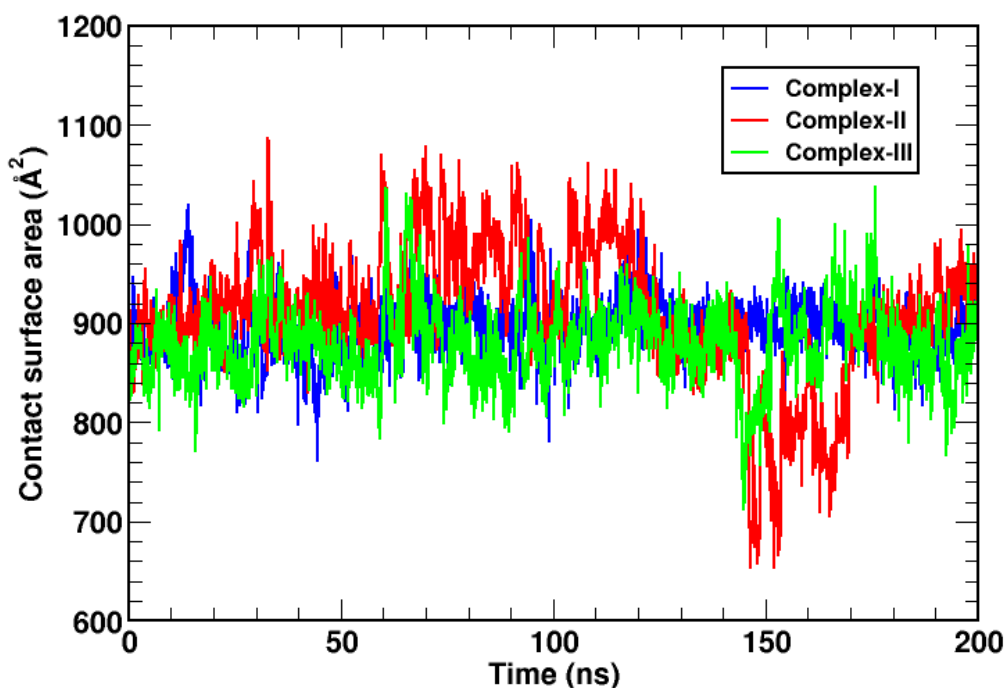


Figure 28: Contact surface area of DNA and AGT in all three complexes. The area is highly consistent in complex-I but fluctuating in other two complexes.

4.7 Possible Guanine Base Flipping Mechanism

DNA base flipping mechanism involved in genomics repair and epigenic control of gene expression is often occurred. During the DNA replication and transcription, DNA coding at the flipping region may be missed and the newly formed DNA strand can be mutated. This mechanism is essential for methylation damage repair process. The methylation at nitrogen base is a chemical modification, which in turn affects to the structural modification. The structural modification provides the space for AGT to bind in appropriate position of DNA. Till date, the driving force of base flipping in nucleotide containing methyl adduct has not been fully understood (Lemkul et al., 2014).

Considering the base flipping mechanism is one of the essential steps in methyl transfer process, we modelled two molecular systems taking only DNA from PDB ID 1T38: one DNA containing a methyl adduct at seventh residue and another a normal DNA (i.e., replacing the methyl adduct from seventh residue of DNA). Both the systems were propagated 100 ns simulation run under NPT condition. These investigation would provide the insight of the major cause of base flipping in methylated guanine and reason of stability of orphaned nucleotide after the flipping its nucleotide pair partner. In our

system, GUA7 is the flipping base and CYT20 is the pair partner of flipped base. For this, we investigated the flipping mechanism taking two basic conditions: effect on hydrogen bonding between GUA7 and CYT20 during methylated and non-methylated conditions and the role of sidechain of intercalated ARG128 in AGT that degrades the hydrogen bonding between methyl GUA7 and CYT20.

This section is basically dealt with the causes of flipping out of GUA7 from the base stack and the stability of orphaned CYT20. We have prepared the new molecular systems to deal the flipping mechanism and previously performed 200 ns NPT simulation utilized to study the stability of CYT20.

4.7.1 Bond Breaking due to Methylation

In this part, we study basic reason of weakening the hydrogen bonds between 6OG7 (methylated GUA7) and the CYT20 due to methylation at GUA7. The weak hydrogen bond is vulnerable in dissociation of corresponding nucleotide pairs. Our fundamental interest is to flip out 6OG7 to the vicinity of active cysteine (CYS145) so that the methyl transfer mechanism would be possible. If the base does not flip out from the base region, it is almost impossible to receive the methyl adduct by active CYS145, i.e., 6OG7 should lie within the active packet CYS145 of AGT. That is why, to access the alkylated base 6OG7 with AGT active site, the base flipping is essential. As explained in the previous section, ARG128 forms the hydrogen bonds with CYT20 and facilitates in flipping out 6OG7 extrahelically and then makes methyl adduct approaching in active packet cysteine.

To examine the possible causes of the DNA base flipping mechanism, we have prepared the molecular systems picking only the double-stranded DNA by removing the AGT from protein structure 1T38.pdb. We modelled the system by flipping in methyl guanine (6OG7) from the DNA base stack such that methyl guanine could approach as the corresponding pair partner, CYT20. Then, two different systems were generated by forming: first GUA7 containing methyl adduct at O6 point and another the normal GUA7 (without methyl adduct) as shown in Figure 29. Figure 29(a) contains a double stranded DNA with normal guanine-cytosine pairs and Figure 29(b) has the methyl guanine pairing with cytosine.

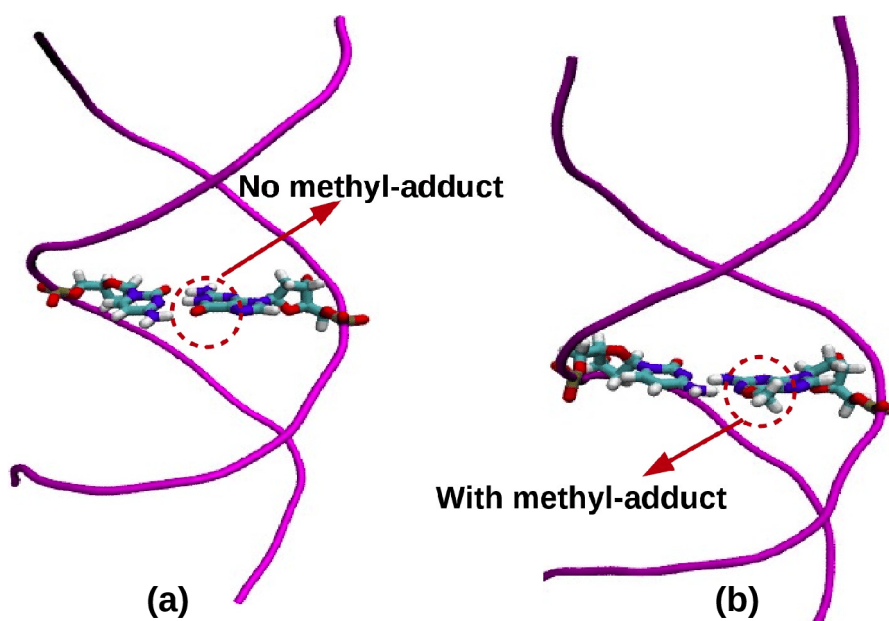


Figure 29: Only DNA system (a) structure of non-methyl GUA7 in dsDNA (b) structure of methyl GUA7 in dsDNA.

In the ordinary condition, cytosine binds to guanine via three hydrogen bonds, however, a different hydrogen bonding scenario was observed during the O6-point methylation at guanine. We would like to inspect the effects on hydrogen bonding due to methylation damage at GUA7. The presumption of the study was that the strength of hydrogen bonds should be weaker in methylation damage condition than that of normal condition so that the base flipping mechanism would be favorable. An investigation was carried out in regard to examining the condition of hydrogen bonds when the O6-point of guanine is methylated. For this, we have performed 100 ns NPT simulations for both the systems. The trajectories obtained from each of 100 ns simulation run was used to study the possible base flipping mechanism due to hydrogen bond breaking in GUA7 and CYT20.

In order to ensure the structural stability, we estimated the RMSD of both systems. The RMSD was calculated for all atoms of DNA backbones without taking hydrogen atoms. The atomic coordinates of the first step of the simulation were taken as the reference coordinates. Then, the RMSD of both systems was plotted as the time-dependent function. The RMSD for non-methyl guanine condition was relatively smaller and stable than that of methyl condition as shown in Figure 30. During the time between ~24 ns to ~60 ns, the RMSD of methylated DNA is relatively larger and sometimes crossed the value of 8 Å. But, after a 60 ns simulation run, the structure was well established and have almost the same value of RMSD in both systems showing the stable structures.

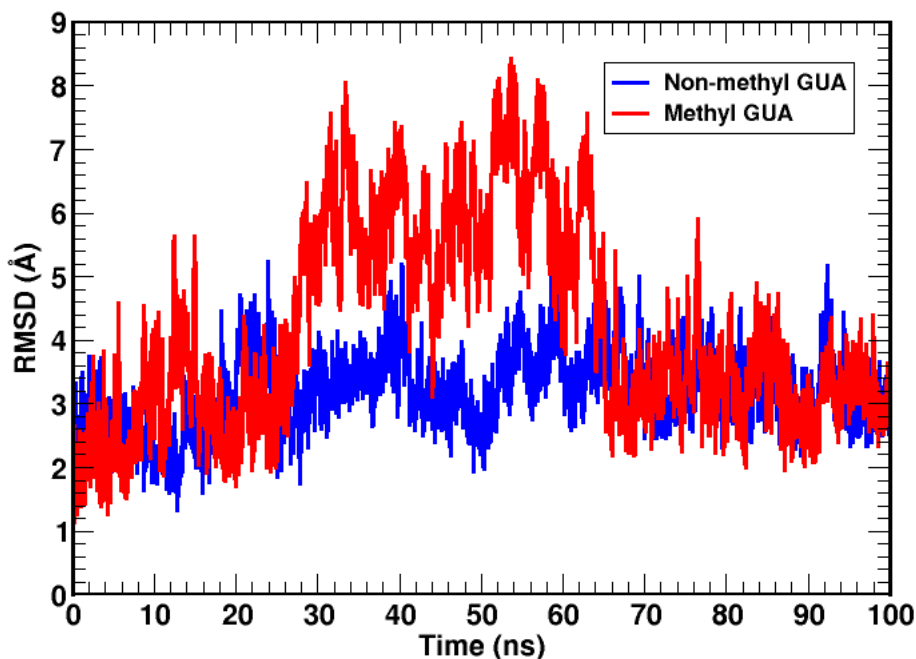


Figure 30: The RMSD of the DNA for methyl guanine and non-methyl guanine system. More fluctuations is found in methyl condition than that of non-methyl condition.

The hydrogen bond pattern of GUA7 and CYT20 are shown in Figure 31. After the simulation, we found that the carbon at CH₃ was slightly rotated away from its hydrogen bond partner N4 of CYT20 and no stable hydrogen bond exists in this position. Furthermore, O2 of CYT20 left the N2 of GUA7 and eventually formed N2-N3 and N1-N3 pairs. Nucleotide pairs in Figure 31 represents the patterns of hydrogen bond formation in methylated and non-methylated conditions.

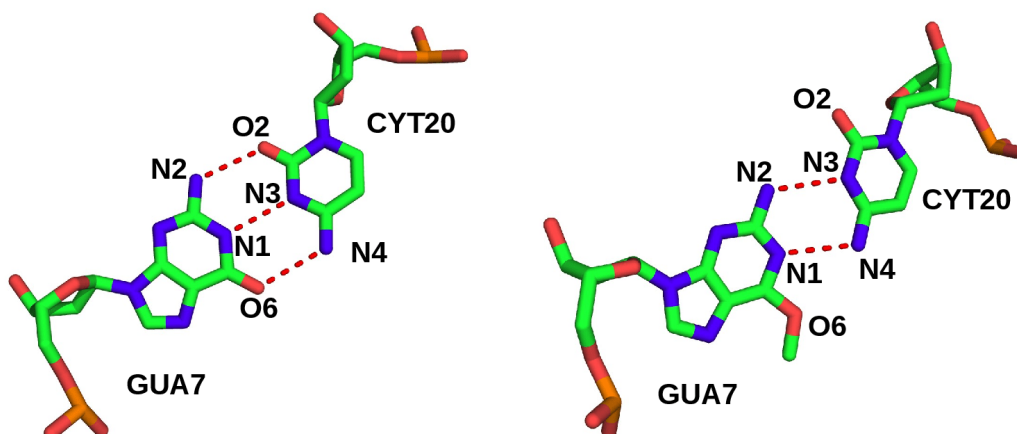


Figure 31: The scheme of formation of hydrogen bonds in non-methylated guanine (GUA7) and CYT20, and methyl guanine (6OG7) and CYT20. All three hydrogen bonds exist in non-methylated guanine condition as the ordinary pairs, however only two hydrogen bonds sustain in methylated guanine.

The number of hydrogen bonds between these two nucleotide pairs, 6OG7(GUA7) and CYT20, is shown in Figure 32. Figure 32 shows that three stable hydrogen bonds exist

in non-methylated GUA7 throughout the simulation, however, only two hydrogen bonds were observed in methyl GUA7. This is one of the main reason of easy rotation of methylated base across the backbone. Even though the number of bonds were fluctuated in the beginning of the simulations, they were stable after 60 ns simulation run. This fluctuation may be arised due to disorientation of flipped in guanine. As soon as the guanine restored the correct orientation, three hydrogen bonds became stable.

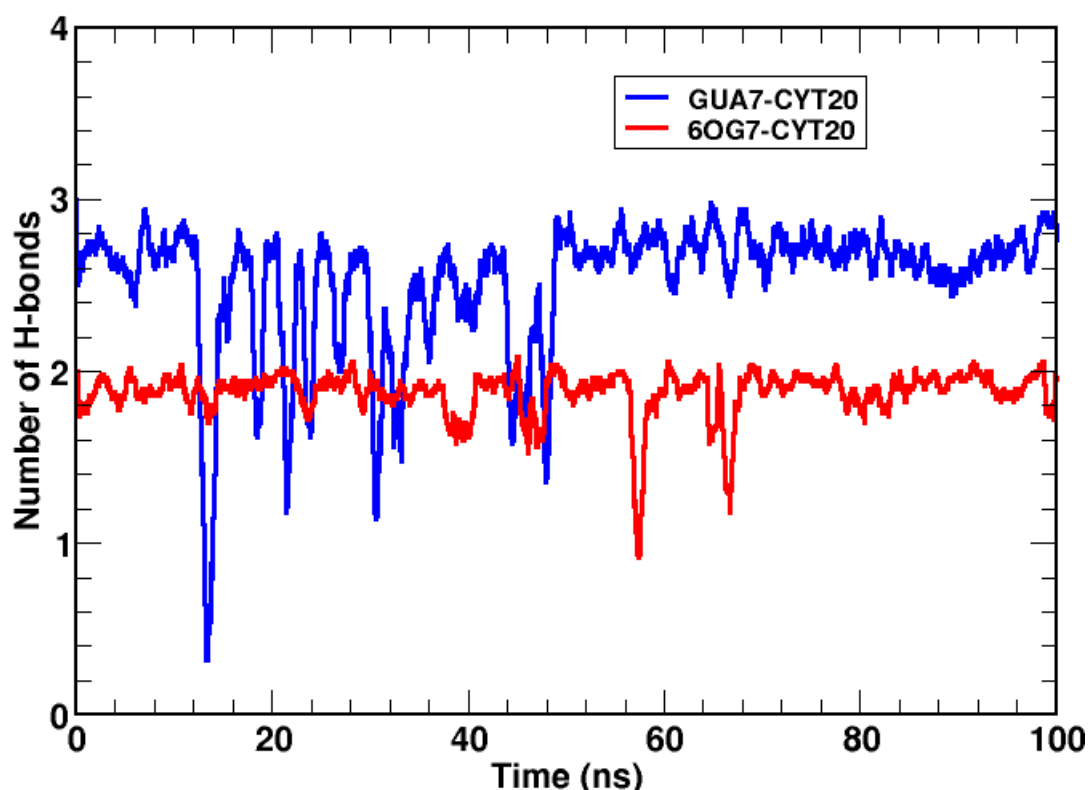


Figure 32: Comparison of the number of hydrogen bonds formed by non-methylated GUA7 and CYT20, and methylated GUA7 and CYT20 pairs. Number of hydrogen bonds are clearly distinguishable in two cases.

The loss of hydrogen bond formation at O6 of guanine can be the G:C to A:T transition in DNA replication, which is a major cause of cell to be carcinogenic. From the above simulation, It has been detected that methylated O6 overlooks in the formation of hydrogen bond to any other nucleotide that comes across it. So, ordinary three hydrogen bonds reduces to only two. This condition not only facilitates in nucleotide flipping but also mutate the newly transcribed DNA strand. In the next subsection, we describe the stability of unpaired CYT20, when GUA7 rotated from the base stack.

4.7.2 Stability of CYT20

During the direct damage repair in DNA, the DNA repair protein like AGT should recognize and trigger the damaged part, and then flip the methyl-damaged nucleotide

out of the base stack. As the O6 point of guanine is the targeted methylation damage, the fundamental requirement of the AGT is to identify the damaged part at the O6 point of guanine. For this, ARG128 of AGT intercalates in between the backbones of the double-stranded DNA and hits the 6OG7 (methylated GUA7), the pair partner of CYT20. The side chain of ARG128 then attempts to form the hydrogen bonds with the CYT20, which is supposed to be broken the hydrogen bonds in the 6OG7-CYT20 pair. When the hydrogen bonds are broken in these nucleotide pair partners, the methylated base rotates out from the backbone and becomes stable at the AGT active packet forming several hydrogen bonds with amino acid residues residing at the interfacial region of the complex (Daniels et al., 2004). Motivated from such interesting role of ARG128, we examined the interaction mechanism of ARG128 with CYT20.

After the rotation of GUA7, its pairing partner CYT20 becomes orphaned, i.e., lost the pair partner and became alone and free as shown in second structure of Figure 15. This orphaned CYT20 when gets free from its binding partner, it may be oriented in any direction, which may loss the structural integrity of the double stranded DNA. But, this was not happened in reality. We then examined how orphaned CYT20 became stable and what the major cause of preventing from the destruction of structural integrity. To find the cause of stability of orphaned nucleotide, we analyzed the trajectories of 200 ns simulation of complex-I targeting the interaction between ARG128 and CYT20.

During the 200 ns NPT run, three points of CYT20 (N4, N3, and O2) were found forming the hydrogen bonds with a point NH2 of ARG128 as shown in Figure 33. These three atoms of CYT20 do not approach within the cutoff distance of 3.5 Å simultaneously to form the hydrogen bonds with ARG128, rather one or two of them form the bonds at a time. O2 was the weakest bond and occasionally approaches within the cutoff distance. At the beginning of the simulations, NH2 is relatively nearer to the N4 point. Over the time evolution of the MD simulation, the hydrogen bond between NH2 and N3 was found to be the strongest and was detected to have the dominating effect than other two bonds.

The radial distributions of contact atoms of CYT20 from NH2 of ARG128 were further investigated. Out of three hydrogen bond pairs, NH2-N4, NH2-N3, and NH2-O2, the second pair remains closer in most of the simulation time as shown in Figure 33. The interaction remained persistent throughout the simulation and stabilized the orphaned CYT20. Thus, the orphaned CYT20 became stable, though its pair partner GUA7 is rotated out from the base stack. The variation of distance of interacting atoms of CYT20, N3, N4 and O2 from the NH2 atom of arginine are also shown in Figure 33(c). This figure shows that distance between NH2-N3 is the most consistent and that of NH2-N4 is most fluctuating. Most of the time, atom O2 lies at the farthest distance from NH2.

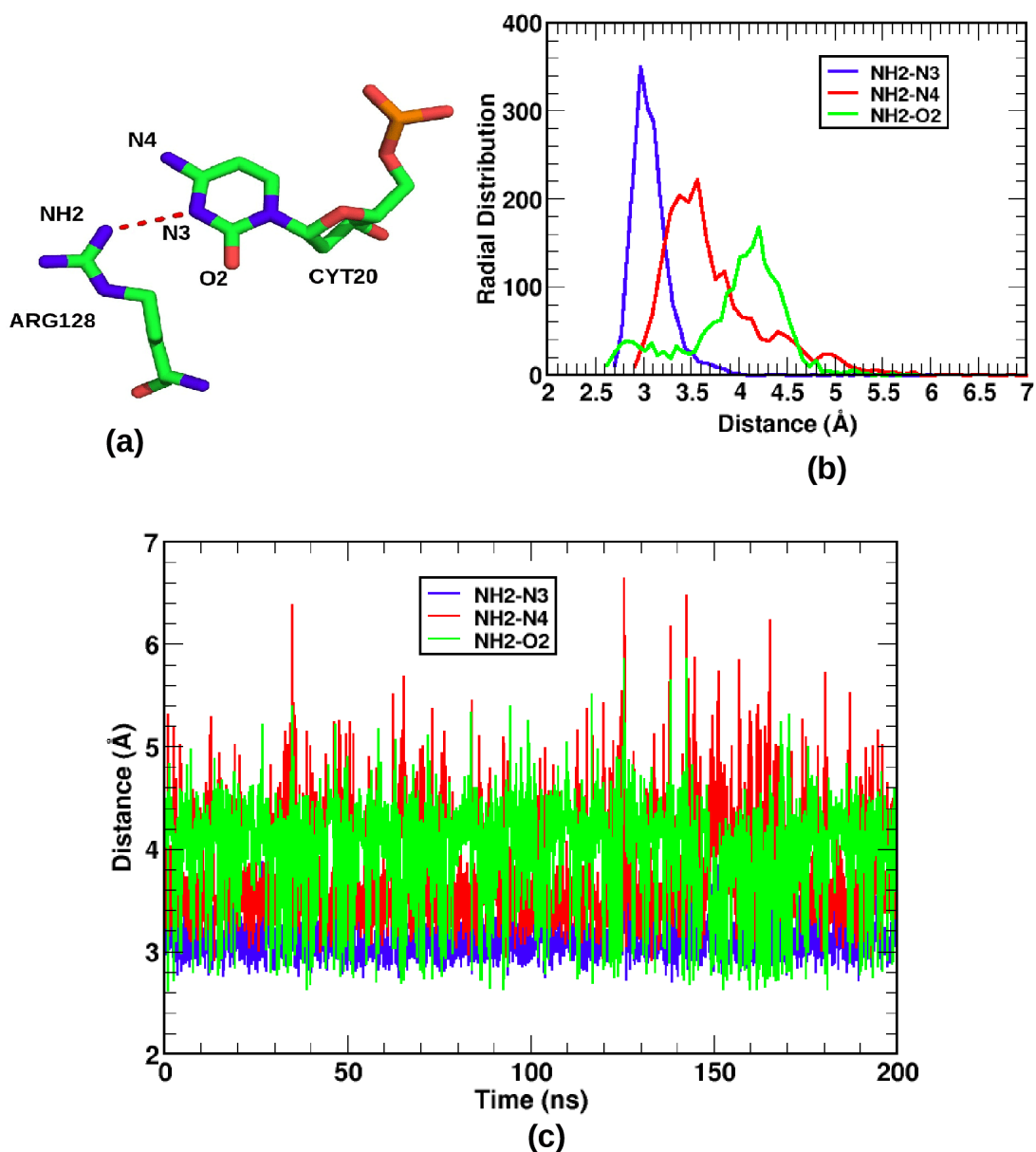


Figure 33: (a) Hydrogen bonding pattern in ARG128-CYT20 pair showing the strongest hydrogen bond between NH2-N3, (b) Radial distribution for NH2-O2, NH2-N3, and NH2-N4 pairs. NH2-N3 pair has the highest distribution and lie within 3.5 Å, but other two pairs occasionally comes within this range. (c) Distance of N3, N4, and O2 atoms of CYT20 from NH2 atom of ARG128.

4.8 Steered Molecular Dynamics (SMD) on AGT

Steered molecular dynamics (SMD) mimics the atomic force microscopy (AFM) experiments and aids to identify the active amino acid residues, involving in protein-ligand, protein-protein and protein-DNA interaction (Phillips et al., 2005). We applied this method to determine the possible interactive mechanism between base flipped DNA and AGT; and also to examine the conformation changes in both DNA and protein residues during the translation of protein from DNA.

In this work, SMD was employed to estimate the solvent accessible surface area (SASA) of DNA, AGT and their complex for both systems. The estimated SASA was implemented to find the contact surface area between AGT and DNA. Further, we performed SMD simulations specifically to find the maximum possible decoupling force (rupture force) during dissociation of AGT from methylated DNA in complex-I and was also performed for the complex-III.

To investigate the aforementioned properties, we have taken the molecular structures from complex-I and complex-III. The original PDB structure contains a methylated DNA (mDNA) at O6 position of GUA7 and DNA repair protein AGT, a complex with non-covalent bonding taken from complex-I, which is named system-I in this sections and another system was prepared by transferring methyl adduct to cysteine (CYS145) in AGT (system-II), which is taken from complex-III.

To perform the SMD simulations, the AGT protein was pulled along positive x direction under constant velocity pulling method with 1 Å/ns taking the DNA as the reference molecule (Isralewitz et al., 2001). The total trajectory length was 30 ns, which was sufficient to evaluate the maximum binding force between AGT and DNA for the supplied uniform velocity. Simulation of each system was repeated for five times and outputs were averaged for corresponding systems. The spring constant was specified to be 1 kcal/(mol·Å²). This procedure was taken the force as the function of displacement. Same procedure has been performed for system-II. The present investigation would be useful for the further study of protein-DNA complex formation, their structural stability and DNA damage repair process.

SMD is often executed to investigate in the binding mechanisms of protein-ligand molecules. Besides this common application of SMD, we applied this technique in the study of protein-DNA binding mechanism, which is expected as the new application for the molecular system and can be the future avenue in the study of DNA-protein interactions.

4.8.1 Pulling of AGT from DNA

We have used the prepared systems, system-I and system-II, to investigate how nucleotide-residue pairs behave during the decoupling AGT from DNA. In both systems, AGT was pulled along positive x-direction taking DNA as the reference molecule. The pulling direction of AGT is shown in Figure 34. Then, a suitable time scale was chosen to evaluate the maximum force required to decouple AGT from DNA. Both systems were simulated as the trial so that this could help us to predict the suitable time scale for the SMD simulations. The trial results guided us to set the simulation time of 30 ns for each

system. Then, both systems were propagated for 30 ns SMD simulations. Each of the systems was repeated five times targeting to get the average value.

In this process, the backbone atoms of DNA were taken as the fixed atoms and alpha carbons in AGT were taken as the dummy atoms; and were pulled from their center of mass (COM) along the positive x-direction with constant velocity ($\vec{v} = d\vec{x}/dt$) in water and ions environment so that the SMD atom experiences the force $\vec{F}(t) = k(\vec{v} t - \Delta\vec{x})$ (Isralewitz et al., 2001), providing the external potential energy (Martínez et al., 2008),

$$U(x, t) = \frac{1}{2}k(\vec{v} t - \Delta\vec{x} \cdot \hat{n})^2 \quad (4.4)$$

where, k ($= 5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$) is the spring constant and gives the stiffness of the applied harmonic restraining force, and $\Delta\vec{x}(t) = \vec{x}(t) - \vec{x}_0$, is the displacement between the molecules being actual position $\vec{x}(t)$ and initial position \vec{x}_0 and \hat{n} is the unit vector along the direction of pulling.

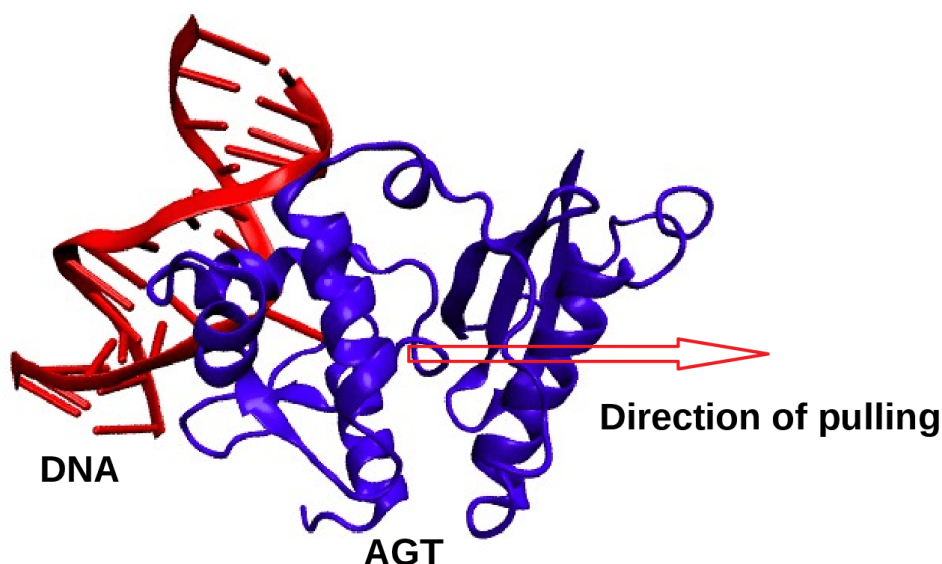


Figure 34: Pulling of AGT along positive x-direction taking DNA as the reference molecule. In this condition, AGT is pulled in both systems, system-I and system-II, along the same direction (positive x direction) in identical condition.

The molecular structures that are obtained from the end of 30 ns SMD simulation are shown in Figure 35. During the simulations, we have detected the decoupling of site 1 in system-I and still in contact in system-II. A part of AGT that binds in the interfacial region of site 2 was found stronger binding than that of other regions. Only a short loop in AGT was detected stretching.

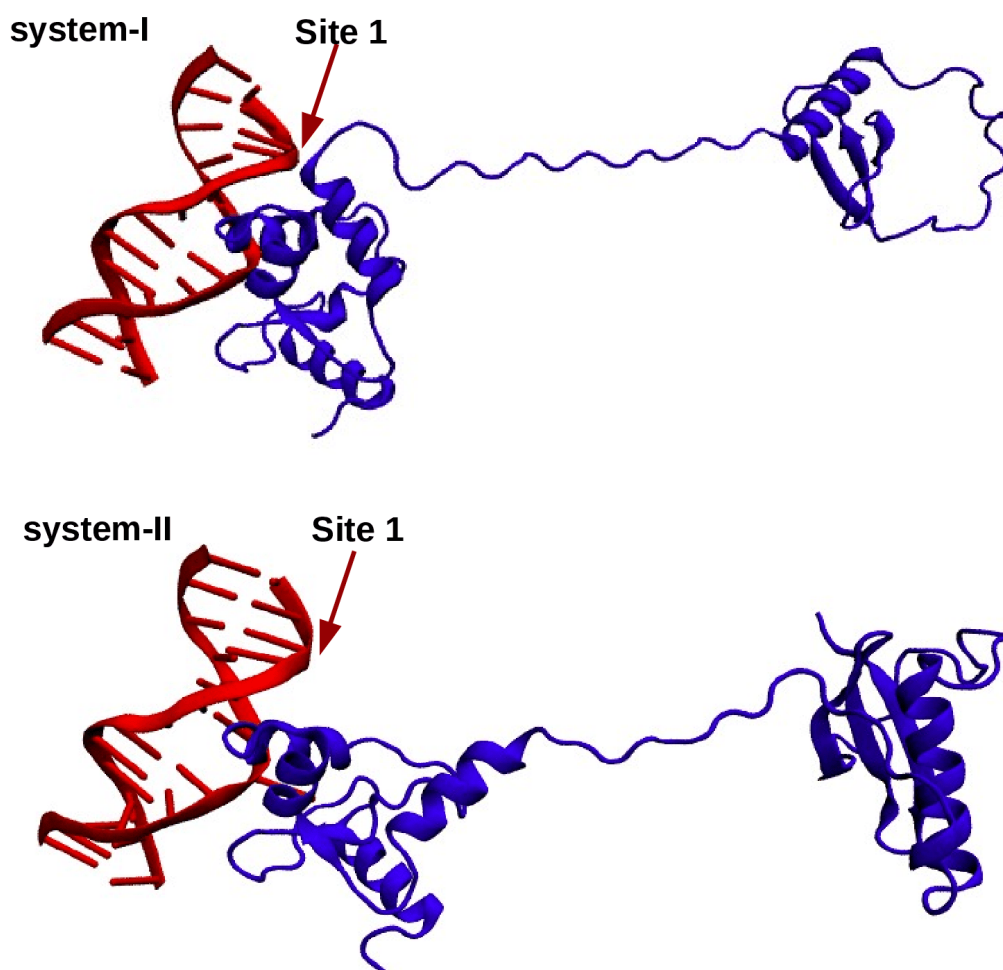


Figure 35: Binding sites of AGT and DNA. Two binding regions (site 1 and site 2) retains upto entire 30 ns simulation in system-I and only one binding region (site 2) exist till the end of the simulation, however, the site 1 was decoupled during the simulation in system-II.

The variation of hydrogen bonds in system-I and system-II during the pulling of AGT from DNA is shown in Figure 36. We detected two specific sites of binding via hydrogen bonds in both of these systems. The strength of hydrogen bonds is greater in site 2 than that of site 1. In the beginning of the simulation, the number of hydrogen bonds was obtained almost equal in both systems, but the number was found slightly decreasing faster in system-II immediately after 1 ns simulation run. In both systems, the number remains almost consistent up to 13 ns, however a sudden down jump was observed in system-II for a certain time. After 22 ns, number of hydrogen bonds remains consistent upto the 30 ns as shown in Figure 36.

The sudden down jump supports the low occupancy percentage of hydrogen bonds in complex-III than that of complex-I as explained in Table 3. In addition, the dissociation of site 1 is also the important region of down jump of H-bonds as shown in Figure 35. Due to the lower strength of hydrogen bonding in system-II (taken form complex-III), the dissociation is easier. Despite the lower occupancy percentage of hydrogen bonds, site 2

was not dissociated due to the fact that there are many number of nucleotide-residue pairs. This region can be considered the strongest binding region. No significant down jump in system-I (taken from complex-I) shows that complex formation mechanism is more favourable in pre-methyl condition than that of post-methyl transfer.

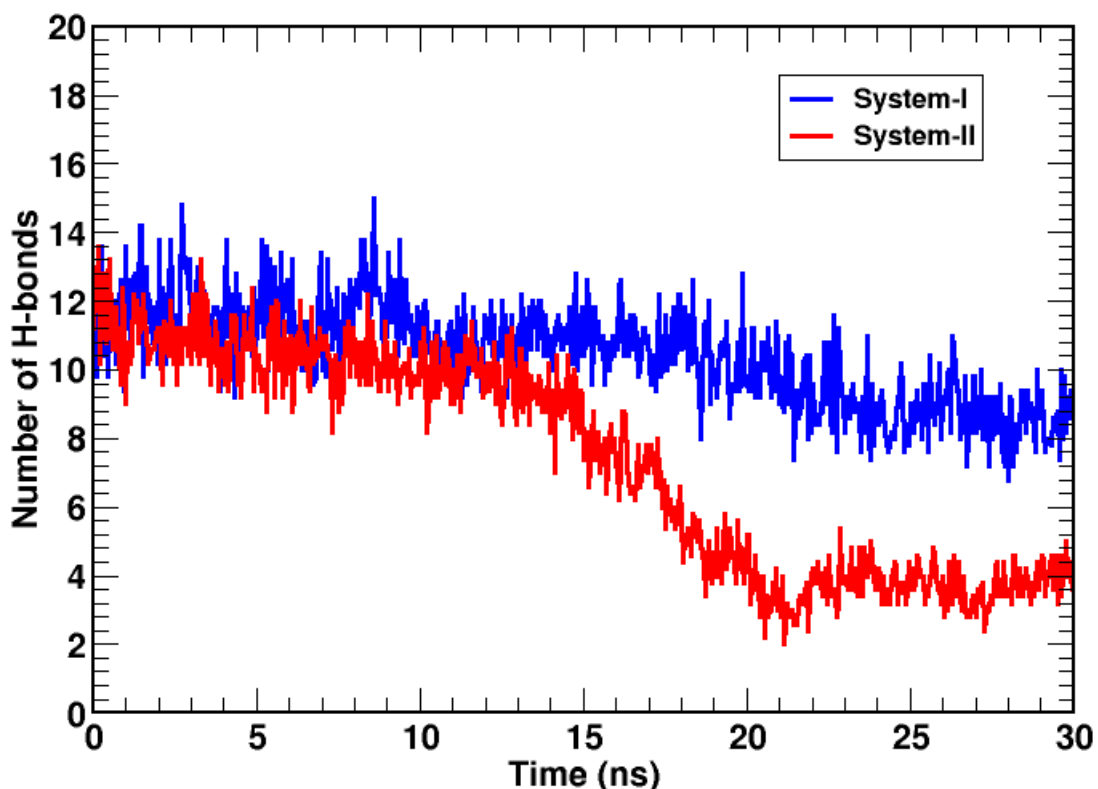


Figure 36: The variation of hydrogen bonds in system-I and system-II. The sharp decrease in the number of hydrogen bonds in system-II occurred due to the decoupling of hydrogen bonds from site 1.

As soon as the site 1 was dissociated from DNA, the binding loop of AGT stretched away and suddenly lowered the binding strength. The effect of lower occupancy of hydrogen bonds can be visualized in site 1 of this system. In site 2, several nucleotides GUA7, THY9, were detected that actively participated in binding with interfacial amino acid residues of AGT. So, hydrogen bonds in site 2 were continuously working to retain the molecules in the form of a complex in both systems. These evidences of major interacting residues in site 2 also favours the the major hydrogen bonds at the interface region of DNA and AGT as explain in Table 3.

Furthermore, an intra-molecular hydrogen bond breaking was detected within the AGT molecule. The hydrogen bond dissociation region of AGT is shown in Figure 37. Although equal force is applied in each of the alpha carbon, there was a certain stretching produced in the AGT molecule. The hydrogen bond dissociation in GLU172-TYR69, HSD171-PHE69, HSD171-PHE70 pairs and aromatic-aromatic interaction breaking in PHE70-TRP167 pair in site 3 contributes to stretching the AGT. Intra-molecular hydrogen

bond breaking is the main region of stretching the loop in AGT.

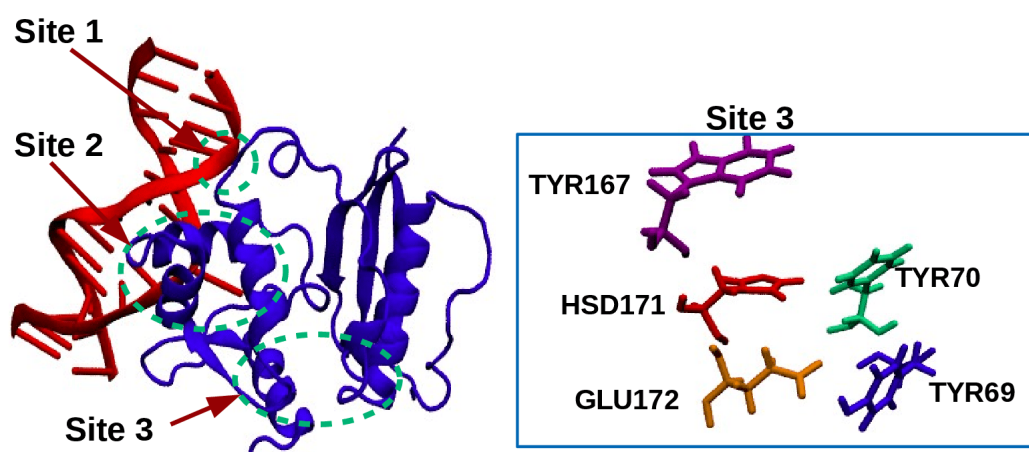


Figure 37: The intramolecular decoupling region of AGT. Breaking of hydrogen bonds in region 3 facilitates the AGT to stretch with the certain part.

4.8.2 Solvent Accessible Surface Area (SASA) and Variation of Contact Surface Area during SMD Pulling

Solvent accessible surface area (SASA) is the geometric measure of the exposed surface of a molecule to the solvent. It quantifies the extent of non-covalent interactions in a molecular complex and provides valuable insight into the rearranging mechanisms of side chains of a protein in an aqueous environment (Scheurer et al., 2018). A molecule with hydrophobic side chains lies on the surface tend to minimize the surface area, whereas the hydrophilic side chains access to increase the surface area (Iwase et al., 1985; Moret & Zebende, 2007). The estimation of SASA is useful to examine the contact area of two molecules and hence to assist in finding the binding affinity of two molecules. In this following subsection, SASA is used to estimate the contact surface area during the translation of AGT from DNA.

At the beginning of pulling of AGT from DNA, the contact surface area in both the systems was not altered significantly (pq and p'q' regions of graphs in Figure 38). After the simulation time lapsed ~13 ns, AGT in system-II was detected to be stretched away by detaching the intramolecular hydrogen bonds in site 3. The same event was detected after ~22 ns in system-I. The descending parts, qr and q'r' of graphs in Figure 38 of the corresponding systems show the decreasing the contact surface area. The intramolecular binding mechanism gives the definite shape in protein. During SMD, the force was applied on AGT that pulled away from the DNA. Countering with such pulling force, the intermolecular binding strength tends to retain the complex in original form. Thus, intramolecular bond dissociation took place due to a force acting on AGT from opposite directions: attraction of the binding region towards the DNA and SMD pulling

force applied in AGT along the opposite direction. The hydrogen bond dissociation in GLU172-TYR69, HSD171-PHE69, HSD171-PHE70 pairs and aromatic-aromatic interaction breaking in PHE70-TRP167 pair in site 3 consequence in stretching the AGT as shown in Figure 38. The loop stretching no longer continued so that almost a constant surface area was marked at regions rs (in system-I) and r's' (in system-II). During the simulation, we have detected that binding at both sites (site 1 and site 2) were stronger in system-I than that of system-II as shown in Figure 38. This is the main reason behind the sharply decreased contact surface area in system-II forming a large gap in the contact surface area between rs and r's' sections.

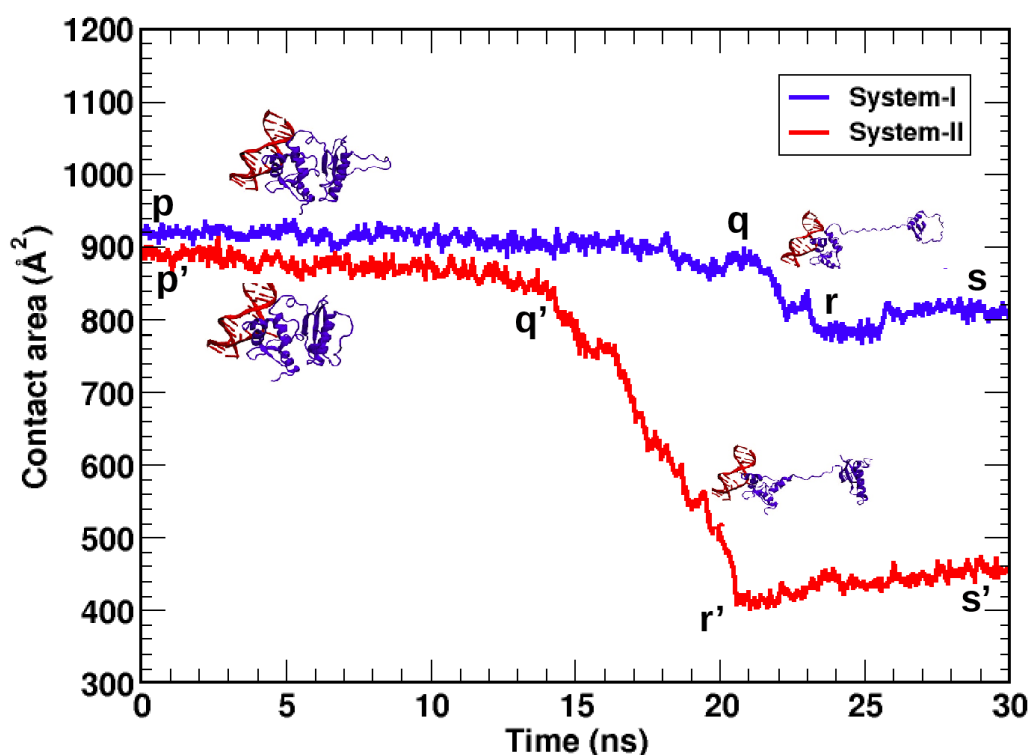


Figure 38: Contact surface area between DNA and AGT for system-I and system-II. pq and p'q' regions of the graph represent the stable binding, qr and q'r' show sharp down jump due to the intramolecular breaking region in AGT and the fast descending of the contact surface area. rs and r's' regions are constantly running parts that refer to the constant contact surface area between the molecules due to the strong binding in site 2 as shown in molecular structure accompanied with the plot.

4.8.3 Maximum Decoupling Force

Steered molecular dynamics (SMD) was also utilized to estimate maximum decoupling force during the separation of AGT and DNA. Initially, the center of mass of AGT lies at almost same position in both systems. During pulling, the AGT structure was gradually stretched away from DNA, however, the major interacting residues were attempted to restore the complex into the original form and kept nucleotide-residue pair partner in contact.

We performed five different simulations by replicating for each system. The outputs from each of the corresponding system are not significantly different as shown in Figures 39(a,b) and averaged the outputs and plotted as shown in Figure 39(c).

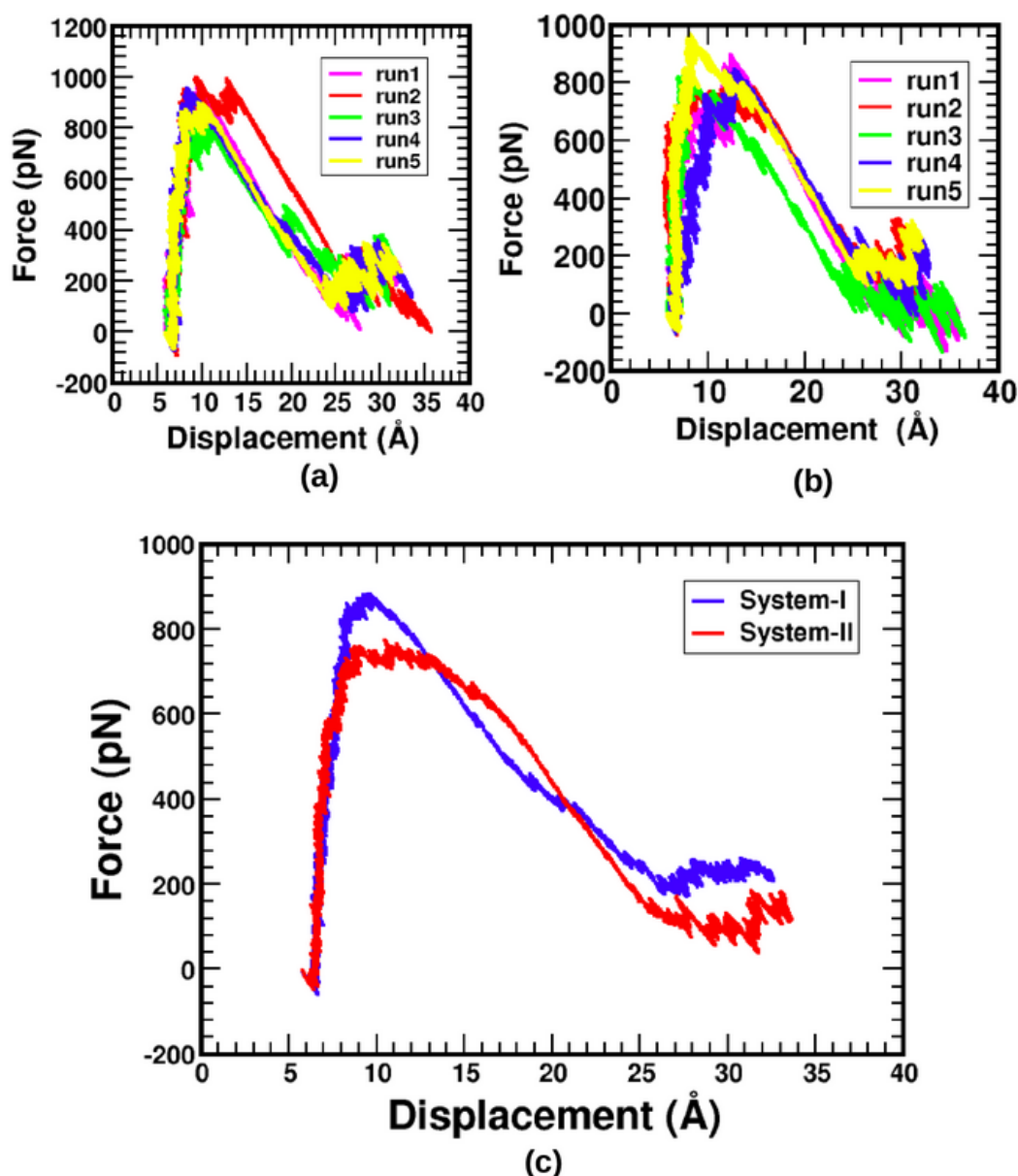


Figure 39: Decoupling force versus COM displacement of AGT by taking DNA as the reference molecule. Figures (a) and (b) are the force versus displacement curves for five identical simulations of system-I and system-II, respectively. (c) the average values of force versus displacement curve obtained from averaging (a) and (b) separately. The greater value of force in system-I than that of system-II shows more favorable binding of methylation damage DNA with AGT than the binding after methyl transfer.

Figure 39 shows the typical force-displacement profile for pulling AGT from the DNA of both systems. At the beginning, the force between AGT and DNA was linearly increased according to the Hooke's law of elasticity. The force was initially observed to be increasing instead of decreasing. After the displacement of ~ 4 Å from the initial position, the force started decreasing i.e., COM distance was 28 Å. In reality, although the

separation was increased, the major interacting residues were not found to be decoupled from their pair partners, rather they arrange to retain the molecules to form a more stable complex.

The decoupling force between the molecules in turn defines the binding force between them. Indeed, the SMD pulling of AGT from DNA can estimate the strongest energy barrier between them. To estimate the force, the data obtained from five independent simulations were used to find the average value of force in each system. The plots for all the replicated windows are shown in Figure 39. The averaged value shows that the strongest energy barrier for decoupling force was estimated to be ~ 900 pN in system-I and ~ 750 pN in the case of system-II as shown in Figure 39. Although the decoupling force is smaller in system-II, this force is sufficient to bind the molecules, and hence, the result shows that AGT does not leave the DNA, even if the interaction is degraded. Our results obtained from molecular dynamics simulations agree with the Atomic Force Microscopy (AFM) experiment of Tessmer & Fried (2014) and Density Functional Theory (DFT) results of Jena et al. (2009).

4.9 Estimation of Free Energy

Free energy is the energy which is available to do some specific work in biomolecular interactions as well as in the interaction with the surrounding aqueous environment. It is performed for the quantitative measurements of intermolecular interactions between ligand-protein, protein-protein, protein-DNA, etc. The quantitative determination of binding free energy provides information about the process of formation of molecular complexes at a given temperature and pressure. More importantly, the binding free energy of a drug candidate to its biomolecular target is extremely useful in drug designing. It also can estimate the bound states and unbound states between the molecules in a complex.

We performed the free energy calculation based on all-atom molecular dynamics simulations to study the possibility of methyl adduct transfer from methylated DNA to AGT. The methyl adduct was at the O6 point of the seventh residue guanine of double-stranded DNA. This type of methylation at guanine is a damage in DNA and can be the main cause of cancer. During the DNA replication and transcription, the methyl point fails to form the hydrogen bond with its pair partner cytosine so that three hydrogen bonds in the G:C pair is reduced to two hydrogen bonds, and hence, the nucleotide pair transition occurs, i.e., G:C pair to A:T pair (Kyrtopoulos et al., 1997). This transition transforms the cell into mutagenic and carcinogenic.

The DNA repair protein AGT is aimed to protect the living cells from methylation

damage. While AGT plays a crucial role in repairing the DNA damage and protect the normal cells from tumor development, it can resist the anticancer drugs to interact with the cancer cells. Given its importance as a drug target for overcoming anticancer drug resistance to improve chemotherapeutic efficacy, structural models of the AGT-DNA complex in different functional states of the methyl transfer process are desired for biophysical insights as well as for the silico screening.

To gain more insights into the DNA repair mechanism, we performed the umbrella sampling (US) (Torrie & Valleau, 1977; Koirala et al., 2021a) technique to estimate the free energy change taking the COM distance as a reaction coordinate. The umbrella sampling is a powerful technique for the computational estimation of free energy of binding between the molecules. There are many configuration states of binding between the molecules. If the molecular simulation is performed to cover all these states, it may take thousands of years to obtain the results because of high energy barriers encounters in the reaction pathway. To overcome such a problem, certain samples were taken which were supposed to represent all the energy barriers in phase space that come up during the interaction.

For the umbrella sampling, a biasing potential has been applied to produce oscillation of the molecules within a certain displacement. It is modeled by connecting the molecules with a virtual spring assigning certain force constant. The displacement of spring virtually mimics the oscillation displacement of molecules connecting at its two ends. The force constant (k) determines the displacement of molecules from their mean position.

The artificial harmonic potential $\omega_i(\xi)$ that is designed for the oscillation of molecules with a certain amplitude is calculated from,

$$\omega_i(\xi) = \frac{1}{2}k_i(\xi_i - \xi_i^{\text{ref}})^2 \quad (4.5)$$

where, ξ is the reaction coordinate along the path of reaction progress, and ξ_i is the center of reaction coordinate (mean position of harmonic motion) for i^{th} window.

The free energy $A_i(\xi)$ during the translation of molecule has been calculated by using the relation,

$$A_i(\xi) = -\frac{1}{\beta} \ln P_i^b(\xi) - \omega_i(\xi) + F_i \quad (4.6)$$

For the free energy calculation from umbrella sampling, many samples are chosen in such a way that the distribution of the molecules in a window should be overlapped to the neighboring windows, which ensures the covering of phase space along the direction of translation. This method is considered as one of the alternative reliable computational techniques of a laboratory experiment in the free energy calculation.

To identify the appropriate value of force constant k for the center to center distance of molecules, we have propagated short simulation runs taking different values of k for some representative samples taking fixed value of x_0 . Comparing patterns of distributions obtained from such simulations, we have taken the value, $k = 1.5 \text{ kcal}/(\text{mol}\cdot\text{\AA}^2)$ for all umbrella sampling simulation of both complexes.

In this method, some representative samples were chosen at different COM distances from AGT by translating DNA (Luzhkov, 2017). We prepared the windows for umbrella sampling by manually shifting the DNA taking protein as a reference molecule as shown in Figure 40. Every successive window was prepared by translating the double-stranded DNA (dsDNA) by 1 \AA distance along negative x-direction resulting in a total of 19 windows for both systems: methyl-adduct is at DNA (pre-methyl transfer condition) and after transferring the adduct to cysteine of AGT (post-methyl transfer condition).

The choice of an appropriate order parameter is important to interpret the results obtained from the simulation. Given the importance of the proximity of the methyl group from CYS145, the distance between S of CYS145 and O of GUA7 is a good choice. Since the CH_3 group is either with O or S depending on the pre-or post-transfer, S-C distance for the complex-I and C-O distance for complex-III could serve as a good reaction parameter, however, a direct comparison can be difficult since these are different complexes. Once the methyl group leaves the cavity, which occurs at a relatively short distance, a more global parameter such as COM distance between the protein and DNA may represent the complex better. We tracked the S-O distance for each window for the COM distance and they follow each other quite well. Additionally, since the comparison of the free energy of the complex is relative to the fully separated state where there is no appreciable interaction between the protein and DNA, we believe that COM as an order parameter works reasonably well.

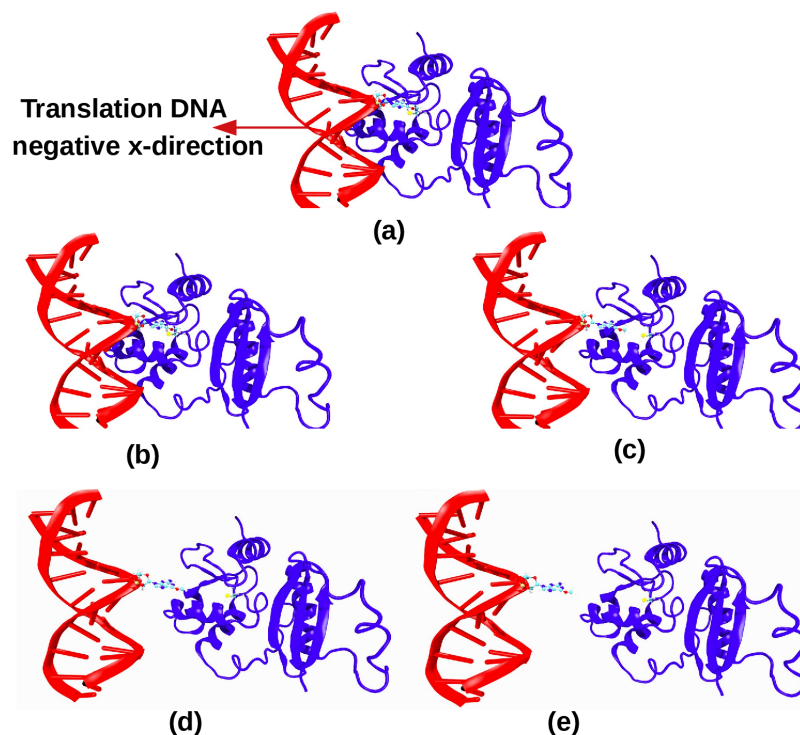


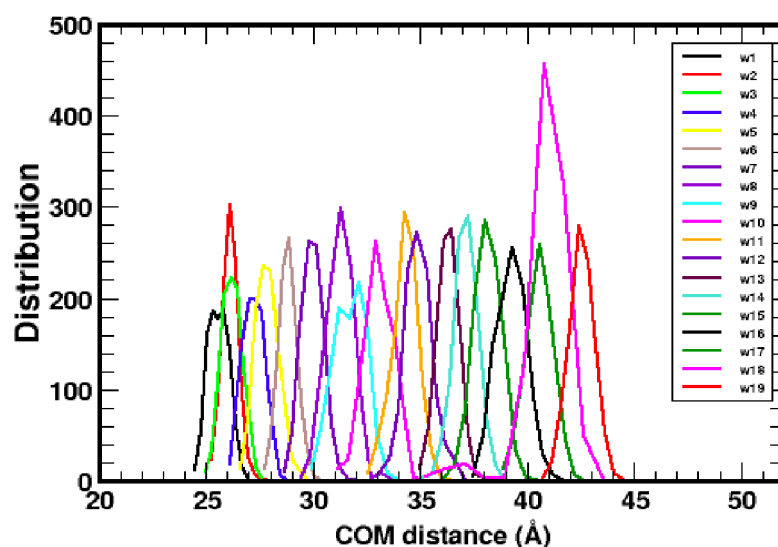
Figure 40: (a) Direction of DNA translation and representative snapshots of displacement of methylated DNA from AGT in the complex-I at COM distances of (b) 24.8 Å (c) 30.8 Å (d) 36.8 Å and (e) 42.8 Å.

In the original structure, DNA and AGT lied at COM distance ~ 24.8 Å binding at the minor groove of DNA. Then, the orientation of the complex was chosen such that the displacement was virtually based on the position of the intercalated ARG128 sidechain as the flipped GUA7 base and when displaced, they are minimally obstructed by other residues. The complex was positioned nearly parallel to the ARG128 sidechain and the flipped base and directly away from AGT. The complex was then re-oriented so that the direction of the displacement lies along the x-axis for convenience. Further, this system was employed to build the samples of 19 windows for each of complex-I and complex-III.

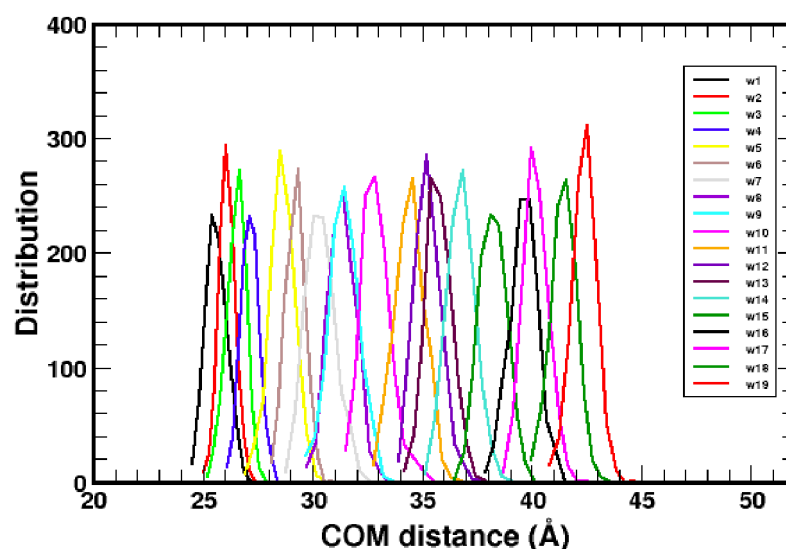
As the base flipping was observed to occur at a rate of $k = 350 \text{ s}^{-1}$ (in the order of millisecond), base flipping and associated conformation transitions upon AGT binding were not accessible in our simulation times cales. To resolve such complication, we used starting structure as the pre-catalytic complex which is formed the methylated GUA7 that is taken originally flipped out and approached into the AGT active site. Considering the reaction between DNA and AGT as S_N2 reaction, we did not take this state for the free energy calculation. Then, another complex was taken for the free energy estimation in which methyl is transferred to CYS145 of AGT. When methyl adduct is transferred to CYS145, the repaired DNA undergoes slow dissociation from AGT. It is also observed that the methylated AGT is bounded to DNA even at the post-methyl transfer condition and undergoes ubiquitination and degradation. This led us to investigate basically the

free energy calculation representing above two conditions.

