

**MOLECULAR MODELING AND DESIGN
TOWARDS UNDERSTANDING MICROSCOPIC
INSIGHTS OF
BIOMOLECULAR STRUCTURE AND FUNCTION**

THESIS SUBMITTED TO



**MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY,
WEST BENGAL
(Formerly WEST BENGAL UNIVERSITY OF TECHNOLOGY)**

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BIOINFORMATICS

PIYA PATRA

(Registration no: PhD/Tech/BI003/2016)

**MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY,
WEST BENGAL
(Formerly WEST BENGAL UNIVERSITY OF TECHNOLOGY)**

Main Campus: NH 12, Haringhata, District- Nadia

Post Office - Simhat, Pin - 741249, West Bengal

City Campus: BF -142, Sector - I, Salt Lake, Kolkata - 700 064

BI003/2016

Maulana Abul Kalam Azad University of Technology, West Bengal
(Formerly known as West Bengal University of Technology)

MAULANA ABUL KALAM AZAD
UNIVERSITY OF TECHNOLOGY,
WEST BENGAL



REGISTRATION CERTIFICATE

To:

PIYA PATRA

GOURANGO DANGA COLONY,
RAMBAGAN, P.O.- SEARSOLE RAJBARI,
DIST. – BURDWAN, RANIGANJ,
PIN – 713358.

Dear Sir/Madam,

I am glad to inform you that you have been granted registration for Ph. D programme under this University in **BIOINFORMATICS** for the Degree of Doctor of Philosophy (Ph. D.).

Your Registration No. **PhD/Tech/BI003/2016** is with effect from **13/01/2016**.

This registration shall remain valid for five years with effect from the date of registration as indicated above.

Title of the Thesis

**MOLECULAR MODELING AND DESIGN TOWARDS
UNDERSTANDING MICROSCOPIC INSIGHTS OF
BIOMOLECULAR STRUCTURE AND FUNCTION**

Dr. Raja Banerjee
(Supervisor)

Prof. Jaydeb Chakrabarti
(Joint Supervisor)

(Associate Supervisor)

Yours faithfully,

January 21, 2017

Registrar

Maulana Abul Kalam Azad University of Technology, West Bengal
BF -142, Sector 1, Salt Lake, Kolkata - 700064

This Registration Certificate will be deemed cancelled if any case of ineligibility is detected.



MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL

(Formerly known as WEST BENGAL UNIVERSITY OF TECHNOLOGY)

BF - 142, Sector - 1, Salt Lake, Kolkata - 700 064, INDIA

Website : www.wbut.ac.in

Prof. Raja Banerjee

Professor, Department of Biotechnology
Co-ordinator, BIF-DBT at WBUT

Ref. No.:

Date :

Certificate from the Supervisor

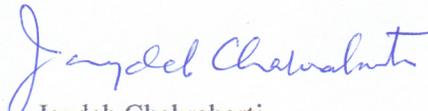
This is to certify that the thesis entitled “**MOLECULAR MODELING AND DESIGN TOWARDS UNDERSTANDING MICROSCOPIC INSIGHTS OF BIOMOLECULAR STRUCTURE AND FUNCTION**” submitted by Ms. Piya Patra (Registration no. PhD/Tech/BI003/2016), who got her name registered on 13/01/2016 for the award of Ph.D (Technology) degree in Bioinformatics of the Maulana Abul Kalam Azad University of Technology, West Bengal (formerly known as West Bengal University of Technology) is absolutely based on the original research work done by her, under the joint supervision of Prof. Jaydeb Chakrabarti, S. N. Bose National Centre for Basic Sciences, Kolkata and myself. She has pursued the work at Department of Bioinformatics, Maulana Abul Kalam Azad University of Technology, West Bengal and Department of Chemical, Biological and Macro-Molecular Science, S. N. Bose National Centre for Basic Sciences, Kolkata. This work has not been submitted previously for the award of any degree or diploma at any other Institute or University.

Raja Banerjee

सत्येन्द्र नाथ बसु राष्ट्रीय मौलिक विज्ञान केन्द्र
SATYENDRA NATH BOSE NATIONAL
CENTRE FOR BASIC SCIENCES
সত্যেন্দ্র নাথ বসু জাতীয় মৌল বিজ্ঞান কেন্দ্র

Certificate from the Joint Supervisor

This is to certify that the thesis entitled “**MOLECULAR MODELING AND DESIGN TOWARDS UNDERSTANDING MICROSCOPIC INSIGHTS OF BIOMOLECULAR STRUCTURE AND FUNCTION**” submitted by Ms. Piya Patra (Registration no. PhD/Tech/BI003/2016), who got her name registered on 13/01/2016 for the award of Ph.D (Technology) degree in Bioinformatics of the Maulana Abul Kalam Azad University of Technology, West Bengal (formerly known as West Bengal University of Technology) is absolutely based on the original research work done by her, under the joint supervision of Prof. Raja Banerjee, Maulana Abul Kalam Azad University of Technology, West Bengal and myself. She has pursued the work at Department of Bioinformatics, Maulana Abul Kalam Azad University of Technology, West Bengal and Department of Chemical, Biological and Macro-Molecular Science, S. N. Bose National Centre for Basic Sciences, Kolkata. This work has not been submitted previously for the award of any degree or diploma at any other Institute or University.


Jaydeb Chakrabarti

Prof. Jaydeb Chakrabarti
Senior Professor

Department of Chemical, Biological & Macro-Molecular Sciences

ब्लॉक - जे.डी., सेक्टर - III, सोल्ट लेक, कोलकाता - 700 106 / Block - JD, Sector - III, Salt Lake, Kolkata - 700 106

दूरभाष / Phones : (00) 91 - (0) 33 - 2335 5706-8, 2335 3057 / 61, 2335 0312 / 1313

टेलीफैक्स / TELEFAX : +91 -33-2335 3477 / 2335 1364 / 2335 9176

वेबसाइट / Website: <http://www.bose.res.in>

भारत सरकार के विज्ञान एवं प्रौद्योगिकी विभाग के अंतर्गत एक स्वायत्त संस्थान

AN AUTONOMOUS INSTITUTE UNDER DEPARTMENT OF SCIENCE & TECHNOLOGY, GOVERNMENT OF INDIA



*Dedicated to Maa, Baba
Buni (Riya) & Manas*

*Who motivate, inspired and guided me
always to take challenge and learn
something new...*

Acknowledgements

Acknowledgements:

At this memorable moment, I would like to express my gratitude to my Ph.D. supervisors **Prof. Raja Banerjee** of Maulana Abul Kalam Azad University of Technology, WB and **Prof. Jaydeb Chakrabarti** of S. N. Bose National Centre for Basic Sciences, Kolkata for their continuous support and guidance throughout my research work. They not only gave me valuable suggestions and the chance to work independently, but also worked alongside me with the same earnestness. I am extremely thankful to **Prof. Jaydeb Chakrabarti** for sharing his knowledge with me. I am also thankful to **Dr. Mahua Ghosh** of S. N. Bose National Center for Basic Sciences, Kolkata for her support and guidance in parallel.

I am thankful to the honorable vice-chancellor, **Prof. Saikat Maitra**, the registrar, **Dr. Partha Pratim Lahiri** of Maulana Abul Kalam Azad University of Technology, WB, and Ph.D. convener **Dr. Ranjita Biswas (MAKAUT, WB)** for providing me with the amenities essential for my work in MAKAUT, WB. I am grateful to all my teachers during my M.Tech, Ph.D. coursework and Ph.D. committee members, especially **Prof. Indranil Mukherjee (MAKAUT, WB)**, **Dr. Jaya Bandyopadhyay (MAKAUT, WB)**, **Prof. Ansuman Lahiri (CU)**, **Dr. Soumalee Basu (CU)**, **Prof. Dhananjay Bhattacharyya (SINP)**, **Prof. Kamallesh Bhowmik (SINP)**, **Prof. Debnath Pal (IISc Bangalore)** and **Prof. Rabi Majumder (SINP)** for support and friendly assistance since my initial days in research.

I have started my Ph.D. work in Bioinformatics Department (DBT-BIF), Maulana Abul Kalam Azad University of Technology, WB and CBMS Department, S. N. Bose National Centre for Basic Sciences, Kolkata. I am sincerely grateful to **Prof. Samit Kumar Ray**, the director, **Prof. Biswajit Chakraborty**, the dean (Academic Programme) of S. N. Bose National Centre for Basic Sciences, Kolkata, who have allowed me to avail all the computational facilities for working freely in S. N. Bose Centre. Without these facilities, performing the long MD simulation runs would have been a tough assignment.

I formally acknowledge **DBT-BIF, TEQIP-II, TEQIP-III and ICMR, Govt. of India** for providing me the fellowship in my Ph.D. tenure at different time and other support of my research.

Acknowledgements

I am thankful to my husband, **Dr. Manas Mondal**, Research Associate Professor, Institute of Systems Biology, Shenzhen Bay Laboratory, China for his constant support, guidance and the valuable suggestions. His unconditional love, friendship, inspiration, motivation and mental support help me a lot to overcome all the difficulties during my Ph.D. tenure.

It is a pleasure for me to remember all my friends and colleagues during different phases of my research tenure. I am particularly obliged to my seniors **Dr. Tridip Sheet, Soumya Kanti Ghosh, Dr. Nikhat Saba, Dr. Divanshu Shukla and Samar Naskar** of MAKAUT, WB, **Dr. Mausumi Ray, Dr. Sutapa Dutta, Dr. Sankar Das, Dr. Paramita Saha, Dr. Samapan Sikdar, Dr. Sayan Bayan, Dr. Buddhadeb Pal and Dr. Suman Dutta** of S. N. Bose Centre, **Dr. Debasree Sarkar, Dr. Arijita Sarkar, Dr. Prerna Priya, Tanmoy Jana** of Bose Institute, Kolkata and **Dr. Rahul Majumder, Dr. Soumyabrata Banerjee** of Calcutta University for their valued guidance and assistance. I am indebted to my beloved juniors – **Riju Pal, Sasthi Charan Mandal, Shubhadip Moulick, Biswajit Pabi, Shubhrasish Mukherjee, Rafiqul Alam, Didhiti Bhattacharya, Taniya Basu, Saikat Pal, Abhik Ghosh Moulick, Rahul Karmakar, Edwin Tendong, Raghu, Dipika Mandal, Subrata Ghosh, Dipanjan Maity, Darshan, Agneesh, Asith, Tausif, Arunava** at S. N. Bose National Centre for Basic Sciences, Kolkata and **Shubhankar Sahu, Subhrendu Ghosh, Indra Kundu, Abhirupa Ghosh, Srijan, Tulika, Bipasha, Sourav Hati** at MAKAUT, WB – for always helping me out of critical situations with their supports and suggestions, both in personal and professional life. The atmosphere within our lab has always been a pleasant experience to bear with the difficulties of research life. I will always treasure the companionship of my B.Tech, M.Tech and Ph.D. course work batchmates – **Suvankar Ghosh, Rinki Banik, Kritika Sharma, Dr. Sumit Mukherjee, Samrat Ghosh, Swagata Roy Chowdhury, Deep Shubhra, Asmita Roy, Sadip Midya, Ishita Bhakta**. Companionship, chatting and interactions with seniors, juniors and friends in the different institutes during my Ph.D. tenure will be a memorable experience for me.

Bioinformatics Department, Ph.D. section, Finance section of MAKAUT, WB and CBMS Department, academic section and computer section of S. N. Bose National Centre for Basic Sciences, Kolkata has been a great place to flourish my work, thanks to the friendly milieu maintained by its scientific members specially **Dr. Atri Bhowmik (Finance Officer,**

Acknowledgements

MAKAUT, WB), Dr. Soumen Kr. Pati (Bioinformatics Department, MAKAUT, WB), Dr. Smarajit Das (HOD, Bioinformatics Department, MAKAUT, WB). I also wish to thank non-scientific members of MAKAUT, WB and S. N. Bose Centre specially **Tanima, Pramod, Suman Kar, Biswajit Das, Shreya Sen, Soumya Bhattacharya, Buddhadeb Da, Mahato Da** of MAKAUT, WB and **canteen staffs** of S. N. Bose Centre.

I would also like to thanks my school, college teachers - **Jonaki Basu, Malancha Pal, Aparajita madam, Rabin Gandhi, Bipod Taran Ghosh, Sudhangshu sir, Tapati Chattapadhyay, Sanghamitra madam, Dr. Sanchita Bandyopadhyay, Prof. Debnath Pal, Dr. Subhrangshu Supakar, Dr. Rahul Banerjee, Dr. Koushik Pal** for their guidance and for motivating me to undertake a career in research. Along with them I also want to acknowledge **Prof. Mrinal Kanti Poddar (CU), Dr. Sudipto Saha (Bose Institute)** for their encouragement during my B.Tech and M.Tech project.

I also wish to thank my childhood friends, **Umesh, Pinky, Shrestha, Kka, Nandita, Dipanjali, Siddhartha, Anupama** for always being there for me available. I also wish to thank my friends **Sampurna, Kakali, Roop, Pallab, Samrat, Debjani Di, Rose, Aritra Da, Soumili Di, Lipika Sheet, Preranna, Sibun Da, Padmaja Di, Mahes Mondal.** I must thank them for making my experiences really memorable.

I sincerely thank my family for always being there with me through all successes and failures. I wish to thank **my parents** and sister **Riya** for their constant and unconditional love, support and comfort. Also, there are many others, who by their co-operation and support have made this success achievable. Lastly, I express my academic indebtedness to all the authors whose works I have been cited in this thesis.

Finally, I thank the almighty God for being with me all the way.

Date:

Piya Patra

Department of Bioinformatics
Maulana Abul Kalam Azad University of Technology, WB
Haringhata, Nadia, Pin- 741249, West Bengal, India

Abstract

Proteins and peptides play integral role in regulating key biological and biochemical processes within a living cell. Quite often function of a protein depends on the binding of small molecules or ligands to the active site of the protein. Microscopic understanding of ligand-protein interactions is important not only to understand the cellular process and but also to design proteins/peptides for drug delivery and other targeted applications. However, how ligands recognize the particular binding site or active site of the protein molecules and trigger the biological functions is not understood well. Experimental tools to probe biomolecular structure and function at microscopic level are limited, which makes the *in-silico* studies important. In this thesis I focus on the interactions of charged ligands, like anions with peptides of different functional proteins and cell penetrating cationic peptides. All-atom molecular dynamics simulations and quantum chemical calculations show that sequence dependent conformational preference of motif residues is crucial for anion recognition by anion binding C^αNN motif in functional proteins. The studies further reveal that there is an intimate connection between coordination and conformational preferences of the motif residue. I also study condensation of counter-anion around cationic Tat peptide using all-atom molecular dynamics simulations. It is found that anion condensation influences the peptide conformation and absorption properties of the Tat peptide on the lipid bilayer surface. The condensation is maximum at an optimum anionic concentration in solution. Simplified statistical mechanics-based model study suggests that the maximum condensation at an optimum solvent condition is generically due to competition between attraction by the positive charged groups of the polypeptide and repulsion between the anions. Tat peptide in presence of anion is also found induced stress on the lipid bilayer surface that can facilitate the translocation of Tat peptide through bilayer. These studies will provide further insight to understand the cell penetrating and antimicrobial activity of Tat peptide in presence of anion, which is helpful for drug delivery and other biomedical applications.

Abbreviations

List of Abbreviations:

DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
ATP	Adenosine triphosphate
GTP	Guanosine triphosphate
PLP	Proteolipid protein
NADP	Nicotinamide adenine dinucleotide phosphate
FAD	Flavin adenine dinucleotide
CoA	Coenzyme A
NAD	Nicotinamide adenine dinucleotide
FMN	Flavin mononucleotide
AMP	Antimicrobial peptide
fs	Femtosecond
ps	Picosecond
ns	Nanosecond
MD	Molecular Dynamics
MC	Monte Carlo
PDB	Protein Data Bank
VMD	Visual Molecular Dynamics
FSSP	Fold Classification Based on Structure-Structure Alignment of Protein
A	Alanine
C	Cysteine
D	Aspartic acid
E	Glutamic acid
F	Phenylalanine
G	Glycine

Abbreviations

H	Histidine
I	Isoleucine
K	Lysine
L	Leucine
M	Methionine
N	Asparagine
P	Proline
Q	Glutamine
R	Arginine
S	Serine
T	Threonine
V	Valine
W	Tryptophan
Y	Tyrosine
GROMACS	GRoningen Machine for Chemical Simulations
AMBER	Assisted Model Building with Energy Refinement
VASP	Vienna ab-initio Simulation Package

Table of Contents:

CHAPTER	PAGE NUMBER
CHAPTER 1	1 - 25
1. Introduction:	
<i>Anion binding motif in protein</i>	3 – 8
<i>Influence of anion on cationic antimicrobial peptide</i>	9 – 15
<i>References</i>	16 – 25
CHAPTER 2	26 - 87
2. Structural basis of anion specificity in C^αNN motif at the functional interface of proteins	
2.1 Introduction	27 – 30
2.2 Material & Methods	30 – 33
2.2.1 Systems for study	30 – 31
2.2.2 Molecular Dynamics(MD) Simulation	31 – 32
2.2.3 Trajectory analysis	32 – 33
2.3 Results & Discussion	34 – 57
2.3.1 Anion induced conformational preferences of the motif residues	37 – 43
2.3.2 Conformational preferences of flanking residues	43 – 46
2.3.3 Motion of the anion around the motif	46 – 48
2.3.4 Solvent structure around the motif	49 – 51
2.3.5 Basis of conformational preferences of C ^α NN motif	51 – 57
2.4 Conclusion	58 – 58
Reference	59 – 64
Appendix	65 – 87
A2.1 Dihedral angle and secondary structure assignment of protein	65 – 67
A2.2 Molecular Dynamics Simulation	68 – 83
A2.2.1 Molecular Mechanics Force Fields	68 – 69
A2.2.2 Energy Minimization	69 – 73
A2.2.3 Equation of motion	73 – 77
A2.2.4 Periodic Box and Minimum Image Convention	77 – 78
A2.2.5 Thermodynamic Ensembles & Equilibration	78 – 83
A2.3 Conformational Thermodynamics	84 – 87

CHAPTER 3	88 - 131
3. Quantum chemical studies on anion binding and specificity to C^αNN motif	
3.1 Introduction	89 – 90
3.2 Material & Methods	90 – 92
3.2.1 Systems for study	90 – 91
3.2.2 Quantum chemical calculation	91 – 91
3.2.3 Theoretical analysis	92 – 92
3.3 Results	93 – 113
3.3.1 Optimized geometry of motif fragments	93 – 95
3.3.2 Coordination of anion	95 – 104
3.3.3 Stabilization energy	104 – 105
3.3.4 Replacement of anion	105 – 109
3.3.5 Benchmark calculations	109 – 113
3.4 Discussion	114 – 117
3.5 Conclusions	117 – 118
References	119 – 122
Appendix	123 – 131
A3.1 Density Functional Theory for electronic structure calculations	123 – 127
A3.2 Quantum chemical calculation using VASP	128 – 128
A3.3 Quantum chemical calculation using GAUSSIAN	129 – 131
 CHAPTER 4	 132 - 162
4. Anion scavenging by cationic antimicrobial peptide	
4.1 Introduction	133 – 135
4.2 Material & Methods	135 – 137
4.2.1 Molecular Dynamics (MD) simulation	135 – 136
4.2.2 Theoretical analysis	136 – 137
4.3 Results of all-atom simulations	137 – 147
Structural properties of peptide	137 – 140
Anion condensation	140 – 147
4.4 Model Calculation	147 – 150
4.5 Discussion	150 – 151
4.6 Conclusion	151 – 152
References	153 – 159
Appendix	160 – 162
A4.1 Metropolis algorithm & Monte Carlo simulation	160 – 161
A4.2 Calculation of η_s	162 – 162

Table of Contents

CHAPTER 5	163 - 193
<i>5. Interaction of cationic polypeptide with lipid bilayer</i>	
<i>5.1 Introduction</i>	164 – 166
<i>5.2 Material & Methods</i>	166 – 170
<i>5.2.1 Modeling of peptide-lipid system</i>	<i>166 – 167</i>
<i>5.2.2 Molecular Dynamics (MD) simulation</i>	<i>167 – 168</i>
<i>5.2.3 Theoretical analysis</i>	<i>168 – 170</i>
<i>5.3 Results</i>	170 – 177
<i>Lipid bilayer structure</i>	<i>170 – 173</i>
<i>Interaction of peptide with lipid bilayer</i>	<i>173 – 177</i>
<i>5.4 Discussion</i>	178 – 184
<i>5.5 Conclusion</i>	185 – 185
<i>References</i>	186 – 191
<i>Appendix</i>	192 – 193
<i>A5 Molecular Dynamics simulation setup for peptide-lipid system</i>	
CHAPTER 6	194 - 197
<i>Conclusions & future perspective</i>	195 – 197
PUBLICATIONS	198

LIST OF PUBLICATIONS

Publications arising from this thesis

- **Patra P**, Ghosh M, Banerjee R, Chakrabarti J; “Quantum chemical studies on anion specificity of C^αNN motif in functional proteins”, **J Comput. Aided Mol. Des. 2018; 32(9):929-936.**
- **Patra P**, Ghosh M, Banerjee R, Chakrabarti J; “Anion induced conformational preference of C^αNN motif residues in functional proteins”, **Proteins 2017; 85(12):2179-2190.**
- **Patra P**, Banerjee R, Chakrabarti J; “Control of solvent exposure of cationic polypeptides in anionic environment” (Under review)
- **Patra P**, Banerjee R, Chakrabarti J; “Membrane properties in presence of HIV1-TAT polypeptide” (Communicated)

Other publications apart from this thesis

- Dutta S, **Patra P**, Chakrabarti J; “Self-assembly in amphiphilic macromolecules with solvent exposed hydrophobic moieties, **Biopolymers 2019; DOI: 10.1002/bip.23330.**

PRESENTATIONS IN CONFERENCES/WORKSHOPS

- **Poster presentation:** “*Microscopic studies on anion condensation around arginine rich cationic antimicrobial peptide*”, **Piya Patra**, Raja Banerjee, Jaydeb Chakrabarti; **STATPHYS- KOLKATA X**, Presidency University, Kolkata, 26th - 29th November, 2019
- **Poster presentation:** “*Molecular dynamics study of anion scavenging by cationic antimicrobial peptide*”, **Piya Patra**, Raja Banerjee, Jaydeb Chakrabarti; **Indian Biophysical Society Meeting (IBS 2019)**, IISER Kolkata, 15th - 17th march, 2019
- **Poster presentation:** “*In silico study of C^αNN motif in functional proteins towards recognition of anion*”, **Piya Patra**, Mahua Ghosh, Raja Banerjee, Jaydeb Chakrabarti; **INDO-US conference on Multiscale Simulation & Mathematical Modelling of Complex Biological Systems (MSMM 2019)**, Jawaharlal Nehru University, New Delhi, 30th January - 1st February, 2019.
- **Poster presentation:** “*Anion recognition and specificity of C^αNN motif in functional proteins*”, **Piya Patra**, Mahua Ghosh, Raja Banerjee, Jaydeb Chakrabarti; **International Conference on Complex and Functional Materials (ICCFM 2018)**, Biswa Bangla Convention Centre, Kolkata, 13th - 16th December, 2018.
- **Poster presentation:** “*Anion induced secondary structural facet of anion recognition C^αNN motif*”, **Piya Patra**, Mahua Ghosh, Raja Banerjee, Jaydeb Chakrabarti; **Soft matter and chemical physics in Kolkata region (SMCPK) meeting**, S. N. Bose National Centre for Basic Sciences, Kolkata, 15th September, 2018.
- **Poster presentation:** “*Sulphate induced secondary structural facet of anion recognition C^αNN motif*”, **Piya Patra**, Mahua Ghosh, Raja Banerjee, Jaydeb Chakrabarti; **'6th Indian Peptide Symposium'**, February, 2017, Homi Bhabha Centre for Science Education (HBCSE), Mumbai, India.



CHAPTER 1

INTRODUCTION



Chapter 1

Biomolecular structure and functions are integrally connected. Among the biomolecules, proteins play integral role in regulating key biological and biochemical processes within a living cell. At microscopic level, often function of protein depends on the interaction of ligands with it. Such interactions are important in DNA–protein complexes¹, active transport of ions and other ligands², structural equilibrium of deoxyhemoglobin³, competitive inhibition of triosephosphate isomerase⁴ and phosphoglycerate kinase⁵ and many more⁶⁻¹⁰. How ligands recognize the particular binding site or active site of the protein molecules and trigger the biological functions is not understood well.

Microscopic understanding of ligand-protein interactions is important to design proteins/peptides for drug delivery and other targeted applications. Different ligand recognition motifs are found in the functionally important regions of protein structures. The control of the residue conformational preference is therefore important to trigger the biological functions and enzymatic activity of the protein. In some cases, specific protein motif also participates in the formation of the functional interface important for binding another protein. Understanding the structural and functional significance of such motif will help in protein engineering and modulating the protein-protein interaction related to the biological functions. The central to the potential pharmacologic application of antimicrobial peptides is the degree to which they differentiate, or may be engineered to differentiate between microbial targets and normal host cells¹¹⁻¹³. Designing biomolecules with targeted function is a major endeavor in current day research with huge potential therapeutic applications¹⁴⁻¹⁹. Microscopic understanding of the ligand binding and cell penetrating activity of antimicrobial peptide will help to design potential antimicrobial peptides for therapeutical targets, drug delivery and other biomedical applications. Experimental tools to probe biomolecular structure and function at

microscopic level are limited. This leads one to use *in-silico* model to explore structure and function connections in variety of contexts. Molecular modeling and *in-silico* study of the peptide-ligand interactions will facilitate new opportunities and approaches to drug design and develop pharmacologic agents.

In this thesis I focus on interactions of charged ligands, like anions with peptides of different functional proteins. I consider a couple of different cases to this end: (1) specific interactions of anions like sulphate and biphosphate with polypeptides having anion binding C^oNN motif, and (2) influence of biphosphate on guanidium-rich human immunodeficiency virus type 1 (HIV-I) Tat polypeptide on antimicrobial activity. My primary objective is to provide *in-silico* insights to design polypeptides for specific applications.

Anion binding motif in protein:

Three dimensional structures of proteins tend to be dominated by the classical structures of alpha helix (α -helix) and beta sheet (β -sheet). However, many other functional or hydrogen-bonded structural elements, termed as motifs, occur which form functional interfaces of protein. Protein motifs can be defined by their primary sequence or by the arrangement of secondary structure elements²⁰⁻²¹. In zinc finger motif which is found in a widely varying family of DNA-binding proteins, the conserved cysteine and histidine residues coordinate with a zinc ion and this coordination is essential to stabilize the tertiary structure. Sequence motifs can often be recognized by simple inspection of the amino acid sequence of a protein, and when detected provide strong evidence for biochemical function. The second, equally common, use of the term motif refers to a set of contiguous secondary structure elements that either have a particular functional significance or define a portion of an independently folded domain²¹.

Chapter 1

An example of such motif is the helix-turn-helix motif which is found in many DNA-binding proteins. Another motif called the EF hand motif is specific for calcium binding and is present in parvalbumin, calmodulin, troponin-C and thereby regulate cellular activities. The simplest motif involving β -strands, is adjacent anti-parallel strands joined by a loop, called the hairpin β -motif. These occur quite frequently in protein structure and are present in most anti-parallel β structures both as an isolated ribbon and as part of the more complex β -sheet.

In last few years extensive structural studies based on the crystallography and NMR show different kinds of ligand-recognition motifs in proteins²²⁻³¹. In order to trigger the anion–protein interaction, usually proteins use short peptide-based motif(s) to recognize the anions which usually act as the hydrogen bond acceptors. An anion is held in a very limited space close to the NH direction of a peptide group and the strongly localized nature of the interaction imparts rigidity to the binding. There are patterns in the binary and ternary interactions involving various anions. It is found from the anion-bound (especially phosphate and sulphate ion) protein crystals that protein segments comprised of the main chain atoms along with the side chains constitute the functional interface towards recognizing the anions and in most of the cases such interactions are found to be dictated by the specific conformation of the protein backbone in the functional interface^{22, 24}. Nest²⁷⁻²⁸ is a common three residue motif in proteins in which anionic or negatively charged atoms or groups bridge pairs of main chain NH atoms and this is common within larger motifs like the Schellman loop which participate in binding anionic groups such as phosphates and iron–sulphur centers. Another three to four residue motif is Niche³² that bind K^+ , Na^+ or Ca^{2+} occur in some functional contexts in the cyclic peptides valinomycin and antamanide. In several enzymes which are allosterically activated by Na^+ or K^+ ; and in the calcium pump, a Niche is integrally involved in the ion transport. In

Nest main-chain NH groups form a potential anion-binding site which is characterized by the conformation of backbone dihedral angles of consecutive protein residues. Backbone conformation of adjacent residues helps to form an atom-sized cavity where the main chain NH groups pointing inwards, ideal for an anion-binding site (Figure 1.1(a))²⁷. The first and third NH groups are well positioned for hydrogen bonding to an atom in the cavity generated, while the middle NH sometimes also does so, though it points more to the side. Nest is found to occur in the P-loops of nucleotide triphosphate-binding proteins and in the loops that bind [2Fe2S] or [4Fe4S] in iron-sulphur proteins. P-loop is the glycine-rich sequence that binds the triphosphates in a large superfamily of ATP and GTP-binding proteins.

Several other anion recognition motifs in protein structures are reported in the literature²²⁻³⁰ based only on main chain atoms of either a three or four residue segment of the polypeptide chain. These motifs include: (1) a motif for the recognition of the free phosphate ion or sulphate; (2) a motif for the recognition of adenine, adenine-containing nucleotides, and some other nucleotides and nucleotide analogs; (3) the “structural P-loop” motif for the recognition of the phosphate group in nucleotides; and (4) the phosphate group binding “cup” in pyridoxal phosphate-dependent enzymes. Each of these motifs has been observed to occur in multiple families having different folds. It was found from the phosphate group binding cup in PLP-dependent enzymes with different folds^{21, 33}, that the cup is partially formed by three structurally invariant main chain atoms. These invariant positions lie along a three residues segment. The first invariant position contains the alpha carbon atom (C^α) of the first amino acid residue of the segment, while the second and the third positions are formed by the backbone amide nitrogen atoms of the second and third amino acid residues of the segment (Figure 1.1(b)). This particular structural motif is termed as 'C^αNN' motif³⁴. Unlike other

motifs, where the interaction with anion is mediated either via the positively charged side chains and/or via the main chain amides of the anion recognition motif of proteins^{28, 34-35}, participation of the three main chain atoms (C_{i-1} , N_i , N_{i+1}) of the structurally invariant three consecutive residues at the functional interface^{6-8, 36} and especially the presence of C^α atom in the motif makes the ‘ $C^\alpha NN$ ’ motif unique. This motif has also been observed to occur in multiple families having different folds like adenine-binding proteins, ATP-binding enzymes and further different families of ATP, NAD(P), FAD and CoA-binding proteins³⁴.

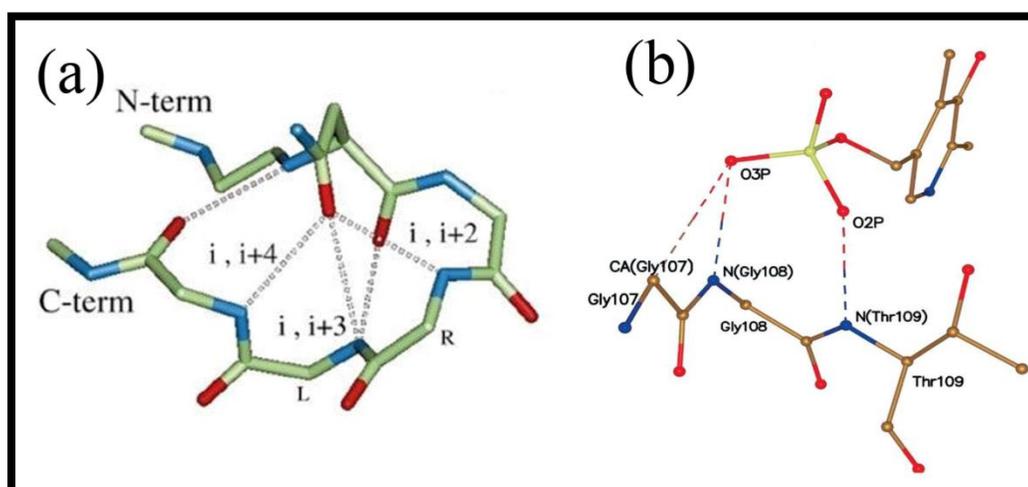


Figure 1.1: Main chain anion binding motif in protein: (a) Nest motif as found in β -bulge loop²⁷ (b) $C^\alpha NN$ motif as found in PLP³⁴.

In PLP-dependent enzymes and other proteins, phosphate and sulphate groups are often bound near the amino terminus of positively charged α -helix and in the presence of several peptide NH groups that are free to hydrogen bond^{22, 37-40}. Novel sulphate or phosphate binding structural ‘ $C^\alpha NN$ ’ motif also found to present mostly in a loop region preceding an anchoring helix^{31, 34}. It was found from the crystal structures that in most of the cases ligand atoms interacting with the motif are oxygen atoms derived from either phosphate or sulphate.

Although interactions of phosphate/sulphate with the motif occur solely through protein backbone atoms, the conservation of amino acid types in the motif sequence indicates that side chains may also have an important influence on conformation, specificity and/or the strength of the interactions. Bioinformatics studies suggest that the consensus sequence-pattern, for the C^αNN motif is [Gly/Ser]-Xaa-[Thr/Ser] with right-handed β_{αα} or β_{αβ} conformation, where Xaa is any amino acid^{22, 34}. The most frequently observed conformation for the C^αNN motif is β_{αα} followed in frequency by β_{αβ}, indicates that phosphate often binds at the amino terminus of an α-helix. The second and the third amino acid residues of the β_{αα}-tripeptide constitute the beginning of this anchoring α-helix, whereas the first position of the β_{αα}-tripeptide corresponds to the N-cap residue⁴¹. Previous biophysical studies on the C^αNN motif appended at the N-terminal of a model helical peptide indicate that the motif segment recognizes the anions (both sulphate and phosphate ions) through local interaction via context free environment and the presence of polar residues enhances the anion recognition⁴²⁻⁴⁴.

Interestingly, several features are shared by the anion binding C^αNN motif, structural P-loop and Nest motifs, like involvement of main chain interactions with the ligand, preference of amino acid residues in the motif, and special conformational properties of the motif residues. The detailed understanding of the microscopic basis of anion recognition and influence of anionic sulphate/phosphate in the stability and conformation of the different peptide segments in the anion-binding motifs is lacking yet.

In this thesis I consider sulphate and biphosphate interaction with C^αNN motif. I carry out classical all-atom Molecular Dynamics (MD) simulation to understand microscopic basis of interactions of anionic sulphate and biphosphate with C^αNN motif in functional proteins, as discussed in chapter 2. MD simulation study shows that anion induced conformational

preference of C^αNN motif residues are sequence specific. For a given peptide sequence, the conformational preferences of the residues are dependent on the nature of the anion as well. The residues showing no shift in conformational preferences by the anion are responsible for anion recognition, which are mostly polar and basic residues.

The charged species tend to polarize the neutral atoms which cannot be taken into classical force field calculations. This leads us to perform quantum chemical calculations on this system as reported in chapter 3. Quantum chemical calculation reveals that stability of anion bound C^αNN motif is governed by the coordination of motif atoms with oxygen atoms of anion. An intimate connection between coordination and conformational preferences of the motif residues is also established. The residues which do not undergo conformational switch in presence of anion are found to coordinate with the anion at lowest energy.

C^αNN motif containing residues are functionally important, typically part of a substrate, cofactor or protein-binding site. Conserved residues in this motif can also recognize and interact with the phosphate group of a DNA nucleotide. In functional region of proteins, C^αNN motif is also found to interact with one or several phosphate groups that belong to a biologically important ligand such as FAD, NAD, FMN, ATP, PLP and control the enzymatic activity³⁴. These studies may also be relevant to protein engineering, anion reception, sensing and binding to ligand or DNA to control the enzymatic activity and biological functions.

Influence of anion on cationic antimicrobial peptide:

Another class of peptides whose functions are based on electrostatic interactions is the antimicrobial peptides (AMP) which act as host defense effector molecules found in all organisms. They display remarkable structural and functional diversity⁴⁵. Most AMPs have the ability to kill microbial pathogens directly, whereas others act indirectly by modulating the host defense systems. AMPs are evolutionary conserved in the genome. In higher organisms, AMPs constitute important components of the innate immunity, protecting the host against infections¹⁴⁻¹⁵. In contrast, bacteria produce AMPs in order to kill other bacteria competing for the same ecological niche¹⁶. The AMPs differ widely in sequence and source⁴⁶⁻⁴⁸. Currently, more than 2,000 AMPs have been reported in antimicrobial peptide database⁴⁷⁻⁴⁸.

Most AMPs are relatively short, commonly consisting of 10–50 amino acids, display an overall positive charge, and contain a substantial proportion (typically 50%) of hydrophobic residues^{11, 17, 49}. The AMPs are classified as those that are enriched in one or more amino acid residues e.g., proline-arginine or tryptophan-rich⁵⁰. AMPs are also classified based on their secondary structure into α -helical, β -sheet, or peptides with extended/random coil structure^{12, 49, 51} (Figure 1.2). The α -helical and β -sheet peptides are two largest groups, whereas the α -helical antimicrobial peptides are abundant in the extracellular fluids of insects and frogs and frequently exist as extended or unstructured conformers in solution. However, many of these peptides become helical upon interaction with amphipathic phospholipid membranes⁵². Two of the most studied α -helical AMPs are LL-37^{49, 53} and human lactoferricin⁵⁴ (Figure 1.2).

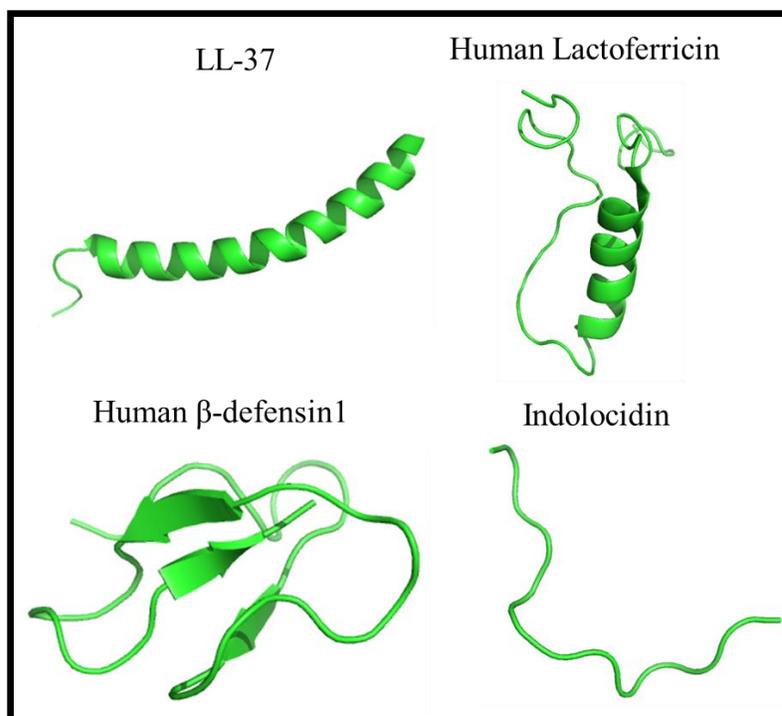


Figure 1.2: Secondary structural conformation of AMPs: LL-37 and human lactoferricin having α -helical peptides, human β -defensin1 with β -sheet peptides, and indolocidin with extended/random-coil structures. Structures are from Protein Data Bank⁵⁵ (PDB id: 2k6o, 1z6v, 1kj5, and 1g89 respectively).