The estimation of free energy can be estimated only when the distribution of molecules in a neighbouring sample windows overlap. To ensure the sufficient overlapping of neighboring windows, we plotted the distributions of molecules with respect to the COM displacement for both complexes. The distributions of molecules are shown in Figure 41(a) for complex-I and Figure 41(b) for complex-III. They show sufficient overlapping distribution so that the calculation of free energy was appropriate to be evaluated. The distributions are shown in different colour patterns.



(a)



(b)

Figure 41: Plots for distribution of DNA from AGT during the free energy estimation for all 19 windows of each of complex-I and complex-III, respectively. The distribution ensures the sufficient overlapping of windows and are suitable for the free energy calculation. (a) distributions in complex-I and (b) distributions in complex-III

To ensure the structural stability of each of all the 19 windows in both complexes, we estimated RMSD of the backbone atoms of each complex without taking hydrogen atoms. The RMSD plots of each sample are shown in Figure 42. In our simulation, only a slight increase in RMSD was observed in complex-III (post-transfer) compared to complex-I (pre-transfer) and the process of conformation changes resulting in the repaired base release is difficult to obtain in the simulation time scales. Here, the post-transfer separation of DNA-AGT was studied with the manual translation of DNA relative to AGT with an increment of 1 Å displacement along the negative x-direction.

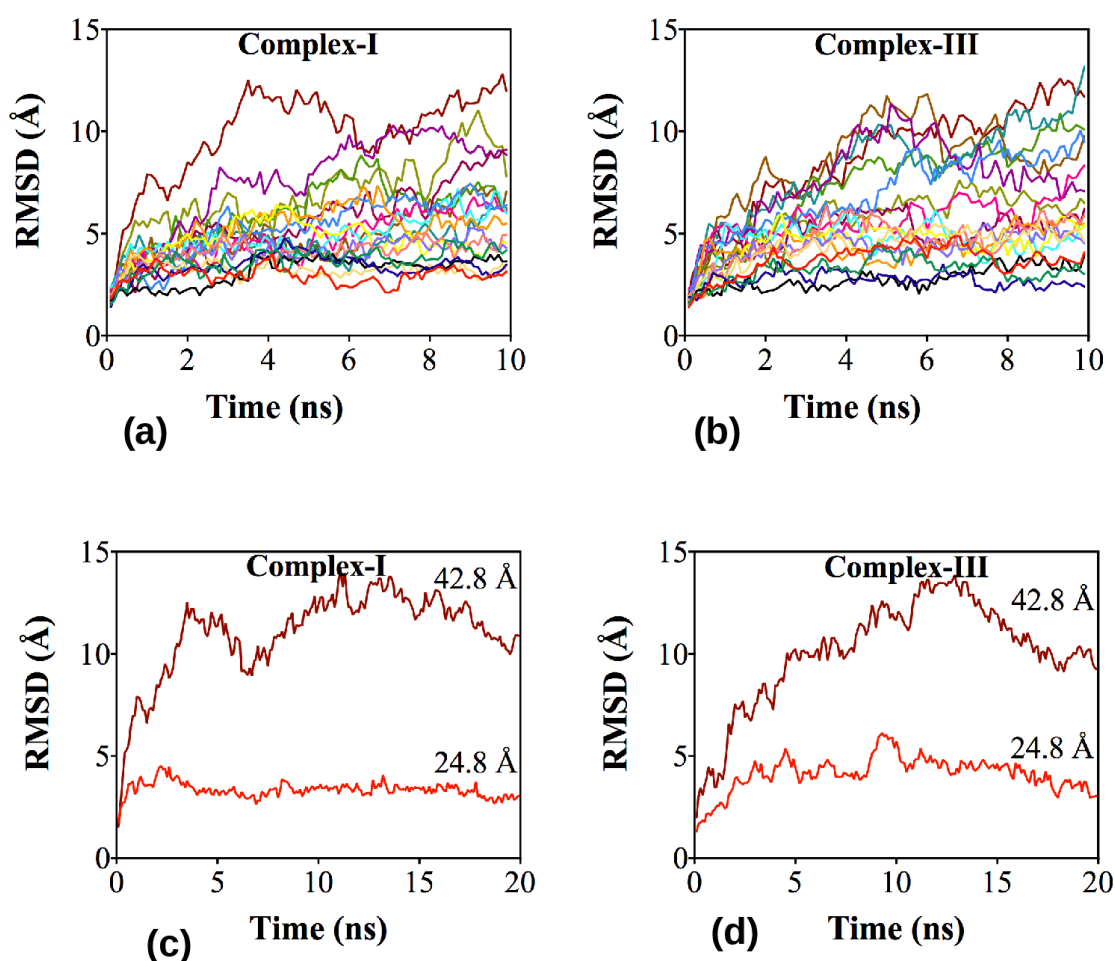


Figure 42: RMSD measurements for 19 windows obtained from the umbrella sampling for complex-I and complex-III. The RMSD measurements are fairly stable, especially for an essential interaction such as when AGT and DNA were in close contact.

Figure 42 shows the RMSD measurements from the MD simulation trajectories obtained for umbrella sampling. The upper panel (Figures a and b) shows the RMSD measurements for all windows of the systems set for umbrella sampling for the complex-I and complex-III. The RMSD measurements are fairly stable, especially for an essential interaction such as when AGT and DNA were in close contact (Koirala et al., 2021). We also extended few simulations up to 20 ns as shown in Figure 42(c,d) and no alteration

in the measurement data. We, therefore, believe that our free energy calculations from 10 ns simulation trajectories are sufficient for a good sampling.

As COM is a good order parameter for comparing the stability of the protein-DNA complex, we note that the distance between the sulfur atom of CYS145 and the oxygen atom of the GUA7 base (S-O distance) can play a significant role as a reaction coordinate. Figure 43 shows the change in free energy for approaching methylated GUA7 to the active site (complex-I) as well as for separating demethylated GUA7 from the active site (complex-III). To compare the free energies of the complexes with GUA7 at the active site, we aligned the curves when the DNA and AGT were fully separated, at which point the free energies for both systems should be comparable as the interaction between DNA and AGT was negligible for the COM distance $> 42 \text{ \AA}$, after 18th window. As shown in Figure 43, the change in free energy for the complex-I, between the complexes and separated states was found to be $\sim 11.1 \text{ kcal/mol}$ and that for the complex-III, was $\sim 9.8 \text{ kcal/mol}$. Comparison of the lowest free energy values at $\sim 26 \text{ \AA}$ COM distance, which represents the state in which the GUA7 or 6OG7 lies in the catalytic cavity, shows that the pre-transfer 6OG7 is more favorable in the cavity compared to the post-transfer GUA7, providing a free-energy advantage for 6OG7 of $\sim 1.3 \text{ kcal/mol}$ over GUA7. This may facilitate the dissociation of the GUA7 from the cavity once the methyl group is transferred to CYS145. Also, the shallower free energy curve beyond $\sim 28 \text{ \AA}$ for complex-III indicates easier dislodgement of the post-transfer GUA7. These observations explain the affinity of 6OG7 over GUA7 in the catalytic cavity, thereby giving an insight into the base demethylation mechanism of AGT (Koirala et al., 2021).

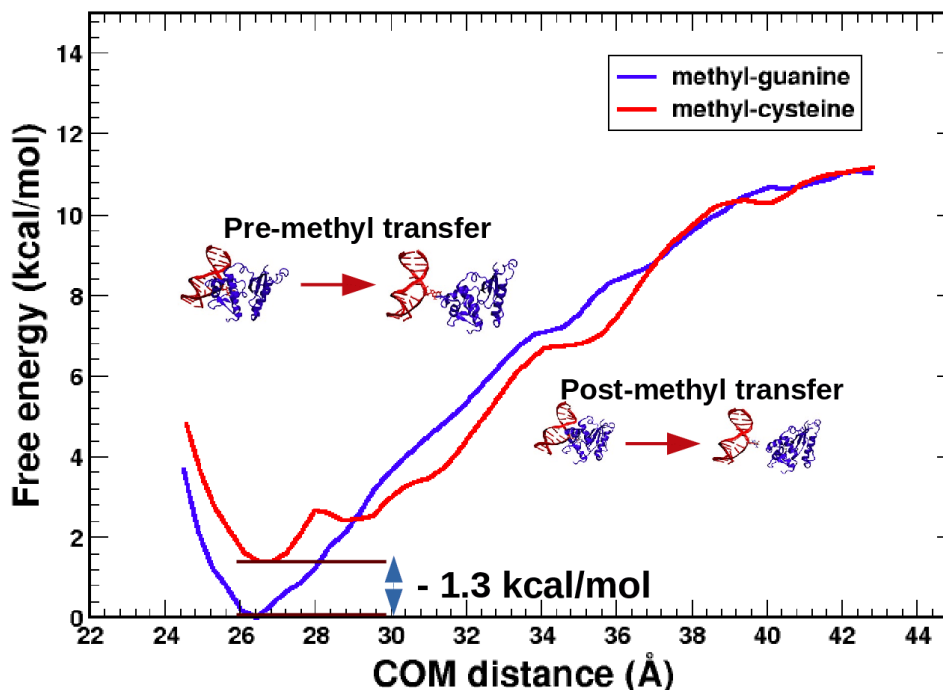


Figure 43: Plots for free energy estimation: Change in free energy for complex-I and complex-III taking COM distance as the reaction coordinates.

We chose 10 ns for each window (total simulation, 10 ns \times 19 windows \times 2 complexes). The estimated value may depend on duration of simulations. Long simulation is preferable in most of the cases, however it is not reasonable in some conditions. Therefore, convergence test for simulations at different time scales. Addressing these inconsistency of outcomes due to the simulation time scale, we examined the convergence test. For the convergence test of 10 ns simulations, we repeated the sampling with 20 ns for complex-I. The result obtained from the 20 ns simulation was not changed significantly as compared to the 10 ns sampling. This ensures that 10 ns simulations of umbrella sampling is sufficient to estimate the free energy. A plot comparing the results of 10 ns versus 20 ns sampling is shown in Figure 44.

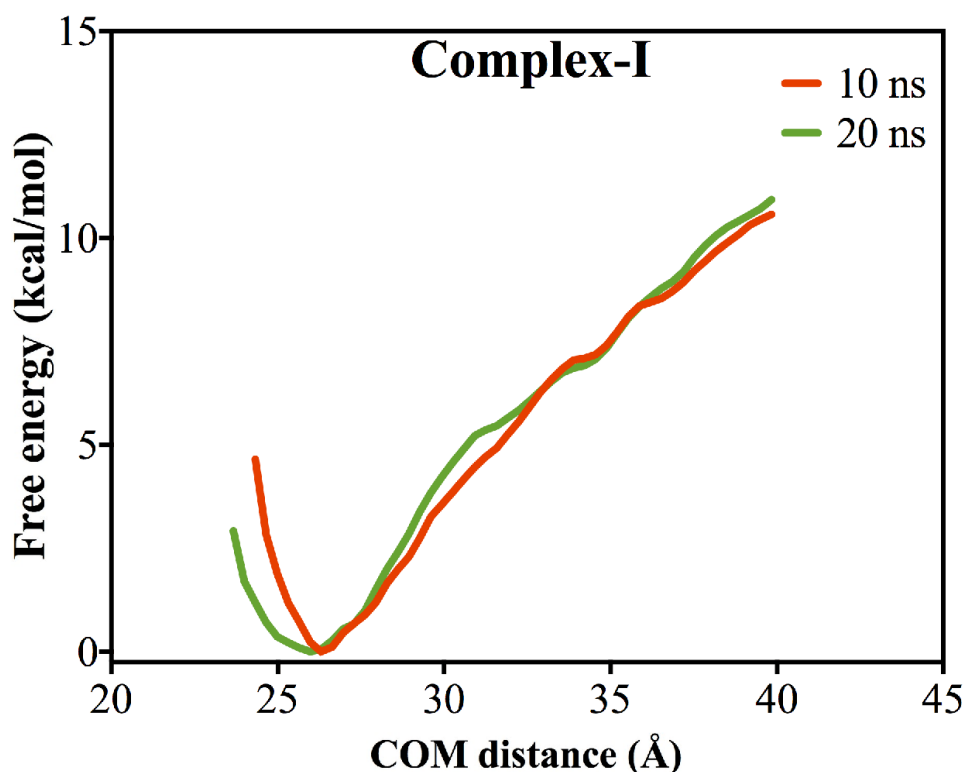


Figure 44: Free energy comparison in 10 ns and 20 ns for complex-I. The free energy change is almost the same showing convergence of 10 ns simulation run.

We compared our findings with the previously reported results. The difference of free energy in complex-I and complex-III agree with the outcomes of Hu et al., 2007. Similarly, our results about the stable binding during the formation of complex have been agreed with several previous works (Kaina et al., 2007; Rasimas et al., 2003; D. Fu et al., 2012).

4.9.1 Electrostatics and van der Waals Interactions during the Translation of DNA

Electrostatics and van der Waals energies were estimated by taking the prepared sample windows in umbrella sampling. The estimation of such non-bonded energies at different configuration states could provide valuable information about the possible distance at which association and dissociation of DNA and AGT. To determine the variation of such energies; electrostatics, van der Waals energies, and their sum were determined in each window as shown in Figure 45. Then, the average value of such non-bonded energies was determined and corresponding standard deviations were also calculated for both complex-I and complex-III. Then, the corresponding energy was plotted by considering COM distance as the reaction coordinate.

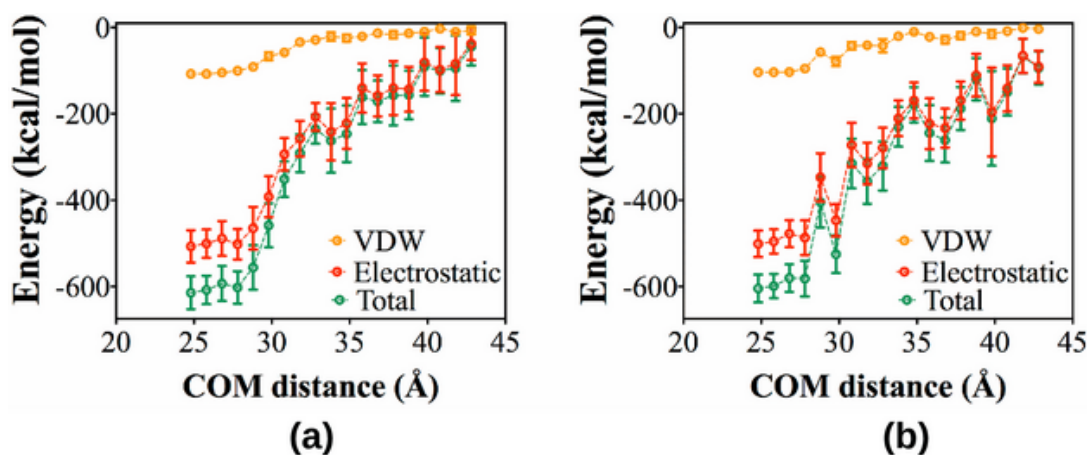


Figure 45: Variation of non-bonded interaction energies with the displacement of DNA from AGT; electrostatic interaction energy (E, red color), van der Waal interaction energy (VDW, orange color), and total non-bonded energy (E+VDW, green color) calculated by using NAMD energy plugin within VMD: (a) for complex-I and (b) for complex-III.

During the simulations, in the first few samples, (i.e., COM distance smaller than 32.8 Å), van der Waals interaction energy has a significant effect in binding the molecules, however, the interaction was almost negligible when COM distance surpassed 32.8 Å. Some fluctuations were observed in the non-bonded energy curve because of the rotation of DNA. As the COM distance between the molecules increases, DNA has got relatively free space for the rotation so that the residues in contact were found to be changing. Notably, ARG128 residue in AGT that originally intercalated at the backbone of DNA had a significant interaction until the COM separation was about 42.8 Å.

Up to the COM distance of about 32.8 Å, AGT interacts with methylated DNA at a minor groove. The nucleotides at the minor groove of DNA 6OG7 and THY9 strongly interact with different residues of the protein like TYR114, SER151, and CYS145. When the distance was gradually increased from 32.8 Å, they no longer bind tightly and protein leaves the minor groove of DNA. Hence, the DNA residues which previously interacted with protein residues got detached and new residues come into contact and play a crucial role in enhancing the non-bonded interaction.

The fluctuation in non-bonded energy was found to be greater in complex-III than in complex-I, which also corroborates the fact that the formation of the molecular complex is more favorable than of the dissociation. After the COM distance ~ 33 Å, van der Waals interaction tends to zero and electrostatic energy also has a negligibly smaller value with respect to the original separation of 24.8 Å. Moreover, the contribution of electrostatic energy has the dominating effect in comparison of such two non-bonded energy analyses.

The interaction energy due to the electrostatic and van der Waals contributions were compared for complex-I and complex-III. The overall result shows that the contribution

of both electrostatics and vdW energies are greater in complex-I than in complex-III as shown in Table 5 (Koirala et al., 2021). However, the energies are sometimes greater in complex-III due to the availability of more flexible molecules. In this complex, energy fluctuations are higher indicating that the system is relatively unstable than complex-I. The standard deviation (S.D.) at each window shows the relative fluctuation of the energy.

Table 5: Variation of electrostatics and van der Waals energy in complex-I and complex-III during the translation of DNA from AGT. Center of mass (COM) distance is the reaction coordinate. The first column contains the COM distance for every successive windows, second and third column are the electrostatics interaction energy in complex-I and complex-III. Fourth and fifth column are the van der waals interaction energy for corresponding complexes.

COM distance (Å)	Electrostatics interaction energy (kcal/mol)		van der Waals interaction energy (kcal/mol)	
	complex-I	complex-III	complex-I	complex-III
24.8	-506.95 ± 37.53	-511.12 ± 30.55	-107.44 ± 4.91	-103.52 ± 5.85
25.8	-500.59 ± 32.93	-495.50 ± 28.06	-107.45 ± 4.75	-103.37 ± 4.76
26.8	-488.81 ± 40.37	-477.53 ± 31.37	-104.28 ± 4.52	-103.00 ± 5.55
27.8	-501.90 ± 35.24	-486.62 ± 40.26	-100.20 ± 6.37	-94.99 ± 5.26
28.8	-464.67 ± 49.21	-347.27 ± 55.80	-91.33 ± 7.54	-57.04 ± 9.18
29.8	-391.80 ± 47.18	-446.44 ± 37.36	-66.47 ± 9.47	-78.56 ± 10.80
30.8	-293.88 ± 37.58	-272.56 ± 51.40	-57.74 ± 7.51	-42.48 ± 8.98
31.8	-257.34 ± 41.28	-315.21 ± 48.30	-33.89 ± 6.25	-41.16 ± 7.94
32.8	-207.10 ± 32.10	-279.00 ± 46.81	-28.56 ± 4.64	-41.85 ± 15.12
33.8	-241.63 ± 66.50	-210.33 ± 41.74	-20.94 ± 9.76	-20.03 ± 7.15
34.8	-222.33 ± 58.96	-169.00 ± 41.64	-24.22 ± 8.34	-9.97 ± 3.71
35.8	-140.33 ± 57.25	-223.14 ± 58.78	-20.78 ± 6.07	-21.42 ± 8.12
36.8	-158.25 ± 47.67	-233.07 ± 45.16	-12.98 ± 3.92	-27.94 ± 9.05
37.8	-140.55 ± 62.52	-168.93 ± 44.34	-16.64 ± 8.63	-18.95 ± 8.54
38.8	-147.07 ± 52.16	-110.38 ± 49.79	-13.35 ± 6.67	-9.48 ± 4.07
39.8	-80.83 ± 65.98	-195.89 ± 102.72	-9.75 ± 4.06	-14.95 ± 9.54
40.8	-97.54 ± 52.50	-141.00 ± 53.78	-2.78 ± 3.58	-8.31 ± 3.76
41.8	-84.19 ± 72.58	-65.37 ± 39.92	-10.03 ± 5.15	-0.81 ± 1.24
42.8	-37.55 ± 38.31	-90.50 ± 37.40	-6.60 ± 9.32	-3.44 ± 2.54

4.10 Intrahelical Flipping of GUA7 after Methyl Transfer

From the outcomes of umbrella sampling simulations of various samples, we have determined that DNA and AGT were almost set free from the mutual interaction after COM distance of 42.8 Å. In this condition, AGT was completely free from binding with the minor groove of DNA so that GUA7 could rotate freely across the base stack. In such condition, we examined whether the GUA7 would return within the backbone to form stable hydrogen bonds with ordinary nucleotide pair partner CYT20 or not. To study the nucleotide inward flipping mechanism, we have taken the nineteenth window of complex-III and analyzed the rotation activity of GUA7. The outcomes of MD simulation show the possibility of returning of flipped out guanine into its original configuration as shown in Figure 46(a).

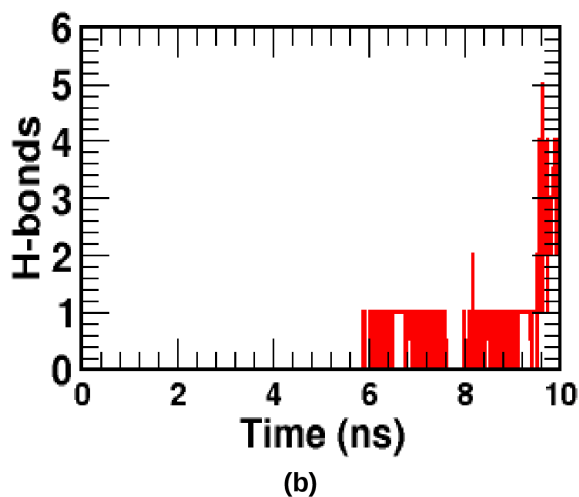
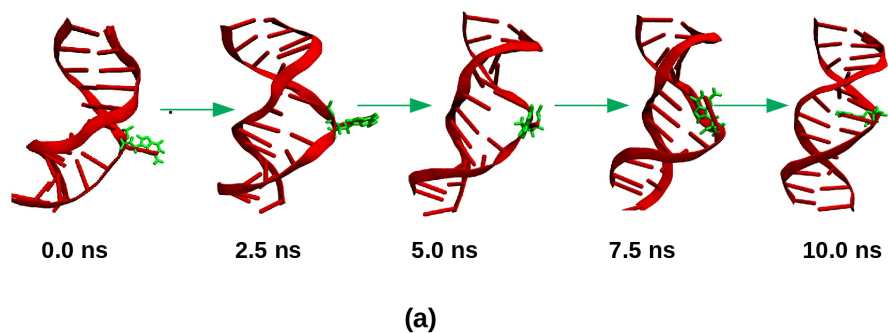


Figure 46: (a) Intrahelical returning of flipped guanine after the methyl adduct removal condition. Five snapshots are taken from the trajectories at 0.0 ns, 2.5 ns, 5.0 ns, 7.5 ns and 10.0 ns. (b) the hydrogen bond pattern between GUA7 and CYT20, hydrogen bond formation occurred only after the flipping back of GUA7 into intrahelical region.

In this molecular structure, GUA7 was originally taken flipping out from the DNA base stack. The MD simulation was performed for that molecular system taking a COM distance of 42.8 Å between DNA and AGT for 10 ns. During the simulation, DNA was observed relatively mobile than the AGT. The flipped guanine, GUA7, was not turned inward up to 5.0 ns time scale, just after 5 ns, it was attempted to recover into the base stack, the rotation is clearly recognized at 7.5 ns. Finally, The non-methylated GUA7 was interestingly returned into the backbone of the corresponding strand just before the completion of 10 ns simulation as shown in the fifth snapshot (Figure 46 (a)). It concludes that when the guanine is released from the interaction with the protein, flipped residue tends to recover to its normal structure. We further analyzed the hydrogen bonds between GUA7 and CYT20 during 10 ns simulation run. To examine the hydrogen bonds, we have taken the cutoff distance 3.5 Å and the angle 30°. In the beginning of the simulation, no hydrogen bond was detected, since these two bases were at distance almost 19 Å far as shown in region p of Figure 46 (b). After the lapse of time of 6 ns, hydrogen bonds came into play to connect the nucleotide pairs and finally came into the contact of all three atoms in the ordinary condition.

Figure 47(a) shows the distance between two targeted atoms, O6, N1, N2 of GUA7 and N4, O2, N3 of CYT20. The simulation was started at the condition of flipped out GUA7 so that the distance between these two atoms were about 19 Å apart. Before 5 ns simulation time scale, their distance was not significantly changed showing that there was no flipping of GUA7 within the base stack as shown in Figure 47(a). The sharp decrease of distance took place in between 5 ns to 6 ns. This is the event of rotation of GUA7 to come back into the backbone. This is the very important event for the recovery of DNA in order to form the ordinary base pairing of GUA7 and CYT20 as occurs in the ordinary condition. After 6 ns simulation time scale, the distance between O6 and N4 of corresponding nucleotides was detected almost the constant, which depicts that GUA7 and CYT20 came back to form the normal base pair. This mechanism is also analyzed in terms of distribution of GUA7 and CYT20 during the 10 ns. This evidence also verifies the pattern of distance variation at the same instant as shown in Figure 47(b).

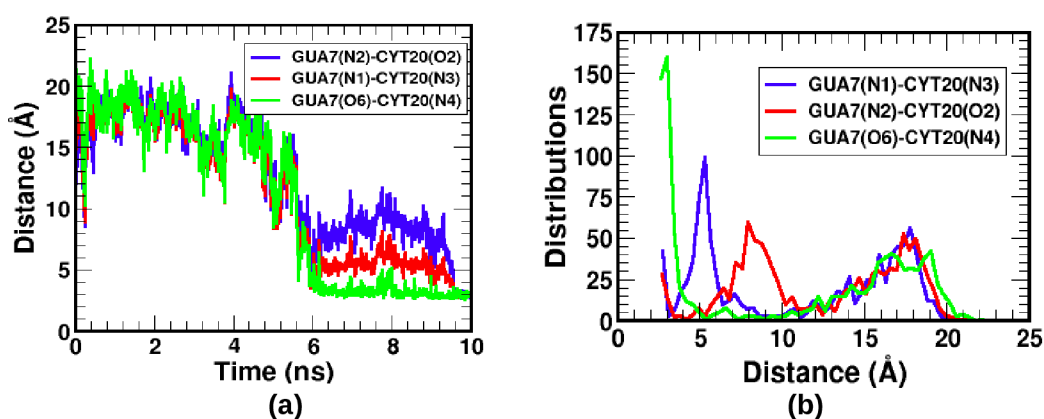


Figure 47: Intrahelical flipping analysis (a) distance between GUA7 and CYT20 (b) the distribution of GUA7 from CYT20, during the returning back of the distribution is significantly varying.

All of the above three different aspects of analysis concludes that the flip in of methyl repaired nucleotide searches its original conformation. The simulation procedure was also repeated for complex-I in the identical condition, however GUA7 was not flipped back into the base stack so that no hydrogen bond was formed even in entire simulation. This depicts that the flipped out guanine returns back into the backbone after removal of methyl adduct from the nucleotide base.

CHAPTER 5

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The confirmation of the DNA-AGT complex is a novel architecture of the biomolecular complex. Despite the usual binding of protein at the major groove in DNA, AGT binds at the minor groove of methylated DNA and serves as the methylation damage repair agent. The study of such a mechanism paves the way for understanding the complicated biological mechanism in our body. In this study, we explored the methyl transfer mechanism in DNA-protein complexes using the computational simulation method and expect to provide an insight so as to form a baseline for future avenues in the field.

This work was carried out in three basic steps: mechanism of complex formation in methylated DNA and AGT, intermediate transient state for methyl transfer from DNA to AGT, and the possible dissociation of methyl-AGT from repaired DNA. These three steps were studied by estimating several physical parameters like interaction energies, decoupling forces, contact surface area, the free energy of binding. Before the calculation of such mechanical and thermodynamic parameters, the molecular stability was examined via several structural parameters including root mean square deviation (RMSD), root mean square fluctuation (RMSF), contact surface area between DNA and AGT, etc. Likewise, several interaction methods between the molecules in the molecular complex such as hydrogen bonding, electrostatic and van der Waals contacts were studied by adopting MD trajectories to deal with these properties in the DNA-protein complexes. The free energy calculations, maximum possible decoupling force needed to separate DNA and AGT have been study to examine the thermodynamic properties of the DNA-AGT complex structure.

In order to study the systematic pathways during the methylation damage repair process in all three steps mentioned above, two different molecular structures were extracted from the protein data bank (PDB). Here, one complex contains a non-covalent bonding structure (PDB ID 1T38) whereas, other complex contains a covalently linked structure (PDB ID 1T39). The non-covalent structure was comprised of the DNA and AGT

molecules in which O6 point of guanine in DNA had been methylated and AGT was undamaged. Likewise, the covalently linked structure comprised methylated guanine at the O6 point and AGT in which the sidechain of sulfur at active site cysteine (CYS145) was linked to methyl adduct carbon at O6 point of guanine located at seventh residue in DNA. Furthermore, the non-covalent complex was modified to generate the third complex: the molecular complex after the methyl transfer condition. The molecular complexes for such three conditions were designated as complex-I (pre-methyl transfer condition or entry of AGT to methylated DNA), complex-II (transient intermediate state during the methyl transfer process), and complex-III (post-methyl transfer or the dissociation stage of methyl-AGT from repaired DNA).

The molecular dynamics (MD) simulations were carried out by using NAMD and most of the outcomes obtained from the simulation were compared between complex-I and complex-III so as to draw the conclusions for methylation damage repair mechanism. In the first step, each of the three complexes were subjected to 200 ns molecular dynamics simulations under NPT condition. Then, RMSD, RMSF, and total energies were evaluated to examine the structural stability of the complexes. In addition to these parameters, the structural stability was examined by calculating the distance between the main interacting atoms of DNA and AGT. Furthermore, the hydrogen bond consistency of these two molecules were explored. During the simulations, the RMSD of complex-I was found to be the most consistent complex indicating the most favorable structure among the three systems. Besides, the RMSD of complex-II and complex-III were slightly higher than complex-I, however, they lie under the condition of a stable structure.

The RMSF determined for alpha carbons of amino acid residues in AGT shows that the amino acids that lie at the interface of DNA are more stable than that of the residues lying at the remote side of DNA showing the strong binding affinity of interfacial residues during the complex formation. Moreover, the total energy of the entire system was found to be constant throughout the simulation in each complex. Likewise, the analysis of the distance between the specific atoms of highly interacting residues was almost constant which further verifies the stability of the complexes.

Hydrogen bonding analysis reveals that some residue-nucleotide pairs contribute to having a significant role in the formation of a stable complex. Out of many noticeable hydrogen bond pairs, TYR114-GUA7(6OG7) and SER151-THY9 had dominating effect with large hydrogen bond occupancy percentages forming consistent hydrogen bonds throughout the simulations. These two residue-nucleotide pairs may have a key role in binding the AGT with DNA at its minor groove. Besides these two pairs, the contributions of many other residue-nucleotide pairs were significant. As such, electrostatic and van der Waals interactions were found to have a significant contribution in the complex formation. The electrostatic and van der Waals interaction energies were estimated quan-

titatively for complex-I and complex-III and then, compared. The quantitative results show that electrostatic interaction energy plays a dominant role among these non-bonded interactions.

The 200 ns simulation run was further studied to investigate the replacement effect of ARG128 in place of GUA7 considering as the pair partner with CYT20. In this condition, ARG128 mimics being the nucleotide pair partner to the orphaned CYT20. The simulation outcomes showed that the NH₂ atom of ARG128 forms hydrogen bonds with N₃, N₄, and O₂ atoms of CYT20, and NH₂-N₃ atoms formed the most stable pair partner. Since ARG128 acts as the nucleotide base partner with CYT20, the 6OG7 gets free from the binding and tends to rotate across the backbone. This in turn facilitates the GUA7 to be flipped out with an extra-helical scheme.

The MD simulations were further carried out to inspect the necessary conditions for flipping out of methyl guanine (6OG7) from its base stack. To investigate the flipping condition, new systems were generated by only picking up the DNA system and removing the AGT from the complex. These two systems contain a non-methyl guanine system (system-I) and a methyl guanine system (system-II). Each system was subjected to the MD simulations for 100 ns under the NPT condition and the formation of hydrogen bonds between nucleotide partners CYT20-6OG7 (methyl guanine condition) and CYT20-GUA7 (non-methyl guanine condition) were examined. The outputs from hydrogen bond analysis clearly demonstrated the breaking of the ordinary hydrogen bond between N₄ and O₆ and the formation of new N₄-N₁ and N₃-N₂ hydrogen bond pairs by rejecting the O₂ from CYT20 in case of methyl guanine condition. However, all three possible hydrogen bonds were detected in the case of non-methyl guanine pair with CYT20. This bond-breaking due to the presence of methyl adduct at the O₆ point of guanine causes nucleotide base transition, G:C to A:T, during DNA replication and transcription.

Considering the contact surface area as a parameter in examining the stability of a complex, variation of the contact area of DNA and AGT was determined during the pulling of AGT from DNA. In such investigation, AGT was translated along positive x-direction. At the beginning of the simulation, both the systems had almost equal contact surface area. The contact area started to decrease gradually in both systems right after the beginning of the simulation, but the rate was higher in system-II than that of system-I. There was a down jump of surface area owing to the decoupling of some residue-nucleotide pairs. A relatively large gap was detected in system-II because of the breaking of hydrogen bonds in THR95-THY23 pairs, which was not been broken in system-I.

To investigate the decoupling mechanism between DNA and AGT, steered molecular dynamics (SMD) with constant velocity pulling was applied. Five identical runs were

propagated in each system and averaged over them. The maximum average decoupling force estimated from the SMD run shows that ~ 900 pN is required to decouple the molecules in system-I and ~ 750 pN is required for system-II, concluding the fact that there is a greater binding affinity in the pre-methyl transfer system than in the post-methyl system. In both systems, the major binding region of AGT is the minor groove of DNA.

The estimation of free energy in complex-I and complex-III is the measure of the binding affinity of the molecules in the corresponding complexes. On the other hand, this physical parameter estimates the possibility of methyl transfer from DNA to AGT. Umbrella sampling technique was utilized thereby forming 19 different samples for both complexes by displacing DNA along negative x-direction having a sample width of 1 \AA for each successive window. The similar short time simulation runs were carried out at different values of k taking the center of the mass (COM) distance being the reaction coordinates. The reaction coordinates were chosen in such a way that the distributions of molecules were sufficiently overlapped with the neighboring windows to ensure the representation of phase space along the translating direction.

The free energy curves obtained for complex-I and complex-III show that free energy change is greater in pre-methyl transfer condition (i.e., 11.1 kcal/mol) than in post-methyl transfer condition (i.e., 9.8 kcal/mol) with a net ~ 1.3 kcal/mol difference in the mutual difference of two systems. In addition to free energy estimation, RMSD for each window were determined which further verified the stability of both complexes during the simulation.

Finally, the extra-helical flipping of GUA7 was analyzed to examine if the complex returns back into the backbone after the completion of the methyl transfer process. For this experiment, the molecular complex was set up such that the center of mass (COM) distance between DNA and AGT is 42.8 \AA . Since AGT is beyond the interacting region, the extra-helical GUA7 is observed to be flexible and thus, can be easily rotated across the base stack. At the end of the 10 ns NPT simulation run, this nucleotide was observed to be turning back into the backbone and formed a hydrogen bond with its nucleotide pair partner CYT20. Our findings regarding the structural stability, decoupling force and free energy estimation were compared with the previously reported results. We found that our findings have the close agreement with these results.

5.2 Recommendations

This work paves the several pathways for the future works. Most importantly, we have designed force fields for modified structures in DNA and protein as well, which can be applied to new research work in the related fields. Newly designed force fields could

be useful for DNA alkylation condition and alkyl damage transfer process. Further, The force field designed for the covalent linkage in DNA-AGT complex provides the new perspective in the study of covalent binding between DNA and protein.

The investigation on methyl transfer mechanism can contribute in drug designing against cancer. The result of this work suggests that the methyl transfer mechanism is possible from DNA to AGT. The study on these DNA damage would be really support for the experiment examinations. Moreover, the work can be extended for the ethyl damage reversal. Further, the work can be extended to investigate several other damages in DNA caused by interstrand nucleotide mismatching, nucleotide deletion etc.

The complex-II is the intermediate state of methyl transfer mechanism. It is transient state which can remain very short time in the order of 100 ns and difficult to realize in practice. We assumed the sp^3 hybridization mechanism during the reaction of methylated DNA and AGT. This mechanism can be a method of investigation of reaction mechanism. In addition, the reaction can be the nucleophilic reaction likely proceeds via S_N2 mechanism and a regular covalent link may not be sufficient to study the reaction mechanism of transition state. To study the S_N2 mechanism is a challenging task, since pentavalent carbons should be realized and needed to parameterize. For this, we should involve a transient -S-CH₃-O- bonding with CH₃ in a planar geometry. However, it is a much bigger undertaking of analyzing the accurate transition state, with the force field parameterization of pentacoordinate carbon and we believe that our use of a more relatively straightforward parameters for regular covalent -S-CH₂-O- link is a close representation for our purpose as most of the work with the energy comparison is made only for the pre- and post- transfer processes. Realization of S_N2 reaction and parameterization of pentacarbon coordinates can be an important extension of this research work.

The molecular dynamics study of biomolecules is an emerging field of research in physics, chemistry and biology as well. However, there are several limitations regarding on this field. First of all, it is the theoretical work which should be carried out in wet laboratory to validate the result. Specifically, in this work, the interaction between DNA and protein is a challenging task in both molecular dynamics study and lab experiment. The outputs of MD simulations merely depend on the input information that we supplied to the software program. Force fields that designed for entire biomolecular system may shift from the actual result. The lab experiment is expensive and complicated. Moreover, the limitations of the programmings and system setup inherent in the output, so multiple simulations are required to confirm the validity of the result.

CHAPTER 6

6. SUMMARY

6.1 Summary

We have studied the molecular interaction between DNA and AGT (a DNA repair protein). The DNA comprises the methylation damage at O6-point of guanine base. Such type of methyl induced damage in DNA base is ubiquitous in cell nucleus and can be a major cause of cancer. O6-alkylguanine-DNA alkyl transferase (AGT) is considered as an important candidate for methylation damage reversal in DNA at O6 point of guanine base. Indeed, the atomic-level investigations in the interface region of the AGT and methylated DNA (mDNA) complex can provide an in-depth understanding of the binding mechanism between these molecules. This allows to evaluate the silico-drug nature of AGT and is useful in removing the methylation damage in DNA. Thus, our study encompasses the process of recognition, complex formation, and possible deformation of AGT and DNA after methyl repair.

We have performed the molecular dynamics (MD) simulations to investigate the molecular binding mechanism in three complexes: complex of mDNA and AGT, during the covalent linkage condition and the complex after methyl transfer from DNA to AGT. We performed the MD simulations utilizing the NAMD package. For this, we prepared three basic complexes representing the pre-methyl transfer condition (complex-I), intermediate transient state during the covalent bonding between DNA and AGT (complex-II) and post-methyl transfer condition (complex-III). All of three complexes are the modified structures. Therefore, we have designed required force fields (i.e., topology and parameters) for the modified structures.

We have investigated the key nucleotide-residue pairs in DNA and AGT at the interface region and examine the binding affinity of the molecules with various aspects. We have studied in detail the structural stability and major interfacial interactions, contact surface area, free energy changes and decoupling force of each complex. The structural stability test reveals that complex-I is the most stable. The total energy and hydrogen bond analysis for molecular stability also shows the favourable binding between DNA

and AGT in all complexes. Similarly, two basic regions of binding between DNA and AGT in which minor groove binding is most strong. Moreover, greater decoupling force in complex-I than that of complex-III, calculated from pulling AGT from methylated DNA, in turn estimates the maximum binding affinity in complex-I. The investigation on free energy changes reveals that the methyl transfer mechanism is possible from DNA to AGT. This mechanism is further verified by examining other physical parameters like rupture force calculation, contact area calculation etc.

All the outcomes from the simulations of three complexes are compared and thereby drawn the conclusions. In overall comparison complex-I represents the complex formation and complex-III represents the complex dissociation. We investigated the structural stability of intermediate state (complex-II). We believe that the use of parameters for the amino acid and nucleotide modifications and for the protein-DNA covalent bond will allow investigations of the DNA repair mechanism as well as the exploration of cancer therapeutics targeting the AGT-DNA complexes at various functional states as well as explorations via stabilization of the complex.

REFERENCES

- Acharya, K., Koirala, R. P., & Pantha, N. (2021). Diffusion of oxytocin in water: a molecular dynamics study. *BIBECHANA*, *18*(1), 108–117.
- Aftabuddin, M., & Kundu, S. (2007). Hydrophobic, hydrophilic, and charged amino acid networks within protein. *Biophysical Journal*, *93*(1), 225–231.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). DNA-binding motifs in gene regulatory proteins. In *Molecular biology of the cell. 4th edition*. Garland Science, New York, USA.
- Ali, R. B., Teo, A. K.-C., Oh, H.-K., Chuang, L. S.-H., Ayi, T.-C., & Li, B. F. (1998). Implication of localization of human DNA repair enzyme O 6-methylguanine-DNA methyltransferase at active transcription sites in transcription-repair coupling of the mutagenic O 6-methylguanine lesion. *Molecular and Cellular Biology*, *18*(3), 1660–1669.
- Allen, M. P., & Tildesley, D. J. (1991). *Computer simulation of liquids*. Oxford university press, New York, USA.
- Allen, M. P., & Tildesley, D. J. (2012). *Computer simulation in chemical physics* (Vol. 397). Springer Science & Business Media, Berlin, Germany.
- Ames, B. N. (1989). Mutagenesis and carcinogenesis: endogenous and exogenous factors. *Environmental and Molecular Mutagenesis*, *14*(S16), 66–77.
- Ames, B. N., Shigenaga, M. K., & Gold, L. S. (1993). DNA lesions, inducible DNA repair, and cell division: three key factors in mutagenesis and carcinogenesis. *Environmental Health Perspectives*, *101*(suppl 5), 35–44.
- Apel, K., & Hirt, H. (2004). Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annu. Rev. Plant Biol.*, *55*, 373–399.
- Arsham, M. S., Barch, M. J., & Lawce, H. J. (2017). *The AGT cytogenetics laboratory manual. 4th edition*. John Wiley & Sons, USA.
- Avery, O. T., MacLeod, C. M., & McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of

- transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *The Journal of Experimental Medicine*, 79(2), 137–158.
- Banani, S. F., Lee, H. O., Hyman, A. A., & Rosen, M. K. (2017). Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology*, 18(5), 285–298.
- Banavali, N. K., & MacKerell Jr, A. D. (2002). Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. *Journal of molecular biology*, 319(1), 141–160.
- Beranek, D. T. (1990). Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 231(1), 11–30.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., . . . Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, 28(1), 235–242.
- Berman, H. M., Westbrook, J., Feng, Z., Iype, L., Schneider, B., & Zardecki, C. (2002). The nucleic acid database. *Acta Crystallographica Section D: Biological Crystallography*, 58(6), 889–898.
- Best, R. B., Zhu, X., Shim, J., Lopes, P. E., Mittal, J., Feig, M., & MacKerell Jr, A. D. (2012). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles. *Journal of Chemical Theory and Computation*, 8(9), 3257–3273.
- Bigelow, C. C. (1967). On the average hydrophobicity of proteins and the relation between it and protein structure. *Journal of Theoretical Biology*, 16(2), 187–211.
- Blake, C., Geisow, M., Oatley, S., Rerat, B., & Rerat, C. (1978). Structure of prealbumin: secondary, tertiary and quaternary interactions determined by Fourier refinement at 1.8 Å. *Journal of Molecular Biology*, 121(3), 339–356.
- Bornot, A., Etchebest, C., & De Brevern, A. G. (2011). Predicting protein flexibility through the prediction of local structures. *Proteins: Structure, Function, and Bioinformatics*, 79(3), 839–852.
- Braun, E., Gilmer, J., Mayes, H. B., Mobley, D. L., Monroe, J. I., Prasad, S., & Zuckerman, D. M. (2019). Best practices for foundations in molecular simulations [Article v1. 0]. *Living Journal of Computational Molecular Science*, 1(1), 1–28.

- Bruner, S. D., Norman, D. P., & Verdine, G. L. (2000). Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature*, *403*(6772), 859–866.
- Bulaj, G. (2005). Formation of disulfide bonds in proteins and peptides. *Biotechnology Advances*, *23*(1), 87–92.
- Cadet, J., Sage, E., & Douki, T. (2005). Ultraviolet radiation-mediated damage to cellular DNA. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *571*(1-2), 3–17.
- Carugo, O., & Pongor, S. (2001). A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Science*, *10*(7), 1470–1473.
- Chandler, D. (1978). Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *The Journal of Chemical Physics*, *68*(6), 2959–2970.
- Chandler, D. (2005). Interfaces and the driving force of hydrophobic assembly. *Nature*, *437*(7059), 640–647.
- Chatterjee, N., & Walker, G. C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and Molecular Mutagenesis*, *58*(5), 235–263.
- Chen, L., Wang, R.-S., & Zhang, X.-S. (2009). *Biomolecular networks: methods and applications in systems biology* (Vol. 10). John Wiley & Sons, USA.
- Chou, P. Y., & Fasman, G. D. (1978). Empirical predictions of protein conformation. *Annual Review of Biochemistry*, *47*(1), 251–276.
- Chowdhury, N., & Bagchi, A. (2015). An Overview of DNA-Protein Interactions. *Current Chemical Biology*, *9*(2), 73–83.
- Christmann, M., Verbeek, B., Roos, W. P., & Kaina, B. (2011). O6-Methylguanine-DNA methyltransferase (MGMT) in normal tissues and tumors: enzyme activity, promoter methylation and immunohistochemistry. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, *1816*(2), 179–190.
- Cleaver, J. E., Lam, E. T., & Revet, I. (2009). Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. *Nature Reviews Genetics*, *10*(11), 756–768.
- Cleaves II, H. J., Scott, A. M., Hill, F. C., Leszczynski, J., Sahai, N., & Hazen, R. (2012). Mineral–organic interfacial processes: potential roles in the origins of life. *Chemical Society Reviews*, *41*(16), 5502–5525.
- Cotterill, R. (2003). *Biophysics: An Introduction*. John Wiley & Sons, England.

- Coulocheri, S. A., Pigis, D. G., Papavassiliou, K. A., & Papavassiliou, A. G. (2007). Hydrogen bonds in protein–DNA complexes: where geometry meets plasticity. *Biochimie*, 89(11), 1291–1303.
- Coulondre, C., & Miller, J. H. (1977). Genetic studies of the lac repressor: IV. Mutagenic specificity in the lacI gene of Escherichia coli. *Journal of molecular biology*, 117(3), 577–606.
- Curuksu, J., & Zacharias, M. (2009). Enhanced conformational sampling of nucleic acids by a new Hamiltonian replica exchange molecular dynamics approach. *The Journal of Chemical Physics*, 130(10), 03B610.
- Dadou, S. M., El-Barghouthi, M. I., Alabdallah, S. K., Badwan, A. A., Antonijevic, M. D., & Chowdhry, B. Z. (2017). Effect of protonation state and N-acetylation of chitosan on its interaction with xanthan gum: a molecular dynamics simulation study. *Marine drugs*, 15(10), 298.
- Dahm, R. (2005). Friedrich Miescher and the discovery of DNA. *Developmental Biology*, 278(2), 274–288.
- Daniels, D. S., Mol, C. D., Arvai, A. S., Kanugula, S., Pegg, A. E., & Tainer, J. A. (2000). Active and alkylated human AGT structures: a novel zinc site, inhibitor and extrahelical base binding. *The EMBO Journal*, 19(7), 1719–1730.
- Daniels, D. S., Woo, T. T., Luu, K. X., Noll, D. M., Clarke, N. D., Pegg, A. E., & Tainer, J. A. (2004). DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nature Structural & Molecular Biology*, 11(8), 714–720.
- Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12), 10089–10092.
- De Bont, R., & Van Larebeke, N. (2004). Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis*, 19(3), 169–185.
- DeLano, W. L. (2002). The PyMOL molecular graphics system. <http://www.pymol.org>.
- Dexheimer, T. S. (2013). *DNA repair pathways and mechanisms*. Springer, USA.
- Dey, B., Thukral, S., Krishnan, S., Chakrobarty, M., Gupta, S., Manghani, C., & Rani, V. (2012). DNA–protein interactions: methods for detection and analysis. *Molecular and Cellular Biochemistry*, 365(1), 279–299.
- Dill, K., & Bromberg, S. (2010). *Molecular driving forces: statistical thermodynamics in biology, chemistry, physics, and nanoscience. 2nd edition*. Garland Science, New York, USA.

- DiStasio, R. A., Gobre, V. V., & Tkatchenko, A. (2014). Many-body van der Waals interactions in molecules and condensed matter. *Journal of Physics: Condensed Matter*, *26*(21), 213202.
- Drabløs, F., Feyzi, E., Aas, P. A., Vaagbø, C. B., Kavli, B., Bratlie, M. S., . . . Krokan, H. E. (2004). Alkylation damage in DNA and RNA—Repair mechanisms and medical significance. *DNA Repair*, *3*(11), 1389–1407.
- Duguid, E. M., Mishina, Y., & He, C. (2003). How do DNA repair proteins locate potential base lesions? A chemical crosslinking method to investigate O6-alkylguanine-DNA alkyltransferases. *Chemistry & Biology*, *10*(9), 827–835.
- Dyson, H. J., & Wright, P. E. (1991). Defining solution conformations of small linear peptides. *Annual Review of Biophysics and Biophysical Chemistry*, *20*(1), 519–538.
- Dyson, H. J., & Wright, P. E. (1993). Peptide conformation and protein folding. *Current Opinion in Structural Biology*, *3*(1), 60–65.
- Ebert, J. C., & Altman, R. B. (2008). Robust recognition of zinc binding sites in proteins. *Protein Science*, *17*(1), 54–65.
- Emamjomeh, A., Choobineh, D., Hajieghrari, B., MahdiNezhad, N., & Khodavirdipour, A. (2019). DNA–protein interaction: identification, prediction and data analysis. *Molecular Biology Reports*, *46*(3), 3571–3596.
- Erickson, L. C., Laurent, G., Sharkey, N. A., & Kohn, K. W. (1980). DNA cross-linking and monoadduct repair in nitrosourea-treated human tumour cells. *Nature*, *288*(5792), 727–729.
- Essigmann, J., Croy, R., Nadzan, A., Busby, W., Reinhold, V., Büchi, G., & Wogan, G. (1977). Structural identification of the major DNA adduct formed by aflatoxin B1 in vitro. *Proceedings of the National Academy of Sciences*, *74*(5), 1870–1874.
- Esteller, M. (2007). Cancer epigenomics: DNA methylomes and histone-modification maps. *Nature Reviews Genetics*, *8*(4), 286–298.
- Falnes, P. Ø., Bjørås, M., Aas, P. A., Sundheim, O., & Seeberg, E. (2004). Substrate specificities of bacterial and human AlkB proteins. *Nucleic Acids Research*, *32*(11), 3456–3461.
- Feller, S. E., & MacKerell, A. D. (2000). An improved empirical potential energy function for molecular simulations of phospholipids. *The Journal of Physical Chemistry B*, *104*(31), 7510–7515.

- Fernández, I., Uggerud, E., & Frenking, G. (2007). Stable pentacoordinate carbocations: Structure and bonding. *Chemistry—A European Journal*, *13*(30), 8620–8626.
- Fleck, A., & Munro, H. (1966). The determination of nucleic acids. *Methods of biochemical analysis*, *14*, 113–176.
- Freeman, S. E., Blackett, A. D., Monteleone, D. C., Setlow, R. B., Sutherland, B. M., & Sutherland, J. C. (1986). Quantitation of radiation-, chemical-, or enzyme-induced single strand breaks in nonradioactive DNA by alkaline gel electrophoresis: application to pyrimidine dimers. *Analytical Biochemistry*, *158*(1), 119–129.
- Frenkel, D., & Smit, B. (2002). *Understanding molecular simulation: from algorithms to applications* (Vol. 1). Academic Press, USA.
- Friedberg, E. C., McDaniel, L. D., & Schultz, R. A. (2004). The role of endogenous and exogenous DNA damage and mutagenesis. *Current Opinion in Genetics & Development*, *14*(1), 5–10.
- Friedman, R., Boye, K., & Flatmark, K. (2013). Molecular modelling and simulations in cancer research. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, *1836*(1), 1–14.
- Fu, D., Calvo, J. A., & Samson, L. D. (2012). Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nature Reviews Cancer*, *12*(2), 104–120.
- Fu, Y., Bernasconi, L., & Liu, P. (2021). Ab Initio Molecular Dynamics Simulations of the SN1/SN2 Mechanistic Continuum in Glycosylation Reactions. *Journal of the American Chemical Society*, *143*(3), 1577–1589.
- Fuglebakk, E., Echave, J., & Reuter, N. (2012). Measuring and comparing structural fluctuation patterns in large protein datasets. *Bioinformatics*, *28*(19), 2431–2440.
- Gallo, M. T., Grant, B. J., Teodoro, M. L., Melton, J., Cieplak, P., Phillips Jr, G. N., & Stec, B. (2009). Novel procedure for thermal equilibration in molecular dynamics simulation. *Molecular Simulation*, *35*(5), 349–357.
- Gapsys, V., Khabiri, M., de Groot, B. L., & Freddolino, P. L. (2018). Comment on “Deficiencies in Molecular Dynamics Simulation-Based Prediction of Protein-DNA Binding Free Energy Landscapes”. *The Journal of Physical Chemistry B*, *124*(6), 1115–1123.
- Gavrieli, Y., Sherman, Y., & Ben-Sasson, S. A. (1992). Identification of programmed cell death in situ via specific labeling of nuclear DNA fragmentation. *Journal of Cell Biology*, *119*(3), 493–501.

- Gerson, S. L. (2002). Clinical relevance of MGMT in the treatment of cancer. *Journal of Clinical Oncology*, 20(9), 2388–2399.
- Gerson, S. L., Trey, J. E., Miller, K., & Berger, N. A. (1986). Comparison of O⁶-alkylguanine-DNA alkyltransferase activity based on cellular DNA content in human, rat and mouse tissues. *Carcinogenesis*, 7(5), 745–749.
- Grasso, G., Rebella, M., Muscat, S., Morbiducci, U., Tuszynski, J., Danani, A., & Deriu, M. A. (2018). Conformational dynamics and stability of U-shaped and S-shaped amyloid β assemblies. *International Journal of Molecular Sciences*, 19(2), 571.
- Grimme, S., Diedrich, C., & Korth, M. (2006). The Importance of Inter- and Intramolecular van der Waals Interactions in Organic Reactions: the Dimerization of Anthracene Revisited. *Angewandte Chemie*, 118(4), 641–645.
- Gruzdev, A., & Kishchenko, G. (1978). Fluorescence polarization of stretched polytene chromosomes stained with acridine orange. *Biophysics of Structure and Mechanism*, 4(2), 97–110.
- Gullingsrud, J. R., Braun, R., & Schulten, K. (1999). Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations. *Journal of Computational Physics*, 151(1), 190–211.
- Halliwell, B., & Aruoma, O. I. (1991). DNA damage by oxygen-derived species Its mechanism and measurement in mammalian systems. *FEBS letters*, 281(1-2), 9–19.
- Harris, L.-A., Williams, L. D., & Koudelka, G. B. (2014). Specific minor groove solvation is a crucial determinant of DNA binding site recognition. *Nucleic Acids Research*, 42(22), 14053–14059.
- Harteis, S., & Schneider, S. (2014). Making the bend: DNA tertiary structure and protein-DNA interactions. *International Journal of Molecular Sciences*, 15(7), 12335–12363.
- Hendrix, D. K., Brenner, S. E., & Holbrook, S. R. (2005). DNA structural motifs: building blocks of a modular biomolecule. *Quarterly Reviews of Biophysics*, 38(3), 221–243.
- Hovmöller, S., Zhou, T., & Ohlson, T. (2002). Conformations of amino acids in proteins. *Acta Crystallographica Section D: Biological Crystallography*, 58(5), 768–776.
- Hu, J., Ma, A., & Dinner, A. R. (2008). A two-step nucleotide-flipping mechanism enables kinetic discrimination of DNA lesions by AGT. *Proceedings of the National Academy of Sciences*, 105(12), 4615–4620.

- Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B. L., . . . MacKerell, A. D. (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods*, *14*(1), 71–73.
- Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of Molecular Graphics*, *14*(1), 33–38.
- Ishiguro, K., Zhu, Y.-L., Shyam, K., Penketh, P. G., Baumann, R. P., & Sartorelli, A. C. (2010). Quantitative relationship between guanine O6-alkyl lesions produced by Onrigin and tumor resistance by O6-alkylguanine-DNA alkyltransferase. *Biochemical Pharmacology*, *80*(9), 1317–1325.
- Isralewitz, B., Gao, M., & Schulten, K. (2001). Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology*, *11*(2), 224–230.
- Iwase, K., Komatsu, K., Hirono, S., Nakagawa, S., & Moriguchi, I. (1985). Estimation of hydrophobicity based on the solvent-accessible surface area of molecules. *Chemical and Pharmaceutical Bulletin*, *33*(5), 2114–2121.
- Jackson, S. P., & Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature*, *461*(7267), 1071–1078.
- Jaeckle, K. A., Eyre, H. J., Townsend, J. J., Schulman, S., Knudson, H. M., Belanich, M., . . . Schold, S. C. (1998). Correlation of tumor O6 methylguanine-DNA methyltransferase levels with survival of malignant astrocytoma patients treated with bis-chloroethylnitrosourea: a Southwest Oncology Group study. *Journal of Clinical Oncology*, *16*(10), 3310–3315.
- Jayaram, B., McConnell, K., Dixit, S. B., & Beveridge, D. (1999). Free energy analysis of protein–DNA binding: the EcoRI endonuclease–DNA complex. *Journal of Computational Physics*, *151*(1), 333–357.
- Jena, N. R., Shukla, P., Jena, H. S., Mishra, P., & Suhai, S. (2009). O6-methylguanine repair by O6-alkylguanine-DNA alkyltransferase. *The journal of Physical Chemistry B*, *113*(51), 16285–16290.
- Jones, S., Shanahan, H. P., Berman, H. M., & Thornton, J. M. (2003). Using electrostatic potentials to predict DNA-binding sites on DNA-binding proteins. *Nucleic Acids Research*, *31*(24), 7189–7198.
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, *79*(2), 926–935.

- Kaina, B., Christmann, M., Naumann, S., & Roos, W. P. (2007). MGMT: key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. *DNA Repair*, 6(8), 1079–1099.
- Kannam, S. K., Oehme, D. P., Doblin, M. S., Gidley, M. J., Bacic, A., & Downton, M. T. (2017). Hydrogen bonds and twist in cellulose microfibrils. *Carbohydrate polymers*, 175, 433–439.
- Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 9(9), 646–652.
- Karplus, M., & Petsko, G. A. (1990). Molecular dynamics simulations in biology. *Nature*, 347(6294), 631–639.
- Kästner, J. (2011). Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(6), 932–942.
- Kauffman, C., & Karypis, G. (2012). Computational tools for protein–DNA interactions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1), 14–28.
- Kelley, M. R., & Fishel, M. L. (2008). DNA repair proteins as molecular targets for cancer therapeutics. *Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents)*, 8(4), 417–425.
- Khanal, S. P., Koirala, R. P., Mishra, E., & Adhikari, N. P. (2021). Molecular dynamics study of structural properties of γ -aminobutyric acid (GABA). *BIBECHANA*, 18(1), 67–74.
- Klauda, J. B., Venable, R. M., Freites, J. A., O’Connor, J. W., Tobias, D. J., Mondragon-Ramirez, C., . . . Pastor, R. W. (2010). Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The Journal of Physical Chemistry B*, 114(23), 7830–7843.
- Koirala, R. P., Bhusal, H. P., Khanal, S. P., & Adhikari, N. P. (2020). Effect of temperature on transport properties of cysteine in water. *AIP Advances*, 10(2), 025122.
- Koirala, R. P., Khanal, S. P., Shiwakoti, S., & Adhikari, N. P. (2020a). Intermolecular Interaction of hTHYN1 Protein with Double Methylated DNA at 5m-Cytosine Nucleotide. *Journal of Institute of Science and Technology*, 25(1), 37–44.
- Koirala, R. P., Pokhrel, R., Baral, P., Tiwari, P. B., Chapagain, P. P., & Adhikari, N. P. (2021). Structural insights into the repair mechanism of AGT for methyl-induced DNA damage. *Biological Chemistry*.

- Koirala, R. P., Thapa, B., Khanal, S. P., Adhikari, R. P., & Adhikari, N. P. (2020b). Intra-molecular Conformational Stability in Human Growth Hormone. *Journal of Nepal Physical Society*, 6(2), 41–49.
- Koirala, R. P., Thapa, B., Khanal, S. P., Powrel, J., Adhikari, R. P., & Adhikari, N. P. (2021a). Binding of SARS-CoV-2/SARS-CoV spike protein with human ACE2 receptor. *Journal of Physics Communications*, 5(3), 035010.
- Krimm, S., & Bandekar, J. (1986). Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *Advances in Protein Chemistry*, 38, 181–364.
- Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., & Kollman, P. A. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, 13(8), 1011–1021.
- Kuzmanic, A., & Zagrovic, B. (2010). Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. *Biophysical Journal*, 98(5), 861–871.
- Kyrtopoulos, S. A., Anderson, L. M., Chhabra, S. K., Souliotis, V. L., Pletsa, V., Valavanis, C., & Georgiadis, P. (1997). DNA adducts and the mechanism of carcinogenesis and cytotoxicity of methylating agents of environmental and clinical significance. *Cancer Detection and Prevention*, 21(5), 391–405.
- Leach, A. R., & Leach, A. R. (2001). *Molecular modelling: principles and applications, 2nd edition*. Pearson education, UK.
- Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., . . . others (2016). CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *Journal of Chemical Theory and Computation*, 12(1), 405–413.
- Lemkul, J. A. (2020). Pairwise-additive and polarizable atomistic force fields for molecular dynamics simulations of proteins. *Progress in Molecular Biology and Translational Science*, 170, 1–71.
- Lemkul, J. A., Savelyev, A., & MacKerell Jr, A. D. (2014). Induced polarization influences the fundamental forces in DNA base flipping. *The Journal of Physical Chemistry Letters*, 5(12), 2077–2083.
- Lengyel, P., & Söll, D. (1969). Mechanism of protein biosynthesis. *Bacteriological Reviews*, 33(2), 264.

- Leontis, N. B., & Westhof, E. (2001). Geometric nomenclature and classification of RNA base pairs. *Rna*, 7(4), 499–512.
- Li, W., & Sancar, A. (2020). Methodologies for detecting environmentally induced DNA damage and repair. *Environmental and Molecular Mutagenesis*, 61(7), 664–679.
- Li, X., & Heyer, W.-D. (2008). Homologous recombination in DNA repair and DNA damage tolerance. *Cell Research*, 18(1), 99–113.
- Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*, 362(6422), 709–715.
- Lindahl, T., Demple, B., & Robins, P. (1982). Suicide inactivation of the E. coli O6-methylguanine-DNA methyltransferase. *The EMBO Journal*, 1(11), 1359–1363.
- Lindahl, T., & Wood, R. D. (1999). Quality control by DNA repair. *Science*, 286(5446), 1897–1905.
- Liu, L., & Gerson, S. L. (2006). Targeted modulation of MGMT: clinical implications. *Clinical Cancer Research*, 12(2), 328–331.
- Lowe, E. J., Lowe, J., et al. (1998). *The possibility of metaphysics: Substance, identity, and time* (Vol. 181). Clarendon Press Oxford, England.
- Lucas-Lenard, J., & Lipmann, F. (1971). Protein biosynthesis. *Annual Review of Biochemistry*, 40(1), 409–448.
- Lumry, R., & Eyring, H. (1954). Conformation changes of proteins. *The Journal of Physical Chemistry*, 58(2), 110–120.
- Luscombe, N. M., Austin, S. E., Berman, H. M., & Thornton, J. M. (2000). An overview of the structures of protein-DNA complexes. *Genome Biology*, 1(1), 1–37.
- Luzhkov, V. B. (2017). Molecular modelling and free-energy calculations of protein–ligand binding. *Russian Chemical Reviews*, 86(3), 211.
- MacKerell Jr, A. D., Bashford, D., Bellott, M., Dunbrack Jr, R. L., Evanseck, J. D., Field, M. J., . . . others (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of Physical Chemistry B*, 102(18), 3586–3616.
- Makov, G., & Payne, M. (1995). Periodic boundary conditions in ab initio calculations. *Physical Review B*, 51(7), 4014.

- Margison, G. P., Povey, A. C., Kaina, B., & Santibáñez Koref, M. F. (2003). Variability and regulation of O6-alkylguanine–DNA alkyltransferase. *Carcinogenesis*, *24*(4), 625–635.
- Marnett, L. J. (2000). Oxyradicals and DNA damage. *Carcinogenesis*, *21*(3), 361–370.
- Martínez, L. (2015). Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. *PloS One*, *10*(3), e0119264.
- Martínez, L., Polikarpov, I., & Skaf, M. S. (2008). Only subtle protein conformational adaptations are required for ligand binding to thyroid hormone receptors: simulations using a novel multipoint steered molecular dynamics approach. *The Journal of Physical Chemistry B*, *112*(34), 10741–10751.
- Matthew, J., & Ohlendorf, D. (1985). Electrostatic deformation of DNA by a DNA-binding protein. *Journal of Biological Chemistry*, *260*(10), 5860–5862.
- Mattossovich, R., Merlo, R., Miggiano, R., Valenti, A., & Perugino, G. (2020). O6-alkylguanine-DNA Alkyltransferases in Microbes Living on the Edge: From Stability to Applicability. *International Journal of Molecular Sciences*, *21*(8), 2878.
- McCulloch, S. D., & Kunkel, T. A. (2008). The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Research*, *18*(1), 148–161.
- McDonald, I., & Singer, K. (1967). Machine calculation of thermodynamic properties of a simple fluid at supercritical temperatures. *The Journal of Chemical Physics*, *47*(11), 4766–4772.
- McDonald, I., & Singer, K. (1969). Examination of the adequacy of the 12–6 potential for liquid argon by means of Monte Carlo calculations. *The Journal of Chemical Physics*, *50*(6), 2308–2315.
- McIlwraith, M. J., Vaisman, A., Liu, Y., Fanning, E., Woodgate, R., & West, S. C. (2005). Human DNA polymerase η promotes DNA synthesis from strand invasion intermediates of homologous recombination. *Molecular Cell*, *20*(5), 783–792.
- McKeague, M., Otto, C., Ráz, M. H., Angelov, T., & Sturla, S. J. (2018). The Base Pairing Partner Modulates Alkylguanine Alkyltransferase. *ACS Chemical Biology*, *13*(9), 2534–2541.

- Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., & Beckstein, O. (2011). MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, 32(10), 2319–2327.
- Miescher, F. (1869). Die Histochemischen und Physiologischen Arbeiten von Friedrich Miescher—Aus dem Wissenschaftlichen Briefwechsel von F. Miescher. *FCW Vogel (Leipzig)*, 1, 33–38.
- Miescher, F. (1897). *Die histochemischen und physiologischen Arbeiten von Friedrich Miescher* (Vol. 2). Vogel.
- Mikles, D. C., Bhat, V., Schuchardt, B. J., Deegan, B. J., Seldeen, K. L., McDonald, C. B., & Farooq, A. (2013). pH modulates the binding of early growth response protein 1 transcription factor to DNA. *The FEBS Journal*, 280(15), 3669–3684.
- Mitchell, D. L., & Clarkson, J. M. (1981). The development of a radioimmunoassay for the detection of photoproducts in mammalian cell DNA. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis*, 655(1), 54–60.
- Modrich, P. (2006). Mechanisms in eukaryotic mismatch repair. *The Journal of biological chemistry*, 281(41), 30305–30309.
- Moret, M., & Zebende, G. (2007). Amino acid hydrophobicity and accessible surface area. *Physical Review E*, 75(1), 011920.
- Moutevelis, E., & Woolfson, D. N. (2009). A periodic table of coiled-coil protein structures. *Journal of Molecular Biology*, 385(3), 726–732.
- Murugan, R. (2010). Theory of site-specific DNA-protein interactions in the presence of conformational fluctuations of DNA binding domains. *Biophysical Journal*, 99(2), 353–359.
- Myers, J. K., Pace, C. N., & Scholtz, J. M. (1997). Helix propensities are identical in proteins and peptides. *Biochemistry*, 36(36), 10923–10929.
- Neveu, E., Popov, P., Hoffmann, A., Migliosi, A., Besseron, X., Danoy, G., . . . Grudin, S. (2018). RapidRMSD: Rapid determination of RMSDs corresponding to motions of flexible molecules. *Bioinformatics*, 34(16), 2757–2765.
- Nimrod, G., Szilágyi, A., Leslie, C., & Ben-Tal, N. (2009). Identification of DNA-binding proteins using structural, electrostatic and evolutionary features. *Journal of Molecular Biology*, 387(4), 1040–1053.
- Noll, D. M., & Clarke, N. D. (2001). Covalent capture of a human O⁶-alkylguanine alkyltransferase–DNA complex using N¹, O⁶-ethanoxanthosine, a mechanism-based crosslinker. *Nucleic Acids Research*, 29(19), 4025–4034.

- Ostling, O., & Johanson, K. J. (1984). Microelectrophoretic study of radiation-induced DNA damages in individual mammalian cells. *Biochemical and Biophysical Research communications*, 123(1), 291–298.
- Pace, C. N., Fu, H., Lee Fryar, K., Landua, J., Trevino, S. R., Schell, D., . . . others (2014). Contribution of hydrogen bonds to protein stability. *Protein Science*, 23(5), 652–661.
- Passagne, I., Evrard, A., Winum, J.-Y., Depeille, P., Cuq, P., Montero, J.-L., . . . Vian, L. (2003). Cytotoxicity, DNA damage, and apoptosis induced by new fote-mustine analogs on human melanoma cells in relation to O6-methylguanine DNA-methyltransferase expression. *Journal of Pharmacology and Experimental Therapeutics*, 307(2), 816–823.
- Patra, N., Ioannidis, E. I., & Kulik, H. J. (2016). Computational Investigation of the Interplay of Substrate Positioning and Reactivity in Catechol O-Methyltransferase. *PloS One*, 11(8), e0161868.
- Pauling, L., & Corey, R. B. (1953). A proposed structure for the nucleic acids. *Proceedings of the National Academy of Sciences of the United States*, 39(2), 84.
- Pauling, L., Corey, R. B., & Branson, H. R. (1951). The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences*, 37(4), 205–211.
- Pegg, A. E. (1990). Mammalian O6-alkylguanine-DNA alkyltransferase: regulation and importance in response to alkylating carcinogenic and therapeutic agents. *Cancer research*, 50(19), 6119–6129.
- Pegg, A. E., Morimoto, K., & Dolan, M. E. (1988). Investigation of the specificity of O6-alkylguanine-DNA-alkyltransferase. *Chemico-biological Interactions*, 65(3), 275–281.
- Pfeifer, G. P., Drouin, R., Riggs, A. D., & Holmquist, G. P. (1991). In vivo mapping of a DNA adduct at nucleotide resolution: detection of pyrimidine (6-4) pyrimidone photoproducts by ligation-mediated polymerase chain reaction. *Proceedings of the National Academy of Sciences*, 88(4), 1374–1378.
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., . . . Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16), 1781–1802.
- Ramachandran, G. T., & Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Advances in Protein Chemistry*, 23, 283–437.

- Rapaport, D. C. (2004). *The art of molecular dynamics simulation. 2nd edition.* Cambridge university press, UK.
- Rasimas, J. J., Pegg, A. E., & Fried, M. G. (2003). DNA-binding Mechanism of O⁶-Alkylguanine-DNA Alkyltransferase: Effects of Protein and DNA Alkylation on Complex Stability. *Journal of Biological Chemistry*, 278(10), 7973–7980.
- Reuter, S., Gupta, S. C., Chaturvedi, M. M., & Aggarwal, B. B. (2010). Oxidative stress, inflammation, and cancer: how are they linked? *Free Radical Biology and Medicine*, 49(11), 1603–1616.
- Ringvoll, J., Nordstrand, L. M., Vågbø, C. B., Talstad, V., Reite, K., Aas, P. A., . . . others (2006). Repair deficient mice reveal mABH2 as the primary oxidative demethylase for repairing 1meA and 3meC lesions in DNA. *The EMBO Journal*, 25(10), 2189–2198.
- Robertson, K. D. (2005). DNA methylation and human disease. *Nature Reviews Genetics*, 6(8), 597–610.
- Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B., & Mann, R. S. (2010). Origins of specificity in protein-DNA recognition. *Annual Review of Biochemistry*, 79, 233–269.
- Roos, W., Baumgartner, M., & Kaina, B. (2004). Apoptosis triggered by DNA damage O⁶-methylguanine in human lymphocytes requires DNA replication and is mediated by p53 and Fas/CD95/Apo-1. *Oncogene*, 23(2), 359–367.
- Rosinski, J. A., & Atchley, W. R. (1999). Molecular evolution of helix–turn–helix proteins. *Journal of Molecular Evolution*, 49(3), 301–309.
- Rupert, C. S. (1960). Photoreactivation of transforming DNA by an enzyme from bakers' yeast. *The Journal of General Physiology*, 43(3), 573–595.
- Rydberg, B., & Lindahl, T. (1982). Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. *The EMBO Journal*, 1(2), 211–216.
- Sabharwal, A., & Middleton, M. R. (2006). Exploiting the role of O⁶-methylguanine-DNA-methyltransferase (MGMT) in cancer therapy. *Current Opinion in Pharmacology*, 6(4), 355–363.
- Sanger, F., & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, 94(3), 441–448.

- Scheurer, M., Rodenkirch, P., Siggel, M., Bernardi, R. C., Schulten, K., Tajkhorshid, E., & Rudack, T. (2018). PyContact: Rapid, customizable, and visual analysis of noncovalent interactions in MD simulations. *Biophysical Journal*, *114*(3), 577–583.
- Serdyuk, I. N., Zaccai, N. R., Zaccai, J., & Zaccai, G. (2017). *Methods in molecular biophysics. 2nd edition*. Cambridge University Press, UK.
- Shimada, J., Kaneko, H., & Takada, T. (1994). Performance of fast multipole methods for calculating electrostatic interactions in biomacromolecular simulations. *Journal of Computational Chemistry*, *15*(1), 28–43.
- Shuck, S. C., Short, E. A., & Turchi, J. J. (2008). Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology. *Cell Research*, *18*(1), 64–72.
- Singer, M. F. (1968). 1968 Nobel Laureate in Medicine or Physiology. *JSTOR*, *162*(3852), 433-436.
- Sinha, B. K. (1995). Topoisomerase inhibitors. *Drugs*, *49*(1), 11–19.
- Sinha, R. P., & Häder, D.-P. (2002). UV-induced DNA damage and repair: a review. *Photochemical & Photobiological Sciences*, *1*(4), 225–236.
- Soll, J. M., Sobol, R. W., & Mosammamparast, N. (2017). Regulation of DNA alkylation damage repair: lessons and therapeutic opportunities. *Trends in Biochemical Sciences*, *42*(3), 206–218.
- Souaille, M., & Roux, B. (2001). Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Computer Physics Communications*, *135*(1), 40–57.
- Srivenugopal, K. S., Yuan, X.-H., Friedman, H. S., & Ali-Osman, F. (1996). Ubiquitination-dependent proteolysis of O 6-methylguanine-DNA methyltransferase in human and murine tumor cells following inactivation with O6-benzylguanine or 1, 3-bis (2-chloroethyl)-1-nitrosourea. *Biochemistry*, *35*(4), 1328–1334.
- Steffen, N. R., Murphy, S., Toller, L., Hatfield, G. W., & Lathrop, R. H. (2002). DNA sequence and structure: direct and indirect recognition in protein-DNA binding. *Bioinformatics*, *18*(suppl_1), S22–S30.
- Sugita, Y., Kitao, A., & Okamoto, Y. (2000). Multidimensional replica-exchange method for free-energy calculations. *The Journal of Chemical Physics*, *113*(15), 6042–6051.

- Sun, P. D., Foster, C. E., & Boyington, J. C. (2004). Overview of protein structural and functional folds. *Current protocols in protein science*, 35(1), 17–1.
- Tessmer, I., & Fried, M. G. (2014). Insight into the cooperative DNA binding of the O6-alkylguanine DNA alkyltransferase. *DNA Repair*, 20, 14–22.
- Thakuria, R., Sarma, B., & Nangia, A. (2017). 7.03-Hydrogen Bonding in Molecular Crystals. *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering. Comprehensive Supramolecular Chemistry II*, 25–48.
- Thompson, A. P., Plimpton, S. J., & Mattson, W. (2009). General formulation of pressure and stress tensor for arbitrary many-body interaction potentials under periodic boundary conditions. *The Journal of Chemical Physics*, 131(15), 154107.
- Tiwari, P. B., Chapagain, P. P., Banda, S., Darici, Y., Üren, A., & Tse-Dinh, Y.-C. (2016). Characterization of molecular interactions between Escherichia coli RNA polymerase and topoisomerase I by molecular simulations. *FEBS Letters*, 590(17), 2844–2851.
- Tiwari, P. B., Chapagain, P. P., & Üren, A. (2018). Investigating molecular interactions between oxidized neuroglobin and cytochrome c. *Scientific Reports*, 8(1), 1–9.
- Torgovnick, A., & Schumacher, B. (2015). DNA repair mechanisms in cancer development and therapy. *Frontiers in Genetics*, 6, 1–15.
- Torrie, G. M., & Valleau, J. P. (1977). Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2), 187–199.
- Tubbs, J. L., Pegg, A. E., & Tainer, J. A. (2007). DNA binding, nucleotide flipping, and the helix-turn-helix motif in base repair by O6-alkylguanine-DNA alkyltransferase and its implications for cancer chemotherapy. *DNA Repair*, 6(8), 1100–1115.
- Tuckerman, M. (2010). *Statistical mechanics: theory and molecular simulation*. Oxford university press, England.
- Van Gunsteren, W. F., & Mark, A. E. (1992). On the interpretation of biochemical data by molecular dynamics computer simulation. *European Journal of Biochemistry*, 204(3), 947–961.
- Varghese, A., & Wang, S. Y. (1967). Cis-syn thymine homodimer from ultraviolet irradiated calf thymus DNA. *Nature*, 213(5079), 909–910.

- Vermeulen, W., De Boer, J., Citterio, E., Van Gool, A., Van der Horst, G., Jaspers, N., ... others (1997). Mammalian nucleotide excision repair and syndromes. *Biochemical Society Transactions*, 25(1), 309–314.
- Vickery, H. B., & Schmidt, C. L. (1931). The history of the discovery of the amino acids. *Chemical Reviews*, 9(2), 169–318.
- Virnau, P., & Müller, M. (2004). Calculation of free energy through successive umbrella sampling. *The Journal of Chemical Physics*, 120(23), 10925–10930.
- Wagner, I., & Musso, H. (1983). New naturally occurring amino acids. *Angewandte Chemie International Edition in English*, 22(11), 816–828.
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9), 1157–1174.
- Waterlow, J. C., Garlick, P. J., Mill Ward, D., et al. (1978). *Protein turnover in mammalian tissues and in the whole body*. Elsevier/North-Holland Biomedical Press.
- Watson, J. D. (2004). *Molecular biology of the gene. 5th edition*. Pearson Education, India.
- Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 171(4356), 737–738.
- Watson, J. D., & Crick, F. H. (2010). *1953. A structure for deoxyribose nucleic acid*. University of Chicago Press.
- Wilkins, M., & Randall, J. (1953). *Biochim et Biophysic. Acta*, 10, 192.
- Wood, R. D., Mitchell, M., Sgouros, J., & Lindahl, T. (2001). Human DNA repair genes. *Science*, 291(5507), 1284–1289.
- Wyatt, M. D., & Pittman, D. L. (2006). Methylating agents and DNA repair responses: Methylated bases and sources of strand breaks. *Chemical research in toxicology*, 19(12), 1580–1594.
- Yunta, M. (2017). It is important to compute intramolecular hydrogen bonding in drug design. *Am. J. Model. Optim*, 5(1), 24–57.
- Zak, P., Kleibl, K., & Laval, F. (1994). Repair of O6-methylguanine and O4-methylthymine by the human and rat O6-methylguanine-DNA methyltransferases. *Journal of Biological Chemistry*, 269(1), 730–733.

- Zang, H., Fang, Q., Pegg, A. E., & Guengerich, F. P. (2005). Kinetic analysis of steps in the repair of damaged DNA by human O6-alkylguanine-DNA alkyltransferase. *Journal of Biological Chemistry*, 280(35), 30873–30881.
- Zavala, A. G., Morris, R. T., Wyrick, J. J., & Smerdon, M. J. (2013). High-resolution characterization of CPD hotspot formation in human fibroblasts. *Nucleic Acids Research*, 42(2), 893–905.
- Zou, X., Ma, W., Solov'Yov, I. A., Chipot, C., & Schulten, K. (2012). Recognition of methylated DNA through methyl-CpG binding domain proteins. *Nucleic Acids Research*, 40(6), 2747–2758.
- Zuckerman, D. M. (2010). *Statistical physics of biomolecules: an introduction*. CRC Press, USA.

APPENDIX

A. Articles published in International Journals

- Koirala, R. P., & Adhikari, N. P. (Submitted 2021). Binding Mechanisms of AGT and Methyl Damage DNA at O6 Point of Guanine before and after Methyl-transfer. *Molecular Simulation*.
- Koirala, R. P., Bhusal, H. P., Khanal, S. P., & Adhikari, N. P. (2020). Effect of temperature on transport properties of cysteine in water. *AIP Advances*, 10(2), 025122.
- Koirala, R. P., Dawanse, S., & Pantha, N. (2021). Diffusion of Glucose in water: a molecular dynamics study. *Journal of Molecular Liquids*.
- Koirala, R. P., Pokhrel, R., Baral, P., Tiwari, P. B., Chapagain, P. P., & Adhikari, N. P. (2021). Structural insights into the repair mechanism of AGT for methyl-induced DNA damage. *Biological Chemistry*.
- Koirala, R. P., Thapa, B., Khanal, S. P., Powrel, J., Adhikari, R. P., & Adhikari, N. P. (2021a). Binding of SARS-CoV-2/SARS-CoV spike protein with human ACE2 receptor. *Journal of Physics Communications*, 5(3), 035010.

B. Articles published in National Journals

- Acharya, K., Koirala, R. P., & Pantha, N. (2021). Diffusion of oxytocin in water: a molecular dynamics study. *BIBECHANA*, 18(1), 108–117.
- Khanal, S. P., Koirala, R. P., Mishra, E., & Adhikari, N. P. (2021). Molecular dynamics study of structural properties of γ -aminobutyric acid (GABA). *BIBECHANA*, 18(1), 67–74.
- Khanal, S. P., Poudel, B., Koirala, R. P., & Adhikari, N. P. (2021). Solvation free energy of protonated lysine: Molecular dynamics study. *Journal of Nepal Physical Society*, 7(2), 69–75.
- Koirala, R. P., Khanal, S. P., & Adhikari, N. P. (2019). Transport properties of cysteine dimer in water. *Himalayan Physics*, 8, 11–18.
- Koirala, R. P., Khanal, S. P., Shiwakoti, S., & Adhikari, N. P. (2020a). Intermolecular Interaction of hTHYN1 Protein with Double Methylated DNA at 5m-Cytosine Nucleotide. *Journal of Institute of Science and Technology*, 25(1), 37–44.

Koirala, R. P., Thapa, B., Khanal, S. P., Adhikari, R. P., & Adhikari, N. P. (2020). Intra-molecular conformational stability in human growth hormone. *Journal of Nepal Physical Society*, 6(2), 41–49.

Koirala, R. P., & Adhikari, N. P. (Submitted 2021). Flipping Back of Extrahelical Guanine after Methyl Repair. *BIBECHANA*.

C. Participation in Conferences

- **Investigating AGT-DNA covalent and non-covalent interactions in methyl-induced DNA damage repair.**

Rajendra P. Koirala, Rudramani Pokhrel, Prabin Baral, Purushottam B. Tiwari, Prem P. Chapagain, and Narayan P. Adhikari. APS March meeting conference, March 15-19, 2021. Presenter: Rajendra P. Koirala

- **Repair Mechanism of Methyl Lesion in DNA Transferring Methyl Adduct to O6-alkylguanine-DNA alkyl transferase (AGT).**

Rajendra P. Koirala, Prem P. Chapagain, and Narayan P. Adhikari. ANPA conference, July 17-19, 2020. Presenter: Rajendra P. Koirala

- **Molecular Interaction of O6-methyl Guanine with DNA Repair Protein AGT.**

Rajendra P. Koirala, Prem P. Chapagain, and Narayan P. Adhikari. 1st NRN GLOBAL Knowledge Convention, July 17-19, 2018.
Presenter: Rajendra P. Koirala

- **Comparative Study of O6-methyl Guanine DNA and Normal DNA in Bonded and non-bonded Interaction.**

Rajendra P. Koirala, Prem P. Chapagain, and Narayan P. Adhikari. International Conference on Explorations in Physics (ICEP-2018), May 29-31, 2018.
Presenter: Rajendra P. Koirala

- **Free Energy Comparison before and after Methyl Transfer from Methyl-induced DNA to AGT.**

Rajendra P. Koirala, Prem P. Chapagain, and Narayan P. Adhikari. International Conference on Nanosciences and High Energy Physics (ICNHEP-2019), February 4-6, 2019. Presenter: Rajendra P. Koirala

Effect of temperature on transport properties of cysteine in water



Cite as: AIP Advances 10, 025122 (2020); doi: 10.1063/1.5132777

Submitted: 24 October 2019 • Accepted: 22 January 2020 •

Published Online: 12 February 2020



Rajendra Prasad Koirala,^{a)} Hem Prasad Bhusal,^{b)} Shyam P. Khanal,^{c)} and Narayan Prasad Adhikari^{d)}

AFFILIATIONS

Central Department of Physics, Tribhuvan University, Kirtipur, Kathmandu, Nepal

^{a)} rpkoirala@tucdp.edu.np

^{b)} hemprasadbhusal@gmail.com

^{c)} shyamkhanal1989@gmail.com

^{d)} Author to whom correspondence should be addressed: npadhikari@tucdp.edu.np

ABSTRACT

Molecular dynamics simulations have been performed to study the transport properties of the dilute solution of cysteine in water at different temperatures. Structural analysis of the system has been carried out using radial distribution functions between different atoms of the solvent and solute. The self-diffusion coefficients of the solute and solvent are estimated from the slope of the mean square displacement vs the time plot using Einstein's equation and their binary diffusion coefficients from Darken's relation. The temperature dependency of diffusion is demonstrated via Arrhenius plots. We have further extended our work to study the effects of the system size on diffusion and to calculate the viscosity coefficients of both the solvent and solution. The calculated values are in close agreement with the previously reported results available in literature.

© 2020 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/1.5132777>

INTRODUCTION

Cysteine, a non-essential amino acid, contains a sulfur atom in its side chain ($\text{CH}_2\text{-S-H}$). The covalent link between the cysteine molecules is of two types: ordinary peptide bond (CO-NH) and disulfide bond ($\text{R-S-S-R}'$). Although methionine, an amino acid, also contains a sulfur atom in its side chain, it does not form a disulfide bond, neither by itself nor with any other molecules. Thus, cysteine possesses identical behavior in the formation of a covalent link during polypeptide chain synthesis. The disulfide bond plays a significant role in protein folding, stability, and functional variation.^{1,2}

Cysteine is a white crystalline solid having a molar mass of 121.15 g/mol and a melting point of 513 K. Its solubility in water is 16 g per 100 ml at 288 K. However, cysteine exhibits a hydrophobic nature, due to which it generally resides in the interior of proteins. Cysteine is essential for the synthesis of highly anti-oxidative glutathione, which is important in the detoxification and protection of various tissues and organs in the body. Furthermore, cysteine contributes to the absorption of nutrients from the intestinal wall and in the metabolism of lipids. It enhances fertility and strengthens the

immune system, thus aiding prevention of dementia, multiple sclerosis, and Parkinson's disease.³ It is also recognized as an anti-aging amino acid. All these functions place cysteine in a special position that cannot be substituted by any other amino acid.⁴

The cysteine molecule as a residue in the protein chain plays a crucial role in the DNA-protein interaction. In the DNA methylation repair mechanism, cysteine works in the suicidal reaction in methyl-transfer from methyl-DNA to cysteine itself.⁵ Cysteine not only plays a role in biomolecular interactions, but also acts as the catalytic agent in the electroreduction process of metals such as bismuth and gold in the appropriate solvents.⁶ Therefore, the study of the mechanical properties of cysteine is necessary in biological and material sciences.

The term transport phenomenon means the process by which the mass, linear momentum, angular momentum, energy, and charge are transferred from one part of the system to another due to non-uniformity or inhomogeneity of the system. Diffusion, an important transport property, is the phenomenon in which mass is transferred as a result of random molecular motion. Various experimental techniques such as the peak-height method,⁷ nuclear magnetic resonance (NMR), and molecular dynamics (MD)

simulations⁸ have been performed to study the diffusion phenomenon of amino acids in water. These studies were mainly concerned about the effect of concentration, polarity, and temperature on the diffusion of amino acids.

Based on the consideration of chemical parallel of its sulfhydryl (R-S-H) and hydroxyl (O-H) groups in the side chain of other polar amino acids, the cysteine molecule appears due to the hydrophilic nature; however, free cysteine molecules are found in the hydrophobic region of proteins.^{9,10} Since the cysteine side chain is hydrophobic in nature, it tends to enhance the diffusion in water. However, the sulfur atom at the cysteine side chain tends to reduce the diffusion, as it has a relatively higher atomic mass than basic elements in organic compounds, such as carbon, nitrogen, and oxygen.¹¹⁻¹³ The study of the effect of the hydrophobic interaction for relatively heavy molecule on the diffusion coefficient would be exciting. Importantly, the thiol -SH site of the cysteine residue in antibodies is functionally active in the interaction with metals such as gold and in bio-sensing. The cysteine molecule, after breakage of the disulfide bond in the peptide chain, plays a very important role in functionalizing the gold surface and immobilizing the antibody. Although gold is an inert metal, it can be made chemically active by interacting with a peptide sequence, basically interacting with sulfur available in the thiol group -SH of the cysteine residue.¹⁴ Moreover, as the cysteine in combination with tryptophan can act as the strong link to bind refractory bio-receptors, it has wide applications in bio-sensing.¹⁵

Since amino acids are the building blocks of protein molecules, they have many similarities with more complex biomolecules such as antibiotics. Antibiotics are widely used in medicines and nutrients.¹⁶ Therefore, the measurement of diffusivity of amino acids is important in designing the drugs. Moreover, its efficiency of movement in solution can be quantitatively measured by determining the coefficient of viscosity in aqueous solutions.

Thus, a comprehensive study of the diffusion process and viscous property of amino acid molecule in water is essential to understand life processes and the physical mechanism of inorganic compounds. Many researchers have already studied the mechanical properties of some amino acids.^{6,17} To the best of our knowledge, the diffusion coefficient and coefficient of viscosity of the cysteine molecule in water using MD simulation has not yet been studied. Therefore, we intend to study these properties of cysteine.

DIFFUSION

Diffusion is a dynamic property of matter in which its particles are transported from the higher concentration region to the lower concentration region. It occurs due to the concentration inhomogeneity and thermal agitation of particles.¹⁸ Diffusion plays many important roles in non-living substances as well as in living organisms. The diffusion in a homogeneous system having no chemical concentration gradient is called self-diffusion, and the corresponding diffusion coefficient is termed self-diffusion coefficient.¹⁹ Einstein's equation is used to calculate the self-diffusion coefficients, which relates the diffusion coefficient with the mean square displacement (MSD) of the particles,^{20,21}

$$D = \lim_{t \rightarrow \infty} \frac{\langle [r(t) - r(0)]^2 \rangle}{6t}. \quad (1)$$

In Eq. (1), $r(t) - r(0)$ is the displacement of the particle from the reference point during the course of time t , $[r(t) - r(0)]^2$ is the square of displacement, and $\langle \dots \rangle$ represents the ensemble average, and hence, $\langle [r(t) - r(0)]^2 \rangle$ gives the MSD of the particle.

Binary diffusion is the diffusion of particles in the mixture of two different substances. It is the quantitatively measured diffusion coefficient using Darken's relation²²

$$D_{12} = N_2 D_1 + N_1 D_2. \quad (2)$$

In Eq. (2), D_{12} is the binary diffusion coefficient, D_1 and D_2 are the self-diffusion coefficients of substances 1 and 2, respectively, and N_1 and N_2 are the corresponding mole fractions.

In order to estimate the diffusion coefficients, simulations are carried out using periodic boundary conditions (PBC). Under PBC, the diffusion strongly depends on the size of the simulation box as suggested by Yeh and Hummer²³ due to the long range nature of hydrodynamics interaction. The effect of the system size on the diffusion coefficient (D_{PBC}) under periodic boundary conditions is accounted for by²³⁻²⁵

$$D_0 = D_{PBC} + \frac{2.84 k_B T}{6\pi\eta L},$$

where D_0 is the system-size independent value of the diffusion coefficient, D_{PBC} is the simulated value of the diffusion coefficient in the cubic box of size L under periodic boundary conditions (PBC), k_B is the Boltzmann constant, T is the absolute temperature of the system, and η is the shear viscosity of the solvent,

$$\text{or, } D_{PBC} = D_0 - \frac{2.84 k_B T}{6\pi\eta L}. \quad (3)$$

From the intercept and slope of Eq. (3), the values of D_0 and η are estimated.

COMPUTATIONAL DETAILS

Molecular dynamics simulations were performed for the system of 3 cysteine and 1039 water molecules in a cubic box of size 3.17 nm at five different temperatures; 288 K, 293 K, 303 K, 313 K, and 323 K. The Extended Simple Point Charge (SPC/E) water model and Optimized Potentials for Liquid Simulations-All Atom (OPLS-AA) force field parameters were used in the simulations. All the bonded and non-bonded interaction parameters are assigned in the OPLS-AA force field by default, and the parameters for SPC/E²⁶ water model are included in the file *spe.itp* inherent to GROMACS 5.1.1.²⁷ In addition, the same atom possesses different partial charges based on the group of attachment. The Coulomb interaction occurs due to the partial charge existing in the atoms/molecules. Likewise, the van der Waal's interaction occurs as a result of the induced dipole interaction.

The coordinates assigned for the molecules in the .pdb file may not be the equilibrium structures, rather they are initial guess from the electron probability density map produced by x-ray diffraction (XRD) or nuclear magnetic resonance (NMR). In addition, molecules may have been under steric hindrance, which may produce unnecessary strain in the system. In order to remove the effects and to bring the system at the minimum potential energy state, the energy minimization process was carried out using the Steepest-descent algorithm.²⁷

After energy minimization, the system is ready to study the dynamic properties. However, the dynamical properties, such as diffusion and viscosity, usually vary with the parameters such as temperature, pressure, and density.²⁸ Therefore, before starting the production run, these aforementioned parameters should be kept constant during simulation, and the system under study is to be brought in the state of thermodynamic equilibrium, which is known as equilibration of the system. To bring the system to the state of thermodynamic equilibrium, we performed the equilibration run for each system in NPT ensemble. The velocity rescaling thermostat with a temperature coupling time of 0.01 ps and the Berendsen barostat with a coupling time of 0.8 ps were used during each simulation.²⁷ The isothermal compressibility of $4.6 \times 10^{-5} \text{ bar}^{-1}$ was taken. The Linear Constraint Solver (LINCS) constraint algorithm was applied to convert all bonds into constraints during the equilibration run.²⁷ The cutoff distance of 1 nm was taken for both Coulomb and Lennard-Jones (LJ) interactions, and the long range Coulomb interaction was accounted for using the PME (Particle-mesh Ewald) method with a Fourier spacing of 0.12 nm. To start up the equilibration run, the velocity of the molecules should be assigned. Maxwell-Boltzmann distribution was used to generate the initial velocities for the particles in the system. To obtain new positions and velocities of the particles after each time step, the leapfrog algorithm²⁷ was chosen. Each equilibration run was performed for 50 ns with a 1 fs time step.

Consequently, the production run of each system was performed to calculate the transport properties of the system in the NVT ensemble for 50 ns with a time step of 1 fs taking velocity-rescaling thermostat with a coupling time of 0.01 ps. Furthermore, it is not required to generate the initial velocities in the NVT run as the simulation continues with the velocities generated in the equilibration run.

RESULTS AND DISCUSSION

In this section, we present structural analysis and transport properties of the system at different temperatures.

STRUCTURE OF THE SYSTEM

The radial distribution functions (RDF) between the pair of atoms are used to analyze the structural properties of the system. For this, we have plotted the RDF between oxygen atoms of water molecules [$g_{\text{OW-OW}}(r)$] and carboxyl oxygen of cysteine and oxygen of water [$g_{\text{OC-OW}}(r)$] at five different temperatures.

Figure 1 shows the RDF between oxygen atoms of water molecules at the following temperatures: 288 K, 293 K, 303 K, 313 K, and 323 K.

The values of the excluded region (ER), first peak position (FPP), first peak value (FPV), second peak position (SPP), second peak value (SPV), third peak position (TPP), and third peak value (TPV) of RDF $g_{\text{OW-OW}}(r)$ are presented in Table I.

In Fig. 1, there are three distinct peaks. The first peak, which is located at the separation of about 0.27 nm from centered atom's position, is the highest and sharpest. This implies that, at this position, the maximum number of oxygen atoms is clustered from the reference oxygen atom. In other words, the probability of finding oxygen atoms at the first peak position is the highest. This is the

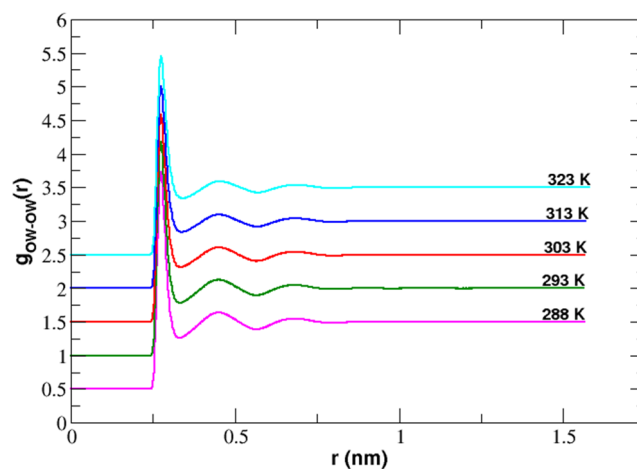


FIG. 1. RDF plot of oxygen-oxygen of water molecules, $g_{\text{OW-OW}}(r)$, at different temperatures.

most preferable position or minimum energy position from the centered atom. The value of the Lennard-Jones parameter σ of oxygen in water is 0.3166 nm and the van der Waal's radius is $(2\frac{1}{2}\sigma) \approx 0.36$ nm. However, the FPP in our system is 0.27 nm less than 0.36 nm. This reveals the fact that there is not only the LJ interaction between oxygen atoms of water, but also other interactions such as Coulomb and bonded interactions are present.

The second and third peaks are relatively shorter and wider, which are located approximately at positions 0.45 nm and 0.68 nm, respectively. The excluded region, in which RDF is zero, has extended up to 0.24 nm from the center of the reference oxygen atom. Any other oxygen atom cannot exist within the excluded region due to strong repulsive forces, namely, the r^{-12} term of the LJ interaction and repulsive Coulomb interactions.²⁹ We have also studied the effect of temperature on RDF. With the increase in temperature, the peak positions are shifted to right, heights of the peaks are decreased, and widths are increased (see Table I). This reflects that our system has become less organized with the increase in temperature. The increase in thermal agitation of atoms in the system with rising temperature accounts for this fact. Furthermore, beyond the third peak graph is the straight line possessing the unit value on

TABLE I. Details of RDF of oxygen-oxygen atoms of water molecules at different temperatures.

T (K)	RDF analysis of OW-OW atoms						
	ER (nm)	FPP (nm)	FPV	SPP (nm)	SPV	TPP (nm)	TPV
288	0.240	0.274	3.230	0.450	1.140	0.682	1.049
293	0.240	0.274	3.175	0.450	1.129	0.680	1.048
303	0.240	0.274	3.077	0.450	1.110	0.686	1.043
313	0.240	0.276	3.000	0.450	1.096	0.690	1.041
323	0.240	0.276	2.945	0.450	1.091	0.686	1.037

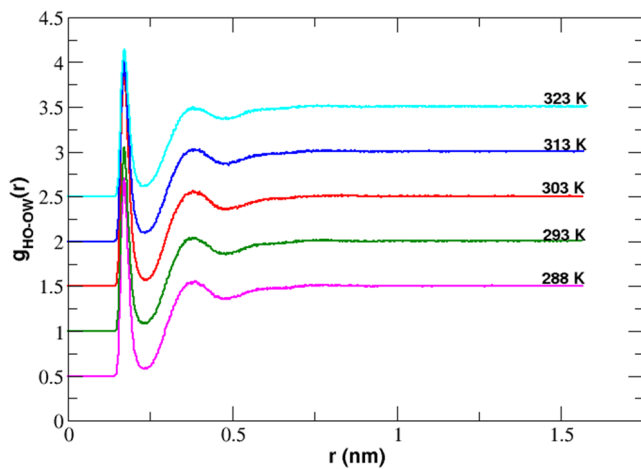


FIG. 2. RDF plot of carbonyl oxygen of cysteine and oxygen of water, $g_{OC-OW}(r)$, at different temperatures.

average. This indicates that there is no pair correlation of the oxygen atoms.

The RDF $g_{OC-OW}(r)$ gives the insight into how the carbonyl oxygen atoms of cysteine are organized around the oxygen atom of water. Figure 2 represents the RDF $g_{OC-OW}(r)$ at the above-mentioned temperatures. From the figure, it is clearly seen that the RDF has two noticeable peaks. The values of ER, FPP, FPV, SPP, and SPV are tabulated in Table II.

The first peak, which is located at the separation of about 0.33 nm from the position of the reference oxygen atom of water, is the highest and sharpest. This implies that, at this position, maximum number of carbonyl oxygen atoms of cysteine clustered from the reference oxygen atom. Therefore, this is the most preferred position of carbonyl oxygen atoms to cluster around the oxygen atom of water. The second peak is relatively shorter and wider, which is located approximately at the position of 0.58 nm. The excluded region extends up to 0.24 nm from the center of the reference oxygen atom. It is not possible to find any carbonyl oxygen within the excluded region due to strong repulsive forces. Beyond the second peak, there is no pair correlation of carbonyl oxygen atoms and the reference oxygen atom of water.

TABLE II. Details of RDF of carbonyl oxygen of cysteine and oxygen of water at different temperatures.

RDF analysis of OC–OW atoms					
T (K)	ER (nm)	FPP (nm)	FPV	SPP (nm)	SPV
288	0.240	0.336	1.125	0.584	1.018
293	0.242	0.338	1.094	0.590	1.020
303	0.242	0.336	1.185	0.578	1.031
313	0.242	0.336	1.092	0.580	1.026
323	0.242	0.336	0.985	0.586	1.003

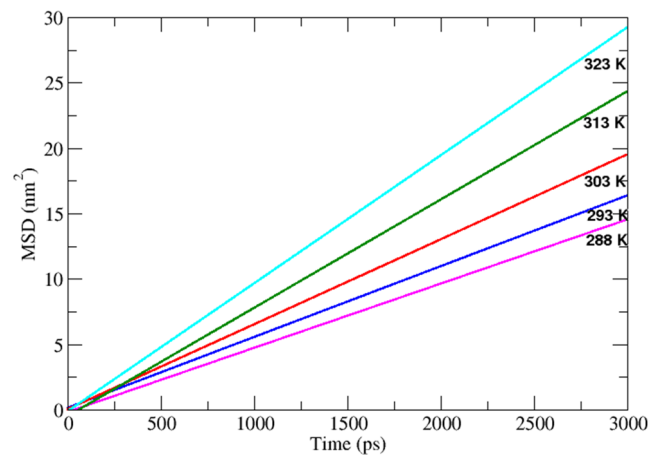


FIG. 3. MSD vs time plot of cysteine at different temperatures.

DIFFUSION COEFFICIENTS

The self-diffusion coefficients of cysteine and water have been calculated for five different temperatures by using the corresponding MSD curves. We have determined the self-diffusion coefficient from the slope of the MSD plot according to Einstein's equation (1).

We have plotted the MSD curves for 3 ns for all temperatures, even though the production run was done for 50 ns as statistics is better at the beginning than the ending region of the plot. Figures 3 and 4 show the MSD vs time plot for cysteine and water at 288 K, 293 K, 303 K, 313 K, and 323 K temperatures. The study has shown that as the temperature increases, the slope of the MSD curves also increases, which in turn increases the self-diffusion coefficient. The estimated values of self-diffusion coefficients of cysteine and water and their binary diffusion coefficients are presented in Table III with previously reported experimental values. Table III demonstrates that the simulated values of the

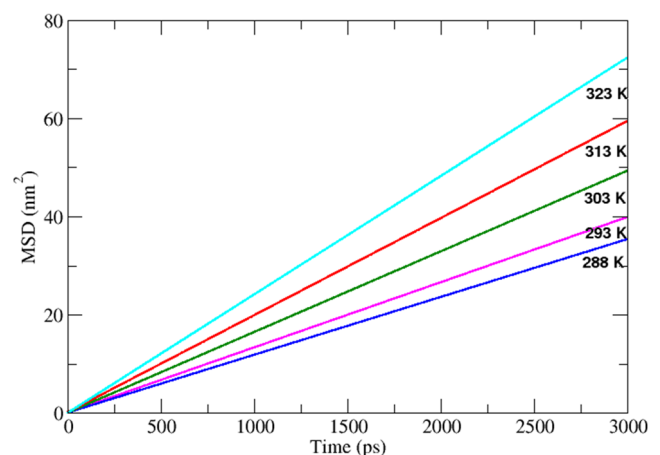


FIG. 4. MSD vs time plot of water at different temperatures.

TABLE III. Estimated values of self and binary diffusion coefficients at different temperatures.

S.No.	Temp. (K)	Diffusion coefficients (D_{PBC}) ($10^{-10} \text{ m}^2 \text{ s}^{-1}$)							
		Self					Binary		
		For cysteine		For water			Cal.	Expt. ⁷	Error (%)
		MSD	MSD	Expt. ³⁰	Error (%)				
1	288	8.17 ± 1.10	19.63 ± 0.01	17.66	11.16	8.20	7.90	3.80	
2	293	9.01 ± 0.27	22.12 ± 0.14	20.25	9.23	9.05	
3	303	10.81 ± 0.41	27.38 ± 0.03	25.97	5.43	10.86	
4	313	13.78 ± 2.79	32.95 ± 0.20	32.22	2.26	13.84	
5	323	16.29 ± 0.51	40.19 ± 0.05	39.83	0.90	16.36	

self-diffusion coefficients are in agreement with that of experimental values within 12% error. The error in the experimental values of self-diffusion coefficient of water, as reported by Holz,³⁰ is less than 1%.

The self-diffusion coefficients of cysteine and water at particular temperatures obtained in the previous sections are now used for the calculation of binary diffusion coefficients by using Darken's relation (2). We have simulated three cysteine molecules and 1039 water molecules, 1042 molecules in total. Thus, the mole fraction of cysteine is 0.003 and that of water is 0.997. The calculated values of binary diffusion coefficients and the corresponding experimental values are shown in Table III. The calculated value of the binary diffusion coefficient agrees within 4%⁷ with the experimental value at 288 K. In addition, the calculated values of diffusion coefficients increase with the increase in temperature because the thermal energy of molecules increases with the increase in temperature but the density of the system decreases, which in turn increases the available space for diffusion. Thus, the molecular movement in the system is enhanced, and hence, the diffusion coefficient increases at higher temperatures.

Temperature dependency of diffusion

As observed in Table III, the diffusion phenomenon is strongly dependent on temperature. The temperature dependent behavior of diffusion is given by the Arrhenius equation²⁹

$$D = D_0 e^{-\frac{E_a}{N_A k_B T}}. \quad (4)$$

In Eq. (4), D is the diffusion coefficient, D_0 represents the pre-exponential factor, E_a is the activation energy for diffusion, N_A is the Avogadro's number whose value is $6.022 \times 10^{23} \text{ mol}^{-1}$, k_B is the Boltzmann's constant whose value is $1.38 \times 10^{-23} \text{ J K}^{-1}$, and T is the absolute temperature. On taking natural logarithm in Eq. (4), we get

$$\ln D = \ln D_0 - \frac{E_a}{N_A k_B T}. \quad (5)$$

The activation energy E_a for diffusion can be obtained from the slope of $\ln D$ vs $\frac{1}{T}$ plot (Arrhenius plot) as

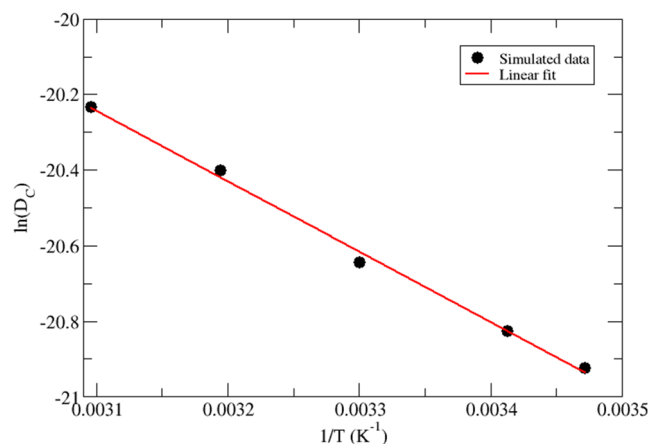
$$E_a = -N_A k_B \frac{\partial \ln D}{\partial (1/T)}. \quad (6)$$

The intercept when extrapolated to the $1/T \rightarrow 0$ in the Arrhenius plot gives the pre-exponential factor.

Figure 5 depicts the Arrhenius plot of the simulated values of self-diffusion of cysteine. The activation energy for self-diffusion of cysteine calculated using the slope of the linear fit of simulated values is found to be $15.49 \text{ kJ mol}^{-1}$.

Figure 6 portrays the Arrhenius plot of both the simulated and experimental values of self-diffusion of water. The activation energies for self-diffusion of water calculated using the corresponding slope of the linear fit of simulated values and experimental values are found to be $15.67 \text{ kJ mol}^{-1}$ and $17.88 \text{ kJ mol}^{-1}$, respectively.

Figure 7 displays the Arrhenius plot of simulated values of binary diffusion of cysteine in water. The activation energy for binary diffusion of cysteine in water calculated using the slope of the linear fit of simulated values is found to be $15.50 \text{ kJ mol}^{-1}$.

**FIG. 5.** Arrhenius diagram for self-diffusion coefficients of cysteine.

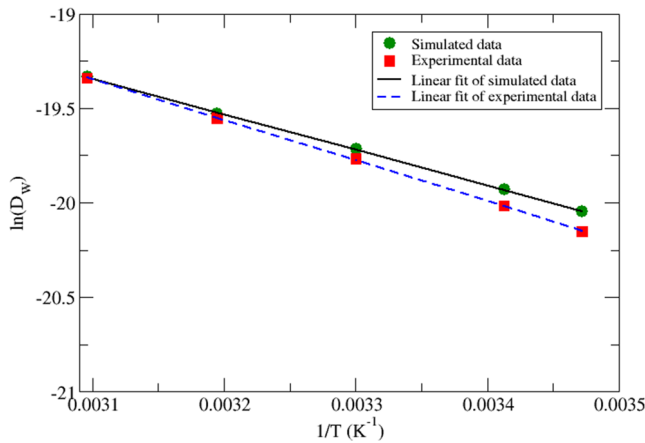


FIG. 6. Arrhenius diagram for self-diffusion coefficients of water

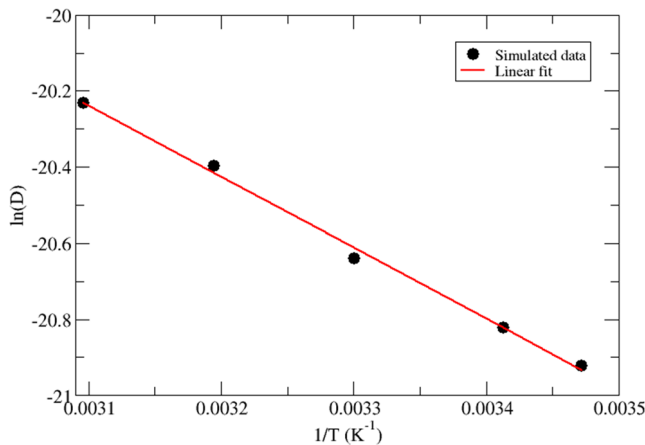


FIG. 7. Arrhenius diagram for binary diffusion coefficients of the binary mixture of cysteine and water.

Figures 5–7 demonstrate the temperature dependency of diffusion. From these plots, it is seen that the diffusion coefficients are found to increase with temperature. We have calculated the activation energies for diffusion of cysteine, water, and their binary mixture by using the slopes of the respective Arrhenius plots, which are tabulated in Table IV.

From Table IV, it is observed that the activation energies for self-diffusion of cysteine and for the binary diffusion of cysteine

TABLE IV. Activation energies for diffusion.

System	Activation energy (E_a) (kJ mol ⁻¹)		
	Simulated	Experimental	Error (%)
Cysteine	15.49		
Water	15.67	17.88 ³⁰	12.36
Binary mixture	15.50		

TABLE V. Estimated values of diffusion coefficients for different sizes (L) of systems at 288 K.

L (nm)	Diffusion coefficients (D_{PBC}) (10^{-10} m ² s ⁻¹)					
	Self		Binary			
	For cysteine	For water	MSD	Expt.	Calculated	Expt.
2.76	8.08	19.36	17.66 ³⁰	8.11	7.90 ⁷	2.7
3.17	8.17	19.63		8.20		3.8
3.75	8.47	20.21		8.50		7.6

in water are almost same. This implies that the concentration of cysteine in the system is infinitesimal. Furthermore, the activation energy calculated for simulated and experimental values of self-diffusion of water is in agreement with the error of 13%.

Effect of system size on diffusion

Moreover, the diffusion coefficient under periodic boundary conditions (PBC) also depends on the size of the system.²³ In the above calculation, the diffusion coefficient has been calculated at different temperatures under periodic boundary conditions. Now, the simulation was extended to find how the diffusion coefficients vary by changing the size of the box. For this, other two systems were set up: (i) 2 cysteine in 693 water molecules in the box of size 2.76 nm and (ii) 5 cysteine in 1732 water molecules in the box of size 3.75 nm. The estimated values of diffusion coefficients under periodic boundary conditions with simulation boxes of different sizes are tabulated in Table V.

In addition, the values of the viscosity coefficient of water and the solution of cysteine in water were determined at 288 K. They were calculated by plotting D_{PBC} vs $1/L$ in accordance with Eq. (3) as in Figs. 8 and 9. The estimated values of the viscosity coefficient of water at targeted temperature are in agreement with the experimental value within 38%. This error might be the nature of the water

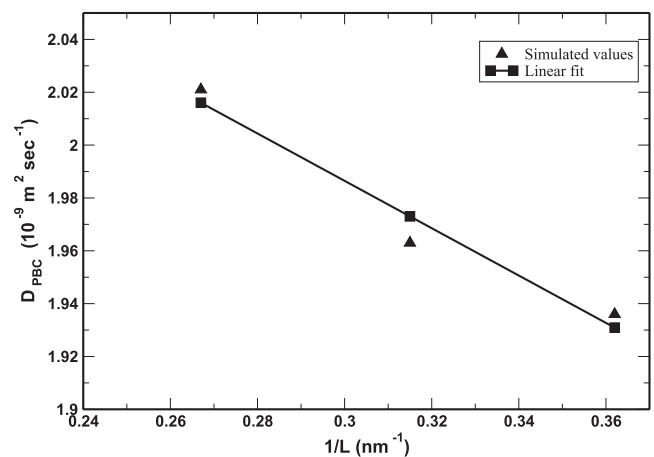


FIG. 8. D_0 vs $(1/L)$ plot for water at 288 K.

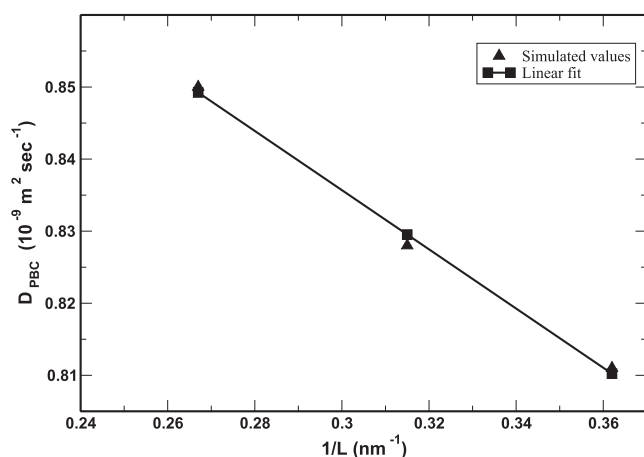


FIG. 9. D_0 vs $(1/L)$ plot for solution of cysteine in water at 288 K.

TABLE VI. System-size independent value of the diffusion coefficient (D_0) and viscosity coefficient (η) for water and solution at 288 K.

System	D_0 ($10^{-10} \text{ m}^2 \text{ s}^{-1}$)			η ($10^{-4} \text{ Nm}^{-2} \text{ s}$)		
	Cal.	Expt.	Error (%)	Cal.	Expt.	Error (%)
Water	22.39	17.66 ³⁰	26.8	7.10	11.37	37.6
Solution	9.59	7.90 ⁷	21.4	14.57		

model SPC/E used for solvation, which underestimates the value of viscosity.³¹ The viscosity coefficient of solution is found to be greater than that of water as presented in Table VI.

CONCLUSIONS

In this work, we have performed the molecular dynamics study of transport properties of cysteine in SPC/E water molecules at 288 K, 293 K, 303 K, 313 K, and 323 K temperatures using the GROMACS 5.1.1 software package. We used OPLS-AA force field parameters throughout the simulation. The structures of the solute and solvent of the system are studied via radial distribution functions between atoms. The analysis of the RDF plots at different temperatures reveals that the system becomes less organized with the increase in temperature. Moreover, Lennard-Jones and Coulomb interactions, including many body effects, contribute to the structural properties of the system.

The self-diffusion coefficients of both cysteine and water are determined by using the Einstein's equation. In addition, the diffusion of the binary mixture of cysteine and water is calculated by using Darken's relation. The simulated values obtained are compared with the corresponding experimental values. The simulated values of self-diffusion coefficients of water show excellent agreements with experimental values, especially at higher temperatures and with a small deviation ($\sim 11\%$) at low temperatures (288 K). Likewise, the simulated values of the binary diffusion coefficient of cysteine in water were compared with the available experimental values at 288 K. This comparison shows very little deviation of about 8%.

Furthermore, the estimated values of diffusion coefficients increase with the increase in temperature, which follows the Arrhenius plots. We also estimate the activation energies of diffusion. We have compared the calculated value of activation energy for self-diffusion of water with the experimental value, which is in agreement with the error of 13%. In addition, we have studied the effect of the system size on diffusion, and the viscosity coefficients of both water and solution are estimated at 288 K.

As the further extension of this work, we have a plan to study the transport properties of polycysteine in aqueous medium.

ACKNOWLEDGMENTS

R.P.K. and S.P.K. acknowledge the partial financial support from the Nepal Academy of Science and Technology (NAST). H.P.B. acknowledges the master thesis grant support from the University Grants Commission (UGC), Nepal. N.P.A. acknowledges the UGC Award No. CRG-73/74-S&T-01 and Associate membership of ICTP helped to complete this work. We also acknowledge fruitful discussion with Prem P Chapagain FIU USA.