The beta-sheet peptides represent a highly diverse group of molecules at the level of primary structure⁵⁶. β -sheet peptides are stabilized by disulphide bonds⁵⁷ and are organized to create an amphipathic molecule¹¹. Due to their rigid structure, the β -sheet peptides are more ordered in aqueous solution and do not undergo as drastic conformational change as helical peptides upon membrane interaction¹¹. The best studied β -sheet AMPs are the defensins, which are produced as inactive precursors in neutrophils, macrophages, and epithelial cells^{49, 58}. Less is known about the structures adopted by the proline-arginine rich and tryptophan-rich peptides.

Nearly all antimicrobial peptides form amphipathic structures upon interaction with target membranes¹¹. Interaction with the membrane is a key factor for the direct antimicrobial activity of AMPs, both when the membrane itself is targeted and when an intra cellular target must be reached by means of translocation^{12-13, 18, 59}. Amphipathicity can be achieved via a

multitude of protein conformations; however, one of the simplest and perhaps most elegant is the amphipathic helix⁵². For β -sheet amphipathicity is characterized by a variable number of β -strands, with relatively few or no helical domains, organized to create both polar and non-polar surfaces⁶⁰.

Most of the reported AMPs are cationic, and only a few of them are anionic^{47-48, 61}. The cationic AMPs often contain domains rich in basic residues like Arg and His at normal pH. Electrostatic interactions between the cationic AMPs and the negatively charged bacterial surface are critical for the interaction between peptides and microbial membrane^{11, 13, 18-19}. The cytoplasmic membranes of both Gram-positive and Gram-negative bacteria are rich in the phospholipids, phosphatidylglycerol, cardiolipin, and phosphatidylserine, which have negatively charged anionic head groups, highly attractive for positively charged AMPs^{11, 13, 62}. In contrast to bacteria, the cytoplasmic membrane of mammalian cells is rich in the zwitterionic phospholipids, phosphatidylethanolamine, phosphatidylcholine, and sphingomyelin, providing a membrane with a neutral net charge^{11, 62}. Therefore, interactions between AMPs and mammalian cell membrane occur mainly via hydrophobic interactions, which are relatively weak compared to the electrostatic interactions taking place between AMPs and bacterial membranes. This fundamental difference between microbial and mammalian membranes protects mammalian cells against AMPs and enables selective action of these peptides¹¹. Previous studies indicate that antimicrobial peptides not only interact with biomembranes of specific composition and asymmetry but may also promote remodeling of these membrane properties within target cells. These structure-activity themes of antimicrobial peptides are consistent with their likely roles in antimicrobial host defense¹¹. The central to the potential

pharmacologic application of antimicrobial peptides is the degree to which they differentiate, or may be engineered to differentiate, between microbial targets and normal host cells.

In last few years Arg-rich cationic peptides have drawn considerable interest due to their importance in immunological response to human immunodeficiency virus (HIV). Studies show that a small highly charged Arg-rich region of the HIV-1 Tat protein is capable of translocating cargoes of different molecular sizes, such as proteins, DNA, RNA, or drugs, across the cell membrane in an apparently energy independent manner. Such peptides are known to translocate through biological membranes quite fast. Arg-rich peptides quiet often act as anions scavengers which reduces the charge of the peptide. This results in the partitioning of peptides in the cell membrane which affect the immunological response significantly⁶³. Arginine, which provides electrostatic interactions in addition to hydrogen bond donors to interact with anionic species, is the most prevalent amino acid present in naturally occurring enzymes. It is found that two-thirds of all known enzymes either act on anionic substrates or require anionic coenzymes⁶⁴. The arginine side chains have been implicated in the mechanism of opening and closing voltage sensitive ion channels in response to transmembrane potentials⁶⁵⁻⁶⁶. Pore formation in lipid bilayer by amphiphilic guanidinium-rich polycations are cation rather than anion selective⁶⁷⁻⁷¹. Many examples for anion binding by mono-, oligo-, and polymeric guanidinium cations have been reported⁷²⁻⁷⁸. An example for anion binding by guanidinium-rich supramolecular polymers is phosphate scavenging by bilayer membranes composed of guanidinium amphiphiles⁷⁹.

Charge neutralization increases the lipophilicity of guanidinium oligo/polymers. It is proposed that translocation behavior of oligo- and polyarginines in bilayer membranes, which is termed as “arginine magic” may originate from counter anion scavenging^{63, 67-71}. Significant

efforts are being employed to use the guanidinium-rich polypeptides as drug carriers⁸⁰⁻⁸⁵. It is reported that anion-mediated variability of charge and solubility makes guanidinium-rich oligo/polymers adaptable to many environments⁶³. Poly- and hexa-arginine are found to phase transferred from water into chloroform in the presence of amphiphilic anions such as monomeric sodium dodecyl sulphate (SDS), egg yolk phosphati-dylglycerol (EYPG), cholesterol sulphate, pyrenebutyrate, and stearate⁶³. Refined combinations of hydrophilic anions, like phosphate, trifluoroacetic acid (TFA), heparin inhibited phase transfer of 5(6)-carboxyfluorescein (CF)-polyarginine complexes into and across chloroform and across lipid bilayer membranes. Binary anion cocktails are found to activate polyarginine as a carrier in bulk chloroform membranes⁶³. The activities of polyarginine in bulk and lipid membranes in presence of counter anion therefore suggest the biological significance of anion-mediated adaptability of the solubility of guanidinium-rich oligo/polymers. Due to the presence of multiple guanidinium cations in arginine rich HIV-1 Tat peptide, it can act as anion scavenger with antimicrobial activity.

The charged domains of the peptide interact with the hydrophilic head groups of the phospholipids, while the hydrophobic domains of the peptide interact with the hydrophobic core of the lipid bilayer, there by driving the AMP deeper into the membrane⁶². Several models have been proposed describing the next events occurring at the bacterial cytoplasmic membrane, which ultimately lead to membrane permeabilization and translocation^{51,61}. Further study of the mechanisms employed by antimicrobial peptides will significantly improve our understanding of how these molecules act to defend against infection. As well as it will facilitate new opportunities and approaches to discover and develop pharmacologic agents that enhances or optimize immune mechanisms and suppress the ability of pathogens to subvert

these mechanisms. Besides antimicrobial function, AMPs also serve as drug delivery vectors, antitumor agents, mitogenic agents, contraceptive agents, and signaling molecules in signal transduction pathways. Certain large hydrophilic drugs cannot easily penetrate through the cell membrane barriers. In such cases, AMPs with efficient membrane translocating property, which could enter the cells without causing damage to the membranes, were used as drug delivery vectors. The main feature of AMPs to serve as delivery vector is that they should be able to penetrate the cell membrane at very low concentrations (micro molar) without any specific receptors and capable of efficiently delivering electrostatically or covalently bound biologically active cargoes such as drugs into the cell interior.

This thesis reports the condensation of anionic biphosphate around HIV-1 Tat peptide using all-atom Molecular Dynamics (MD) simulation and model Monte Carlo simulation as discussed in chapter 4. The all-atom MD simulations reveal that anion condensed primarily around guanidinium group of arginine and there is an optimum concentration of biphosphate ions in solution where the condensation is maximum, while the solvent exposed area is the minimum. Anion condensation is also found to influence the α -helical character of the peptide residues. The model Monte Carlo Simulation supports the non-monotonic dependence of anion condensation.

MD simulations of peptide-lipid bilayer system in chapter 5 reveal that anion can influence the absorption properties of the Tat peptide on the lipid bilayer surface. In absence of anion arginine rich positively charged peptide is absorbed on the negatively charged bilayer surface without changing the overall structural properties of lipid bilayer. Condensation of anion around basic arginine and lysine residues of peptide alters the absorption properties of the Tat peptide on the bilayer surface. Alteration of absorption property is associated with the induced

Chapter 1

stress on the bilayer surface that leads to uneven thickness adaptation of the lipid bilayer. This study point to stress induced facilitation of the translocation of Tat in bilayer. My studies on Tat peptides in presence of anion in solvent and biomembrane may facilitate new opportunities and approaches to develop pharmacologic agents by controlling interactions and hydrophilicity of charged polypeptides. This will further help to design potential antimicrobial peptides for drug delivery and other biomedical applications.

Thesis is organized as follows. Chapter 2 describes all-atom simulation on anion bound C^oNN motif in functional proteins. Chapter 3 describes quantum chemical studies on anion binding and specificity to C^oNN motif. Chapter 4 reports the condensation of anionic biphosphate around cationic HIV-1 Tat peptide using all-atom simulation and model Monte Carlo simulation. Chapter 5 describes all-atom simulations of Tat peptide-lipid bilayer system in presence and absence of anion. Overall conclusions and future perspective of the thesis are summarized in Chapter 6.

References:

1. Pabo, C. O.; Sauer, R. T. Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **1992**, *61*, 1053-95.
2. Ledvina, P. S.; Tsai, A. L.; Wang, Z.; Koehl, E.; Quioco, F. A. Dominant role of local dipolar interactions in phosphate binding to a receptor cleft with an electronegative charge surface: equilibrium, kinetic, and crystallographic studies. *Protein Sci.* **1998**, *7* (12), 2550-9.
3. Kavanaugh, J. S.; Rogers, P. H.; Case, D. A.; Arnone, A. High-resolution X-ray study of deoxyhemoglobin Rothschild 37 beta Trp-Arg: a mutation that creates an inter subunit chloride-binding site. *Biochemistry* **1992**, *31* (16), 4111-21.
4. Verlinde, C. L.; Noble, M. E.; Kalk, K. H.; Groendijk, H.; Wierenga, R. K.; Hol, W. G. Anion binding at the active site of trypanosomal triosephosphate isomerase. Monohydrogen phosphate does not mimic sulphate. *Eur. J. Biochem.* **1991**, *198* (1), 53-7.
5. Harlos, K.; Vas, M.; Blake, C. F. Crystal structure of the binary complex of pig muscle phosphoglycerate kinase and its substrate 3-phospho-D-glycerate. *Proteins* **1992**, *12* (2), 133-44.
6. Abdullaev, Z.; Bodrova, M. E.; Chernyak, B. V.; Dolgikh, D. A.; Kluck, R. M.; Pereverzev, M. O.; Arseniev, A. S.; Efremov, R. G.; Kirpichnikov, M. P.; Mokhova, E. N.; Newmeyer, D. D.; Roder, H.; Skulachev, V. P. A cytochrome c mutant with high electron transfer and antioxidant activities but devoid of apoptogenic effect. *Biochem. J.* **2002**, *362* (Pt 3), 749-54.
7. Barrett, T. E.; Savva, R.; Panayotou, G.; Barlow, T.; Brown, T.; Jiricny, J.; Pearl, L. H. Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell* **1998**, *92* (1), 117-29.

8. Cobessi, D.; Tete-Favier, F.; Marchal, S.; Azza, S.; Branlant, G.; Aubry, A. Apo and holo crystal structures of an NADP-dependent aldehyde dehydrogenase from *Streptococcus mutans*. *J. Mol. Biol.* **1999**, *290* (1), 161-73.
9. Martinez-Liarte, J. H.; Iriarte, A.; Martinez-Carrion, M. Inorganic phosphate binding and electrostatic effects in the active center of aspartate aminotransferase apoenzyme. *Biochemistry* **1992**, *31* (10), 2712-9.
10. Meiering, E. M.; Bycroft, M.; Fersht, A. R. Characterization of phosphate binding in the active site of barnase by site-directed mutagenesis and NMR. *Biochemistry* **1991**, *30* (47), 11348-56.
11. Yeaman, M. R.; Yount, N. Y. Mechanisms of antimicrobial peptide action and resistance. *Pharmacol. Rev.* **2003**, *55* (1), 27-55.
12. Takahashi, D.; Shukla, S. K.; Prakash, O.; Zhang, G. Structural determinants of host defense peptides for antimicrobial activity and target cell selectivity. *Biochimie* **2010**, *92* (9), 1236-41.
13. Malmsten, M. Interactions of Antimicrobial Peptides with Bacterial Membranes and Membrane Components. *Curr. Top. Med. Chem.* **2016**, *16* (1), 16-24.
14. Hancock, R. E. Cationic antimicrobial peptides: towards clinical applications. *Expert Opin. Investig. Drugs* **2000**, *9* (8), 1723-9.
15. Hancock, R. E.; Diamond, G. The role of cationic antimicrobial peptides in innate host defenses. *Trends Microbiol.* **2000**, *8* (9), 402-10.
16. Hassan, M.; Kjos, M.; Nes, I. F.; Diep, D. B.; Lotfipour, F. Natural antimicrobial peptides from bacteria: characteristics and potential applications to fight against antibiotic resistance. *J. Appl. Microbiol.* **2012**, *113* (4), 723-36.

17. Hancock, R. E.; Sahl, H. G. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nat. Biotechnol.* **2006**, *24* (12), 1551-7.
18. Yeung, A. T.; Gellatly, S. L.; Hancock, R. E. Multifunctional cationic host defence peptides and their clinical applications. *Cell. Mol. Life Sci.* **2011**, *68* (13), 2161-76.
19. Giuliani, A., Pirri, G., and Nicoletto, S. Antimicrobial peptides: an overview of a promising class of therapeutics. *Cent. Eur. J. Biol.* **2007**, *2*, 1-33.
20. Aitken, A. Protein consensus sequence motifs. *Mol. Biotechnol.* **1999**, *12* (3), 241-53.
21. de la Cruz, X.; Thornton, J. M. Factors limiting the performance of prediction-based fold recognition methods. *Protein Sci.* **1999**, *8* (4), 750-9.
22. Chakrabarti, P. Anion binding sites in protein structures. *J. Mol. Biol.* **1993**, *234* (2), 463-82.
23. Kobayashi, N.; Go, N. ATP binding proteins with different folds share a common ATP-binding structural motif. *Nat. Struct. Biol.* **1997**, *4* (1), 6-7.
24. Kinoshita, K.; Sadanami, K.; Kidera, A.; Go, N. Structural motif of phosphate-binding site common to various protein super families: all-against-all structural comparison of protein-monomonucleotide complexes. *Protein Eng.* **1999**, *12* (1), 11-4.
25. Denessiouk, K. A.; Johnson, M. S. When fold is not important: a common structural framework for adenine and AMP binding in 12 unrelated protein families. *Proteins* **2000**, *38* (3), 310-26.
26. Denessiouk, K. A.; Rantanen, V. V.; Johnson, M. S. Adenine recognition: a motif present in ATP-, CoA-, NAD-, NADP-, and FAD-dependent proteins. *Proteins* **2001**, *44* (3), 282-91.
27. Watson, J. D.; Milner-White, E. J. A novel main-chain anion-binding site in proteins:

the nest. A particular combination of phi, psi values in successive residues gives rise to anion-binding sites that occur commonly and are found often at functionally important regions. *J. Mol. Biol.* **2002**, *315* (2), 171-82.

28. Milner-White, E. J.; Nissink, J. W.; Allen, F. H.; Duddy, W. J. Recurring main-chain anion-binding motifs in short polypeptides: nests. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 1935-42.

29. Denesyuk, A. I.; Denessiouk, K. A.; Korpela, T.; Johnson, M. S. Functional attributes of the phosphate group binding cup of pyridoxal phosphate-dependent enzymes. *J. Mol. Biol.* **2002**, *316* (1), 155-72.

30. Denessiouk, K. A.; Johnson, M. S. "Acceptor-donor-acceptor" motifs recognize the Watson-Crick, Hoogsteen and Sugar "donor-acceptor-donor" edges of adenine and adenosine-containing ligands. *J. Mol. Biol.* **2003**, *333* (5), 1025-43.

31. Sheet, T.; Supakar, S.; Banerjee, R. Conformational preference of 'C^αNN' short peptide motif towards recognition of anions. *PLoS One* **2013**, *8* (3), e57366.

32. Torrance, G. M.; Leader, D. P.; Gilbert, D. R.; Milner-White, E. J. A novel main chain motif in proteins bridged by cationic groups: the niche. *J. Mol. Biol.* **2009**, *385* (4), 1076-86.

33. Denesyuk, A. I.; Denessiouk, K. A.; Korpela, T.; Johnson, M. S. Phosphate group binding "cup" of PLP-dependent and non-PLP-dependent enzymes: leitmotif and variations. *Biochim. Biophys. Acta* **2003**, *1647* (1-2), 234-8.

34. Denessiouk, K. A.; Johnson, M. S.; Denesyuk, A. I. Novel C^αNN structural motif for protein recognition of phosphate ions. *J. Mol. Biol.* **2005**, *345* (3), 611-29.

35. Bianchi, A.; Giorgi, C.; Ruzza, P.; Toniolo, C.; Milner-White, E. J. A synthetic hexapeptide designed to resemble a proteinaceous P-loop nest is shown to bind inorganic

phosphate. *Proteins* **2012**, *80* (5), 1418-24.

36. Ledvina, P. S.; Yao, N.; Choudhary, A.; Quioco, F. A. Negative electrostatic surface potential of protein sites specific for anionic ligands. *Proc. Natl. Acad. Sci. USA* **1996**, *93* (13), 6786-91.

37. Hol, W. G.; van Duijnen, P. T.; Berendsen, H. J. The alpha-helix dipole and the properties of proteins. *Nature* **1978**, *273* (5662), 443-6.

38. Hol, R. K. W. C. H. D. M. G. J. Interaction of pyrophosphate moieties with alpha-helices in dinucleotide-binding proteins. *Biochemistry* **1985**, *24*, 1346-1357.

39. Copley, R. R.; Barton, G. J. A structural analysis of phosphate and sulphate binding sites in proteins. Estimation of propensities for binding and conservation of phosphate binding sites. *J. Mol. Biol.* **1994**, *242* (4), 321-9.

40. Denessiouk, K. A.; Denesyuk, A. I.; Lehtonen, J. V.; Korpela, T.; Johnson, M. S. Common structural elements in the architecture of the cofactor-binding domains in unrelated families of pyridoxal phosphate-dependent enzymes. *Proteins* **1999**, *35* (2), 250-61.

41. Richardson, J. S.; Richardson, D. C. Amino acid preferences for specific locations at the ends of alpha helices. *Science* **1988**, *240* (4859), 1648-52.

42. Sheet, T.; Banerjee, R. Sulfate ion interaction with 'anion recognition' short peptide motif at the N-terminus of an isolated helix: A conformational landscape. *J. Struct. Biol.* **2010**, *171* (3), 345-52.

43. Sheet, T.; Banerjee, R. The 'C^αNN' motif: an intrinsic lover of sulfate and phosphate ions. *RSC Adv.* **2016**, *6* (59), 54129-54141.

44. Sheet, T.; Ghosh, S.; Pal, D.; Banerjee, R. Computational design of model scaffold for anion recognition based on the 'C^αNN' motif. *Biopolymers* **2017**, *108* (1), 1-15.

45. Andrade, A. C.; Van Nistelrooy, J. G.; Peery, R. B.; Skatrud, P. L.; De Waard, M. A. The role of ABC transporters from *Aspergillus nidulans* in protection against cytotoxic agents and in antibiotic production. *Mol. Gen. Genet.* **2000**, *263* (6), 966-77.
46. Angel, C. S.; Ruzek, M.; Hostetter, M. K. Degradation of C3 by *Streptococcus pneumoniae*. *J. Infect. Dis.* **1994**, *170* (3), 600-8.
47. Wang, G.; Li, X.; Wang, Z. APD2: The updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res.* **2009**, *37* (Database issue), D933-7.
48. Wang, Z.; Wang, G. APD: The Antimicrobial Peptide Database. *Nucleic Acids Res.* **2004**, *32* (Database issue), D590-2.
49. Pasupuleti, M.; Schmidtchen, A.; Malmsten, M. Antimicrobial peptides: key components of the innate immune system. *Crit. Rev. Biotechnol.* **2012**, *32* (2), 143-71.
50. Hancock, R. E. Peptide antibiotics. *Lancet* **1997**, *349* (9049), 418-22.
51. Nguyen, L. T.; Haney, E. F.; Vogel, H. J. The expanding scope of antimicrobial peptide structures and their modes of action. *Trends Biotechnol.* **2011**, *29* (9), 464-72.
52. Zhang, L.; Benz, R.; Hancock, R. E. Influence of proline residues on the antibacterial and synergistic activities of alpha-helical peptides. *Biochemistry* **1999**, *38* (25), 8102-11.
53. Epand, R. M.; Vogel, H. J. Diversity of antimicrobial peptides and their mechanisms of action. *Biochim. Biophys. Acta* **1999**, *1462* (1-2), 11-28.
54. Hunter, H. N.; Demcoe, A. R.; Jenssen, H.; Gutteberg, T. J.; Vogel, H. J. Human lactoferricin is partially folded in aqueous solution and is better stabilized in a membrane mimetic solvent. *Antimicrob. Agents Chemother.* **2005**, *49* (8), 3387-95.
55. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235-

42.

56. Kobayashi, S.; Hirakura, Y.; Matsuzaki, K. Bacteria-selective synergism between the antimicrobial peptides alpha-helical magainin 2 and cyclic beta-sheet tachyplesin I: toward cocktail therapy. *Biochemistry* **2001**, *40* (48), 14330-5.

57. Powers, J. P.; Hancock, R. E. The relationship between peptide structure and antibacterial activity. *Peptides* **2003**, *24* (11), 1681-91.

58. Lai, Y.; Gallo, R. L. AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense. *Trends Immunol.* **2009**, *30* (3), 131-41.

59. Jenssen, H.; Hamill, P.; Hancock, R. E. Peptide antimicrobial agents. *Clin. Microbiol. Rev.* **2006**, *19* (3), 491-511.

60. Hirakura, Y.; Kobayashi, S.; Matsuzaki, K. Specific interactions of the antimicrobial peptide cyclic beta-sheet tachyplesin I with lipopolysaccharides. *Biochim. Biophys. Acta* **2002**, *1562* (1-2), 32-6.

61. Mahlapuu, M.; Hakansson, J.; Ringstad, L.; Bjorn, C. Antimicrobial Peptides: An Emerging Category of Therapeutic Agents. *Front. Cell. Infect. Microbiol.* **2016**, *6*, 194.

62. Ebenhan, T.; Gheysens, O.; Kruger, H. G.; Zeevaart, J. R.; Sathekge, M. M. Antimicrobial peptides: their role as infection-selective tracers for molecular imaging. *Biomed. Res. Int.* **2014**, *2014*, 867381.

63. Sakai, N.; Matile, S. Anion-mediated transfer of polyarginine across liquid and bilayer membranes. *J. Am. Chem. Soc.* **2003**, *125* (47), 14348-56.

64. Riordan, J. F. Arginyl Residues and Anion Binding-Sites in Proteins. *Mol. Cell Biochem.* **1979**, *26* (2), 71-92.

65. Jiang, Y.; Ruta, V.; Chen, J.; Lee, A.; MacKinnon, R. The principle of gating charge

movement in a voltage-dependent K⁺ channel. *Nature* **2003**, 423 (6935), 42-8.

66. Schmidt, D.; Jiang, Q. X.; MacKinnon, R. Phospholipids and the origin of cationic gating charges in voltage sensors. *Nature* **2006**, 444 (7120), 775-9.

67. Sakai, N.; Sorde, N.; Das, G.; Perrottet, P.; Gerard, D.; Matile, S. Synthetic multifunctional pores: deletion and inversion of anion/cation selectivity using pM and pH. *Org. Biomol. Chem.* **2003**, 1 (7), 1226-31.

68. Lee, S.; Tanaka, T.; Anzai, K.; Kirino, Y.; Aoyagi, H.; Sugihara, G. Two mode ion channels induced by interaction of acidic amphipathic alpha-helical peptides with lipid bilayers. *Biochim. Biophys. Acta* **1994**, 1191 (1), 181-9.

69. Iwata, T.; Lee, S.; Oishi, O.; Aoyagi, H.; Ohno, M.; Anzai, K.; Kirino, Y.; Sugihara, G. Design and synthesis of amphipathic 3(10)-helical peptides and their interactions with phospholipid bilayers and ion channel formation. *J. Biol. Chem.* **1994**, 269 (7), 4928-33.

70. Agawa, Y.; Lee, S.; Ono, S.; Aoyagi, H.; Ohno, M.; Taniguchi, T.; Anzai, K.; Kirino, Y. Interaction with phospholipid bilayers, ion channel formation, and antimicrobial activity of basic amphipathic alpha-helical model peptides of various chain lengths. *J. Biol. Chem.* **1991**, 266 (30), 20218-22.

71. Anzai, K.; Hamasuna, M.; Kadono, H.; Lee, S.; Aoyagi, H.; Kirino, Y. Formation of ion channels in planar lipid bilayer membranes by synthetic basic peptides. *Biochim. Biophys. Acta* **1991**, 1064 (2), 256-66.

72. Schmidtchen, F. P.; Berger, M. Artificial Organic Host Molecules for Anions. *Chemical Rev.* **1997**, 97 (5), 1609-1646.

73. Snowden, T. S.; Anslyn, E. V. Anion recognition: synthetic receptors for anions and their application in sensors. *Curr. Opin. Chem. Biol.* **1999**, 3 (6), 740-6.

74. Wiskur, S. L.; Ait-Haddou, H.; Lavigne, J. J.; Anslyn, E. V. Teaching old indicators new tricks. *Acc. Chem. Res.* **2001**, *34* (12), 963-72.
75. Peczuh, M. W.; Hamilton, A. D. Peptide and protein recognition by designed molecules. *Chemical Rev.* **2000**, *100* (7), 2479-94.
76. Austin, R. J.; Xia, T.; Ren, J.; Takahashi, T. T.; Roberts, R. W. Designed arginine-rich RNA-binding peptides with picomolar affinity. *J. Am. Chem. Soc.* **2002**, *124* (37), 10966-7.
77. Rومان, M.; Lievin, J.; Buisine, E.; Wintjens, R. Cation- π /H-bond stair motifs at protein-DNA interfaces. *J. Mol. Biol.* **2002**, *319* (1), 67-76.
78. General, S.; Thunemann, A. F. pH-sensitive nanoparticles of poly (amino acid) dodecanoate complexes. *Int. J. Pharm.* **2001**, *230* (1-2), 11-24.
79. Ariga, K.; Kunitake, T. Molecular recognition at air-water and related interfaces: Complementary hydrogen bonding and multisite interaction. *Acc. Chem. Res.* **1998**, *31* (6), 371-378.
80. Herce, H. D.; Garcia, A. E. Molecular dynamics simulations suggest a mechanism for translocation of the HIV-1 TAT peptide across lipid membranes. *Proc. Natl Acad. Sci. USA* **2007**, *104* (52), 20805-20810.
81. Chauhan, A.; Tikoo, A.; Kapur, A. K.; Singh, M. The taming of the cell penetrating domain of the HIV Tat: myths and realities. *J. Control. Release* **2007**, *117* (2), 148-62.
82. Henriques, S. T.; Melo, M. N.; Castanho, M. A. How to address CPP and AMP translocation? Methods to detect and quantify peptide internalization in vitro and in vivo. *Mol. Membr. Biol.* **2007**, *24* (3), 173-84.
83. Pooga, M.; Langel, U. Synthesis of cell-penetrating peptides for cargo delivery. *Methods Mol. Biol.* **2005**, *298*, 77-89.

Chapter 1

84. Wadia, J. S.; Dowdy, S. F. Transmembrane delivery of protein and peptide drugs by TAT-mediated transduction in the treatment of cancer. *Adv. Drug Deliv. Rev.* **2005**, *57* (4), 579-96.
85. Al-Taei, S.; Penning, N. A.; Simpson, J. C.; Futaki, S.; Takeuchi, T.; Nakase, I.; Jones, A. T. Intracellular traffic and fate of protein transduction domains HIV-1 TAT peptide and octaarginine. Implications for their utilization as drug delivery vectors. *Bioconjug. Chem.* **2006**, *17* (1), 90-100.



CHAPTER 2

**STRUCTURAL BASIS OF ANION
SPECIFICITY IN C^αNN MOTIF AT THE
FUNCTIONAL INTERFACE OF
PROTEINS**



2.1 Introduction:

Proteins play an integral role in regulating key biological and biochemical processes within a living cell. Quite often functions of proteins depend on their interaction with ligands, including small molecules, ions, peptides and so on. Such interactions pertain a pivotal role in DNA–protein interaction¹, active transport of phosphate², structural equilibrium of deoxyhemoglobin³, competitive inhibition of triosephosphate isomerase⁴ and phosphoglycerate kinase⁵ and many more⁶⁻¹⁰. How these small molecules or ligands and ions recognize the particular binding site or active site of the protein molecules and trigger the biological functions is not understood well.

The molecular structures revealed by X-ray crystallography and NMR often show various ligand-recognition motifs in proteins¹¹⁻²⁴. Among the motifs found in different proteins involved in ion binding, some are specific to cation binding^{21, 23-24} while some other recognize anions^{11, 17, 20, 22}. It was found that anion binding motifs in proteins are typically conserved in sequence and conformation^{14, 17, 20}. Crystal structure based studies have shown that such motifs often occur in loop regions preceding a helix and interaction with the anions can induce their well-defined conformations. Detailed bioinformatics studies¹¹, with a number of proteins show that among the anion binding motifs, the C^αNN motif, comprised of backbone C_{i-1}^α(C^α), N_i(N1) and N_{i+1}(N2) atoms of three successive residues for recognising anions, like sulphate (SO₄²⁻) and phosphate (PO₄³⁻), are present in functional interface¹¹. Amino acid analysis on these proteins using the fold classification based on structure-structure alignment of proteins (FSSP database)²⁵ indicates preferences of residue type in the motif (Table 2.1).

Table 2.1: Percentage of occurrence of different amino acids in C^α, N1 and N2 position of C^αNN motif according to FSSP database¹¹

Amino Acid as one letter code	% in 1 st position	% in 2 nd position	% in 3 rd position
A	9.8	9.8	3.9
C	-	1.9	1.9
D	-	11.7	5.8
E	-	5.8	5.8
F	-	3.9	-
G	41.1	17.6	11.7
H	1.9	-	5.8
I	1.9	5.8	-
K	3.9	11.7	3.9
L	5.8	5.8	3.9
M	1.9	-	1.9
N	3.9	1.9	1.9
P	-	-	-
Q	3.9	3.9	3.9
R	-	7.8	-
S	25.4	-	17.6
T	-	3.9	19.6
V	-	3.9	5.8
W	-	1.9	1.9
Y	-	1.9	5.8

The occurrence of Glycine (G) and Serine (S) as the first residue of C^αNN motif is high (41% and 25% respectively). There is preference for hydrophobic residue (51%) at the middle site. The third position is either a polar residue (50%) with mostly S (18%) and Threonine (T, 20%) or a hydrophobic (32%) residue. The motif residues further show structural preferences²⁶⁻²⁷ (see Appendix A2.1 for details). The first amino acid of the three residues sequence usually adopts beta (β) conformation, while the second and the third residue adopt alpha (α)

conformation. According to the crystal structure the geometry of C^αNN motif bound to sulphate¹¹, out of four oxygen atoms (O) of anion (sulphate), two participate in interaction with the motif. One O atom concurrently interacts with the C^α and N1, while the other interacts only with N2¹¹. It is further reported that if the first residue is serine, side chain oxygen atom at the gamma position (OG) of serine forms hydrogen bond with oxygen atom of the anion¹¹.

Recent biophysical and computational studies on the C^αNN motif appended at the N-terminal of a model helical peptide indicate that the motif segment recognizes the anions (both sulphate and the phosphate ions) through local interaction via context free environment and the presence of polar residues enhances the anion recognition²⁸⁻³¹. The C^αNN motif residues in functional proteins are also having large proportion of polar residues, although the motif sequences are quite diverse. Moreover, they are impregnated in polypeptides with variety of secondary structures. Several aspects of anion recognition by the C^αNN motif in functional proteins, like the role played by the polar residues, sequence variability, anion specificity and the solvent distribution around the motif atoms, are not understood very well.

Using the all-atom molecular dynamic (MD) simulations³²⁻³⁶ (see Appendix A2.2 for details) I study the C^αNN motif containing polypeptides from known crystal structure of several functional proteins at neutral pH condition. In choosing the relevant polypeptide, the succeeding and the preceding structural elements are retained. At physiological pH, the phosphate (PO₄³⁻) is mostly available in dibasic (HPO₄²⁻) form. So HPO₄²⁻ are considered instead of PO₄³⁻ in the model system¹². Each polypeptide is simulated both in presence and absence of anionic sulphate and biphosphate. The simulations show that the conformations both with and without anions are similar in some motif residues, mostly polar and basic; while there are shifts in conformational preferences in the two cases for most other residues. The

anion induced conformations depend strongly on the sequence and the type of anion, but not so sensitive to the presence of water molecules in the vicinity of the motif atoms and the flanking residues containing the motif.

In this chapter, I have discussed material & methods in section 2.2, results and discussions in section 2.3 and conclusion in section 2.4.

2.2 Material & Methods:

2.2.1 Systems for study:

The model systems are chosen considering the relevant polypeptide having C^αNN motif in functional proteins and retain the succeeding and the preceding structural elements of the motif. Table 2.2 shows polypeptide sequences containing C^αNN motif of functional proteins with known X-ray crystallographic structures at neutral pH (=7.0) which have been studied in this work. I have named the sequences as per one letter code of the constituent motif residues. The proteins are: (1) Rhesus rotavirus VP4, (PDB ID: 1KQR³⁷) which acts as antigenic epitope corresponding to SQT fragment; (2) Pyridoxal 5'-phosphate (PLP) dependent enzyme 1-Aminocyclopropane-1-carboxylate deaminase, (PDB ID: 1F2D³⁸) involved in transamination reactions, corresponding to SNQ fragment; (3) Extracellular domain of growth hormone receptor (PDB ID: 1AXI³⁹), corresponding to GIH fragment; (4) Polynucleotide phosphorylase, (PDB ID: 1E3H⁴⁰), a polyribonucleotide nucleotidyl transferase involved in degrading prokaryotic mRNA, corresponding to LYD fragment; and (5) Dihydrofolate reductases (PDB ID: 1VDR⁴¹), a halo-phillic protein, corresponding to SRS fragment. I studied each polypeptide both by removing the anion (denoted by WOS and WOP for sulphate and biphosphate respectively) and retaining the anion (denoted by WS and WP for sulphate and biphosphate

respectively). At physiological pH, the phosphate (PO_4^{3-}) is mostly available in dibasic (HPO_4^{2-}) form. So, I consider HPO_4^{2-} instead of PO_4^{3-} in my study.

Table 2.2: Simulated polypeptides having tri residue segment motif

Serial no	PDB Ids	Polypeptide	Motif	Simulation Box size (nm)
1	1KQR	T124 to K145	SQT fragment: S134(C $^\alpha$)Q135(N1)T136(N2)	5.78
2	1F2D	H69 to G92	SNQ fragment: S78(C $^\alpha$)N79(N1)Q80(N2)	5.24
3	1AXI	L141 to R156	GIH fragment: G148(C $^\alpha$)I149(N1)H150(N2)	4.98
4	1E3H	I124 to A148	LYD fragment: L132(C $^\alpha$)Y133(N1)D134(N2)	5.58
5	1VDR	D54 to V82	SRS fragment: S65(C $^\alpha$)R66(N1)S67(N2)	5.01

2.2.2 Molecular Dynamic (MD) Simulation:

The polypeptide is kept in a cubic box (box lengths for different cases given in Table 2.2) and placed at least 10 Å from the box edge and the box is filled with TIP3P water for solvation of the peptide. The hydrogen atoms are taken to be united atoms with C $^\alpha$, N, C, O, OG atoms. I add required numbers of Na $^+$ and Cl $^-$ ions for neutralizing the charged system. The resulting structure is relaxed through energy minimization. The energy minimized structure is subject to MD simulations³²⁻³⁶ (see Appendix A2.2 for details). I use GRONINGEN MACHiNE for Chemical Simulations (GROMACS) 4.6.7⁴²⁻⁴³ tool and Amber99ILDN force field⁴⁴ for simulation. The systems are equilibrated in two phases- (i) NVT equilibration and (ii) NPT equilibration. The system is equilibrated at the desired temperature (300 K) and pressure (1

Bar) using the periodic boundary conditions with 1 femto-second time step. I use the particle-mesh Ewald method for treating the long ranged electrostatic interactions. I check the equilibration from the saturation of root-mean-square deviation (RMSD) of peptide segments considering only backbone heavy atoms. The trajectories are visualized with visual molecular dynamics (VMD)⁴⁵.

2.2.3 Trajectory analysis:

The following analysis are performed using the equilibrated portion of the trajectories as confirmed from the RMSD of the peptide segments with respect to its initial energy minimized structure:

φ-ψ correlation plot: The correlation plots are generated for the backbone torsion angles (ϕ and ψ) for the residues of C^αNN motif from the equilibrated trajectory both in presence and absence of sulphate and biphosphate. I also calculate the relative frequency of occurrence of different secondary structural elements like β strand, right handed alpha helix (RH), left-handed alpha helix (LH) and random coil (C).

Distance distribution: The distances between sulphur (S) of sulphate and phosphorus (P) of biphosphate and C^α, N1 and N2 atom of C^αNN motif are calculated from simulated trajectories. Then the frequency distributions of distances between S of sulphate or P of biphosphate and C^α, N1 and N2 atoms of this motif respectively of each polypeptide over the equilibrated trajectory are calculated, using bins with size 0.5 Å.

Water distribution around motif residues: The radial distribution functions $g(r)$ ³³ of the oxygen atom of water around the C^α, N1 and N2 atoms are calculated. The radial distribution function $g(r)$ is defined as: $g(r) = \langle \Delta N(r) \rangle / (4\pi N\rho\Delta r)$, where $\langle \Delta N(r) \rangle$ is the number of oxygen atom of

water molecules averaged over time, within a distance $r \pm \Delta r/2$ from the atom of interest, ρ is the density of water and N is the total number of water molecules within the system.

Conformational thermodynamics: Conformational thermodynamics of the peptide-anion interaction are extracted by the histogram based method (HBM) using the dihedral angle distributions of peptide fragment (see Appendix A2.3 for details)⁴⁶⁻⁴⁸. The equilibrium conformational free energy cost associated with any peptide dihedral angle θ in the sulphate bound state as compared to the biphosphate bound state is defined as $\Delta G^{conf}(\theta) = -k_B T \ln[H_s^{max}(\theta)/H_p^{max}(\theta)]$, where $H_s(\theta)$ and $H_p(\theta)$ is the distribution of dihedral angle θ in sulphate and biphosphate bound state respectively and the ‘max’ superscript denotes the peak values of the histograms. On the other side the conformational entropy for a particular dihedral can be defined by the Gibbs entropy formula as $S^{conf}(\theta) = -k_B \sum_i H_i(\theta) \ln H_i(\theta)$, where the sum is taken over the histogram bins. The conformational entropy change for the dihedral can be obtained from the given expression $\Delta S^{conf}(\theta) = S_s^{conf}(\theta) - S_p^{conf}(\theta)$. The conformational free energy and entropy cost for the individual peptide residues can be obtained by taking the sum of the contribution of free energy and entropy of each dihedral angles of that residue.