REFERENCES

- G. Bulaj, *Biotechnol. Adv.* **23**, 87–92 (2005).
- W. J. Wedemeyer, E. Welker, M. Narayan, and H. A. Scheraga, *Biochemistry* **39**, 4207–4216 (2000).
- R. M. Canet-Avilés, M. A. Wilson, D. W. Wilson, R. Ahmad, M. Rili, C. McLendon, S. Bandyopadhyay, M. J. Baptista, D. Ringe, G. A. Petsko, and M. R. Cookson, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9103–9108 (2004).
- S. I. Rizvi and R. Jha, *Expert Opin. Drug Discovery* **6**, 89–102 (2011).
- D. S. Daniels, T. T. Woo, K. X. Luu, D. M. Noll, N. D. Clarke, A. E. Pegg, and J. A. Tainer, *Nat. Struct. Mol. Biol.* **11**, 714 (2004).
- A. Nosal-Wiercińska, *J. Electroanal. Chem.* **681**, 103–108 (2012).
- W. Jin and H. Chen, *Chromatographia* **52**, 17 (2000).
- A. S. Virk, T. S. Gardner, S. A. Willis, A. M. Torres, and W. S. Price, *Front. Phys.* **3**, 1 (2015).
- P. Heitmann, *Eur. J. Biochem.* **3**, 346–350 (1968).
- N. Nagano, M. Ota, and K. Nishikawa, *FEBS Lett.* **458**, 69–71 (1999).
- E. B. Liu and M. Hilpert, *Comput. Geosci.* **15**, 379–384 (2011).
- D. Sandrin, D. Wagner, C. E. Sitta, R. Thoma, S. Felekyan, H. E. Hermes, C. Janiak, N. de Sousa Amadeu, R. Kühnemuth, H. Löwen *et al.*, *Phys. Chem. Chem. Phys.* **18**, 12860–12876 (2016).
- Y. Ma, C. Zhu, P. Ma, and K. T. Yu, *J. Chem. Eng. Data* **50**, 1192–1196 (2005).
- B. Della Ventura, A. Ambrosio, A. Fierro, R. Funari, F. Gesuele, P. Maddalena, D. Mayer, M. Pica Ciamarra, R. Velotta, and C. Altucci, *ACS Appl. Mater. Interfaces* **8**, 21762–21769 (2016).
- B. Della Ventura, L. Schiavo, C. Altucci, R. Esposito, and R. Velotta, *Biomed. Opt. Express* **2**, 3223–3231 (2011).
- M. Ibba, C. Stathopoulos, and D. Söll, *Curr. Biol.* **11**, R563–R565 (2001).
- L. G. Korotchkina, E. M. Ciszak, and M. S. Patel, *Arch. Biochem. Biophys.* **429**, 171–179 (2004).
- J. Crank, *The Mathematics of Diffusion*, 2nd ed. (Oxford University Press, Ely House, London, WI, 1975).
- H. Hirakawa, Y. kamei, and O. Yasumichi, *Bull. Chem. Soc. Jpn.* **46**, 2659 (1973).
- D. Frenkel and B. Smith, *Understanding Molecular Simulation from Algorithms to Applications*, 2nd ed. (Academic Press, USA, 2002).
- M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids* (Oxford University Press, USA, 1989).
- L. S. Darken, *Trans. AIME* **175**, 184 (1948).

- ²³I. C. Yeh and G. Hummer, *J. Phys. Chem. B* **108**, 15873–15879 (2004).
- ²⁴B. Dünweg and K. Kremer, *J. Phys. Chem. B* **99**, 6983–6997 (1993).
- ²⁵S. H. Jamali, L. Wolff, T. M. Becker, A. Bardow, T. J. H. Vlugt, and O. A. Moulton, *J. Chem. Theory Comput.* **14**, 2667–2677 (2018).
- ²⁶H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, *J. Phys. Chem.* **91**, 6269–6271 (1987).
- ²⁷D. Spoel *et al.*, *GROMACS User Manual, version 5.1.1*, 2016.
- ²⁸S. P. Khanal, Y. P. Kandel, and N. P. Adhikari, *AIP Adv.* **90**, 065303 (2019).
- ²⁹I. Poudyal and N. P. Adhikari, *J. Mol. Liq.* **198**, 77 (2014).
- ³⁰M. Holz, S. R. Heil, and A. Sacco, *Phys. Chem. Chem. Phys.* **2**, 4740 (2000).
- ³¹A. P. Markesteijn, R. Hartkamp, S. Luding, and J. Westerweel, *J. Chem. Phys.* **136**, 134104 (2012).



PAPER

OPEN ACCESS

RECEIVED

20 November 2020

REVISED

18 February 2021

ACCEPTED FOR PUBLICATION

26 February 2021

PUBLISHED

22 March 2021

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Binding of SARS-CoV-2/SARS-CoV spike protein with human ACE2 receptor

Rajendra P Koirala¹, Bidhya Thapa², Shyam P Khanal¹, Jhulan Powrel¹, Rajendra P Adhikari³ and Narayan P Adhikari¹

¹ Central Department of Physics, Tribhuvan University, Kathmandu, Nepal

² Padma Kanya Multiple Campus, Tribhuvan University, Kathmandu, Nepal

³ Department of Physics, Kathmandu University, Dhulikhel, Nepal

E-mail: narayan.adhikari@cdp.tu.edu.np

Keywords: spike, SARS-COV-2, hACE2, SARS-COV, Umbrella Sampling, Free energy,

Supplementary material for this article is available [online](#)

Abstract

SARS-CoV-2 virus is the serious health concern throughout the world. A comprehensive investigation of binding of SARS-CoV-2 active site with host receptor protein hACE2 is important in designing effective drugs. In the present work, the major amino acid binding partners between the virus CTD and host receptor have been studied and are compared with SARS-CoV RBD binding with hACE2. Our investigation show that some unique hydrogen bond pairs which were not reported in previous work. Along with hydrogen bonding, salt-bridges, hydrophobic interactions and contributions of electrostatic and van der Waals contacts play significant role in binding mechanism. The binding affinity of SARS-CoV-2 CTD/hACE2 is greater than SARS-CoV RBD/hACE2. This outcome is also verified from the free energy estimation by using umbrella sampling.

1. Introduction

Corona virus disease (COVID-19) pandemic, caused by severe acute respiratory syndrome (SARS)-like corona virus (SARS-CoV-2), is a serious health concern for the global community [1–4]. Although the origin of the virus is still unclear, it has been spread all over the world threatening the human civilizations after its initial outbreak from China in 2019. Till date, more than 52 millions infected population has been reported globally and more than 12 hundred thousands people have lost their lives [5]. There is no approved drug or vaccine against the COVID-19, even though several antiviral drugs have been proposed and are also in clinical trials [6]. Understanding more about interactions of this virus with human body is highly demanding at this pandemic time to design drugs. SARS-CoV and other viruses had also threatened the human society at different periods; however the influence of SARS-CoV-2 is significantly higher than other viruses throughout the globe.

SARS-CoV-2 has more than 70 percent of structural similarity with SARS-CoV; most of the residues at binding interface are similar [7, 8]. SARS-CoV-2 similar with SARS-CoV and other corona viruses utilize same human angiotensin converting enzyme 2 (hACE2) receptor to enter into human cell. This entry process is mediated by the spike(s) glycoprotein which are embedded in the capsid of SARS-CoV-2 [9]. The spike protein is subdivided into two receptor binding entities S1 and S2. S1 is responsible in the detection of receptor, whereas S2 plays important role in membrane fusion. Similar to SARS-CoV, C-Terminal Domain (CTD) of S1 subunit of spike protein in SARS-CoV-2 acts as receptor binding domain (RBD) [10, 11]. Even though both SARS-CoV and SARS-CoV-2 have same binding domain, the binding affinity of SARS-CoV-2 is different from that of SARS-CoV [12, 13].

Immediately after the COVID-19 outbreak, several researches have been carried out to identify the nature and location of binding of SARS-CoV-2 CTD with hACE2 using static crystal structure [14, 15]. Although, these researches have attempted to detect the entry process and binding mechanism of SARS-CoV-2 with hACE2, the breakthrough on drug designing is still challenging. Several works are in the way of hopeful future, exploration of detail binding mechanism is still being essential. Moreover, the detail dynamics of SARS-CoV-2 and hACE2 can be very important to design the drug. When we were independently working on the binding mechanism of SARS-CoV-2 with host receptor

hACE2, in the mean time similar type of research works have been published [12]. However our technique as well as some results are different from previous works.

We focus on the estimation of free energy difference of virus protein and hACE2 complex. The free energy calculation provides in-depth insight on the binding mechanism between the protein molecules [16]. There are several experimental techniques of measuring binding free energy such as isothermal titration calorimetry (ITC) [17], fluorescence resonance energy transfer (FRET) [18], nuclear magnetic resonance (NMR) [19], surface plasmon resonance (SPR) [20] and many others. The computational approach can be the best complement for large scale investigations [21–23]. Out of many computational approach, umbrella sampling is one of the widely used method for the estimation of free energy in large molecular systems [24, 25]. It improves the sampling system by designing and implementing the biasing potentials as a function of reaction coordinates [26, 27]. If an energy barrier exists in between two regions of configuration states, there may be poor sampling, despite the long simulation run being carried out. The applied biasing potential bridges such configuration states and makes it easier in searching local or global minima, which can be considered as the structurally favorable state in the molecular complex [28]. Besides these techniques, free energy can be calculated directly from steered molecular dynamics (SMD) [29, 30].

In the present work, we have carried out molecular dynamics (MD) simulations for the comprehensive study of binding mechanism of SARS-CoV-2 CTD with hACE2 and also compared with SARS-CoV-RBD/hACE2. In addition, umbrella sampling method has been executed to estimate the binding free energy of SARS-CoV-2 CTD/hACE2. Required windows for the umbrella sampling have been taken from steered molecular dynamics (SMD) [31] simulations. In SMD, SARS-CoV-2 CTD has been translated taking the hACE2 as the reference molecule. The quantitative estimation of binding affinity between the targeted molecules facilitates in silico-drug designing. We have also performed comparative study of various interactions such as hydrogen bonding, salt bridges, hydrophobic, electrostatic and van der Waals interactions at the binding interface of SARS-CoV-2 and SARS-CoV with hACE2. Furthermore, the contact surface area of these complexes have been estimated and compared to investigate the stability.

2. Methods

2.1. System set up

Two molecular structures, PDB IDs 6LZG and 2AJF, were taken for the molecular dynamics simulations. The PDB ID 6LZG contains the complex of SARS-CoV-2 CTD and hACE2 receptor protein (i.e., SARS-CoV-2 CTD/hACE2 complex) and that of PDB ID 2AJF contains the complex of SARS-CoV RBD and hACE2 receptor protein (i.e., SARS-CoV RBD/hACE2 complex) [32]. CHARMM-GUI [33] was used to create the pdb and psf files of these complexes. Then, both the complex structures were solvated using TIP3P [34] water and electrically neutralized by adding NaCl. We have added the NaCl in the system with concentration 0.15 M by using CHARMM-GUI. In SARS-CoV-2 CTD/hACE2 complex system 220 Na⁺ ions and 197 Cl⁻ ions were added to neutralize the system. Similarly in SARS-CoV RBD/hACE2 complex system 214 Na⁺ and 188 Cl⁻ ions were added so that the system became neutral. A cubical box of dimensions 144 × 144 × 144 Å³ was prepared for NPT simulation of the complex SARS-CoV-2 CTD/hACE2 and another cubical box of dimensions 131 × 131 × 131 Å³ was prepared for NPT simulation of the complex SARS-CoV RBD/hACE2. Further, two equal sized orthorhombic simulation boxes were prepared in order to estimate the free energy differences of above complexes by changing the dimensions to 250 × 90 × 90 Å³ and electrically neutralized by adding NaCl with concentration 0.15 M.

2.2. Molecular dynamics simulation

All atom molecular dynamics (MD) simulations were performed using NAMD [31] simulation package. The CHARMM36m [35] force field was used for each simulations. The Particle Mesh Ewald (PME) [36] was used for the long-range interactions with a 12.0 Å non-bonded cut-off. The energy minimization was performed for 10,000 steps, using the conjugate gradient and line search algorithm [37, 38]. The system was then equilibrated at 310 K for 10 ns with harmonically restrained heavy atoms taking 1 fs time step. Finally, the production run was propagated for 250 ns simulation under NPT conditions by using Langevin dynamics with a damping constant of 1 ps⁻¹ taking time step of 2 fs.

2.3. Molecular dynamics and umbrella sampling

To perform the umbrella sampling, sample windows were chosen from steered molecular dynamics (SMD) trajectories. During SMD, CTD/RBD of SARS-CoV-2 CTD/SARS-CoV RBD were pulled correspondingly towards the negative x-direction with constant velocity pulling method of velocity 0.00005 Å/fs. In this process, the alpha carbons of hACE2 protein were taken as the fixed atoms and alpha carbons in CTD/RBD part of spike protein of the systems were taken as the dummy atoms. CTD/RBD of spikes were pulled from their center of mass (COM) along the negative x-direction with constant velocity ($\vec{v} = d\vec{x}/dt$) in water and ions environment. Then the SMD atom experiences the force $F(\vec{t}) = k(\vec{v} t - \Delta\vec{x})$, providing the external potential energy,

$$U(x, t) = \frac{1}{2}k(\vec{v} t - \Delta\vec{x} \cdot \hat{n})^2 \quad (1)$$

where, k ($=5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$) is the spring constant and gives the stiffness of the applied harmonic restraining force, and $\Delta\vec{x}(t) = \vec{x}(t) - \vec{x}_0$, is the displacement of SMD molecules from initial position \vec{x}_0 to instantaneous position $\vec{x}(t)$ and \hat{n} is the unit vector along the direction of pulling.

Out of many other techniques of free energy calculations [39], umbrella sampling was performed to investigate the free energy difference during the translation of SARS-CoV-2 CTD from hACE2 protein for system SARS-CoV-2 CTD/hACE2; and identical condition is applied for system SARS-CoV RBD/hACE2. SMD trajectories were used to select the appropriate windows. Identifying the information on the termination of molecular interactions from SMD, we estimated the number of umbrella windows in both the systems. Ten windows were prepared in SARS-CoV-2 CTD/hACE2 and six windows were prepared for SARS-CoV RBD/hACE2 complexes. Every successive window was taken from the SMD trajectories during the translation of 1 \AA along the negative x -direction. The window size ensures the sufficient overlapping of successive windows to cover the entire reaction coordinate space. The reaction coordinate was chosen as the distance between the center of mass (COM) of hACE2 and CTD/RBD spike along the negative x -axis. To make the necessary overlapping reaction coordinates, a bias potential of the i^{th} window $V_i(x)$ was used to force the system to fluctuate in coordinate space, which is given by,

$$V_i(x) = \frac{1}{2}k(x_i - x_0)^2 \quad (2)$$

where x_0 is the harmonic constraint defining a center of window i ($i=1$ to 10 for SARS-CoV-2 and 1 to 6 for SARS-CoV), and force constant k is the window width. We used the force constant of $1.5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$.

2.4. Data analysis

Visual Molecular Dynamics (VMD) [40] and Pymol [41] were used to visualize as well as generate images of the complex structures. VMD analysis tools were used to identify and analyze non-bonded interactions by using the simulation trajectories. The NAMD energy plugin, available in VMD, was used to calculate the non-bonded interaction energy contributions. Pycontact [42] software package was used to analyze the hydrophobic interactions and salt bridges between the targeted protein residues in CTD/RBD of spike protein and ACE2. Weighted Histogram Analysis Method (WHAM) program [43] was used to estimate the free energy from umbrella sampling simulation. The free energy calculation of large molecular system is generally computationally demanding. This method minimizes the statistical errors as well as increases the efficiency of computational simulation. Moreover, it has the advantage of multiple overlapping of probability distributions for obtaining better estimation of free energy calculations.

3. Results

3.1. Conformational variation in complexes

To examine the conformational variation during the dynamics, we have estimated the root mean square deviation (RMSD) of each molecule on SARS-CoV-2 CTD/hACE2 and SARS-CoV RBD/hACE2 complexes. Besides RMSD, contact surface area between the molecules within the each complex has also been calculated for both complexes.

We have calculated the RMSD for all atoms of proteins backbone without taking hydrogen atoms. The structure from first step of the simulation was taken as the reference to calculate the RMSD. The RMSD of hACE2 and spike CTD/RBD have been compared separately to evaluate the structural integrity of the molecules. Figures 1(a) shows the RMSD of hACE2 molecule in SARS-CoV-2 CTD/hACE2 and SARS-CoV RBD/hACE2 complexes and figure 1(b) represents the RMSD of spike CTD/RBD. From the figure, it is observed that RMSD of hACE2 of SARS-CoV-2 CTD/hACE2 is smaller than that of SARS-CoV RBD/hACE2. RMSD of both the systems are stable with the values below 3.0 \AA and 4.5 \AA for CoV-2 and CoV respectively. SARS-CoV-2 CTD has the RMSD of 2.5 \AA , whereas the RMSD of SARS-CoV RBD is 4.0 \AA . This shows the large rearrangements of structure in SARS-CoV RBD, while SARS-CoV-2 CTD structure remains relatively stable.

To get more insight into stability, we also analyzed the contact surface area between the spike protein CTD/RBD and hACE2 receptor using MD trajectories. Contact area is the surface buried at the interface between two proteins which contributes to bind and stabilize the protein-protein complexes. Larger contact surface indicates greater stability of the structure [44]. The estimated values of contact surface area are presented in table 1. From the table 1, it has been observed that SARS-CoV-2 CTD/hACE2 has larger contact area than SARS-CoV RBD/hACE2 by $77.02 \pm 2.46 \text{ \AA}^2$. The contact surface area for SARS-CoV-2 CTD is more in comparison to SARS-CoV RBD indicating the greater binding affinity of SARS-CoV-2 with receptor.

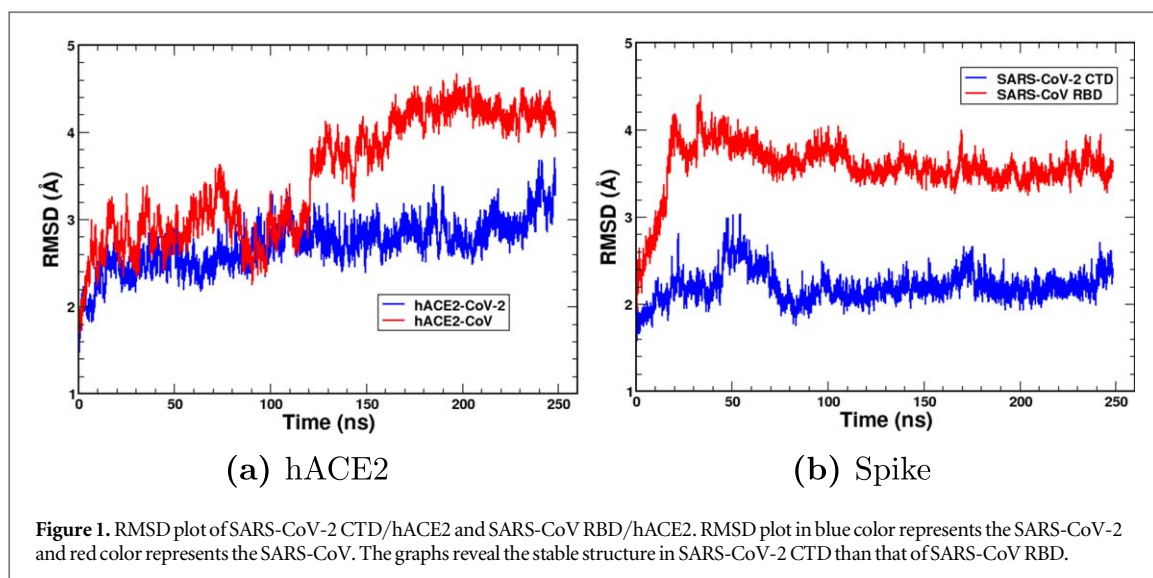


Table 1. Calculation of contact surface area; A_1 , A_2 & A_3 are the average solvent accessible surface area (SASA) of hACE2, SARS-CoV-2 CTD and their complexes; and A'_1 , A'_2 & A'_3 are the average solvent accessible surface area (SASA) of hACE2, SARS-CoV RBD and their complexes.

Complex	SASA (\AA^2) for				Net contact area (A)[44]
SARS-CoV-2 CTD/hACE2	hACE2 (A_1)	SARS-CoV-2 CTD (A_2)	complex (A_3)	$A_1 + A_2 - A_3$	$(A_1 + A_2 - A_3)/2$
	29 100.36	11227.97	38 549.02	1779.31	889.66 ± 1.07
SARS-CoV RBD/hACE2	hACE2 (A'_1)	SARS-CoV RBD (A'_2)	complex (A'_3)	$A'_1 + A'_2 - A'_3$	Net contact area (A')
	29 092.62	10859.54	38326.88	1625.27	812.64 ± 2.22

3.2. Non-bonded interactions

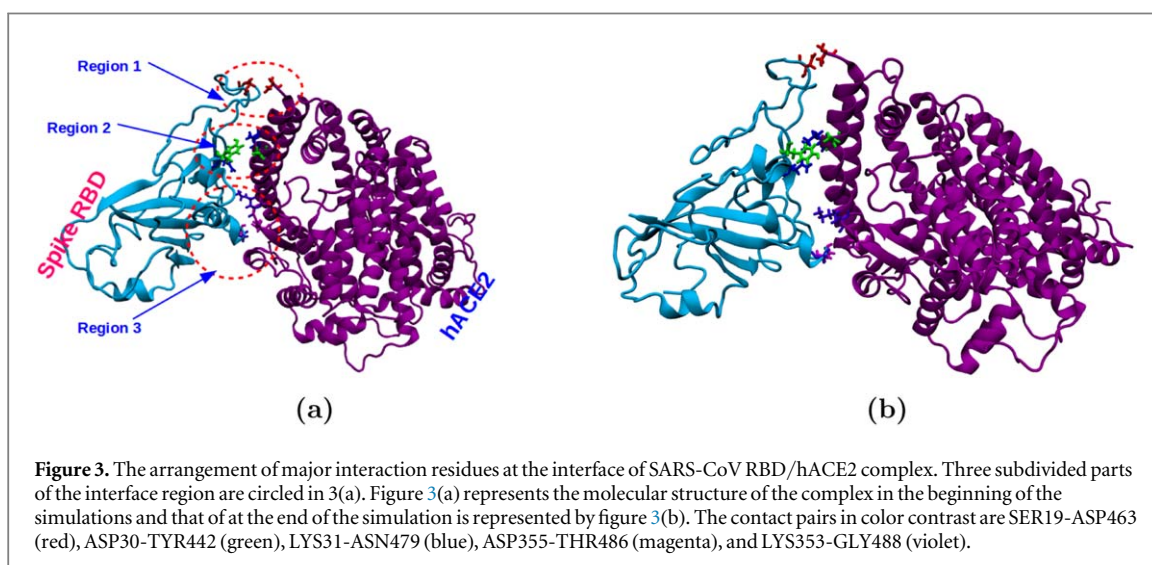
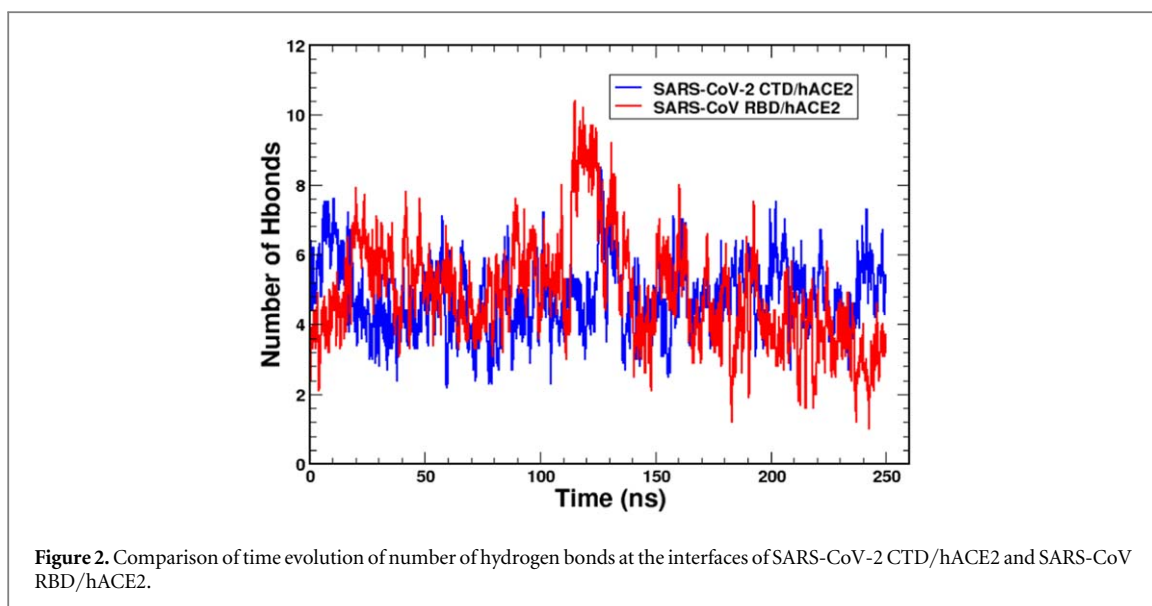
Furthermore, we studied in details the hydrogen bonds, salt-bridges, hydrophobic, electrostatic and vdw interactions between the residues residing at the interface between SARS-CoV-2 CTD/hACE2 and obtained results are compared with SARS-CoV RBD/hACE2.

3.2.1. Hydrogen bonds

At the interface region, hydrogen bonds play pivotal role in binding the molecules to form a stable complex. Wang *et al* (2020) and Lan *et al* (2020) have studied the atomic interactions at the interface of static crystal structure of SARS-CoV-2 CTD/hACE2 complex [14, 15], whereas we have investigated the hydrogen bonds at the interface of two complexes by analyzing the MD simulations trajectories. The cut-off distance for hydrogen bond was taken to be 3.5 Å [14]. We monitored the time evolution of number of hydrogen bonds formed at the interface between SARS-CoV-2 CTD/hACE2 and also compared with that of SARS-CoV RBD/hACE2 as shown in figure 2 (also see supplementary table S 1). Hydrogen bonds were found to be consistently existing in both complexes. Total hydrogen bonds formed during the simulations were seen to be more in case of SARS-CoV; however the strength and life time of potential hydrogen bonds were found to be greater in case of SARS-CoV-2.

Many research works have been published by analyzing the hydrogen bonds pair partners between the molecules in the complexes. Even though our investigations regarding the hydrogen bonds in the complexes are in consistent with those published papers, some pair partners are not consistent with these previously published outcomes. Ali *et al* reported three unique hydrogen bonds in SARS-CoV-2 CTD/hACE2 complex, which were not detected in SARS-CoV RBD/hACE2 complex [12]. We found consistent result in GLU35-TYR449 pair, however the result is not consistent with other two pair partners: TYR449-ASP38 and GLN498-LYS353. We have clearly observed ASP38-TYR436 and LYS353-GLY488 pairs in SARS-CoV-RBD/hACE2. Furthermore, a strong hydrogen bond has been detected between GLN498 with GLN42. In static structures, no hydrogen bond was formed by SER19 of hACE2 with ASP463 residue of SARS-CoV RBD [14, 15]. Our investigation shows two potential hydrogen bonds formed between main-main & main-side chains of SER19 of hACE2 with side chain of ASP463 of SARS-CoV RBD and similar type of bond has been detected between SER19 and ALA475 in SARS-CoV-2 CTD/hACE2 complex (see figure 4(a) and supplementary figures S1 and S2).

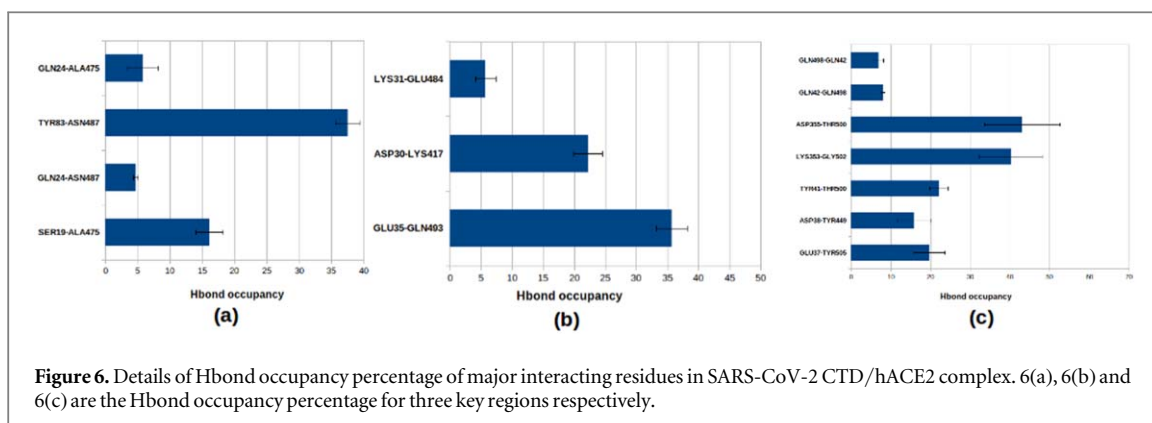
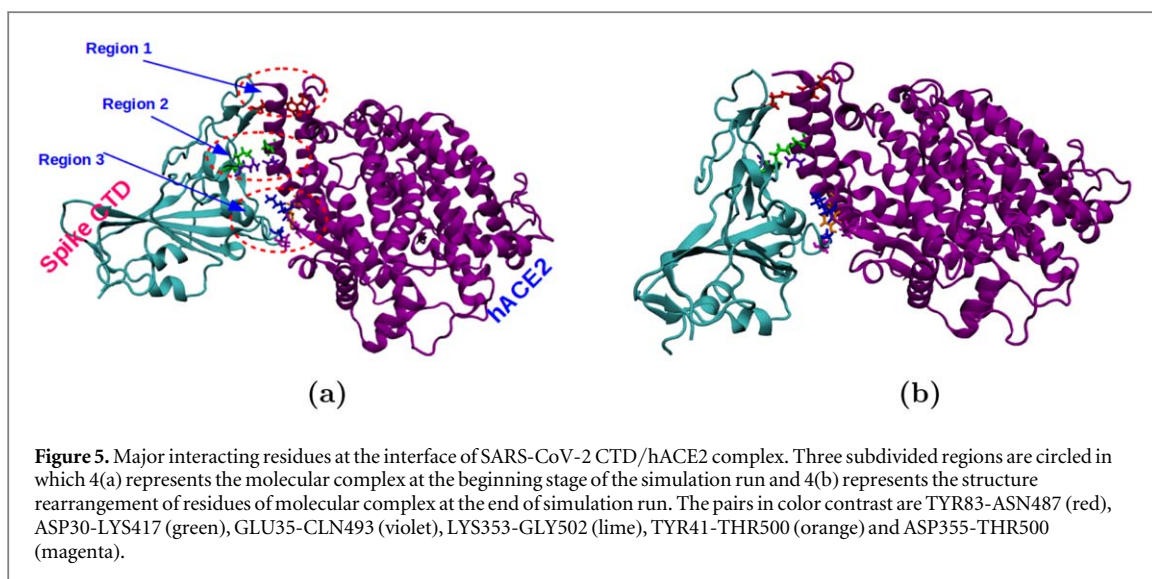
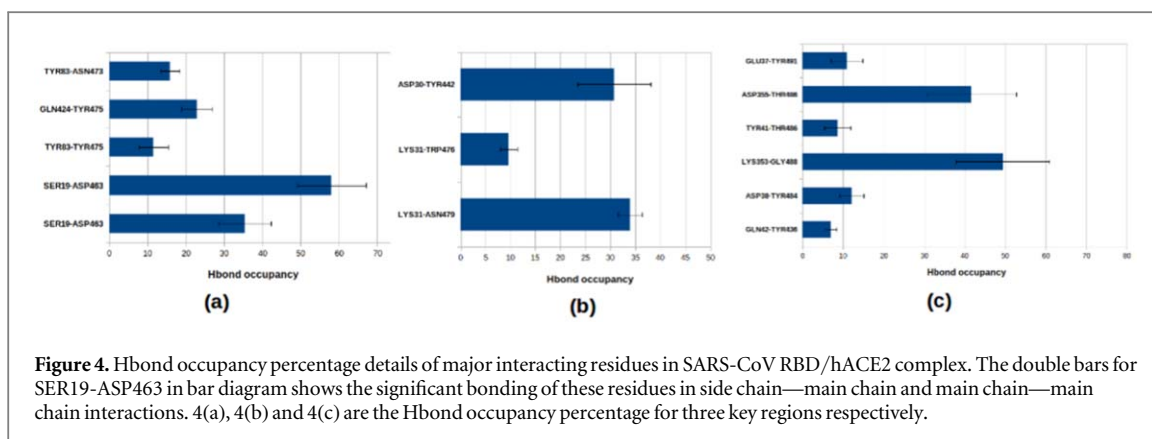
In the present work, interactions of molecules in each complex has been observed considering three main regions where the interfacial residues of hACE2 take part actively in binding with the spike CTD/RBD section of



the virus molecule as shown in supplementary figures S1(a) and S2(a). The hydrogen bonds formed at three key regions of interface between SARS-CoV RBD/hACE2 in the beginning of the simulation run are shown in figures (see supplementary figures: S1(a), S1(b) & S1(c)). Some residue pairs whose hydrogen bond (Hbond) occupancy percentage greater than 20% is shown in figure 3 and detail of Hbond occupancy percentage in three key regions are shown in figures 4(a), (b) and (c). We have observed some differences in the atomic interactions at the interface of both virus proteins and hACE2 than that of static crystal structure. The two different approaches might be the reason of variation in the number of interactions.

The rearrangement of aminoacids at the binding vicinity have been detected in both the complexes. Different color contrast have been used to identify the pair partners as shown in figures 3 and 5. The binding affinity of the complex was observed increasing due to the alignments of pair partners at the interface.

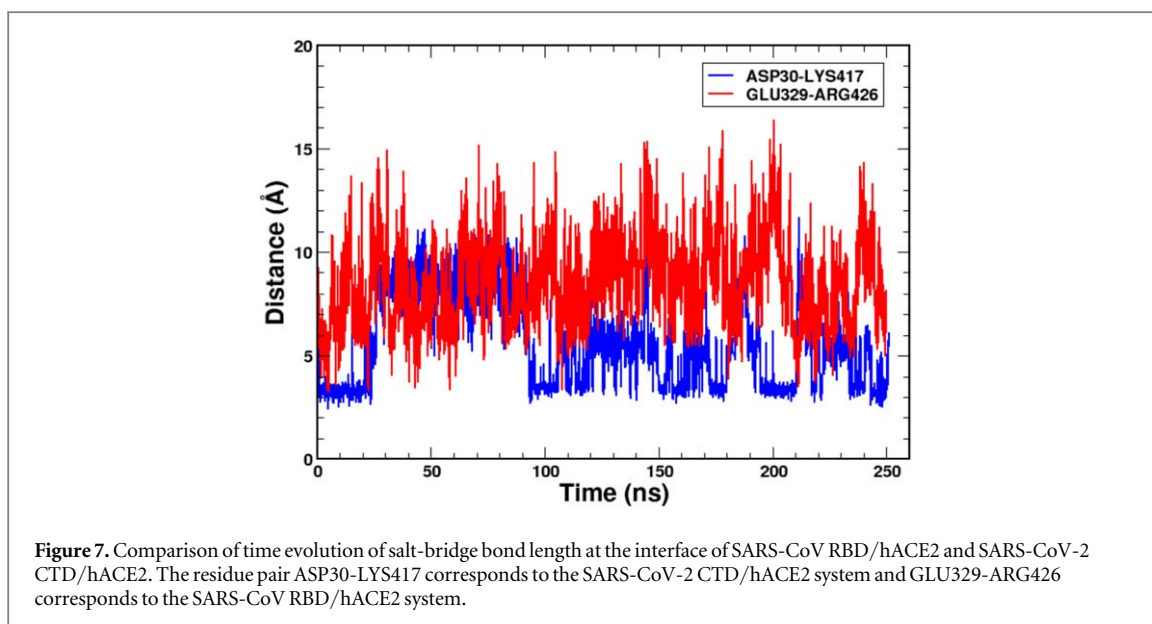
At the binding interface of SARS-CoV-2 CTD and hACE2 receptor, three key regions where most of the polar contacts are actively participated to form the hydrogen bonds. The interactions at the interface in which Hbond occupancy greater than 20% are shown in figure 5, and also the details of Hbond occupancy in these regions are shown in figures 6(a), (b) and (c). In region 1, SER19, GLN24 and TYR83 of hACE2 form hydrogen bonds with ALA475 and ASN487 of SARS-CoV-2 CTD as in figures (see supplementary figures: S2(a), S3(a), S4(a), S5(a) & S6(a)). In region 2, there are interactions between the residues LYS417, TYR453 and GLN493 of SARS-CoV-2 CTD forming hydrogen bonds with ASP30, LYS31 and GLU35 of hACE2 (see supplementary figures: S2(b), S3(b), S4(b), S5(b) & S6(b)). Similarly, in region 3, there is extensive network of hydrogen bonds between SARS-CoV-2 CTD residues TYR449, GLN498, THR500, ASN501, GLY502 and TYR505 with the hACE2 residues GLU37, ASP38, TYR41, GLN42, LYS353 and ASP355 (see supplementary figures: S2(c), S3(c), S4(c), S5(c) & S6(c)). Because of dynamical nature of our system, there is continuous formation, breaking and



reformation of hydrogen bonds during the simulations. The variation of hydrogen bonds formed during the five representative frames of 250 ns simulation of SARS-CoV-2 has been shown in supplementary figures S2–S6.

3.2.2. Salt-bridges

In addition to extensive network of interfacial hydrogen bonds, another important contributions to protein-protein binding comes from salt-bridge interactions. MD trajectory analysis has shown three salt-bridges, having different bond length and strength, formed at the interface of SARS-CoV-2 CTD/hACE2. The salt-bridge formed between the residue LYS417 of SARS-CoV-2 CTD with ASP30 of hACE2 is found to be the strongest one among them owing to its short bond length. The remaining residues GLU484 and LYS458 of



SARS-CoV-2 CTD have formed salt-bridges with LYS31 and GLU23 of hACE2 respectively. In contrast, we found only one salt-bridge formed between ARG426 of SARS-CoV RBD and GLU329 of hACE2 which is weaker than that of SARS-CoV-2 because of larger bond length as in figure 7.

3.2.3. Hydrophobic interactions

3.2.4. Electrostatic and van der Waals (vdw) interactions

The electrostatic and van der Waals (vdw) interactions in two complexes SARS-CoV RBD/hACE2 and SARS-CoV-2 CTD/hACE2 have been studied. Supplementary figure S7 depicts the comparative analysis of energy due to electrostatic and vdw interactions as a function of time. In the beginning of simulations, the electrostatic contributions of SARS-CoV-2 CTD/hACE2 was distinctly higher than SARS-CoV RBD/hACE2, however most of the simulation time, the contributions were almost equal. In addition, the potential energy contributed by vdw interactions were consistently almost equal for both the systems throughout the simulations. It reveals that electrostatic and vdw interactions are almost equally contributed in binding both the complexes.

3.3. Free energy

To investigate the energetic difference in binding of hACE2 with SARS-CoV-2 CTD and SARS-CoV RBD, the free energy differences have been estimated using umbrella sampling technique. Umbrella windows were taken from the trajectories of SMD simulations. The interactions between the molecules in SARS-CoV-2 CTD/hACE2 were found terminated after traversing 9 Å distance away from the original position. To incorporate all interacting pathways, ten windows with 1 Å distance separation were taken for every successive window. On the other hand, the interactions between the molecules in SAR-CoV RBD/hACE2 were found ceased after traversing 5 Å distance from the original position. Therefore, six windows were prepared separating 1 Å distance away for every successive window. To ensure the overlapping of consequent data sets in umbrella sampling, we have plotted the distributions of data obtained from the simulations. We found sufficient overlapping of data sets. The graphs for distribution versus COM distance have been included in the supplementary figures S8 and S9. Figure 8 shows the change in free energy during the translation of virus CTD/RBD from active pocket of hACE2 for both complexes. The center of mass (COM) distance as a reaction coordinate allows us to track the free energy changes for SARS-CoV-2 CTD in complex with hACE2 and SAR-CoV RBD in complex with hACE2. Free energy has also been used to compare the differences in the binding affinity for the two complexes. The SARS-CoV-2 CTD in complex with hACE2 is found to have the greater binding free energy of ~ 1.91 kcal/mol compared to the SAR-CoV RBD in complex with hACE2. This, as well as the nature of the free-energy curve, provides an insight on binding mechanisms of the complexes.

4. Discussion

COVID-19 pandemic has seriously threatened public health throughout the globe. Since there is no approved drug till date to combat against the SARS-CoV-2 virus, more comprehensive study is essential through the

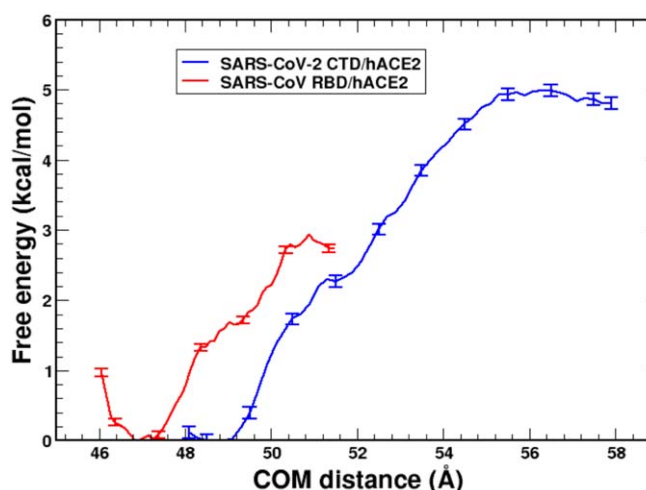


Figure 8. Free energy curve for SARS-CoV-2 CTD/hACE2 and SARS-CoV RBD/hACE2 during the translation of spikes CTD/RBD from hACE2.

various aspects at molecular level. The fundamental necessity is to understand the entry mechanism of the virus into the human cell, which is really helpful to discover the drug against the virus. To deal the entry mechanisms and dynamical characteristics of the virus cell in complex with hACE2 receptor, we used various computational techniques. C-Terminal Domain (CTD) of S1 subunit of spike protein, being the active interacting region, has been taken into consideration in SARS-CoV-2. We performed the comparative analysis of the key residues and atomic interactions responsible for the binding of the SARS-CoV-2 CTD and SARS-CoV RBD with human ACE2 receptor.

Estimation of structural variation during the simulation is the foremost judgement of molecular stability in molecular dynamics study. RMSD is the measure of stability of molecular structure in the cellular environment. Well equilibrated system with consistent RMSD ensures us to proceed for the further study of binding affinity and energy variations of the molecular complexes. Moreover, contact surface area between the molecules identifies the binding strength of the complex. Therefore, we have obtained the contact surface area of both complexes calculating the solvent accessible surface area (SASA). SASA has been determined from time evolution data generated from the 250 ns NPT run. Then, average value of contact area for both the systems have been presented in table 1 and are interpreted graphically in figure 2. Larger contact surface area in SARS-CoV-2 CTD/hACE2 complex depicts the stronger binding of this complex than that of SARS-CoV RBD/hACE2 [45].

Our results show considerable similarity in the binding sites, interfacial residues and important atomic interactions in both viral protein receptor binding domain (i.e., SARS-CoV-2 CTD and SARS-CoV RBD). However, there are some variations in loop between two structures in the binding region and some residues at the binding sites are different. This facilitates more and stronger atomic contacts between SARS-CoV-2 CTD and hACE2 interface and thereby enhancing its binding affinity. Polar residues residing at the interface form an extensive network of hydrogen bonds and salt-bridge interactions [46–49]. Our study reveals that interfacial hydrogen bonds, salt-bridges and hydrophobic interactions play an important role in the binding of SARS-CoV-2 CTD to host cell receptor. Furthermore, comparative analysis of the binding mechanism of two viral proteins with hACE2 show that binding affinity of SARS-CoV-2 is greater than that of SARS-CoV. Notably, more residues are engaged in the binding of SARS-CoV-2 CTD with hACE2. We find the greater number of potential hydrogen bonds formed in the case of SARS-CoV-2 CTD which contributes to higher binding affinity. More and stronger salt-bridges formed in case of SARS-CoV-2 CTD establish stronger binding to the receptor than SARS-CoV RBD. Additionally, we observe hydrophobic interactions are stronger in case of SARS-CoV-2 which also contribute to enhanced binding.

The contributions of electrostatic and vdw contacts are significant to form a stable protein-protein complex [50, 51]. The potential energy in binding the virus CTD/RBD and host receptor are compared in both the systems. Though, initially the electrostatic energy is observed relatively larger in SARS-CoV-2 CTD/hACE2 than that of SARS-CoV RBD/hACE2, the dynamical results show almost equal contributions in both the complexes. This shows that the contributions of hydrogen bonds, salt bridges and hydrophobic interactions are responsible to provide the greater binding strength in SARS-CoV-2 CTD/hACE2.

The binding mechanisms of the complexes are further analyzed to estimate the free energy differences from umbrella sampling method. SMD trajectories are taken for the appropriate samples that ensure the sufficient

overlapping on windows [52]. In SMD, the virus CTD/RBD are pulled upto that distance, beyond which no interactions persists. We find the interactions of molecules in complex SARS-CoV RBD/hACE2 have been terminated after the displacement of RBD by 5 Å from host receptor, whereas the interactions sustain upto 9 Å displacement from the initial position in SARS-CoV-2 CTD/hACE2. Comparisons of free energy of two complexes have provided the insight of bonding affinity between the virus CTD/RBD and hACE2 molecules. The greater free energy difference between SARS-CoV-2 CTD in complex with hACE2 depicts the stronger binding strength than the complex of SARS-CoV RBD and hACE2. As the further investigation, we plan to calculate the solvation free energy of SARS-CoV-2 and SARS-CoV molecule in the aqueous environment.

Acknowledgments

RPK & SPK acknowledge the partial financial support from Nepal Academy of Science and Technology (NAST). JP acknowledges the partial financial support from university grant commission (UGC). NPA acknowledges the UGC Award no. CRG-73/74-S&T-01 and TWAS research grants RG 20-316. We acknowledge the computing facilities of Supercomputer Centre Kathmandu University, which was established with equipment donated by CERN and We thank Prof. Prem Chapagain and Prabin Baral at Florida International University for their help with computing facilities for MD simulations.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

References

- [1] Yan R *et al* 2020 Structural basis for the recognition of the 2019-nCoV by human ACE2 *Science* **367** 1444–8
- [2] Guan W *et al* 2020 Clinical characteristics of coronavirus disease 2019 in China *N. Engl. J. Med.* **382** 1708–20
- [3] Zhu N *et al* 2020 A novel coronavirus from patients with pneumonia in China, 2019 *N. Engl. J. Med.* **382** 727–33
- [4] Wells C R *et al* 2020 Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak *Proc. Natl. Acad. Sci.* **117** 7504–9
- [5] <https://covid19.who.int/table>.
- [6] Folegatti P M *et al* 2020 Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial *The Lancet* **396** 887–97
- [7] Shang J *et al* 2020 Clinical characteristics of coronavirus disease 2019 in China *Nature* **581** 221–4
- [8] Hoffmann M *et al* 2020 SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor *Cell* **181** 271–80
- [9] Lu G, Wang Q and Gao G F 2015 Bat-to-human: spike features determining ‘host jump’ of coronaviruses SARS-CoV, MERS-CoV, and beyond *Trends Microbiol.* **23** 468–78
- [10] Lu G *et al* 2013 Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26 *Nature* **500** 227–31
- [11] Li F *et al* 2005 Structure of SARS coronavirus spike receptor-binding domain complexed with receptor *Science* **309** 1864–8
- [12] Ali A and Vijayan A 2020 Dynamics of the ACE2–SARS-CoV-2/SARS-CoV spike protein interface reveal unique mechanisms *Sci. Rep.* **581** 1–12
- [13] Spinello A Saltalamacchia A and Magistrato A 2020 Is the Rigidity of SARS-CoV-2 Spike Receptor-Binding Motif the Hallmark for Its Enhanced Infectivity? Insights from All-Atoms Simulations *J. Phys. Chem. Lett.* **11** 4785–90
- [14] Wang Q *et al* 2020 Structural and functional basis of SARS-CoV-2 entry by using human ACE2 *Cell* **181** 894–904
- [15] Lan J *et al* 2020 Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor *Nature* **581** 215–20
- [16] Peng Y *et al* 2018 Predicting protein-DNA binding free energy change upon missense mutations using modified MM/PBSA approach: SAMPDI webserver *Bioinformatics* **34** 779–86
- [17] Velázquez-Campoy A *et al* 2004 Isothermal titration calorimetry *Curr. Protoc. Cell Biol.* **23** 17–8
- [18] Hillisch A, Lorenz M and Diekmann S 2001 Recent advances in FRET: distance determination in protein-DNA complexes *Curr. Opin. Struct. Biol.* **11** 201–7
- [19] Campagne S, Gervais V and Milon A 2011 Nuclear magnetic resonance analysis of protein-DNA interactions *J. R. Soc. Interface.* **8** 1065–78
- [20] Teh H *et al* 2007 Characterization of protein-DNA interactions using surface plasmon resonance spectroscopy with various assay schemes *Biochemistry* **46** 2127–35
- [21] Donald J E, Chen W W and Shakhnovich E I 2007 Energetics of protein-DNA interactions *Nucleic Acids Res.* **35** 1039–74
- [22] Jones S *et al* 1999 Protein-DNA interactions: a structural analysis *J. Mol. Biol.* **287** 877–96
- [23] Zhang C *et al* 2005 A knowledge-based energy function for protein-ligand, protein-protein, and protein-DNA complexes *J. Med. Chem.* **48** 2325–35
- [24] Torrie G M and Valleau J P 1977 Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling *J. Comput. Phys.* **23** 187–99
- [25] Kästner J 2011 Umbrella sampling *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1** 932–42
- [26] Zheng L, Chen M and Yan W 2008 Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems *Proc. Natl. Acad. Sci.* **105** 20227–32
- [27] Yang M *et al* 2014 Combine umbrella sampling with integrated tempering method for efficient and accurate calculation of free energy changes of complex energy surface *J. Chem. Phys.* **141** 07B618_1

- [28] Yang Y et al 2016 Efficient sampling over rough energy landscapes with high barriers: A combination of metadynamics with integrated tempering sampling *J. Chem. Phys.* **144** 094105
- [29] Sun Z and Zhang J Z 2021 Thermodynamic insights of base flipping in TNA duplex: force fields, salt concentrations, and free-energy simulation methods *CCS Chemistry* **3** 1026–39
- [30] Sun Z et al 2019 Sulfur-substitution-induced base flipping in the DNA duplex *Phys. Chem. Chem. Phys.* **21** 14923–40
- [31] Phillips J et al 2005 Scalable molecular dynamics with NAMD *J. Comput. Chem.* **26** 1781–802
- [32] Berman H et al 2000 The protein data bank *Nucleic Acids Res.* **28** 235–42
- [33] Lee J et al 2016 CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field *J. Chem. Theory Comput.* **12** 405–13
- [34] Jorgensen W et al 1983 Comparison of simple potential functions for simulating liquid water *J. Chem. Phys.* **79** 926–35
- [35] Huang J et al 2017 CHARMM36m: an improved force field for folded and intrinsically disordered proteins *Nat. Methods* **14** 71–71
- [36] Harvey M J, Giupponi G and Fabritiis G De 2009 ACEMD: accelerating biomolecular dynamics in the microsecond time scale *Nat. Methods* **5** 1632–9
- [37] Khanal S P, Kandel Y P and Adhikari N P 2019 Transport properties of zwitterion glycine, diglycine, and triglycine in water *AIP Adv.* **9** 065303–10
- [38] Koirala R P, Bhusal H P, Khanal S P and Adhikari N P 2020 Effect of temperature on transport properties of cysteine in water *AIP Adv.* **10** 025122
- [39] Wang X and Sun Z 2019 Determination of Base-Flipping Free-Energy Landscapes from Nonequilibrium Stratification *J. Chem. Inf. Model* **59** 2980–94
- [40] Humphrey W et al 1996 VMD: visual molecular dynamics *J. Mol. Graph.* **14** 33–8
- [41] Schrödinger L L C 2015 The PyMOL Molecular Graphics System *Version 1.8*.
- [42] Scheurer M et al 2018 PyContact: Rapid, customizable, and visual analysis of noncovalent interactions in MD simulations *Biophys. J.* **114** 577–83
- [43] Kumar S et al 1992 The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method *J. Comput. Chem.* **13** 1011–21
- [44] Zou X et al 2012 Recognition of methylated DNA through methyl-CpG binding domain proteins *Nucleic Acids Res.* **40** 2747–58
- [45] Ma B et al 2003 Protein–protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces *Proc. Natl. Acad. Sci.* **100** 5772–7
- [46] Teague S J 2003 Implications of protein flexibility for drug discovery *Nat. Rev. Drug Discov.* **2** 527–41
- [47] Chen J, Sawyer N and Rayan L 2013 Protein–protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area *Protein Sci.* **22** 510–5
- [48] Jones S and Thornton J M 1997 Analysis of protein–protein interaction sites using surface patches *J. Mol. Biol.* **272** 121–32
- [49] Xu D, Tsai C and Nussinov R 1997 Hydrogen bonds and salt bridges across protein–protein interfaces *Protein Eng.* **10** 999–1012
- [50] Jones S et al 2003 Using electrostatic potentials to predict DNA-binding sites on DNA-binding proteins *Nucleic Acids Res.* **31** 7189–98
- [51] DiStasio R A Jr., Gobre V V and Tkatchenko A 2014 Many-body van der Waals interactions in molecules and condensed matter *J. Phys. Condens. Matter* **26** 213202
- [52] Isralewitz B et al 2001 Steered molecular dynamics investigations of protein function *J. Mol. Graph. Model.* **19** 13–25

Rajendra P. Koirala, Rudramani Pokhrel, Prabin Baral, Purushottam B. Tiwari*,
Prem P. Chapagain* and Narayan P. Adhikari*

Structural insights into the repair mechanism of AGT for methyl-induced DNA damage

<https://doi.org/10.1515/hsz-2021-0198>

Received August 9, 2020; accepted May 29, 2021;

published online June 30, 2021

Abstract: Methylation induced DNA base-pairing damage is one of the major causes of cancer. O⁶-alkylguanine-DNA alkyltransferase (AGT) is considered a demethylation agent of the methylated DNA. Structural investigations with thermodynamic properties of the AGT-DNA complex are still lacking. In this report, we modeled two catalytic states of AGT-DNA interactions and an AGT-DNA covalent complex and explored structural features using molecular dynamics (MD) simulations. We utilized the umbrella sampling method to investigate the changes in the free energy of the interactions in two different AGT-DNA catalytic states, one with methylated GUA in DNA and the other with methylated CYS145 in AGT. These non-covalent complexes represent the pre- and post-repair complexes. Therefore, our study encompasses the process of recognition, complex formation, and separation of the AGT and the damaged (methylated) DNA base. We believe that the use of parameters for the amino acid and nucleotide modifications and for the protein-DNA covalent bond will allow investigations of the DNA repair mechanism as well as the exploration of cancer therapeutics targeting the AGT-DNA complexes at various functional states as well as explorations via stabilization of the complex.

Keywords: AGT-DNA complex; DNA methylation; DNA repair; MD simulations; umbrella sampling.

Introduction

DNA base-pairing damage involves chemical modifications such as base mismatch and methylation of DNA bases (Crone et al. 1996; Kyrtopoulos et al. 1997; Maser and DePinho, 2002) and is one of the major causes of cancer (Jackson and Bartek 2009; Paulsen and Ferguson-Smith 2001). Methylation of a DNA base can cause carcinogenesis of a living cell (De Bont and van Larebeke 2004; Lindahl and Barnes 2000; Tessmer and Fried 2014; Warren et al. 2006). O⁶-alkylguanine-DNA alkyltransferase (AGT) functions as the demethylation agent for the O⁶-alkylguanine and O⁴-alkylthymine DNA bases (Gerson 2002; Harris et al. 1992; Kelley and Fishel 2008; McKeague et al. 2018; Zak et al. 1994). The interaction of AGT with DNA is considered as a novel interaction mechanism for the recognition of DNA damage and repair (Daniels et al. 2004; Musarrat et al. 1995; Perugino et al. 2012; Rossi et al. 2018). After locating the DNA damage, AGT accesses the alkylated base by flipping the nucleotide into the protein so that the methylated GUA approaches the enzyme active site. Once the methylated GUA is in the active site, it donates its methyl group to the CYS145 residue in AGT (Ali et al. 1998; Duguid et al. 2005).

Because it confers protection against DNA damage, AGT plays a crucial role to protect normal cells from tumorigenesis. For example, it is shown to protect against methylation-induced skin cancer (Becker et al. 1996), liver cancer (Nakatsuru et al. 1993), lung cancer (Liu et al. 1999; Sakumi et al. 1997), and thymic lymphomas (Dumenco et al. 1993). However, AGT can also provide undesired protection to cancer cells from damage due to methylating and chloroethylating anticancer drugs (Fu et al. 2012; Kaina et al. 2007; Rasimas et al. 2003), thus fostering drug resistance. Therefore, AGT is an important drug target for improving the efficacy of chemotherapy.

Despite previous investigations of this functionally important DNA-AGT complex, structural investigations with thermodynamic properties of the complex are still lacking. Computational investigations of the demethylation process

*Corresponding authors: **Purushottam B. Tiwari**, Department of Oncology, Georgetown University, Washington, D.C., USA, E-mail: pbt7@georgetown.edu. <https://orcid.org/0000-0002-3140-0723>; **Prem P. Chapagain**, Department of Physics, Florida International University, Miami, FL, USA; and Biomolecular Sciences Institute, Florida International University, Miami, FL, USA, E-mail: chapagap@fiu.edu. <https://orcid.org/0000-0002-0999-4975>; and **Narayan P. Adhikari**, Central Department of Physics, Tribhuvan University, Kathmandu, Nepal, E-mail: narayan.adhikari@cdp.tu.edu.np. <https://orcid.org/0000-0003-4535-1368>

Rajendra P. Koirala, Central Department of Physics, Tribhuvan University, Kathmandu, Nepal, E-mail: rajendra.koirala@cdp.tu.edu.np

Rudramani Pokhrel and Prabin Baral, Department of Physics, Florida International University, Miami, FL, USA, E-mail: rpokh002@fiu.edu (R. Pokhrel), pbara006@fiu.edu (P. Baral)

of methylated DNA bases by AGT are limited due to the lack of pre-defined force-field parameters. In this work, we utilized computational approaches with CHARMM36m (Huang et al. 2017) force field parameters to create covalent and non-covalent complexes and explored their suitability to investigate structural features of the complex using molecular dynamics (MD) simulations. Such structural investigations have increasingly been used in recent years for investigating complexes of proteins with other partners (Magnen et al. 2018; Tiwari et al. 2016, 2018, 2020), including the protein-DNA covalent complex with implications in drug discovery (Tiwari et al. 2020). We used the umbrella sampling method to calculate the changes in the free energies in AGT-DNA complexes in two non-covalent complexes in different states of methylation, pre- and post-repair structures, one with methylated GUA in DNA and the other with methylated CYS145 in AGT. This technique allows the determination of energetically favorable states and has been used in studies of many bio-molecular systems (Luzhkov 2017; Pokhrel et al. 2019). The use of the parameters for the amino acid and nucleotide modifications and for the protein-DNA covalent bond allows computational investigations of the AGT-DNA complex structures before and after methyl transfer, as well as the AGT-DNA covalent complex. Such investigations will be useful for future explorations of cancer therapeutics targeting the AGT-DNA complexes at various functional states as well as explorations via stabilization of the complex. For example, the catalytic site in the transient state of the DNA-AGT complex may provide a novel target for *in silico* drug screening to identify AGT inhibitors that inhibit the methyl transfer, potentially overcoming the anticancer drug resistance to improve the efficacy of chemotherapy.

Results

Dynamics of the AGT-DNA covalent and non-covalent complexes

We generated all three complexes – complex-I, complex-II, and complex-III as explained in the Methods section. The complex-I has methylated GUA7 (6OG7) in DNA, complex-II represents a possible model of a transient, intermediate state in which methylated GUA7 covalently bonded to CYS145, and complex-III has methylated CYS (ORT145) in AGT. In the complex-I, GUA7 is methylated at O⁶ point and this methyl group is supposed to be transferred to CYS145 in AGT whereas the methyl group in GUA7 is transferred to CYS145 in complex-III. In this transfer, the sulfur atom of

CYS145 side chain becomes deprotonated and the methyl group gets attached to it. Figure 1A shows a representative AGT-DNA complex (complex-I). Figure 1B displays the covalent bonding between the 6OG7 and CYS145. It is to be noted that in the AGT-DNA covalent complex in the protein data bank (PDB ID 1T39), CYS145 is cross-linked to the mechanistic inhibitor N¹, O⁶-ethanoxanthosine (Daniels et al. 2004) as shown in Figure 1C. This inhibitor was purposely used to prevent the methyl transfer and achieve a stable covalent complex, resulting in the covalent bonding seen in Figure 1C in contrast to Figure 1B, which represents the transient, intermediate catalytic state during the methyl transfer process. Figure 1D shows the pre-catalytic state with methylated GUA7 (6OG7) and Figure 1E shows the post-catalytic state with methylated CYS145, both without a covalent bond between the GUA and CYS. The CYS145 residue (with or without the hydrogen attached to the sulfur) is enclosed within a circle in Figure 1B–E.