2.3 Results & Discussions:

The C^αNN motif containing sulphate and biphosphate bound polypeptide fragments used in this study are shown in Figures 2.1(a)-(e). In case of SQT fragment the entire polypeptide containing the polar motif residues Serine (S134), Glutamine (Q135), Threonine (T136) is a loop (Figure 2.1(a)). In SNQ fragment motif residues S78, Asparagine (N79), Q80 are polar as well and the N-terminal flanking residues of the polypeptide in this structure form a loop, while the C-terminal flanking residues form a helix (Figure 2.1(b)). In GIH fragment none of the motif residues Glycine (G148), Isoleucine (I149), Histidine (H150) are polar and the entire polypeptide containing the motif is a loop (Figure 2.1(c)). In LYD fragment the motif residues are Leucine (L132), Tyrosine (Y133), Aspartic acid (D134) where the middle residue is polar. In this case the N-terminal flanking residues form loop, while those in C-terminal form a helix (Figure 2.1(d)). In SRS fragment the biphosphate binding motif residues are S65, Arginine (R66), S67 belonging to a loop, the middle residue being basic and the rest two polar (Figure 2.1(e)).

Equilibrations of the polypeptide systems both with and without the anion are ensured from the root-mean-square deviation (RMSD) plots, shown for all the cases both with and without anions in Figures 2.2(a)-(k). The saturation of the RMSD is highly system dependent and typically faster in the presence of the anions. The equilibrated regions of trajectories for each polypeptide are considered for detailed analysis.

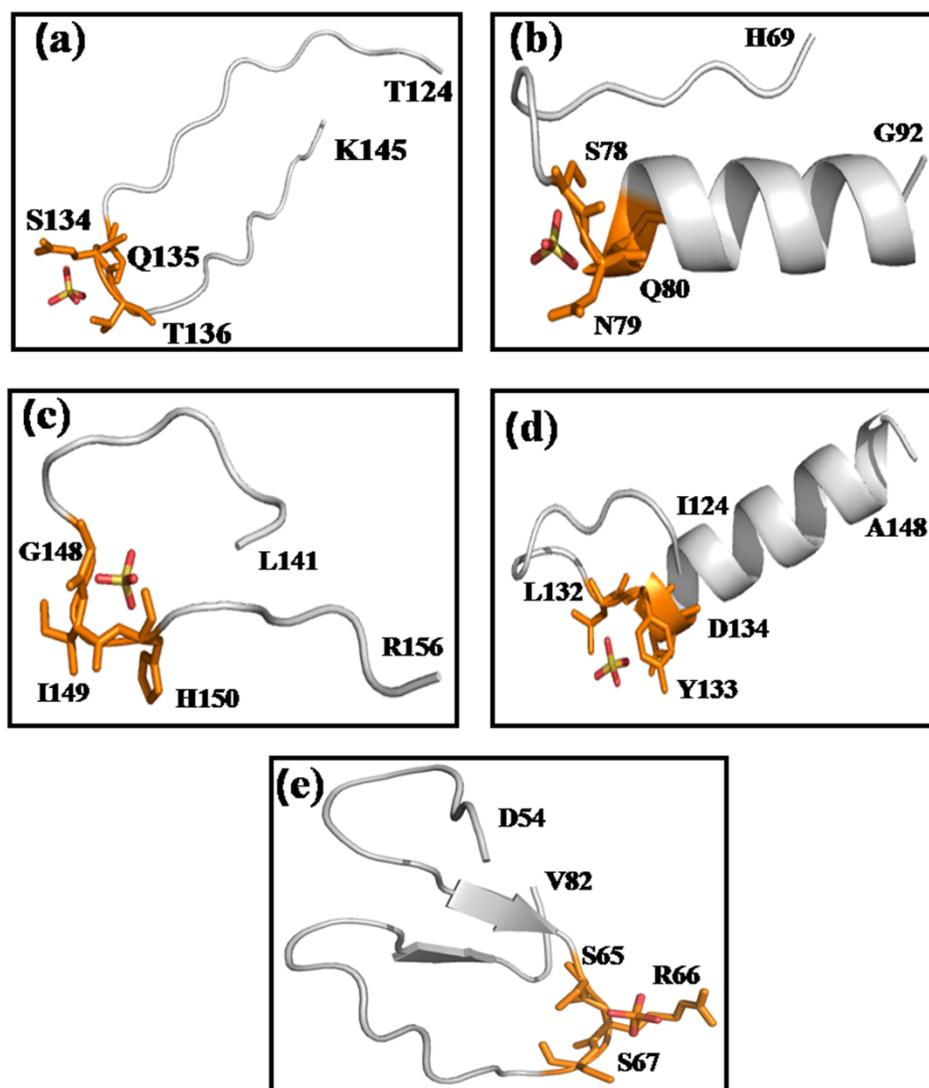


Figure 2.1: Crystal structure of polypeptide containing: (a) SQT fragment (PDB ID: 1KQR), (b) SNQ fragment (PDB ID: 1F2D), (c) GIH fragment (PDB ID: 1AXI), (d) LYD fragment (PDB ID: 1E3H) and (e) SRS fragment (PDB ID: 1VDR), where the residues of the motif are shown in “stick” (orange) representation.

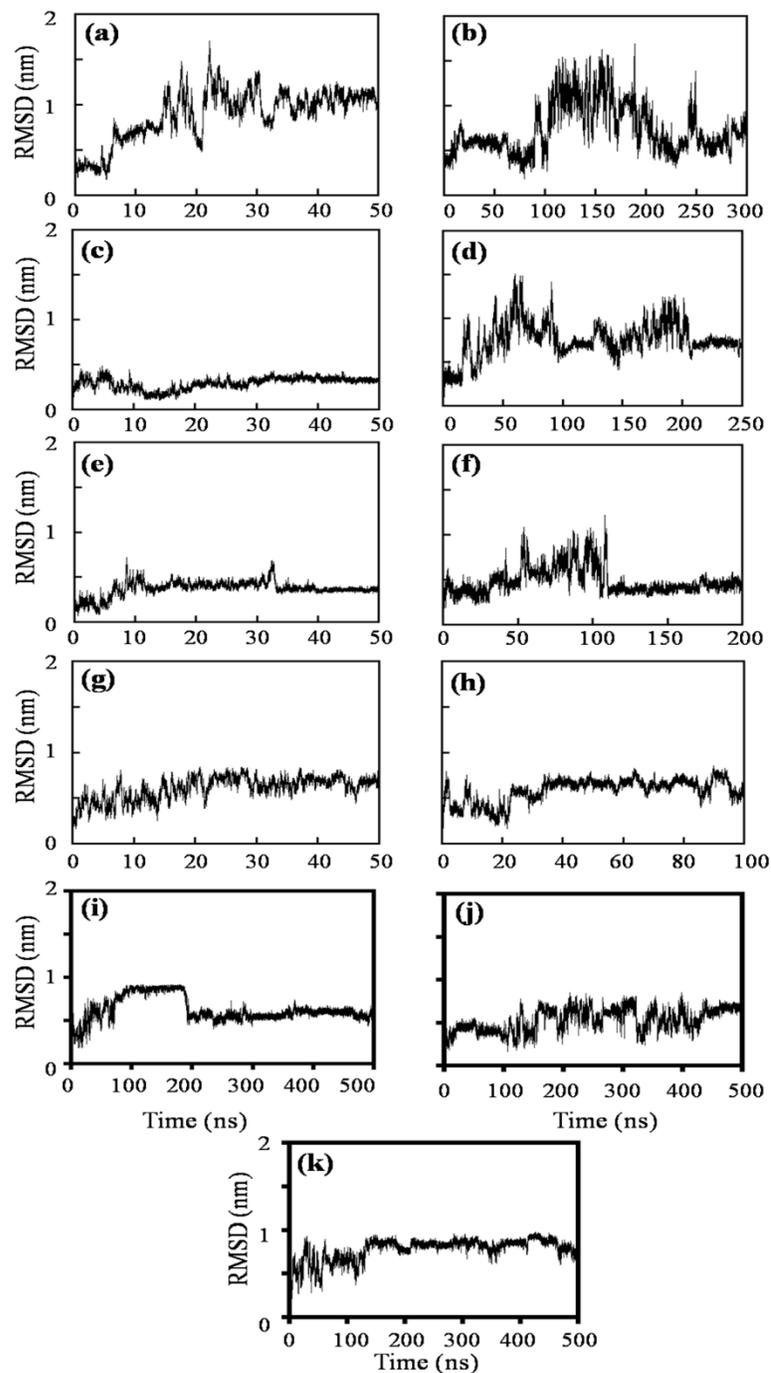


Figure 2.2: RMSD versus time plots of the peptide fragment: (a) and (b) for SQT fragment in WS and WOS respectively, (c) and (d) for SNQ fragment in WS and WOS respectively, (e) and (f) for GIH fragment in WS and WOS respectively, (g) and (h) for LYD fragment in WS and WOS respectively, (i) and (j) for SRS fragment in WP and WOP respectively and (k) for LYD fragment in WP.

2.3.1 Anion induced conformational preferences of the motif residues:

Conformational preferences of the motif residues are analyzed based on correlation plots between the backbone torsion angles (ϕ and ψ) of motif residues over the simulated equilibrated trajectory in terms of secondary structural elements, like right handed helix (RH), left handed helix (LH), beta (β) and random coil (C) conformation. In absence and presence of anion the correlation plots of the torsion angles for motif residues in SQT, SNQ, GIH, LYD and SRS fragments are shown in Figure 2.3-2.7.

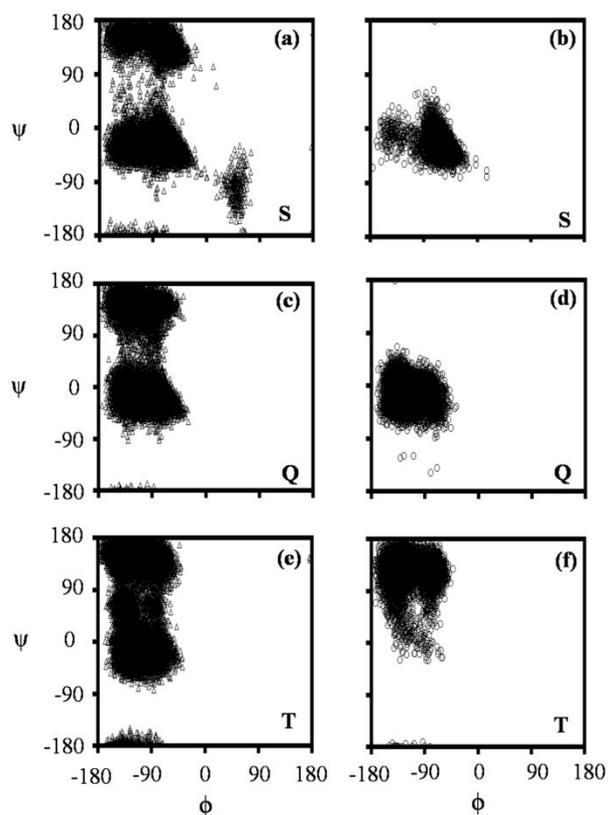


Figure 2.3: The ϕ - ψ correlation plots of the residues in SQT fragment: S134 in (a) WOS and (b) WS conditions; Q135 in (c) WOS and (d) WS conditions; and T136 in (e) WOS and (f) WS conditions.

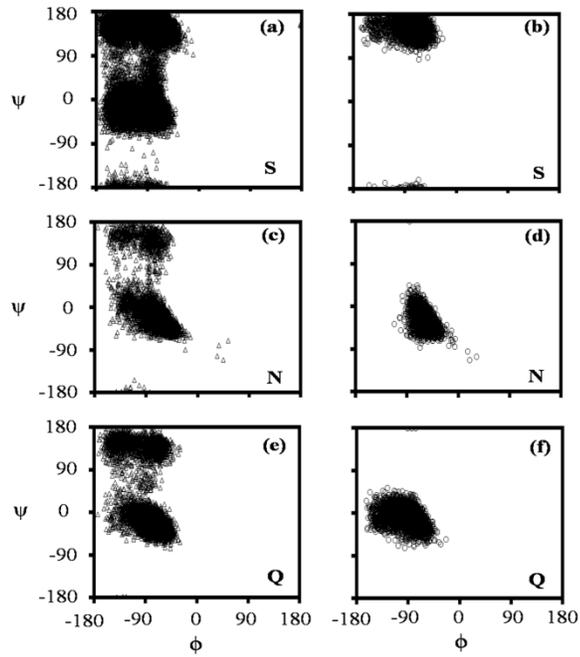


Figure 2.4: The ϕ - ψ correlation plots of the residues in SNQ fragment: S78 in (a) WOS and (b) WS conditions; N79 in (c) WOS and (d) WS conditions; and Q80 in (e) WOS and (f) WS conditions.

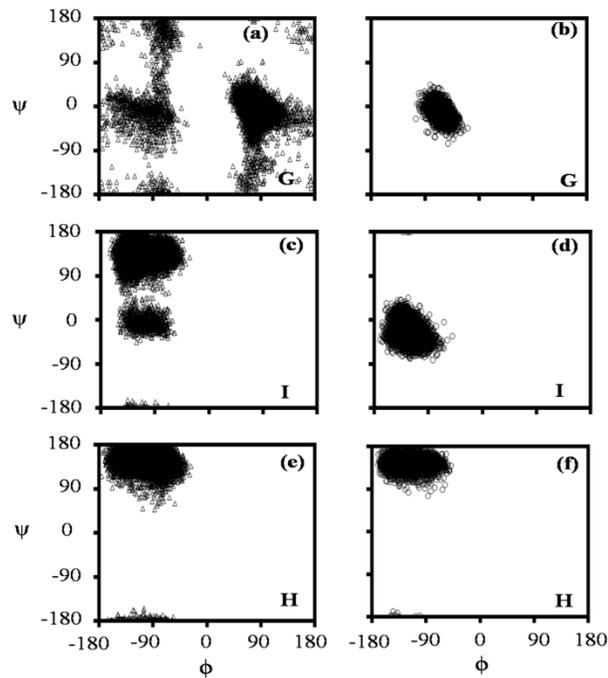


Figure 2.5: The ϕ - ψ correlation plots of the residues in GIH fragment: G148 in (a) WOS and (b) WS conditions; I149 in (c) WOS and (d) WS conditions; and H150 in (e) WOS and (f) WS conditions.

Chapter 2

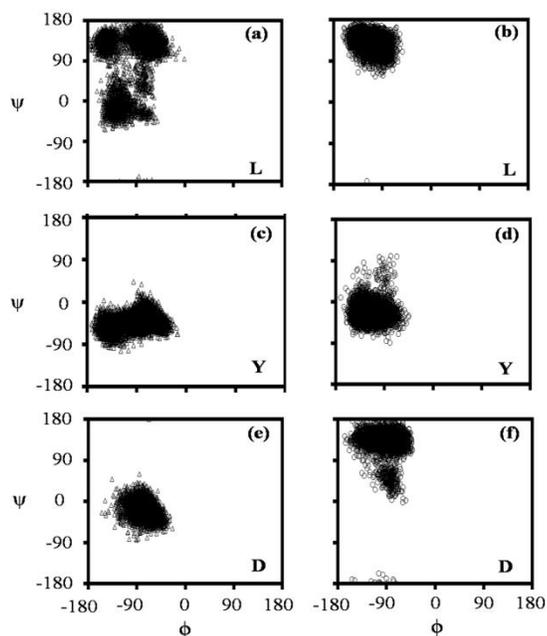


Figure 2.6: The ϕ - ψ correlation plots of the residues in LYD fragment: L132 in (a) WOS and (b) WS conditions; Y133 in (c) WOS and (d) WS conditions; and D134 in (e) WOS and (f) WS conditions.

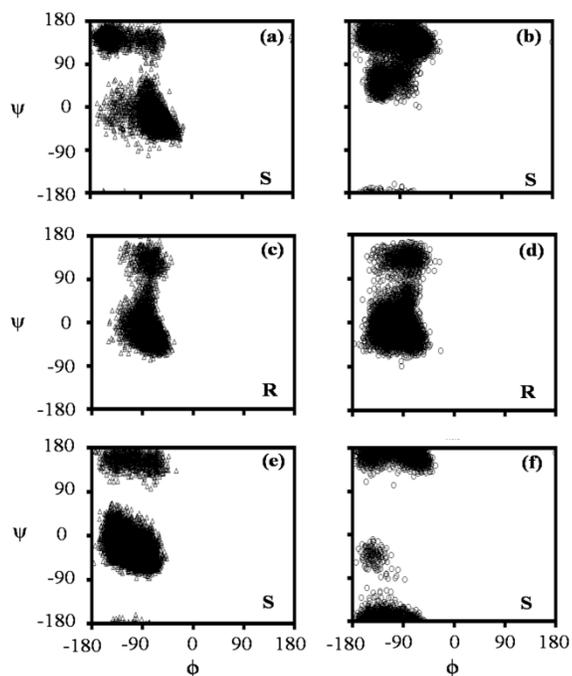


Figure 2.7: The ϕ - ψ correlation plots of the residues in SRS fragment: S65 in (a) WOP and (b) WP conditions; R66 in (c) WOP and (d) WP conditions; and S67 in (e) WOP and (f) WP conditions.

Using the ϕ - ψ correlation plot of motif residues in each peptide fragment, the percentages of different secondary structural elements of the motif residues are calculated in the equilibrated trajectory in presence and absence of anion. The detailed secondary structural preference of motif residues in different peptide fragment in absence and presence of sulphate and biphosphate are tabulated in Table 2.3 and Table 2.4, respectively. Secondary structural conformation of motif residues in different peptide fragments as found in their crystal structure along with the values of ϕ and ψ are also shown in the tables.

The case of SQT is discussed in details below. In crystal structure (1KQR) S134 shows β conformation, whereas both Q135 and T136 have RH conformation. The preference for conformation of S fluctuates between RH, β and C in WOS condition ((Figure 2.3(a)). The percentages of occurrences of these structures over the entire trajectory (Table 2.3) show that the major WOS conformation is RH (60%), despite having substantial conformational flexibility. However, the correlation plot shows that S in SQT for WS is in RH conformation (Figure 2.3(b)) over entire trajectory. The WS conformation is in agreement to the crystal structure data. Thus, the major WOS conformation is stabilized by the sulphate in WS state, indicating similar conformational preferences in both cases. I find that Q switches between RH and β conformations with RH having slightly higher preference to β (Table 2.3) in the WOS case (Figure 2.3(c)) and is in RH conformation in WS (Figure 2.3(d)). Here again there is no shift in conformational preference. On the other hand, T has propensities for both RH (60%) and β (40%) conformations (Figure 2.3(e)) in WOS condition and is in primarily β conformation (Figure 2.3(f)) in WS condition. In contrast to the first two motif residues, the minor conformation in WOS is stabilized by the sulphate so that there is a shift in conformational preference.

Table 2.3: The secondary structural elements of the motif residues in crystal structure and simulated structure for WOS, WS and sulphate replaced by biphosphate (RP) cases. (“-” indicates no occurrence at that particular position)

PDB Ids	Residues	System	RH	β	LH	C	Crystal Structure (ϕ, ψ)
SQT fragment (1KQR)	S (Polar)	WOS	60	30	-	10	β (-71.9, 134.7)
		WS	100	-	-	-	
		RP	79	21	-	-	
	Q (Polar)	WOS	55	45	-	-	RH (-74.6, -29.1)
		WS	100	-	-	-	
		RP	51	49	-	-	
	T (Polar)	WOS	60	40	-	-	RH (-124.7, -16.2)
		WS	3	97	-	-	
		RP	95	5	-	-	
SNQ fragment (1F2D)	S (Polar)	WOS	47	53	-	-	β (-61.5, 144.7)
		WS	-	100	-	-	
		RP	-	100	-	-	
	N (Polar)	WOS	91	9	-	-	RH (-90.9, -48.2)
		WS	100	-	-	-	
		RP	99	-	-	1	
	Q (Polar)	WOS	73	27	-	-	RH (-50.3, -47.6)
		WS	100	-	-	-	
		RP	100	-	-	-	
GIH fragment (1AXI)	G (Hydrophobic)	WOS	20	15	27	38	β (-66.4, 157.7)
		WS	100	-	-	-	
		RP	21	31	19	29	
	I (Hydrophobic)	WOS	23	77	-	-	RH (-87.6, -19.9)
		WS	100	-	-	-	
		RP	27	73	-	-	
	H (Basic)	WOS	-	100	-	-	β (-117, 163.8)
		WS	-	100	-	-	
		RP	21	26	19	34	
LYD fragment (1E3H)	L (Hydrophobic)	WOS	45	55	-	-	β (-107.2, 119.3)
		WS	-	100	-	-	
		RP	100	-	-	-	
	Y (Polar)	WOS	100	-	-	-	RH (-80.9, -21.9)
		WS	97	3	-	-	
		RP	48	52	-	-	
	D (Acidic)	WOS	100	-	-	-	RH (-48.6, -57.4)
		WS	17	83	-	-	
		RP	100	-	-	-	

Chapter 2

For SNQ fragment, polar S shows slightly larger preference for β conformation than RH in WOS case (Table 2.3). S adopts β conformation in WS condition which is similar to that in the crystal structure. Thus, there is no shift in conformational preference induced by sulphate. The latter two residues (N and Q) do not show any shift in conformational preferences (both RH) in WOS and WS conditions as in the crystal structure (Figure 2.4). According to the conformational preferences of GIH fragment as shown in Table 2.3 the hydrophobic G residue lacks any dominant conformational preference in WOS condition. In WS condition the conformation shifts to RH in contrast to β conformation in crystal structure. This is not so surprising, for G in general lacks well defined secondary structure. Hydrophobic I in GIH fragment shows change from predominant β conformation in WOS condition to predominant RH conformation in WS condition. The basic residue H does not show any shift in conformational preferences (Figure 2.5). In case of LYD, it is evident from Table 2.3 that hydrophobic L shows no particular conformational preference between RH and β in WOS, while having bias to β conformation for WS. Polar Y does not change its conformational preferences, while acidic D conformation switches from RH in WOS to β conformation in WS (Figure 2.6).

In case of biphosphate bound SRS fragment Table 2.4 shows that both S are predominantly in RH conformation in the WOP case. Both S are mostly in β conformation in WP condition. However, the residue shows substantial fluctuations with minor contributions from other conformations. Thus, both S show shifts in conformational preferences in the WP case compared to the WOP case. However, the basic R does not show any change in preference, being in RH conformation for both the cases (Figure 2.7).

Table 2.4: The secondary structural elements of the motif residues in crystal structure and simulated structure for WOP and WP cases (“-” indicates no occurrence at that particular position)

PDB Ids	Residues	System	RH	β	LH	C	Crystal Structure (ϕ, ψ)
SRS fragment (1VDR)	S (Polar)	WOP	70	30	-	-	β (-154.8, 150.8)
		WP	11	89	-	-	
	R (Basic)	WOP	89	11	-	-	RH (-63.8, -64.3)
		WP	85	15	-	-	
	S (Polar)	WOP	83	17	-	-	β (-45.7, 140.9)
		WP	3	97	-	-	

Therefore, secondary structural preferences of motif residues in absence and presence of anion as mentioned in Table 2.3 and Table 2.4 clearly indicate: (1) the conformational preferences of the residues fluctuate without anion and (2) the anion typically stabilizes one of the residue conformations experienced in absence of the anion. The conformations of the motif residues are also in good agreement to their crystal structure conformations.

2.3.2 Conformational preferences of flanking residues:

The conformational preferences are examined for a couple of flanking residues in both N- and C-terminal in the vicinity of the motif residues as mentioned in Table 2.5a and 2.5b. The data in these tables show that the flanking residues undergo conformational fluctuations without the anion. However, the anions give stability to one of the conformations without anion. The anion induced conformations are in general agreement to the crystal structure data. In SQT fragment N132 and alanine (A133) are N-terminal and Q137 and tryptophan (W138) are C-terminal residues, all of which have loop conformation in sulphate bound crystal structure. Table 2.5.a shows shift in conformational preferences in all the residues except A133. In SNQ fragment, the N-terminal residues are R76 and Q77, both being in loop conformation and the C-terminal residues are T81 and R82 both having helix conformation in the crystal structure.

The simulated trajectories show no shift in conformational preferences in all these residues except R82 (Table 2.5.a).

Table 2.5.a: The secondary structural elements of flanking residues around the motif residues in simulated structure for both WOS and WS conditions (SQT, SNQ and GIH fragments) (“-” indicates no occurrence at that particular position)

PDB Ids	Residues	System	RH	β	LH	C
SQT fragment (1KQR)	N132	WOS	62	38	-	-
		WS	10	90	-	-
	A133	WOS	34	66	-	-
		WS	12	88	-	-
	Q137	WOS	92	8	-	-
		WS	12	88	-	-
	W138	WOS	23	77	-	-
		WS	80	20	-	-
SNQ fragment (1F2D)	R76	WOS	75	25	-	-
		WS	99	1	-	-
	Q77	WOS	61	39	-	-
		WS	100	-	-	-
	T81	WOS	73	27	-	-
		WS	100	-	-	-
	R82	WOS	37	63	-	-
		WS	100	-	-	-
GIH fragment (1AXI)	L146	WOS	6	94	-	-
		WS	100	-	-	-
	T147	WOS	6	35	58	1
		WS	7	93	-	-
	A151	WOS	1	99	-	-
		WS	2	98	-	-
	D152	WOS	84	16	-	-
		WS	2	98	-	-

The N-terminal flanking residues of GIH (L146 and T147), and C-terminal residues (A151 and D152) are all having loop conformation in crystal structure. Out of these residues only A151 does not show any shift in conformation (Table 2.5.a). None of the LYD N-terminal residues (D130 and H131, loop conformation in crystal structure) and C-terminal residues of these fragments (valine, V135 and V136, helix in crystal structure), shows shift in conformational preferences (Table 2.5.b).

Table 2.5.b: The secondary structural elements of flanking residues around the motif residues in simulated structure for LYD fragment in WOS and WS conditions and for SRS fragment in WOP and WP conditions (“-” indicates no occurrence at that particular position)

PDB Ids	Residues	System	RH	β	LH	C	
LYD fragment (1E3H)	D130	WOS	99	1	-	-	
		WS	99	1	-	-	
	H131	WOS	37	63	-	-	
		WS	1	99	-	-	
	V135	WOS	100	-	-	-	
		WS	98	2	-	-	
	V136	WOS	100	-	-	-	
		WS	99	1	-	-	
	SRS fragment (1VDR)	V63	WOP	2	98	-	-
			WP	6	94	-	-
M64		WOP	4	96	-	-	
		WP	11	89	-	-	
E68		WOP	5	95	-	-	
		WP	89	11	-	-	
R69		WOP	78	22	-	-	
		WP	63	37	-	-	

In SRS fragment, V63 and Methionine, M64 are the N-terminal residues and Glutamine, E68 and R69 are the C-terminal residues, all being loop in the crystal structure (Table 2.5.b). Here, except E68, no other residue shows shift in conformation preference between WP and WOP

conditions. Overall, the conformational preferences of the flanking residues do not seem to be well correlated to those of motif residues.

2.3.3 Motion of the anion around the motif:

Motion of the anion around the motif residues is analyzed based on the distribution of distance between motif atoms and anionic species in the equilibrated trajectory. The distances between the sulphur atom and C^α ($d_{C^\alpha-S}$), N1 (d_{N1-S}) and N2 (d_{N2-S}) atom of the motif for different fragments are considered here. The frequencies of these distances, $f(d_{C^\alpha-S})$, $f(d_{N1-S})$ and $f(d_{N2-S})$ for SQT, SNQ, GIH and LYD fragments are shown in Figures 2.8(a)-(d).

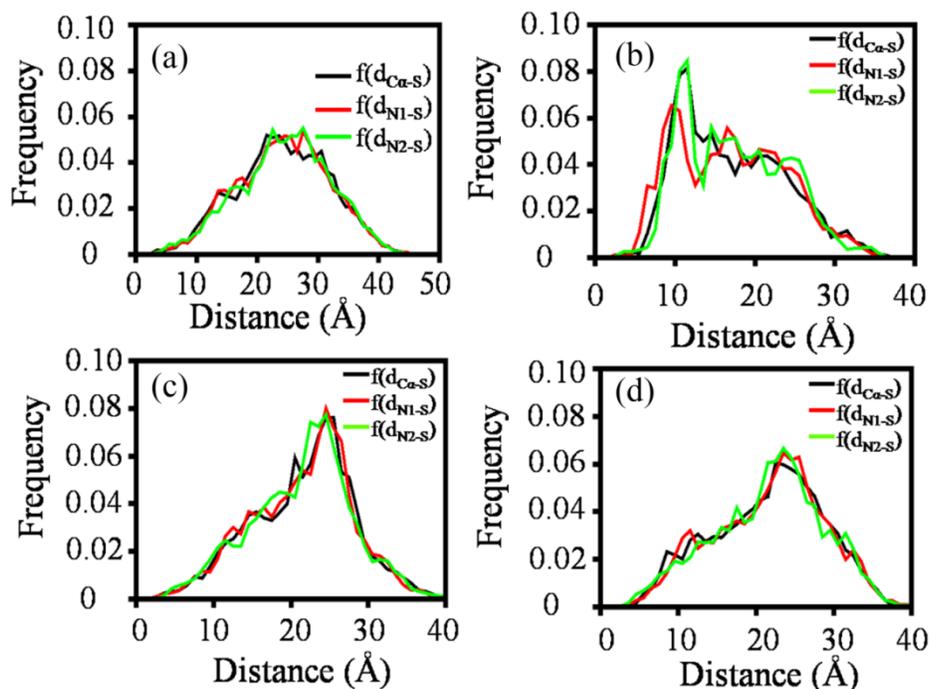


Figure 2.8: $f(d_{C^\alpha-S})$, $f(d_{N1-S})$ and $f(d_{N2-S})$ between C^α , N1 and N2 atoms of $C^\alpha NN$ motif and sulphur atom of sulphate in (a) SQT fragment, (b) SNQ fragment, (c) GIH fragment and (d) LYD fragment.

All these distributions indicate that sulphate ion prefers to stay at large distances from the $C^\alpha NN$ motif atoms. The time dependent fluctuations in the anion distances from motif atoms have also been examined for all these motif fragments. Figure 2.9 shows $d_{C^\alpha-S}$, d_{N1-S} and d_{N2-S}

as a function of time over the simulated trajectory in case of SQT, SNQ and LYD peptide fragments. It indicates that sulphate ion moves back and forth in the neighborhood of the motif ($\leq 5 \text{ \AA}$) to larger distances ($\geq 10 \text{ \AA}$).

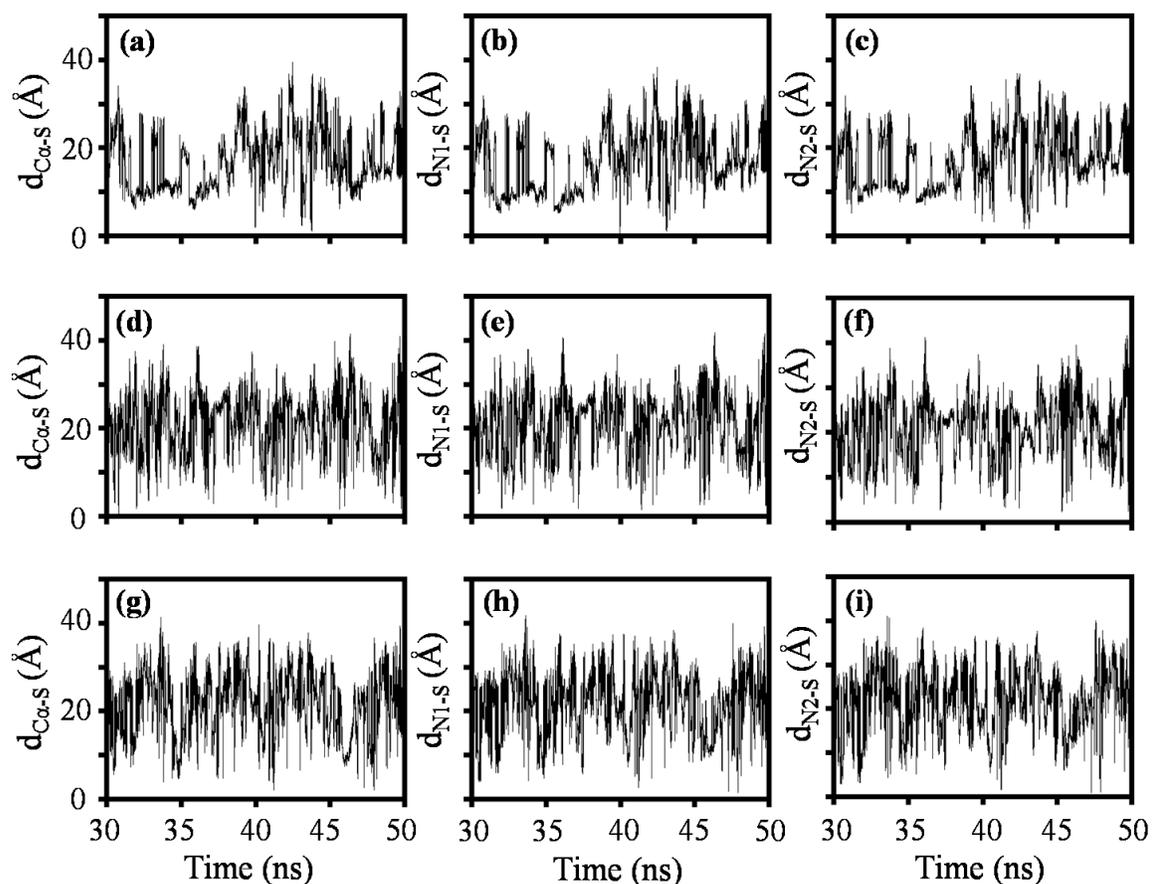


Figure 2.9: The distances ($d_{C\alpha-s}$), (d_{N1-s}), (d_{N2-s}) between C^α , N1 and N2 and sulphate as functions of time- (a),(b),(c) for SQT fragment, (d),(e),(f) for SNQ fragment and (g),(h),(i) for LYD fragment respectively.

Frequencies of distances C^α ($f(d_{C^\alpha-P})$), N1 ($f(d_{N1-P})$) and N2 ($f(d_{N2-P})$) from phosphorus atom of the biphosphate ion in case of biphosphate bound SRS fragment is shown in Figure 2.10(a). The peak of $f(d_{C^\alpha-P})$, $f(d_{N1-P})$ and $f(d_{N2-P})$ is around 6 \AA , 5 \AA and 4 \AA respectively. This indicates unlike sulphate, biphosphate ion stabilizes in the vicinity of the motif. However, $d_{C^\alpha-P}$ as a function of time over the simulated trajectory (Figure 2.10(b)) indicates that biphosphate ion

also moves back and forth in the neighborhood of the motif, while it prefers to stay mostly ≤ 10 Å from the motif.

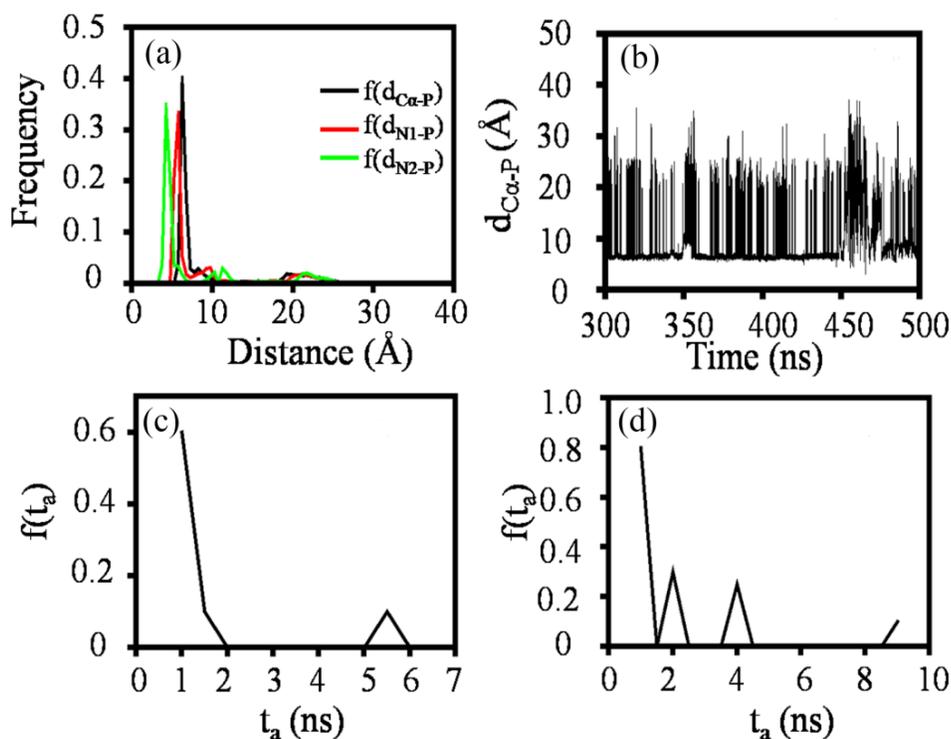


Figure 2.10:(a) The frequency distributions of distances; $f(d_{C\alpha-P})$, $f(d_{N1-P})$ and $f(d_{N2-P})$ between phosphorous atom of biphosphate and C^α , N1 and N2 atoms, respectively of $C^\alpha NN$ motif in SRS fragment. (b) $d_{C\alpha-P}$ as a function of time in SRS fragment. (c) The distribution $f(t_a)$ of time interval, t_a between two successive approaches by sulphate to the vicinity of C^α atom ($d_{C\alpha-S} \leq 5\text{Å}$) in SQT fragment. (d) The distribution $f(t_a)$ of time interval, t_a between two successive approaches by biphosphate to the vicinity of C^α atom ($d_{C\alpha-P} \leq 5\text{Å}$) in SRS fragment.

The distributions ($f(t_a)$) of time interval, t_a between any two successive approaches of sulphate and biphosphate to the vicinity of C^α atom (≤ 5 Å) are shown in (Figure 2.10(c) and (d)). The distribution is sharply peaked for low t_a with mean time $\langle t_a \rangle$ (~ 0.5 ns) which is much smaller compared to the time scale of secondary structural relaxation⁴⁹⁻⁵⁰. Thus, the anion induced structure cannot relax due to multiple approach of the anion to the vicinity of the motif in very short time scale.

2.3.4 Solvent structure around the motif:

The solvent structure around the motif is given by radial distribution of oxygen (O) atoms of solvent water molecules around C^α, N1 and N2 atoms of the motif in presence and absence of sulphate and biphosphate. For SQT fragment, a prominent first peak in distribution $g_{C\alpha}^{WS}(r)$ is observed around $r=4$ Å with respect to C^α atom in WS case (Figure 2.11(a)). In absence of sulphate the distribution $g_{C\alpha}^{WOS}(r)$ with respect to C^α atom is also similar as $g_{C\alpha}^{WS}(r)$ (Figure 2.11(b)). There is no shift in conformational preference for S residue corresponding to this atom. There is no peak in radial distributions $g_{N1}^{WS}(r)$ and $g_{N2}^{WS}(r)$ with respect to N1 and N2 atom respectively in presence of sulphate. In WOS case also there is no peak in distributions $g_{N1}^{WOS}(r)$ and $g_{N2}^{WOS}(r)$. However, conformational shift is observed in T containing N2 but no shift is observed for Q having N1. In case of SNQ fragment, first peak of $g_{C\alpha}^{WS}(r)$ is around $r = 4$ Å (Figure 2.11(c)). In Figure 2.11(d), there is also a peak in distribution of $g_{C\alpha}^{WOS}(r)$ around $r = 4$ Å. Conformational shift is not observed at this position for the motif residue S. There is a peak of $g_{N1}^{WS}(r)$ at $r = 3$ Å but there is no peak in distribution of $g_{N1}^{WOS}(r)$. Conformational shift is also not observed for N residue at this position. There are no peak in distributions of $g_{N2}^{WS}(r)$ and $g_{N2}^{WOS}(r)$. Here also no conformational shift takes place for Q residue. In case of SRS fragment, a prominent first peak in distribution $g_{C\alpha}^{WP}(r)$ is observed around $r=4$ Å with respect to C^α atom in WP case (Figure 2.11(e)). In WOP case the distribution $g_{C\alpha}^{WOP}(r)$ has peak around $r = 3$ Å (Figure 2.11(f)). Conformational shift is observed for the residue S at this position. There is no peak in distribution $g_{N1}^{WP}(r)$ and $g_{N2}^{WP}(r)$ with respect to N1 and N2 respectively in presence of biphosphate. In WOP case the distributions, $g_{N1}^{WOP}(r)$ and $g_{N2}^{WOP}(r)$ are peaked around $r = 3$ Å (Figure 2.11(f)). Conformational shift is not observed in case of R residue at N1 position, but observed for S residue at N2 position.

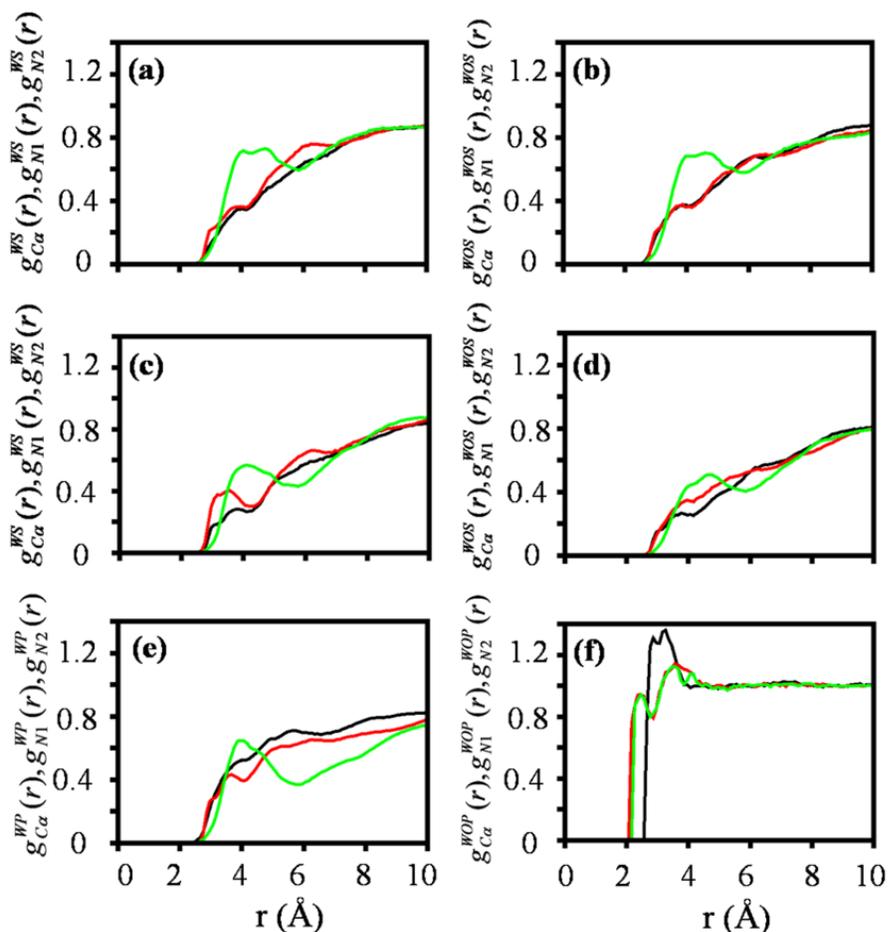


Figure 2.11: Radial distributions of the oxygen (O) atom of solvent water molecule around C^α (green), N1 (red) and N2 (black) atoms of the motif for SQT fragment (a) in presence ($g_{C^\alpha}^{WS}(r)$, $g_{N1}^{WS}(r)$, $g_{N2}^{WS}(r)$) and (b) in absence ($g_{C^\alpha}^{WOS}(r)$, $g_{N1}^{WOS}(r)$, $g_{N2}^{WOS}(r)$) of sulphate; (c) and (d) similar quantities for SNQ fragment; and (e) for SRS fragment in presence ($g_{C^\alpha}^{WP}(r)$, $g_{N1}^{WP}(r)$, $g_{N2}^{WP}(r)$) and (f) in absence ($g_{C^\alpha}^{WOP}(r)$, $g_{N1}^{WOP}(r)$, $g_{N2}^{WOP}(r)$) of biphosphate.

The distributions of water molecules in the presence and absence of the anion do not show difference in GIH and LYD fragments as shown in Figure 2.12. However conformational shift is observed for I with N1 in case of GIH fragment and for D with N2 in case of LYD fragment. Thus, changes in water distributions around the motif C^α , N1 and N2 atoms are not correlated to the changes in conformational preferences of the motif residues.

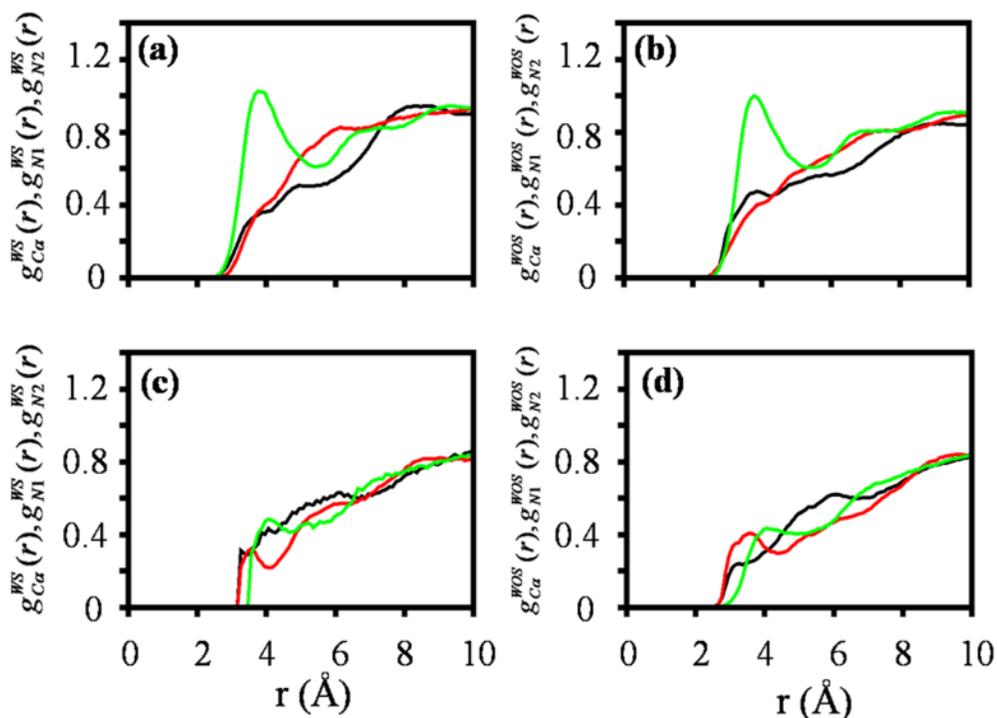


Figure 2.12: Radial distributions of the oxygen (O) atom of solvent water molecule around C^α (green), $N1$ (red) and $N2$ (black) atoms of the motif for GIH fragment in (a) presence ($g_{C^\alpha}^{WS}(r), g_{N1}^{WS}(r), g_{N2}^{WS}(r)$) and (b) absence ($g_{C^\alpha}^{WOS}(r), g_{N1}^{WOS}(r), g_{N2}^{WOS}(r)$) of sulphate; (c) and (d) similar quantities for LYD fragment.

2.3.5 Basis of conformational preferences of $C^\alpha NN$ motif:

The factors responsible for conformational preferences and anion specificity of different motif residues are detailed below.

Effect of sequence

The shifts in conformational preferences due to sulphate or biphosphate are sensitive to the sequence of the motif residues. The same residue shows different conformational preferences in WS condition depending on the sequence. For instance, the conformational preference of S in SQT fragment (RH) is quite in contrast to that of S in SNQ fragment (β). The polar S behaves distinctly in the presence of different anions. In WS condition S in SQT and SNQ fragments shows no shift in conformational preference with respect to the WOS

condition. But, in case of SRS fragment both S show biphosphate induced shifts in conformational preferences.

Effect of anion

The residue conformations of a given motif are sensitive to the type of anion. In order to probe this, sulphate is replaced by biphosphate in different fragments to see the effect of anions on conformational preference of C^αNN motif. The case of LYD fragment is illustrated in details. This fragment is chosen, for the fragment has hydrophobic, polar and charged residues. It is observed that secondary structural preference of the motif residues is different in the sulphate replaced by biphosphate (RP) case with respect to WS case (Table 2.3). Without any anion L132 fluctuates between RH (45%) and β (55%) conformations. In WS case L132 is primarily β conformation but it prefers RH conformation in RP case selecting the minor WOS conformation. Thus, L132 shows shifts in conformation from almost equally probable WOS conformations. However, the shift is dependent on the type of anion. Y133 primarily prefers RH conformation both in WS and WOS conditions. In RP case Y133 fluctuates between RH (48%) and β (52%) conformations (Table 2.3). Thus, Y133 lacks particular conformational preference in RP case, unlike that of WS. D134 prefers RH in WOS, β conformation in WS and RH conformation in RP condition, indicating shift in conformation in WS but no shift in RP.

The distance frequencies, $f(d_{C^{\alpha}-S})$, $f(d_{N1-S})$ and $f(d_{N2-S})$ in Figure 2.8(d) and $d_{C^{\alpha}-S}$ as a function of time in Figure 2.9(g) for LYD fragment in WS condition indicate that sulphate ion prefers to stay at large distances from the C^αNN motif atoms with back and forth motion in the neighborhood of the motif. When sulphate is replaced by biphosphate, the frequencies $f(d_{C^{\alpha}-P})$, $f(d_{N1-P})$ and $f(d_{N2-P})$ in Figure 2.13(a) and $d_{C^{\alpha}-P}$ as a function of time in Figure 2.13(b) show similar behavior of the biphosphate as the sulphate in WS condition. This behavior is in striking

contrast with that of biphosphate in the SRS fragment in WP condition (Figure 2.10(a)). This may be due to the enhanced flexibility experienced by Y133 due to replacement with biphosphate.

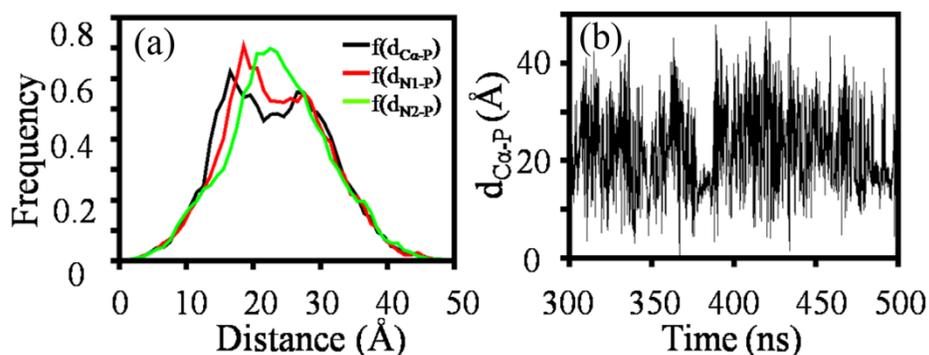


Figure 2.13: (a) The frequency distributions of distances; $f(d_{C^{\alpha}-P})$, $f(d_{N1-P})$ and $f(d_{N2-P})$ between phosphorus atom of biphosphate and C^α, N1 and N2 atoms, respectively of C^αNN motif in LYD fragment. (b) The distances ($d_{C^{\alpha}-P}$) between C^α and biphosphate as functions of time for LYD fragment.

The differences in fluctuations of the backbone dihedrals are shown in Figure 2.14 for LYD segment both in WS and replacement by biphosphate conditions. Distribution of backbone dihedral ϕ for the motif residues L, Y and D are defined as $H_L(\phi)$, $H_Y(\phi)$ and $H_D(\phi)$, and that of ψ are defined as $H_L(\psi)$, $H_Y(\psi)$ and $H_D(\psi)$, respectively. $H_L(\phi)$ shows single peak distribution both in WS and RP condition and not alter much due to replacement of sulphate by biphosphate (Figure 2.14(a)). $H_L(\psi)$ also shows single peak distribution both in WS and RP condition (Figure 2.14(b)). Depending on the anion, the peak positions are different: $\psi \sim 140^\circ$ in WS and $\psi \sim -20^\circ$ in RP condition. $H_Y(\phi)$ shows single broad peak in WS, but becomes bimodal distribution in RP condition (Figure 2.14(c)). $H_Y(\psi)$ shows single peak distribution both in WS and RP condition (Figure 2.14(d)). Here again the peak positions are different, $\psi \sim -10^\circ$ in WS and $\psi \sim 170^\circ$ in RP condition. The splitting of $H_Y(\phi)$ peak is consistent with fluctuations of conformation of Y133 in RP condition. $H_D(\phi)$ and $H_D(\psi)$ show single peak

distribution both in WS and RP condition (Figure 2.14(e) and (f)). $H_D(\varphi)$ peaks nearly coincide ϕ and $H_D(\psi)$ peaks are $\psi \sim 135^\circ$ in WS and $\psi \sim -10^\circ$ in RP condition.

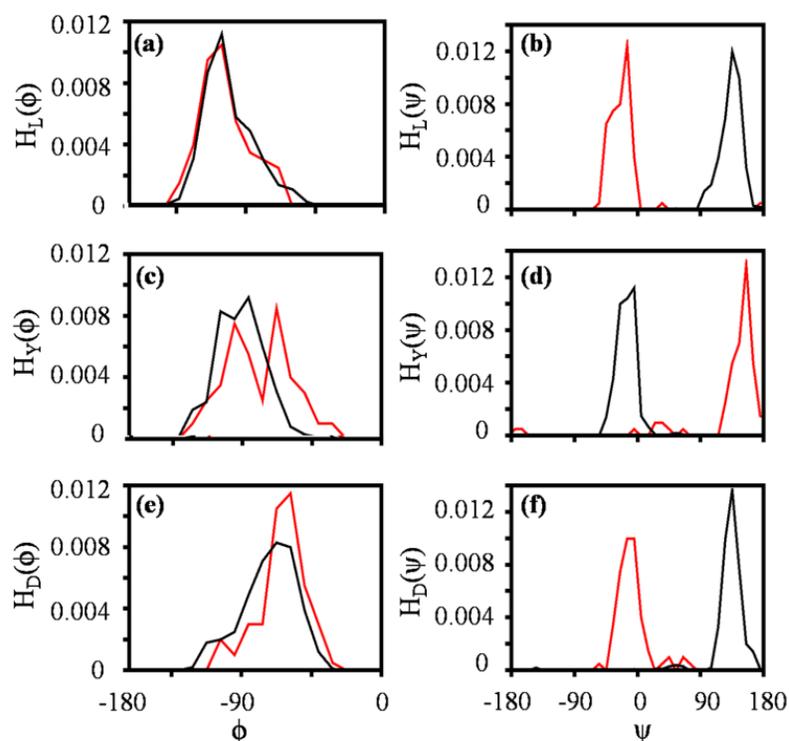


Figure 2.14: Distributions of backbone dihedral angle in WS (black) and RP (red) conditions for motif residues in LYD fragment: (a) $H_L(\varphi)$, (b) $H_L(\psi)$, (c) $H_Y(\varphi)$, (d) $H_Y(\psi)$, (e) $H_D(\varphi)$ and (f) $H_D(\psi)$

Similar sensitivities of conformational preferences depending on the type of anion have been observed for other fragments as well (Table 2.3). In SQT fragment S does not show any shift in conformation both in WS and RP cases compared to WOS. However, this residue shows substantial fluctuations in conformation in RP condition. Q shifts to RH conformation in WS, but lacks conformational preference in RP case. T shifts to β conformation in WS condition, but in RP condition there is no shift in conformation. In SNQ fragment S shifts to β conformation both in WS and RP cases as compared to WOS. N and Q do not show any shift in conformation both in WS and RP cases compared to WOS. In GIH fragment G shifts to RH conformation in WS, but lacks conformational preference in RP case. Isoleucine shifts to RH

conformation in WS condition, but in RP condition there is no shift in conformation. H does not show any shift in conformation in WS but lacks conformational preference in RP case.

Upon ligand binding the changes in fluctuations in dihedral angles cost conformational free energy and entropy⁴⁶. Earlier experimental and theoretical works show that the conformational thermodynamics govern the stability of ligand bound proteins. I study the conformational free energy and entropy changes of the motif residues due to replacement of sulphate by biphosphate from the backbone dihedral distributions of motif residues, $H_\alpha(\phi)$ and $H_\alpha(\psi)$ ⁴⁶⁻⁴⁸. Conformational thermodynamics data for the backbone dihedrals of the biphosphate bound motif residues with respect to the sulphate bound case for different peptide fragment are shown in Table 2.6. ΔG_i^{conf} and $T\Delta S_i^{conf}$ indicates the change in conformational free energy and entropy of the i-th residue, obtained by summing over the backbone dihedrals. $\Delta G_i^{conf}(\phi)$, $\Delta G_i^{conf}(\psi)$ and $T\Delta S_i^{conf}(\phi)$, $T\Delta S_i^{conf}(\psi)$ indicates the change in conformational free energy and entropy of the i-th motif residues due to backbone dihedral ϕ and ψ , respectively. The total change in conformational free energy (ΔG^{conf}) and entropy ($T\Delta S^{conf}$) of the motif are obtained by sum of the changes of individual residues of the motif. The negative change in conformational free energy and entropy indicates the conformational stability due to replacement by biphosphate as well as ordering, whereas the positive change indicates the instability and disorder due to biphosphate replacement.

Table 2.6 shows that the free energy changes are not significant in the residues in case of SQT. There is a large amount of disorder due to both of the backbone dihedrals of S. Overall, the motif shows slight increase in free energy and large amount of disorder by the anion replacement. In SNQ fragment free energy change is negative for S and positive for Q due to backbone dihedral ψ . However, backbone dihedral ψ introduce disorder in both S and Q. Due

to anion replacement this motif gets the conformational stability along with large amount of disorder. In GIH fragment G and I destabilized and disordered due to backbone dihedral ϕ and ψ . However, H is stabilized and ordered due to backbone dihedral ψ and ϕ , respectively. Overall large positive change in free energy and entropy indicates instability and enhanced disorder of this motif due to anion replacement. In LYD fragment all the motif residues destabilized due to backbone dihedral distribution, particularly ϕ (Figure 2.14). There is also disorder of L and D due to the backbone dihedral ψ and ϕ , respectively.

Table 2.6: Conformational thermodynamics change of motif residues of peptide fragment and motif in presence of biphosphate with respect to presence of sulphate

Peptide fragment	Motif residue	Change in conformational free energy			Change in conformational entropy			(a) Total change in conformational free energy (ΔG^{conf}) (b) Total change in conformational entropy ($T\Delta S^{conf}$)
		ΔG_i^{conf} in kJ/mol	$\Delta G_i^{conf}(\phi)$ in kJ/mol	$\Delta G_i^{conf}(\psi)$ in kJ/mol	$T\Delta S_i^{conf}$ In kJ/mol	$T\Delta S_i^{conf}(\phi)$ in kJ/mol	$T\Delta S_i^{conf}(\psi)$ in kJ/mol	
SQT	S134	0.46	0.10	0.36	2.76	1.55	1.21	(a) 0.48 (b) 3.45
	Q135	0.04	0.02	0.02	0.88	0.73	0.15	
	T136	-0.02	0.36	-0.38	-0.19	0.36	-0.55	
SNQ	S78	-2.09	1.41	-3.5	2.19	0.81	1.38	(a) -1.389 (b) 4.29
	N79	-0.002	-0.019	0.017	-0.57	-0.034	-0.536	
	Q80	0.703	0.08	0.623	2.67	0.64	3.31	
GIH	G148	3.41	3.41	0	3.03	0.84	2.19	(a) 13.38 (b) 5.01
	I149	14.6	0.71	13.89	3.19	1.26	1.93	
	H150	-4.63	8.54	-13.17	-1.21	-1.27	0.06	
LYD	L132	3.46	6.30	-2.84	1.51	0.38	1.13	(a) 13.98 (b) 1.98
	Y133	7.65	7.47	0.18	-1.63	1.06	-2.69	
	D134	2.87	2.73	0.14	2.10	2.73	-0.63	

Overall due to anion replacement in this case also this motif gets conformational instability and disorder. Overall the thermodynamic data indicate that the free energy costs are typically very large and positive due to the presence of biphosphate instead of sulphate in case of LYD and GIH fragment. This means that the anion replacement is not conformationally favourable

for these fragments, while SQT and SNQ are largely unaffected due to replacement. The anion replacement also introduces disorder in all the cases.

The shift of conformation in the presence of anion may be interpreted as the conformation without the anion is not favourable; and the residue needs to undergo conformational rearrangement to accommodate the anion. On the other hand, no shift in conformational preference indicates that the anion can be readily accommodated. This way one can identify the residues primarily responsible for anion recognition. In case of SNQ fragment where all the residues are polar, the major conformational state in WOS condition has been stabilized by the sulphate ion in WS condition. This implies that the WOS conformation of all three residues is favourable for sulphate binding. In case of SQT fragment the first two residues do not show shift in conformational preferences in WS condition compared to the WOS condition. Similarly, basic H in GIH fragment, hydrophobic L and polar Y in LYD and basic R in SRS do not show shifts in conformation. All these results point to the fact that the polar residues are favourable for sulphate recognition as found for the synthetic peptides^{28, 30-31}. Interestingly, the basic residues also adopt favourable conformation for anion recognition. The conformational preferences for majority of the motif residues shows shift in conformation in the presence of anions. This indicates that the anions can act as conformational switches in these residues. More importantly, the switching of the conformations can be controlled by the residue sequences and the type of anion for a given residue.

2.4 Conclusion:

Overall studies based on classical force field based MD simulations show that anion induced conformational preference in C^αNN motif residues changes are sequence specific. The motif residues fluctuate between different conformations without anion, one of which is stabilized by the anion. Although the sulphate ion is not stabilized in the vicinity of the motif atoms, the back and forth motion of the ion with very short time scale between successive approaches aids the structural changes. On the other hand, the biphosphate ion is stabilized in the vicinity of the motif residues, leading to their conformational preferences. For a given sequence, the conformational preferences of the residues are dependent on the nature of the anion as well. However, the preference of the conformation is not sensitive to the water distribution around the motif residues and the flanking residues. The residues showing no shift in conformational preferences by the anion are likely to be the ones responsible for anion recognition. These are mostly polar and basic residues. The microscopic method of identifying the anion recognizing residues may be useful to understand ligand recognition in general. The possibility of tuning conformational preferences by different anions may be useful for device applications based on conformational tunability of bio-macromolecules⁵¹⁻⁵².

References:

1. Pabo, C. O.; Sauer, R. T. Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **1992**, *61*, 1053-95.
2. Ledvina, P. S.; Tsai, A. L.; Wang, Z.; Koehl, E.; Quioco, F. A. Dominant role of local dipolar interactions in phosphate binding to a receptor cleft with an electronegative charge surface: equilibrium, kinetic, and crystallographic studies. *Protein Sci.* **1998**, *7* (12), 2550-9.
3. Kavanaugh, J. S.; Rogers, P. H.; Case, D. A.; Arnone, A. High-resolution X-ray study of deoxyhemoglobin Rothschild 37 beta Trp--Arg: a mutation that creates an intersubunit chloride-binding site. *Biochemistry* **1992**, *31* (16), 4111-21.
4. Verlinde, C. L.; Noble, M. E.; Kalk, K. H.; Groendijk, H.; Wierenga, R. K.; Hol, W. G. Anion binding at the active site of trypanosomal triosephosphate isomerase. Monohydrogen phosphate does not mimic sulphate. *Eur. J. Biochem.* **1991**, *198* (1), 53-7.
5. Harlos, K.; Vas, M.; Blake, C. F. Crystal structure of the binary complex of pig muscle phosphoglycerate kinase and its substrate 3-phospho-D-glycerate. *Proteins* **1992**, *12* (2), 133-44.
6. Abdullaev, Z.; Bodrova, M. E.; Chernyak, B. V.; Dolgikh, D. A.; Kluck, R. M.; Pereverzev, M. O.; Arseniev, A. S.; Efremov, R. G.; Kirpichnikov, M. P.; Mokhova, E. N.; Newmeyer, D. D.; Roder, H.; Skulachev, V. P. A cytochrome c mutant with high electron transfer and antioxidant activities but devoid of apoptogenic effect. *Biochem. J.* **2002**, *362* (Pt 3), 749-54.
7. Barrett, T. E.; Savva, R.; Panayotou, G.; Barlow, T.; Brown, T.; Jiricny, J.; Pearl, L. H. Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell* **1998**, *92* (1), 117-29.

Chapter 2

8. Cobessi, D.; Tete-Favier, F.; Marchal, S.; Azza, S.; Branlant, G.; Aubry, A. Apo and holo crystal structures of an NADP-dependent aldehyde dehydrogenase from *Streptococcus mutans*. *J. Mol. Biol.* **1999**, *290* (1), 161-73.
9. Martinez-Liarte, J. H.; Iriarte, A.; Martinez-Carrion, M. Inorganic phosphate binding and electrostatic effects in the active center of aspartate aminotransferase apoenzyme. *Biochemistry* **1992**, *31* (10), 2712-9.
10. Meiering, E. M.; Bycroft, M.; Fersht, A. R. Characterization of phosphate binding in the active site of barnase by site-directed mutagenesis and NMR. *Biochemistry* **1991**, *30* (47), 11348-56.
11. Denessiouk, K. A.; Johnson, M. S.; Denesyuk, A. I. Novel CalphaNN structural motif for protein recognition of phosphate ions. *J. Mol. Biol.* **2005**, *345* (3), 611-29.
12. Chakrabarti, P. Anion binding sites in protein structures. *J. Mol. Biol.* **1993**, *234* (2), 463-82.
13. Kobayashi, N.; Go, N. ATP binding proteins with different folds share a common ATP-binding structural motif. *Nat. Struct. Biol.* **1997**, *4* (1), 6-7.
14. Kinoshita, K.; Sadanami, K.; Kidera, A.; Go, N. Structural motif of phosphate-binding site common to various protein superfamilies: all-against-all structural comparison of protein-monomonucleotide complexes. *Protein Eng.* **1999**, *12* (1), 11-4.
15. Denessiouk, K. A.; Johnson, M. S. When fold is not important: a common structural framework for adenine and AMP binding in 12 unrelated protein families. *Proteins* **2000**, *38* (3), 310-26.
16. Denessiouk, K. A.; Rantanen, V. V.; Johnson, M. S. Adenine recognition: a motif present in ATP-, CoA-, NAD-, NADP-, and FAD-dependent proteins. *Proteins* **2001**, *44* (3),

282-91.

17. Watson, J. D.; Milner-White, E. J. A novel main-chain anion-binding site in proteins: the nest. A particular combination of phi,psi values in successive residues gives rise to anion-binding sites that occur commonly and are found often at functionally important regions. *J. Mol. Biol.* **2002**, *315* (2), 171-82.
18. Denesyuk, A. I.; Denessiouk, K. A.; Korpela, T.; Johnson, M. S. Functional attributes of the phosphate group binding cup of pyridoxal phosphate-dependent enzymes. *J. Mol. Biol.* **2002**, *316* (1), 155-72.
19. Denessiouk, K. A.; Johnson, M. S. "Acceptor-donor-acceptor" motifs recognize the Watson-Crick, Hoogsteen and Sugar "donor-acceptor-donor" edges of adenine and adenosine-containing ligands. *J. Mol. Biol.* **2003**, *333* (5), 1025-43.
20. Milner-White, E. J.; Nissink, J. W.; Allen, F. H.; Duddy, W. J. Recurring main-chain anion-binding motifs in short polypeptides: nests. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60* (Pt 11), 1935-42.
21. Torrance, G. M.; Leader, D. P.; Gilbert, D. R.; Milner-White, E. J. A novel main chain motif in proteins bridged by cationic groups: the niche. *J. Mol. Biol.* **2009**, *385* (4), 1076-86.
22. Wintjens, R.; Rooman, M. Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. *J. Mol. Biol.* **1996**, *262* (2), 294-313.
23. Struhl, K. Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. *Trends Biochem. Sci.* **1989**, *14* (4), 137-40.
24. Berg, J. M. Zinc fingers and other metal-binding domains. Elements for interactions between macromolecules. *J. Biol. Chem.* **1990**, *265* (12), 6513-6.
25. Holm, L.; Sander, C. The FSSP database: fold classification based on structure-

structure alignment of proteins. *Nucleic Acids Res.* **1996**, 24 (1), 206-9.

26. Albert L. Lehninger, M. C., David L. Nelson. *Principles of Biochemistry*. Palgrave Macmillan: **2004**.

27. Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **1963**, 7, 95-9.

28. Sheet, T.; Supakar, S.; Banerjee, R. Conformational preference of 'C^αNN' short peptide motif towards recognition of anions. *PLoS One* **2013**, 8 (3), e57366.

29. Sheet, T.; Banerjee, R. Sulfate ion interaction with 'anion recognition' short peptide motif at the N-terminus of an isolated helix: A conformational landscape. *J. Struct. Biol.* **2010**, 171 (3), 345-52.

30. Sheet, T.; Banerjee, R. The 'C^αNN' motif: an intrinsic lover of sulfate and phosphate ions. *Rsc Advances* **2016**, 6 (59), 54129-54141.

31. Sheet, T.; Ghosh, S.; Pal, D.; Banerjee, R. Computational design of model scaffold for anion recognition based on the 'C^αNN' motif. *Biopolymers* **2017**, 108 (1), 1-15.

32. Leach, A. R. *Molecular Modelling: Principles and Applications* 2nd ed.; Pearson Education: **2009**.

33. M.P. Allen, D. J. T. *Computer Simulation of Liquids*. Oxford University Press: **1991**; p 400.

34. Daan Frenkel, B. S. *Understanding Molecular Simulation: From Algorithms to Applications*. 2nd ed.; Academic Press: **2001**.

35. W, K. *Modern Methods of Optimization*. Springer: Heidelberg/Berlin, **1990**.

36. D.A, M. *Statistical Mechanics*. University Science Books: Sansalito, USA, **2000**.

37. Dormitzer, P. R.; Sun, Z. Y.; Wagner, G.; Harrison, S. C. The rhesus rotavirus VP4 sialic

acid binding domain has a galectin fold with a novel carbohydrate binding site. *EMBO J.* **2002**, *21* (5), 885-97.

38. Yao, M.; Ose, T.; Sugimoto, H.; Horiuchi, A.; Nakagawa, A.; Wakatsuki, S.; Yokoi, D.; Murakami, T.; Honma, M.; Tanaka, I. Crystal structure of 1-aminocyclopropane-1-carboxylate deaminase from *Hansenula saturnus*. *J. Biol. Chem.* **2000**, *275* (44), 34557-65.

39. Atwell, S.; Ultsch, M.; De Vos, A. M.; Wells, J. A. Structural plasticity in a remodeled protein-protein interface. *Science* **1997**, *278* (5340), 1125-8.

40. Symmons, M. F.; Jones, G. H.; Luisi, B. F. A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation. *Structure* **2000**, *8* (11), 1215-26.

41. Pieper, U.; Kapadia, G.; Mevarech, M.; Herzberg, O. Structural features of halophilicity derived from the crystal structure of dihydrofolate reductase from the Dead Sea halophilic archaeon, *Haloferax volcanii*. *Structure* **1998**, *6* (1), 75-88.

42. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435-47.

43. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26* (16), 1701-18.

44. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78* (8), 1950-8.

45. Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33-8, 27-8.

46. Das, A.; Chakrabarti, J.; Ghosh, M. Conformational contribution to thermodynamics of binding in protein-peptide complexes through microscopic simulation. *Biophys. J.* **2013**, *104* (6), 1274-84.
47. Das, A.; Chakrabarti, J.; Ghosh, M. Thermodynamics of interfacial changes in a protein-protein complex. *Mol. Biosyst.* **2014**, *10* (3), 437-45.
48. Sikdar, S.; Chakrabarti, J.; Ghosh, M. A microscopic insight from conformational thermodynamics to functional ligand binding in proteins. *Mol. Biosyst.* **2014**, *10* (12), 3280-9.
49. Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **2007**, *450* (7171), 913-6.
50. Charlier, C.; Khan, S. N.; Marquardsen, T.; Pelupessy, P.; Reiss, V.; Sakellariou, D.; Bodenhausen, G.; Engelke, F.; Ferrage, F. Nanosecond time scale motions in proteins revealed by high-resolution NMR relaxometry. *J. Am. Chem. Soc.* **2013**, *135* (49), 18665-72.
51. Lee, T. S.; Radak, B. K.; Harris, M. E.; York, D. M. A Two-Metal-Ion-Mediated Conformational Switching Pathway for HDV Ribozyme Activation. *Acs Catal.* **2016**, *6* (3), 1853-1869.
52. Knipe, P. C.; Thompson, S.; Hamilton, A. D. Ion-mediated conformational switches. *Chem. Sci.* **2015**, *6* (3), 1630-1639.

Appendices of Chapter 2

A2.1 Dihedral angle and secondary structure assignment of protein:

Secondary structure refers to regular, recurring arrangements of adjacent amino acids of a polypeptide chain in space considering steric constraints and many weak interactions. The protein main chain can be divided into repeating units, called the peptide units, each ranging from one C^α atom to the next C^α atom. As the peptide units are effectively rigid groups that are linked into a chain by covalent bonds at the C^α atoms. The only degrees of freedom they have are rotations around the C^α -C' and N- C^α bonds. By convention, the angle of rotations around the N- C^α bond is called phi (ϕ) and the angle of rotation around C^α -C' bond is called psi (ψ) (Figure A2.1). In this way each amino acid is associated with two conformational angles, ϕ and ψ . The third possible torsion angle within the protein backbone (called omega, ω) is essentially flat and fixed to 180 degrees. This is due to the partial double-bond character of the peptide bond, which restricts rotation around the C-N bond, placing two successive alpha-carbons and C, O, N and H between them in one plane²⁶.

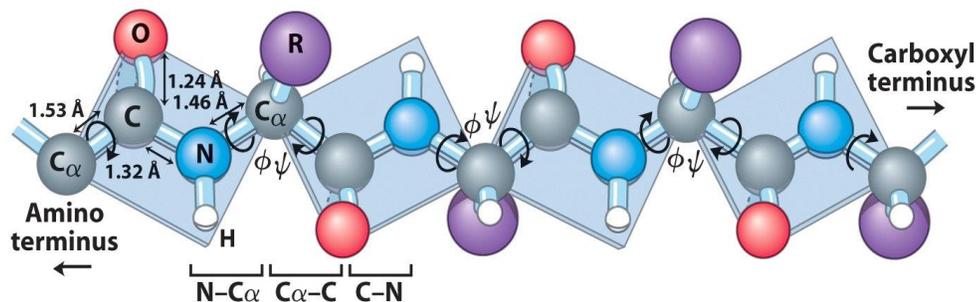


Figure A2.1: Schematic diagram of peptide backbone with side chains designated as R and showing the phi (ϕ) and psi (ψ) torsion angles. (Adopted from “Principles of Biochemistry”, 4th Edition, Lehninger²⁶)

Rotation of the protein chain can be described as rotation of the peptide bond planes relative to each other. As ϕ and ψ are the only degrees of freedom, conformation of whole main chain

of the polypeptide is completely determined when ϕ and ψ torsion angles for each amino acid are determined with high accuracy. These ϕ and ψ torsion angles of the polypeptide chain, also called Ramachandran angles. The ϕ - ψ correlation plot which is well known as Ramachandran plot (Figure A2.2) provides an easy way to view the distribution of torsion angles of a protein structure²⁶⁻²⁷. It considered atoms as a hard sphere with a definite radius (their van der Waals radius) and no two atoms overlap limits greatly the possible bond angles in a polypeptide chain. This constraint and other steric interactions severely restrict the variety of three dimensional (3D) arrangements of atoms (or conformations) that are possible. Ramachandran plot provides an overview of allowed and disallowed regions of backbone torsion angle values, serving as an important indicator of the quality of protein three dimensional structures²⁷.

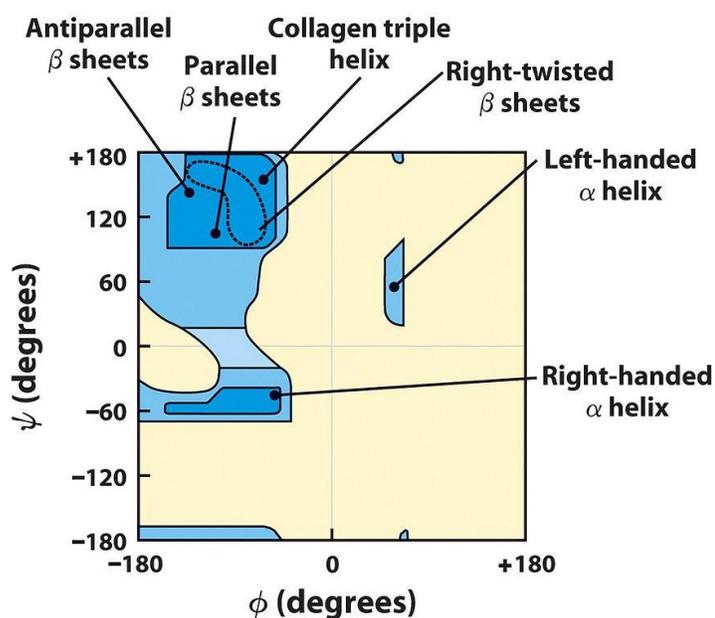


Figure A2.2: Ramachandran plot: The values of ϕ and ψ for various allowed secondary structures. The shaded regions correspond to conformations where there are no steric clashes (Adopted from “Principles of Biochemistry”, 4th Edition, Lehninger²⁶).

Chapter 2

Side chain dihedral angles are defined as the rotation around the bonds between following atoms $C^\alpha-C^\beta$ for χ_1 , $C^\beta-C^\gamma$ for χ_2 , $C^\gamma-C^\delta$ for χ_3 , $C^\delta-N^\epsilon$ for χ_4 and $N^\epsilon-C^Z$ for χ_5 . Dihedral angles are the most important local structural parameters that control protein folding essentially, and if we would have a way to predict the Ramachandran angles for a particular protein, we would be able to predict its 3D folding. When the 3D structures of many different protein molecules are compared, it becomes clear that, although the overall conformation of each protein is unique, two regular folding patterns are often found in parts of them. These two common folding patterns α -helix and β -sheet are the classic secondary structural elements of protein, result from hydrogen-bonding between the N–H and C=O groups in the polypeptide backbone, without involving the side chains of the amino acids²⁶. Thus, they can be formed by many different amino acid sequences. In each case, the protein chain adopts a regular, repeating conformation. Certain amino acids favour either α -helices or β -sheets, while others favour formation of loop regions. According to the Ramachandran plot, amino acids with backbone dihedral angles in the range $-180^\circ < \phi < 0^\circ$, $-100^\circ < \psi < 45^\circ$ are considered as to be in the right handed α -helical region (RH) and amino acids with backbone dihedral angles in the range $-180^\circ < \phi < -45^\circ$, $45^\circ < \psi < 180^\circ$ are considered as to be in the β -sheet region²⁷. ϕ angle confined to a narrow range around 50° while ψ angle within $+20^\circ$ to $+100^\circ$ is considered as left handed α -helical region (LH) and residues having dihedral angles in the range of $0^\circ < \phi < 180^\circ$, $-90^\circ < \psi < 90^\circ$ are defined as coil (C) conformation.