We performed 200 ns of MD simulations for each of the three complexes to ensure the stability of the complexes and to investigate structural dynamics. A representative simulation movie of the 200 ns CPT production run for the AGT-DNA covalent complex, complex-II, is given in the Supplementary Material (Section S3). The DNA appears to be flexible at the ends, especially at the lower end. We calculated the total interaction energies using the 200 ns simulation trajectories of each of the three complexes and results are shown in Figure 2A. There is no significant fluctuation in the total energy for all three systems, suggesting the stability of the complexes (Tiwari et al. 2018). Figure 2B also shows that there was no significant fluctuation in RMSD in all the three complexes. The total number of hydrogen bonds formed between AGT and DNA as a function of simulation time shows a similar level of hydrogen bonding in all three complexes (Figure 2C). A detailed analysis of the AGT-DNA hydrogen bonding shows that residues TYR114 and CYS145 (or ORT145) in AGT consistently form hydrogen bonding with 6OG7 (or GUA7). These residues and nucleotides may be important in the formation of the complexes. Table 1 shows all residues in AGT and nucleotides in DNA that establish hydrogen bonding with occupancy greater than 40%. We have also calculated hydrogen-bonding distances for the two residue-nucleotide pairs and presented in Figure 2D–F, for complex-I, complex-II, and complex-III, respectively. As expected the weaker hydrogen bonding was observed for the complex-III (post-catalytic state) as compared to complex-I (pre-catalytic state) as can be seen in Figure 2D, F as well as in Table 1. Even though TYR114-GUA7 in the complex-III bonding looks as stable as TYR114-6OG7 in the complex-I, the occupancy of this hydrogen bonding is much less (~59%) in the complex-III as compared to the

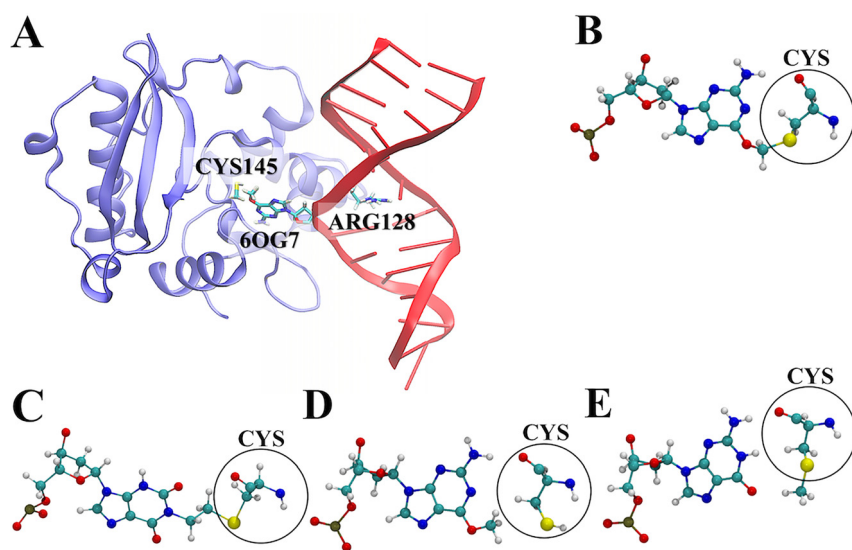


Figure 1: (A) A representative structure of the AGT-DNA complex (Complex I). The intercalating residue ARG128 in AGT, O⁶-methylguanine (6OG7) in DNA, and nearby CYS145 in AGT are shown in stick representations. (B) Covalent bonding between AGT and DNA at the biological substrate O⁶-methylguanine (6OG7). (C) Covalent bonding between AGT and DNA cross-linked to the mechanistic inhibitor N¹, O⁶-ethanoxanthosine as presented in the PDB ID 1T39 (Daniels et al. 2004). (D) CYS145 in AGT and 6OG7 in DNA without the covalent bonding. (E) Methylated CYS145 (ORT145) in AGT and GUA7 in DNA without the covalent bonding. The AGT residues and DNA nucleotides forming the covalent bonding in Figure 1A and the residues and nucleotides in Figure 1B–E are shown in colored CPK representation. The CYS145 portions (with or without hydrogen atom bound to the sulfur atom) in Figure 1B–E are enclosed within circles.

complex-I (~88%). We believe that the stronger hydrogen bonding in the complex-I corresponds to the recognition of DNA by AGT for the methyl transfer process. We also observed that three water molecules reside within interfacial

region between AGT and DNA during the 200 ns simulation of the complex-I and form two water-mediated hydrogen bonds between ARG128 and CYT8, and one between LYS125 and CYT21, as shown in Figure S1 (Supplementary Material

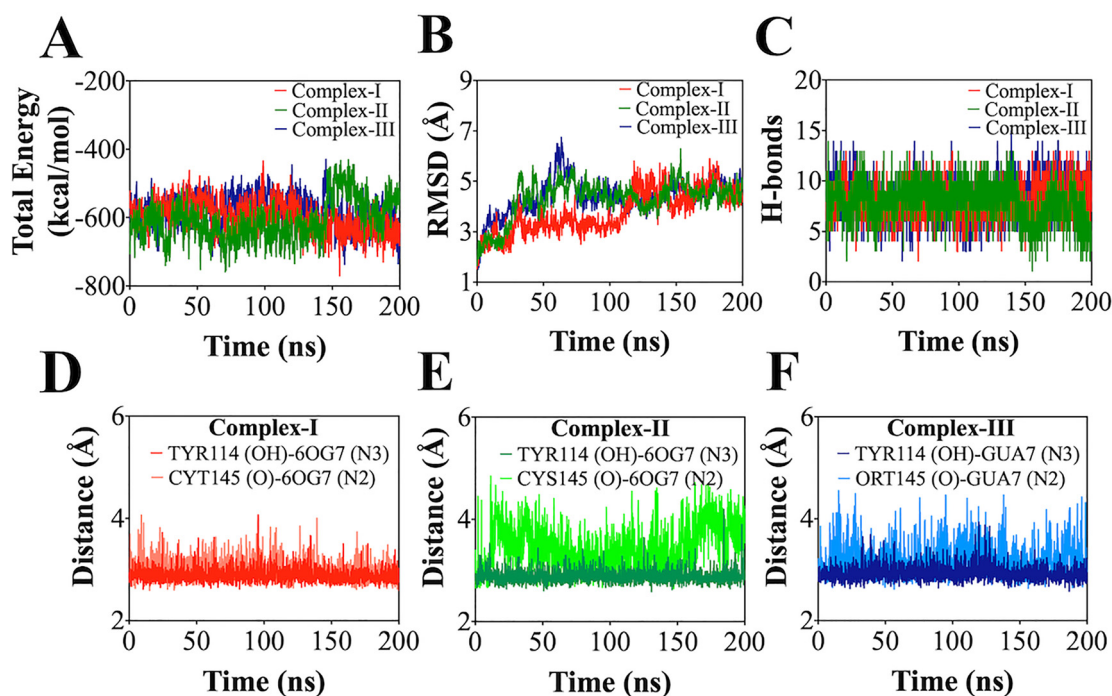


Figure 2: (A) Total energy, (B) RMSD measurements, (C) number of hydrogen bonds between AGT and DNA, and (D)–(F) distances between atoms in residues and nucleotides forming hydrogen bonds in the three AGT-DNA complexes (complex-I, complex-II, and complex-III) calculated using 200 ns simulation trajectories.

Table 1: AGT residues and DNA nucleotides establishing hydrogen bonding between AGT and DNA with occupancy greater than 40%.

AGT-DNA complex	Residue-nucleotide pairs		Occupancy
	AGT	DNA	
Complex-I	TYR114 (OH)	6OG7 (N3)	87.6%
	CYS145 (O)	6OG7 (N2)	87.1%
	SER151 (OG)	THY9 (O1P)	85.6%
	THR95 (OG1)	THY23 (O1P)	80.3%
	SER151 (N)	THY9 (O2P)	69.6%
	THR95 (N)	THY23 (O1P)	62.8%
	GLN115 (N)	THY9 (O1P)	53.2%
Complex-II	SER151 (OG)	THY9 (O1P)	90.0%
	TYR114 (OH)	6OG7 (N3)	71.6%
	THR95 (OG1)	THY23 (O1P)	69.2%
	SER151 (N)	THY9 (O2P)	68.4%
	THR95 (N)	THY23 (O1P)	56.1%
	CYS145 (O)	6OG7 (N2)	53.8%
	GLN115 (N)	THY9 (O1P)	48.9%
Complex-III	ARG128 (NH2)	CYT20 (N3)	47.8%
	TYR114 (OH)	GUA7 (N3)	59.0%
	SER151 (N)	THY9 (O1P)	55.8%
	PHE94 (N)	THY23 (O1P)	48.4%
	ORT145 (O)	GUA7 (N1)	47.8%
	ARG135 (NH1)	GUA7 (O5')	44.6%
	ASN157 (ND2)	GUA7 (O4')	43.7%
	ORT145 (O)	GUA7 (N2)	40.9%

The atoms that are predicted by VMD to be responsible for establishing the hydrogen bondings are presented inside parentheses.

Section S4). While we did not observe these water-mediated hydrogen bonding in complexes II and III during the simulated timescales, such water occupancy may occur in longer timescales.

DNA-AGT complex formation and dissociation

To investigate the underlying mechanism of AGT-DNA complex formation (pre-catalytic, formation of complex-I) and dissociation (post-catalytic, complex-III) process, we performed umbrella sampling for different separation distances between AGT and DNA. For this, DNA was translated, along the negative x -axis, relative to AGT by 1 Å window for 19 windows. The direction of the displacement was chosen visually based on the orientation of the intercalated ARG128 side-chain as well as the flipped GUA7 base so that when displaced, they are minimally obstructed by other residues (i.e. nearly parallel to the ARG128 sidechain and the flipped base and directly away from AGT). The complex was then re-oriented so that the direction of the displacement lies along the

x -axis for convenience. Figure S2A in the Supplementary Material (Section S5) depicts the direction of translation. Two sets of these windows were prepared, one with methylated GUA7 (6OG7) and the other with methylated CYS145 (ORT145). The access of AGT to the alkylated base only occurs when alkylated base is flipped out of the base stack and into the AGT active site for extrahelical repair and ARG128 facilitates the base flipping (Daniels et al. 2004). As the base flipping is observed to occur at a rate of $k = 350 \text{ s}^{-1}$ (Zang et al. 2005) (i.e. in the order of milliseconds), base flipping and the associated conformational transitions upon AGT binding are not accessible in our simulated timescales. Therefore, we used the pre-catalytic complex as our starting structure, which is formed when the methylated GUA base is flipped and captured into the AGT active site. Once the complex is formed and the methyl is transferred to CYS145, the repaired DNA undergoes slow dissociation from AGT (Zang et al. 2005). It is also observed that post-transfer methylated AGT can remain bound to DNA and undergoes ubiquitination and degradation (Tessmer et al. 2012). The set with the methylated CYS145 represents the process during the dissociation of the complex. We calculated the non-bonded interaction energies as a function of the separation distance (COM distance) for both of these processes. Figure 3 shows the contribution of van der Waals and electrostatic energies averaged over 10 ns simulations at various COM distances. As shown in Figure 3A (complex-I) and Figure 3B (complex-III), the electrostatic energy was found to have the major contribution to the interaction energies. A modest contribution from the van der Waals interaction is observed for COM distance $< 32.8 \text{ Å}$ as the methylated GUA7 approaches the active site (Figure 3A) or GUA7 separates from the active site (Figure 3B). The Representative snapshots of the complex-I at different COM distances are shown in the Supplementary Material (Figure S2, Section S5). When the distance is gradually increased from 32.8 Å, AGT dissociates from the minor groove of DNA, losing the original AGT-DNA interactions and forming new interactions. The intercalating residue ARG128 interacts strongly with GUA7 when the COM distance $> 37 \text{ Å}$. The RMSD measurements (shown in Figure S3, Section S6) are fairly stable in each of the 10 ns windows, indicating the convergence of simulations.

To understand the energetic differences in the approach of the damaged base versus separation of the repaired base, we calculated the free energies for these processes using umbrella sampling along the COM as the reaction coordinate. While COM is a good order parameter for comparing the stability of the protein-DNA complex, we note that the distance between the sulfur atom of CYS145 and the oxygen atom of the GUA7 base (S-O distance) can also be chosen as a reaction coordinate. Figure 4 shows the change in free energy for approaching

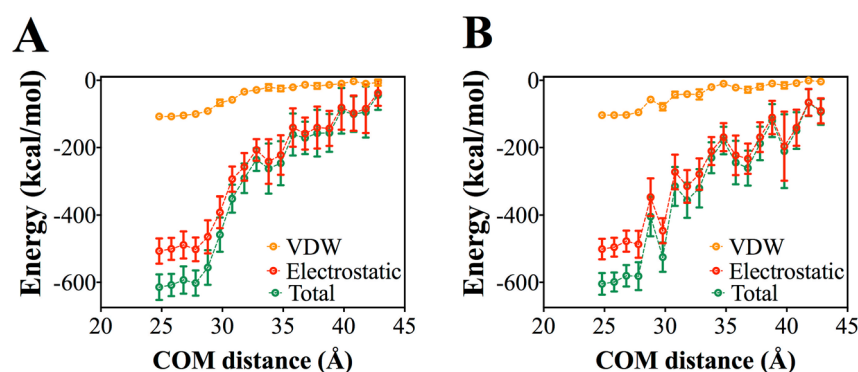


Figure 3: Variation of non-bonded interaction energies with the displacement of DNA from AGT. Van der Waal interaction energy (VDW, orange color), electrostatic interaction energy (E, red color) and total non-bonded energy (E+VDW, green color) was calculated using NAMD energy plugin within VMD for complex-I (A) and for complex-III (B). Symbols and error bars represent average and standard deviation, respectively, calculated from the 10 ns NAMD simulation trajectories for the complexes at each COM distance. Dotted lines are guides to eyes.

methylated GUA7 to the active site (complex-I) as well as for separating demethylated GUA7 from the active site (complex-III). To compare the free-energies of the complexes with GUA7 at the active site, we aligned the curves when the DNA and AGT are fully separated, at which point the free energies for both systems should be comparable as the interaction between DNA and AGT is negligible for the COM distance >42 Å, after 18th window. As shown in Figure 4, the change in free energy for the complex-I, ΔG_1 , between the complexed and separated states was found to be ~ 11.1 kcal/mol and that for the complex-III, ΔG_2 , was ~ 9.8 kcal/mol. Comparison of the lowest free energy values at ~ 26 Å COM distance, which represents the state in which the GUA7 or 6OG7 in the catalytic cavity, shows that the pre-transfer 6OG7 is more favorable in the cavity compared to the post-transfer GUA7, giving a free-energy advantage for 6OG7 of $\Delta\Delta G \sim -1.3$ kcal/mol over GUA7. This may facilitate the dissociation of the GUA7 from the

cavity once the methyl group is transferred to CYS145. Also, the shallower free energy curve beyond ~ 28 Å for complex-III indicates an easier dislodgement of the post-transfer GUA7. These observations explain the affinity of 6OG7 over GUA7 in the catalytic cavity, giving an insight into the base demethylation mechanism of AGT.

Discussion

AGT is considered as the demethylation agent to repair the methyl-damaged DNA (Hu et al. 2008; Tessmer and Fried 2014; Tubbs et al. 2007). The entry of AGT near the damaged part of DNA is the fundamental requirement to initiate the methyl transfer process (McKeague et al. 2018) by forming a transient complex between the two molecules (Daniels et al. 2004; Duguid et al. 2005; Fang et al. 2008). Methylation of a DNA base is one of the major causes of the carcinogenesis of a living cell (Lindahl and Barnes 2000; Tessmer and Fried 2014). While AGT plays a crucial role in repairing the DNA damage and protect the normal cells from tumor development, it can also allow anticancer drug resistance by protecting the cancer cells from anticancer drugs (Kaina et al. 2007; Rasimas et al. 2003; Fu et al. 2012) that aim at damaging the DNA. Given its importance as a drug target for overcoming anticancer drug resistance to improve chemotherapeutic efficacy, structural models of the AGT-DNA complex in different functional states of the methyl transfer process are desired for biophysical insights as well as for in silico screening. For example, small molecules interfering the covalent complex formation and release can potentially inhibit the DNA repair process, improving the efficacies of the anticancer drugs given in combination. Motivated by this, we have created the covalent complex between AGT and DNA and performed subsequent MD simulations to investigate its structural features and we believe that the complex-II

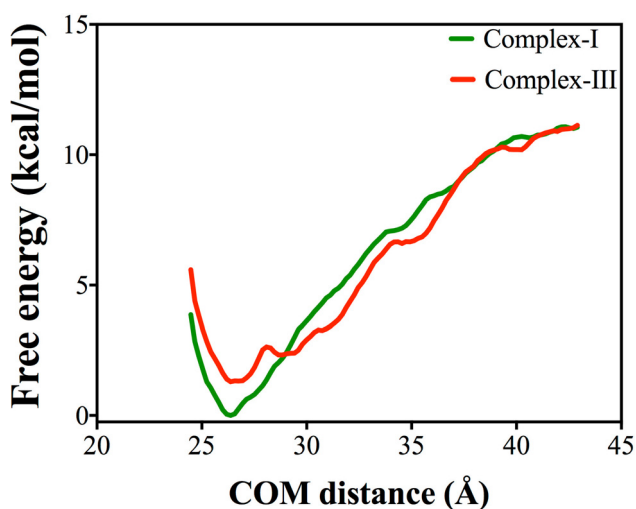


Figure 4: Change in free energy for complex-I and complex-III as a function of COM distance.

generated in this work with the force field modifications may be useful in the structure-based discovery of anticancer agents. We note that our transient intermediate model uses the covalently linked sp^3 hybridized carbon between S and O (i.e. $-S \cdots CH_2 \cdots O-$). However, the actual transfer of methyl group to sulfur from methylguanine is considered to occur via a S_N2 -like mechanism (Daniels et al. 2000; Mattosovich et al. 2020), which involves a planar, penta-coordinated carbon ($-S \cdots CH_3 \cdots O-$) (Fernández et al. 2007) as the transient intermediate with short lifetime (Fu et al. 2021). While the orientations of the groups and the nature of the cavity that likely occur in the order of 100 ns timescales (Patra et al. 2016) may be sufficiently represented by our model for the purpose of stabilizing the complex, further work with appropriate force-field modifications for penta-coordinate carbon is needed for a more detailed analysis of the geometry and energetics of the transition state.

To gain insights into the DNA repair mechanism of AGT, we performed umbrella sampling and calculated free energy and presented results in Figure 4 from 10 ns simulations. We also performed the sampling with 20 ns up to COM distance of 40 Å for Complex-I and we did not see much change compared to the 10 ns sampling. This comparison is shown in the Supplementary Material (Figure S4, Section S7). Figure 4 shows the free energy as a function of separation distance between AGT and DNA for both systems – with methylated GUA and with methylated CYS, representing the pre- and post-transfer of the methyl group. In actual post-transfer process, slight displacement of the AGT helix 6 in the helix-turn-helix motif resulting from the steric hindrance due to methylation of CYS145 may facilitate the release of the repaired DNA from the cavity, and subsequent ubiquitination and protein degradation (Daniels et al. 2000). In our simulations, only a slight increase in RMSD is observed (Figure 2B) in complex III (post-transfer) compared to complex I (pre-transfer) and the process of conformational changes resulting in the repaired base release is difficult to obtain in our computational timescales. Here, the post-transfer separation of DNA-AGT was studied with manual translation of DNA relative to AGT with an increment of 1 Å displacement along the negative x -direction.

The COM distance as a reaction coordinate allows us to track the free-energy changes for both of these processes and compare the differences in the affinity for the methylated versus demethylated GUA in the catalytic cavity. The interaction energies (shown in Figure 2A) were not able to clearly differentiate the structural changes, which is understandable but the free-energy calculations, more relevant to the binding assays, do show the difference between pre- and

post-transfer. The methylated GUA is found to have an affinity of $\Delta\Delta G \sim -1.3$ kcal/mol in the cavity compared to the unmethylated GUA after the methyl group is transferred. This, as well as the nature of the free-energy curve, provides an insight into the base demethylation mechanism of AGT. Finally, our study encompasses the pre-transfer, model intermediate and post-transfer complexes, and separation of the AGT and the damaged (methylated) DNA base and provides models for computational investigations of the DNA repair mechanism as well as for the exploration of cancer therapeutics targeting these AGT-DNA complexes at various functional states.

Materials and methods

System setup

The molecular structures, PDB IDs 1T38 and 1T39 (Daniels et al. 2004), were taken from the protein data bank to create the input structures for MD simulations. The PDB structure 1T38 contains the protein-DNA non-covalent complex with a methyl group attached to the O^6 -position of GUA7 (named as 6OG7) of the DNA (complex-I). The PDB structure 1T39 contains the protein-DNA covalent-complex (complex-II). CHARMM-GUI (Lee et al. 2016) was used to fulfill the missing residues in both PDB structures. We used VMD (Humphrey et al. 1996) to create the non-covalent AGT-DNA system with methylated GUA7 (complex-I) with CHARMM36m (Huang et al. 2017) force-field. The residue SER145 in the PDB ID 1T38 was replaced with CYS. The GUA methylation required additional force field parameters, which are given in Supplementary Material (Section S1). Similarly, for generating the AGT-DNA covalent system (complex-II), we first converted the residue E1X7 in the PDB ID 1T39 to methylated GUA7 and then applied a patch to connect the methylated GUA7 and CYS145, resulting in a covalent bond between the protein and DNA. Finally, the non-covalent AGT-DNA system with methylated CYS145 (complex-III) was set up using the PDB ID 1T38 with additional set of force-field parameters for CYS methylation. The additional topology and parameters used to generate all the three complexes are given in Supplementary Material (Sections S1 and S2). Each of these complexes was solvated with TIP3P water in cubic box and electrically neutralized by adding NaCl. The non-covalent complexes-I and -III used in the umbrella sampling simulations were solvated in orthorhombic boxes.

Molecular dynamics simulation

All-atom molecular dynamics (MD) simulations were performed using the NAMD simulation package (Phillips et al. 2005). The CHARMM36m (Huang et al. 2017) force field was used in simulations of all the complex structures. The additional force field parameters that were used to simulate the three different complexes are given in the Supplementary Material (Section S2). The Particle Mesh Ewald (PME) was used for the long-range interactions with a 12.0 Å non-bonded cutoff.

The energy minimization was performed for 10,000 steps, using the conjugate gradient and line search algorithm. Each system was then equilibrated with harmonically restrained heavy atoms at 300 K using 1 fs time step. The production runs were performed using Langevin dynamics with a damping constant of 1 ps^{-1} under CPT conditions.

Umbrella sampling

Umbrella sampling was performed for the complex-I and complex-III systems. We prepared 19 different 1 \AA windows for each system with the AGT protein taken as a reference molecule and DNA was translated along the negative x -direction by manual displacement. The window size ensures the sufficient overlapping of successive windows to cover the entire reaction coordinate space (Banavali and MacKerell 2002; Luzhkov 2017; Sugita et al. 2000). The reaction coordinate was chosen as the distance between the center-of-mass (COM) of AGT and DNA along the negative x -axis. To make the necessary overlapping reaction coordinates, a bias potential $V(x)$ was used to force the system to fluctuate in coordinate space, which is given by, $V(x) = \frac{1}{2}k_i(x - x_0)^2$, where x_0 is the harmonic constraint defining a center of window i ($i = 1$ to 19), and force constant k_i is the window width. Although the harmonic potential fluctuates the system to overlap the reaction coordinates, the windows are still unbiased. We used the force constant of $1.5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$.

Data analysis

VMD (Humphrey et al. 1996) was used to analyze the simulation trajectories and visualize the structures. The NAMD energy plugin, available in VMD was used to calculate the non-bonded interaction energies; electrostatics (E), van der Waals (VDW) contributions. The potential energy contributions to electrostatic and VDW energy are given by $U_E = \frac{q_1q_2}{4\pi\epsilon_0 r}$ and $U_{VDW} = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]$, respectively (Phillips et al. 2005), where q_1 and q_2 are charges of the interacting pairs separated at a distance r , ϵ_0 is the permittivity of the free space, ϵ is the depth of the potential well, and σ is the distance at which the potential energy is zero. VMD was used to analyze hydrogen bonding between AGT residues and DNA nucleotides. The free energy was estimated by the Weighted Histogram Analysis Method (WHAM) (Kumar et al. 1992) program. GaphPad prism (San Diego, CA) was used to plot the graphs.

Acknowledgments: We sincerely thank Professor Alexander D. MacKerell Jr. at the Computer-Aided Drug Design Center, University of Maryland, USA, for his valuable input and help with the force-field parameters. R.P.K. acknowledges the partial financial support from the Nepal Academy of Science and Technology (NAST). N.P.A. acknowledges the UGC Award No. CRG-73/74-S&T-01 and Associate membership of The Abdus Salam International Center for Theoretical Physics, Trieste, Italy.

Author contributions: P.B.T., P.P.C., and N.P.A. guided the project. R.P.K., R.P., P.B., and P.B.T. performed MD simulations. R.P.K., R.P., P.B., P.B.T., P.P.C., and N.P.A. analyzed the data. R.P.K., P.B.T., and P.P.C. wrote the

manuscript. R.P., P.B., and N.P.A. contributed to the manuscript editing.

Research funding: This work was supported by University Grants Commission- Nepal (CRG-73/74-S&T-01).

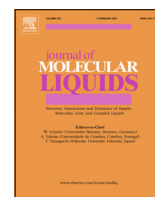
Conflict of interest statement: The authors declare no conflict of interest.

References

- Ali, R.B., Teo, A.K., Oh, H.K., Chuang, L.S., Ayi, T.C., and Li, B.F. (1998). Implication of localization of human DNA repair enzyme O⁶-methylguanine-DNA methyltransferase at active transcription sites in transcription-repair coupling of the mutagenic O⁶-methylguanine lesion. *Mol. Cell. Biol.* 18: 1660–1669.
- Banavali, N.K. and MacKerell, A.D., Jr. (2002). Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. *J. Mol. Biol.* 319: 141–160.
- Becker, K., Dosch, J., Gregel, C.M., Martin, B.A., and Kaina, B. (1996). Targeted expression of human O(6)-methylguanine-DNA methyltransferase (MGMT) in transgenic mice protects against tumor initiation in two-stage skin carcinogenesis. *Canc. Res.* 56: 3244–3249.
- Crone, T.M., Goodtzova, K., and Pegg, A.E. (1996). Amino acid residues affecting the activity and stability of human O⁶-alkylguanine-DNA alkyltransferase. *Mutat. Res.* 363: 15–25.
- Daniels, D.S., Mol, C.D., Arvai, A.S., Kanugula, S., Pegg, A.E., and Tainer, J.A. (2000). Active and alkylated human AGT structures: a novel zinc site, inhibitor and extrahelical base binding. *EMBO J.* 19: 1719–1730.
- Daniels, D.S., Woo, T.T., Luu, K.X., Noll, D.M., Clarke, N.D., Pegg, A.E., and Tainer, J.A. (2004). DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat. Struct. Mol. Biol.* 11: 714–720.
- De Bont, R. and van Larebeke, N. (2004). Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis* 19: 169–185.
- Duguid, E.M., Rice, P.A., and He, C. (2005). The structure of the human AGT protein bound to DNA and its implications for damage detection. *J. Mol. Biol.* 350: 657–666.
- Dumenco, L.L., Allay, E., Norton, K., and Gerson, S.L. (1993). The prevention of thymic lymphomas in transgenic mice by human O⁶-alkylguanine-DNA alkyltransferase. *Science* 259: 219–222.
- Fang, Q., Noronha, A.M., Murphy, S.P., Wilds, C.J., Tubbs, J.L., Tainer, J.A., Chowdhury, G., Guengerich, F.P., and Pegg, A.E. (2008). Repair of O⁶-G-alkyl-O⁶-G interstrand cross-links by human O⁶-alkylguanine-DNA alkyltransferase. *Biochemistry* 47: 10892–10903.
- Fernández, I., Uggerud, E., and Frenking, G. (2007). Stable pentacoordinate carbocations: structure and bonding. *Chemistry* 13: 8620–8626.
- Fu, D., Calvo, J.A., and Samson, L.D. (2012). Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat. Rev. Canc.* 12: 104–120.
- Fu, Y., Bernasconi, L., and Lu, P. (2021). Ab initio molecular dynamics simulations of the S_N1/S_N2 mechanistic continuum in glycosylation reactions. *J. Am. Chem. Soc.* 143: 1577–1589.

- Gerson, S.L. (2002). Clinical relevance of MGMT in the treatment of cancer. *J. Clin. Oncol.* 20: 2388–2399.
- Harris, L.C., Potter, P.M., and Margison, G.P. (1992). Site directed mutagenesis of two cysteine residues in the *E. coli* OGT O⁶-alkylguanine DNA alkyltransferase protein. *Biochem. Biophys. Res. Commun.* 187: 425–431.
- Hu, J., Ma, A., and Dinner, A.R. (2008). A two-step nucleotide-flipping mechanism enables kinetic discrimination of DNA lesions by AGT. *Proc. Natl. Acad. Sci. U. S. A.* 105: 4615–4620.
- Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B.L., Grubmüller, H., and MacKerell, A.D. (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* 14: 71–73.
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14: 27–28.
- Jackson, S.P. and Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature* 461: 1071–1078.
- Kaina, B., Christmann, M., Naumann, S., and Roos, W.P. (2007). MGMT: key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. *DNA Repair* 6: 1079–1099.
- Kelley, M.R. and Fishel, M.L. (2008). DNA repair proteins as molecular targets for cancer therapeutics. *Anti Canc. Agents Med. Chem.* 8: 417–425.
- Kumar, S., Rosenberg, J.M., Bouzida, D., Swendsen, R.H., and Kollman, P.A. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13: 1011–1021.
- Kyrtopoulos, S.A., Anderson, L.M., Chhabra, S.K., Souliotis, V.L., Pletsas, V., Valavanis, C., and Georgiadis, P. (1997). DNA adducts and the mechanism of carcinogenesis and cytotoxicity of methylating agents of environmental and clinical significance. *Canc. Detect. Prev.* 21: 391–405.
- Lee, J., Cheng, X., Swails, J.M., Yeom, M.S., Eastman, P.K., Lemkul, J.A., Wei, S., Buckner, J., Jeong, J.C., Qi, Y., et al. (2016). CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theor. Comput.* 12: 405–413.
- Lindahl, T. and Barnes, D.E. (2000). Repair of endogenous DNA damage. *Cold Spring Harbor Symp. Quant. Biol.* 65: 127–133.
- Liu, L., Qin, X., and Gerson, S.L. (1999). Reduced lung tumorigenesis in human methylguanine DNA-methyltransferase transgenic mice achieved by expression of transgene within the target cell. *Carcinogenesis* 20: 279–284.
- Luzhkov, V.B. (2017). Molecular modelling and free-energy calculations of protein–ligand binding. *Russ. Chem. Rev.* 86: 211.
- Magnen, M., Elsasser, B.M., Zbodakova, O., Kasperek, P., Gueugnon, F., Petit-Courty, A., Sedlacek, R., Goettig, P., and Courty, Y. (2018). Kallikrein-related peptidase 5 and seasonal influenza viruses, limitations of the experimental models for activating proteases. *Biol. Chem.* 399: 1053–1064.
- Maser, R.S. and DePinho, R.A. (2002). Connecting chromosomes, crisis, and cancer. *Science* 297: 565–569.
- Mattosovich, R., Merlo, R., Miggiano, R., Valenti, A., and Perugino, G. (2020). O⁶-alkylguanine-DNA alkyltransferases in microbes living on the edge: from stability to applicability. *Int. J. Mol. Sci.* 21: 2878.
- McKeague, M., Otto, C., Raz, M.H., Angelov, T., and Sturla, S.J. (2018). The base pairing partner modulates alkylguanine alkyltransferase. *ACS Chem. Biol.* 13: 2534–2541.
- Musarrat, J., Wilson, J.A., Abouissa, H., and Wani, A.A. (1995). O⁶-alkylguanine DNA alkyltransferase activity levels in normal, benign and malignant human female breast. *Biochem. Biophys. Res. Commun.* 208: 688–696.
- Nakatsuru, Y., Matsukuma, S., Nemoto, N., Sugano, H., Sekiguchi, M., and Ishikawa, T. (1993). O⁶-methylguanine-DNA methyltransferase protects against nitrosamine-induced hepatocarcinogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 90: 6468–6472.
- Patra, N., Ioannidis, E.I., and Kulik, H.J. (2016). Computational investigation of the interplay of substrate positioning and reactivity in catechol O-methyltransferase. *PLoS One* 11: e0161868.
- Paulsen, M. and Ferguson-Smith, A.C. (2001). DNA methylation in genomic imprinting, development, and disease. *J. Pathol.* 195: 97–110.
- Perugino, G., Vettone, A., Illiano, G., Valenti, A., Ferrara, M.C., Rossi, M., and Ciaramella, M. (2012). Activity and regulation of archaeal DNA alkyltransferase: conserved protein involved in repair of DNA alkylation damage. *J. Biol. Chem.* 287: 4222–4231.
- Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26: 1781–1802.
- Pokhrel, R., Pavadai, E., Gerstman, B.S., and Chapagain, P.P. (2019). Membrane pore formation and ion selectivity of the Ebola virus delta peptide. *Phys. Chem. Chem. Phys.* 21: 5578–5585.
- Rasimas, J.J., Pegg, A.E., and Fried, M.G. (2003). DNA-binding mechanism of O⁶-alkylguanine-DNA alkyltransferase. Effects of protein and DNA alkylation on complex stability. *J. Biol. Chem.* 278: 7973–7980.
- Rossi, F., Morrone, C., Massarotti, A., Ferraris, D.M., Valenti, A., Perugino, G., and Miggiano, R. (2018). Crystal structure of a thermophilic O⁶-alkylguanine-DNA alkyltransferase-derived self-labeling protein-tag in covalent complex with a fluorescent probe. *Biochem. Biophys. Res. Commun.* 500: 698–703.
- Sakumi, K., Shiraishi, A., Shimizu, S., Tsuzuki, T., Ishikawa, T., and Sekiguchi, M. (1997). Methylnitrosourea-induced tumorigenesis in MGMT gene knockout mice. *Canc. Res.* 57: 2415–2418.
- Sugita, Y., Kitao, A., and Okamoto, Y. (2000). Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* 113: 6042–6051.
- Tessmer, I. and Fried, M.G. (2014). Insight into the cooperative DNA binding of the O(6)-alkylguanine DNA alkyltransferase. *DNA Repair* 20: 14–22.
- Tessmer, I., Melikishvili, M., and Fried, M.G. (2012). Cooperative cluster formation, DNA bending and base-flipping by O⁶-alkylguanine-DNA alkyltransferase. *Nucleic Acids Res.* 40: 8296–8308.
- Tiwari, P.B., Chapagain, P.P., Banda, S., Darici, Y., Üren, A., and Tse-Dinh, Y.-C. (2016). Characterization of molecular interactions between *Escherichia coli* RNA polymerase and topoisomerase I by molecular simulations. *FEBS Lett.* 590: 2844–2851.
- Tiwari, P.B., Chapagain, P.P., Seddek, A., Annamalai, T., Üren, A., and Tse-Dinh, Y.-C. (2020). Covalent complex of DNA and bacterial topoisomerase: implications in antibacterial drug development. *ChemMedChem* 15: 623–631.
- Tiwari, P.B., Chapagain, P.P., and Üren, A. (2018). Investigating molecular interactions between oxidized neuroglobin and cytochrome c. *Sci. Rep.* 8: 10557.
- Tubbs, J.L., Pegg, A.E., and Tainer, J.A. (2007). DNA binding, nucleotide flipping, and the helix-turn-helix motif in base repair by O⁶-

- alkylguanine-DNA alkyltransferase and its implications for cancer chemotherapy. *DNA Repair* 6: 1100–1115.
- Warren, J.J., Forsberg, L.J., and Beese, L.S. (2006). The structural basis for the mutagenicity of O⁶-methyl-guanine lesions. *Proc. Natl. Acad. Sci. U. S. A.* 103: 19701–19706.
- Zak, P., Kleibl, K., and Laval, F. (1994). Repair of O⁶-methylguanine and O⁴-methylthymine by the human and rat O⁶-methylguanine-DNA methyltransferases. *J. Biol. Chem.* 269: 730–733.
- Zang, H., Fang, Q., Pegg, A.E., and Guengerich, F.P. (2005). Kinetic analysis of steps in the repair of damaged DNA by human O⁶-alkylguanine-DNA alkyltransferase. *J. Biol. Chem.* 280: 30873–30881.
-
- Supplementary Material:** The online version of this article offers supplementary material (<https://doi.org/10.1515/hsz-2021-0198>).



Diffusion of glucose in water: A molecular dynamics study

Rajendra Prasad Koirala, Sita Dawanse, Nurapati Pantha*

Central Department of Physics, Tribhuvan University, Kirtipur, Kathmandu, Nepal



ARTICLE INFO

Article history:

Received 21 July 2021

Revised 8 October 2021

Accepted 11 October 2021

Available online 16 October 2021

Keywords:

Glucose

Diffusion coefficient

Radial distribution function

Transport properties

Activation energy

ABSTRACT

A glucose molecule, simple structure of carbohydrate, has many functions in our body. It is transported to every cell and served as of energy source. The investigation of transport properties of glucose in aqueous medium would be useful to understand the body metabolism. In this research work, molecular dynamics (MD) simulations have been utilized to study the transport and structural properties of glucose in water at different temperature (298.15 K, 303.15 K, 308.15 K, and 312.15 K). The self diffusion coefficients of glucose and water have been estimated from mean square displacement plot (MSD) by using Einstein's relation. Further, Darken's relation has been used to find the binary diffusion coefficients of the mixture. The structural analysis of the system has been estimated with the help of radial distribution function (RDF) of its constituent atom/molecule pairs. The outcomes from the MD simulations depicts that the temperature dependence of self diffusion coefficients of solute and solvent follow Arrhenius behavior, which further assist to calculate the activation energy of the system. The results from the present work agree well with the previously reported experimental values.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Carbohydrates are the basic biomolecules that serve as the major source of energy in biological cells. They contribute as the essential structural components in tissues and have central role in nutrition [1,2]. These molecules perform the collaborative works with the other biomolecules like proteins, nucleic acids and lipids [3]. They are found in all types of organisms ranging from bacteria to whale and in plants as well [4]. The sugars (except the pentose sugar in nucleic acids), starch and cellulose belong in this molecular category [5].

The molecular structural of a carbohydrate molecule is represented by $C_m(H_2O)_n$ with same or different numerical values of m and n . In this research work, we have taken the condition of $m = n = 6$ that gives a single glucose molecule ($C_6H_{12}O_6$) with a subcategory of carbohydrate [6,7]. A glucose molecule has molecular weight $180.16 \text{ g mol}^{-1}$ with approximate density of 1.54 g cm^{-3} . It is most abundant form of monosaccharide ($C_nH_{2n}O_n$). It can remain in the acyclic as well as cyclic form [8]. In our work, we have taken the cyclic (ring form) as shown in Fig. 1. It travels in body cells through the bloodstream.

The food we eat is the major source of glucose [9]. It is received from plants in the form of complex molecular structures. These complicated molecules are dissolved by the digestive juices in

the digestive tracts and a glucose molecule (a type of monosaccharide molecule), a subcategory of carbohydrate, is formed. Then, it dissolves in the blood stream and transported into the various cells of body tissues [10,11]. As the transportation of glucose takes the path of our blood stream to our cells, it is commonly called the blood glucose or blood sugar [5]. Insulin hormone is responsible in the movement of glucose. Insufficient insulin production may cause the health problems [12]. Glucose is to be transported into every cell of the body. Many organs within our body perform works without interruption throughout the life [13]. Brain cells consume glucose in high rate, since it controls every automatic system in the body. Liver, heart, kidney and intestine perform their work even during the sleeping period. Glucose provides energy for their regular work [14,15]. Besides such essential functions, unnecessary deposition of glucose in body cell causes health complications. Diabetes is one of the serious health problems due to the excessive flow of glucose in blood stream. In the worst condition, immoderate concentration of glucose in body cell causes the cancer [16].

Physiological process influences when the blood glucose decreases below normal level. The consequences can be coma, permanent brain damage, trouble talking, confusion, seizure or death [17]. Lack of proper concentration of glucose in tissues causes the serious physiological consequences like kidney failure, cardiovascular diseases, blindness and many more [18]. Determining the diffusivity and permeability coefficient are very notable for diagnosis and therapy of various human diseases [19]. Furthermore, several

* Corresponding author.

E-mail address: mnrurapati@gmail.com (N. Pantha).

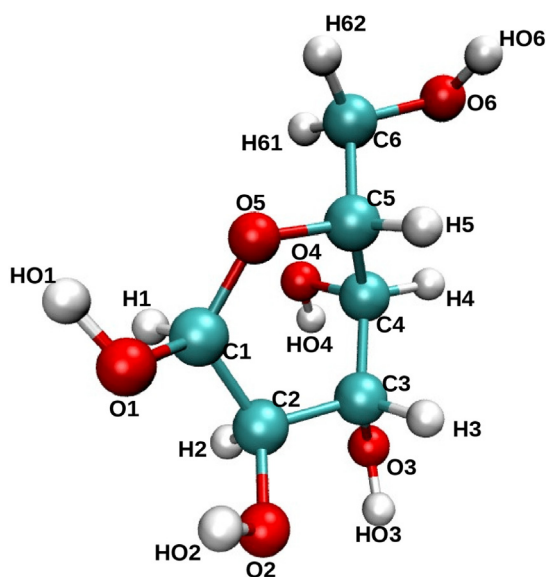


Fig. 1. Schematic representation of a glucose molecule. Its five hydroxyl (OH) groups are arranged in a specific way along its six-carbon back.

research studies show that hyperglycemia is associated with COVID-19 mortality [20].

Glucose works as the body fuel in order to carry out the cell metabolism in correct way. It has the significant contribution in energy storage at different parts of the body [21]. It stores the energy in the form of glycogen, which is essential in driving the internal body parts like liver, kidneys, muscles, lungs, and many parts of nervous system. In addition to the energy storage function, it is an important substrate in the biochemical reaction and also acts as the supplementary stuffs of proteins and fats as a build material [22]. It is transported throughout the body to perform such important functions [23]. Motivated by these important functions of glucose in various tissues, we have investigated transport property of glucose within aqueous environment. We have performed molecular dynamics simulations to examine diffusion coefficient of glucose in water. We believe that this work provides an alternative method of study on transport property of glucose in living body.

The structure of this article is so organized that immediately after this part, the theoretical background of "Diffusion Coefficient" is presented. Then, molecular model, its system setup and simulation procedures are explained in the "Methods and Methodology" section. The outcomes of entire work are presented, analyzed and discussed in the "Results and Discussion" section. Conclusions and possible extension of the work is briefly discussed in "Conclusions and Concluding Remarks section".

1.1. Diffusion coefficient

Several food substrates are transported randomly (i.e., diffused) throughout our body via water medium. Glucose, as a the substrate in the body, is to be approached in every cell to burn with oxygen, which is essential to generate the energy. The diffusion of such substrate molecules is remarkably affected by temperature variation and concentration inhomogeneity [24,25]. In cellular environment, a small variation in blood temperature may affect the significant variation of diffusion of molecular contents [26,27]. Diffusion coefficient is utilized to measure the transportation phenomenon of substrate in the aqueous medium. The coefficient can be taken within the medium of identical molecules or in the different molecules.

Self-diffusion depicts the migration of a molecule within the group of identical molecules such as the diffusion of water molecule within the water environment [28]. The self-diffusion coefficient is calculated by Einstein's relation and can be utilized to estimate the self-diffusion coefficient of a substance by using mean square displacement (MSD) plot against time (t). For three dimensional system, self diffusion coefficient is deduced from mean square displacement [25,29] and is written as,

$$D = \lim_{t \rightarrow \infty} \frac{\langle [\mathbf{r}(t) - \mathbf{r}(0)]^2 \rangle}{6t} \quad (1)$$

where, $[\mathbf{r}(t) - \mathbf{r}(0)]$ represents the displacement of particle from reference point at time t . The symbol $\langle \dots \rangle$ indicates the ensemble average. The ensemble average is taken over all atoms of the component in the simulation and all time origins [30]. In Eq. (1), $[\mathbf{r}(t) - \mathbf{r}(0)]^2$ gives the MSD of diffusing particle and the best fitted line for MSD versus time gives the straight line, which is employed to calculate self diffusion coefficient (D).

The diffusion of a substance in another substance is, of course, a binary or mutual diffusion and is quantitatively studied in terms of binary diffusion coefficient [31]. In our work, we have calculated the self diffusion coefficients of glucose (solute) and water (solvent) separately by using Einstein's relation and then, these calculated self-diffusion coefficients were employed to estimate binary diffusion coefficient using Darken's phenomenological relation [32],

$$D_{AB} = N_B D_A + N_A D_B \quad (2)$$

where, D_{AB} is the binary diffusion coefficient, D_A is self-diffusion coefficient of species A , D_B is self diffusion coefficient of species B , N_A is the mole fraction of species A and N_B is the mole fraction of species B .

2. Methods and methodology

This section incorporates basic concepts on modeling of glucose molecule, molecular system setup in aqueous environment and simulation procedure.

2.1. Molecular models

Glucose molecule ($C_6H_{12}O_{12}$) is electronically neutral. It contains aldehyde group (CH_2OH) and double bonded oxygen, which are covalently bonded with 6C atoms. Of course, the different atoms present in this molecule contain different partial charges, though it is electrically neutral. So, total potential energy of the system comprises both bonded and non-bonded interactions. Bonded interaction includes bond length stretching, bond angle variation and dihedral (improper and proper) angle, which contribute in bonded potential. Besides, bonded interactions, non-bonded interactions also play crucial role to provide the potential energy in the molecule. Non-bonded interaction are coulomb potential and Lennard-Jones potential that are pairwise additive [29]. These potentials are essential to form a stable glucose molecule.

2.2. System setup and simulations

Three glucose molecules are extracted from protein data bank with PDB ID 4I3U [33]. There were several other molecules exist in the 4I3U complex, we have chosen the glucose molecules as our target of investigation. Then, these glucose molecules were solvated in a cubical box of dimensions $4.5 \times 4.5 \times 4.5 \text{ nm}^3$ having 1670 number of water molecules as shown in Fig. 2. The extended simple charge (SPC/E) [34] water was used. No ions were added,

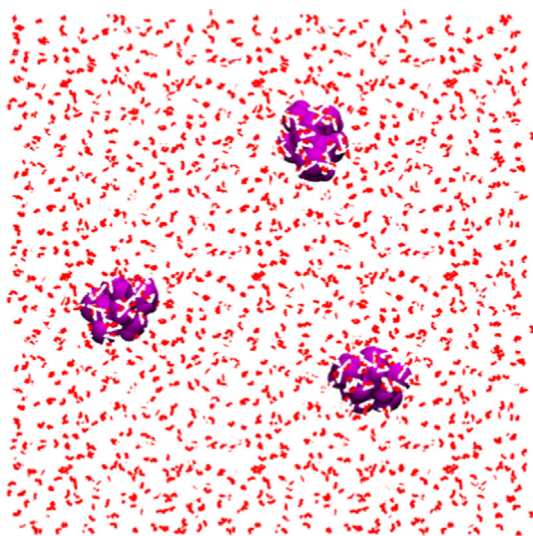


Fig. 2. Structure of simulation box with molecules after energy minimization.

since the system was already chargeless. The bonded interaction: angles, bond length, dihedrals as well as non bonded interaction: partial charges, sigma epsilon, are parameterized in the Optimized Potentials for Liquid Simulations-All Atom (OPLS-AA) force fields by default. We performed the molecular dynamics (MD) simulations to estimate the transport properties of glucose. GROMINGEN MACHINE for Chemical Simulations 5.1.1 (GROMACS 5.1.1) software program [35] was used for the MD simulations. The simulations were performed at four different temperatures; 298.15 K, 303.15 K, 308.15 K, and 312.15 K. OPLS-AA force field parameters were used for entire simulations work. The MD simulations were performed in three steps: energy minimization run, equilibration run and production run. The PDB structure of the glucose molecules that we have taken were modeled from X-ray diffraction (XRD) technology. The coordinates taken from this method may not be precisely accurate so that the molecule may not be at the minimum potential energy state. Moreover, PDB structure does not contain the hydrogen atoms, they are to be filled with the software program and are the sole guessed coordinates. These guessed coordinates also produce the high potential energy [36]. To eliminate all such types of errors in coordinates, energy minimization run was performed.

To restore the molecular system in cellular environment, we have propagated the equilibration run. This simulation run maintains the system in suitable temperature and pressure [37]. Pressure coupling was done using Berendsen barostat to pressure 1 bar. $\tau_t = 0.01$ ps and $\tau_p = 0.8$ ps was taken to maintain constant temperature and pressure coupling time. The initial velocity was generated according to Maxwell-Boltzmann distribution function at various temperature. Particle Mesh Ewald (PME) type is used for long range coulomb interaction with fourier spacing of 0.12 nm with cut off distance of 1.0 nm [29,30]. Then, each system was carried out production run 100 ns under NVT condition. The outputs obtained from the simulations were visualized and analyzed by using analysis tools in visual molecular dynamics (VMD) [38]. Xmgrace was used to plot the graphs.

3. Results and discussion

In this work, we have carried out molecular dynamics simulations to estimate the bonded and non-bonded energy profile of the glucose molecule. Furthermore, we have estimated self diffusion coefficient of water and glucose, and then binary diffusion coefficient as well as structural properties of glucose molecules in water.

3.1. Energy minimization and equilibration

The molecule taken for the consideration is usually in higher energy state. In the beginning of the MD simulations, the potential energy of the molecule should be drawn at the local minimum energy state to ensure that the system for simulation was in one of the local minima. If the molecule at higher energy state is directly carried out for the production run, system may not be energetically converged, which might mislead the outcomes from the actual result. Therefore, we have performed the energy minimization run so that the steric hindrance among the atoms within the molecules could be removed.

After energy minimization, equilibration run was performed for 100 ns under NPT condition. Equilibration run could maintain the temperature and pressure at the realistic condition that can mimic the real condition of the living cell. We maintained the pressure and temperature of the systems at our targeted values. In addition, we have estimated the density of the system for all four different temperature. The density of the system was found comparable value of water, since most of the part is occupied with water. The density of the system was observed decreasing as the temperature increasing.

Table 1 lists the equilibrium temperatures and densities of the system at respective coupling temperature.

3.2. Energy profile

Bonded and non-bonded interactions contribute to form a stable glucose molecule. The bond stretching, angle variation, and dihedral angle change exist in bonded interaction. This type of interaction is contributed by chemical bonds, specifically the covalent bonds in glucose molecule. Besides, these bonded interactions, both short range and long range non-bonded interactions provide leading role to form the molecule. Thus, total potential energy is the sum of bonded and non-bonded potential energies. The molecule at temperature greater than absolute zero possesses certain kinetic energy. Therefore, total energy of the system is the sum of the total kinetic energy and the potential energy of all the atoms. From the Fig. 3, the energies due to bonded interactions have relatively small positive values in comparison to overall energy profile. Likewise, Lennard-Jones (LJ) interaction energy is positive with an average of $(15324.64.20 \pm 1.55)$ kJ mol⁻¹. The coulomb potential energy is negative with value of (-93539.45 ± 1.47) kJ mol⁻¹. Hence, the total potential energy was found to be negative with the value (-76526.90 ± 1.10) kJ mol⁻¹. This illustrates that coulomb interaction has the dominating effect in overall contributions of bonded and non-bonded interactions. Further, the value of kinetic energy was observed (12594.50 ± 0.16) kJ mol⁻¹. Even though, the kinetic energy is also positive, the total energy of the system is negative because of the larger negative magnitude of potential energy. As a result, the total energy is (-63932.40 ± 0.13) kJ mol⁻¹. This indicates that our system is bound and in stable equilibrium.

3.3. Self diffusion coefficient of glucose

The self diffusion coefficients of glucose molecule have been estimated for four different temperature. The value of self diffusion coefficient was calculated in according to the Einstein's equation as given in Eq. 1. Although the production run was done for 100 ns, the MSD curves were plotted for 6 ns to get the better statistics and to obtain linear nature of MSD curve. Fig. 4 depicts the MSD versus time plot of glucose for 6 ns with their respective linear fit at different temperature. It is found that, slope of MSD increases with increase in temperature, which in turn implies that self diffusion coefficient increases with increase in temperature. The self diffusion coefficient of glucose molecule at different temperature has been presented in Table 2.

Table 1
Temperature and Density of system after equilibration run.

System	Coupling Temperature (K)	Equilibrium Temperature (K)	Equilibrium Density (kgm^{-3})	Reference Density (kgm^{-3}) [39]
I	298.15	298.150 \pm 0.007	995.621 \pm 0.044	997.040
II	303.15	303.140 \pm 0.006	992.975 \pm 0.033	995.650
III	308.15	312.141 \pm 0.009	990.200 \pm 0.036	994.049
IV	312.15	312.139 \pm 0.007	987.001 \pm 0.046	992.597

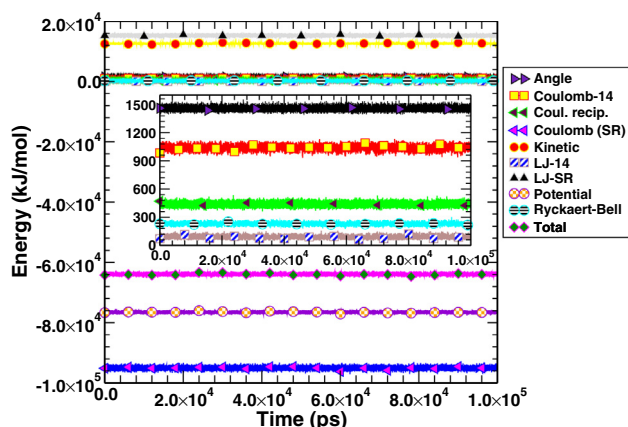


Fig. 3. The energy profile of the system at $T = 298.15$ K.

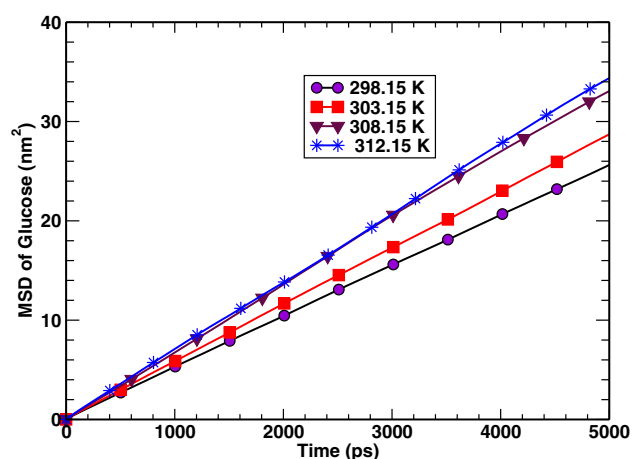


Fig. 4. MSD plots for glucose obtained from simulation and their linear fit at different temperatures.

Table 2
Self-diffusion coefficient of glucose at different temperature obtained from simulation.

Temperature (K)	Self Diffusion of glucose ($D_{Glc}^{self} \times 10^{-9}$) $\text{m}^2 \text{s}^{-1}$
298.15	0.84790 \pm 0.00012
303.15	0.95378 \pm 0.00011
308.15	1.09517 \pm 0.00004
312.15	1.13171 \pm 0.00002

3.4. Self diffusion coefficient of water

We have estimated self diffusion coefficient of water at different temperature using the same approach of Einstein's relation. Even though several researches have been carried out to estimate the self diffusion coefficient of water, we also determined here to validate our results for glucose molecule [40,41]. Moreover, we need self diffusion coefficient of both glucose and water to find the mutual diffusion of glucose in water. Fig. 5 shows the MSD

plots of water at different temperature for 6 ns with the corresponding linear fit. From Fig. 5, it is lucid that slope of MSD plot increases with increase in temperature, it successively infers that self diffusion coefficient of water increases with increase in temperature. The self diffusion coefficient of water obtained from the simulation was also compared with the experimental value. As the simulation and experimental values in water were found comparable, this also validated the result of self diffusion coefficients of glucose. The self diffusion coefficient of water at different temperature are also presented in Table 3. The percentage error of experimental and simulated results are also mentioned in the Table 3. The result of simulated value is compared with the experimental paper and are in good agreement within 8.5% error.

3.5. Binary diffusion coefficient

The binary diffusion coefficient is calculated using Darken's relation and is close to self diffusion coefficient of solute due to low concentration of solute in the solution. We have simulated 3 glucose molecules and 1670 water molecules in a cubical simulation box, that contains total 1673 molecules. Therefore, the mole fraction of glucose molecule is 0.00179 and that of water molecule is 0.99821. Here, mole fraction of water dominates at all, from the Darken's relation the binary diffusion coefficient of glucose becomes very much close to the self-diffusion coefficient of glucose. Table 4 represents the simulated values of binary diffusion coefficient of glucose-water system at different temperatures. The simulated results in this work are compared with the experimentally reported value and it is found that the results of the simulation are within the reliable error range and the maximum deviation is about 31.8%.

3.6. Temperature dependence of diffusion

The movement of molecules in a system depends on its temperature. As the temperature increases, kinetic energy of the molecules increases, which ultimately enhances the molecular diffusion in the same medium or in the mixture. This temperature

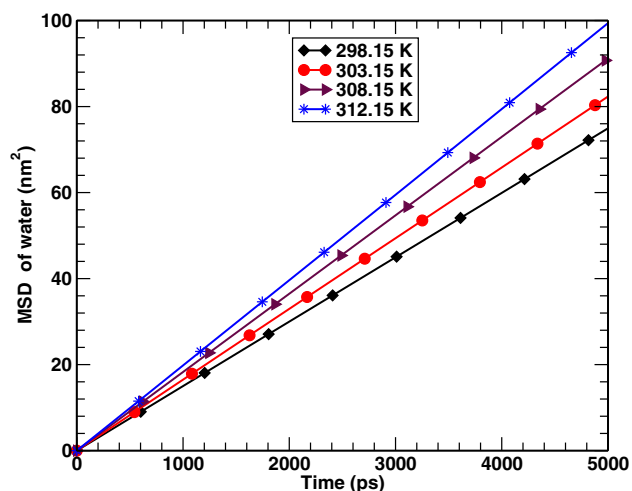


Fig. 5. Linear fit of MSD of water at different temperature.

Table 3
Self-diffusion coefficient of water at different temperature obtained from simulation.

Temperature (K)	Experimental ($D_{\text{wat}}^{\text{Exp}} \times 10^{-9}$) m^2s^{-1}	Simulated ($D_{\text{wat}}^{\text{Sim}} \times 10^{-9}$) m^2s^{-1}	% Error
298.15	2.299 [42]	2.49633 ± 0.00003	8.50
303.15	2.597 [42]	2.74113 ± 0.00004	5.54
308.15	2.895 [42]	3.03713 ± 0.00002	4.91
312.15	3.183 [43]	3.28914 ± 0.00011	3.33

Table 4
Binary-diffusion coefficient of glucose in water at different temperature obtained from simulation.