A2.2 Molecular Dynamics Simulation:

A2.2.1 Molecular Mechanics Force Fields:

Force fields permit to calculate the potential energy of a molecular system, V . The energy of molecular system depends on the atomic positions of all the atoms in the system which is usually expressed in term of Cartesian coordinates. Consideration of molecular mechanics force field helps to calculate the energy of a system only as a function of the nuclear positions. Molecular Mechanics are invariably used to perform all-atom calculations on systems containing a significant number of atoms. A typical force field represents each atom in the system as a single point and energies as a sum of two-, three-, and four-atom interactions such as bond stretching and angle bending. Although, simple functions (e.g. Hooke's Law) considering harmonic potential are used to describe these interactions, the force field can work quite well³². The potential energy of a certain interaction is described by an equation which involves the positions of the particles and some parameters (e.g force constants, charge) which have been determined experimentally or by quantum mechanical calculations. Several types of force fields exist. Two of those may use an identical functional form yet have very different parameters and thus bring about different energies for the same system. Moreover, force fields with the same functional form but different parameters, and force fields with different functional forms, may give close results³².

The value of the potential energy $V(r)$ is expressed as a sum of internal or bonded terms, which describe the bonds, angles and bond rotations in a molecule, and a sum of external or non-bonded terms, which account for interactions between non-bonded atoms or atoms separated by three or more covalent bonds. Therefore,

$$V(r) = V_{bonded}(r) + V_{non-bonded}(r) \quad (2.1)$$

Now considering all the bonded and non-bonded interactions total potential energy of a system can be expressed as³²,

$$\begin{aligned}
 V_{total}(r) = & \frac{1}{2} \sum_{bonds} k_l (l - l_0)^2 + \frac{1}{2} \sum_{angles} k_\theta (\theta - \theta_0)^2 + \frac{1}{2} \sum_{torsions} k_\phi [1 + \cos(n\phi - \delta)] + \frac{1}{2} \sum_{impropers} k_\epsilon \alpha^2 \\
 & + \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_{i,j} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]
 \end{aligned} \tag{2.2}$$

Here the first four terms are due to bonded interactions and last two terms are due to non-bonded interactions comprising electrostatic interactions and van der Waals interactions. Force field parameters in different molecular mechanics force field includes the equilibrium bond lengths and angles for bond stretching, angle bending and dihedral angle terms in the potential energy function as well as the atomic charge, force constants and the Lennard-Jones 12-6 potential parameters which are derived from the high-level quantum chemical calculations. The force field equation given above is only one variant of the many force fields. One can include other potential energy functions which contain more terms for a more precise calculation.

A2.2.2 Energy Minimization:

The MD simulations begin from energy minimized initial structure. Potential energy is a multi-dimensional function of the coordinate. For a system with N atoms the energy is thus a function of 3N-6 internal or 3N Cartesian coordinates. Variation of energy with the coordinates is usually referred to as the potential energy surface. Minimum energy arrangements of the atoms in the potential surface correspond to the stable states of the system; any movement away from a minimum gives a configuration with higher energy. There may be a very large number of minima on the energy surface. The minimum with the very lowest energy is known as the global energy minimum. To identify those geometrics of the system

that correspond to minimum points on the energy surface we use an energy minimization algorithm.

Let consider a multidimensional function $V(r_i)$ which depends on several variables $r_1, r_2, r_3, \dots, r_i$. The problem is to find the values of these variables where $V(r_i)$ gives minimum values. Minimum values are those points in an energy landscape where we will reach higher positions in any direction. So, the first derivative of the function with respect to every single variable is zero, i.e. $\frac{\partial V}{\partial r_i} = 0$ with r_i being the coordinates of the atoms; and the matrix $\frac{\partial V}{\partial r_i}$ is positive definite. Derivatives provide useful information about the shape of the energy landscape and so can significantly accelerate the search process. That is why methods which use derivatives are more appropriate to minimize the complex functions describing the energy landscape of a molecule³². The direction of the first derivative, the gradient, of a function tells us in which way we will find a minimum and its magnitude indicates the steepness of the slope at a given point of the function. The second derivative provides information about the curvature of the energy landscape.

To describe derivative methods, it is useful to write the energy function as a Taylor expansion around point r_k ³²⁻³⁴. In one dimension we can write

$$V(r) = V(r_k) + (r - r_k) \cdot V'(r_k) + \frac{1}{2} [(r - r_k)^2 \cdot V''(r_k)] + \dots \quad (2.3)$$

where V' is the first derivative and V'' is the second derivative of the energy function V .

In the case of a multidimensional function the variable r corresponds to a vector \vec{r} and the derivatives are replaced by matrices. For a system with N atoms $V(\vec{r})$ is a function of $3N$ coordinates³². So \vec{r} has $3N$ components and the gradient, $\vec{g} = V'(\vec{r})^T$, accordingly is a vector with $3N$ dimensions as well, with each element being the partial derivative of V with respect to a single coordinate, $\frac{\partial V}{\partial r_i}$. The second derivative $V''(\vec{r})$ is a $(3N \times 3N)$ -matrix. Every element

(i, j) corresponds to the partial second derivate of V with respect to the coordinates r_i and r_j ,

$\frac{\partial^2 V}{\partial r_i \partial r_j}$. This is a symmetric matrix which is called Hessian Matrix. Thus, the multidimensional

Taylor expansion is written as follows³⁵:

$$V(\vec{r}) = V(\vec{r}_k) + V'(\vec{r}_k) \cdot (\vec{r} - \vec{r}_k) + \frac{1}{2} [(\vec{r} - \vec{r}_k)^T \cdot V''(\vec{r}_k) \cdot (\vec{r} - \vec{r}_k)] + \dots \quad (2.4)$$

This is a quadratic function and thus can only be seen as an approximation for an energy function. However, the area close to a minimum is well approximated by this Taylor expansion.

First-order minimization methods use the information of the first derivative, the gradient \vec{g}_k , to find local minima of the function $V(r)$. Steepest Descent and Conjugate Gradient are the most important and most frequently used first-order minimization methods for energy minimization in molecular modeling and simulation.

Steepest Descent (SD)

In my simulation, steepest descent method is used for energy minimization. The steepest descent method moves along the negative gradient $-\vec{g}_k$ downhill the energy landscape beginning from a starting point of interest. As this is an iterative method the search is done stepwise:

- (i) First calculate the gradient \vec{g}_k at a starting point \vec{r}_k and take a step in the direction of the negative gradient $-\vec{g}_k$, the point of arrival being the starting point for the next iteration.
- (ii) Knowing the direction of the gradient the next thing to do is to determine the length of the step to take. This is achieved by performing a line search in the direction of the gradient³²⁻³⁴. If we consider a cross-section through the energy function in direction of \vec{g}_k we see that there will be a (one-dimensional) minimum, which is the optimal starting point for the next step, \vec{g}_{k+1} . If we take

the next step from this minimum point, the directions, \vec{v}_k , and the gradients, \vec{g}_k , of successive steps will always be orthogonal³²⁻³⁴,

$$\begin{aligned}\vec{g}_k \cdot \vec{g}_{k+1} &= 0 \\ \vec{v}_k \cdot \vec{v}_{k+1} &= 0\end{aligned}\tag{2.5}$$

Finding the minimum via line search can be achieved in several ways. One possibility, for example, is to perform a one-dimensional quadratic approximation in the direction of the gradient.

Conjugate Gradient (CG)

The SD approach encounters severe problems when used with functions other than quadratic functions especially when they have the shape of a narrow valley. In this case it gives undesirable behaviour³⁵ taking only very short steps and thus extremely increasing computation. Conjugate gradient method is another similar approach that can overcome this difficulty. In contrast to SD, in this method the gradients of successive steps are not orthogonal. Instead the directions at each point \vec{v}_k are orthogonal to gradients of the following steps \vec{g}_k provided that a line search has been performed:

$$\vec{v}_k \cdot \vec{g}_{k+1} = 0\tag{2.6}$$

For every direction \vec{v}_k the direction of the previous step \vec{v}_{k-1} is taken into account,

$$\vec{v}_k = -\vec{g}_k + \gamma_k \cdot \vec{v}_{k-1}\tag{2.7}$$

with γ being a scalar constant given by

$$\gamma_k = \frac{\vec{g}_k \cdot \vec{g}_k}{\vec{g}_{k-1} \cdot \vec{g}_{k-1}}\tag{2.8}$$

Thus, new directions are linear combinations of the current gradient \vec{g}_k and the previous direction \vec{v}_{k-1} . As there is no previous direction for the first step, the conjugate gradient methods starting direction is the same as for SD.

In molecular modeling and simulation in general we start minimization from experimentally obtained structures with very high energy. Steepest descent and the conjugate gradient methods are first applied as they reach lower energies very fast due to their low demands on computation. However, when reaching regions of lower energy, where slopes are less steep, other, more sophisticated methods such as Newton-Raphson can be applied to speed up the search for low energy points.

A2.2.3 Equation of motion:

In molecular dynamics successive configurations of the system are generated by integrating Newton's laws of motion. Time evolution of the positions and velocities of the particles in the system is derived based on Newton's equation of motion³²⁻³⁴:

$$F_i = m_i \cdot a = m_i \cdot \frac{dv_i}{dt} = m_i \cdot \frac{d^2r_i}{dt^2} \quad (2.9)$$

It describes the motion of a particle of mass m_i with coordinate r_i and F_i being the force on the particle. This is used to calculate the motion of a finite number of atoms or molecules, respectively, under the influence of a force field that describes the interactions inside the system with a potential energy function, $V(r)$, where r corresponds to the coordinates of all atoms in the system. The relationship of the potential energy function and applied force is given by

$$F(r_i) = -\nabla_i V(r), \quad (2.10)$$

with $F(r_i)$ being the force acting on a particle due to a potential, $V(r)$. Combining these two equations gives

$$\frac{dV(r)}{dr_i} = -m_i \cdot \frac{d^2r_i}{dt^2} \quad (2.11)$$

This relates the derivative of the potential energy to the changes of the atomic coordinates in time. As the potential energy is a complex multidimensional function this equation can only be solved numerically with some approximations.

With the acceleration being $a = -\frac{1}{m} \cdot \frac{dV}{dr}$ we can then calculate the changes of the system in time by just knowing (i) the potential energy $V(r)$, (ii) initial coordinates $r_{i,0}$ and (iii) an initial distribution of velocities, $v_{i,0}$. Thus, using this deterministic method we can predict the state of the system at any point of time in the future or the past.

The initial distribution of velocities is usually randomly chosen from a Gaussian or Maxwell-Boltzmann distribution³⁵⁻³⁶, which gives the probability of atom i having the velocity in the direction of r_i at the temperature T by:

$$p(v_{i,r}) = \left(\frac{m_i}{2\pi k_b T}\right)^{\frac{1}{2}} \cdot \exp\left(-\frac{1}{2} \frac{m_i v_{i,r}^2}{k_b T}\right) \quad (2.12)$$

Velocities are then corrected so that the overall momentum of the system equals a zero:

$$P = \sum_{n=1}^N m_i \vec{v}_i = 0 \quad (2.13)$$

Integration Algorithms

The solution of the equation of motion as mentioned above is a rather simple one which is only sufficiently good over a very short period of time, in which the velocities and accelerations can be regarded as constant. So, algorithms were introduced repeatedly performing small time steps, thus propagating the system's properties (positions, velocities and accelerations) in time. Time steps are typically chosen in the range of 1 fs³²⁻³⁴. It is necessary to use such a small time step, as many molecular processes occur in such small periods of time that they cannot be resolved with larger time steps. A time series of coordinate sets calculated

this way is referred to as a trajectory and a single coordinate set as a frame. All algorithms assume that system's properties (positions, velocities and accelerations) can be approximated by a Taylor series expansion³²⁻³⁴:

$$\begin{aligned}
 r(t + \delta t) &= r(t) + \delta t.v(t) + \frac{1}{2}\delta t^2.a(t) + \dots \\
 v(t + \delta t) &= v(t) + \delta t.a(t) + \frac{1}{2}\delta t^2.b(t) + \dots \\
 a(t + \delta t) &= a(t) + \delta t.b(t) + \frac{1}{2}\delta t^2.c(t) + \dots
 \end{aligned}
 \tag{2.14}$$

with r , v and a being the positions, the velocities and the accelerations of the system. The series expansion is usually truncated after the quadratic term.

Verlet Algorithm

The most widely used algorithm for integrating the equations of motion in MD simulations is the Verlet algorithm³²⁻³⁴. It can be derived by simply summing the Taylor expressions for the coordinates at the time $(t + \delta t)$ and $(t - \delta t)$:

$$\begin{aligned}
 r(t + \delta t) &= r(t) + \delta t.v(t) + \frac{1}{2}\delta t^2.a(t) + \dots \\
 r(t - \delta t) &= r(t) - \delta t.v(t) + \frac{1}{2}\delta t^2.b(t) - \dots \\
 \Rightarrow r(t + \delta t) &= 2r(t) - r(t - \delta t) + \delta t^2.a(t)
 \end{aligned}
 \tag{2.15}$$

Thus, it uses the position $r(t)$ and acceleration $a(t)$ at time t and the positions from the previous step $r(t - \delta t)$ to calculate new positions $r(t + \delta t)$. In this algorithm velocities are not explicitly calculated but can be obtained in several ways. One is to calculate mean velocities between the positions $r(t + \delta t)$ and $r(t - \delta t)$.

$$v(t) = \frac{1}{2\delta t} \cdot [r(t + \delta t) - r(t - \delta t)]
 \tag{2.16}$$

The advantages of this algorithm are that it is straight forward and has modest storage requirements, comprising only two sets of positions $r(t)$ and $r(t - \delta t)$, and the accelerations

$a(t)$. The disadvantage, however, is its moderate precision, because the positions are obtained by adding a small term $[\delta t^2 \cdot a(t)]$ to the difference of two much larger terms $[2r(t) - r(t - \delta t)]$. This results in rounding errors due to numerical limitations of the computer. Furthermore, this is not a self-starting algorithm. New positions $r(t + \delta t)$ are obtained from the current positions $r(t)$ and the positions at the previous step $r(t - \delta t)$. So, at $t = 0$ there are no positions for $r(t - \delta t)$ and therefore it is necessary to provide another way to calculate them. One way is to use the Taylor expansion truncated after the first term:

$$\begin{aligned} r(t + \delta t) &= r(t) + \delta t \cdot v(t) + \frac{1}{2} \delta t^2 \cdot a(t) + \dots \\ \Rightarrow a(-\delta t) &= r(0) - \delta t \cdot v(0) \end{aligned} \quad (2.17)$$

Leap-Frog Algorithm

There are several variations of the Verlet algorithm trying to avoid its disadvantages. One example is the leap-frog algorithm³²⁻³⁵. Leap-frog algorithm is used in my simulation. It uses the following equations:

$$\begin{aligned} v(t + \frac{1}{2} \delta t) &= v(t - \frac{1}{2} \delta t) + \delta t \cdot a(t) \\ r(t + \delta t) &= r(t) + \delta t \cdot v(t + \frac{1}{2} \delta t) \end{aligned} \quad (2.18)$$

where $a(t)$ is obtained using

$$a(t) = -\frac{1}{m} \cdot \frac{dV}{dr} \quad (2.19)$$

First, the velocities $v(t + \frac{1}{2} \delta t)$ are calculated from the velocities at $v(t - \frac{1}{2} \delta t)$ and the accelerations $a(t)$. Then the positions $r(t + \delta t)$ are deduced from the velocities just calculated and the positions at time t . In this way the velocities first ‘leap-frog’ over the positions and then the positions leap over the velocities. The leap-frog algorithm’s advantages

over the Verlet algorithm are the inclusion of the explicit velocities and the lack of the need to calculate the differences between large numbers.

An obvious disadvantage, however, is that the positions and velocities are not synchronized. This means it is not possible to calculate the contribution of the kinetic energy (from the velocities) and the potential energy (from the positions) to the total energy simultaneously.

A2.2.4 Periodic Box and Minimum Image Convention:

I use bulk simulation for my MD calculation. So, periodic box and minimum image convention is important for my calculation. For any size of the simulated system, the number of atoms N would be negligible as compared with the number of atoms contained in a macroscopic piece of matter. The ratio between the number of surface atoms and the total number of atoms would be very large, causing surface effects to be dominant. For this *Periodic Boundary Condition* (PBC) is used in which particles are enclosed in a box, and the box is replicated to infinity by rigid translation in all the three cartesian directions, completely filling the space³²⁻³³. The application of PBC allows us to simulate equilibrium system properties with a manageable number of atoms by eliminating surface effects. The basic idea behind the PBC is that if an atom moves in the original simulation box, all its images move in a concerted manner by the same amount and in the same fashion. The computational advantage of this method is that we need to keep track of the original image only as representative of all other images. As the simulation evolves, atoms can move through the boundary of the simulation cells. When this happens, an image atom from one of the neighboring cells enters to replace the lost particle. As a result of applying PBC the number of interacting pairs increases enormously. This is because of each particle in the simulation box not only interacts with other

particles in the box but also with their images. This problem can be handled by choosing a finite range potential within the criteria of *minimum image convention*³²⁻³³. The essence of the minimum image criteria is that it allows only the nearest neighbors of particle images to interact. Assuming potential range to be short, a *minimum image convention* is adopted that each atom interacts with the nearest atom or image in the periodic array. In the course of the simulation, if an atom leaves the basic simulation box, attention can be switched to the incoming image. In practice, the mechanism of doing so is to use the potential in a finite range such that the interaction of two distant particles at or beyond a finite length can be neglected. This maximum length must be equal to or less than the half of the box length used in the simulation. A cutoff distance R_c (half the box length) or potential cut off is defined when a particle is separated by a distance equal or larger than R_c , two particles do not interact with each other. This helps to avoid expensive force calculation³²⁻³³.

A2.2.5 Thermodynamic Ensembles & Equilibration:

Equilibration of the system is performed considering three main thermodynamics ensembles³⁶:

Microcanonical ensemble (NVE ensemble): In this ensemble there is no transfer of energy or matter between the system and surroundings, and the volume of the system remains fixed. In other words, this is a physical system where the total number of particles (N), the total volume (V), and the total energy (E) are constant.

Canonical ensemble (NVT ensemble): In the canonical ensemble, the volume of the system is fixed, and energy can transfer across the boundary between system and surroundings, but the matter cannot. The canonical ensemble applies to systems of any size considering system immersed in a heat bath at a temperature (T), where the heat bath is much larger than the system.

Chapter 2

As the system and surroundings are in thermal contact, the system will transfer heat to and from the surroundings until they are in thermal equilibrium. Therefore, unlike the NVE ensemble, the temperature of the canonical ensemble can be defined constant (T). This ensemble is abbreviated in terms of constant number of particles (N), total volume (V), and temperature (T).

Isothermal-Isobaric ensemble (NPT ensemble): In the isothermal-isobaric ensemble, energy can transfer across the boundary, but matter cannot. The volume of the system can change such that the initial pressure of the system matches the pressure exerted on the system by its surroundings. Similar to the canonical ensemble, isothermal-isobaric ensemble can be described as a system immersed in a heat bath at a temperature (T), where the heat bath is much larger than the system. This ensemble maintains a constant total number of particles (N), and constant temperature (T) and pressure (P), typically abbreviated NPT ensemble.

Equilibration of a system is often conducted in two phases. The first phase is conducted under an NVT ensemble, where the system is equilibrated at a desired temperature T keeping the N and V constant. In the second phase equilibration of pressure is conducted under an NPT ensemble. After the system is well equilibrated at the desired temperature and pressure production Molecular Dynamics is usually performed at NPT ensemble which most closely resembles the experimental conditions.

Temperature Control:

The temperature of the system is related to the time average of the kinetic energy which is given by:

$$\langle K \rangle_{N,V,T} = \frac{3}{2} N k_B T \quad (2.20)$$

The easiest way to scale the temperature is to rescale the velocities i.e. if the temperature at time t is $T(t)$ then the velocities are multiplied by a factor λ so that the associated change in temperature may be given as³²⁻³³:

$$\begin{aligned}\Delta T &= \frac{1}{2} \sum_{i=1}^N \frac{2}{3} \frac{m_i (\lambda v_i)^2}{Nk_B} - \frac{1}{2} \sum_{i=1}^N \frac{2}{3} \frac{m_i (v_i)^2}{Nk_B} \\ \Delta T &= (\lambda^2 - 1)T(t) \\ \lambda &= \sqrt{\frac{T_{new}}{T(t)}}\end{aligned}\tag{2.21}$$

Another alternative way to maintain the temperature is to couple the system to an external bath that is fixed at desired temperature. The bath acts as a source of thermal energy, removing or supplying heat from the system as appropriate. Here the velocities are scaled at each step in such a way that the rate of change of temperature is proportional to the difference in temperature between bath and system.

$$\frac{dT(t)}{dt} = \frac{1}{\tau} (T_{bath} - T(t))\tag{2.22}$$

τ is the coupling parameter whose magnitude defined the how closely the bath and system are coupled with each other³². So, the change in temperature between successive steps is

$$\Delta T = \frac{\delta t}{\tau} [T_{bath} - T(t)]\tag{2.23}$$

Thus, the scaling factor may be defined as

$$\lambda^2 = 1 + \frac{\delta t}{\tau} \left[\frac{T_{bath}}{T(t)} - 1 \right]\tag{2.24}$$

A large value of τ the coupling is weak, while a smaller value indicates strong coupling. As the value of the coupling constant reaches onetime step, the algorithm reduces to a simple velocity scaling method. Based on previous work a coupling constant of 0.4 ps is seemed to be

useful for a time step of 1 fs³².

Noose Hover Dynamics

In both NVT and NPT ensemble the temperature needs to be maintained. Noose dynamics is a method for performing constant temperature dynamics³²⁻³³. In this approach an extra thermal reservoir variable is inserted into the dynamical equations to control the temperature. This needs a time step of 0.5 fs with the Verlet method in order to approach within 3% of the target temperature. The smaller the time step, the closer it approaches the target temperature.

$$\dot{p}_i = f_i - \xi p_i$$

$$\dot{\xi} = \frac{\sum_{i\alpha} p_i^2 / m_i - Nk_B T}{W} = v_T^2 \left[\frac{\sum_{i\alpha} p_i^2 / m_i}{Nk_B T} - 1 \right] = v_T^2 \left[\frac{\tau}{T} - 1 \right] \quad (2.25)$$

Here ξ is the friction coefficient which is allowed to vary in time. W is the thermal inertia parameter which is replaced by v_T , a decay time for thermal fluctuations. N is the number of degrees of freedom. T stands for instantaneous ‘mechanical temperature’. If $\frac{\tau}{T} > 1$, then the system temperature is hot, thus the friction coefficient ξ will increase and vice versa.

Langevin Dynamics

A molecular system in the real world is unlikely to be present in vacuum. The interaction of molecules causes friction and the occasional high velocity collision will perturb the system. Langevin dynamics attempts to extend MD to allow for these effects. This mimics the viscous effect similar to the procedure of Noose Hover dynamics and control the temperature by introducing damping and random forces the system³²⁻³³.

Pressure Control

The pressure is calculated by the virial theorem of Claussius. It is defined as the expectation value of the sum of products of the coordinates of the particles and the forces acting on them³²⁻³³. This is written as

$$W = \sum_{i=1}^N x_i \frac{dp}{dx_i} \quad (2.26)$$

Here the term derivative term is the derivative of the momentum (p) along the x direction of the i^{th} particle. According to the theorem this product is equal to $-3Nk_B T$. For an ideal gas where the interaction occurs only between the gas particles and the container the virial is equal to $-3PV$. However, there are forces which occur between the particles in a real gas and thus the total virial of a real system is equal to the sum of an ideal part i.e. $-3PV$ and the interactions between particles³².

$$W = -3PV + \sum_{i=1}^N \sum_{j=i+1}^N r_{ij} \frac{dv(r_{ij})}{dr_{ij}} = -3Nk_B T \quad (2.27)$$

Here the term $\frac{dv(r_{ij})}{dr_{ij}}$ represents the force f_{ij} between particle i and j . We thus have the following

$$\text{expression for the pressure: } P = \frac{1}{V} \left[Nk_B T - \frac{1}{3} \sum_{i=1}^N \sum_{j=i+1}^N r_{ij} f_{ij} \right] \quad (2.28)$$

Constant pressure simulation requires periodic boundary conditions. Pressure is controlled by dynamically adjusting the size of the cell and rescaling all atomic coordinates during the simulation. In constant pressure method sometime an additional degree of freedom s is considered, corresponding to the volume of the cubic simulation cell, which adjusts itself to equalize the internal and applied pressures³². This degree of freedom effectively serves as a piston and is given a mass W having units of mass.length⁻⁴. This degree of freedom mimics the

action of a piston on a real system. This is associated with a kinetic energy with velocity V :

$$K_v = \frac{1}{2}W\dot{V}^2 \quad (2.29)$$

Thus, we have:

$$\frac{dP}{dt} = \frac{(P_d - P)}{t_p} \quad (2.30)$$

Here P_d is the desired pressure and t_p is the time constant for pressure fluctuations. At each step the volume of the cell is scaled by a factor of χ and the molecular centre of mass by a factor $\chi^{0.33}$. Thus, we have: $\chi = 1 - \beta \frac{\delta t}{t_p} (P_d - P)$ where, $\beta = 1/k_B T$. (2.31)

This method when combined with a method of temperature control creates trajectories in the NPT ensemble. For the NPT ensemble according to the Noose Hoover dynamics the following equations are for pressure control³²⁻³³.

$$\begin{aligned} \dot{s} &= \frac{p}{mV^{1/3}} \\ \dot{p} &= f - (\chi + \xi)p \\ \chi &= \frac{\dot{V}}{3V} \\ \dot{\chi} &= \frac{(P - P_d)V}{t_p^2 k_B T} \end{aligned} \quad (2.32)$$

Here ξ is given by the above equation, k_B is the Boltzmann constant and f is the frictional force. Here we have to specify the desired pressure P_d , time constant t_p , decay time ν_T and instantaneous temperature of the piston. In addition, the damping coefficients ξ and desired temperature of the molecule T are also specified for maintaining the constant temperature.

A2.3 Conformational Thermodynamics:

Dihedrals angles of protein can be considered as the conformational variables. Multidimensional histograms of the dihedral angle distributions can estimate the conformational free energy and entropy for the protein-protein and protein-ligand interactions. In biomolecules, the long-ranged dihedral correlations have been found to be negligible except for some short-ranged correlations among the side chain torsions⁴⁶. These observations practically illustrate the importance of completely reduced one dimensional histograms based on a single dihedral angle⁴⁶. The histograms can be treated as the probability of finding the system in a given conformation; they can be interpreted as given by the Boltzmann factors of the corresponding effective free energies, while the entropies are given by the Gibbs formula³⁶. For a system with conformational variable set $\{\xi_1, \xi_2, \dots, \xi_i, \dots, \xi_N\}$, the normalized probability distribution is given by

$$P(\{\xi_i\}) = \frac{1}{Z} \exp \left[-\frac{H(\{\xi_i\})}{k_B T} \right] \quad (2.33)$$

where k_B is the Boltzmann constant, T is the absolute temperature, $H(\{\xi_i\})$ is the Hamiltonian, and Z is the partition function of the system. The reduced probability distribution for a given conformational variable ξ can be obtained by integrating over the other variables in Eq. 2.33.

$$P(\xi_i = \xi) = \frac{1}{Z} \int \exp \left[\frac{-H(\{\xi_i\})}{k_B T} \right] \delta(\xi_i - \xi) \prod_i d\xi_i = \frac{1}{Z} \exp \left[\frac{-G(\xi)}{k_B T} \right] \quad (2.34)$$

It defines the effective free energy $G(\xi)$ or the potential of mean force associated with ξ . We can consider ξ as a function of θ_p and θ_l , where θ_p and θ_l are the dihedral angles for protein p and ligand l , respectively. Therefore, we can define following effective free energies from Eq. 2.34:

$$\begin{aligned}
 P_{p+l}^c(\xi) &= \frac{1}{Z} \exp\left[\frac{-G_{p+l}^c(\xi)}{k_B T}\right], \\
 P_p^f(\theta_p) &= \frac{1}{Z} \exp\left[\frac{-G_p^f(\theta_p)}{k_B T}\right], \\
 P_l^f(\theta_l) &= \frac{1}{Z} \exp\left[\frac{-G_l^f(\theta_l)}{k_B T}\right]
 \end{aligned} \tag{2.35}$$

where subscript $p+l$, p and l indicate quantities associated with the complex, the protein and the ligand, respectively, while the superscripts c and f denote the complex and free state, respectively.

Now the free energy change for ξ due to complexation is

$$\Delta G_{conf}(\xi) = -k_B T \ln \left[\frac{P_{p+l}^c(\xi)}{P_p^f(\theta_p) P_l^f(\theta_l)} \right] \tag{2.36}$$

If dihedral correlations are negligibly small, the conformational variables can be considered independent⁴⁶ and we can write

$$G_{p+l}^c(\xi) = G_p^c(\theta_p) + G_l^c(\theta_l) \text{ and } P_{p+l}^c(\xi) = P_p^c(\theta_p) \cdot P_l^c(\theta_l)$$

Therefore, we can write from Eq. 2.36

$$\Delta G_{conf}(\xi) = -k_B T \ln \left[\frac{P_p^c(\theta_p)}{P_p^f(\theta_p)} \right] - k_B T \ln \left[\frac{P_l^c(\theta_l)}{P_l^f(\theta_l)} \right] \tag{2.37}$$

The above expression implies that the thermodynamics is given separately in terms of the individual dihedrals. If we sum over all the dihedral angles of protein and peptide, we get the total conformational free energy change as follows⁴⁶:

$$\Delta G_{conf}^{tot} = -k_B T \sum_{\theta_p} \ln \frac{P_p^c(\theta_p)}{P_p^f(\theta_p)} - k_B T \sum_{\theta_l} \ln \frac{P_l^c(\theta_l)}{P_l^f(\theta_l)} = \Delta G_{conf}^{prot} + \Delta G_{conf}^{lig} \tag{2.38}$$

From the equilibrated molecular dynamics simulation trajectory we can generate the normalized probability distribution of protein dihedral θ which is given by the histograms $H_p^c(\theta_p)$ and $H_p^f(\theta_p)$ and that for a ligand dihedral by $H_l^c(\theta_l)$ and $H_l^f(\theta_l)$ in the bound and the free states, respectively. The peak of the histogram defines the equilibrium value of the relevant dihedrals. Then the equilibrium conformational free energy cost associated with any protein dihedral θ is

$$\Delta G_{conf}^{eq}(\theta) = -k_B T \ln \left[\frac{H_{p,max}^c(\theta)}{H_{p,max}^f(\theta)} \right] \quad (2.39)$$

where the subscript *max* denotes the maximum of histogram. The free energy contributions from the neighborhood of the maximum can be accounted for within a Quasi Harmonic expansion about the maximum at $\theta = \theta_0$ up to the quadratic term⁴⁶,

$$H(\theta) = H_{max} - \frac{1}{2} H''(\theta_0)(\theta - \theta_0)^2,$$

where $H''(\theta_0) = |C(\theta_0)|$ is the curvature near the maximum.

For multimodal histograms, free energies are computed by taking an average, weighted by the maximum values of the peaks⁴⁶.

The conformational entropy for a particular dihedral can be estimated directly using the Gibbs entropy formula, given for a dihedral ξ by

$$S_{conf}(\xi) = -k_B \sum_i H_i(\xi) \ln H_i(\xi) \quad (2.40)$$

where the sum is taken over the histogram bins i with a nonzero value of $H_i(\xi)$. Therefore, the conformational entropy change for the dihedral is

$$\Delta S_{conf}(\xi) = -k_B \left[\sum_i H_i^c(\xi) \ln H_i^c(\xi) - \sum_i H_i^f(\xi) \ln H_i^f(\xi) \right] \quad (2.41)$$

Chapter 2

In the Quasi Harmonic limit, the entropy associated with the histogram of any dihedral ξ can be expressed in terms of the entropy of a harmonic oscillator fitted to the peak⁴⁶. The frequency of the oscillator of mass μ and force constant k is given by $\omega = \sqrt{k/\mu} = \sqrt{C/\mu}$.

Therefore, entropy is given by

$$S_{conf}(\xi) = k_B \left(1 + \ln \frac{2\pi k_B T}{h\sqrt{C/\mu}} \right)$$

with h being the Planck's constant. If C_f and C_c are the curvatures near maxima for the dihedral-histogram in free and complexed states, respectively, then

$$\Delta S_{conf}(\xi) = \frac{1}{2} k_B \ln \left(\frac{C_f}{C_c} \right) \quad (2.42)$$

For multi-peak histograms, Quasi Harmonic entropies are obtained by weighted average over the peaks with finite curvature around the maxima⁴⁶.

The thermodynamics of conformational changes of a given residue are finally obtained by adding all the associated dihedral contributions. Total changes in ΔG_{conf} and ΔS_{conf} of protein or polypeptide are computed by adding the all residue contributions.



CHAPTER 3

QUANTUM CHEMICAL STUDIES ON ANION BINDING AND SPECIFICITY TO C^αNN MOTIF



3.1 Introduction:

All-atom molecular dynamics (MD) simulation studies on C^αNN motif containing peptide fragments of functional proteins as discussed in chapter 2 indicate that anion induced conformational preference of the motif residues depends on the sequence and the type of anion¹. Backbone atoms of the C^αNN motif which do not possess any distinct electrical character mainly interact with the anion. Therefore, the stability of the anion is likely to associate with the charge polarization of the backbone motif atoms due to anion. Previous simulation studies based on classical force field calculations cannot account the charge polarization effects. Quantum chemical (QC) calculations considering electrons of different atomic species in the system can capture the charge polarization effects quite well²⁻⁶. However, in large biomolecular system QC calculations are challenging due to the involvement of a large number of electrons. In such case electronic structure calculations by QC methods are usually done by truncating the system to only active parts⁴⁻⁵ and considered the solvent as a continuum²⁻⁶.

Here I report on QC calculations on anion (sulphate or phosphate) coordinated C^αNN motif fragments along with two flanking residues on both terminals, taken from the crystal structures of the same functional proteins as considered in Chapter 2. I consider all the fragments both with (denoted by WS and WP with sulphate and biphosphate, respectively) and without (denoted by WOS and WOP) anion. At physiological pH, the phosphate (PO₄³⁻) is mostly available in dibasic (HPO₄²⁻) form. So, I consider HPO₄²⁻ instead of PO₄³⁻ in my system as done in my previous MD simulation studies detailed in chapter 2¹. I also study the effect of anion replacement by replacing SO₄²⁻ with HPO₄²⁻ (SRP) and HPO₄²⁻ with SO₄²⁻ (PRS) in the corresponding motif fragment. The QC calculations are performed in vacuum, as earlier studies

show that metal ion coordination to protein fragments are well described by vacuum QC calculations⁵. Although the proteins structures are experimentally known, the polypeptide structures used in QC calculations are all optimized.

I show that the stability of C^αNN motif is governed by the coordination of motif atoms with oxygen atoms of anion. The sequence of the motif as well as the type of anion influences the coordination pattern and consequently, the stabilization energy. I compare the QC results with those from force field calculations reported in chapter 2. In this chapter, material & methods are discussed in section 3.2, results are in section 3.3, discussions are in section 3.4 and conclusion are in section 3.5.

3.2. Material & methods:

3.2.1 Systems for study:

Anion (sulphate or phosphate) coordinated C^αNN motif fragments along with two flanking residues on both terminals are taken from the crystal structures of different functional proteins as shown in details in Chapter 2: (1) SNQ fragment, (2) SQT fragment, (3) LYD fragment, (4) GIH fragment; and (5) SRS fragment. Here the first four fragments bind to sulphate and the last fragment is phosphate binding motif. The systems are shown in Table 3.1. Neutral caps are added, N-terminus by acetyl (-COCH₃) group and the C-terminus by N-methylamide (-NHCH₃) group to the flanking residues, as done earlier⁷⁻¹⁰. The missing hydrogen atoms are added using visual molecular dynamics (VMD) based on CHARMM27 topologies¹¹.

Table 3.1: Systems taken for studies (Motif residues are shown in bold)

System	PDB Ids	Fragment	Sequences
1	1F2D	SNQ	(77)Gln- Ser(C^α)-Asn(N)-Gln(N) -Thr(81)
2	1KQR	SQT	(133)Ala- Ser(C^α)-Gln(N)-Thr(N) -Gln(137)
3	1E3H	LYD	(131)Hse- Leu(C^α)-Tyr(N)-Asp(N) -Val(135)
4	1AXI	GIH	(147)Thr- Gly(C^α)-Ile(N)-Hse(N) -Ala(151)
5	1VDR	SRS	(64)Met- Ser(C^α)-Arg(N)-Ser(N) -Glu(68)

3.2.2 Quantum chemical calculation:

The peptide fragments are optimized in vacuum using density functional form of electronic ground state¹²⁻²¹, explained in Appendices A3.1 and A3.2, in the Vienna Ab initio Simulation Package (VASP) with plane wave (PW)^{14, 20} basis and projector augmented-wave (PAW)^{19, 21} potentials, detailed in Appendix A3.2. DFT-PW calculations have been used to describe systems, like biomolecules in vacuum²²⁻²³ and water clusters²⁴, where non-bonded interactions are better described. I use the plane wave basis with cutoff energy of 400 eV and periodic box size of $30 \times 30 \times 30 \text{ \AA}^3$ in the VASP^{17, 20-21, 25}. I considered PAW potential with PBE as exchange-correlational functional and a gamma centered κ mesh of $1 \times 1 \times 1$. I perform ionic relaxation for all the peptide fragment atoms followed by self-consistent calculations for relaxing the electronic degrees of freedom to determine the electronic ground state energy. Some benchmark calculations have been performed using the Gaussian 03 package, explained in Appendix A3.3¹⁸ considering B3LYP functional with 6-31G (2d, 2p) basis set which is normally used for biomolecular systems^{4, 26-28}.

3.2.3 Theoretical analysis:

Root-mean-square deviation (RMSD) calculation:

I calculate the RMSD between the crystal structure of motif fragment and optimized structure of motif fragment in presence and absence of anion by superposing the two structures using the Pymol²⁹ software and evaluating root-mean-square displacement between the atoms of the two motif fragments.

Calculation of hydrogen bond:

For studying the hydrogen bonding interaction, I consider motif atoms as a hydrogen bond donor and oxygen atoms of anion as a hydrogen bond acceptor. I use the criteria of donor (D) to acceptor (A) atom distance $< 3.5 \text{ \AA}$ and D-H-A angle $> 120^\circ$ for consideration of hydrogen bonding interaction.

Calculation of partial density of state (PDOS):

I calculate the density of states (DOS), representing the number of electronic states per unit of energy (in eV), in the optimized geometry. The total DOS is calculated from the energy levels of each individual state by projecting the DOS on all the orbitals. I then compute the different contribution of the different orbitals for each and individual atoms and calculate the partial density of state (PDOS) of the atom.

Calculation of stabilization energy:

I calculate the single point energy of the ion-electron system in the optimized structure of peptide fragments both in presence and absence of anion. The stabilization energy (SE), ΔE_{SE} is defined as the energy difference between the anion bound motif and the free components. I also calculate the energy difference by replacing the anion.