Temperature (K)	Experimental [44] ($D_{\text{wat}}^{\text{Exp}} \times 10^{-9}$) m^2s^{-1}	Simulated ($D_{\text{wat}}^{\text{Sim}} \times 10^{-9}$) m^2s^{-1}	error % Error
298.15	0.651 ± 0.001	0.85085 ± 0.00012	30.6
303.15	0.740 ± 0.007	0.95664 ± 0.00011	29.3
308.15	0.832 ± 0.006	1.09865 ± 0.00003	31.8
312.15	0.924 ± 0.007	1.13557 ± 0.00002	22.8

dependent nature of the diffusion coefficient is formulated by the Arrhenius relation, [26].

$$D = D_0 \exp\left(\frac{-E_a}{N_A k_B T}\right) \quad (3)$$

where, D_0 is the pre-exponential factor also called frequency factor, E_a is the activation energy for diffusion, T is the absolute temperature, N_A is the Avogadro's number, and k_B is the Boltzmann constant. D_0 and E_a are called the activation parameters of diffusion. The activation energy of diffusion process is given by the Eq. 4.

$$E_a = -N_A k_B \frac{\partial \ln D}{\partial (1/T)} \quad (4)$$

The relation in Eq. 4 corresponds to negative slope of the Arrhenius diagram, which is a plot between $\ln(D)$ and reciprocal of absolute temperature. The intercept of extrapolated Arrhenius line $\frac{1}{T} \rightarrow 0$, yields the pre-exponential factor D_0 . In this work, activation energy evaluated from Arrhenius relation that gives the minimum possible energy required to diffuse itself (self-diffusion) and in water (binary diffusion) of glucose molecule. Fig. 6 shows the Arrhenius diagram for self-diffusion coefficient of glucose. This diagram is very important to evaluate the activation energy of the molecules in a system. Activation energy gives the minimum energy required to proceed the reaction. The value of this energy relies on the energy barriers encounter during the reaction process. The activation energy for self-diffusion coefficient of glucose estimated from the slope is equal to $16.74 \text{ kJ mol}^{-1}$. In the molecular system which we have taken for the consideration contains maximum part of water. Considering the water as the reference molecule to examine the validity of the result regarding the activation energy of glucose, we also estimated the activation energy of water. Significant correlation of estimated value of activation energy with the experimental value ensures the validity of our result in case of glucose molecule as same system was taken for the simulations. Fig. 7 shows the Arrhenius diagram for self-diffusion coefficient of water. The activation energy for self-diffusion coefficient of water estimated from the slope is equal to $15.46 \text{ kJ mol}^{-1}$.

After the estimation of activation energy of glucose and water, we deduced the binary diffusion coefficient of glucose in water at different temperature. The estimated values at various temperature have been compared with the experimental value. Fig. 8 shows the relation of binary diffusion coefficient of water-glucose system with temperature. Also, the activation energy of glucose in water deals the energy required to diffuse glucose in aqueous environment. The activation energy for binary-diffusion

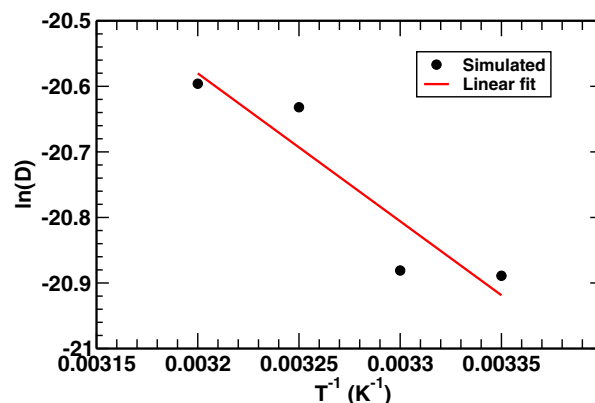


Fig. 6. Arrhenius plot of the self-diffusion coefficient of glucose.

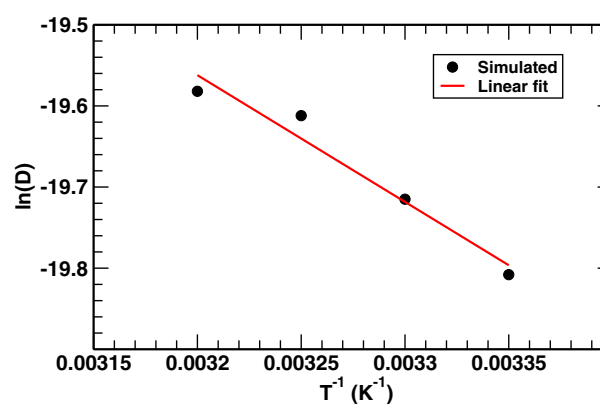


Fig. 7. Arrhenius plot of the self-diffusion coefficient of water.

coefficient of water-glucose mixture estimated from the slope is equal to $16.78 \text{ kJ mol}^{-1}$ and presented in Table 5. The simulated value of activation energy for binary mixture of glucose agreed with the experimental value within 5.35% error. The estimated value from simulation has the good agreement with experimental value within an error of 13.10%. Such error occurred in water might be due to the fact that our system consists of water and glucose molecules, whereas the experimental result was recorded for pure water.

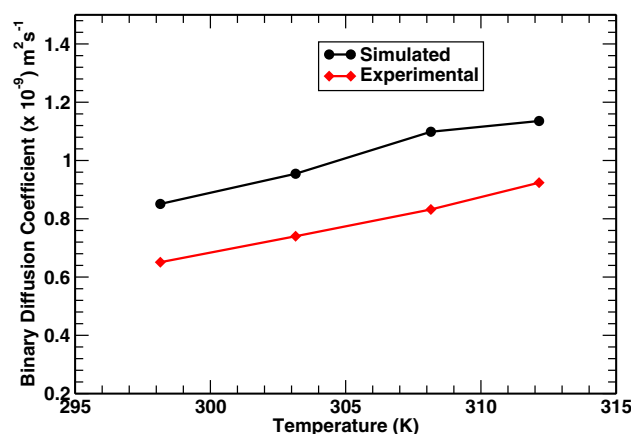


Fig. 8. Arrhenius plot of the binary diffusion coefficient of glucose in water obtained from simulation and experiment [44].

Table 5
Activation energy of the system.

System	Activation Energy (E_a) kJmol^{-1}		Percentage Error
	Simulated	Experimental	
Water	15.46	17.79 [45]	13.10
Glucose	16.74		
Binary mixture	16.78	17.75 [44]	5.35

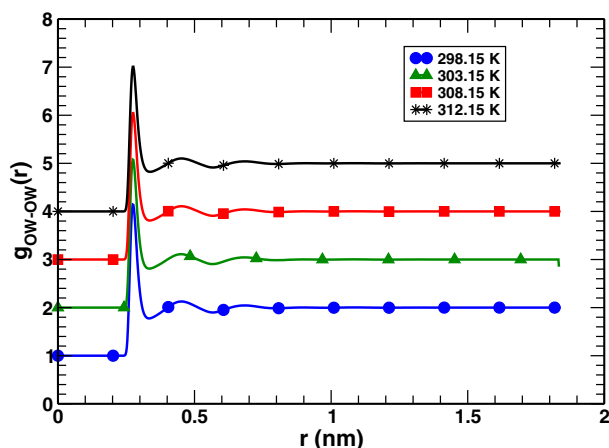


Fig. 9. Plot of RDF g_{OW-OW} at different temperatures.

3.7. Radial Distribution Function (RDF)

Pair correlation and structure of system was studied from radial distribution function (RDF). It provides the preferred position of one particle with respect to other particle. Structure of solvent can be interpreted using RDF as it depicts how density and probability vary with distance [46].

3.7.1. Radial distribution function of solvent

RDF of oxygen atom of water (OW) in reference to oxygen atom of another water molecule (OW) is presented in Fig. 9. The value of $g(r)$ is zero up to certain region from origin, that is called excluded region. It is the region, where the probability of finding particle with respect to the reference particle is zero. The value of σ for OW – OW is 0.32 nm, and its corresponding van der Waals radius is $2^{1/6}\sigma \approx 0.36$ nm. This implies that the atoms do not come closer to each other than van der Waals radius. The exclusion region is due to LJ interaction and repulsive coulomb interaction. From Table 6, it is found that value of excluded region is less than that of van der Waals radius and thus, it agrees with the theory that there is zero probability of finding particles in excluded region. In the Fig. 9, three distinct peaks of height 3.142, 1.127, and 1.045 corresponding to distance 0.274 nm, 0.452 nm, and 0.682 nm at temperature 298.15 K were observed. This explains radius of first coordinate shell of oxygen of water around reference oxygen of water is 0.274 nm. Beyond third peaks, the value of $g(r)$ was obtained unity indicating that there is no correlation between

Table 6
Details of simulation data for the RDF $g_{OW-OW}(r)$ at different temperatures.

Temperature (K)	ER (nm)	FPP (nm)	FPV	SPP (nm)	SPV	TPP (nm)	TPV
298.15	0.240	0.274	3.142	0.452	1.127	0.682	1.045
303.15	0.240	0.274	3.077	0.454	1.110	0.684	1.044
308.15	0.240	0.276	3.044	0.452	1.108	0.686	1.042
312.15	0.240	0.276	3.008	0.452	1.101	0.688	1.040

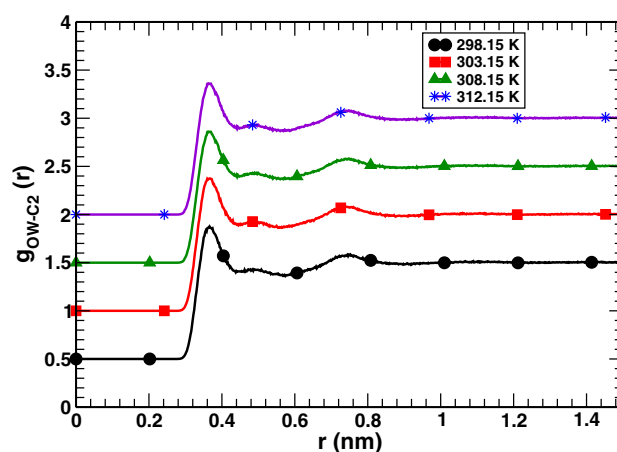


Fig. 10. Plot of RDF g_{OW-C2} at different temperature.

oxygen in water molecules. From the Table 6, it is clear that the magnitude of peak decreases but width increases with increase in temperature. Such effect in the graph accounts for the fact that thermal agitation increases with increase in temperature.

In Table 6, ER represents Excluded Region, FPP represents First Peak Position, FPPV represents First Peak Value, SPP represents Second Peak Position, SPV represents Second Peak Value, TPP represents Third Peak Position and TPV represents the Third Peak Value.

3.7.2. Radial distribution function of OW-C2

In the present work, C2 represents the carbon atom of the aldehyde group of the glucose molecule, and OW is the oxygen of the water molecule. Both C2 and OW are negatively charged; there exists repulsive coulomb interaction between them. The value of σ for OW – C2 is 0.350 nm, and its corresponding van der Waals radius is $2^{1/6}\sigma \approx 0.392$ nm. The increase in van der Waals radius between OW and C2 increases the ER, FPP and SPP in comparison to the g_{OW-OW} curves as represented in Fig. 9. RDF at different temperature is shown in Fig. 10 and the details of the curves are listed

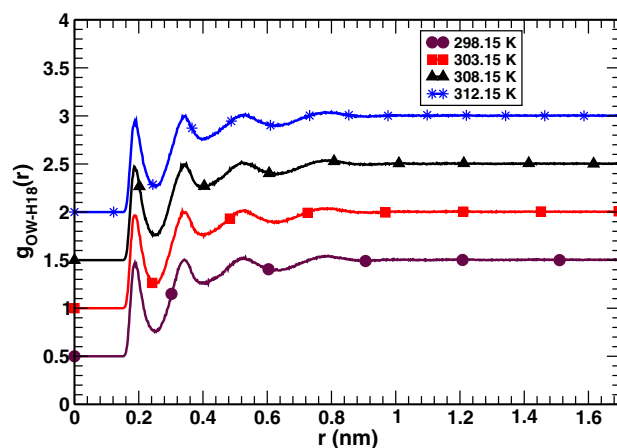


Fig. 11. Plot of RDF g_{OW-H18} at different temperature.

Table 7
Details of simulation data for the RDF $g_{C_2-OW}(r)$ at different temperature.

Temperature (K)	ER (nm)	FPP (nm)	FPV	SPP (nm)	SPV
298.15	0.280	0.366	1.376	0.732	1.072
303.15	0.280	0.366	1.375	0.742	1.082
308.15	0.280	0.368	1.360	0.748	1.081
312.15	0.280	0.364	1.359	0.750	1.076

Table 8
Details of simulation data for the RDF $g_{OW-H18}(r)$ at different temperature.

Temperature (K)	ER (nm)	FPP (nm)	FPV	SPP (nm)	SPV	TPP (nm)	TPV
298.15	0.150	0.188	0.971	0.344	1.002	0.526	1.024
303.15	0.150	0.190	0.967	0.346	0.999	0.530	1.016
308.15	0.150	0.190	0.962	0.348	0.998	0.532	1.009
312.15	0.150	0.192	0.946	0.348	0.996	0.534	1.006

in Table 7. After the second peak there is no prominent peak, and the curve is straight line beyond 1 nm, which implies that there is no pair correlation. From the Table 7, it is clear that the magnitude of peak decreases but width increases with increase in temperature. Such effect in the graph accounts for the fact that thermal agitation increases with increase in temperature. From the 7 it is clear that the excluded region is 0.280 nm which is less than van der Waals radius and thus, it agrees with the theory that there is zero probability of finding particles in excluded region.

3.7.3. Radial distribution function of OW-H18

In this part, H18 is the hydrogen of aldehyde group of glucose molecule and OW is the oxygen of the water molecule. H18 has partially positive charge, whereas OW is negatively charged, therefore highly attractive coulomb interaction exists between them. This leads to decrease in ER, FPP, SPP and TPP values in comparison to the g_{OW-OW} curves as presented in Fig. 9. The RDF curve at different temperature is illustrated in Fig. 11. At temperature 298.15 K the FPP is 0.188 nm with the corresponding peak value 0.971, the SPP is 0.344 nm with the corresponding peak value 1.002, and the TPP is 0.526 nm with the corresponding peak value 1.024. After the third peaks, there is no prominent peak and is almost straight line beyond 1 nm, which implies there is no pair correlation. Similar trend is followed by the RDF at 303.15 K, 308.15 K, and 312.15 K with peak position and peak height as listed in Table 8. The height of the peak is slightly decreasing and the width is increasing with increasing in temperature of the system. Table 8 shows that there is excluded region, which indicates that other factors rather than van der Waals interaction; like as Coulomb interactions, bonded interactions and many-body effects are highly dominated.

4. Conclusions and concluding remarks

This work has been carried out under classical molecular dynamics simulation in order to study the diffusion of binary mixture of 3 glucose molecules and 1670 water molecules at different temperature; 298.15 K, 303.15 K, 308.15 K and 312.15 K. The molar concentration fixes to 0.1 mol dm⁻³ to mimic the experimental concentration. Density variation with temperature and energy profiles of the system have been studied at different temperature mentioned above to confirm and test the equilibrium state of the system. Equilibrium structure of solute and solvent is studied via radial distribution function of different component of the system. The analysis of RDF plots at different temperature reveals that

the system becomes less organized with increase in temperature. Self diffusion coefficient of water is in good agreement with experimental data within 8.5% of deviation at most. The simulated value of binary diffusion coefficients of the systems are also in agreement with experimental data with maximum deviation of 31.8%. Such error is obtained due to different water model used in experimental and in simulation process and due to some interaction parameters present in simulation box. They follow the Arrhenius Behavior and the activation energy of water and the binary mixture (of water and glucose) calculated from the present work agree well with the available experimental values.

Through this study, we can infer that outcomes from the molecular dynamics simulation is reliable and is useful in the future study. In future, we intent to extend our research work to study other transport properties like viscosity, thermal conductivity at different temperature with different concentrations. Further more, this work can be extended by varying pressure, temperature, and solvent apart from water. In addition, we plan to study large and complex structure of carbohydrates in water and in blood to understand the biological phenomenon within the body.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

R. P. Koirala acknowledges the Nepal Academy of Science and Technology (NAST) and S. Dawanse acknowledges the University Grants Commission (UGC) for partial financial support. The authors acknowledge the University Grants Commission (UGC) Nepal, Award No. CRG-73/74-S & T-01 for computational facility and Prof. Dr. N. P. Adhikari for useful discussion.

References

- [1] J.B. German, C.J. Dillard, Composition, structure and absorption of milk lipids: a source of energy, fat-soluble nutrients and bioactive molecules, *Critical Rev. Food Sci. Nutr.* 46 (1) (2006) 57–92, <https://doi.org/10.1080/10408690590957098>.
- [2] Å. Krogdahl, G.I. Hemre, T.P. Mommsen, Carbohydrates in fish nutrition: digestion and absorption in postlarval stages, *Aquaculture Nutr.* 11 (2) (2005) 103–122, <https://doi.org/10.1111/j.1365-2095.2004.00327.x>.
- [3] J. Stenger, M. Cowman, F. Eggers, E.M. Eyring, Molecular dynamics and kinetics of monosaccharides in solution. A broadband ultrasonic relaxation study, *J. Phys. Chem. B* 104 (19) (2000) 4782–4790, <https://doi.org/10.1021/jp9940194>.

- [4] K.J. Niklas, *Plant Biomechanics: An Engineering Approach to Plant Form and Function*, University of Chicago Press, 1992.
- [5] B. Singh, T. Garg, A.K. Goyal, G. Rath, Development, optimization, and characterization of polymeric electrospun nanofiber: a new attempt in sublingual delivery of nicorandil for the management of angina pectoris, *Artif. Cells Nanomed. Biotechnol.* 44 (6) (2016) 1498–1507, <https://doi.org/10.3109/21691401.2014.937868>.
- [6] J.F. Robyt, F. John, *Essentials of Carbohydrate Chemistry*, Springer Science & Business Media, 1998.
- [7] P. Narayanan, *Essentials of Biophysics*, Anshan, 2010.
- [8] Y. Ho, *Patient-Specific Controller for an Implantable Artificial Pancreas*, Springer, New York, 2019.
- [9] W.R. Russell, A. Baka, I. Björck, N. Delzenne, D. Gao, H.R. Griffiths, et al., Impact of diet composition on blood glucose regulation, *Critical Rev. Food Sci. Nutrition.* 56 (4) (2016) 541–590, <https://doi.org/10.1080/10408398.2013.792772>.
- [10] G.E. Lienhard, J.W. Slot, D.E. James, M.M. Mueckler, How cells absorb glucose, *Scientific Am.* 266 (1) (1992) 86–91, <http://www.jstor.org/stable/24938903>.
- [11] M.S. McAllister, L. Krizanac-Bengez, F. Macchia, R.J. Naftalin, K.C. Pedley, M.R. Mayberg, et al., Mechanisms of glucose transport at the blood–brain barrier: an in vitro study, *Brain Res.* 904 (1) (2001) 20–30, [https://doi.org/10.1016/S0006-8993\(01\)02418-0](https://doi.org/10.1016/S0006-8993(01)02418-0).
- [12] C.S. Brennan, Molecular nutrition & food research, Dietary fibre, glycaemic response, and diabetes, *Mol. Nutr. Food Res.* 49 (6) (2005) 560–570, <https://doi.org/10.1002/mnfr.200500025>.
- [13] P. Mergenthaler, U. Lindauer, G.A. Dienel, A. Meisel, Sugar for the brain: the role of glucose in physiological and pathological brain function, *Trends Neurosci.* 36 (10) (2013) 587–597, <https://doi.org/10.1016/j.tins.2013.07.001>.
- [14] J.J. Harris, R. Jolivet, D. Attwell, Synaptic energy use and supply, *Neuron* 75 (5) (2012) 762–777, <https://doi.org/10.1016/j.neuron.2012.08.019>.
- [15] R.S. Haber, S.P. Weinstein, E. O'Boyle, S. Morgello, Tissue distribution of the human GLUT3 glucose transporter, *Endocrinology* 132 (6) (1993) 2538–2543, <https://doi.org/10.1210/en.132.6.2538>.
- [16] B.C. Martin, J.H. Warram, A.S. Krolewski, J.S. Soeldner, C.R. Kahn, R.N. Bergman, Role of glucose and insulin resistance in development of type 2 diabetes mellitus: results of a 25-year follow-up study, *The Lancet.* 340 (8825) (1992) 925–929.
- [17] J. Bousquière, C. Michon, C. Bonazzi, *Food Hydrocolloids* 70 (2017) 304–312, <https://doi.org/10.1016/j.foodhyd.2017.04.010>.
- [18] T.S. Altannavch, K. Roubalova, P. Kucera, M. Andel, Effect of High Glucose Concentrations on Expression of ELAM-1, VCAM-1 and ICAM-1 in HUVEC with and without Cytokine Activation, *Physiol. Res.* 53 (1) (2004) 77–82.
- [19] I. Kushner, J.A. Omerville, Permeability of human synovial membrane to plasma proteins. Relationship to molecular size and inflammation, *Arthritis Rheumat.: Off. J. Am. College Rheumatol.* 14 (5) (1971) 560–570, <https://doi.org/10.1002/art.1780140503>.
- [20] P. Yang, N. Wang, J. Wang, A. Leo, Y. Tu, Admission fasting plasma glucose is an independent risk factor for 28-day mortality in patients with COVID-19, *J. Medical Virol.* 93 (4) (2021) 2168–2176, <https://doi.org/10.1002/jmv.26608>.
- [21] E.D. Rosen, B.M. Spiegelman, Adipocytes as regulators of energy balance and glucose homeostasis, *Nature.* 444 (7121) (2006) 847–853, <https://doi.org/10.1038/nature05483>.
- [22] N.N. Potter, J.H. Hotchkiss, *Food Science*, Springer Science & Business Media, 2012.
- [23] J. Dupont, R.J. Scaramuzzi, Insulin signalling and glucose transport in the ovary and ovarian function during the ovarian cycle, *Biochem. J.* 473 (11) (2016) 1483–1501, <https://doi.org/10.1042/BCJ20160124>.
- [24] J. Crank, *The Mathematics of Diffusion*, 2nd ed., Oxford University Press, Oxford, 1975.
- [25] D. Frenkel, B. Smit, *Understanding Molecular Simulations from Algorithms to Applications*, second ed., Academic Press, 2002.
- [26] H. Mehrer, *Diffusion in Solids Fundamental, Method, Material, Diffusion controlled process*, Springer, Berlin, 2007.
- [27] R.P. Koirala, H.P. Bhusal, S.P. Khanal, N.P. Adhikari, Effect of temperature on transport properties of cysteine in water, *AIP Adv.* 10 (2) (2020) 025122, <https://doi.org/10.1063/1.5132777>.
- [28] H. Hirakawa, Y. Kamei, M. Sugisaki, Y. Oishi, Relationship between Self-Diffusion and Interdiffusion in Gaseous Systems, *Bull. Chem. Soc. Japan* 46 (9) (1973) 2659–2662, <https://doi.org/10.1246/bcsj.46.2659>.
- [29] M.P. Allen, D.J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, USA, 1989.
- [30] S. Pokharel, N. Pantha, N.P. Adhikari, Diffusion coefficients of nitric oxide in water: A molecular dynamics study, *Int. J. Mod. Phys. B* 30 (27) (2016) 1650205, <https://doi.org/10.1142/S0217979216502052>.
- [31] G.S. Hartley, J.F. Crank, Some fundamental definitions and concepts in diffusion processes, *Trans. Faraday Soc.* 45 (1949) 801–818, <https://doi.org/10.1039/TF9494500801>.
- [32] L.S. Darken, Diffusion, mobility and their interrelation through free energy in binary metallic systems, *Trans. AIME.* 175 (1948) 184–201.
- [33] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, et al., The Protein Data Bank, *Nucleic Acids Res.* 28 (1) (2000) 235–242, <https://doi.org/10.1093/nar/28.1.235>.
- [34] H.J.R. Berendsen, J.R. Grigera, T.P. Straatsma, Molecular dynamics with coupling to an external bath, *J. Phys. Chem.* 91 (24) (1987) 6269–6271, <https://doi.org/10.1063/1.448118>.
- [35] M.J. Abraham, D. van der Spoel, E. Lindahl, B. Hess and the GROMACS development team, *GROMACS User Manual*, version 5.0.4, www.gromacs.org (2014).
- [36] E.B. Walton, K.J. VanVliet, Equilibration of experimentally determined protein structures for molecular dynamics simulation, *Phys. Rev. E* 74 (6) (2006) 061901, <https://doi.org/10.1103/PhysRevE.74.061901>.
- [37] M.T. Gallo, B.J. Grant, M.L. Teodoro, J. Melton, P. Cieplak, G.N. Phillips Jr, B. Stec, Novel procedure for thermal equilibration in molecular dynamics simulation, *Mol. Simulat.* 35 (2009) 349–357, <https://doi.org/10.1080/08927020802647272>.
- [38] W. Humphrey, A. Dalke, Andrew, K. Schulten, VMD: Visual molecular dynamics, *J. Mol. Graphics.* 14 (1) (1996) 33–38, [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- [39] <https://antoine.frostburg.edu/chem/senese/javascript/>.
- [40] P. Mark, N. Lennart, Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K, *J. Phys. Chem. A* 105 (43) (2001) 9954–9960, <https://doi.org/10.1021/jp003020w>.
- [41] G. Guevara-Carrion, J. Vrabec, H. Hasse, Prediction of self-diffusion coefficient and shear viscosity of water and its binary mixtures with methanol and ethanol by molecular simulation, *J. Chem. Phys.* 134 (7) (2011) 074508, <https://doi.org/10.1063/1.3515262>.
- [42] M. Holz, S.R. Heil, A. Sacco, Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate 1H NMRPFG measurements, *Phys. Chem. Chem. Phys.* 2 (20) (2000) 4740–4742, <https://doi.org/10.1039/B005319H>.
- [43] <https://dtrx.de/od/diff/> (accessed 9 April 2021).
- [44] A.C.F. Ribeiro, O. Ortona, et al., Binary mutual diffusion coefficients of aqueous solutions of β -cyclodextrin at temperatures from 298.15 to 312.15 K, *J. Chem. Eng. Data* 51 (4) (2006) 1368–1371, <https://doi.org/10.1021/jje060092t>.
- [45] A.J. Easteal, W.E. Price, L.A. Woolf, Diaphragm cell for high-temperature diffusion measurements. Tracer diffusion coefficients for water to 363 K, *J. Chem. Soc. Faraday Trans. 1.* 85 (5) (1989) 1091–1097, <https://doi.org/10.1039/F19898501091>.
- [46] D.A. McQuarrie, *Statistical Mechanics*, University Science Books, USA, 2000.



INTERMOLECULAR INTERACTION OF HTHYN1 PROTEIN WITH DOUBLE METHYLATED DNA AT 5M-CYTOSINE NUCLEOTIDE

Rajendra Prasad Koirala, Shyam Prakash Khanal, Sudip Shiwakoti, Narayan Prasad Adhikari*

Central Department of Physics, Tribhuvan University, Nepal

*Corresponding author: npadhikari@gmail.com

(Received: March 5, 2020; Revised: April 20, 2020; Accepted: April 30, 2020)

ABSTRACT

Human thymocyte nuclear protein 1 (hTHYN1) is one of the DNA binding proteins. It is essential for the regulation of Pax5 expression and the development of B cells in humans. Its thermodynamic and biological functions have been unclear yet. The study of the binding mechanism of hTHYN1 protein with DNA is essential to understand various biochemical functions in the human body. In this work, molecular dynamics (MD) simulations have been performed to understand the binding mechanisms of double methylated DNA (dmDNA) at cytosine nucleotide with hTHYN1 protein. Hydrogen bonding and other non-bonded (electrostatics and van der Waals) interactions among the residue-nucleotide pairs have been observed during the MD simulations and are also found responsible to form protein-DNA complex and to provide the stability of the structure. No salt bridges and hydrophobic interactions have been detected. Some of the protein residues in hTHYN1 have been found to strongly cooperate in the formation of the DNA-protein complex. Arginine residue of hTHYN1 has been observed as a major contributor in binding to the DNA. Many other residues also have significant roles in binding with DNA.

Keywords: Contact score, DNA-protein complex, Hydrogen bonding, Non-bonded interactions

INTRODUCTION

Deoxyribose nucleic acid (DNA) and protein interactions play important roles in various biological processes like DNA damage repair, DNA replication, transcription, recombination, regulation of gene expression, and packaging of chromosomal DNA (Narayanan, 2010). DNA binding proteins act as an enzyme that catalyzes the biological reaction during the interaction with DNA. DNA-protein interactions may be specific or non-specific. In a specific binding mechanism, the specific site of a protein interacts or wraps the DNA helix groove; however, the protein may interact at a random position of DNA (Jackson & Bartek, 2009; Tubbs *et al.*, 2007). In present work, a double methylated DNA at 5m-cytosine (methylation at 5th carbon of nucleobase) and human DNA binding protein hTHYN1 complex has been taken into account.

Human thymocyte nuclear protein 1 (hTHYN1) contains 167 amino acid residues. It is mostly found in the cell nucleus. It is a highly conserved protein, i.e., remains unchanged, and is useful in the development of B cells in mammals (Fujita *et al.*, 2018; Yu *et al.*, 2009). The experimental finding reveals that it is dispensable for regulation of Pax5 gene expression and development of B cells at least in two mammals, humans, and mice. Even though a few cellular functions of hTHYN1 protein in the human body have been understood, its interacting mechanisms with DNA are still unclear (Kitaura *et al.*, 2019). Methylation of DNA nucleotide can cause DNA damage. Methylation occurs due to the covalent addition of methyl-adduct (CH₃) at a point of DNA nucleotide.

DNA methylation may occur on one of the nucleobases or at the backbone of DNA. It may occur in a single point or many points of a DNA molecule. Among several base methylation mechanisms, 5-methylcytosine (5-mcyt) is the most common type of DNA base in both eukaryotes and prokaryotes (Pfeifer *et al.*, 2019; Seiler *et al.*, 2018). Methylation at nucleobase causes cancer, muscular dystrophy, and birth defects (Crone *et al.*, 1996). Besides many adverse effects of DNA methylation in living cells, it is also helpful in several cases of gene expression and changing chromatin structure (Paulsen & Ferguson-Smith, 2001).

Non-covalent binding plays a significant role in molecular recognition and binding. Bio-molecular recognition strongly depends on the convoluted network of non-covalent bonded interactions. Such type of interaction holds for all biological interactions including protein-protein complex, protein-DNA complex and integral membrane proteins (Contreras *et al.*, 2012; Etheve *et al.*, 2016; Wan *et al.*, 2013; Zhang *et al.*, 2016). A study on the influences of these interactions in binding DNA and protein molecules revealed many important body mechanisms (Donald *et al.*, 2007).

DNA and protein interactions among the amino acid residues of protein and nucleotides of DNA may bring structural changes in both protein and DNA molecules. The binding process and the cooperative character of these molecules have great importance to deal with the proper functioning within the human body (Mackerell & Nilsson, 2008; Peng *et al.*, 2018). In this work, we dealt with the residues in protein and nucleotides in DNA that

contribute to non-bonded interactions. Various works on DNA-protein interactions have been carried out so far, however, the molecular binding mechanism of hTHYN1 and dmDNA via non-covalent bonding has not been studied yet (Kitaura *et al.*, 2019). To our best knowledge, MD analysis of hTHYN1-DNA complex would be a novel work in this area.

The details of simulations set up and methods of recognizing the significantly binding residues in the DNA-protein complex, and their non-bonded energy estimation was discussed. We presented the findings and expressed the consequences of the research in mathematical and graphical representations.

MATERIALS AND METHODS

System setup

The crystal structure of human THYN1 protein in complex with 5-methylcytosine containing DNA, PDBID 5J3E.pdb, was taken to create the molecular system for MD simulations. The original pdb structure had a complex of targeted protein-DNA in dimer form. This dimer form was modeled to monomer by using CHARMM-GUI (Lee *et al.*, 2016) online software and then generated new pdb and psf structures. The secondary structure of protein and DNA complex has been generated from the molecular representation view of Visual Molecular Dynamics (VMD) (Humphrey *et al.*, 1996) software is shown in Fig. 1. CHARMM-GUI was also used for the remaining steps of molecular system preparation: solvation and ionization. The complex was solvated by using the TIP3P water model and further, made electrically neutral by adding 21 excess Na^+ ions. Thus, a cubic simulation box with dimensions $80 \times 80 \times 80 \text{ \AA}^3$ was prepared to contain 47971 atoms including solvent molecules and neutralizing ions.

Molecular dynamics simulations

All-atom molecular dynamics (MD) simulations have been carried out by nano-scale molecular dynamics (NAMD) simulation package (Phillips *et al.*, 2005). CHARMM36m force field (Huang *et al.*, 2017; Vanommeslaeghe *et al.*, 2010) has been used for entire simulations. Initializing parameters were supplied through Verlet algorithms. The velocity for each atom was assigned through Maxwell's Boltzmann distribution (Khanal *et al.*, 2019; Koirala *et al.*, 2020; Phillips *et al.*, 2005)

Molecular dynamics simulations have been executed in three major steps: energy minimization run, equilibration run, and production run. The original pdb structure has been estimated from x-ray diffraction (XRD) and nuclear magnetic resonance (NMR) techniques. Moreover, hydrogen atoms are absent in the original pdb structure, so they are to be added by using the molecular system

building software. The coordinates, thus, assigned might have steric hindrance, i.e., atom coordinates might be in the wrong position.

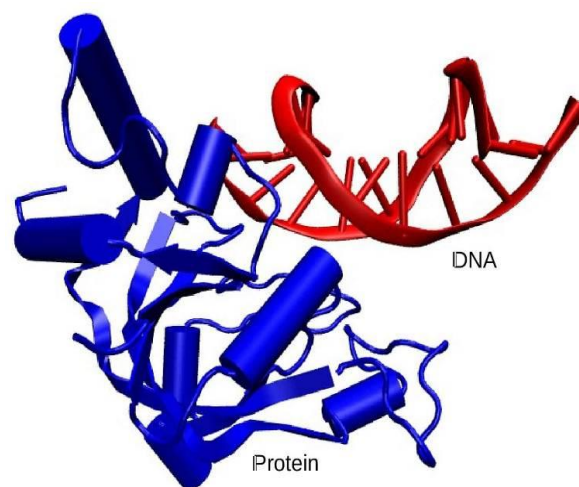


Fig. 1. Double methylated DNA at cytosine and hTHYN1 protein complex

Therefore, before beginning further MD run to estimate the required physical parameters, the structure should be stabilized by keeping at absolute zero temperature, called the energy minimization run (Adhikari *et al.*, 2018; Pokhrel *et al.*, 2019; Walton & Vanvliet, 2006). We performed the energy minimization run for 10,000 steps with a conjugate gradient algorithm. Particle Mesh Elward (PME) was used for the long-range interactions with a 12.0 \AA non-bonded cut-off. Then, the equilibration run was continued for 25 ns through the same simulation file command at 300 K (27 °C) temperature by coupling velocity rescaling thermostat and Berendsen barostat with harmonically restrained condition with 1 fs time step. Then, the system was forwarded for a 100 ps NPT run with a fully unrestrained condition. Finally, the production run was performed for the 50 ns under the NVT ensemble with a 2 fs time step.

Data analysis

Various tools available in the NAMD plugin package in Visual Molecular Dynamics (VMD) (Humphrey *et al.*, 1996) have been used for the analysis of the outcomes after simulations. Also, the Pycontact software package (Scheurer *et al.*, 2018) was used to study the hydrogen bonding between residue-nucleotide pairs that actively involved in binding the molecules. "NAMD Energy" plugin package in VMD was used to estimate the non-bonded energy contributions during the interaction between the targeted molecules. Analysis tools available in VMD have also been used to analyze the conformation change in the DNA-protein complex during the simulations. Xmgrace package was used to plot the simulated data in graphical form.

RESULTS

Root mean square deviation of complex

Molecular stability of the hTHYN1- dmDNA complex was evaluated from the RMSD plot as shown in Fig. 2. In the measurement of molecular deviation, the starting structure has been taken as the reference position for the calculation. The root mean square deviation (RMSD) value has been converged immediately after a short period of MD simulation begun and remains stable throughout the simulation. The black color indicates the RMSD of double methylated DNA (dmDNA) and blue color indicates the RMSD of hTHYN1 protein. DNA was found slightly fluctuating than the protein molecule during the simulation.

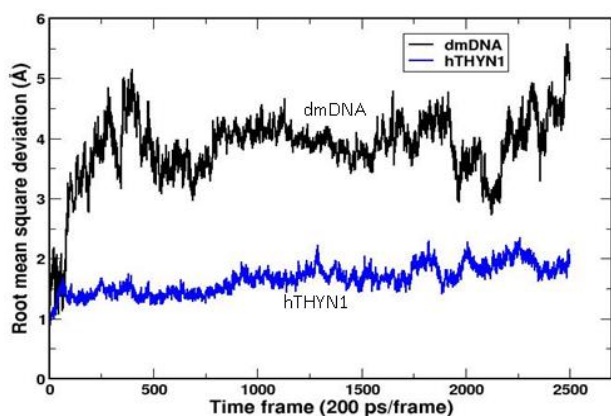


Fig. 2. Root mean square deviation of hTHYN1 protein and dmDNA

Non-bonded interaction and hydrogen bonding between the hTHYN1 and dmDNA

Non-bonded interactions (electrostatic and van der Waals) were analyzed to investigate the binding mechanism in complex formation. During 50 ns MD simulations, the contribution of electrostatic interaction was significantly higher than vdW in total contribution as shown in Fig. 3. Both the electrostatic and van der Waals energy were found in negative values. This shows that the hTHYN1 protein can easily bind to the targeted DNA so that the hTHYN1-dmDNA complex is possible. Moreover, the potential energy provided by electrostatic and vdW interaction was obtained gradually increasing with the simulation time progresses. This depicts that the binding of hTHYN1 protein and dmDNA gradually increased during the simulation. In Fig. 3, black color indicates the plot for electrostatic, green color indicates the plot for vdW, and blue color indicates the sum of electrostatic and vdW interaction energy during the simulation of 50 ns time.

Hydrogen bonding between the hTHYN1 residues and dmDNA nucleotides were estimated by using Pycontact software. Contact scores have been accumulated with

atom distance cut off 5 Å and angle cut off 120°. Higher the contact score implies the stronger hydrogen bonding between corresponding pairs. There were 106 possible contacts observed during the simulation. Out of these, 22 contact pairs have the contact score greater than 1, which were noticeable in binding the molecules to be complex. Specifically, the hydrogen bonding contact between ARG202-GUA10 pair has been found 14.78, which was significantly higher in comparison with all other contact pairs. Next, ARG202 has another contact with THY9 of the DNA with a contact score of 6.36. No hydrophobic and salt bridges were observed. The residue pairs for significant hydrogen bonding are presented in Fig. 4. The positions of most interacting residue-nucleotide pairs are shown in Fig. 5(i) and the locations of partners of maximum contact score are shown in Fig. 5(ii).

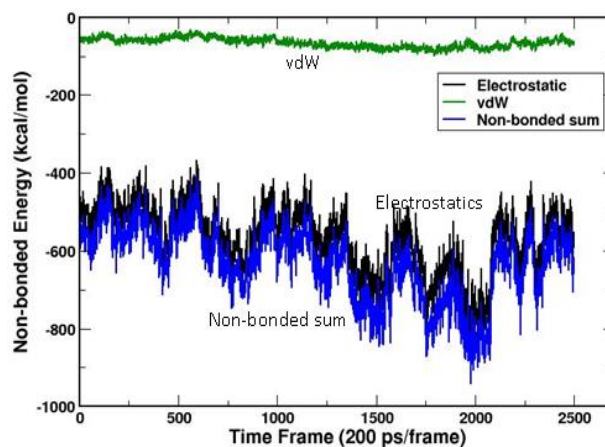


Fig. 3. Non-bonded energy contributions in hTHYN1 protein and dmDNA

Pycontact timeline has been shown in Fig. 4 for 50 frames. The first column shows the labels for contacts, the color code represents the contact positions on the nature of the interaction. Green color represents the side-chain side-chain interaction, yellow color represents the side-chain back-bone interaction and blue color represents the back-bone back-bone interactions. The intensity of the color code shows the strength of the hydrogen bonding mechanism. The residue-nucleotide pair partners and their corresponding contact scores are shown in Fig. 6. Most of the contacts have been side chain - side chain interactions. Some of the interacting partners were observed continuously persist in contact throughout the simulation time as shown in Fig. 6. In contrast, some others left the partners as soon as the simulation started as shown in Fig. 7.

Structural change in hTHYN1 protein and dmDNA

The structural variation in the complex has been observed. The snapshots to investigate the complex for every 10 ns time range during the simulations are shown in Fig. 8. Six pairs, LEU203-THY9, LEU203-GUA10, SER204-

GUA10, ARG202-THY9, ARG202-GUA10, and LYS60-GUA10, were found continuously interacting throughout the simulation, i.e., no break in the interaction. These

pairs significantly contribute to binding the hTHYN1 protein and dmDNA molecules.

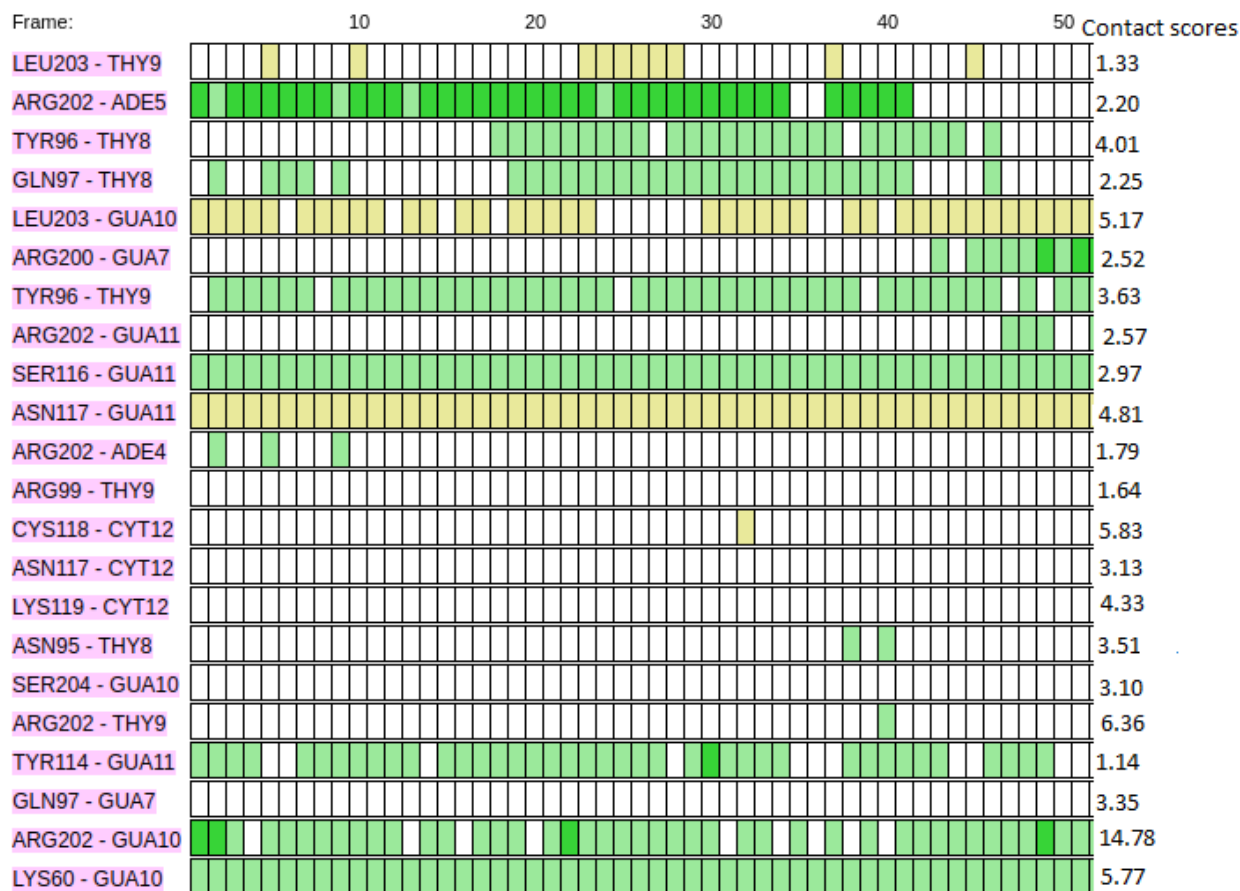


Fig. 4. Contact partners of amino acid residues in hTHYN1 protein and dmDNA nucleotides. ARG202 has interacted strongly with the DNA nucleotides. Corresponding residue-nucleotide pairs are presented in the left hand column and contact scores are presented at the right hand column of Pycontact timeline

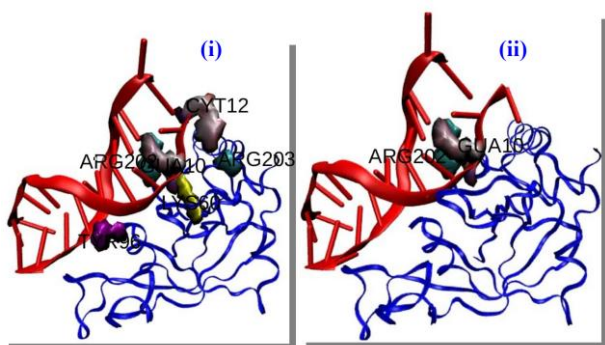


Fig. 5. Representation of secondary structure of DNA-protein complex showing the (i) main interacting residues of hTHYN1 protein and dmDNA and (ii) contact pairs ARG202-GUA10 having maximum contact score

Fig. 8 shows the structural variation of the protein-DNA complex in every 10 ns simulation time steps. During the simulation time frame, hTHYN1 has continuously been interacting with methylated DNA. TYR96 residue attempted to intercalate into the DNA backbones. After 20 ns lapse of simulation time, TYR96 residue was found gradually entering into the DNA strands.

DISCUSSION

We observed the binding mechanism of the hTHYN1 protein and dmDNA. Out of many factors, Coulomb interaction, van der Waals interaction, and hydrogen bonding play a very important role in molecular binding (Jayaram *et al.*, 1999; Raut & Lamichhane, 2017). As soon as protein molecule approaches at an appropriate distance with the most probable state, the non-covalent interactions become effective and the protein forms a molecular complex with DNA. The prerequisite of finding

the molecular stability of the complex is the evaluation of RMSD of the entire molecular system. So, Root Mean Square Deviation (RMSD) has been used for the standard measure of the structural distance between coordinate sets of molecules under MD simulation. In present work, even though the RMSD of DNA of a molecule is greater than the protein molecule, both are stabilized shortly after the simulation started and found stable throughout the simulation.

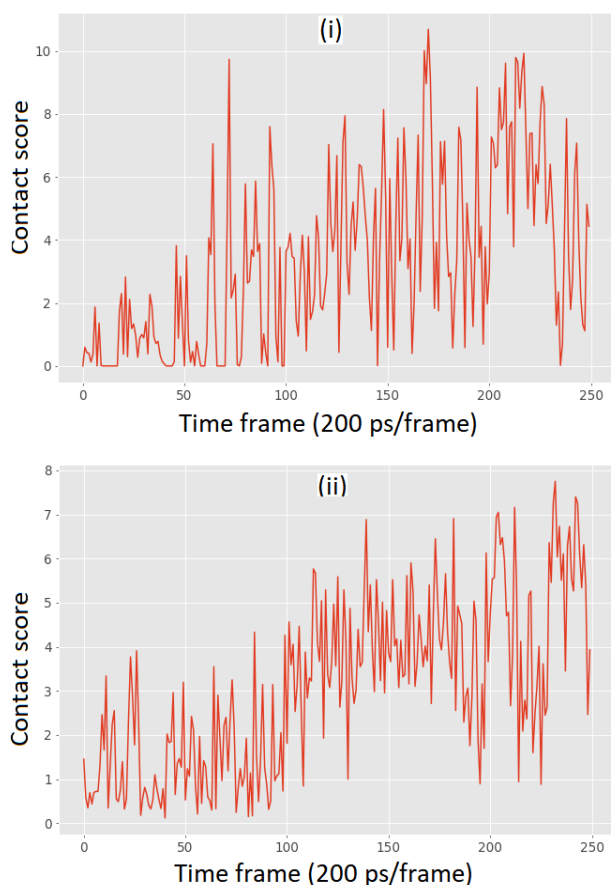


Fig. 6. Plot for contact scores of interacting partners (i) GLN97-GUA7 (ii) SER204-GUA10 pairs for hydrogen bonding strength gradually increasing

Furthermore, non-bonded interactions like electrostatics and van der Waals interactions play a vital role in non-covalent bonding interactions (Pantha & Adhikari, 2014; Sedhain & Kaphle, 2017). The contribution of electrostatics and van der Waals (vdW) interaction energy and their combined effect were analyzed using the NAMD plugin program in VMD. The magnitude of non-bonded interaction energy has been gradually increasing as the simulation time progresses, which indicates favorable in the cooperativity of binding the molecules in protein and DNA.

Hydrogen bonding between the hTHYN1 residues and dmDNA nucleotides has been estimated by using the

Pycontact software package. This package utilizes the structure file and trajectory file that were generated from the MD simulation. Out of many tools available in Pycontact software, we have analyzed the hydrogen bonding and salt bridges between hTHYN1 and dmDNA molecules. The role of hydrogen bonding was observed significantly high to form the complex; however, salt bridge and hydrophobic interaction have not been detected.

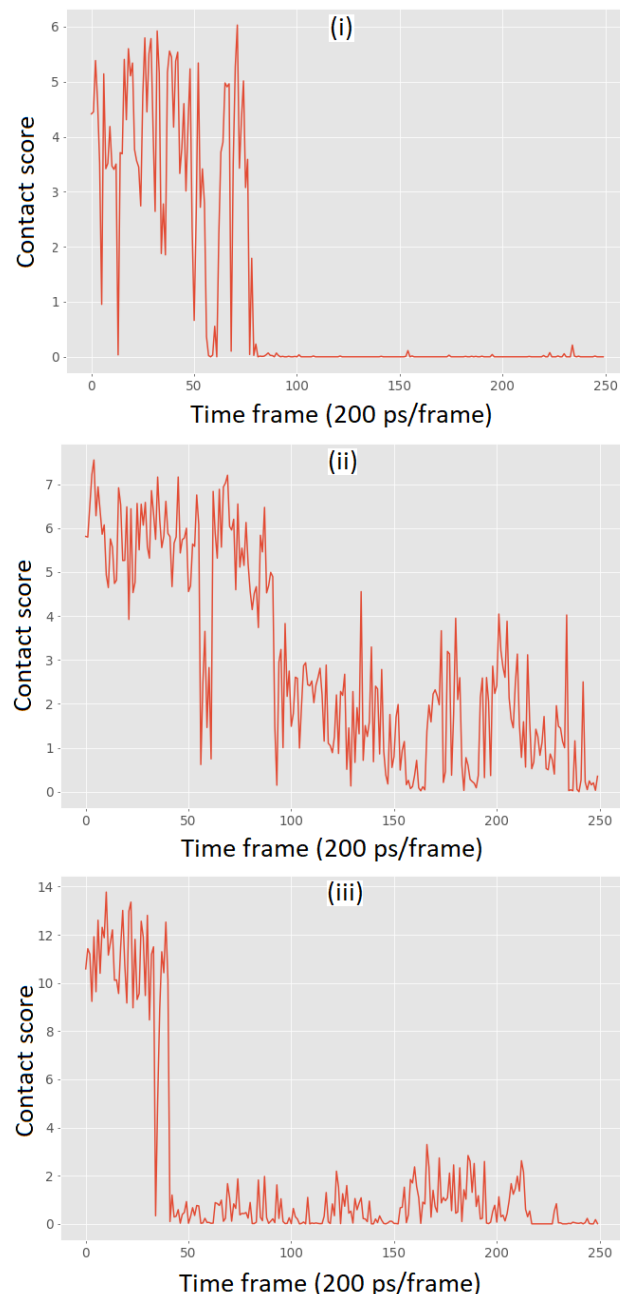


Fig. 7. Plot for contact scores of interacting partners significantly decreasing for hydrogen bonding strength in (i) TYR114-GUA11 pairs (ii) SER116-GUA11 pairs (iii) ARG202-ADE5 pairs

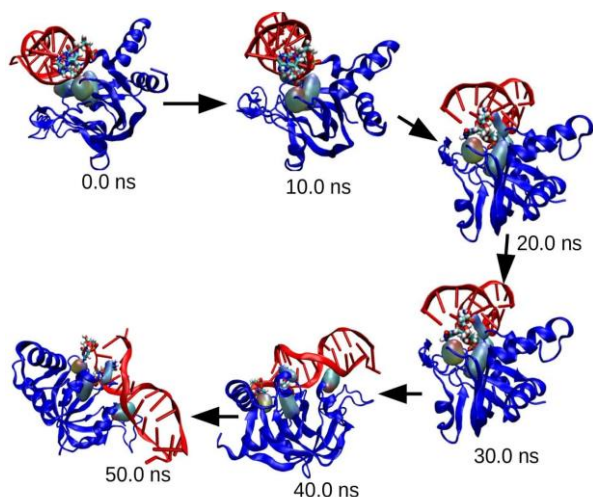


Fig. 8. The simulation snapshots of molecules at different time steps at (i) 0.0 ns (ii) 10.0 ns (iii) 20.0 ns (iv) 30.0 ns (v) 40.0 ns (vi) 50 ns

Some of the residues interact throughout the simulation, however, some others have been found disconnected after a short simulation time. Since Coulomb interaction dominates in overall non-bonded interaction, the partial charge distribution has been unfavorable from an electrostatic point of view, so they move away from each other. The contact score was found at increasing trend in GLN97-GUA7 pairs and SER204-GUA10 pairs; this signifies that they were approaching nearer to each other so that the interaction between these residues became stronger gradually. Also, some other residues have left their interacting partners.

The protein binds to the DNA to perform several biological mechanisms like DNA replication, lesion repair mechanism, and gene expression. The identification of active residues is essential to resolve the complications associated with DNA. That's why we investigated how protein residues interact with the DNA. Our findings are similar to the consequences of the Density Functional Theory (DFT) and Atomic Force Microscopy (AFM). Density Functional Theory (DFT) study on the targeted active residues in a protein that binds to the minor groove of DNA has found that some of the residues are responsible to provide the stable complex between methyl-DNA and DNA repair protein (Jena & Bansal, 2011). The Atomic Force Microscopy (AFM) study for the cooperativity of methyl-DNA and protein residues revealed that conformational change within protein residues has a dominating effect for a stable DNA-protein complex rather than sequence specificity (Daniels *et al.*, 2000; Tessmer & Fried, 2014). We also performed the 50 ns MD simulation to identify the interaction mechanism of residue-nucleotide pairs. During the simulation, some of the residues; LEU203, LEU203, SER204, ARG202, ARG202, and LYS60 were found continuously interacting

with two nucleotides GUA10 and THY9, and enhanced to provide the stable structure of the complex. There were several contacts, out of 106 contacts observed during the simulations, 22 were significant and found to contribute in the complex formation.

CONCLUSIONS

Molecular dynamics (MD) simulations have been performed to study the interaction of methylated-DNA at cytosine with human hTHYN1 protein using the NAMD software package. We performed MD simulation in three steps: energy minimization run, equilibration run, and production run. The energy minimization run was performed for 10,000 steps by using the conjugate gradient algorithm with Particle Mesh Elward (PME) for non-bonded cut off 12 Å. This run has provided us a reasonable starting structure in terms of geometry and solvent orientation for the further MD simulation run. In the next step, the equilibration run was performed for 25 ns by coupling velocity rescaling thermostat and Berendsen barostat. The system was studied at a temperature of 300 K and pressure of 1 bar to make the system thermodynamically stable. The stable system obtained from the equilibration run was subjected to NVT production run for 50 ns.

During the analysis of non-bonded interactions, LEU203-THY9, LEU203-GUA10, SER204-GUA10, ARG202-THY9, ARG202-GUA10, and LYS60-GUA10 interacting partners have been observed throughout the simulation. ARG202-GUA10 pair has the maximum hydrogen bonding contact score. Moreover, some conformational change en found in DNA when the protein residues attempted to interacts more strongly into the nucleobase of DNA.

ACKNOWLEDGMENTS

R.P.K. and S.P.K. acknowledge the partial financial support from the Nepal Academy of Science and Technology (NAST). N.P.A. acknowledges the UGC Award No. CRG-73/74-S&T-01 and Associate membership of ICTP helped to complete this work.

REFERENCES

- Adhikari, N. P., Khanal, S. P., & Pokharel, S. (2018). Solvation free energy of light alkanes in polar and amphiphilic environments. *Bibechana*, 16, 92-105.
- Contreras, F. X., Ernst, A. M., Haberkant, P., Bjorkholm, P., Lindahl, E., Gonen, B., . . . Brugger, B. (2012). Molecular recognition of a single sphingolipid species by a protein's transmembrane domain. *Nature*, 481(7382), 525-529.
- Crone, T. M., Goodtzova, K., & Pegg, A. E. (1996). Amino acid residues affecting the activity and stability of human O6-alkylguanine-DNA

- alkyltransferase. *Mutation Research/DNA repair*, 363(1), 15-25.
- Daniels, D. S., Mol, C. D., Arvai, A. S., Kanugula, S., Pegg, A. E., & Tainer, J. A. (2000). Active and alkylated human AGT structures: a novel zinc site, inhibitor and extrahelical base binding. *The EMBO Journal*, 19(7), 1719-1730.
- Donald, J. E., Chen, W. W., & Shakhnovich, E. I. (2007). Energetics of protein-DNA interactions. *Nucleic Acids Research*, 35(4), 1039-1047.
- Etheve, L., Martin, J., & Lavery, R. (2016). Dynamics and recognition within a protein-DNA complex: a molecular dynamics study of the SKN-1/DNA interaction. *Nucleic Acids Research*, 44(3), 1440-1448.
- Fujita, T., Kitaura, F., Oji, A., Tanigawa, N., Yuno, M., Ikawa, M., . . . Fujii, H. (2018). Transgenic mouse lines expressing the 3xFLAG-dCas9 protein for enChIP analysis. *Genes Cells*, 23(4), 318-325.
- Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., ... & MacKerell, A. D. (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods*, 14(1), 71-73.
- Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of Molecular Graphics*, 14(1), 33-38, 27-38.
- Jackson, S. P., & Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature*, 461(7267), 1071-1078.
- Jayaram, B., McConnell, K., Dixit, S. B., & Beveridge, D. (1999). Free energy analysis of protein-DNA binding: the EcoRI endonuclease-DNA complex. *Journal of Computational Physics*, 151(1), 333-357.
- Jena, N. R., & Bansal, M. (2011). Mutagenicity associated with O6-methylguanine-DNA damage and mechanism of nucleotide flipping by AGT during a repair. *Physical Biology*, 8(4), 046007. doi:10.1088/1478-3975/8/4/046007
- Khanal, S. P., Kandel, Y. P., & Adhikari, N. P. (2019). Transport properties of zwitterion glycine, diglycine, and triglycine in water. *AIP Advances*, 9(6). doi:10.1063/1.5099069
- Kitaura, F., Yuno, M., Fujita, T., Wakana, S., Ueda, J., Yamagata, K., & Fujii, H. (2019). Normal B cell development and Pax5 expression in Thy28/ThyN1-deficient mice. *PLoS One*, 14(7), e0220199. doi:10.1371/journal.pone.0220199
- Koirala, R. P., Bhusal, H. P., Khanal, S. P., & Adhikari, N. P. (2020). Effect of temperature on transport properties of cysteine in water. *AIP Advances*, 10(2). doi:10.1063/1.5132777
- Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., . . . Im, W. (2016). CHARMM-GUI Input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *Journal of Chemical Theory and Computation*, 12(1), 405-413.
- Mackerell, A. D., Jr., & Nilsson, L. (2008). Molecular dynamics simulations of nucleic acid-protein complexes. *Current Opinion in Structural Biology*, 18(2), 194-199.
- Narayanan, P. (2010). *Essentials of biophysics*. New Delhi, India: New Age International Publisher.
- Pantha, N., & Adhikari, N. P. (2014). Structure and symmetrization of hydrogen bonding in ices VIII and X at high pressure: a density functional theory approach. *Journal of Institute of Science and Technology*, 19(2), 14-18.
- Paulsen, M., & Ferguson-Smith, A. C. (2001). DNA methylation in genomic imprinting, development, and disease. *The Journal of Pathology*, 195(1), 97-110.
- Peng, Y., Sun, L., Jia, Z., Li, L., & Alexov, E. (2018). Predicting protein-DNA binding free energy change upon missense mutations using a modified MM/PBSA approach: SAMPDI webserver. *Bioinformatics*, 34(5), 779-786.
- Pfeifer, G. P., Szabo, P. E., & Song, J. (2019). Protein interactions at oxidized 5-methylcytosine bases. *Journal of Molecular Biology*. doi:10.1016/j.jmb.2019.07.039
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., . . . Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16), 1781-1802.
- Pokhrel, R., Pavadai, E., Gerstman, B. S., & Chapagain, P. P. (2019). Membrane pore formation and ion selectivity of the Ebola virus delta peptide. *Physical Chemistry Chemical Physics*, 21(10), 5578-5585.
- Raut, N. K., & Lamichhane, H. P. (2017). Sequential substitution of methyl side group in naphthoquinones and anthraquinone to investigate the sensitivity of the carbonyl band. *Journal of Institute of Science and Technology*, 22(1), 137-146.
- Scheurer, M., Rodenkirch, P., Siggel, M., Bernardi, R. C., Schulten, K., Tajkhorshid, E., & Rudack, T. (2018). PyContact: rapid, customizable, and visual analysis

- of noncovalent interactions in MD simulations. *Biophysical Journal*, 114(3), 577-583.
- Sedhain, R. P., & Kaphle, G. C. (2017). Structural and electronic properties of transition metal dichalcogenides (MX₂) M=(Mo, W) and X=(S, Se) in the bulk state: a first-principles study. *Journal of Institute of Science and Technology*, 22(1), 41-50.
- Seiler, C. L., Fernandez, J., Koerperich, Z., Andersen, M. P., Kotandeniya, D., Nguyen, M. E., . . . Tretyakova, N. Y. (2018). Maintenance DNA methyltransferase activity in the presence of oxidized forms of 5-methylcytosine: structural basis for ten eleven translocation-mediated DNA demethylation. *Biochemistry*, 57(42), 6061-6069.
- Tessmer, I., & Fried, M. G. (2014). Insight into the cooperative DNA binding of the O(6)-alkylguanine DNA alkyltransferase. *DNA Repair (Amst)*, 20, 14-22.
- Tubbs, J. L., Pegg, A. E., & Tainer, J. A. (2007). DNA binding, nucleotide flipping, and the helix-turn-helix motif in base repair by O6-alkylguanine-DNA alkyltransferase and its implications for cancer chemotherapy. *DNA Repair (Amst)*, 6(8), 1100-1115.
- Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., . . . Mackerell, A. D., Jr. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*, 31(4), 671-690.
- Walton, E. B., & Vanvliet, K. J. (2006). Equilibration of experimentally determined protein structures for molecular dynamics simulation. *Physical Review E*, 74(6 Pt 1), 061901. doi:10.1103/PhysRevE.74.061901
- Wan, H., Hu, J. P., Li, K. S., Tian, X. H., & Chang, S. (2013). Molecular dynamics simulations of DNA-free and DNA-bound TAL effectors. *PLoS One*, 8(10), e76045. doi:10.1371/journal.pone.0076045
- Yu, F., Song, A., Xu, C., Sun, L., Li, J., Tang, L., . . . He, J. (2009). Determining the DUF55-domain structure of human thymocyte nuclear protein 1 from crystals partially twinned by tetartohedral. *Acta Crystallographica Section D: Biological Crystallography*, 65(Pt 3), 212-219.
- Zhang, Y., Vukovic, L., Rudack, T., Han, W., & Schulten, K. (2016). Recognition of poly-ubiquitins by the proteasome through protein refolding guided by electrostatic and hydrophobic interactions. *Journal of Physical Chemistry B*, 120(33), 8137-8146.