3.3 Results:

3.3.1 Optimized geometry of motif fragments:

The optimized structures of C^αNN motif fragments are analyzed both in absence and presence of anion. Oxygen atoms of the anion are numbered according to the crystal structure of the corresponding peptide fragments. The optimized geometries of SNQ motif fragment in WOS, WS and SRP conditions are shown in Figures 3.1(a)-(c). With respect to the crystal structure the RMSD of optimized structures of SNQ motif fragment in WOS (Figure 3.1(a)), WS (Figure 3.1(b)) and SRP (Figure 3.1(c)) is 0.585 Å, 0.310 Å and 1.237 Å, respectively. Here in WS condition C^α and N1 atoms of the motif forms hydrogen bonds with O3 and N2 atom of the motif forms hydrogen bond with O4 atom (Figure 3.1(b)). In SRP the hydrogen bond pattern is different; N1 and N2 atoms of the motif form hydrogen bonds with O3 atom of biphosphate and hydrogen bonding interaction between C^α atom of the motif and oxygen atom of anion is disrupted (Figure 3.1(c)). Similar patterns are observed in other motif fragments as shown in Figure 3.2.

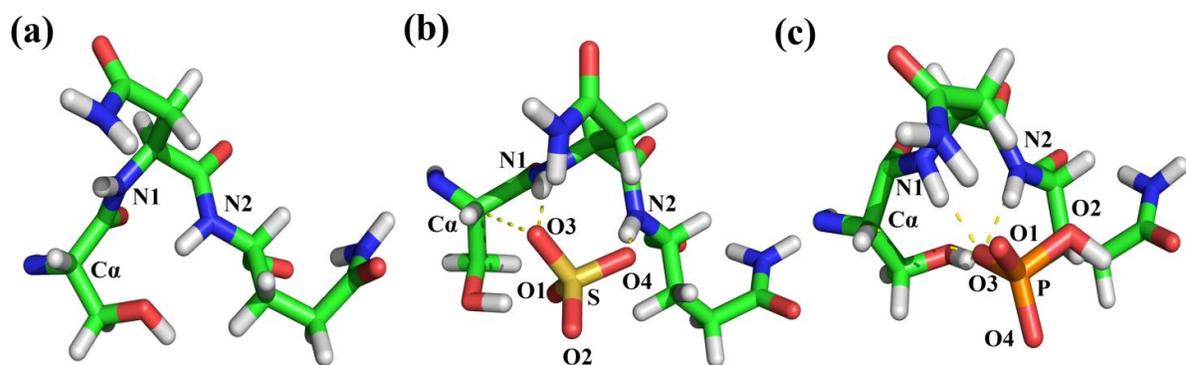


Figure 3.1: Optimized geometry of C^α NN motif residues in SNQ fragment: (a) in absence of sulphate (WOS), (b) in presence of sulphate (WS) and (c) in presence of biphosphate replacing the sulphate (SRP). The hydrogen bonds between motif atoms and oxygen atoms of anion are shown in yellow dotted line.

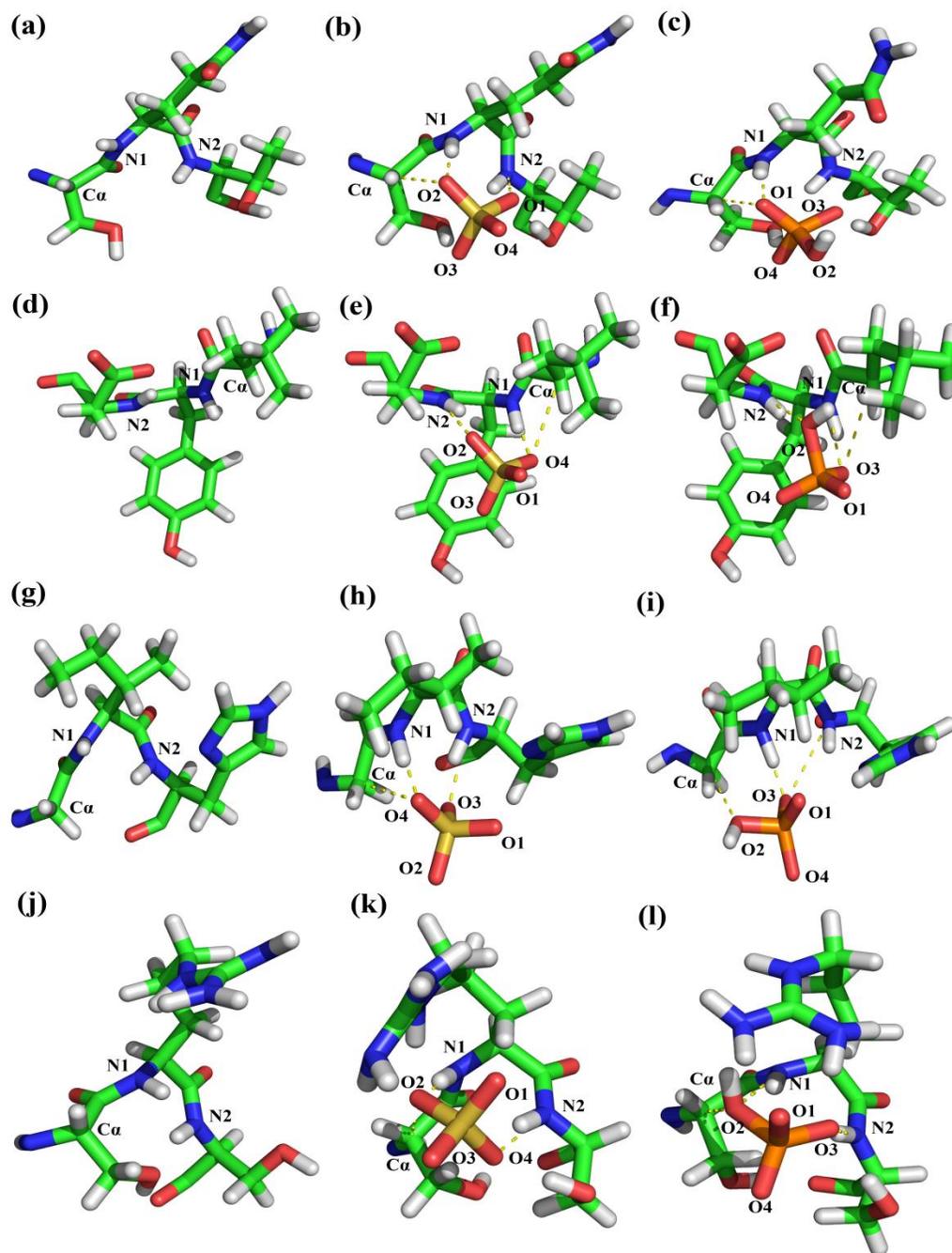


Figure 3.2: Optimized geometry of C $^{\alpha}$ NN motif residues in SQT fragment: (a) in WOS, (b) WS and (c) SRP condition; LYD fragment: (d) in WOS, (e) WS and (f) SRP condition; GIH fragment: (g) in WOS, (h) WS and (i) SRP condition; SRS fragment: (j) in WOP, (k) PRS and (l) WP condition. The hydrogen bonds between motif atoms and oxygen atoms of anion are shown in yellow dotted line.

In SRP - SQT fragment C^α and N1 atoms of the motif form hydrogen bonds with O1 atom of biphosphate, while hydrogen bond between N2 atom of the motif and oxygen atom of anion is disrupted (Figure 3.2). In SRP - LYD case C^α and N1 atoms of the motif form hydrogen bonds with O3 atom and N2 atom of the motif forms hydrogen bond with O2 atom of biphosphate (Figure 3.2). In SRP - GIH case C^α atom of the motif forms hydrogen bond with O2 atom and N1 and N2 atoms of the motif form hydrogen bond with O3 atom of biphosphate (Figure 3.2). In SRS fragment after replacement of biphosphate by sulphate (PRS - SRS) hydrogen bonding interaction between C^α atom of the motif and oxygen atom of anion is disrupted. N1 and N2 atom of the motif forms hydrogen bond with O2 and O4 atom of sulphate, respectively (Figure 3.2). Therefore, hydrogen bonding patterns between motif atoms and anion changes due to replacement of anion. However, in the optimized geometry of WS and WP fragments, most of the hydrogen bonds between anion and motif atoms are improved in terms of hydrogen bond distance and angle as depicted from Table 3.2. C^α and N1 atoms of the motif form hydrogen bonds with one oxygen atom of the anion concurrently and N2 atom of the motif forms hydrogen bond with another oxygen atom of anion.

3.3.2 Coordination of anion:

The PDOS contributions of the peptide and the capping atoms are calculated across the energy spectra. The energy values are given with respect to the Fermi energy. Representative cases are shown for SNQ fragment for WOS (Figure 3.3(a)), WS (Figure 3.3(b)) and SRP (Figure 3.3(c)) conditions. Similarly, data for SRS fragment in WOP, WP and PRS are shown in Figure 3.3(d), Figure 3.3(e) and Figure 3.3(f), respectively. PDOS of all capping atoms (black line) is found to be negligible compared to all peptide atoms (red line) in the entire energy range. Thus, QC results are not affected by the capping.

Chapter 3

Table 3.2: Distance of sulphate/biphosphate O atom from motif atoms and hydrogen bonding angle between motif atom and O atom of sulphate/biphosphate for the systems in WS/WP condition

Motif atom (X)	Initial crystal structure				Optimized Structure			
	X-O1 distance (Å) & X-H..O1 angle	X-O2 distance (Å) & X-H..O2 angle	X-O3 distance (Å) & X-H..O3 angle	X-O4 distance (Å) & X-H..O4 angle	X-O1 distance (Å) & X-H..O1 angle	X-O2 distance (Å) & X-H..O2 angle	X-O3 distance (Å) & X-H..O3 angle	X-O4 distance (Å) & X-H..O4 angle
System: 1F2D (SNQ fragment)								
C ^a	4.23 105.49°	5.59 132.00°	3.34 130.54°	4.92 106.70°	3.23 90.61°	5.11 122.27°	3.39 128.77°	4.74 101.77°
N1	4.04 134.20°	4.91 167.68°	2.75 165.05°	3.43 131.82°	3.21 105.18°	4.69 139.95°	2.68 164.30°	3.32 108.62°
N2	3.87 147.14°	4.95 170.88°	3.96 150.34°	2.64 153.02°	3.22 126.48°	4.78 165.75°	4.06 140.81°	2.66 165.85°
System: 1KQR (SQT fragment)								
C ^a	4.70 106.49°	3.61 136.24°	5.66 134.92°	3.97 115.42°	4.28 97.92°	3.15 130.45°	5.32 127.56°	3.80 105.05°
N1	3.42 122.84°	2.86 168.60°	4.93 158.24°	4.07 127.92°	3.07 109.98°	2.68 164.34°	4.69 149.71°	4.07 123.83°
N2	3.14 128.58°	4.44 134.60°	5.49 140.06°	4.55 170.62°	2.78 132.78°	4.24 150.02°	5.17 148.52°	4.24 165.98°
System: 1E3H (LYD fragment)								
C ^a	4.28 141.33°	3.96 112.50°	5.51 142.17°	3.44 146.04°	4.33 126.37°	3.89 98.94°	5.33 130.18°	3.12 135.60°
N1	4.73 134.0°	3.40 114.96°	4.80 152.02°	2.92 158.94°	4.59 142.21°	3.14 112.16°	4.43 148.80°	2.67 166.88°
N2	5.72 161.43°	3.40 159.90°	5.25 137.02°	4.60 131.43°	5.00 166.68°	2.62 153.90°	4.39 132.29°	4.03 144.44°
System: 1AXI (GIH fragment)								
C ^a	5.40 92.85°	4.64 92.44°	3.54 67.54°	3.39 106.94°	5.13 108.98°	4.25 104.59°	3.08 120.65°	3.45 78.35°
N1	4.46 149.85°	4.85 155.59°	3.28 124.10°	2.88 169.26°	4.34 138.20°	4.57 148.70°	3.02 110.15°	2.65 175.35°
N2	4.30 123.26°	5.40 156.84°	3.13 157.58°	4.17 140.03°	3.89 127.19°	4.95 163.87°	2.65 159.57°	3.81 141.00°
System: 1VDR (SRS fragment)								
C ^a	5.45 108.80°	3.42 120.55°	4.47 91.21°	3.64 82.58°	5.50 116.42°	3.08 112.76°	4.54 88.53°	4.53 95.29°
N1	5.00 142.83°	3.11 162.03°	3.15 112.85°	3.84 115.02°	4.88 165.56°	2.97 157.12°	3.36 138.67°	4.65 154.09°
N2	5.24 161.93°	4.72 134.17°	2.86 162.06°	4.20 126.63°	5.14 172.14°	4.13 150.04°	2.77 160.16°	4.33 148.50°

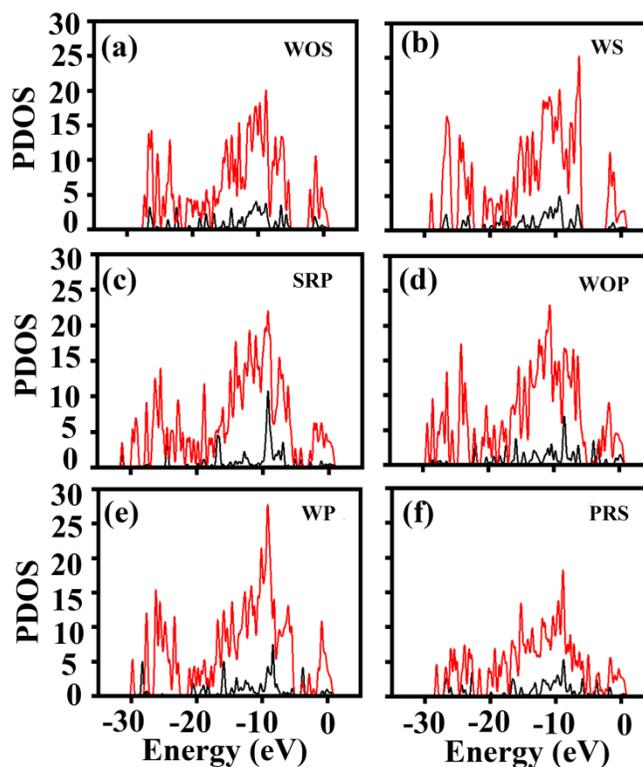


Figure 3.3: PDOS of all peptide atoms barring the capping atoms (red) and the capping atoms (black) for different energy for SNQ fragment : (a) in absence of sulphate (WOS), (b) in presence of sulphate (WS) and (c) in presence of biphosphate replacing the sulphate (SRP); SRS fragment: (d) in absence of biphosphate (WOP), (e) in presence of biphosphate (WP) and (f) in presence of sulphate replacing the biphosphate (PRS).

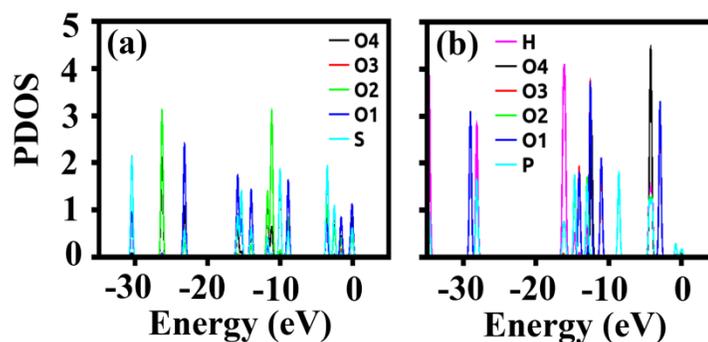


Figure 3.4: PDOS of (a) sulphate and (b) biphosphate for different energy.

The PDOS of different atoms of sulphate and biphosphate are shown in Figure 3.4(a) and (b), respectively. The PDOS peaks of oxygen atoms of anion are mainly distributed in the energy range -10 eV to -30 eV. PDOS of individual atoms of the motif residues S, N and Q in WS - SNQ fragment are shown in Figures 3.5(a)-(c), respectively as representative case. It shows that the major PDOS contributions of the atoms in motif residues come from the energy range -5 eV to -30 eV.

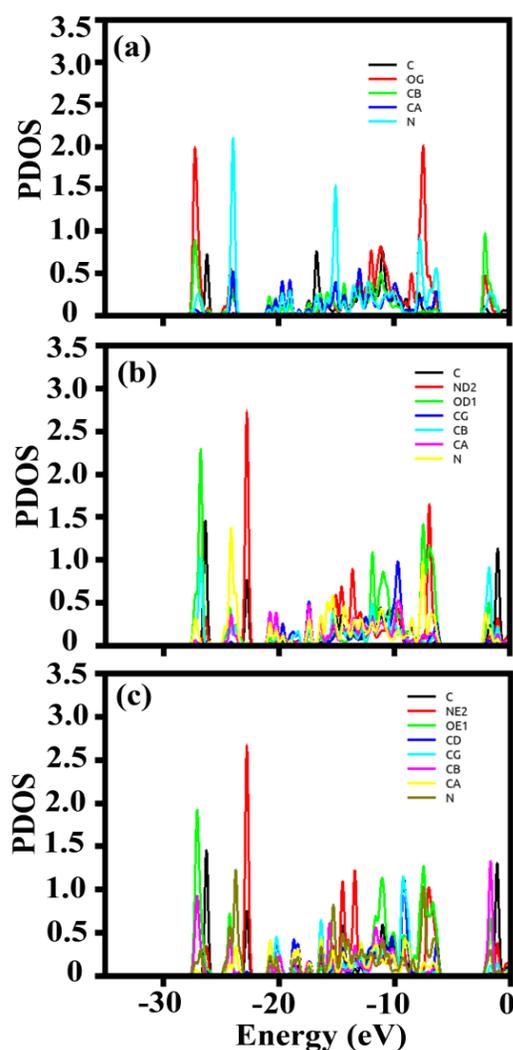


Figure 3.5: PDOS of individual non-hydrogen atoms of the motif residue (a) S78, (b) N79 and (c) Q80 for different energy (E) in case of SNQ fragment in WS condition.

The coordination is given in terms of simultaneous PDOS contributions of motif atoms of the peptide fragment and the atoms of the anionic species in WS and WP cases. If the peaks of PDOS of a motif atom and an atom of the anion lie within 0.04 eV (thermal energy at room temperature), the atom pairs are taken to coordinate at an energy, given by the mean of the peak values. Figures 3.6(a)-(c) show the PDOS of the motif atoms of SNQ fragment along with sulphur and oxygen atoms of sulphate. It shows simultaneous PDOS contribution of the motif atoms and oxygen atoms of the sulphate take place within the energy range -10 eV to -16 eV. The lowest energy coordination takes place between N2 atom of the motif and O4 atom of sulphate at $E = -15.30$ eV. Both C^α and N1 atoms of the motif coordinate with O3 atom of sulphate at a slightly higher energy (-14.27 eV). These oxygen atoms of anion are also found to form hydrogen bond with the motif atoms C^α , N1 and N2 in SNQ fragment. In case of SQT fragment PDOS of the motif atoms along with the sulphur and oxygen atoms of sulphate are shown in Figures 3.6(d)-(f). The lowest energy coordination takes place between C^α and N1 atom of the motif and O2 atom of sulphate both at $E = -11.70$ eV. N2 atoms of the motif coordinate with O1 atom of sulphate at slightly higher energy, $E = -11.04$ eV.

In LYD fragment N1 and C^α atoms of the motif coordinate with O4 atom of sulphate at lowest energies $E = -14.90$ eV and -14.85 eV, respectively (Figures 3.7(a)-(c)). At $E = -13.48$ eV, N2 atom of the motif coordinates with O2. N1 atoms of the motif also coordinate with O4 atom of sulphate at higher energy -13.44 eV and -11.28 eV. In GIH fragment lowest energy coordination takes place between N2 atom of the motif and O3 atom of sulphate at $E = -16.52$ (Figures 3.7(d)-(f)). At $E = -15.01$ eV, N2 atom of the motif also coordinates with O3 atom of sulphate. N1 atoms of the motif coordinate with O4 atom of sulphate at higher energy, $E = -10.06$ eV.

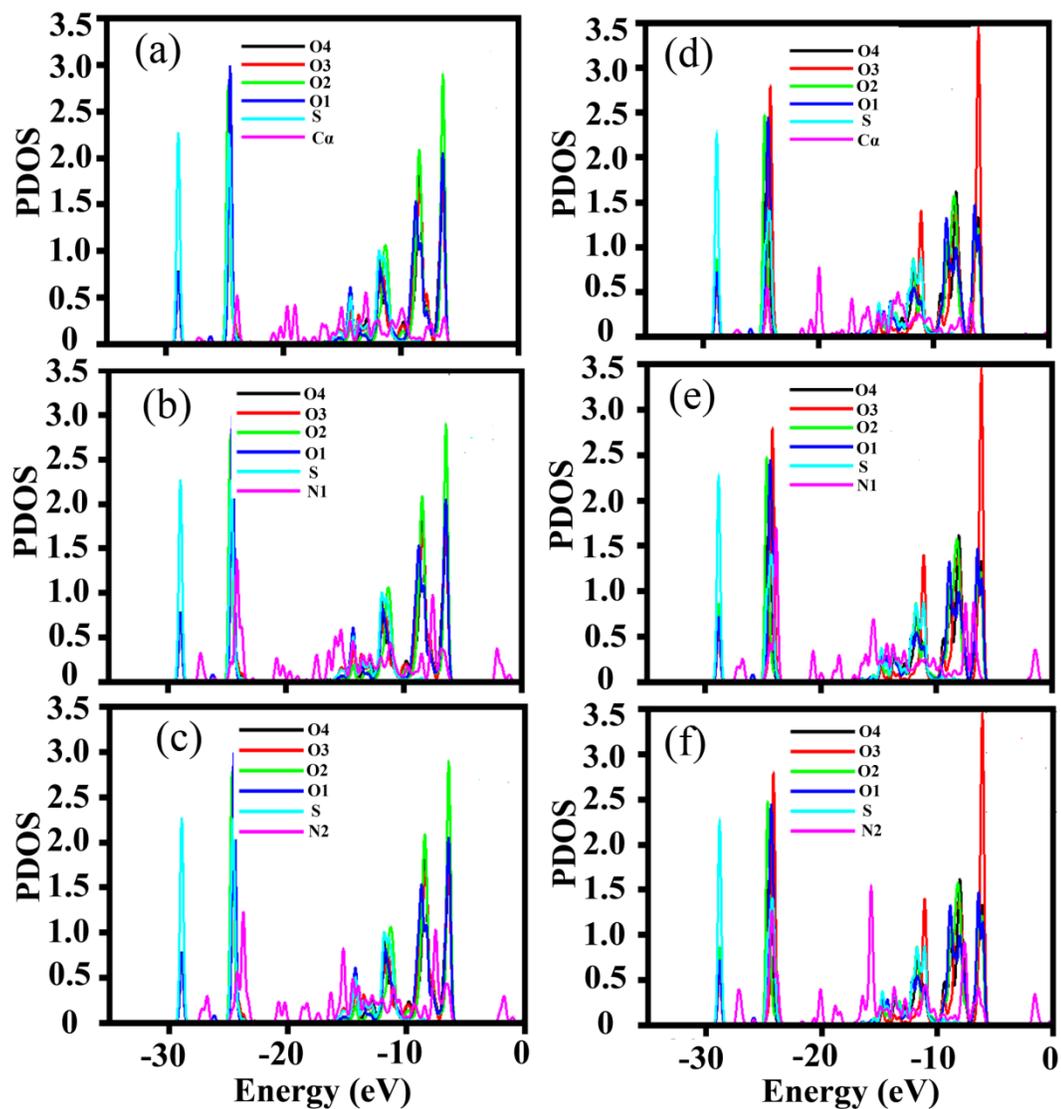


Figure 3.6: PDOS of S (cyan), O1 (blue), O2 (green), O3 (red) and O4 (black) atoms of bound sulphate and motif atoms of peptide (magenta): (a) C α of S (b) N1 of N and (c) N2 of Q in case of SNQ fragment and (d) C α of S (e) N1 of Q and (f) N2 of T in case of SQT fragment in WS condition as function of energy.

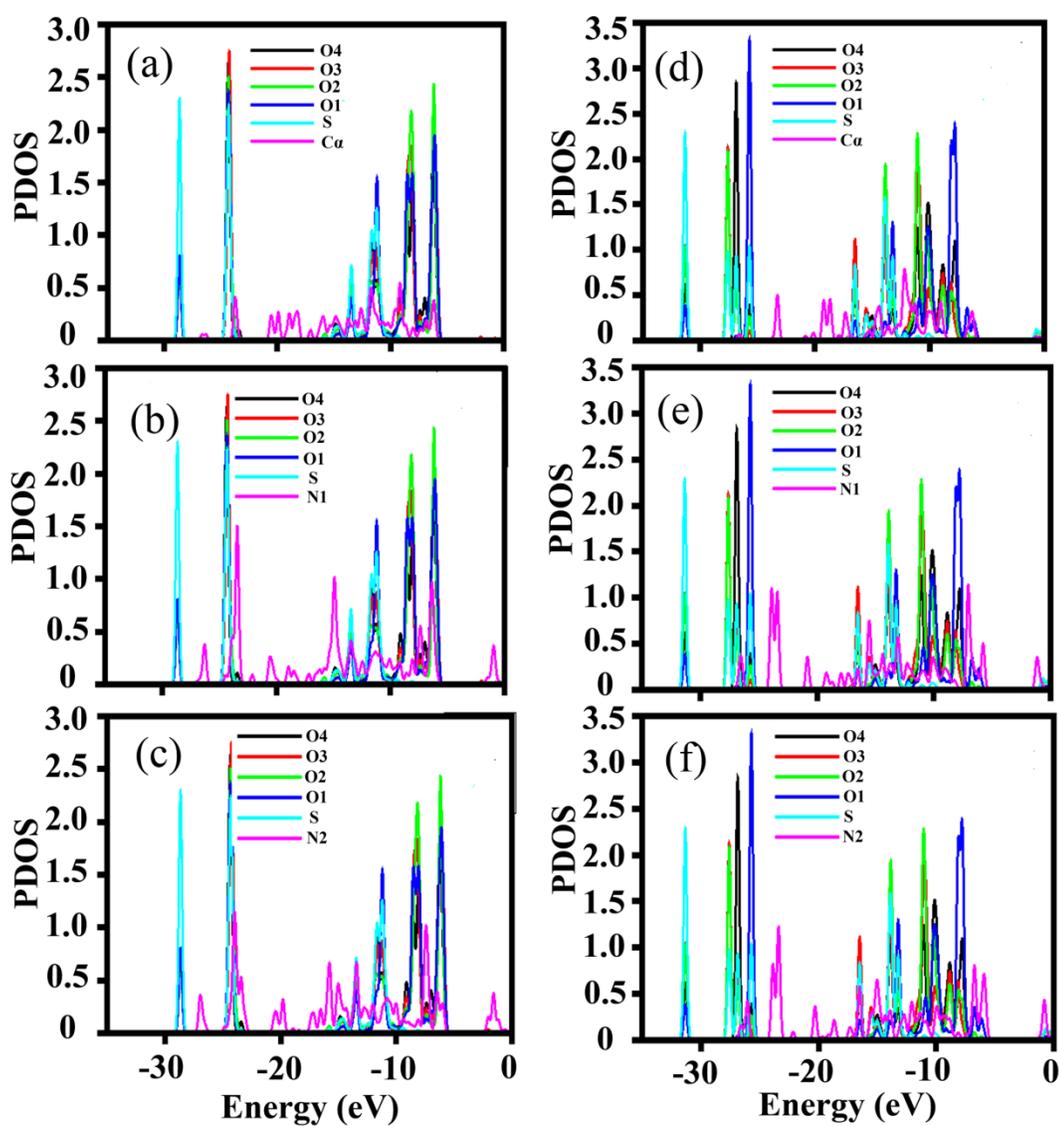


Figure 3.7: PDOS of S (cyan), O1 (blue), O2 (green), O3 (red) and O4 (black) atoms of bound sulphate and motif atoms of peptide (magenta): (a) C α of L (b) N1 of Y and (c) N2 of D in case of LYD fragment and (d) C α of G (e) N1 of I and (f) N2 of H in case of GIH fragment in WS condition as function of energy.

Figures 3.8(a)-(c) show the PDOS of the motif atoms of biphosphate bound SRS fragment along with the phosphorus and oxygen atoms of biphosphate. Here, the lowest energy coordination takes place between N2 atom of the motif and O3 atom of biphosphate at $E = -11.95$ eV. Both C^α and N1 atoms of the motif coordinate with O2 atom of biphosphate at a slightly higher energy (-11.12 eV).

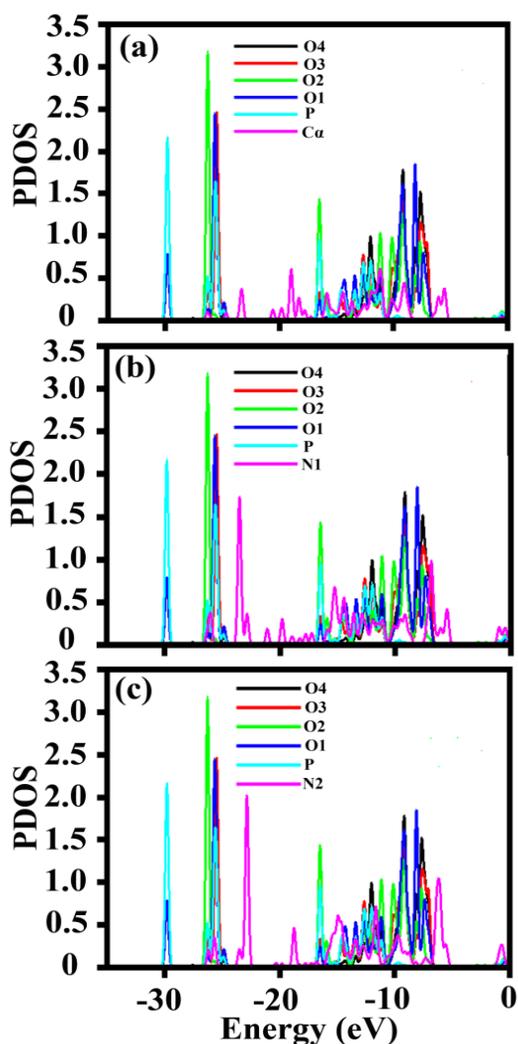


Figure 3.8: PDOS of P (cyan), O1 (blue), O2 (green), O3 (red) and O4 (black) atoms of bound biphosphate and motif atoms of peptide (magenta): (a) C^α of S (b) N1 of R and (c) N2 of S in case of SRS fragment in WP condition as function of energy.

Overall in case of SNQ, GIH and SRS fragments the lowest energy coordination takes place between oxygen atom of anion and N2 atom of the motif. For SQT and LYD fragments C^α and N1 atoms of motif coordinate with one oxygen atom of the anion simultaneously at lowest energy. In all these motif fragments, the coordinated oxygen atoms of anion are also found to be hydrogen bonded with the corresponding motif atoms (Table 3.2). I also study the simultaneous PDOS contributions of side chain atoms of the motif residues and the oxygen atoms of the anion. In case of S of SNQ fragment, other than C^α atom of motif residue side chain atoms C^β and O^γ coordinate with O1 atom at $E = -14.5$ eV (Figure 3.9(a)). Side chain $N^{\delta 2}$ atom of motif residue N also coordinate with O3 atom of sulphate at $E = -13.8$ eV (Figure 3.9(b)). Thus, the side chain atom of S in SNQ fragment participates in coordination with sulphate. However, side chain atom coordination in other motif sequences is not observed.

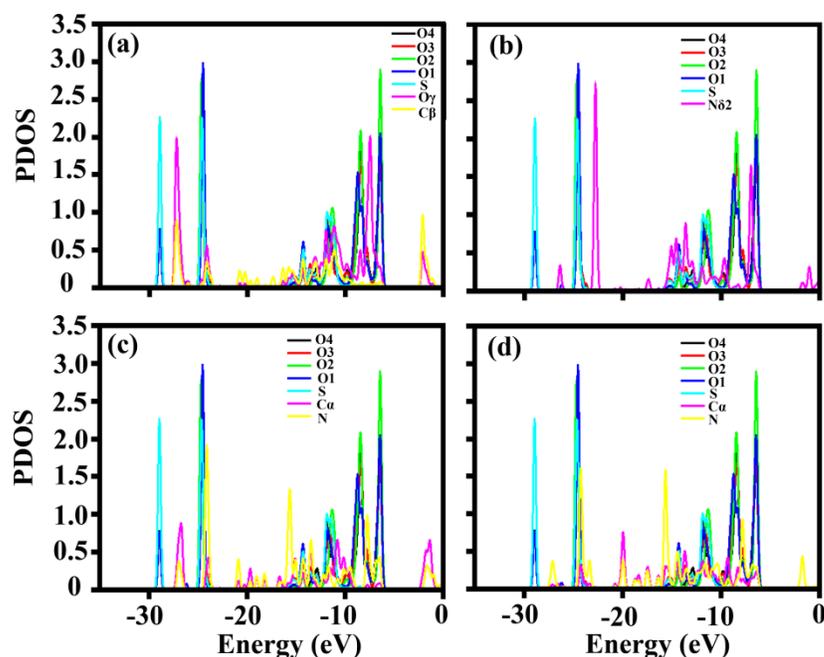


Figure 3.9: PDOS of S (cyan), O1 (blue), O2 (green), O3 (red) and O4 (black) atoms of bound sulphate and (a) side chain atom of S, (b) side chain atom of N, (c) C^α and N atoms of the flanking residue Q77 and (d) C^α and N atoms of the flanking residue T81 in case of SNQ fragment in WS condition as function of energy.

PDOS of individual atoms of terminal flanking residues Q77 and T81 in case of SNQ fragment along with atoms of sulphate are shown in Figures 3.9(c)-(d). Simultaneous PDOS contributions of oxygen atoms of sulphate and any of these atoms of flanking residue Q77 and T81 is not observed (Figures 3.9(c)-(d)). In the other cases also terminal flanking residues do not contribute simultaneously with the PDOS of oxygen atoms of anion in the ground state energy range. Therefore, the terminal flanking residues do not participate to coordinate with the oxygen atoms of anion.

3.3.3 Stabilization energy:

Stabilization energy ΔE_{SE} of the systems due to presence of anion is shown in Table 3.3. Table 3.3 indicates that in all the cases the stability of the system enhances in presence of anion. However, ΔE_{SE} is sequence dependent: SNQ>SQT>LYD>GIH. In case of SNQ and SQT motif fragments where stabilization energy is higher all the coordinating residues are polar. The additional stability of SNQ compared to SQT may be due to coordination of side chain atom of S as well. LYD fragment gives more stabilization energy as compared to GIH fragment. In LYD fragment L is non-polar, Y is polar and D is acidic in nature. In GIH fragment G and I is non-polar and H is also non-polar at physiological pH. Presence of polar residue in LYD fragment gives larger stabilization energy as compared to GIH fragment. In biphosphate bound SRS fragment stabilization energy is less than sulphate bound cases, though there are two polar and one basic residue.

Table 3.3: Stabilization energy of C^αNN motif containing fragments in presence of the anion

Fragment containing the C^αNN motif residues	Stabilization energy in kcal/mol $\Delta E_{SE} = E_{WS/WP} - (E_{WOS} + E_{SO4/EHPO4})$
SNQ fragment (WS)	-439.4
SQT fragment (WS)	-436.6
LYD fragment (WS)	-428.8
GIH fragment (WS)	-422.1
SRS fragment (WP)	-257.3

3.3.4 Replacement of anion:

I also study the coordination of the motif atoms with the oxygen atoms of anion by replacement of anion. The coordination scenario changes due to anion replacement. In case of SRP - SNQ fragment PDOS of N atoms of the motif are found to appear simultaneously with the PDOS of O3 atom of biphosphate (Figure 3.10(a)-(c)). The lowest energy coordination takes place between N1 atom of the motif and O3 atom of biphosphate at $E = -7.526$ eV. N2 atom of the motif coordinate with O3 atom of biphosphate at higher energy, $E = -6.982$ eV (Figure 3.10(b)-(c)). On the other hand, in presence of sulphate (WS – SNQ) motif atoms coordinate with oxygen atoms of sulphate at the energy range -16 eV to -14 eV. Thus, in contrast to the sulphate bound case, here the coordination takes place at higher energy. I find similar trends in all other cases. In SRP - SQT fragment energetically the deepest coordination takes place between N1 atom of the motif and O1 atom of biphosphate at $E = -7.142$ eV. N2 atom of the motif coordinate with O3 atom of biphosphate at higher energy, $E = -6.738$ eV (Figures 3.10(d)-(f)).

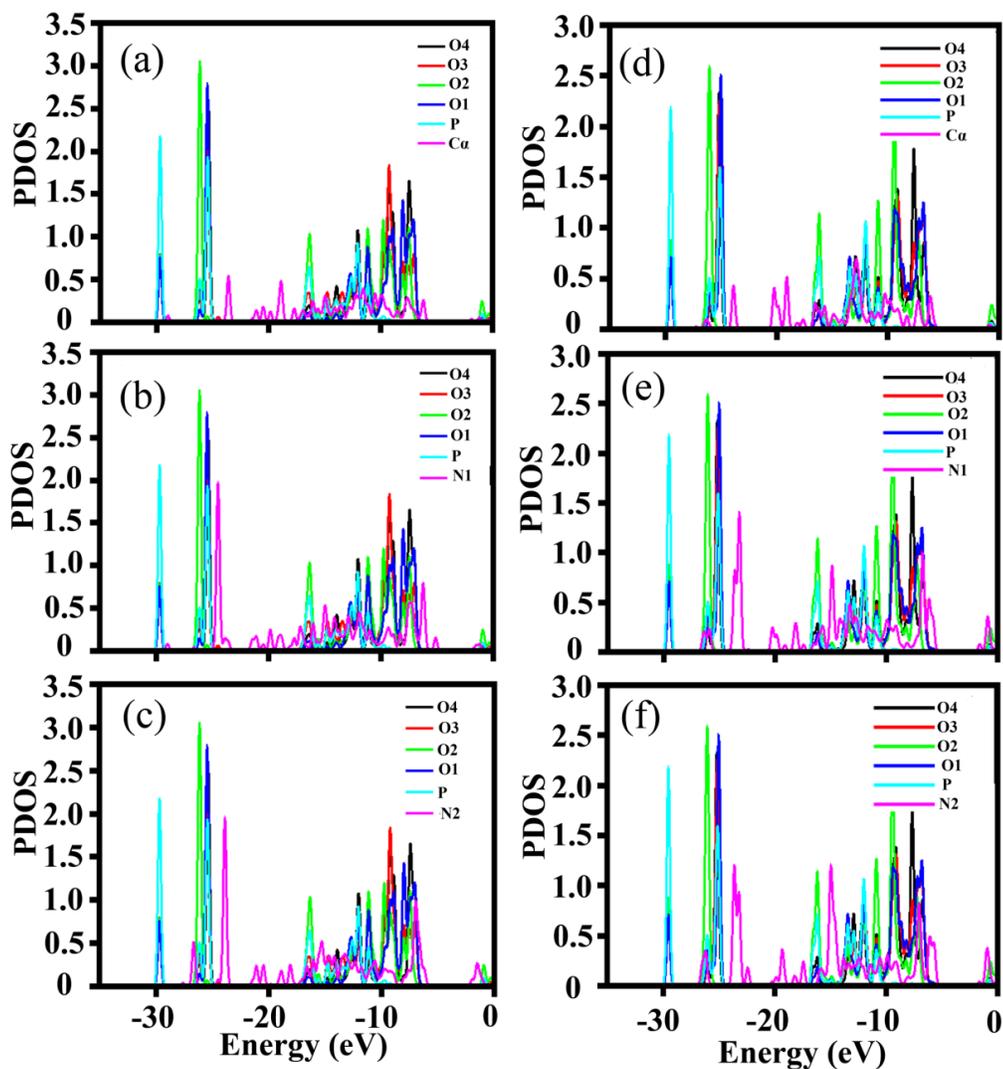


Figure 3.10: PDOS of P (cyan), O1 (blue), O2 (green), O3 (red) and O4 (black) atoms of bound biphosphate and motif atoms of peptide (magenta): (a) C^α of S (b) N1 of N and (c) N2 of Q for SNQ fragment in different energy and (d) C^α of S (e) N1 of Q and (f) N2 of T for different energy in case of SQT fragment in SRP case.

In SRP - LYD fragment the lowest energy coordination takes place between N1 atom of the motif and O3 atom of biphosphate at $E = -7.939$ eV. N2 atom of the motif coordinates with O2 atom of biphosphate at slightly higher energy (-7.883 eV) (Figures 3.11(a)-(c)). The lowest energy coordination takes place between N1 atom of the motif and O3 atom of biphosphate at $E = -11.123$ eV in case of SRP - GIH fragment (Figures 3.11(d)-(f)). In case of PRS - SRS

fragment lowest energy coordination takes place between N2 atom of the motif and O4 atom of sulphate at $E = -8.427$ eV. N1 atom of the motif coordinates with O2 atom of sulphate at higher energy (-7.603 eV) (Figures 3.12(a)-(c)).

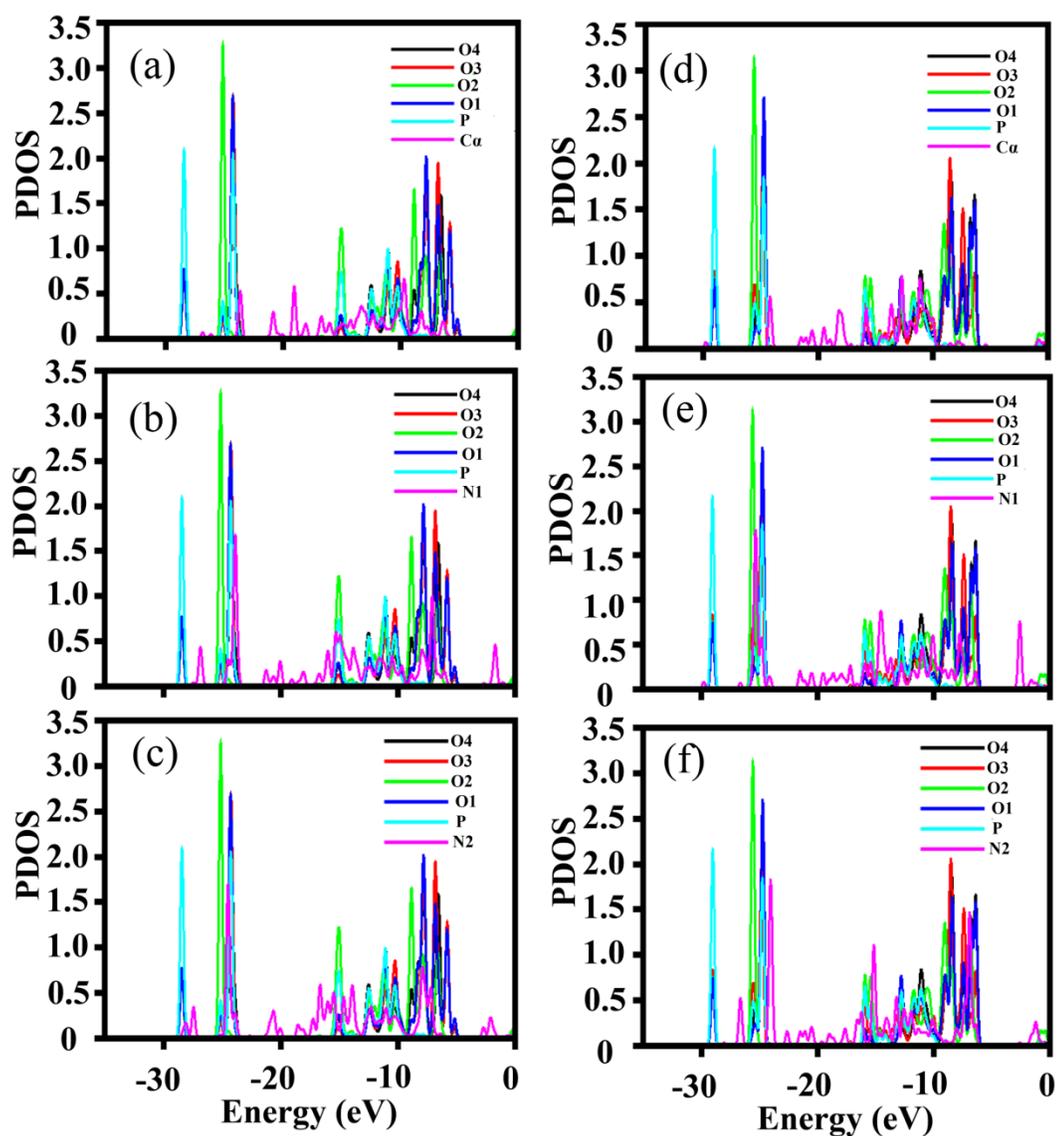


Figure 3.11: PDOS of P (cyan), O1 (blue), O2 (green), O3 (red) and O4 (black) atoms of bound biphosphate and motif atoms of peptide (magenta): (a) C^α of L (b) N1 of Y and (c) N2 of D for different energy in case of LYD fragment and (d) C^α of G (e) N1 of I and (f) N2 of H for different energy in case of GIH fragment in SRP case.

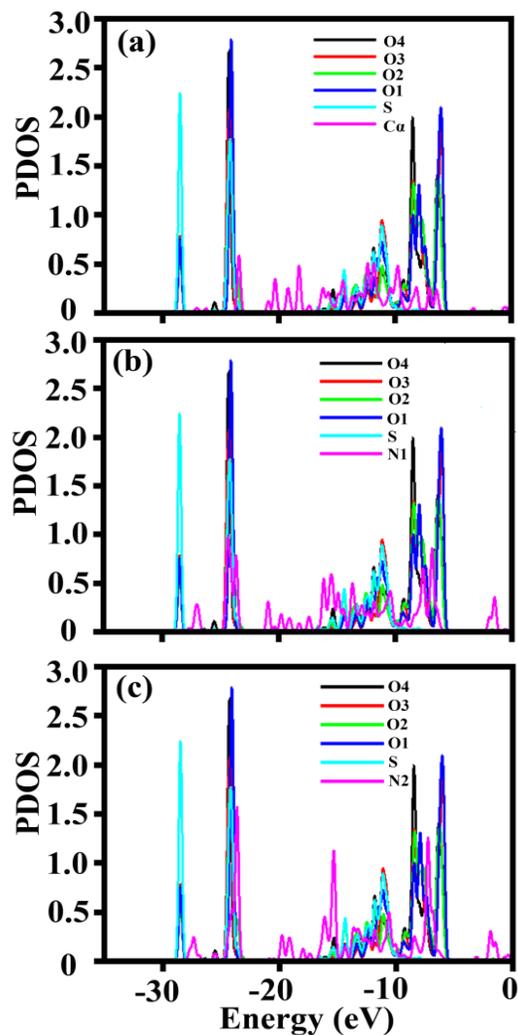


Figure 3.12: PDOS of S (cyan), O1 (blue), O2 (green), O3 (red) and O4 (black) atoms of bound sulphate and motif atoms of peptide (magenta): (a) C^α of S (b) N1 of R and (c) N2 of S for different energy in case of SRS fragment in PRS case.

Thus, due to replacement of anion lowest energy coordination between motif atoms and oxygen atom of anion takes place at higher energy for all the fragments. Due to anion replacement in the motif fragment stabilization energy decreases and relative stabilization energy also becomes positive (Table 3.4) in all the cases, indicating energetically unfavorable conditions.

Table 3.4: Stabilization energy of C^αNN motif containing fragments on replacement of anion

Fragment containing the C^αNN motif residues	Relative stabilization energy due to replacement of anion in kcal/mol (E_{SRP/PRS} + E_{SO4/HPO4}) - (E_{WS/WP} + E_{HPO4/SO4})
SNQ fragments (SRP)	441.2
SQT fragment (SRP)	280.0
LYD fragment (SRP)	308.9
GIH fragment (SRP)	531.2
SRS fragment (PRS)	104.5

3.3.5 Benchmark calculations:

I also perform benchmark calculations for SQT fragment using Gaussian 03¹⁸ considering B3LYP functional with 6-31G (2d, 2p) basis set which is normally used for biomolecular systems^{4,26-28}. I find that the dihedral angles (ϕ , ψ , χ_1) in the VASP optimized and crystal structures are similar (Table 3.5). Consequently, I take the crystal structure and perform partial optimization of SQT peptide fragment both in presence and absence of sulphate by fixing all the non-hydrogen atoms. In the optimized geometry I analyze the orbital contributions of motif atoms and anion over the molecular orbitals of the system. I also study the charge distribution in the optimized geometry both in presence and absence of sulphate using natural population analysis (NPA).

Table 3.5: Dihedral conformations (ϕ , ψ , χ_1) of motif residues in SQT fragment

Residue Name	Dihedral conformations (ϕ, ψ, χ_1) in crystal structure	Dihedral conformations (ϕ, ψ, χ_1) in VASP optimized geometry
S	-71.9, 134.7, 130.4	-73.1, 135.1, 131.2
Q	-74.6, -29.1, -111.9	-73.9, -30.5, -112.0
T	-124.7, -16.2, 154.5	-123.9, -16.5, 157.1

I compute the occupancies of all the atomic orbitals in the molecular orbitals. The calculations indicate that the simultaneous contribution of motif atoms and oxygen atoms of sulphate takes place within the energy range -14 to -18 eV in case of SQT fragment (Figure 3.13). C $^\alpha$ and N1 atoms of the motif coordinate with O2 atom of sulphate at E = -15.52 eV and -15.98 eV, respectively. O1 atom of sulphate coordinate with N2 atom of the motif at E = -16.12 eV. According to VASP results, C $^\alpha$ and N1 atom of the motif together coordinate with O2 atom of sulphate at E = -11.70 eV and N2 atom of the motif coordinates with O1 atom of sulphate at slightly larger energy, E = -11.04 eV. The coordination is comparable to the VASP results, although the coordination take place at lower energies. I also calculate the change in natural charges on the atoms of the motif residues due to presence of sulphate. I find that the amount of charges lost by sulphate (0.218) in bound state is distributed over the peptide fragments as shown in Table 3.6.

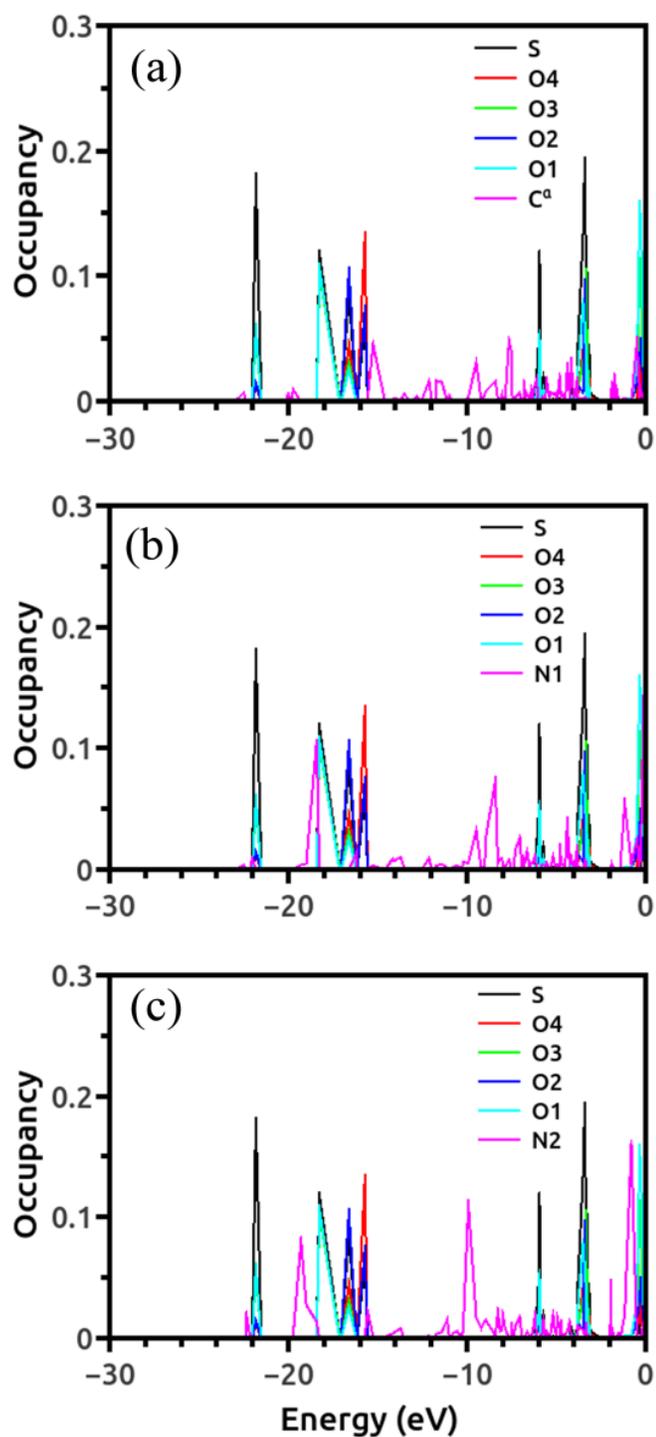


Figure 3.13: Orbital occupancy of S (black), O1 (cyan), O2 (blue), O3 (green) and O4 (red) atoms of bound sulphate and motif atoms of peptide (magenta): (a) C^α of S, (b) N1 of Q and (c) N2 of T in case of SQT fragment in WS condition as function of energy.

Table 3.6: Changes of natural charge of the atoms in motif residues and bound sulphate for SQT peptide fragment in case of WS and WOS/free conditions

Atom Name	Residue Name	Natural charge in WS condition	Natural charge in WOS/free condition	Change of natural charge
N	S134	-0.6222	-0.63542	0.013
C ^α	S134	-0.18054	-0.1594	-0.021
C	S134	0.67262	0.67435	-0.002
O	S134	-0.65435	-0.58382	-0.071
C ^β	S134	-0.11464	-0.11562	0.001
H	S134	0.39815	0.41813	-0.020
H ^α	S134	0.30443	0.24561	0.059
H ^{β1}	S134	0.19988	0.22031	-0.020
H ^{β2}	S134	0.22589	0.26239	-0.036
H ^{γ1}	S134	0.54182	0.51505	0.027
O ^γ	S134	-0.80379	-0.79491	-0.009
N1	Q135	-0.67696	-0.67019	-0.007
C ^α	Q135	-0.13226	-0.14284	0.011
C	Q135	0.68144	0.67384	0.008
O	Q135	-0.68023	-0.62575	-0.054
C ^β	Q135	-0.48232	-0.47017	-0.012
C ^γ	Q135	-0.54625	-0.55054	0.004
C ^δ	Q135	0.66216	0.66036	0.002
N ^{ε2}	Q135	-0.86435	-0.84843	-0.016
O ^{ε1}	Q135	-0.63355	-0.65	0.016
H	Q135	0.47638	0.4287	0.048
H ^α	Q135	0.25695	0.29273	-0.036
H ^{β1}	Q135	0.2809	0.27334	0.008

Chapter 3

H ^{β2}	Q135	0.27936	0.23849	0.041
H ^{γ1}	Q135	0.26018	0.27198	-0.012
H ^{γ2}	Q135	0.24554	0.2656	-0.020
H ^{ε2}	Q135	0.41648	0.43298	-0.016
H ^{ε2}	Q135	0.39294	0.40891	-0.016
N2	T136	-0.67237	-0.66635	-0.006
C ^α	T136	-0.16513	-0.17257	0.007
C	T136	0.69499	0.68653	0.008
O	T136	-0.69016	-0.65312	-0.037
C ^β	T136	0.09068	0.08502	0.006
C ^{γ2}	T136	-0.72753	-0.72615	-0.001
O ^{γ1}	T136	-0.76483	-0.78412	0.019
H	T136	0.47876	0.42652	0.052
H ^α	T136	0.2838	0.30611	-0.022
H ^β	T136	0.20607	0.23997	-0.034
H ^{γ1}	T136	0.49869	0.49131	0.007
H ^{γ2}	T136	0.21428	0.23613	-0.022
H ^{γ2}	T136	0.22906	0.25747	-0.028
H ^{γ2}	T136	0.2741	0.23005	0.044
S	SO4	2.66452	2.62088	0.044
O1	SO4	-1.06616	-1.12756	0.061
O2	SO4	-1.11814	-1.18002	0.062
O3	SO4	-1.14681	-1.18244	0.036
O4	SO4	-1.11587	-1.13087	0.015

3.4 Discussion:

Since the same systems I have studied by classical force field simulations¹, it is worth to compare the present results to those from simulations. Classical force field based molecular dynamics (MD) simulation studies show anion induced conformational preference in the motif residues¹. A conformational preference of a residue has been defined as percentage of RH, β , LH and coil (C) over simulated conformation in equilibrium. It is observed that conformation of motif residues fluctuate without anion and anion typically stabilizes one of the residue conformations observed in absence of the anion¹. If the conformational preference in the presence of anion is one as the minor population in the absence of anion, the residue is said to undergo conformational switching induced by the anion, else the residue does not undergo conformational switching. Table 3.7 and Table 3.8 show in details the conformational preference of motif residues and the coordination pattern in these residues. The motif residue in each peptide fragment which does not show conformational switching due to presence of anion coordinate with the oxygen atom of the anion at lowest energy. These residues are mostly polar in nature. The residues undergoing conformational switching have coordinated at higher energy levels in general. Thus, the residues having no conformational switching are the ones in energetically favorable condition for coordination.

The connection between conformational switch and coordination energy is revealed by anion replacement as well, although the coordination energies are shifted to higher values. Due to anion replacement conformational switching of the motif residues is not observed in case of SNQ, SQT and GIH fragment (Table 3.8) where I observe coordination with anion oxygen, albeit shifted at higher energy than WS/WP cases. Only L in LYD fragment undergoes conformational switching in SRP condition. This residue does not participate in coordination

(Table 3.8). Similarly, S in SRS fragment undergoes conformational switching in PRS condition and it also does not participate in coordination.

Table 3.7: Conformational preference of C^αNN motif residues in different peptide fragments as observed in previous study¹. Major conformational preference is marked in bold font

Motif fragment	Residue	Conformational preference		
		WOS	WS/WP	SRP/PRS
SNQ	S	RH, β	β	β
	N	RH	RH	RH
	Q	RH , β	RH	RH
SQT	S	RH , β	RH	RH , β
	Q	RH , β	RH	RH , β
	T	RH , β	β	RH
LYD	L	RH, β	β	RH
	Y	RH	RH	RH, β
	D	RH	β	RH
GIH	G	RH, β, LH, C	RH	RH, β , LH, C
	I	RH, β	RH	RH, β
	H	B	β	RH, β, LH, C
SRS	S	RH , β	β	β
	R	RH	RH	RH
	S	RH	β	RH

Table 3.8: Conformational switch of C^αNN motif residues in different peptide fragments as observed in previous study¹ and coordination energy of motif atoms with the oxygen atoms of anion observe in QC calculation. Deepest coordination energy is marked in bold font

Motif fragment	Residue	Conformational switch		Coordination Energy with motif atoms and O atom of anion in eV	
		Presence of anion (WS/WP)	Replacement of anion (SRP/PRS)	Presence of anion (WS/WP)	Replacement of anion (SRP/PRS)
SNQ	S	No	No	-14.30	
	N	No	No	-14.292	-7.526
	Q	No	No	-15.30	-6.982
SQT	S	No	No	-11.699	
	Q	No	No	-11.699	-7.142
	T	Yes	No	-11.043	-6.738
LYD	L	No	Yes	-14.852	
	Y	No	No	-14.90	-7.939
	D	Yes	No	-13.484	-8.883
GIH	G	Yes	No		
	I	Yes	No	-10.059	-11.123
	H	No	No	-16.616	
SRS	S	Yes	Yes	-11.155	
	R	No	No	-11.123	-7.603
	S	Yes	No	-11.955	-8.427

According to the bioinformatics study³⁰ based on the crystal structure database analysis of the sulphate/phosphate binding proteins, the anion binding C^αNN motif interacts with the anionic oxygen via backbone C^α and N atoms of the consecutive motif residues in a specific pattern, where C^α and N1 atom of the motif interact with one oxygen atom of the anion concurrently and N2 atom of the motif interacts with another oxygen atom of anion. In QC study I find that this stable interaction pattern mainly stabilizes the anion in the motif and gives the anion selectivity. My studies also support the observation from previous biophysical studies³¹⁻³² based on the C^αNN motif containing synthetic peptide fragments that polar residues are the primary ones to give stability to the motif. Based on this study and previous MD simulation studies I propose that the mode of binding interactions and anion recognition mechanism of C^αNN motif for various functional proteins and model synthetic peptides is universal one.

3.5 Conclusion:

The quantum chemical calculations show that the stability of C^αNN motif is governed by the coordination of motif atoms with oxygen atoms of anion. The sequence of the motif influences the coordination energy and stabilization energy. I find larger stabilization energy in case of motif having more polar residues. Stability of the motif also depends on the type of anion. The most nontrivial aspect of my results is the intimate connection between coordination and conformational preferences of the residues which has not been established so far to the best of my knowledge. I consider the motif residues from different proteins which have different functional sites, and, therefore, this study may be useful to get the microscopic

Chapter 3

insights into their functions. Similar analysis can be useful for understanding stability of other ligand binding motifs as well.

References:

1. Patra, P.; Ghosh, M.; Banerjee, R.; Chakrabarti, J. Anion induced conformational preference of C^αNN motif residues in functional proteins. *Proteins* **2017**, *85* (12), 2179-2190.
2. Kagawa, H.; Mori, K. Molecular orbital study of the interaction between MgATP and the myosin motor domain: The highest occupied molecular orbitals indicate the reaction site of ATP hydrolysis. *J. Phys. Chem. B* **1999**, *103* (34), 7346-7352.
3. Blomberg, M. R. A.; Siegbahn, P. E. M. A quantum chemical approach to the study of reaction mechanisms of redox-active metalloenzymes. *J. Phys. Chem. B* **2001**, *105* (39), 9375-9386.
4. Sikdar, S.; Ghosh, M.; De Raychaudhury, M.; Chakrabarti, J. Quantum Chemical Studies on Stability and Chemical Activities in Calcium Ion Bound Calmodulin Loops. *J. Phys. Chem. B* **2015**, *119* (46), 14652-9.
5. Sikdar, S.; Ghosh, M.; De Raychaudhury, M.; Chakrabarti, J. Quantum chemical studies on the role of residues in calcium ion binding to Calmodulin. *Chem. Phys. Lett.* **2014**, *605*, 103-107.
6. Fukui, K. Role of frontier orbitals in chemical reactions. *Science* **1982**, *218* (4574), 747-54.
7. Lepsik, M.; Field, M. J. Binding of calcium and other metal ions to the EF-hand loops of calmodulin studied by quantum chemical calculations and molecular dynamics simulations. *J. Phys. Chem. B* **2007**, *111* (33), 10012-22.
8. Wu, E. L.; Mei, Y.; Han, K.; Zhang, J. Z. Quantum and molecular dynamics study for binding of macrocyclic inhibitors to human alpha-thrombin. *Biophys. J.* **2007**, *92* (12), 4244-53.
9. He, X.; Zhang, J. Z. The generalized molecular fractionation with conjugate

caps/molecular mechanics method for direct calculation of protein energy. *J. Chem. Phys.* **2006**, *124* (18), 184703.

10. Zhang, D. W.; Chen, X. H.; Zhang, J. Z. Molecular caps for full quantum mechanical computation of peptide-water interaction energy. *J. Comput. Chem.* **2003**, *24* (15), 1846-52.

11. Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoseck, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30* (10), 1545-614.

12. Leach, A. R. *Molecular Modelling: Principles and Applications*. 2nd ed.; Pearson Education: **2009**.

13. Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*. 2nd ed.; Wiley: **2004**.

14. Car, R.; Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **1985**, *55* (22), 2471-2474.

15. M. C. Payne, M. P. T., D. C. Allan, T. A. Arias, and J. D. Joannopoulos. Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.* **1992**, *64*, 1045-1097.

16. M D Segall, P. J. D. L., M J Probert, C J Pickard, P J Hasnip, S J Clark and M C Payne First-principles simulation: ideas, illustrations and the CASTEP code. *J. Phys. Condens. Matter* **2002**, *14*, 2717-2744.

17. Kresse, G.; Furthmuller, J. Efficient iterative schemes for ab initio total-energy

calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54* (16), 11169-11186.

18. Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A. *Gaussian 03, Revision C.02*, Gaussian, Inc., Wallingford CT: **2004**.

19. Blochl, P. E. Projector augmented-wave method. *Physical Rev. B, Condens. Matter* **1994**, *50* (24), 17953-17979.

20. Kresse, G.; Hafner, J. Ab-Initio Molecular-Dynamics for Open-Shell Transition-Metals. *Phys. Rev. B* **1993**, *48* (17), 13115-13118.

21. Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **1999**, *59* (3), 1758-1775.

22. Henkelman, G.; LaBute, M. X.; Tung, C. S.; Fenimore, P. W.; McMahon, B. H. Conformational dependence of a protein kinase phosphate transfer reaction. *Proc. Nat. Acad.*

Sci. USA **2005**, *102* (43), 15347-51.

23. Adhikari, P.; Wen, A. M.; French, R. H.; Parsegian, V. A.; Steinmetz, N. F.; Podgornik, R.; Ching, W. Y. Electronic structure, dielectric response, and surface charge distribution of RGD (1FUV) peptide. *Sci. Rep.* **2014**, *4*, 5605.

24. Silvestrelli, P. L.; Parrinello, M. Structural, electronic, and bonding properties of liquid water from first principles. *J. Chem. Phys.* **1999**, *111* (8), 3572-3580.

25. Kresse, G.; Hafner, J. Abinitio Molecular-Dynamics for Liquid-Metals. *Phys. Rev. B* **1993**, *47* (1), 558-561.

26. Tirado-Rives, J.; Jorgensen, W. L. Performance of B3LYP Density Functional Methods for a Large Set of Organic Molecules. *J. Chem. Theory Comput.* **2008**, *4* (2), 297-306.

27. Siegbahn, P. E.; Borowski, T. Modeling enzymatic reactions involving transition metals. *Acc. Chem. Res.* **2006**, *39* (10), 729-38.

28. Blomberg, M. R.; Siegbahn, P. E. Quantum chemistry applied to the mechanisms of transition metal containing enzymes -- cytochrome c oxidase, a particularly challenging case. *J. Comput. Chem.* **2006**, *27* (12), 1373-84.

29. *The PyMOL Molecular Graphics System*, Version 2.0; Schrodinger, LLC.

30. Denessiouk, K. A.; Johnson, M. S.; Denesyuk, A. I. Novel CalphaNN structural motif for protein recognition of phosphate ions. *J. Mol. Biol.* **2005**, *345* (3), 611-29.

31. Sheet, T.; Banerjee, R. The '(CNN)-N-alpha' motif: an intrinsic lover of sulfate and phosphate ions. *Rsc. Adv.* **2016**, *6* (59), 54129-54141.

32. Sheet, T.; Ghosh, S.; Pal, D.; Banerjee, R. Computational design of model scaffold for anion recognition based on the 'C(alpha) NN' motif. *Biopolymers* **2017**, *108* (1).

Appendices of Chapter 3

A3.1 Density Functional Theory for electronic structure calculations:

Density functional theory (DFT) is a successful theory to calculate the electronic structure of atoms, molecules, and solids. Traditional electronic structure methods (e.g. Hartree-Fock method) attempt to find approximate solutions to the Schrodinger equation of N interacting electrons moving in an external, electrostatic potential (Coulomb potential generated by the atomic nuclei). However, there are serious limitations of this approach: (i) the problem is highly nontrivial, even for very small numbers N and the resulting wave-functions are complicated objects, (ii) the computational effort grows very rapidly with increasing N , so the description of larger systems becomes prohibitive. A different approach is taken in DFT where, instead of the many-body wave-function, the one-body density is used as fundamental variable. Since the density $n(r)$ is a function of only three spatial coordinates (rather than the $3N$ coordinates of the wave-function), density functional theory is computationally feasible even for large systems. The foundations of density functional theory are the Hohenberg-Kohn and Kohn-Sham theorems¹²⁻¹³.

In ground-state DFT one is interested in systems of N interacting electrons described by the Hamiltonian

$$\begin{aligned} \hat{H} &= \hat{T} + \hat{V} + \hat{V}_{ee} \\ &= -\sum_{i=1}^N \frac{\nabla_i^2}{2} + \sum_{i=1}^N v(r_i) + \frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N \frac{1}{|r_i - r_j|} \end{aligned} \quad (3.1)$$

with the kinetic, potential and interaction energy operators \hat{T} , \hat{V} and \hat{V}_{ee} , respectively. The central statement of density functional theory according to the Hohenberg-Kohn theorem can

be summarized in the following three statements¹³:

1. The ground state electron density $n(r)$ of a system of interacting electrons uniquely determines the external potential $v(r)$ in which the electrons move and thus the Hamiltonian and all physical properties of the system.

2. The ground-state energy E_0 and the ground-state density $n_0(r)$ of a system characterized by the potential $v_0(r)$ can be obtained from a variational principle which involves only the density, i.e., the ground state energy can be written as a functional of the density, $E_{v_0}[n]$, which gives the ground-state energy E_0 if and only if the true ground-state density $n_0(r)$ is considered. For all other densities $n(r)$, the inequality

$$E_0 = E_{v_0}[n_0] < E_{v_0}[n] \quad (3.2)$$

holds.

3. There exists a functional $F[n]$ such that the energy functional can be written as

$$E_{v_0}[n_0] = F[n] + \int d^3r v_0(r)n(r) \quad (3.3)$$

The functional $F[n]$ is universal in the sense that, for a given particle-particle interaction, it is independent of the potential $v_0(r)$ of the particular system under consideration, i.e., it has the same functional form for all systems.

From the Hohenberg-Kohn variational principle, i.e., the second statement given above, the ground-state density $n(r)$ corresponding to the external potential $v(r)$ can be obtained as solution of the Euler equation¹²⁻¹³

$$\frac{\delta E_v[n]}{\delta n(r)} = \frac{\delta F[n]}{\delta n(r)} + v(r) = 0 \quad (3.4)$$

The formal definition of the Hohenberg-Kohn functional $F[n]$ is¹²⁻¹³,

$$F[n] = T[n] + V_{ee}[n] = \langle \psi[n] | \hat{T} | \psi[n] \rangle + \langle \psi[n] | \hat{V}_{ee} | \psi[n] \rangle \quad (3.5)$$

where $\psi[n]$ is that N-electron wave-function which yields the density n and minimizes the expectation value of $\hat{T} + \hat{V}_{ee}$.

Hohenberg-Kohn theorem provides the basic theoretical foundation for the construction of an effective single-particle scheme which allows the calculation of the ground-state density and energy of systems of interacting electrons. The resulting equations, the so-called Kohn-Sham equations, are at the heart of modern density functional theory. They have the form of the single-particle Schrodinger equation¹²⁻¹³

$$\left[-\frac{\nabla^2}{2} + v_s(r) \right] \varphi_i(r) = \varepsilon_i \varphi_i(r) \quad (3.6)$$

The density can then be computed from the N single-particle orbitals occupied in the ground state Slater determinant

$$n(r) = \sum_i^{occ} |\varphi_i(r)|^2 \quad (3.7)$$

The central idea of the Kohn-Sham scheme is to construct the single-particle potential $v_s(r)$ in such a way that the density of the auxiliary non-interacting system equals the density of the interacting system of interest. To this end one can partition the Hohenberg-Kohn functional in the following way¹³

$$F[n] = T_s[n] + U[n] + E_x[n] \quad (3.8)$$

$$\text{Where } U[n] = \frac{1}{2} \int d^3r \int d^3r' \frac{n(r)n(r')}{|r-r'|} \quad (3.9)$$

is the classical electrostatic energy of the charge distribution $n(r)$, and $E_{xc}[n]$ is the exchange-correlation energy which is formally defined by

$$E_{xc}[n] = T[n] + V_{ee}[n] - U[n] - T_S[n] \quad (3.10)$$

From the above definitions one can derive the form of the effective potential using Eq. (3.6)

$$v_s[n](r) = v(r) + \int d^3r' \frac{n(r')}{|r-r'|} + v_{xc}[n](r) \quad (3.11)$$

where the exchange-correlation potential v_{xc} is defined by¹³

$$v_{xc}[n](r) = \frac{\delta E_{xc}[n]}{\delta n(r)} \quad (3.12)$$

Since $v_s[n](r)$ depends on the density, Eq. (3.6), (3.7) and (3.11) have to be solved self-consistently. This is known as the Kohn-Sham scheme of density functional theory.

Formal definition of the exchange-correlation energy is not helpful for practical calculations and one needs to use an approximation for this quantity¹³. The true E_{xc} is a universal functional of the density, i.e., it has the same functional form for all systems. The simplest of all functional, the so-called local density approximation (LDA)¹²⁻¹³ has remained the approximation of choice. In LDA, the exchange-correlation energy is given by

$$E_{xc}^{LDA}[n] = \int d^3r n(r) e_{xc}^{unif}(n(r)) \quad (3.13)$$

where $e_{xc}^{unif}(n(r))$ is the exchange-correlation energy per particle of an electron gas with spatially uniform density n . It can be obtained from quantum Monte Carlo calculations. By its

very construction, the LDA is expected to be a good approximation for spatially slowly varying densities. LDA has proved to be remarkably accurate for a wide variety of systems¹³.

There are two fundamentally different numerical approaches that are used for solving the Kohn–Sham equation as is required in large-scale DFT calculations. The first approach represents the electron density orbitals by a plane wave basis set which is most naturally applied to condensed phase systems in which the imposition of periodic boundary conditions (implicit in the use of plane waves, which are periodic) is appropriate. Plane wave methods have been implemented by two different algorithmic approaches. Car–Parinello methods¹⁴⁻¹⁷ mandate performing *ab-initio* molecular dynamics, evolving the system by means of an extended Lagrangian formalism. This type of approach is most useful in studying dynamics in liquids and molten solids. Conjugate gradient-based optimization approaches, implemented in programs such as Vienna ab-initio Simulation Package (VASP)¹⁷ is typically used to electronic structure calculation in solid state and material science problems. The alternative to plane wave methods is the use of localized, atom-centered Gaussian basis sets, as has been used for many years in solving the Hartree–Fock equations. Indeed, most localized basis implementations of DFT were developed by modifying existing Hartree–Fock codes, as was done, for example, in the GAUSSIAN series of programs¹⁸. The use of Gaussian basis functions allows hybrid DFT functionals to be treated in straightforward fashion. Gaussian calculations are all-electron calculations, while the VASP calculations are full-potential valence-only calculations. As a consequence, Gaussian calculations will be very expensive for larger systems for which plane wave calculations are suited very well.

A3.2 Quantum chemical calculation using VASP:

Complex package for performing *ab-initio* quantum-mechanical molecular dynamics (MD) simulations using pseudo-potentials or the projector-augmented wave method and a plane wave basis set is VASP¹⁷. The approach implemented in VASP is based on the (finite-temperature) local-density approximation with the free energy as variational quantity and an exact evaluation of the instantaneous electronic ground state at each MD time step.

Efficient matrix diagonalisation schemes and an efficient Pulay/Broyden charge density mixing are used in VASP. These techniques avoid all problems possibly occurring in the original Car-Parrinello method, which is based on the simultaneous integration of electronic and ionic equations of motion¹⁷. Ultra-soft Vanderbilt pseudo-potentials (US-PP) or the projector-augmented wave (PAW) method describes the interaction between ions and electrons. In this PAW method, the size of the basis-set can be kept very small even for transition metals and first row elements like C and O. Generally, not more than 100 plane waves (PW) per atom are used for bulk materials, in most cases even 50 PW per atom will be enough for a reliable description. Forces and the full stress tensor can be calculated with VASP and used to relax atoms into their instantaneous ground-state. Pseudo-potentials have been introduced to avoid the need for an explicit treatment of the strongly bound and chemically inert core electrons. They are a necessary ingredient of all plane-wave methods, but they can also be used in local-basis set methods to reduce the computational effort¹⁷.

A3.3 Quantum chemical calculation using GAUSSIAN:

GAUSSIAN is a computational chemistry software program used for performing the quantum chemical calculations. GAUSSIAN calculations are all-electron calculations considering localized, atom-centered Gaussian basis sets with different functional forms¹⁸. The GAUSSIAN program is usually controlled by specific *keywords*, which request given types of calculation. If the *keywords* are used, the program converts them to internal parameters, which then control the execution.

Various steps in GAUSSIAN calculations in this thesis are summarized as follows:

1. The first step is to read in a title for the job plus molecular charge, multiplicity (singlet, triplet, etc.), molecular geometry in the form of a *Z-matrix*. This consists of atomic numbers, bond lengths, bond angles, dihedral angles from which the Cartesian (x, y, z) coordinates are calculated. The information from the *Z-matrix* is then used to work out electronic configuration and the orbital occupancies.
2. In the next step the nuclear repulsion energy which depends on the atomic numbers and the molecular geometry is calculated.
3. The atomic orbitals or basis sets are then assigned to each nucleus. *Ab-initio* programs use an internally stored standard set of coefficients and exponents that define the orbitals (the *basis set*).
4. *Ab-initio* programs next calculate the various one- and two-electron integrals required later in the calculation.
5. *Ab-initio* programs then produce an *initial guess* of molecular orbitals used as a starting point for the SCF calculations. The usual form of initial guess for *ab-initio* programs is that obtained from an extended Hückel calculation on the molecule of interest.

6. The solution to the SCF equations is iterated cycle by cycle until the electronic energy is at a minimum and the density matrix does not change. At this stage the calculation is said to be converged, or to have reached self-consistency, and the program proceeds to the next step.
7. For a geometry optimization the atomic forces are then determined analytically and used to estimate the minimum-energy geometry for the molecular species being calculated.
8. The above process is repeated for each new geometry until the atomic forces are close to zero and the total energy does not change significantly from cycle to cycle.
9. The next stage of the calculation depends on the type of job to be performed. For a single point, the program may either move directly to the *population analysis*, which calculates the atomic charges, overlaps, dipole and higher moments. The correlation energy for DFT is then calculated.

The flow chart for performing the typical *ab-initio* quantum chemical calculation in GAUSSIAN is shown in Figure A3.1.