Intra-molecular Conformational Stability in Human Growth Hormone

R. P. Koirala¹, B. Thapa², S. P. Khanal¹, R. P. Adhikari³ and N. P. Adhikari^{1,*}

¹Central Department of Physics, Tribhuvan University, Kathmandu, Nepal

²Padma Kanya Multiple Campus, Tribhuvan University, Kathmandu, Nepal

³Department of Natural Sciences, Kathmandu University, Dhulikhel, Nepal

*Corresponding Email: narayan.adhikari@cdp.tu.edu.np

Received: 22 October, 2020; Revised: 28 November, 2020; Accepted: 25 December, 2020

Abstract

Human growth hormone (hGH) is synthesized, stored and secreted by somatotrophic cells within the lateral wings of the anterior lobe of pituitary glands; and is transported to other organs of human body. Study of intra-molecular structure and its binding mechanisms within the molecule gives more insight of structural stability of the molecule and is also essential in drug designing. In this article, we have investigated the various bonded and non-bonded interactions that contribute for the conformation of entire structure of the hGH molecule using molecular dynamics (MD) simulation. The MD outcomes show that the molecule is hydrophobic in nature. In its conformation, several types of interactions exist, such as disulphide bridges (bonded) and nonbonded: hydrogen bond, hydrophobic, aromatic-aromatic, ionic, aromatic-sulphur, cation-pi.

Keywords: Intra-binding, Growth hormone, Hydrophobic, Aromatic.

1. INTRODUCTION

Personal height is the highly concerned matter for everyone. It is also a measure of health conditions [1]. Human growth hormone (hGH) is essential for the proper development of height as well as every part of body. It is a single chain peptide hormone that stimulates growth, reproduction and regeneration of cell in human [2]. Although growth hormone is found in other animals, its effect is significant only in human and old world monkey. Many researches have revealed that the variation of amino acid sequence of hGH is significant in different species [3]. Even though the sequence is different, its role in other animals is also similar to that in human body. This hormone contains 191 amino acid residues with 22,124 daltons molecular weight. It is synthesized, stored and secreted by somatotrophic cells within the lateral wings of the anterior lobe of pituitary glands [4]. The hormone after secretion from the gland mixes into the blood stream and is transported into body cells [5]. The structure includes four helices necessary for the functional interaction with the growth hormone (GH) receptor [6]. Human growth hormone is

essential in physical development. Main problem of human growth in childhood is the short stature, i.e., insufficient growth in accordance with age; and delayed sexual maturity in adulthood [7].

Human growth hormone, also known as somatotropin, provides important contribution in human development. It increases the concentration of glucose and free fatty acids [1]. It is legally prescribed as a drug to treat the children's growth disorders and adult growth hormone deficiency. Moreover, it is also used in raising livestock more efficiently in industrial agriculture.

Besides its functions of somatic and bone growth as well as increase in the size and mass of organs and tissues, hGH also influences in the functioning of proteins, carbohydrates and lipids metabolisms [8]. These mechanisms caused by hGH are basically due to its ability to bind with specific target cell receptor. hGH surface specifically favors for the inter-molecular bindings at the interfacial region of receptor. Inter-molecular binding occurs in the expense of intra-molecular binding, which ultimately lowers the conformation stability of the molecule [9, 10]. There are several non-bonded

interactions to form a macromolecule a stable structure. Hydrogen bonding play a pivotal role in the formation of secondary structure of a protein. Likewise, disulphide bond is a covalent bonding to give the shape in tertiary structure. Similarly, aromatic-aromatic interactions between two aromatic rings separated by distance 4.5 Å to 7 Å provide the conformation stability. This interaction is energetically favorable and basically applies to form the tertiary and quaternary structure [11, 12]. Similarly, electrostatic and van der Waals interactions are pervasive in intra- and inter-molecular bindings [13]. Likewise, hydrophobic interaction has influences in conformation changes of protein molecule in aqueous environment [14].

In this work, we have used molecular dynamics simulation to study the intra-molecular interactions and the contributions of hydrogen bonding and several other interactions such as hydrophobic, ionic, aromatic-aromatic, aromatic-sulphur and cation-pi that are responsible to provide the stability of hGH molecule. Moreover, we have compared the intra-molecular interactions of amino acid residues in static structure with the structure obtained from molecular dynamics simulations. To our best knowledge, the comprehensive study on molecular stability of this hormone has not been studied yet via MD run. We believe that this work aids the in-depth knowledge about the binding of amino acid residues in order to form the stable structure of the hormone in cellular environment.

2. MATERIALS AND METHODS

To perform molecular dynamics simulation, one needs initial structure of molecule (pdb file) and force field parameters. In this section, we describe system setup of the hGH molecule, force field parameters used in the present work and simulation details.

System setup: The molecular structure of human growth hormone was taken from protein data bank with PDBID 1HGU.pdb [15]. In original file, two amino acids were missing at the N-terminus and C-terminus positions, which were filled with software program CHARMM-GUI [16]. This software program was also used to generate new protein structure file (psf) and pdb files. To mimic the cellular environment, the hormone molecule was solvated in TIP3P water in a cubical box of dimensions $84 \times 84 \times 84 \text{ \AA}^3$. As TIP3P water model resembles real water closely, we chose this model in this work. Furthermore, the molecular system had originally 4 excess negative electronic charges,

so the system was electrically neutralized by adding 4 Na⁺ ions.

Force fields: Force fields contain topology and parameter files. Topology files are used to generate psf files and parameter files contain the information about parameters of potential energy functions. In the present work, we used CHARMM36m [17] force field.

Molecular Dynamics Simulation: All-atom molecular dynamics (MD) simulations were carried out by using NANOSCALE Molecular Dynamics (NAMD) [18] simulation package. The Particle Mesh Ewald (PME) was used to treat the long-range interactions with a 12.0 Å non-bonded cut off. The energy minimization was performed for 10,000 steps, using the conjugate gradient algorithm. Energy minimization run removes the unwanted hindrances between the atoms in the system. Since this run is performed in 0 K temperature, the system chooses the local minimum energy state [19, 20]. After energy minimization, the system was equilibrated at 310 K under the isothermal-isochoric conditions for 10 ns with 1 fs time step [21,22]. Then, the production run was propagated for 100 ns under NVT simulation run taking time step 2 fs by using Langevin dynamics with a damping constant of 1 ps⁻¹.

The NAMD energy plugin package available in Visual Molecular Dynamics (VMD) [23] was used to estimate structural stability of the hGH molecule in aqueous environment. Protein interaction calculator (PIC) [24] has been used to analyze the intra-molecular interactions. The study on intra-molecular interactions depicts the detailed insights on folded state of the targeted molecule.

3. RESULTS AND DISCUSSION

This research work has been carried out by molecular dynamics simulation to study the amino acid arrangements, their folding mechanisms, the contributions of hydrogen bonding, and various other non-bonded interactions to form such a structure of human growth hormone in aqueous environment. The solvent accessible surface area and the energy profiles of bonded and non-bonded interactions have also been investigated to know the contact area and the stability of hGH in the aqueous environments.

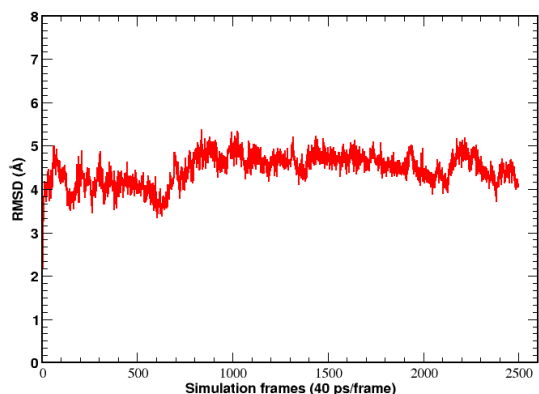
Structural stability of human growth hormone: The pre-requisite of examining the structural stability of any biomolecular system is the estimation of root mean square deviation (RMSD). RMSD measures the structural stability of entire molecule in the given environment. We have taken

the hGH molecule in aqueous environment in order to resemble the cellular condition. RMSD of the molecule has been calculated from VMD analysis tool "RMSD Trajectory Tool". The mathematical relation to estimate the RMSD is,

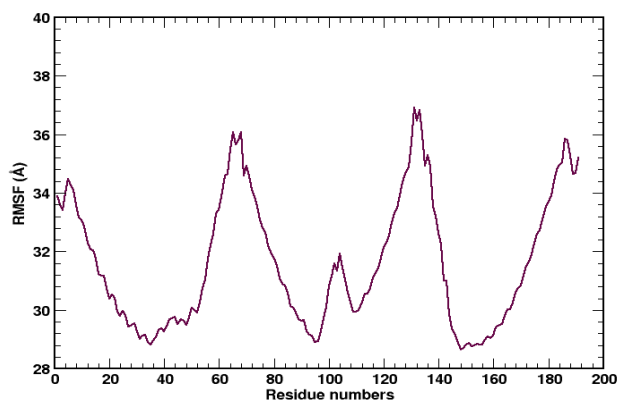
$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i(t) - r_i^{ref})^2}$$

Where, $r_i(t)$ and r_i^{ref} represent the current and reference coordinates of i^{th} atom respectively and N is total number of atoms.

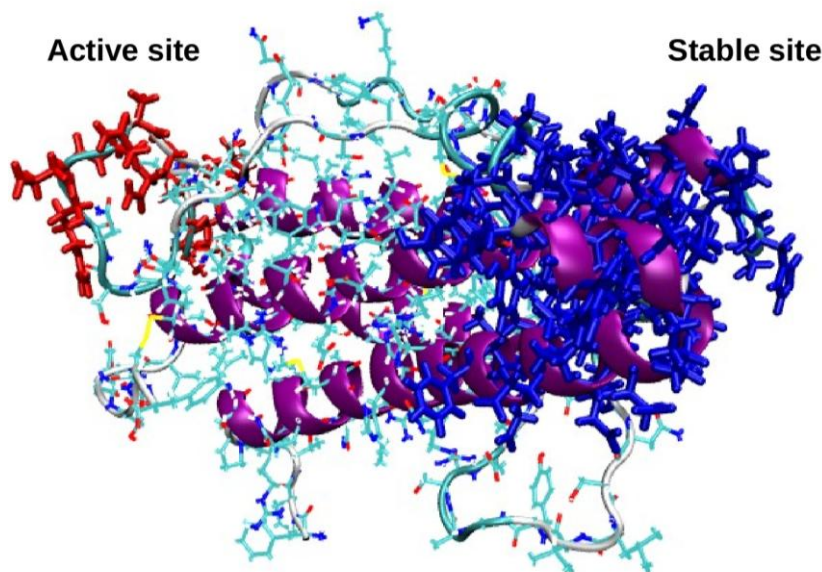
In this process, initial coordinates are taken as the reference and average deviation of molecule from the reference coordinates has been calculated in every frame of simulation. During the 100 ns NVT production run, the RMSD of the hGU was found fairly stable after 4 ns simulation (100 frames) time as shown in Fig. 1(i).



(i)



(ii)



(iii)

Fig. 1: Structural characterization of hGH in aqueous environment (i) RMSD (ii) RMSF and (iii) hGH molecule.

In addition, we have also determined the fluctuation of alpha carbons of hGH residues during the simulations through root mean square fluctuation (RMSF) as shown in Fig. 1(ii), which gives the flexibility of residues within the molecular system. Greater value of RMSF of any alpha carbon implies

the greater flexibility of corresponding amino acid in the protein chain and vice versa. The mathematical formula to find the RMSF is,

$$RMSF = \sqrt{\langle r_i(t) - \langle r_i \rangle \rangle^2}$$

where, $r_i(t)$ and $\langle r_i \rangle$ are the current position and average position respectively.

During the 100 ns simulation run, it has been observed that the regions of amino acid residues 64-68 and 130-136 are most fluctuating and the regions of 23-49, 88-98 and 144-164 are the most stable. The most fluctuating and stable regions are shown in Fig. 1 (iii).

Intra-binding potential energy: There are several interactions existing within a protein molecule to provide it a stable conformation. We have estimated the contributions of bonded and non-bonded potential energy in hGH molecule. Energy

profile for bonded interactions have been determined from 100 ns simulation. From the energy profile, it has been observed that improper dihedral has lowest contributions; and the contributions of harmonic and dihedral angle show almost equal and the largest value as shown in Fig. 2. During the conformation of protein molecules, bonded energy provides the direct covalent link among the atoms, whereas the non-bonded energy contributes to fold the molecule forming the stable structure [25-27]. On analyzing the non-bonded energy, the contributions of electrostatic interaction has found to be significantly higher potential energy than that of van der Waals.

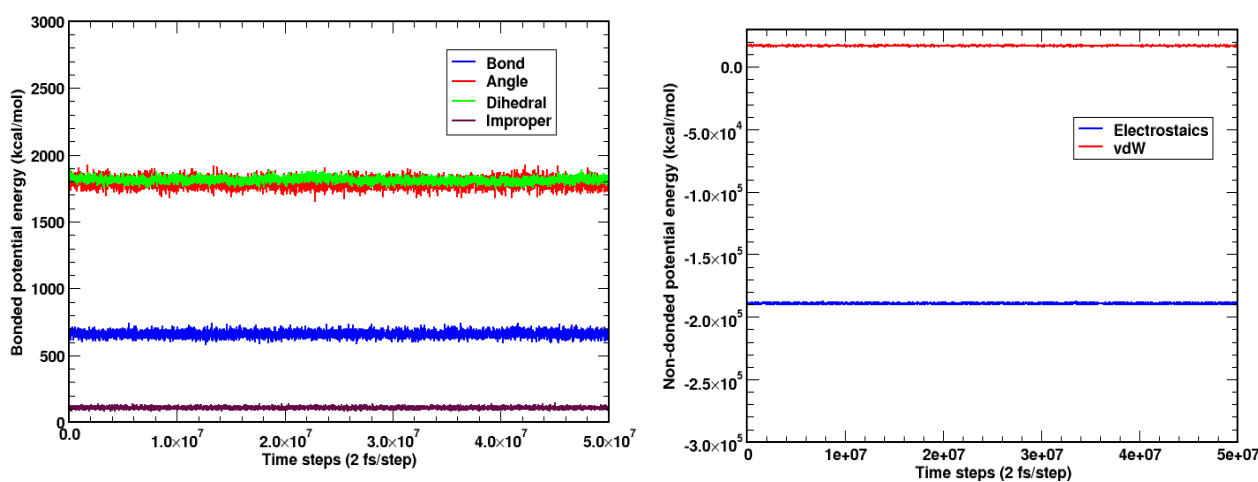


Fig. 2: Energy profiles for hGH (i) bonded (bond, angle, dihedral and improper) interaction (ii) non-bonded (electrostatics and van der Waals) interactions.

Intra-molecular hydrogen bonds: We have also investigated the hydrogen bonding within the molecule to form the hGH molecule. Hydrogen bonding is essential to form secondary structure of protein. The intra-molecular hydrogen bonding provides the important information for drug designing [28, 29]. The number of hydrogen bonds can determine the conformational stability in the molecule. Its role in intermolecular interactions is also very important, which enhances the cooperativity among the molecules in protein-protein, protein-ligand and protein-nucleic acid systems. We have found total 1294 hydrogen bonding in entire 100 ns simulation, and average of 65 hydrogen bonds persisted in each frame of simulation. The number of hydrogen bonds within the cutoff distance of 3.5 Å has been presented in Fig. 3.

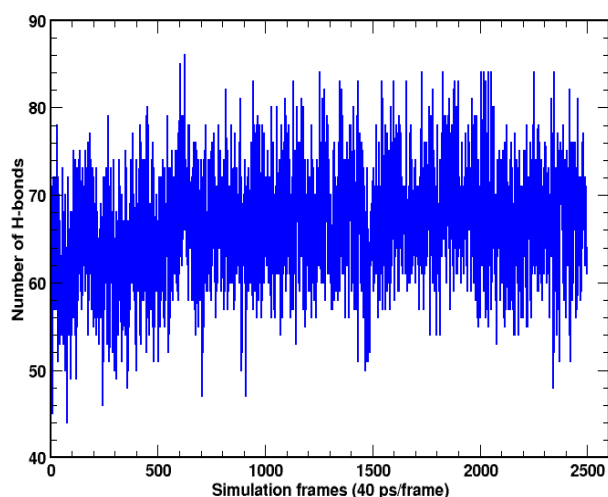


Fig. 3: Time variation of number of hydrogen bonding for intra-molecular binding in hGH.

Many intra-molecular hydrogen bonds were observed in both static and dynamic conditions of the hGH structure. We have shown the hydrogen bonding patterns for three interacting pairs. These three pairs are the representative pairs for main

chain – main chain, side chain – side chain and main chain – side chain as shown in Fig. 4. There are several such type of hydrogen bonds were observed to form the stable structure of the hormone.

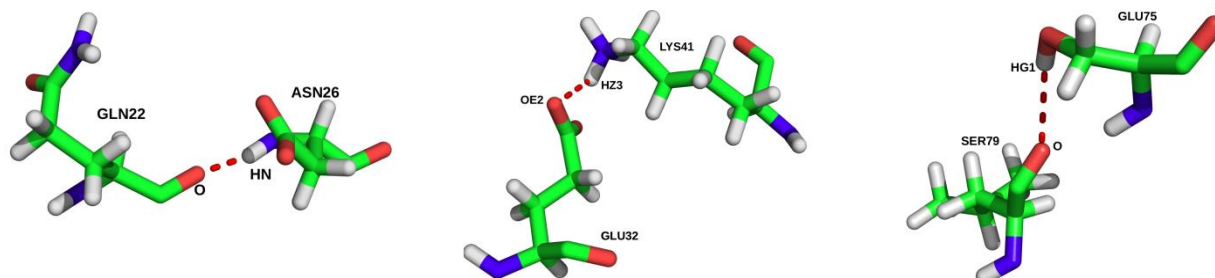


Fig. 4: Intra-molecular hydrogen bonds: main chain – main chain (left), side chain – side chain (middle) and main chain – side chain (right).

Surface accessible surface area (SASA): SASA measures the surface area of a molecule that contacts with the solvent molecules. We have taken water as the solvent to study how the amino acid residues residing on the surface of hGH behave in the aqueous environment. Fig. 5 shows the SASA plot with respect to simulation frame. During the 100 ns simulation, the overall value of SASA has been observed decreasing. The decreasing nature of graph shows the internal rearrangement of amino acid residues that resides on the surface. This shows that the molecule is hydrophobic in nature so that the surface residues tend to aggregate together to minimize the surface area exposed to water.

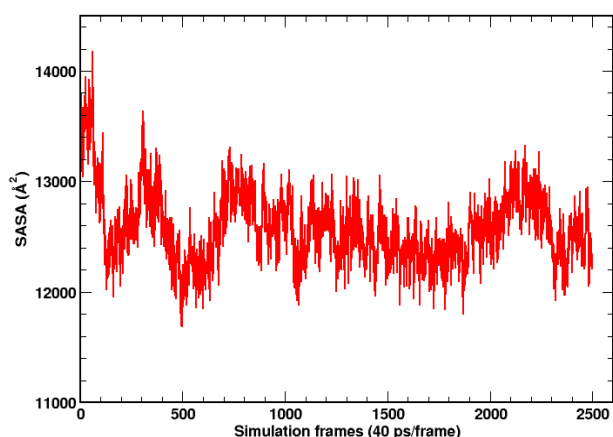


Fig. 5: Solvent accessible surface area (SASA) of hGH in water environment.

Intra-molecular Interactions: We have computed the various interactions present within the human

growth hormone such as hydrophobic interactions, disulphide bridges, ionic interactions, aromatic-aromatic interactions, aromatic-sulphur interactions and cation-pi interactions using protein interaction calculator (PIC) web server. We have compared these intra-molecular interactions in static structure with the most stable structure obtained from molecular dynamics simulations. The most stable structure was taken from the MD trajectories corresponding to the minimum SASA.

There are two disulphide bridges formed between two sulphur atoms of cysteine residues CYS182-CYS189 and CYS53-CYS165 within 2.2 Å in both structures. The distance between two cysteine residues in CYS182-CYS189 and CYS53-CYS165 is 2.02 Å in static structure, whereas 2.06 Å and 2.03 Å respectively in the structure from dynamics.

The hydrophobic interactions are of particular importance for the structural stability of hGH. There are 116 interactions between the hydrophobic residues such as PHE, LEU, ALA, MET, PRO, VAL, ILE, TYR and TRP present in the static structure. We have observed the decrease in the solvent accessible surface area (SASA) of protein during the simulation in aqueous environment that suggests the hydrophobic nature of the molecule. This is corroborated further by the greater number of hydrophobic interactions in the dynamics, which are 143 as compared to 116 in static. The cut off distance of hydrophobic interaction was taken to be 5 Å.

In addition, the ionic interactions between the charged residues lying within the cut off distance of 6 Å present in both structures of hGH protein are depicted in tables 1 and 2. We have observed 29

ionic interactions in structure from dynamics, whereas only 11 in static. Salt bridges are the important interactions in forming the tertiary structure of protein molecules. They are the ionic interactions between charged atoms in residues lying within hydrogen bond distance (3.5 Å). In the structure obtained from MD simulation 7 salt-bridges has been observed whereas only 2 salt-bridges are obtained in static structure.

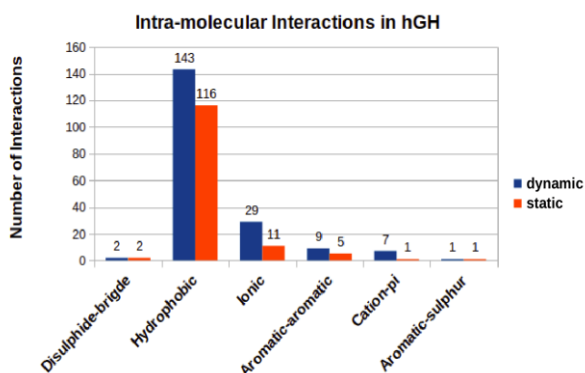


Fig. 6: Bar diagram showing different intra-molecular interactions in hGH molecule.

The bar diagram shown in Fig. 6 summarizes the various intra-molecular interactions present in hGH molecule in both structures i.e., static and dynamic conditions. The comparison diagram shows that hydrophobic interactions have the maximum binding sites and aromatic-sulphur has minimum binding sites.

Table 1: Ionic Interactions within 6 Å (in static structure)

Position	Residue	Position	Residue
16	ARG	107	ASP
16	ARG	116	ASP
18	HIS	174	GLU
21	HIS	174	GLU
32	GLU	41	LYS
94	ARG	109	ASP
115	LYS	118	GLU
145	LYS	169	ASP
154	ASP	158	LYS
167	ARG	171	ASP
171	ASP	172	LYS

Table 2: Ionic Interactions within 6 Å (structure in dynamic)

Position	Residue	Position	Residue
16	ARG	116	ASP
19	ARG	107	ASP
19	ARG	26	ASP
26	ASP	29	GLU
29	GLU	30	GLU
29	GLU	32	GLU
32	GLU	33	GLU
32	GLU	41	LYS
33	GLU	41	LYS
38	LYS	39	GLU
64	ARG	65	GLU
94	ARG	109	ASP
94	ARG	112	ASP
109	ASP	112	ASP
112	ASP	115	LYS
115	LYS	116	ASP
115	LYS	118	GLU
116	ASP	119	GLU
118	GLU	119	GLU
129	GLU	130	ASP
129	GLU	134	ARG
130	ASP	134	ARG
153	ASP	154	ASP
168	LYS	169	ASP
168	LYS	171	ASP
169	ASP	172	LYS
171	ASP	172	LYS
171	ASP	174	GLU
174	GLU	178	ARG

The aromatic-aromatic interactions also play important role in the structural stability of proteins. These interactions are significant within 4.5 to 7 Å range between aromatic residues. In human growth hormone protein, we have detected 9 and 5 aromatic-aromatic interactions for the dynamic and static structures respectively. These interactions are presented in tables 3 and 4. Moreover, there is an aromatic-sulphur interaction within 5.3 Å present between PHE166 and MET170 residues in both

structures. In static structure, the distance between the centroid and sulphur was observed 4.66 Å and the angle was 72.7°, whereas in dynamic structure the distance between the centroid and sulphur is 5.01 Å and the angle is 44.6°.

The cation-pi interactions occur between the cations of the side chains of arginine or lysine residues with the polarizable pi electron of the aromatic ring. These interactions are distance

dependent interactions. Here, we have taken the cut off distance of 6 Å. In static structure, there is only one cation-pi interaction present between the residues LYS41 and TYR164 within the distance of 5.8 Å and at an angle of 119.7°. However, in case of structure obtained from dynamics, we observed 7 cation-pi interactions between the ARG and LYS residues with the aromatic residues PHE and TRP as shown in table 5.

Table 3: Aromatic-Aromatic Interactions within 4.5 and 7 Å (in static structure).

Position	Residue	Position	Residue	D(Centroid-Centroid)	Dihedral Angle
25	PHE	28	TYR	6.93	122.2
31	PHE	35	TYR	4.73	132.8
54	PHE	143	TYR	6.05	25.35
86	TRP	166	PHE	4.93	50.11
160	TYR	164	TYR	6.24	95.19

Table 4: Aromatic-Aromatic Interactions within 4.5 and 7 Å (structure in dynamic).

Position	Residue	Position	Residue	D(Centroid-Centroid)	Dihedral Angle
25	PHE	28	TYR	6.14	144.03
28	TYR	160	TYR	4.74	8.09
28	TYR	164	TYR	6.07	31.76
31	PHE	35	TYR	5	118.96
54	PHE	143	TYR	6.74	50.35
86	TRP	166	PHE	5.63	51.6
86	TRP	97	PHE	6.4	22.9
97	PHE	146	PHE	5.18	123.48
97	PHE	166	PHE	4.79	37.53

Table 5: Cation-Pi Interactions within 6 Å (structure in dynamic)

Position	Residue	Position	Residue	D (Cation-Pi)	Angle
1	PHE	16	ARG	5.15	108.14
1	PHE	8	ARG	5.32	119.02
25	PHE	167	ARG	3.9	13.26
42	TYR	41	LYS	4.54	141.61
111	TYR	94	ARG	3.98	156.72
139	PHE	77	ARG	3.93	149.17
191	PHE	64	ARG	4.45	47.09

4. CONCLUSIONS

We have carried out the molecular dynamics (MD) simulations in order to identify the intra-molecular

contacts and their contributions in forming the stable structure of human growth hormone. The MD run was propagated for 100 ns simulation

under NVT condition at body temperature 310 K. We have studied bonded and non-bonded interactions which contribute in the formation of stable structure in aqueous environment. Many non-bonded contacts are compared in static and dynamic conditions.

The decreasing nature of solvent accessible surface area (SASA) depicts the hydrophobic nature of hGH which makes the structure more stable in water. We have compared the various intra-molecular interactions namely disulphide bonds, hydrophobic, ionic, aromatic-aromatic, aromatic-sulphur, cation-pi interactions between the structures from dynamics and statics. In the structure taken from the simulation at the condition of minimum SASA, we have found hydrophobic, ionic, cation-pi, aromatic-aromatic interactions significantly higher than that of static structure. The hydrogen bonding is observed strongly contributing in intramolecular binding. In addition, the estimation of bonded and non-bonded energy profiles show that electrostatic interactions have higher contributions in non-bonded condition. Similarly, the harmonic angle and dihedral angle have almost equal contribution in the conformation of the molecule. Bond energy contribution has also significant role but smaller than that of angle and dihedral.

ACKNOWLEDGEMENTS

RPK and SPK acknowledge the partial financial support from the Nepal Academy of Science and Technology (NAST). NPA acknowledges the UGC Award No. CRG-73/74-S&T-01. We acknowledge the computing facilities of Supercomputer Center Kathmandu University, which was established with equipment donated by CERN and the Arkansas High Performance Computing Center which is funded through multiple National Science Foundation grants and Arkansas Economic Development Commission.

REFERENCES

- [1] Fayter, D.; Nixon, J.; Hartley, S.; Rithalia, A.; Butler, G.; Rudolf, M.; Glasziou, P.; Bland, M.; Stirk, L.; and Westwood, M. Effectiveness and cost-effectiveness of height-screening programmes during the primary school years: a systematic review, *Archives of disease in childhood*, **93**: 278-284 (2008).
- [2] Greenwood, F.; and Landon, J. Growth hormone secretion in response to stress in man, *Nature*, **210**: 540-541 (1966).
- [3] Aloj, S.; and Edelhoeh, H. The molecular properties of human growth hormone, *Journal of Biological Chemistry*, **247**: 1146-1152 (1972).
- [4] Kohler, M.; Püschel, K.; Sakharov, D.; Tonevitskiy, A.; Schänzer, W.; and Thevis, M. Detection of recombinant growth hormone in human plasma by a 2-D PAGE method, *Electrophoresis*, **29**: 4495-4502 (2008).
- [5] Li, C. H. Human growth hormone: 1974–1981, *Molecular and cellular biochemistry*, **46**: 31-41 (1982).
- [6] Yi, S.; Bernat, B.; Pál, G.; Kossiakoff, A.; and Li, W.-H. Functional promiscuity of squirrel monkey growth hormone receptor toward both primate and nonprimate growth hormones, *Molecular biology and evolution*, **19**: 1083-1092 (2002).
- [7] Koirala, R. P.; Pradhan, S.; and Aryal, S. K. Ultrasonic Measurement of Kidney Length in Nepalese People. *Journal of Nepal Physical Society*, **4**(1): 49-53 (2017).
- [8] Wells, J. A.; Cunningham, B. C.; Fuh, G.; Lowman, H. B.; Ultsch, M.; Devos, A. M.; Bass, S. H.; Mulkerrin, M. G. The molecular basis for growth hormone–receptor interactions, In *Recent Progress in Hormone Research*, 253-275, Elsevier (1993).
- [9] Meiering, E. M.; Serrano, L.; and Fersht, A. R. Effect of active site residues in barnase on activity and stability, *Journal of molecular biology*, **225**: 585-589 (1992).
- [10] Schulga, A. A.; Makarov, A. A.; Levichkin, I. V.; Belousova, Y. V.; Lobachov, V. M.; Protasevich, I. I.; Pace, C. N.; and Kirpichnikov, M. P. Increased stability of human growth hormone with reduced lactogenic potency, *FEBS letters*, **528**: 257-260 (2002).
- [11] Burley, S.; and Petsko, G. A. Aromatic-aromatic interaction: a mechanism of protein structure stabilization, *Science*, **229**: 23-28 (1985).
- [12] Bhattacharyya, R.; Samanta, U.; and Chakrabarti, P. Aromatic–aromatic interactions in and around α -helices, *Protein engineering*, **15**: 91-100 (2002).
- [13] Dahiyat, B. I.; and Mayo, S. L. Protein design automation, *Protein Science*, **5**: 895-903 (1996).
- [14] Kellis, J. T.; Nyberg, K.; and Fersht, A. R. Contribution of hydrophobic interactions to protein stability, *Nature*, **333**: 784-786 (1988).
- [15] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; and Bourne, P. E. The protein data bank, *Nucleic acids research*, **28**: 235-242 (2000).
- [16] Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L. 3rd, MacKerell, A. D.; Jr., Klauda, J. B.; and Im, W. CHARMM-GUI Input Generator for NAMD,

- GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field, *J Chem Theory Comput*, **12**: 405-413 (2016).
- [17] Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; and MacKerell, A. D., Jr. CHARMM36m: an improved force field for folded and intrinsically disordered proteins, *Nat Methods*, **14**: 71-73 (2017).
- [18] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; and Schulten, K. Scalable molecular dynamics with NAMD, *J Comput Chem*, **26**: 1781-1802 (2005).
- [19] Koirala, R. P.; Bhusal, H. P.; Khanal, S. P.; and Adhikari, N. P. Effect of temperature on transport properties of cysteine in water, *AIP Advances*, **10**: 025122 (2020).
- [20] Ramachandran, S.; Kota, P.; Ding, F.; and Dokholyan, N. V. Automated minimization of steric clashes in protein structures, *Proteins*, **79**: 261-270 (2011).
- [21] Pantha, N.; Chauhan, B.; Sharma, P.; and Adhikari, N. P. Tuning Structural and Electronic Properties of Phosphorene with Vacancies. *Journal of Nepal Physical Society*, **6**(1): 7-15 (2020).
- [22] Walton, E. B.; and Vanvliet, K. J. Equilibration of experimentally determined protein structures for molecular dynamics simulation, *Phys Rev E Stat Nonlin Soft Matter Phys*, **74**: 061901 (2006).
- [23] Humphrey, W.; Dalke, A.; and Schulten, K. VMD: visual molecular dynamics, *J Mol Graph*, **14**: 33-38, 27-38 (1996).
- [24] Tina, K.; Bhadra, R.; and Srinivasan, N. PIC: protein interactions calculator, *Nucleic acids research*, **35**: W473-W476 (2007).
- [25] Neopane, S.; and Pantha, N. First-Principles Study of van der Waals Interactions between Halogen Molecules (Cl₂ and I₂). *Journal of Nepal Physical Society*, **5**(1): 19-23 (2019).
- [26] Thakuria, R.; Sarma, B.; and Nangia, A. 7.03 Hydrogen Bonding in Molecular Crystals, *Comprehensive Supramolecular Chemistry*, **II**: 25-48 (2017).
- [27] DiStasio Jr, R. A.; Gobre, V. V.; and Tkatchenko, A. Many-body van der Waals interactions in molecules and condensed matter, *Journal of Physics: Condensed Matter*, **26**: 213202 (2014).
- [28] Yunta, M. It is important to compute intramolecular hydrogen bonding in drug design, *American Journal of Modeling and Optimization*, **5**: 24-57 (2017).
- [29] Hubbard, R. E.; and Haider, M. K. Hydrogen bonds in proteins: role and strength, *eLS* (2010).

Transport properties of cysteine dimer in water

Research Article

Rajendra Pd. Koirala*, Shyam P. Khanal, Narayan P. Adhikari

Central Department of Physics, Tribhuvan University, Kirtipur, Kathmandu, Nepal

Abstract: Disulphide bond in cysteine residues plays vital role in structural stability and functional variation of protein molecules. Study of cysteine dimer linking with disulphide bond reveals the nature of stability of tertiary and quaternary structure in polypeptide chain. In order to study the transport properties of cysteine dimer, the molecular dynamics (MD) simulations have been performed at different temperature. The self diffusion coefficients of both cysteine dimer and TIP3P water model have been estimated at four different temperature from the slope of mean square displacement (MSD) versus time plot using Einstein's relation and their binary diffusion coefficients from Darken's relation.

Keywords: Molecular dynamics • Diffusion coefficient • Disulphide bond • Dimer

1. Introduction

Cysteine is a non-essential amino acid. It contains a thiol group -SH in its side chain. It is a white crystalline solid having physical properties: molar mass 121.15 gram per mole, melting point 513 K and solubility in water is 16 gram per 100 mL at 288 K [1, 2]. Cysteine plays vital role in the absorption of nutrients in the inner wall of small intestine. Its defensive mechanism against some diseases is essential for human body. Cysteine residue in a protein molecule strengthens the immune system and defends against the dementia, parkinson and multiple sclerosis [3, 4].

Two cysteine molecules bond together via two different methods: peptide bond (CO-NH) and disulphide bond ($R-S-S-R'$) [1], where R & R' are the side chains of two cysteine residues as shown in Fig. 1. Peptide bond is ordinary bond to form a polypeptide chain, whereas the disulphide bond is formed between two side chains derived from two thiol groups. Disulphide bond in two cysteine residues are very important components to form tertiary structure of proteins, besides to some weak interactions like hydrogen bonding, hydrophobic interactions salt bridges and weakly polar interaction. Although methionine, another amino acid, contains sulphur atom in its side chain, it does not form disulphide bond. Hence, the study the characteristics of disulphide bond is important. [1, 5].

* Corresponding Author: rpkoirala@tucdp.edu.np

Due to the oxidation of the sulfhydryl group of cysteine residues in a protein molecule, disulfide bond is formed. In such process, the thiol part -SH of cysteine molecules, contributes in disulfide bond formation, is deprotonated and covalent bond is formed between them [5–7]. Two cysteine molecules also covalently linked through peptide bond. Peptide bond is formed after the elimination of a water molecule during the interaction of carboxyl (C-terminus) and amine (N-terminus) regions of two amino acids. This bond has partial double bond character, i.e., stronger than single bond and weaker than double bond. It prevents the rotation of residues in protein molecules [1, 2]. This work is basically focus on disulphide bonding in cysteine dimer.

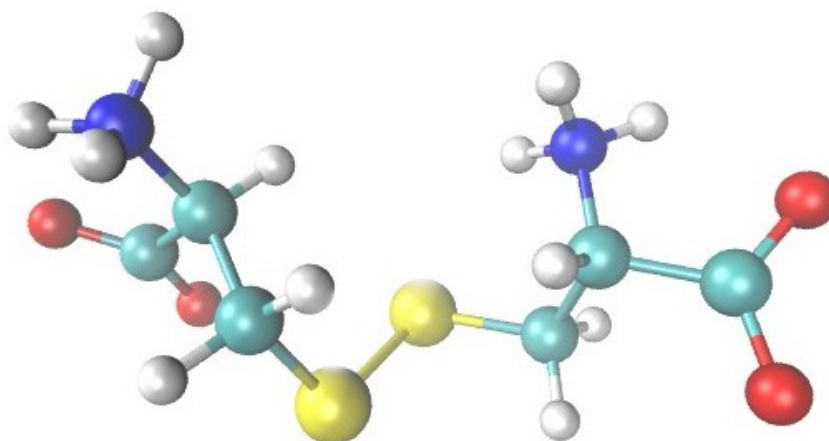


Figure 1. Cysteine dimer.

Diffusion is an important physical phenomenon to understand the rate of transport of mass from the region of higher concentration to lower concentration [8–11]. Many bio-molecules like insulin are stabilized due to the formation of disulphide bond. The detailed understanding of effect of disulphide bond in any molecule and its effect on diffusion phenomenon in aqueous environment plays an important role in different areas including drugs designing. With our best knowledge, the diffusion coefficient of cysteine dimer, covalently bonding with disulphide bond, in water using classical molecular dynamics technique has not been studied yet, which motivate us to study the diffusion phenomenon of cysteine dimer.

2. Methodology

Diffusion

Diffusion is a transport property of matter due to which particles flows from the region of higher concentration to lower concentration due to the concentration inhomogeneity [12]. The knowledge about diffusion phenomenon of biomolecules plays an important role to understand many phenomena in living organisms, like transport of biomolecules at different part of body through body fluids. The diffusion in a homogeneous system in the absence of chemical concentration gradient is called self diffusion and is measured in terms of self diffu-

sion coefficient [13] which is estimated from slope of mean square displacement(MSD) versus time graph using Einstein's equation as [14, 15];

$$D = \lim_{t \rightarrow \infty} \frac{\langle [r(t) - r(0)]^2 \rangle}{6t}. \quad (1)$$

In equation (1), $r(t) - r(0)$ is the displacement of particle from reference point during the course of time t , $[r(t) - r(0)]^2$ is the square of displacement and $\langle \dots \rangle$ represents the ensemble average and hence $\langle [r(t) - r(0)]^2 \rangle$ gives MSD of particle.

Further, the binary diffusion, the diffusion of particles in the mixture of two different species, is measured in terms of binary diffusion coefficient using Darken's relation as [16];

$$D_{12} = N_2 D_1 + N_1 D_2 \quad (2)$$

In equation (2), D_{12} is the binary diffusion coefficient, D_1 and D_2 are the self-diffusion coefficients of substances 1 and 2 respectively, and N_1 and N_2 are the corresponding mole fractions.

Computational Details

In this work, we performed classical molecular dynamics (MD) simulations for the system of 1 cysteine dimer and 1984 water molecules in cubic simulation box of size about 3.90 nm at four different temperature; 288 K, 298 K, 303 K and 308 K using GROMACS 5.1.2 software package [17]. The cysteine dimer was extracted from the insulin molecule (pdb entry: 3i40) taking seventh residue of chain A and also seventh residue of chain B. The input files for MD simulation were generated from CHARMM-GUI online software program [18]. The dimer was solvated in a periodic boundary condition (pbc) box with TIP3P water sample. All the bonded and non-bonded parameters are used, assigned by CHARMM36 force field [19].

MD simulation was begun from energy minimization run using Steepest-descent algorithm [17]. This run removes steric clashes in which the undesirable coordinates of atoms of side chain or backbone that may occupy wrong position in the same coordinate space and bring the system in the condition of minimum potential energy state. Since the transport properties like diffusion depends on the thermodynamic parameters like temperature, pressure etc. of the system [20], the system should be stabilized for thermodynamic parameters. Then, the equilibration run was carried out to stabilize the temperature, pressure, density etc. The equilibrated bio-molecular system saves the computation efforts by removing external unwanted forces [21, 22]. To bring the system in thermodynamics equilibrium, equilibration run of 100 ns with time step of 1 fs was carried out in NPT ensemble for the system at each temperature. During the equilibration run; LINCS algorithms, Berendsen barostat with coupling time of 0.8 ps and Velocity rescaling thermostat with coupling time of 0.01 ps were used to constraint all the bonds, maintain constant pressure and maintain constant temperature respectively [17]. Also, 1 nm cut-off distance was taken for Lennard-Jones and Coulomb interactions; and PME (Particle-mesh Ewald) method was used to handle long range coulomb interaction. Furthermore, Maxwell-Boltzmann distribution and was used to

assign initial velocities of each particle; and the new positions and velocities of the particles was calculated using leapfrog algorithm [17] after each time step respectively.

After equilibration of the system, production run was performed in NVT ensemble for 100 ns with time step of 1 fs for each temperature. During the equilibration run, velocity-rescaling thermostat with coupling time of 0.01 ps was used and the initial velocities of each particle was taken from the final step of equilibration run.

3. Results and Discussion

Diffusion coefficient

In this section, we present the self diffusion coefficients of water and cysteine dimer as well as their binary diffusion coefficient at different temperature. The self diffusion coefficient of both solute as well as solvent are estimated from their MSD versus time graph. Although, all the production run were carried out for 100 ns, we have plotted the MSD graph for 3 ns and 5 ns for cysteine dimer and water respectively due to the region that statistics is better in beginning region of the graph. Figs. 2 and 3 show the msd versus time graph for cysteine dimer and water respectively at four different temperatures: 288 K, 298K, 303 K and 308 K. From the Figures, it is clearly seen that the slope of the graph increases with temperature for both cysteine dimer as well as water. This indicates that self diffusion increases with temperature.

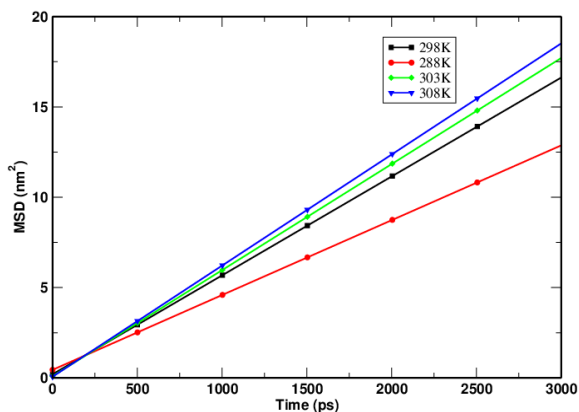


Figure 2. MSD versus time plot of cysteine dimer at four different temperature.

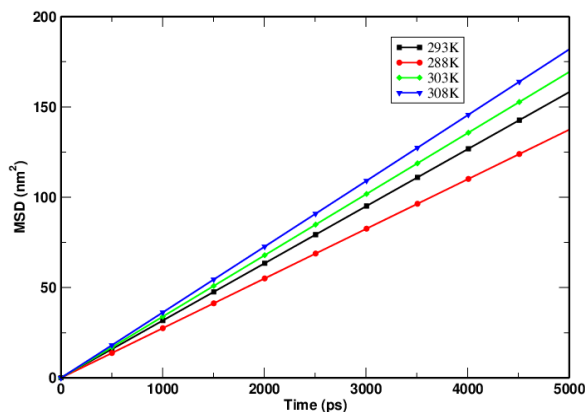


Figure 3. MSD versus time plot of water at four different temperature.

Also, the estimated values of the self diffusion coefficients of cysteine dimer and water estimated from the slope of the MSD versus time graph using Einstein relation; and their binary diffusion coefficients using Darken's relation are presented in Table 1.

From the Table 1, it is seen that both self and binary diffusion coefficients increases with increase in temperature. This is due to the increase in random velocity of the particles and decrease in density of the system with increase in temperature. As a result, the available space for diffusion increases with temperature. The simulated value of self diffusion coefficient at 298 K is in closely agreement with previously reported simulated value

[23]. Also, from the Table, we observe that the estimated values of self diffusion coefficient of water at different temperature are greater than the previously reported experimental values. The reason for this observation is that we used TIP3P water model during the simulations which overestimate the value of diffusion coefficient. Furthermore, we observed that the binary diffusion coefficient of the system equals to corresponding value of self diffusion coefficient of cysteine dimer, which is due to infinite dilute concentration of cysteine dimer in the solution.

Table 1. Estimated values of self and binary diffusion coefficients at different temperature.

SN	Temp. (K)	Diffusion coefficients (D_{PBC})($10^{-9}m^2 s^{-1}$)				
		Self				Binary Calculated
		For cysteine dimer		For Water		
		MSD	MSD Reference[23]	Experiment [24]		
1.	288	0.69	4.58		1.77	0.69
2.	298	0.91	5.27	5.4	2.30	0.91
3.	303	0.98	5.65		2.60	0.98
4.	308	1.02	6.06		2.90	1.02

Temperature Dependency of Diffusion

The Table 1 shows the temperature dependent behavior of diffusion coefficient. In order to check whether the behavior is Arrhenius or not, we have plotted graphs between $\ln(D)$ versus $(1/T)$. Fig. 4 shows the temperature dependent behavior of self diffusion of water which follows the Arrhenius equation [25]:

$$\ln D = \ln D_0 - \frac{E_a}{N_A k_B T} \quad (3)$$

In equation (3), D is the diffusion coefficient, D_0 represents pre-exponential factor, E_a is the activation energy for diffusion, N_A is Avogadro's number whose value is $6.022 \times 10^{23} \text{ mol}^{-1}$, k_B is the Boltzmann's constant whose value is $1.38 \times 10^{-23} \text{ JK}^{-1}$ and T is the absolute temperature. The intercept when extrapolated to the $1/T \rightarrow 0$ in the Arrhenius plot gives the pre-exponential factor.

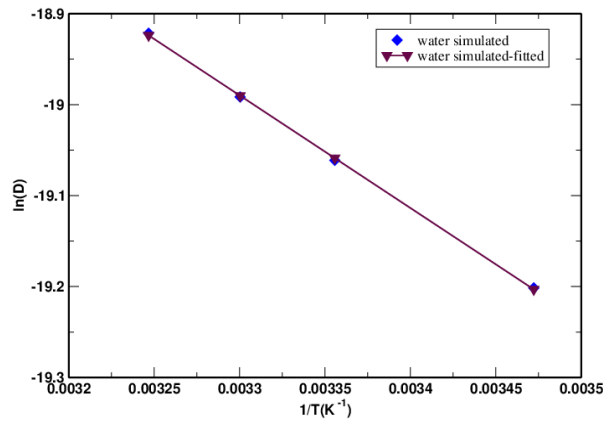


Figure 4. Arrhenius diagram for simulated values of self diffusion coefficient of water.

4. Conclusions

In this work, molecular dynamics simulation has been carried out to study the diffusion properties of cysteine dimer with disulphide bond in infinite dilute solution of water. This study is performed solvating single cysteine dimer molecule in 1984 water molecules at temperature 288 K, 298 K, 303 K, 308 K using GROMACS 5.1.2 software package. CHARMM36 force field parameters and TIP3P water model are used throughout the simulations. Einstein's equation is used to determine the self diffusion coefficient of cysteine dimer and the water. The estimated values of self diffusion coefficients are found to be higher than the previously reported experimental values. This is due to the fact that TIP3P water model overestimate the diffusion coefficient. Also, the diffusion of binary mixture of cysteine dimer and water was calculated from Darken's relation. The temperature dependence of diffusion coefficient has been tested from Arrhenous plot.

In near future, we are intended to study the transport properties of poly-cysteine with peptide bond in aqueous medium.

5. Acknowledgements

RPK & SPK acknowledge the partial financial support from Nepal Academy of Science and Technology (NAST). NPA acknowledges the UGC Award no. CRG-73/74-S&T-01.

References

- [1] Bulaj G. Formation of disulfide bonds in proteins and peptides. *Biotechnology advances*. 2005;23(1):87–92.
- [2] Wedemeyer WJ, Welker E, Narayan M, Scheraga HA. Disulfide bonds and protein folding. *Biochemistry*. 2000;39(15):4207–4216.
- [3] Canet-Avilés RM, Wilson MA, Miller DW, Ahmad R, McLendon C, Bandyopadhyay S, et al. The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization. *Proceedings of the National Academy of Sciences*. 2004;101(24):9103–9108.
- [4] Rizvi SI, Jha R. Strategies for the discovery of anti-aging compounds. *Expert opinion on drug discovery*. 2011;6(1):89–102.
- [5] Privalov P, Khechinashvili N. A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. *Journal of molecular biology*. 1974;86(3):665–684.
- [6] Burley S, Petsko G. Weakly polar interactions in proteins. In: *Advances in protein chemistry*. vol. 39. Elsevier; 1988. p. 125–189.
- [7] Sticke DF, Presta LG, Dill KA, Rose GD. Hydrogen bonding in globular proteins. *Journal of molecular biology*. 1992;226(4):1143–1159.

- [8] Thapa S, Adhikari N. A molecular dynamics study of oxygen gas in water at different temperatures. *International Journal of Modern Physics B*. 2013;27(08):1350023.
- [9] Bhandari D, Adhikari N. Molecular dynamics study of diffusion of krypton in water at different temperatures. *International Journal of Modern Physics B*. 2016;30(11):1650064.
- [10] Ghimire S, Adhikari NP. Study of structural and transport properties of argon, krypton, and their binary mixtures at different temperatures. *Journal of molecular modeling*. 2017;23(3):94–107.
- [11] Poudyal I, Adhikari NP. Temperature dependence of diffusion coefficient of carbon monoxide in water: A molecular dynamics study. *Journal of Molecular Liquids*. 2014;194:77–84.
- [12] Crank J. *The mathematics of diffusion*. Oxford University Press; 1975.
- [13] Hirakawa H, Kamei Y, Sugisaki M, Oishi Y. Relationship between Self-Diffusion and Interdiffusion in Gaseous Systems. *Bulletin of the Chemical Society of Japan*. 1973;46(9):2659–2662.
- [14] Frenkel D, Smit B. *Understanding molecular simulation: from algorithms to applications*. vol. 1. Elsevier; 2002.
- [15] Allen M, Tildesley D. *Computer Simulations of Liquids*, Oxford: Clarendon Press. Oxford University Press; 1989.
- [16] Darken LS. Diffusion, mobility and their interrelation through free energy in binary metallic systems. *Trans Aime*. 1948;175:184–201.
- [17] Abraham M, van der Spoel D, Lindahl E, Hess B, team atGd. *GROMACS User Manual Version 5.1*. 2. GROMACS Development Team. 2016;.
- [18] Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, et al. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *Journal of chemical theory and computation*. 2016;12(1):405–413.
- [19] Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry*. 2010;31(4):671–690.
- [20] Khanal SP, Kandel YP, Adhikari NP. Transport properties of zwitterion glycine, diglycine, and triglycine in water. *AIP Advances*. 2019;9(6):065303.
- [21] Walton EB, VanVliet KJ. Equilibration of experimentally determined protein structures for molecular dynamics simulation. *Physical Review E*. 2006;74(6):061901.
- [22] Gallo MT, Grant BJ, Teodoro ML, Melton J, Cieplak P, Phillips Jr GN, et al. Novel procedure for thermal equilibration in molecular dynamics simulation. *Molecular simulation*. 2009;35(5):349–357.
- [23] Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *The Journal of Physical Chemistry A*. 2001;105(43):9954–9960.
- [24] Holz M, Heil SR, Sacco A. Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate ¹H NMR PFG measurements. *Physical Chemistry Chemical Physics*.

2000;2(20):4740–4742.

- [25] Mehrer H. Diffusion in solids: fundamentals, methods, materials, diffusion-controlled processes. vol. 155. Springer Science & Business Media; 2007.

BIBECHANA

ISSN 2091-0762 (Print), 2382-5340 (Online)

Journal homepage: <http://nepjol.info/index.php/BIBECHANA>

Publisher: Department of Physics, Mahendra Morang A.M. Campus, TU, Biratnagar, Nepal

Diffusion of oxytocin in water: a molecular dynamics study

Khimananda Acharya, Rajendra Prasad Koirala, Nurapati Pantha*

Central Department of Physics, Tribhuvan University, Kathmandu, Nepal

*Email: mrnurapati@gmail.com

Article Information:

Received: June 6, 2020

Accepted: July 11, 2020

Keywords:

Diffusion

Molecular Dynamics

Mean Square Displacement

Activation energy

ABSTRACT

Classical molecular dynamics simulation is performed to estimate the diffusion coefficient of oxytocin in water at different temperatures, 288 K, 300 K, 313 K, 323 K, using GROningen Machine for Chemical Simulations (GROMACS). The simulation is carried out using GROMOS43A1 force field and extended simple point charge (SPC/E) water model. The stability of the system is evaluated from energy profile of potential and kinetic energy, which assures well equilibrated molecular system. The self-diffusion coefficient of oxytocin and water is obtained from Einstein's relation and binary diffusion coefficient is obtained from Darken's relation. As temperature increases the diffusion coefficient also increases as per expectation. The diffusion coefficients of water from the present calculations agree well with the previously reported values, within the 10% of deviation. Furthermore, the activation energy has been studied using Arrhenius Plot.

DOI: <https://doi.org/10.3126/bibechana.v18i1.29316>

This work is licensed under the Creative Commons CC BY-NC License. <https://creativecommons.org/licenses/by-nc/4.0/>

1. Introduction

Biomolecules are the building blocks of living organisms. They have the specific physio-chemical configurations [1]. The structural variability and chemical conformation of the molecules determine the biological functioning in living cells. Biomolecules work independently as well as cooperatively to perform in various body mechanisms [2]. Proteins, nucleic acids, lipids and carbohydrates are the basic category of

biomolecules, also called the macromolecules. Some other molecules are also characterized into this class; primary metabolites, secondary metabolites, natural and artificially synthesized products [3,4].

The poly-condensation of amino acids residues forms a protein molecule. The carboxyl (-COOH) tail of an amino acid when covalently bonded to amine (NH₂-) head of another amino acid by water elimination, a dipeptide molecule is formed [5,6].

Then, the condensation of other amino acids one after another in new carboxyl tail, a long polypeptide chain is formed, which will finally be a functional protein molecule. A long polypeptide chain cannot stay in linear form, i.e., primary structure, because of difficulty in structural stability [7]. To minimize the overall energy profile, the molecule folds into various shapes, basically in alpha-helices and beta-turns. Hydrogen bonds, salt bridges; and hydrophobic and hydrophilic interaction takes place to provide different stable structures of protein molecules [4].

Oxytocin is a short peptide chain of nine amino acids (a nanopeptide) with the sequence cysteine-tyrosine-isoleucine-glutamine-asparagine-cysteine-proline-leucine-glycine (Cys Tyr Ile Gln Asn Cys Pro Leu Gly, or CYIQNCPLG) with additional amide (-NH₂) in the carboxyl terminus of glycine residue [8] as shown in Fig 1(i) and its space-filling calotte (CPK) model for the visualization is shown in Fig. 1(ii). Its molecular formula is C₄₃H₆₆N₁₂O₁₂S₂ with its molecular weight 1007 g per 193 moles (i.e., 1007:193 g/mole). It has white powdery color and is soluble in water and in butanol. A disulphide bridge is present first and sixth cysteine residues of the molecule [9,10].

Oxytocin is also known as love hormones responsible for the feeling love and satisfaction. It plays role in social bonding, sexual reproduction in both sexes, milk production etc. It is best known for its role in childbirth and breast feeding [12]. During labor large amount of oxytocin release causing contraction of the uterus to facilitate birth. When the contractions are not strong to compress the blood vessels, postpartum hemorrhage (PPH) can menace women's life. In this condition uterotonic medicine like oxytocin is given to stimulate contraction and stop the bleeding [11,13].

Oxytocin hormone is synthesized in pituitary gland. It is, then, released to the blood stream through the

posterior lobe of the gland. Some neurons and other parts of brain also produce oxytocin [14,15]. Liquid contents of living body, like water, contribute to transport these hormones to various parts of the cell [16]. We believe that the study of transport properties of this hormone in water is very important to understand its mechanical properties in aqueous medium. Furthermore, diffusion coefficient of nitric oxide in water had been observed through the same technique [17]. To our best knowledge, the transport properties of oxytocin in water have not been studied yet. We have performed the molecular dynamics to deal the transport properties of oxytocin.

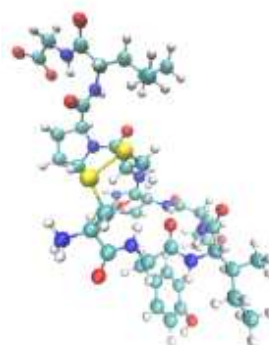
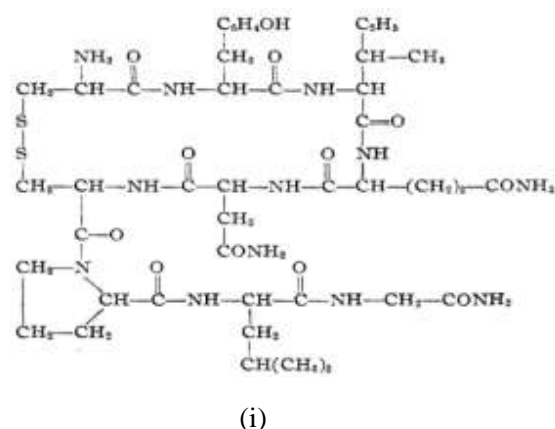


Fig.1: (i) Molecular structure of an oxytocin molecule [11] (ii) CPK model of the oxytocin molecule.

Diffusion

Diffusion is the random process in which the matter is transported from one part of a system to another.

Matter is transported always from higher concentration to the lower concentration. This occurs often due to the concentration inhomogeneity and thermal agitation of the particles in a medium. Diffusion is an imperative process within the human body and it is indispensable to transport of molecule within organ like lungs, kidneys, stomach, eyes etc. We realize several phenomena regarding the diffusion; like mixing of smoke in air, sugar in water, dust in air etc [18,19]. In this work, we estimate the self-diffusion and binary diffusion coefficient of oxytocin in water.

The diffusion in a homogeneous system without any chemical concentration gradient exists in it is referred the self-diffusion and the corresponding diffusion coefficient is known as self-diffusion coefficient. The extent of diffusion is physically measured in terms of self-diffusion coefficient. It is calculated using Einstein's relation [20,21],

$$D = \lim_{t \rightarrow \infty} \frac{\langle [r(t) - r(0)]^2 \rangle}{6t} \quad (1)$$

The term $r(t) - r(0)$ in equation (1) is the displacement of particle from the reference position at time t , $[r(t) - r(0)]^2$ is the square of displacement, and $\langle \dots \rangle$ represents the ensemble average, thus, the expression $\langle [r(t) - r(0)]^2 \rangle$ is the mean square displacement (MSD) of the particle.

For the system with two different substances, binary diffusion takes place. It is measured in terms of diffusion coefficient called binary or mutual diffusion coefficient [22].

$$D_{AB} = N_B D_A + N_A D_B \quad (2)$$

where, D_{AB} is Binary diffusion coefficient, D_A , D_B = Self diffusion coefficient of species A and B respectively, and N_A , N_B = mole fraction of species A and B respectively.

The diffusion phenomenon is the temperature dependence phenomena. The mathematical relation for the temperature dependence is given by Arrhenius equation [23].

$$D = D_0 e^{\frac{-E_a}{RT}} \quad (3)$$

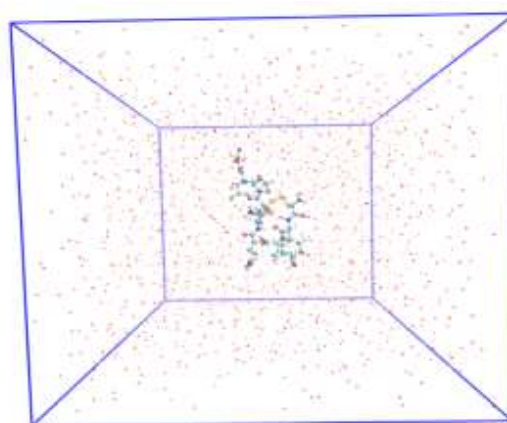
Where, D is diffusion coefficient, E_a is activation energy, D_0 is pre-exponential factor, R is universal gas constant and T is absolute temperature.

2. Computational Details

System setup

A PDBID 2MGO.pdb was taken from the protein data bank (pdb), an authorized pdb website (www.rcsb.org). The molecule was in complete form as our requirement, so no modification was done in the original structure. The topology and parameters for the system setup and molecular dynamics (MD) simulations were taken from GROMOS43A1 force field [23]. While considering the water model, we should care on its compatibility with the force field parameter we have used [25]. The molecule was solvated into a cubic box of $4.07 \times 4.07 \times 4.07 \text{ nm}^3$ dimension with SPC/E [26] water model by GROMACS 5.1.1 [27] command. The molecule was naturally chargeless, so extra ions have not been added. There were 2167 water molecules and one oxytocin molecule in the simulation box as shown in Fig. 2.

Fig. 2: Oxytocin molecule solvated in a PBC box.



Molecular dynamics simulations

All atom molecular dynamics simulation was propagated to study the various mechanical properties of Oxytocin molecule in water.

GROMOS43A1 force field was used for the entire simulation. Three basic steps: energy minimization, system equilibration and production run were carried out to deal the targeted properties.

The original pdb structure has been estimated from x-ray diffraction (XRD) or nuclear magnetic resonance (NMR) techniques. The assigned coordinates for the atoms in the molecule might have steric hindrance due to the mechanical errors. Moreover, hydrogen atoms are absent in original pdb, and they are added by the molecular system building software. Therefore, before begin our MD run, we must be sure that there is no any inappropriate geometry otherwise force acting on particle will be very large and system will be out of equilibrium and MD fails. In order to get solved this inappropriate geometry and kinetic energy; system must be subjected to the process of energy minimization. In our system because of fast and reliable performance, we have decided to choose steepest-descent algorithm [29].

Energy minimization run is often performed in absolute zero Kelvin temperature, which is practically unachievable and unreliable for the natural environment. So, the molecular environment is so created that can be resembled in the practice in room temperature or for any natural process. We basically maintain the system that follows the statistical ensembles with appropriate macroscopic parameters; pressure, temperature and volume [30,31]. Pressure coupling was made to set pressure 1 bar using barendsen barostat. For box re-scaling isothermal compressibility is set to $4.5 \times 10^{-5} \text{ bar}^{-1}$. Particle Mesh Ewald (PME) for coulomb interaction with cut off distance 1.0 nm and Fourier spacing of 0.3 was used. V-rescaling has been used for temperature coupling. Equilibration run was performed at four different temperatures 288 K, 300 K, 313 K, and 323 K and was run 20 ns for each system.

The appropriate system prepared from the equilibration run was propagated to study the molecular diffusion in water at assigned

temperature above. Production run was performed in NVT ensemble. Each of the system was run 60 ns time period.

3. Results and Discussion

We present and discuss about the energy profiles, self-diffusion coefficients of both water and oxytocin; and the output of self-diffusion coefficient is used to evaluate binary diffusion coefficients. Finally, the activation energy of corresponding systems is estimated for each system.

Energy minimization and system equilibration

The fundamental requirement of MD simulation is to bring the molecular system in minimum potential energy so that the atoms in the molecule get free from atomic hindrance. The best way to achieve this condition is energy minimization run and analyzing the outcomes. Energy minimization run is always run at absolute zero temperature, which ensures the minimum potential energy. Fig. 3 shows the potential energy curve during the minimization run. The potential energy has been well converged and we confirmed the system to proceed for the post processing. We found potential energy $-1.2015872 \times 10^4 \text{ kJ mol}^{-1}$ in the most condition and steepest descent converged to $F_{\max} < 50 \times 10^4 \text{ kJ mol}^{-1}$.

Before commencing the MD production run, the molecular system should be maintained at a thermodynamic ensemble. This can be done by MD equilibration run and the outcomes were evaluated determining some thermodynamic parameters like pressure, temperature etc. These parameters have been assigned to resemble the cellular environment in human body. The best way of finding the stability of the molecular system is the calculation of density, which was not assigned for the system. We have calculated density of water and entire system including the oxytocin after equilibration run. The variation of density of water and entire system at different temperatures has been presented in table 1. Because of infinite dilution, the mixture

has similar order of density with that of pure water. The decreasing pattern of density on increasing temperature ensures that the system is appropriately equilibrated and is suitable for the further calculation.

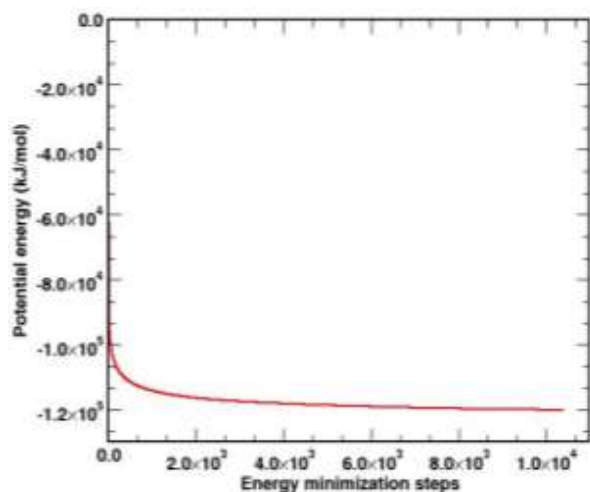


Fig. 3: Potential energy after energy minimization.

Energy profile

We have studied energy profile under the GROMOS43A1 force field in MD simulation. When the system is in MD simulations, several interactions have been taken to calculate energy. The energy profile is contributed by many bonded and non-bonded interactions. Bond interactions are bond stretching, bond angle vibration, proper dihedral and improper dihedral whereas non-bonded interactions are coulomb interaction and L-J interaction. The energy profile is the graphical manifestation of these various energy interactions.

To reduce the computational cost, we have kept bond length constant in our simulation with the help of constraint algorithm LINCS. Therefore, there is no energy profile due to bond stretching. Also we found that bonded interaction energy was positive. The coulomb interaction and L-J interaction are the major contribution on the non-bonded interaction.

These interactions were calculated using the list of non-bonded atom within the certain region. We use periodic boundary condition for such a purpose. The SR coulomb interaction has negative value and has dominant contribution on total potential energy. The value of L-R coulomb and coulomb reciprocal energy are significantly small. Within the cutoff region, the L-J interaction has significant role, however, (L-R) L-J interaction has negligible contribution on total energy. The energy profiles of the system at four different temperatures have been studied. The energy profile for all the interaction energy with total energy at 300 K temperature is shown in Fig. 4.

Self-diffusion coefficient of water and oxytocin

Self-diffusion coefficients of water and oxytocin have been calculated at four different temperatures; 288 K, 300 K, 313 K, 323 K using the Einstein's relation in equation (1). The slope of corresponding mean square displacement (MSD) was utilized to determine the self-diffusion coefficients in both systems. Fig. 5(i) and 5(ii) show the mean square displacement curves for water and oxytocin respectively.

We have carried out 60 ns MD simulation for each temperature; however the linear relationship has to be required to draw the value of slope in MSD versus time graph. So, we have selected 2 ns simulation run in the beginning to get the better statistics. The MSDs, thus, obtained were found linearly increasing with respect to time, and in turn increased the self-diffusion coefficients. The self-diffusion coefficient of water was compared with previously reported experimental value and was found good agreement within 10%. The better agreement at higher temperature indicates that the force field parameters could have been calculated in these temperature regimes which are relevant to physiological phenomenon.

Table 1: Temperatures and corresponding densities of system after equilibration run.

Coupling temperature (K)	Equilibrium temperature (K)	Equilibrium density (kg m ⁻³)	Density of water (kg m ⁻³) [32]
288	287.983±0.007	1008.71±0.07	999.13
300	299.983±0.007	1002.87±0.05	996.56
313	313.005±0.011	995.46±0.02	992.28
323	323.995±0.011	998.96±0.03	988.11

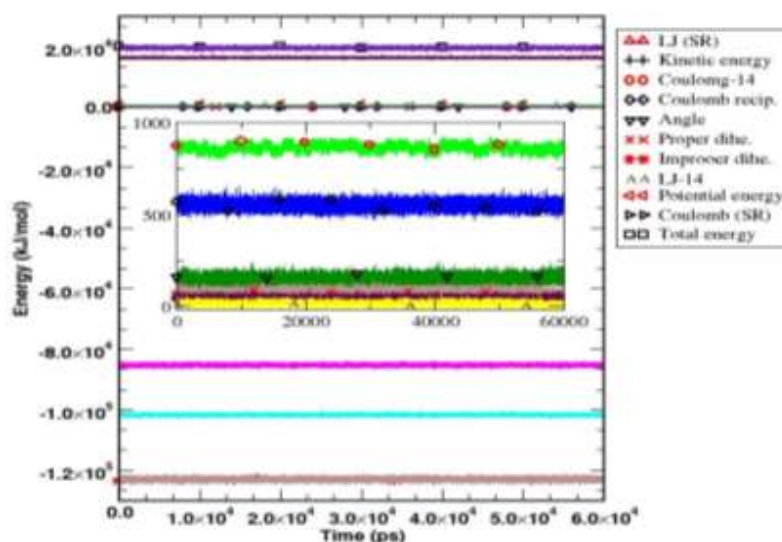


Fig. 4: Total energy profile of oxytocin at 300 K.

The simulated values and experimental values of self-diffusion coefficient of water at different temperatures are listed in the table 2. We also have calculated self-diffusion coefficient of oxytocin from the MSD graph of oxytocin molecule. The simulated values of self-diffusion coefficient of oxytocin at different temperatures are listed in the table 3. In both systems, the self-diffusion coefficients are obtained as, $D_{323\text{ K}} > D_{313\text{ K}} > D_{300\text{ K}} > D_{288\text{ K}}$.

While simulating macromolecule/s in a box, one needs to be careful about its size effect. It is a usual practice to consider the smallest possible simulating box and enlarge the space for moving molecules by applying periodic boundary condition (PBC). However, the size of the box should be sufficient to maintain the basic criteria of molecular simulations. The rule of thumb for this requirement says that the cut-off distance of different interactions under consideration should be less than the half the box size [24]. Furthermore, previous

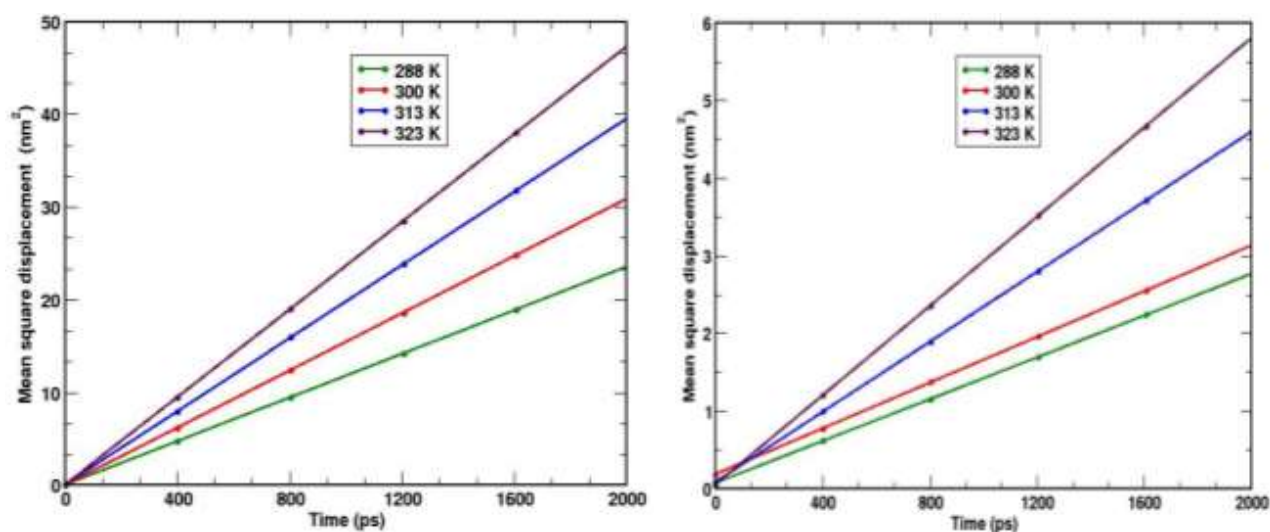


Fig. 5: MSD vs time plot of (i) water (ii) oxytocin at different temperatures.

Table 2: Simulated and experimental self-diffusion coefficients of water at four different temperatures.

Temperature K	Simulated diffusion coefficient D_w^s (10^{-10} m ² /s)	Experimental diffusion coefficient D_w^E (10^{-10} m ² /s) [33]	Difference %
288	19.16±0.05	17.77	9.39
300	25.72±0.05	-	-
313	32.87±0.10	32.40	1.43
323	39.39±0.08	39.68	0.74

studies have shown that as far as these minimum criteria are maintained, the size of the box does not affect the main conclusions of the results [34]. We have taken care of size of the oxytocin molecule and also the cutoff distances, which are less than half-size of the simulating box in each dimension.

Binary diffusion coefficient

The self-diffusion coefficients obtained for water and oxytocin were used to find the binary diffusion

coefficients of oxytocin in water by using Darken's relation in equation (2). The calculated values of binary diffusion coefficients were found increasing on increasing the temperature. When temperature increases there is increase in thermal energy which results the decrease in density of the system and hence there is increase in diffusion coefficient. The estimated binary diffusion coefficients of oxytocin in water have been presented in table 4.

Table 3: Simulated values of self-diffusion coefficients of oxytocin at four different temperatures.

Temperature K	Self-diffusion Coefficient D_0^s (10^{-10} m ² /s)
288	2.24±0.32
300	2.45±0.29
313	3.75±0.26
323	4.79±0.01

Table 4: Simulated binary diffusion coefficients at four temperatures.

Temperature K	Binary Diffusion Coefficient D_0^s (10^{-10} m ² /s)
288	2.23
300	2.46
313	3.76
323	4.81

Temperature dependency of diffusion

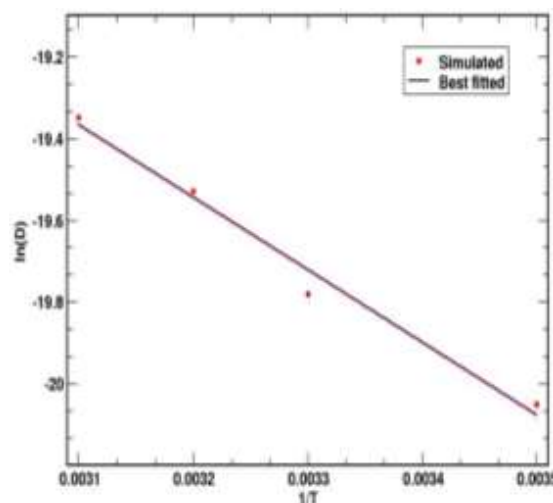
We obtain activation energy from the slope of diffusion versus reciprocal of time as suggested by the equation (3), $D = D_0 e^{\frac{-E_a}{N_A k_B T}}$, and $R = N_A k_B$, where N_A is Avogadro number whose value is 6.023×10^{23} mol⁻¹ and k_B is the Boltzmann constant carrying the value 1.38×10^{-23} JK⁻¹, meaning of remaining parameters are explained in equation (3). On taking the natural logarithm for the expression, we get,

$$\ln D = \ln D_0 - \frac{E_a}{N_A k_B T} \tag{4}$$

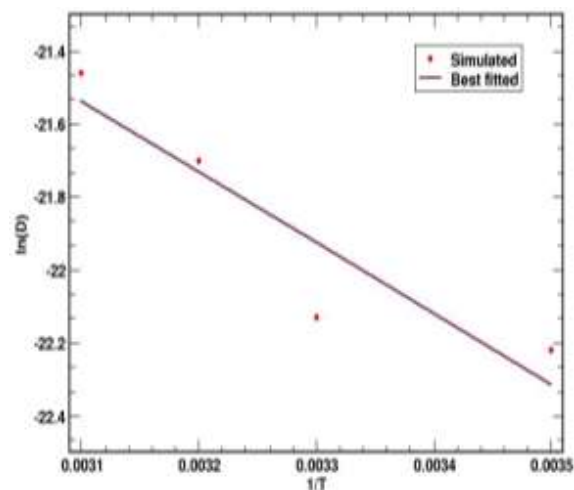
Then, the activation energy E_a for diffusion was obtained from the slope of $\ln D$ versus $\frac{1}{T}$, called the Arrhenius plot as,

$$E_a = -N_A k_B \frac{\partial \ln D}{\partial (1/T)} \tag{5}$$

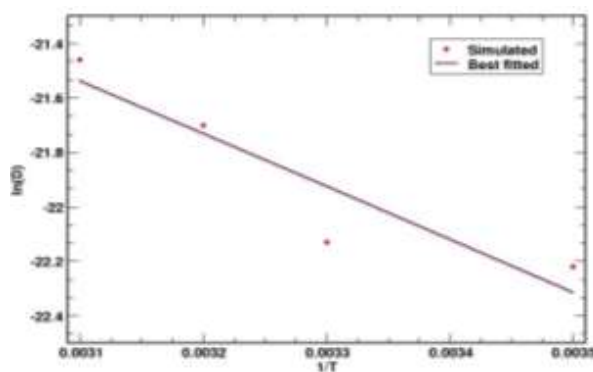
The intercept when extrapolated to the $\frac{1}{T} \rightarrow 0$ in the Arrhenius plot gives the pre-exponential factor.



(i)



(ii)



(iii)

Fig. 6: Arrhenius diagram for self-diffusion coefficient of (i) Oxytocin, (ii) Water and (iii) Binary mixture.

The Arrhenius plot of simulated value of water is shown in Fig. 6(i), and that of oxytocin is 6(ii) and for their binary mixture is shown in Fig. 6(iii), which were used to determine the activation energy for each condition. From the table 5, it is clear that the activation energy for oxytocin molecule and binary mixture is found to be almost equal; this is due to the infinitesimal concentration of oxytocin in our system. Further, we found that there is an acceptable level of agreement between the simulated and experimental values of activation energy of water.

4. Conclusions and concluding remarks

The self-diffusion coefficient of water and oxytocin is obtained using Einstein's relation. The binary diffusion coefficient is obtained using Darken's relation. The results obtained for energy profile, the kinetic energy is proportional to absolute temperature, is as expected. The experimented and simulated value of self-diffusion coefficient of water for all temperature is within error 10 %. Diffusion coefficient of water for water and oxytocin are found increasing with increase in temperature. Furthermore, the dependency of diffusion with temperature is analyzed with Arrhenius plot. Activation energy is also calculated from the slope of these plots. We have found that the activation energy of water obtained from simulated data is differing from experimented value with an error of 17.18 %. We do not have any experimental result for diffusion of oxytocin at present time to compare our final result; however, we can be sure on our accuracy with comparing the results of simulated and experimental self-diffusion coefficient of water.

Acknowledgements

K. Acharya acknowledges the partial support from University Grants Commission (UGC), Nepal. R.P.K. acknowledges the partial financial support from the Nepal Academy of Science and Technology (NAST).

References

- [1] P. Narayanan, Essentials of biophysics, New Age International Publishers, New Delhi (2010).
- [2] C. M. Venkatachalam, G. N. Ramachandran, Conformation of polypeptide chains, Annual Review of Biochemistry 38(1) (1969) 45-82.
- [3] C. Mura, C. E. McAnany, An introduction to biomolecular simulations and docking, Molecular Simulation 40 (10-11) (2014) 732-764.
<https://doi.org/10.1080/08927022.2014.935372>
- [4] W. H. Brown, T. Poon, Introduction to organic chemistry, John Wiley & Sons (2014).
- [5] R. Cotterill, Biophysics: an introduction, John Wiley & Sons (2003).
- [6] G. Ramachandran, V. Sasisekharan, Ado, Protein Chem, 23 (1968) 283.
- [7] S. B. Kent, Total chemical synthesis of proteins, Chemical Society Reviews 38(2) (2009) 338-351.
<https://doi.org/10.1039/B700141J>
- [8] V. J. Hruby, M. S. Chow, D.D. Smith, Conformational and structural considerations in oxytocin-receptor binding and biological activity, Annual Review of Pharmacology and Toxicology 30(1) (1990) 501-534.
- [9] C. A. Burtis, E. R. Ashwood, D. E. Bruns, Tietz textbook of clinical chemistry and molecular diagnostics-e-book, Elsevier Health Sciences (2012).
- [10] <https://pubchem.ncbi.nlm.nih.gov/compound/Oxytocin>
- [11] V. D. Vigneaud, C. Ressler, C. J. M. Swan, C. W. Roberts, P. G. Katsoyannis, S. Gordon, The synthesis of an octapeptide amide with the hormonal activity of oxytocin, Journal of the American Chemical Society 75(19) (1953) 4879-4880.
<https://doi.org/10.1021/ja01115a553>
- [12] S. C. Herpertz, K. Bertsch, Oxytocin effects on brain functioning in humans, Biological Psychiatry 79(8) (2016) 631-632.
<https://doi.org/10.1016/J.blopsych.2016.02.004>
- [13] F. T. Buisman-Pijlman, N. M. Sumracki, J. J. Gordon, P. R. Hull, C. S. Carter, M. Tops, Individual differences underlying susceptibility to addiction: role for the endogenous oxytocin system, Pharmacology Biochemistry and Behavior 119 (2014) 22-38.
<https://doi.org/10.1016/j.pbb.2013.09.005>
- [14] E. Sausville, D. Carney, J. Battey, The human vasopressin gene is linked to the oxytocin gene and

- is selectively expressed in a cultured lung cancer cell line, *Journal of Biological Chemistry* 260 (18) (1985) 10236-10241.
- [15] H. E. Ross, C. D. Cole, Y. Smith, I. D. Neumann, R. Landgraf, A. Z. Murphy, L. J. Young, Characterization of the oxytocin system regulating affiliative behavior in female prairie voles, *Neuroscience* 162(4) (2009) 892-903.
<https://doi.org/10.1016/j.neuroscience.2009.05.055>
- [16] R. P. Koirala, H.P. Bhusal, S. P. Khanal, N. P. Adhikari, Effect of temperature on transport properties of cysteine in water, *AIP Advances* 10(2) (2020) 025122.
<https://doi.org/10.1063/1.5132777>
- [17] S. Pokharel, N. Pantha, and N. Adhikari, Diffusion coefficients of nitric oxide in water: A molecular dynamics study, *International Journal of Modern Physics B* 30 (27) (2016) 1650205.
<https://doi.org/10.1142/S0217979216502052>
- [18] J. Crank, *The mathematics of diffusion* 2nd Edition. Oxford Science Publications (1975).
- [19] H. Hirakawa, Y. Kamei, M. Sugisaki, Y. Oishi, Relationship between Self-Diffusion and Interdiffusion in Gaseous Systems. *Bulletin of the Chemical Society of Japan* 46(9) (1973) 2659-2662.
<https://doi.org/10.1246/bcsj.46.2659>
- [20] D. Frenkel, B. Smit, *Understanding molecular simulation: From algorithms to applications*. *Computational Sciences Series* 1(2002) 1-638.
- [21] J. Lekner, Summation of Coulomb fields in computer-simulated disordered systems, *Physica A: Statistical Mechanics and its Applications* 176(3) (1991) 485-498.
[https://doi.org/10.1016/0378-4371\(91\)90226-3](https://doi.org/10.1016/0378-4371(91)90226-3)
- [22] L. S. Darken, Diffusion, mobility and their interrelation through free energy in binary metallic systems, *Trans. AIME* 175 (1948) 184-201.
- [23] H. Mehrer, *Diffusion in solids: fundamentals, methods, materials, diffusion-controlled processes*, Springer Science & Business Media (2007).
- [24] M. Allen, D. Tildesley, *Computer simulation of liquids*, New York (1987).
- [25] U. Dahal, N. P. Adhikari, "Molecular dynamics study of diffusion of heavy water in normal water at different temperatures," *Journal of Molecular Liquids* 167(2012) 34-39.
<https://doi.org/10.1016/j.molliq.2011.12.008>
- [26] H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, The missing term in effective pair potential, *Journal of Physical Chemistry* 91 (1987) 6269-6271.
<https://doi.org/10.1021/j100308a038>
- [27] D. Spoel et al., *GROMACS User Manual*, version 5.1.1, 2016.
- [28] H. J. Berendsen, J. V. Postma, W. F. van Gunsteren, A. R. H. J. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath, *The Journal of Chemical Physics* 81(8) (1984) 3684-3690.
<https://doi.org/10.1063/1.448118>
- [29] *Gromacs user manual*,
<http://manual.gromacs.org/documentation/2016.4/manual-2016.4.pdf>.
- [30] J. N. Evans, *Biomolecular NMR spectroscopy* (No. 54342 EVA), Oxford University (1995).
- [31] E. B. Walton, K. J. VanVliet, Equilibration of experimentally determined protein structures for molecular dynamics simulation, *Physical Review E* 74(6) (2006) 061901.
<https://doi.org/10.1103/PhysRevE.74.061901>
- [32] M. Tanaka, G. Girard, R. Davis, A. Peuto, N. Bignell, Recommended table for the density of water between 0 C and 40 C based on recent experimental reports, *Metrologia* 38 (4) (2001) 301.
<https://doi.org/10.1088/0026-1394/38/4/3>
- [33] M. Holz, S. R. Heil, and A. Sacco, Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate 1h nmr pfg measurements, *Physical Chemistry Chemical Physics* 2 (2000) 4740- 4742.
<https://doi.org/10.1039/B005319H>
- [34] P. Mark and L. Nilsson, Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K, *Journal of Physical Chemistry A* 105 (2001) 9954-9960.
<https://pubs.acs.org/doi/10.1021/jp003020w>

BIBECHANA

ISSN 2091-0762 (Print), 2382-5340 (Online)

Journal homepage: <http://nepjol.info/index.php/BIBECHANA>

Publisher: Department of Physics, Mahendra Morang A.M. Campus, TU, Biratnagar, Nepal

Molecular dynamics study of structural properties of γ -aminobutyric acid (GABA)

Shyam P. Khanal, Rajendra Prasad Koirala, Esha Mishra, Narayan P. Adhikari*

Central Department of Physics, Tribhuvan University, Kirtipur, Nepal

*Email: narayan.adhikari@cdp.tu.edu.np

Article Information:

Received: June 13, 2020

Accepted: June 25, 2020

Keywords:

Molecular dynamics

GABA

Radial distribution function

van der Waals radius

ABSTRACT

The study of structural conformation of Gamma-aminobutyric acid (GABA) exhibits its biological and chemical activities. The GABA molecule is responsible in neurotransmission from one neuron to another neuron and activates the ion channels to pass the chlorine and sodium ions in nerve cells. Its conformation in solid state and gas state are extremely different and it also shows five different conformations in aqueous solution. The study of its structure in such environment can reveal its activity in cellular environment. We have performed the classical molecular dynamics study of this system of GABA in aqueous medium to deal its structure. Radial distribution function (RDF) has been used to study the structural properties of the system.

DOI: <https://doi.org/10.3126/bibechana.v18i1.29442>

This work is licensed under the Creative Commons CC BY-NC License. <https://creativecommons.org/licenses/by-nc/4.0/>

1. Introduction

Gamma-aminobutyric acid (GABA), a chief inhibitory neurotransmitter, plays major role in reducing the neuronal excitability throughout the central nervous system [1-3]. It is mostly found in nervous system of highly developed brain of mammals. It is, in fact, a chemical messenger that transmits the signals across chemical synapses from one neuron to another neuron, gland cells or to the muscles [4]. It is also used to treat high blood pressure, stress and anxiety; and to stimulate the secretion of natural growth hormone of body. The disorder of GABA in the body may cause the

neurologic and psychiatric conditions. In addition, GABA is also detected in the other part from central nervous system like intestines, kidneys, uterus, ovaries, lungs etc. So, it has several functions in body mechanisms [5].

GABA is synthesized from anion of glutamic acid, called glutamate, via the enzyme glutamate decarboxylase with pyridoxal phosphate as a cofactor [6]. Its chemical formula is $C_4H_9NO_2$ and molar mass 103.120 g/mol. It is a white microcrystalline powder with density 1.11 g/mL. It is soluble in water with solubility value 130 g/100 mL. The temperature difference between solid state

and the gas state is quite narrow, with melting point 203.7°C and boiling point 247.9°C [7,8].

It contains a primary amine group and a carboxylic acid functional group, due to which it is categorized in amino acid group; however, the amino group (-NH₂) does not link to the alpha carbon as the ordinary amino acid contains. That is why, the GABA is not incorporated into the protein molecule. It is mostly found in zwitterion form [9, 10]. The interesting characteristic of GABA is its nature of conformation with its surroundings. It is mostly found highly folded configuration in gas phase and extended form in solid phase. Its character is surprising in solvent. It has high solvent effect with five different conformation [11]. The biochemical functions of biomolecules are greatly influenced by their structural conformation. Since the GABA can be at different conformations in various phases and also in the surrounding conditions, its structural study makes the great sense to understand the neurotransmission in nervous system, the spinal cord and hyperpolarization condition in ion exchange process in nerve cells [12]. Several researches have been carried out to deal the structural conformations of GABA in aqueous solution; however its structural variations in the living body are still unclear. Solvation of GABA molecule in water resembles real body.

Ashby *et al.* studied about different interactions with binding site that involved in GABA binding in molecular level; and Zafar and Jabeen studied the structure and function of GABA transporters (GATs) using computational method [13, 14]. Also, many experimental studies about transport properties of GABA have been already performed. Umecky *et al.* and Yui *et al.* measured the binary diffusion coefficient of GABA in infinitesimal aqueous medium at different temperature using Taylor dispersion method [15,16]. Also, viscosity of aqueous solution of GABA was estimated by Romero and Beltron [17].

Molecular dynamics (MD) study can be considered as an alternative technique to study about many properties including structural analysis [18]. MD also provides guideline for experimental study. To our best knowledge, the structural properties of GABA in water have not been studied using molecular dynamics. We expect that this study will help to learn the structural conformation of the molecule in water environment.

In this paper, we have discussed methods and methodology in section 2. The results of the work are presented and discussed in section 3; and finally conclusions and concluding remarks are presented in section 4.

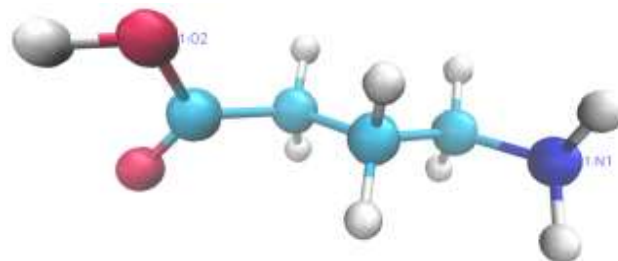


Fig. 1: Snapshot of GABA molecule.

2. Methods and Methodology

Modeling of system

We performed classical molecular dynamics simulations of the system of aqueous solution of GABA. GROMACS 5.1.1 software package was used for the simulations [19]. The OPLS-AA (optimized potentials for liquid simulations – all atom) force field for modeling the GABA molecule and three point SPC/E [20] water model were used during the simulations. In the classical molecular dynamics simulations, we solve the Newton's equation of motion [21]. To account the intra-molecular interactions, the bonded interaction i.e. bond stretching, bond angle and dihedral potentials are considered. And, the non-bonded Coulomb and Lennard-Jones (LJ) interactions contribute to the inter-molecular interactions. The Coulomb interaction arises due to partial charges of the

atoms in GABA and water molecules. For SPC/E water model, the partial charges of hydrogen and oxygen atoms are +0.4238e and -0.8476e respectively where e is elementary charge; and Lennard-Jones (LJ) parameters are 0.316 nm and 78.2k_B. Now, the inter-molecular interaction is caused due to non-bonded interactions which can be expressed as

$$V(r_{ij}) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_m r_{ij}}$$

where r_{ij} is the distance between i^{th} & j^{th} atoms of charges q_i & q_j respectively, ϵ & σ are LJ parameters and ϵ_m is permittivity of medium between the charge.

Computational details

We performed the simulations of the system of 3 γ -aminobutyric acid (GABA) as solute and 1035 water molecules as solvent at 1 atm pressure at five different temperature: 298.2 K, 303.2 K, 313.2 K, 323.2 K and 333.2. The simulations were carried out in cubic simulation box under periodic boundary conditions (PBC). At first, to remove van der Waals bad contact and to obtain the minimum potential energy state, energy minimization of the system was carried out using Steepest-descent

method taking 50 kJ/mole-nm force tolerance [22, 23].

Many properties of the system under study depend upon the parameters like temperature, pressure etc. So, the system must be in thermodynamics equilibrium. For this, the system was equilibrated at each temperature for 200 ns time taking time step of 0.002 ps using isothermal-isobaric (NPT) ensemble. During equilibration run, velocity rescaling thermostat with 0.01 ps coupling time was used to control temperature and Berendsen barostat with coupling time of 0.8 ps was used to keep constant pressure [24]. LINCS algorithm and Maxwell-Boltzmann distribution were taken to constraint all bonds and to assign initial velocities for each particle respectively [24]. Also, Particle Mesh Ewald (PME) method was chosen to account the long range Coulomb interaction; and cut off parameters of 1 nm was taken for both short range Coulomb & Lennard-Jones (LJ) interactions. In order to solve the equations of motion, Leap-frog algorithm was used [24]. Figure 2 represents the temperature and density profiles of the system after equilibration run at temperature of 303.2 K. Also, the simulated values of temperature and density are presented in the table 1.

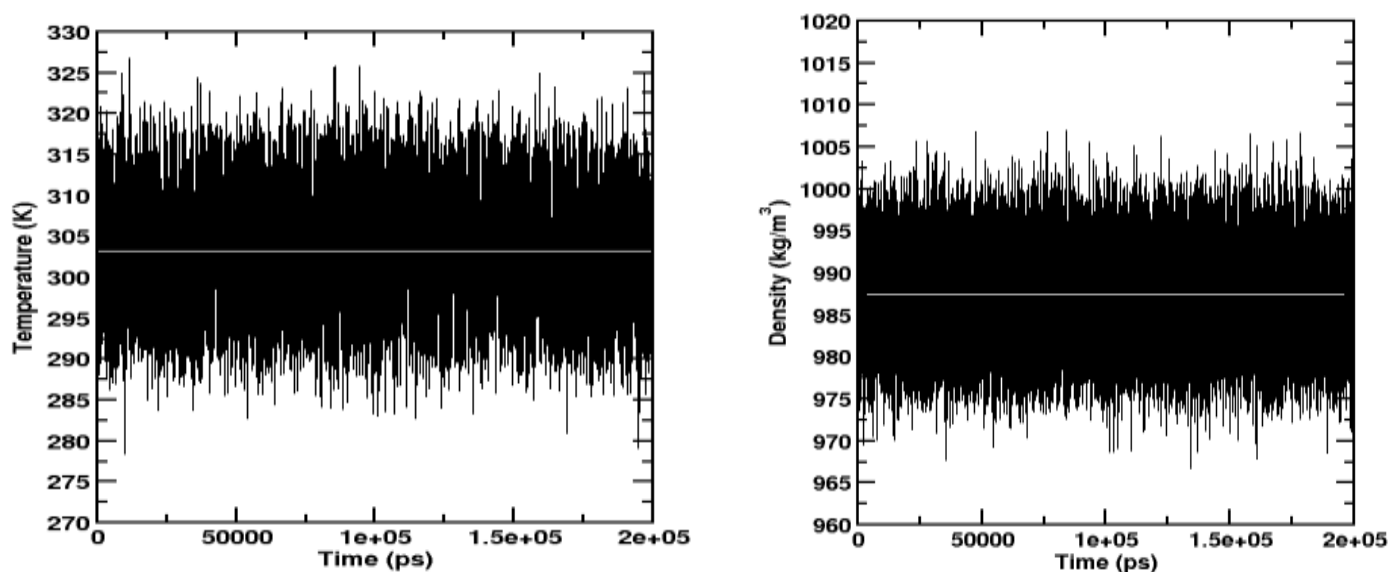


Fig. 2: Temperature (left) and density (right) profiles of the system after equilibration run at 303.2 K temperature.

Table 1: Simulated values of temperature and density at five different coupling temperature.

Coupling Temperature (K)	Simulated Temperature (K)	Simulated Density (kg/m ³)	Experimental Density (kg/m ³)[25]
298.2	298.20±0.01	990.07±0.02	997.03
303.2	303.19±0.00	987.49±0.02	995.63
313.2	313.19±0.04	981.46±0.03	992.19
323.2	323.19±0.01	974.45±0.02	988.19
333.2	333.19±0.02	967.19±0.02	983.17

From the table, it is seen that the simulated values of densities at different coupling temperature agree within 2% with previously reported experimental values.

After the equilibrating the system, the production run was done at each temperature using canonical (NVT) ensemble. During the production run, velocity rescaling thermostat with coupling time of 0.01 ps was used to control temperature. The velocities of final step of equilibration run were taken as initial velocities for production run. Each production was performed for 100 ns with time step 0.002 ps.

3. Results and Discussion

In this section, we present the RDF between different atoms of GABA and Water molecules.

Radial Distribution Function (RDF)

Structural properties of the system has been studied by using radial distribution function (RDF). RDF, that provides the idea of distribution of molecules around another molecule which is taken as reference, gives the probability of finding a pair of atoms located at distance 'r' [26]. For liquid, RDF shows an oscillation up to certain distance and becomes unity which means that there is no correlation between molecules after the distance [27, 28]. The GABA molecule contains the functional groups amide (-NH₂) and carboxyl (-COOH). Thus, to find the structural properties of

the system under study, we have calculated the RDF of oxygen of water & oxygen of water ($g_{OW-O}(r)$), oxygen of water & nitrogen of amide (-NH₂) group of GABA ($g_{OW-N1}(r)$), and oxygen of water & oxygen of carboxylic group (-COOH) of GABA ($g_{OW-O2}(r)$). Figures 3, 4 and 5 represent the $g_{OW-O}(r)$, $g_{OW-N1}(r)$ and $g_{OW-O2}(r)$ at five different temperature 298.2 K, 303.2 K, 313.2 K, 323.2 K and 333.2 K respectively.

In the RDF plots, there is a region from the reference atom up to which the value of RDF is zero. In this region, the probability of finding another atom is zero. Such region is known as excluded region (ER). Beyond the zero probability region, some peaks are observed and the value of RDF becomes unity beyond certain distance from reference atom. The unity value of RDF indicates that no correlation between atoms at that region i.e. correlation between atoms takes place up to certain distance from reference position [23]. Three peaks are observed between excluded and unity regions. The peaks indicate the favorable position of the atoms/molecules from reference. The first peak means the most favorable position of the atoms from reference. The values of excluded region (ER), first peak position (FPP), first peak value (FPV), second peak position (SPP), second peak value (SPV), third peak position (TPP) and third peak value (FPV) for $g_{OW-O}(r)$, $g_{OW-N1}(r)$ and $g_{OW-O2}(r)$ are presented in Tables 2, 3, and 4 respectively.

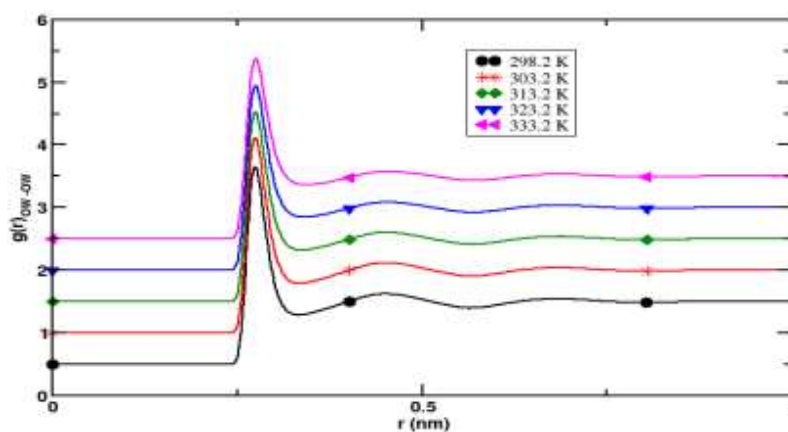


Fig. 3: RDF between oxygen atoms of water ($g_{ow-ow}(r)$) at different temperature.

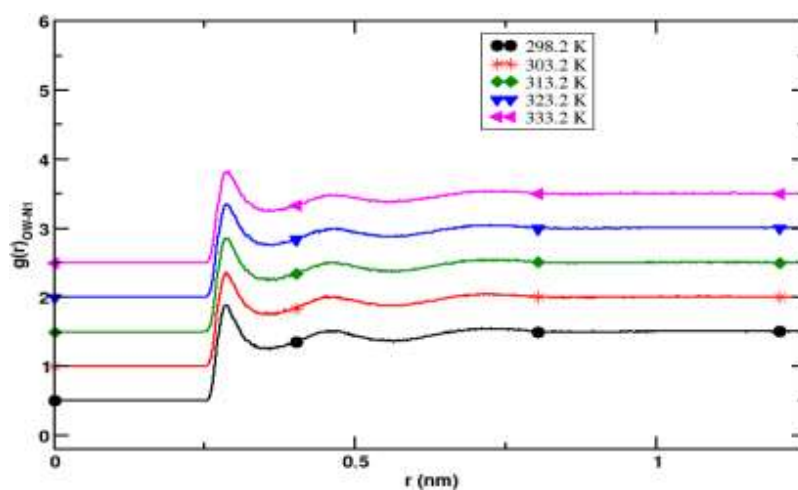


Fig. 4: RDF between nitrogen of GABA and oxygen of water ($g_{ow-N1}(r)$) at different temperature.

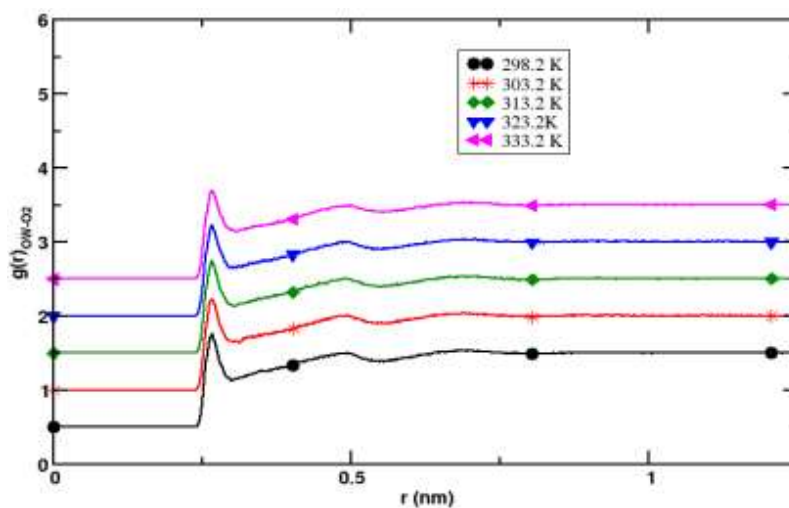


Fig. 5: RDF between oxygen of GABA and oxygen of water ($g_{ow-o2}(r)$) at different temperature.

Table 2: Simulated data for Radial distribution function (RDF) between water molecules ($g_{ow-ow}(r)$) at different temperature.

Temperature(K)	ER(nm)	FPP(nm)	FPV	SPP(nm)	SPV	TPP(nm)	TPV
298.2	0.240	0.274	3.141	0.450	1.126	0.686	1.046
303.2	0.240	0.276	3.104	0.450	1.120	0.690	1.040
313.2	0.240	0.276	3.018	0.450	1.103	0.688	1.037
323.2	0.240	0.276	2.936	0.450	1.087	0.694	1.036
333.2	0.240	0.276	2.874	0.450	1.075	0.686	1.034

Table 3: Simulated data for Radial distribution function (RDF) between nitrogen (N1) of GABA and water molecules ($g_{ow-N1}(r)$) at different temperature.

Temperature(K)	ER(nm)	FPP(nm)	FPV	SPP(nm)	SPV	TPP(nm)	TPV
298.2	0.248	0.284	1.379	0.466	1.016	0.726	1.047
303.2	0.248	0.286	1.352	0.452	1.017	0.712	1.047
313.2	0.246	0.284	1.359	0.470	1.004	0.730	1.048
323.2	0.248	0.286	1.344	0.468	0.997	0.724	1.045
333.2	0.248	0.290	1.321	0.458	0.985	0.726	1.040

Table 4: Simulated data for Radial distribution function (RDF) between oxygen (O2) of GABA and water molecules ($g_{ow-O2}(r)$) at different temperature.

Temperature(K)	ER(nm)	FPP(nm)	FPV	SPP(nm)	SPV	TPP(nm)	TPV
298.2	0.234	0.266	1.258	0.494	1.003	0.700	1.039
303.2	0.236	0.266	1.231	0.494	1.008	0.700	1.039
313.2	0.234	0.266	1.242	0.490	1.006	0.684	1.034
323.2	0.234	0.266	1.228	0.494	0.999	0.690	1.037
333.2	0.234	0.266	1.188	0.488	0.991	0.684	1.031

The values of Lennard-Jones parameter (σ) for OW-OW, OW-N1 and OW-O2 are 0.316 nm, 0.323 nm and 0.308 nm respectively. And, the calculated values of van der Waals radius ($2^{1/6}\sigma$) for OW-OW, OW-N1 and OW-O2 are 0.355 nm, 0.363 nm and 0.346 nm respectively. From the Tables 2-4, it is clearly observed that the values of first peak position (FPP) are smaller than the respective van der Waals radius, which indicates that other potentials along with Lennard-Jones (LJ) also contribute for stability of the system. On the other hand, Coulomb interactions arise in the system due to partial charges of hydrogen and oxygen of

SPC/E water model, and of N1 from -NH₂ & O2 from -COOH group of GABA molecule. Thus, both Coulomb as well as LJ including many body effects are responsible for structural properties of the system [19, 20]. From the Table, it is also observed that the values of excluded region (ER) are less than corresponding values of the van der Waals radius.

4. Conclusions and concluding remarks

We performed classical molecular dynamics of a system of 3 γ -aminobutyric acid (GABA) as solute and 1035 water as solvent at five different

temperature: 298.2 K, 303.2 K, 313.2 K, 323.2 K and 333.2 K using GROMACS 5.1.1 package. During the simulations, SPC/E water model and OPLS-AA force field parameters were used. Radial distribution function (RDF) was taken to analyze the structural properties of the system. For this, we plotted the RDF between different atoms at five different temperature i.e. (i) oxygen atoms of water molecules ($g_{OW-OW}(r)$), (ii) nitrogen of $-NH_2$ group of GABA and oxygen of water ($g_{OW-N1}(r)$) and (iii) oxygen of $-COOH$ group of GABA and oxygen of water ($g_{OW-O2}(r)$). From all the plots, we observe that both excluded region (ER) as well as first peak position (FPP) are smaller than the corresponding van der Waals radius. This indicates that along with Lennard Jones (LJ), Coulomb potential which arises due to the partial charges of different atoms as well as many body effects also contribute to the structural properties of the system.

In near future, we intend to study about the free energy calculation of the system.

Acknowledgements

SPK and RPK acknowledge the partial financial support from Nepal Academy of Science and Technology (NAST). EM acknowledges the master thesis grants support from University Grants Commission (UGC), Nepal. NPA acknowledges the UGC Award no. CRG-73/74-S&T-01.

References

- [1] H. Lodish et al. Neurotransmitters, synapses, and impulse transmission, In *Molecular Cell Biology*, 4th edition. WH Freeman, (2000).
- [2] L. Sivilotti & A. Nistri, = GABA receptor mechanisms in the central nervous system, *Progress in Neurobiology* 36(1) (1991) 35-92.
[https://doi.org/10.1016/0301-0082\(91\)90036-Z](https://doi.org/10.1016/0301-0082(91)90036-Z)
- [3] K. Łątka, J. Jończyk, and M. Bajda, Structure modeling of γ -aminobutyric acid transporters–Molecular basics of ligand selectivity, *Inter. J. Biol. Macromol* (2020).
<https://doi.org/10.1016/j.ijbiomac.2020.04.263>
- [4] T. A. Rocheleau, J. C. Steichen & A. E. Chalmers, A point mutation in a *Drosophila* GABA receptor confers insecticide resistance, *Nature* 363(6428) (1993) 449-451.
<https://doi.org/10.1038/363449a0>
- [5] D. Ma et al., Structure and mechanism of a glutamate–GABA antiporter, *Nature* 483 (7391) (2012) 632-636.
<https://doi.org/10.1038/nature10917>
- [6] P. Rorsman et al., Glucose-inhibition of glucagon secretion involves activation of GABA A-receptor chloride channels, *Nature* 341(6239) (1989) 233-236.
<https://doi.org/10.1038/341233a0>
- [7] A. J. Dobson & R. E. Gerkin, γ -Aminobutyric acid: a novel tetragonal phase, *Acta Crystallographica Section C: Crystal Structure Communications* 52(12) (1996) 3075-3078.
<https://doi.org/10.1107/S0108270196010001>
- [8] O. A. Petroff, Book review: GABA and glutamate in the human brain, *The Neuroscientist* 8(6) (2002) 562-573.
<https://doi.org/10.1177/1073858402238515>
- [9] A. Schousboe & H. S. Waagepetersen, GABA: homeostatic and pharmacological aspects, *Progress in Brain Research* 160 (2007) 9-19.
[https://doi.org/10.1016/S0079-6123\(06\)60002-2](https://doi.org/10.1016/S0079-6123(06)60002-2)
- [10] N. Ottosson, M. Pastorcak, S. T. van der Post, & H. J. Bakker, Conformation of the neurotransmitter γ -aminobutyric acid in liquid water, *Physical Chemistry Chemical Physics* 16(22) (2014) 10433-10437.
<https://doi.org/10.1039/C4CP00671B>
- [11] G. Deniau, et al., Synthesis, Conformation and Biological Evaluation of the Enantiomers of 3-Fluoro- γ -Aminobutyric Acid ((R)- and (S)-3F-GABA): An Analogue of the Neurotransmitter GABA, *Chem Bio Chem* 8(18) (2007) 2265-2274.
<https://doi.org/10.1002/cbic.200700371>
- [12] C. Lee, & A. J. de Silva, Interaction of Neuromuscular Blocking Effects of Neomycin and Polymyxin B, *Anesthesiology: The Journal of the American Society of Anesthesiologists* 50(3) (1979) 218-220.
<https://doi.org/10.1097/0000542-197903000-0001>
- [13] J. A. Ashby, et al., GABA binding to an insect GABA receptor: a molecular dynamics and mutagenesis study, *Biophysical Journal* 103(2012) 2071-2081.
<https://doi.org/10.1016/j.bpj.2012.10.016>
- [14] S. Zafar, and I.t Jabeen, Structure, function, and modulation of γ -aminobutyric acid transporter 1 (GAT1) in neurological disorders: a pharmacoinformatic prospective, *Frontiers in Chemistry* 6 (2018) 397.
<https://doi.org/10.3389/fchem.2018.00397>

- [15] T. Umecky, K. Ehara, S. Omori, T. Kuga, K. Yui, and T. Funazukur, Binary diffusion coefficients of aqueous phenylalanine, tyrosine isomers, and aminobutyric acids at infinitesimal concentration and temperatures from (293.2 to 333.2) K, *Journal of Chemical & Engineering Data* 58 (2013) 1909-1917.
<https://doi.org/10.1021/je3012698>
- [16] K. Yui, Y. Noda, M. Koido, M. Irie, I. Watanabe, T. Umecky and T. Funazukur, Binary Diffusion Coefficients of Aqueous Straight-Chain Amino Acids at Infinitesimal Concentration and Temperatures from (298.2 to 333.2) K, *Journal of Chemical & Engineering Data* 58 (2013) 2848-2853.
<https://doi.org/10.1021/je301370s>
- [17] C. M. Romero and A. Beltrán, Effect of temperature and concentration on the viscosity of aqueous solutions of 3-aminopropanoic acid, 4-aminobutanoic..., *Revista Colombiana de Química* 41 (2012) 123-131.
- [18] M. P. Allen and D. Tildesley, *Computer Simulation of Liquids* New York, USA: Oxford University Press, 1991.
- [19] D. van der Spoel, et al., GROMACS: fast, flexible, and free, *Inc. J Comput Chem* 26 (2005) 1701-1718.
<https://doi.org/10.1002/jcc.20291>
- [20] H. J. C. Berendsen J. R. Grigera, and T. P. Straatsma, The missing term in effective pair potentials, *Journal of Physical Chemistry* 91 (1987) 6269-6271.
- [21] M. E. Tuckerman, Mark, *Statistical mechanics: theory and molecular simulation*, Oxford university press, 2010.
- [22] S. P. Khanal, Y. P. Kandel, and N. P. Adhikari, Transport properties of zwitterion glycine, diglycine, and triglycine in water, *AIP Advances* 9 (2019) 065303.
<https://doi.org/10.1063/1.5099069>
- [23] R. P. Koirala, et al, Effect of temperature on transport properties of cysteine in water, *AIP Advances* 10 (2020) 025122.
<https://doi.org/10.1063/1.5132777>
- [24] M. J. Abraham, et al., *GROMACS User Manual Version 5.1. 1*, GROMACS Development Team (2016).
- [25] www.mrbigler.com/misc/Pvap-H2O.PDF
- [26] D. A. McQuarrie, *Statistical Mechanics*, Harper and Row, New York 1976.
- [27] S. Pokharel, N. Pantha, and N. P. Adhikari, Diffusion coefficients of nitric oxide in water: A molecular dynamics study, *International Journal of Modern Physics B* 30 (2016) 1650205.
<https://doi.org/10.1142/S0217979216502052>
- [28] K. Sharma and N. P. Adhikari, Temperature dependence of diffusion coefficient of nitrogen gas in water: A molecular dynamics study, *International Journal of Modern Physics B* 28 (2014) 1450084.
<https://doi.org/10.1142/S0217979214500842>

Certificate of Attendance

American Physical Society
March Meeting 2021
March 15–19, 2021 | Online

THIS IS TO CERTIFY THAT

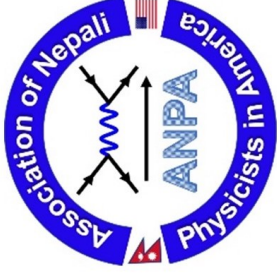
Rajendra Prasad Koirala

attended the American Physical Society's March Meeting.
Pre-meeting tutorials and short courses were held on March 13 and 14.



A handwritten signature in black ink, appearing to read "Don Wise".

Don Wise, Senior Meetings Registrar
March 19, 2021



Certificate of Appreciation

THIS CERTIFICATE IS PRESENTED TO

MR. RAJENDRA P. KOIRALA
Tribhuvan University, Nepal

FOR YOUR PRESENTATION DURING THE ANPA CONFERENCE 2020

DR. CHET R. BHATT
CHAIR, ANPA CONFERENCE 2020

DR. SHREE K. BHATTARAI
PRESIDENT, ANPA



NON-RESIDENT NEPALI ASSOCIATION (NRNA)

presents this

CERTIFICATE OF APPRECIATION

To MR RAJENDRA PRASAD KOIRALA

for presenting a paper at the **1ST NRN GLOBAL KNOWLEDGE CONVENTION**
organized by **NRNA** in partnership with the **GOVERNMENT OF NEPAL**

on 12 - 14 October 2018 in Kathmandu, Nepal

Bhaban Bhatta
President, NRNA

Kumar Panta
Vice-President, NRNA & Chair
1st NRN Global Knowledge Convention

Hem Raj Sharma, PhD
Advisor, NRNA & Chair
1st NRN Global Knowledge Convention



INTERNATIONAL CONFERENCE ON
EXPLORATIONS IN PHYSICS (ICEP-2018)



29-31 May, 2018, Kathmandu, Nepal

Rajendra Prasad Koirala

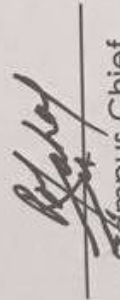
Central Department of Physics, Tribhuvan University, Kirtipur, Kathmandu, Nepal

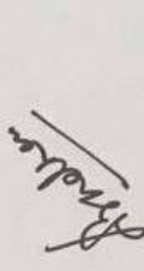
Contributed an oral presentation entitled

Comparative study of O6-Methyl Guanine-DNA and Normal-DNA in bonded and non-bonded interaction

during the conference


Chief Guest
Prof. Dr. Jiba Raj Pokharel
Vice Chancellor, NAST


Campus Chief
Rajesh Mahaju
Amrit Campus



SOC Chair
Assoc. Prof. Dr. Leela Pradhan Joshi
Amrit Campus



**International Conference on
Nanosciences and High Energy Physics
(ICNHEP-2019)
February 4-6, 2019
Central Department of Physics
Tribhuvan University, Kirtipur, Nepal**



Certificate

This is to Certify that **Dr./Mr./Ms. Rajendra Prasad Koirala**..... has actively participated in **International Conference in Nanosciences and High Energy Physics "ICNHEP-2019"** organized by Central Department of Physics, Tribhuvan University, Kirtipur Kathmandu, Nepal on 4-6 February, 2019.

[Signature]

Prof. Dr. Ram Pd. Khatriwada
Dean, IoST, TU, Kirtipur

[Signature]

Dr. Gopi Chandra Kaphle
Convener, ICNHEP-2019

[Signature]

Prof. Dr. Binil Aryal
Head CDP, TU, Kirtipur



Lecture Series on

Research Methodology

6 November – 22 December 2017

Central Department of Physics
Tribhuvan University, Kirtipur, Nepal



Participation Certificate

.....
Rajendra Prasad Koirala.....

.....
Central Department of physics T.U.....

participated in **21 hours lecture series** on

Research Methodology delivered by **Prof. Dr. Subodh R. Shenoy**,
TIFR, India during 6 November to 22 December 2017.

.....
Subodh R. Shenoy.....

Prof. Dr. Subodh R. Shenoy

Guest Speaker

Tata Institute of Fundamental Research, India

.....
Binil Aryal.....

Prof. Dr. Binil Aryal

Head

CDP, TU Kirtipur



Nepal Physical Society

Ghantaghar, Kathmandu

Certificate of Participation

This certificate is awarded to

Mr. Rajendra Prasad Koirala from

**Central Department of Physics, TU, Nepal, for his presentation entitled
Binding of AGT with DNA during Pre- and Pos-methyl Transfer in
DNA-Protein complex**

On Scientific Session of 37th Annual Convention, February 6, 2021.

Prof. Dr. Narayan Prasad Chapagain
President
Nepal Physical Society

Dr. Sunil Babu Shrestha
Chief Guest
Vice – Chancellor, NAST

Prof. Dr. Leela Pradhan Joshi
Chair, Scientific Committee
37th Annual Convention, NPS

February 6, 2021