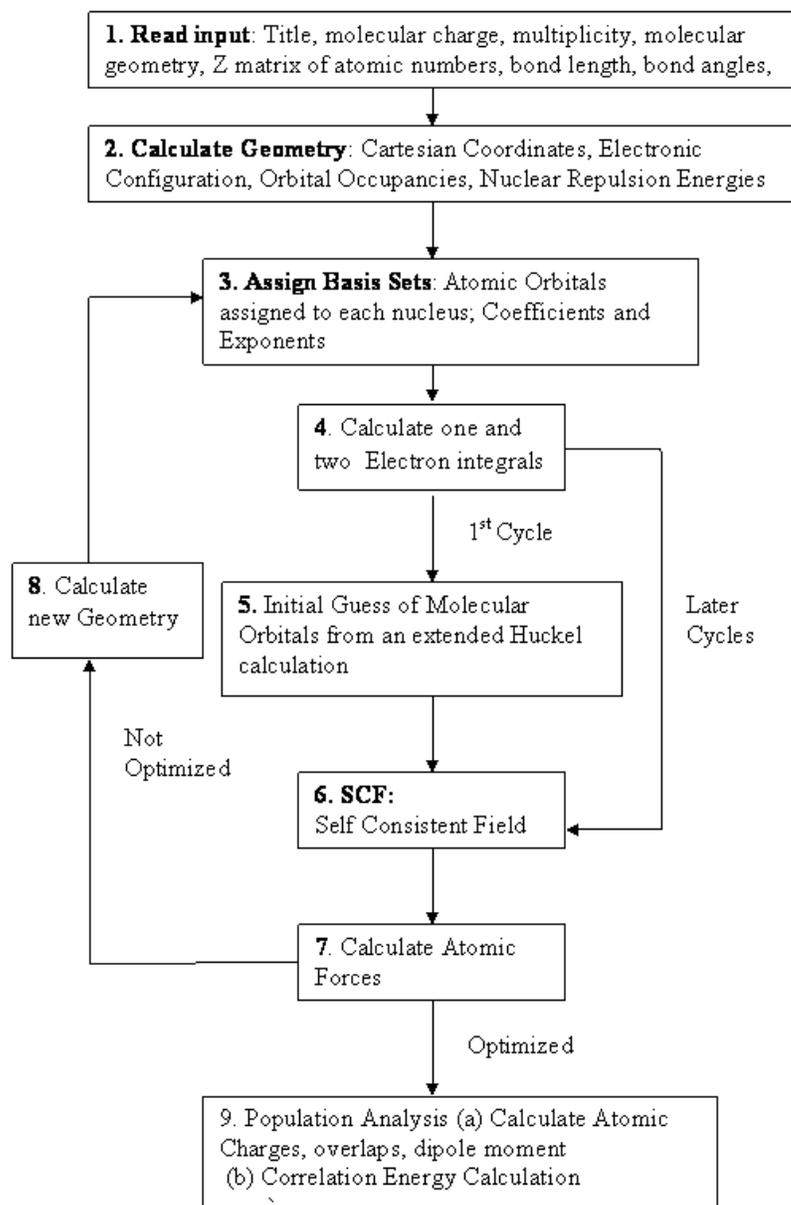


Figure A3.1: Flow chart for performing the *ab-initio* quantum chemical calculation in GAUSSIAN



CHAPTER 4

ANION SCAVENGING BY CATIONIC ANTIMICROBIAL PEPTIDE



4.1 Introduction:

Often bio-macromolecules, like many polypeptides and nucleic acids are highly charged. Counterion condensation in the vicinity of these bio-macromolecules is not only of crucial importance in biological phenomena¹⁻³, but also relevant in applications, like medicine and catalysis⁴⁻⁵ and has been extensively studied⁶⁻¹⁰. Cationic polypeptides, rich in basic residues and involved in anion binding, sensing and transport, have drawn current interest for their ability to penetrate to lipid bilayers. Consequently, cationic polypeptides, also known as cell penetrating peptides (CPP)¹¹⁻¹² participate in defense mechanism by host cells¹³⁻¹⁴ and have potential use as drug carriers¹⁵⁻²⁰. Experimental and simulation studies indicate the penetration by CPP into cell membrane depends on electrostatic interaction of the peptide with phosphate groups of lipid bilayer of cell membrane and is facilitated in presence of negatively charged species in solution²¹. These observations suggest that interaction of anionic species with these cationic polypeptides is of crucial importance.

Unlike nucleic acids with uniform distribution of charges over backbone, charged functional groups, distributed non-uniformly over side chains of the residues. The cationic domains in CPPs are rich in arginine, a hydrophilic amino acid residue with basic guanidinium side chain at normal pH²²⁻²³. A Large number of naturally occurring enzymes²⁴ also contain arginine. Several experiments show anion-mediated variability of charge and solubility of guanidinium-rich poly-peptides, also termed as “arginine magic”²⁵⁻³⁰, make them adaptable to many environments. Anion binding by mono-, oligo-, and polymeric guanidinium cations has also been reported³¹⁻³⁹. Existing theoretical studies which are primarily based on coarse-grained models^{10,40-45} do not capture explicit role of charged functional groups on counterion condensation. For instance, microscopic origin of “arginine magic” is not understood to the

best of our knowledge. Such microscopic information may also be useful to design polypeptides as drug carriers into a cell.

Here I study by all-atom simulations of cationic polypeptide of a CPP in aqueous medium in presence of anions. The main objective is to emphasize the role of functional groups in the side chains and residue conformation in anion condensation. A simple model has also been studied to figure out the generic feature of observations from all-atom simulations. I consider HIV-1 Tat peptide, positively charged transduction domain of HIV-1 trans-activator of transcription (Tat) protein⁴⁶, involved in immunological response to human immunodeficiency virus (HIV). HIV-1 Tat peptide is capable of translocating cargoes of different molecular sizes, such as proteins, DNA, RNA, or drugs, across cell membrane^{15-20, 47-50}. The HIV-1 Tat peptide sequence (residues 48–60) is GRKKRRQRRRPPQ (denoted by one letter code, G = glycine, R = arginine, K = lysine, Q = glutamine and P = proline) with eight basic residues, R and K, a couple of polar Q and hydrophobic G and P⁴⁷⁻⁵⁰. NMR studies suggest that the polypeptide is unstructured in solution⁵¹. While G does not have any side chain, R has guanidium groups, both K and Q has one amide group each in the side chain, and P has a cyclic ring with nitrogen⁵². The anion is taken to be biphosphate (HPO_4^{2-}), hydrated form of phosphate ions, at different concentrations in aqueous medium. It is observed that biphosphate condensation mostly take place around the side chain nitrogen atoms, facilitated by helical conformation preference of residue. The condensation attains a maximum and minimum solvent exposure as a function of anion concentration in the solution. Such microscopic information may be useful to design polypeptides as drug carriers into a cell.

The chapter is organized as follows: material & methods in section 4.2, results of all-atom simulations in section 4.3, model calculation in section 4.4, discussion in section 4.5 and conclusion in section 4.6.

4.2 Material & methods:

4.2.1 Molecular Dynamics (MD) simulation:

The initial coordinates of the Tat peptide are taken from the NMR structure of the HIV-1 Tat protein (PDB code 1JFW⁵¹, model 1) having sequence 48GRKKRRQRRRPPQ60 (Figure 4.1).

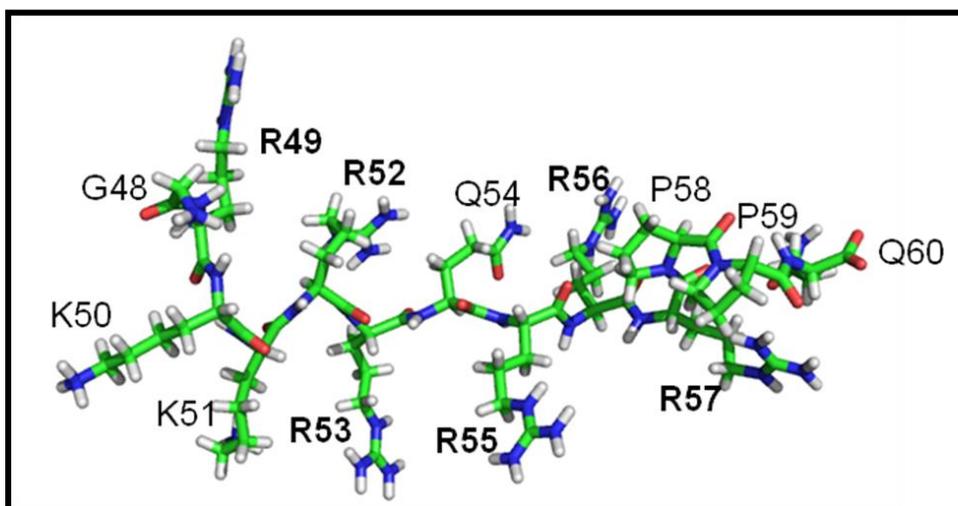


Figure 4.1: Structure of Tat peptide as found in NMR structure of HIV-1 Tat protein (PDB code 1JFW⁵¹). Side chain guanidinium group containing arginine (R) residues are marked in bold font.

All-atom MD simulations (Appendix A2.2) of Tat peptide are carried out both in absence and presence of biphosphate (HPO_4^{2-}) using GROMACS simulation tool⁵³⁻⁵⁴ with Amber99ILDN⁵⁵ force field. I put the peptide in a cubical box having box length 55 Å and solvate the system such that there are at least 10 Å thick layers of water molecules around the solute in all directions.

Along with solvated peptide without biphosphate ions in solution (WOP), four different systems are prepared considering four different concentrations of anionic biphosphate ions in solution (WP). In these four systems I maintain the biphosphate concentration of 10 mM (WP10), 20 mM (WP20), 40 mM (WP40) and 60 mM (WP60) along with required numbers of Na^+ and Cl^- ions for neutralizing the charged system. Each system is relaxed through energy minimization. The long ranged electrostatic interactions are treated by the particle-mesh Ewald method. After energy minimization, the system is subject to 1 microsecond MD run in NPT ensemble using 2 femto-second integration time with periodic boundary conditions over the simulation box at the desired temperature (300 K) and pressure (1 Bar). Altogether I perform 15 microseconds MD simulation in our study. For each system I have three independent MD trajectories. Conformations in each MD run are saved every 1 pico-second from the trajectories for further analysis. The trajectories are visualized with visual molecular dynamics (VMD)⁵⁶.

4.2.2 Theoretical analysis:

Conformational preference of polypeptide residues:

Conformational preference of the amino acid residues in polypeptide chain is analyzed based on the backbone dihedral angles of peptide residues as discussed in Appendix A2.1. The secondary structural preference of polypeptide residues is defined in terms of percentages of right-handed α -helix (RH), left handed α -helix (LH), β and coil (C) conformation over the simulated trajectory.

Solvent accessible surface area (SASA):

Surface area of a polypeptide that is accessible to a solvent can be measured in terms of SASA. With 1.4 Å probe radius of water molecule I calculate average SASA of the

individual peptide residues from the equilibrated MD trajectory using GROMACS tool considering double cubic lattice method^{53-54, 57}.

Anion distribution function:

The distribution of anionic biphosphate around the peptide fragment is analyzed using distribution function $g^{(a)}(r_{N-P})$ of P atom of biphosphate molecule with respect to the backbone and side chain N atom of α -th amino acid residue in the peptide fragment. The distribution function $g^{(a)}(r_{N-P})$ is defined as: $g^{(a)}(r_{N-P}) = \langle N(r) \rangle / (4\pi N \rho \Delta r)^{58}$, where $\langle N(r) \rangle$ is the number of P atom of biphosphate molecules averaged over time, within a distance $r \pm \Delta r/2$ from the atom of interest, ρ is the density of biphosphate and N is the total number of biphosphate molecules in the system.

4.3 Results of all-atom simulations:

Equilibration of the studied systems is ensured from the root-mean-square displacement (RMSD) of the peptide segment considering backbone heavy atoms with respect to its initial energy minimized structure and the equilibrated trajectories are used for further analysis (Figure 4.2). The residues are denoted by one letter code of amino acid along with the residue number. The terminal residues are not considered in the analysis.

Structural properties of peptide:

The conformational preferences of the peptide residues are characterized in terms of secondary structural elements, like right handed α helix (RH), left handed α helix (LH), beta (β) and coil (C) conformation. The propensities of different structural elements of the peptide residues are calculated over the simulated trajectory. Figure 4.3 shows the percentages of different secondary structural elements of the peptide residues in absence of biphosphate and

different concentration of biphosphates. Most of the residues prefer β conformation in solution in WOP condition.

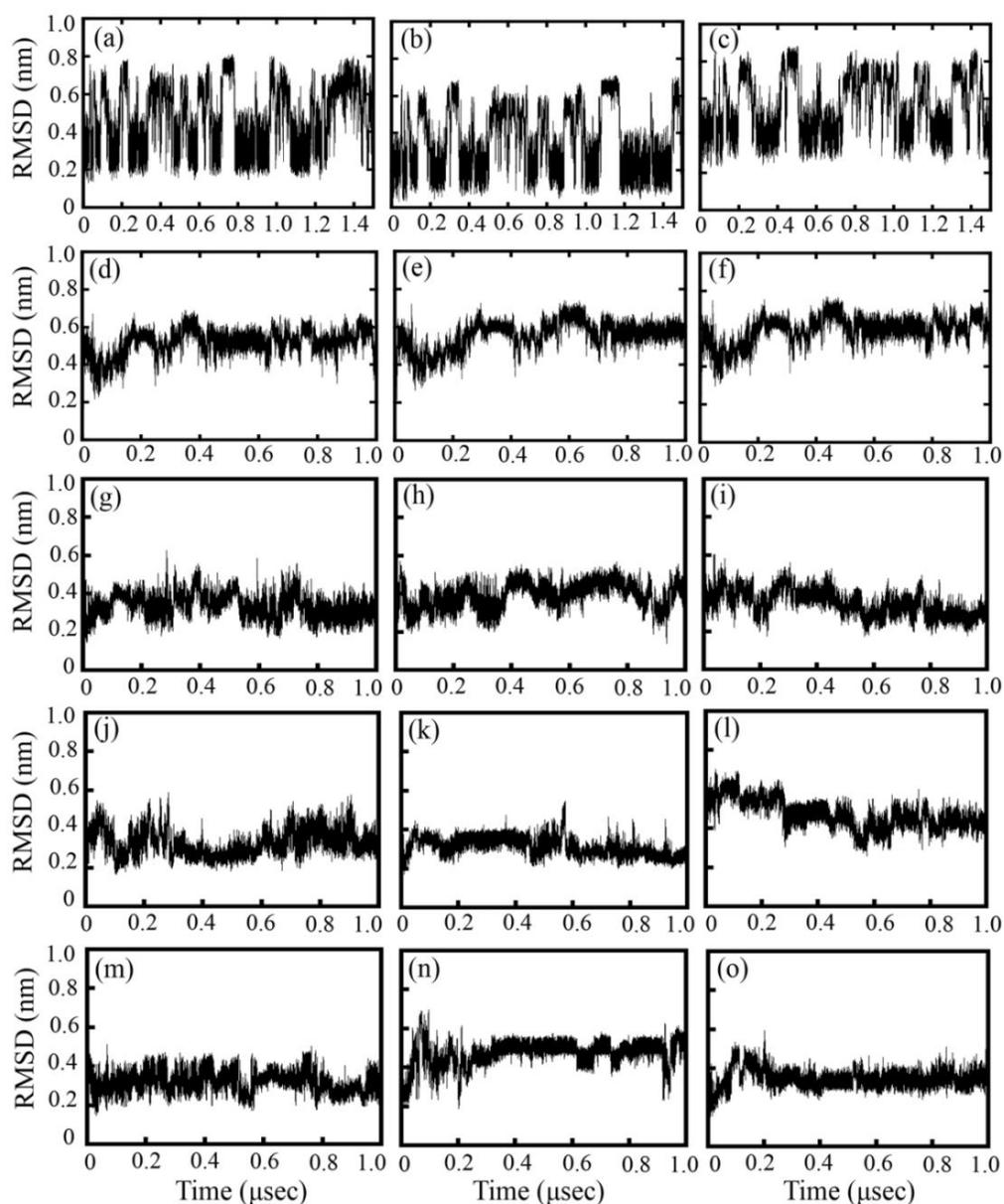


Figure 4.2: Root-mean-square deviations (RMSD) of peptide chain with respect to the initial energy minimized structure in absence (a)-(c) and presence of 10 mM (d)-(f), 20 mM (g)-(i), 40 mM (j)-(l) and 60 mM (m)-(o) biphosphate ions in solution for three independent MD trajectories.

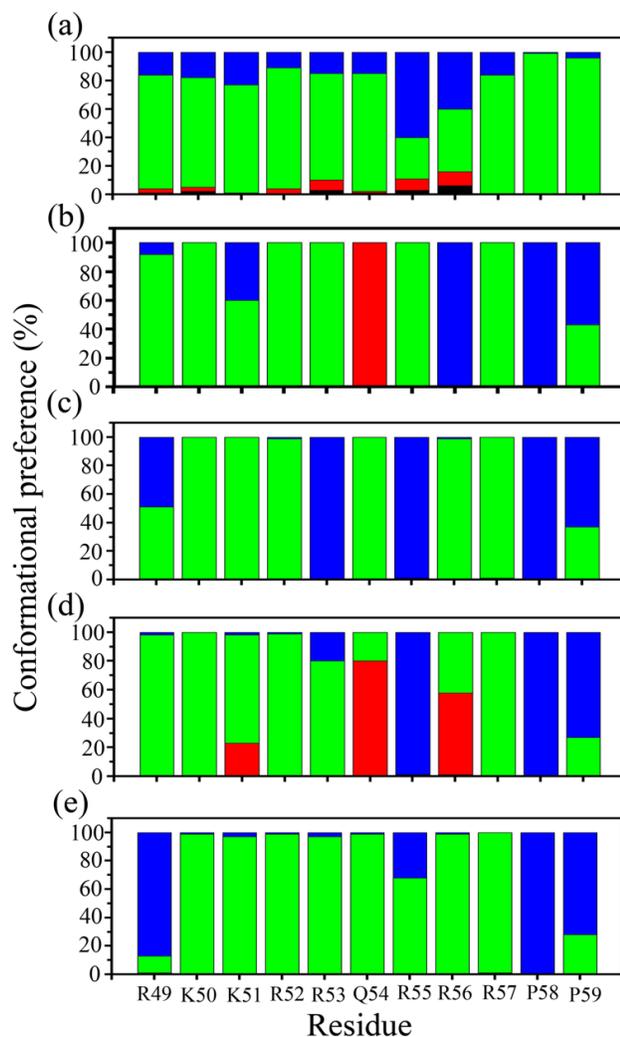


Figure 4.3: Conformational preference (RH: blue, β : green, LH: red, coil: black) of the residues in HIV-1 Tat peptide in (a) absence and presence of (b) 10 mM, (c) 20 mM, (d) 40 mM and (e) 60 mM concentration of biphosphate ions in solution.

In presence of HPO_4^{2-} , the conformational preferences of the residues change depending on the anion concentration. In WP10 system, R49, K50, K51, R52, R53, R55 and R57 remain in β conformation and Q54 switches to the LH conformation. R56, P58 and P59 switch from β to RH conformation. In WP20 condition, R49 fluctuates in between RH and β conformations; R53 prefers RH conformation; R55 acquires RH. Both P58 and P59 switch β to RH conformation and there is no change in conformational preference in the other residue. In

WP40 condition, Q54 switch in LH conformation; R55 is in RH conformation, but R56 fluctuates between LH and β conformations and both P58 and P59 are in RH conformation. The conformational preferences of other residues remain same. In WP60, R49 switches to RH conformation. Both P58 and P59 are in RH conformation. The conformational preferences of other residues are not changed. Overall, it is observed that R53, Q54, R55 and R56 have the maximum helicity percentage at different biphosphate ion concentration.

Anion condensation:

The microscopic details of interaction between peptide residues and biphosphate ions are given in terms of biphosphate distributions, namely, number of biphosphate ions around different atoms in the residues within a spherical shell of radii r and $r+\Delta r$, from the centre of mass of biphosphate ion. The major condensation of the anionic biphosphate takes place around side chain nitrogen atoms, as shown in atomistic representation in Figure 4.4(a).

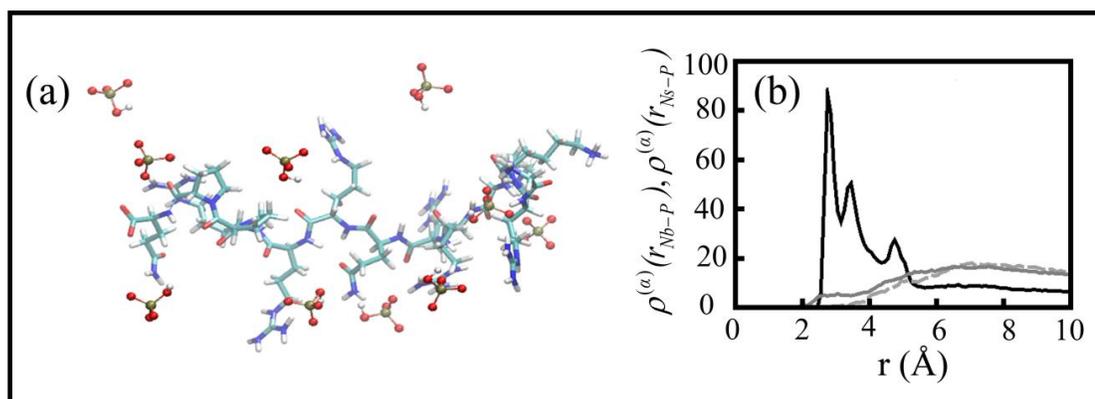


Figure 4.4: (a) Condensation of the anionic biphosphate around side chain nitrogen atoms of peptide residues as found in MD simulation with 20 mM concentration of biphosphate ions in solution. (b) Radial distribution, $\rho^{(\alpha)}(r_{Ns-P})$ and $\rho^{(\alpha)}(r_{Nb-P})$ of P atom of biphosphate ion around side chain guanidinium group (black solid line), side chain NH group (gray dashed line) and backbone N atom (gray solid line) of R52 at 20 mM biphosphate concentration.

Biphosphate distributions around the backbone nitrogen are denoted by $\rho^{(\alpha)}(r_{Nb-P})$ and that around side chain nitrogen atoms by $\rho^{(\alpha)}(r_{Ns-P})$ for the α -th residue. In case of arginine average is taken over both nitrogen atoms in the side chain. Figure 4.4(b) shows that $\rho^{(R52)}(r_{Ns-P})$ has strongly peaked structure near nitrogen atoms of the side chain guanidinium group of R52, whereas $\rho^{(R52)}(r_{Nb-P})$ shows relatively flat profile near the backbone nitrogen atom.

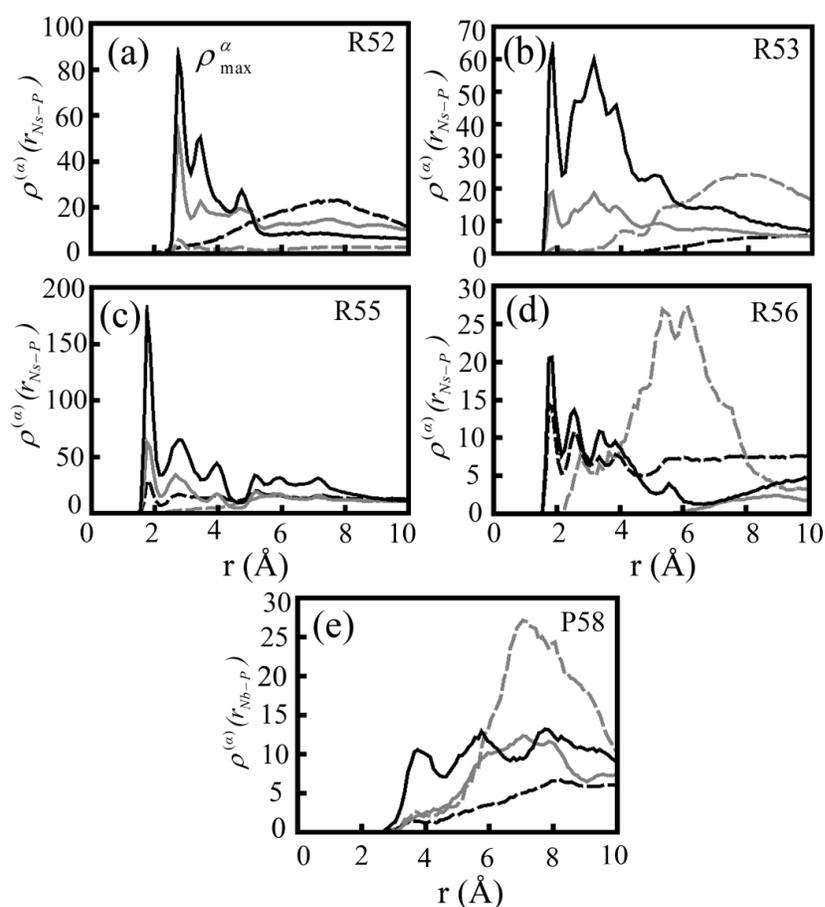


Figure 4.5: (a)-(d) Radial distribution of P atom of biphosphate ion, $\rho^{(\alpha)}(r_{Ns-P})$ around side chain guanidinium group of arginine (R) and (e) $\rho^{(\alpha)}(r_{Nb-P})$ around backbone N atom of proline (P58) at different concentration of biphosphate ions in solution (WP10: gray dashed; WP20: black solid; WP40: gray solid and WP60: black dashed line)

Figure 4.5(a) shows that the $\rho^{(R52)}(r_{Ns-P})$ is sharply peaked in WP20, whereas the peaks are much less pronounced for WP40 and the peaks vanish for both WP10 and WP60. In case of other residues, like R53 (Figure 4.5(b)), R55 (Figure 4.5(c)) and R56 (Figure 4.5(d)) $\rho^{(\alpha)}(r_{Ns-P})$ shows very similar trend with different biphosphate concentrations. Figure 4.5(e) shows the case of the hydrophobic residue P58. Here a peak of relatively lower height in the vicinity of side chain nitrogen is seen only for WP20, whereas in other cases the peaks are quite far from the side chain nitrogen. Figure 4.6 shows additional data for other residues. Figures 4.7 - 4.9 show distributions of biphosphate ion, $\rho^{(\alpha)}(r_{Nb-P})$ and $\rho^{(\alpha)}(r_{Ns-P})$ around different charged and polar peptide residues at different concentrations of biphosphate ions in solution.

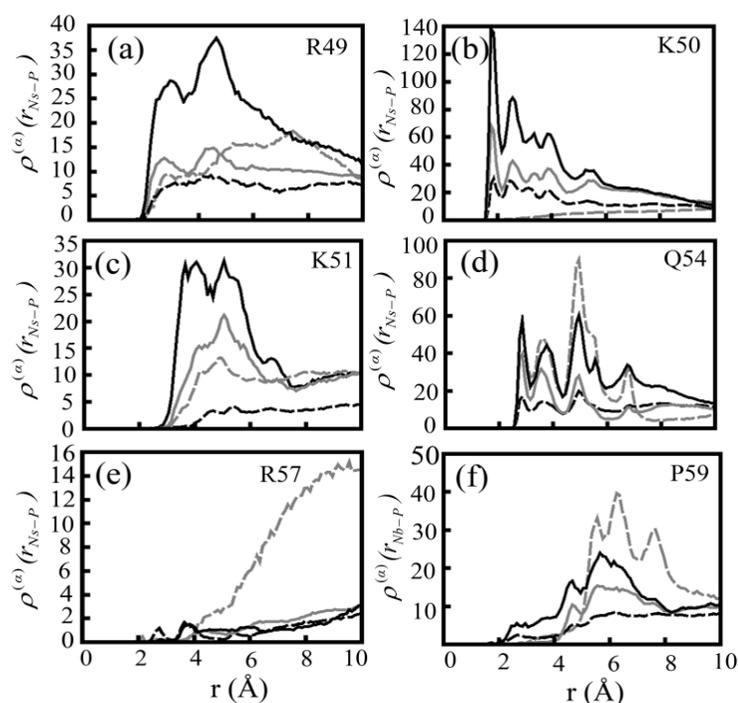


Figure 4.6: Radial distribution, $\rho^{(\alpha)}(r_{Ns-P})$ of P atom of biphosphate ion around side chain guanidinium group of arginine (R), side chain N atom of lysine (K), glutamine (Q54) and backbone N atom of proline (P) at different concentration of biphosphate ions in solution. (Gray dashed: WP10; black solid: WP20; gray solid: WP40 and black dashed: WP60 condition)

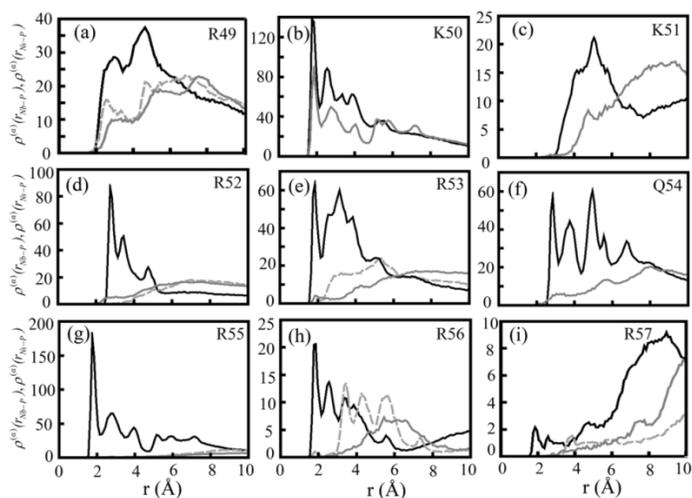


Figure 4.7: Radial distribution of biphosphate ion around backbone N atom, $\rho^{(\alpha)}(r_{Nb-P})$ (gray solid line) and side chain N atoms, $\rho^{(\alpha)}(r_{Ns-P})$ (black solid line: guanidinium group of arginine (R), side chain N atom of lysine (K) and glutamine (Q); gray dashed line: side chain NH group of arginine (R)) of basic and polar residues of Tat peptide in presence of 20 mM biphosphate ions in solution.

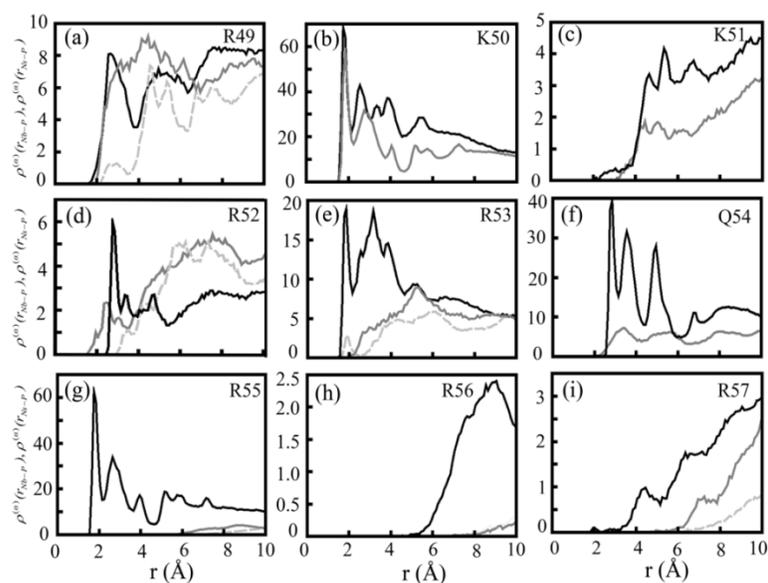


Figure 4.8: Radial distribution of biphosphate ion around backbone N atom, $\rho^{(\alpha)}(r_{Nb-P})$ (gray solid line) and side chain N atoms, $\rho^{(\alpha)}(r_{Ns-P})$ (black solid line: guanidinium group of arginine (R), side chain N atom of lysine (K) and glutamine (Q); gray dashed line: side chain NH group of arginine (R)) of basic and polar residues of Tat peptide in presence of 40 mM biphosphate ions in solution.

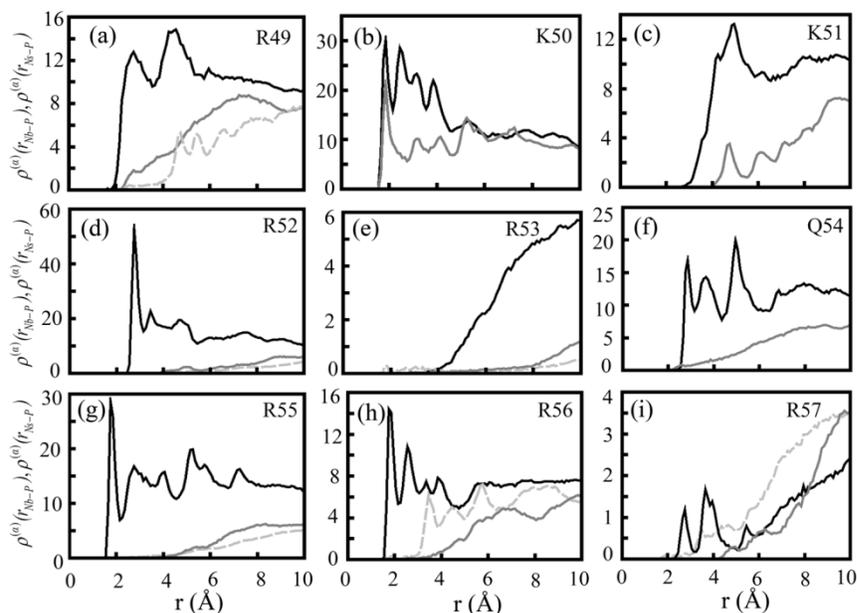


Figure 4.9: Radial distribution of biphosphate ion around backbone N atom, $\rho^{(\alpha)}(r_{Nb-P})$ (gray solid line) and side chain N atoms, $\rho^{(\alpha)}(r_{Ns-P})$ (black solid line: guanidinium group of arginine (R), side chain N atom of lysine (K) and glutamine (Q); gray dashed line: side chain NH group of arginine (R)) of basic and polar residues of Tat peptide in presence of 60 mM biphosphate ions in solution.

Maximum peak intensity of $\rho^{(\alpha)}(r_{Ns-P})$, ρ_{\max}^{α} , marked in Figure 4.5(a) as a function of biphosphate ion concentration is shown in Figure 4.10(a) for arginines, R49, R52, R53, R55 and R56. A non-monotonic dependence of ρ_{\max}^{α} on biphosphate concentration is observed. ρ_{\max}^{α} has maximum at 20 mM concentration of the anions. The data for maximum peak intensity around the side chain nitrogen in case of K50, K51 and Q54, shown in inset Figure 4.10(a), also exhibit similar behavior.

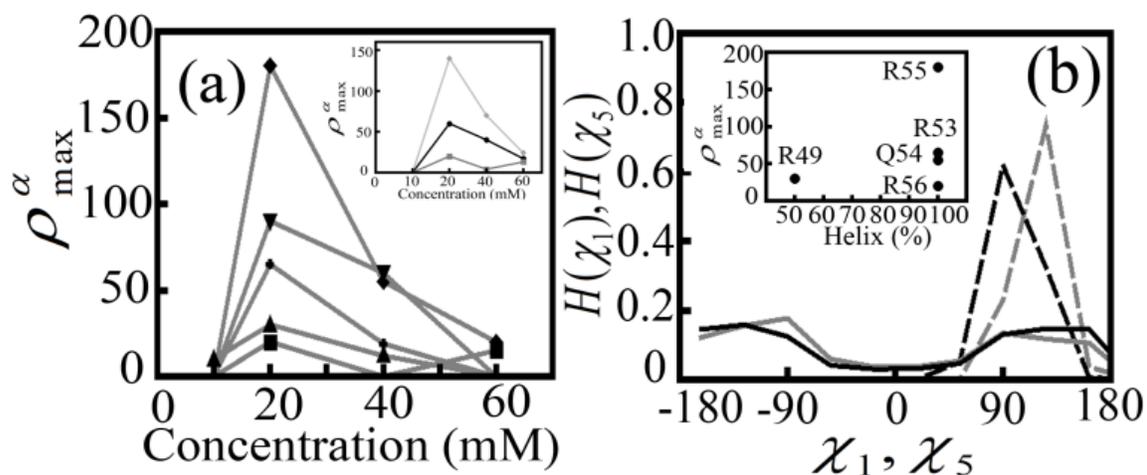


Figure 4.10: (a) Variation of ρ_{\max}^{α} in case of arginine residues (up-pointing triangle: R49, down-pointing triangle: R52, cross: R53, diamond: R55 and square: R56) as shown in main panel and in case of K50 (diamond), K51 (square) and Q54 (sphere) residues of peptide as shown in inset at different concentration of biphosphate ions in solution. (b) Distribution of side chain torsion angle χ_1 and χ_5 in absence (gray solid and gray dashed) and presence (black solid and black dashed) of 20 mM biphosphate ions in solution, and correlation between α -helical propensity of peptide residues and maximum biphosphate ions density around side chain nitrogen group, ρ_{\max}^{α} as shown inset.

The anion concentration in the vicinity of the peptide residues is also sensitive to the α -helix content. R53, Q54, R55 and R56 have maximum helicity percentage at different biphosphate concentration. Inset, Figure 4.10(b) shows that there is a sharp rise in of ρ_{\max}^{α} above a critical value of helical preference. There is a sharp rise in of ρ_{\max}^{α} above a critical value of helical content. The conformation sensitivity can be ascribed to change in side chain fluctuations in different secondary structure. The fluctuations of the side chain dihedrals, χ_1 to χ_5 are considered to this end. The histograms for χ_1 and χ_5 for R53 in WOP and WP20, as exemplary cases, are shown in main panel, Figure 4.10(b). The conformational preference of R53 is β in WOP, while that in WP20 is RH. The histograms in two cases are not so different for χ_1, χ_2 and χ_3 . However, the distributions for dihedrals χ_4 and χ_5 (Figure 4.10(b)) show

differences. The spectrum is broader, suggesting enhanced flexibility of side chains in the helical conformation helps to capture biphosphate ions. The hydrophobic residues P58 and P59 also attain helix conformation. However, the effective hydrophobicity of the side chain does not allow them to have large anion condensation.

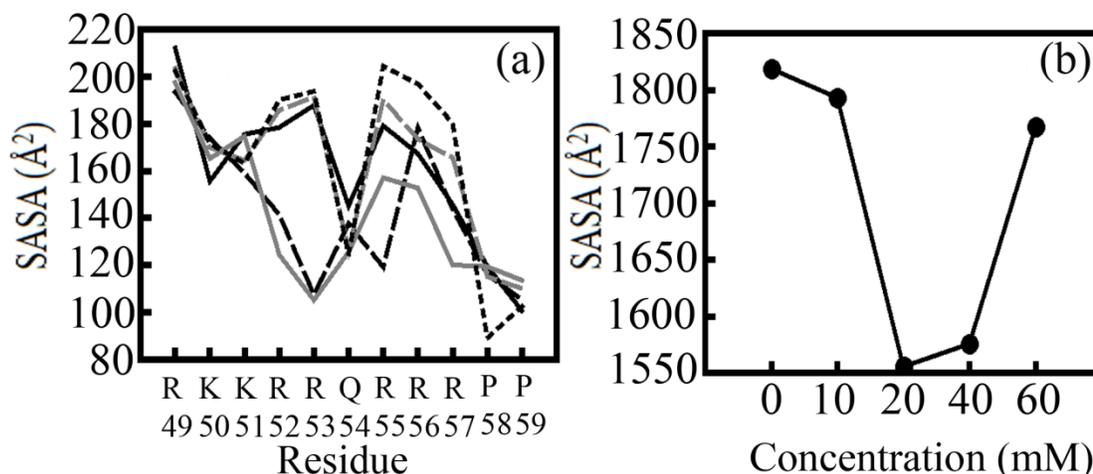


Figure 4.11: (a) Solvent accessible surface area (SASA) of peptide residues in absence (black dashed line) and presence of different concentration of biphosphate ions in solution (10 mM: gray long dashed line, 20 mM: gray solid line, 40 mM: black long dashed line and 60 mM: black solid line). (b) Overall solvent accessible surface area (SASA) of peptide in absence and presence of different concentration of biphosphate ions in solution.

I also calculate the solvent accessible surface area (SASA) of the polypeptide. The residue wise SASA in Figure 4.11(a) shows that the hydrophobic residues behave quite differently from the basic and polar residues. While the latter ones have the least SASA in WP20, the hydrophobic residues P58 and P59 have larger SASA in presence of biphosphate compared to the WOP case. SASA value of arginine residues decreases in WP20 case as compared to the WOP case and increased again in WP60 case. Biphosphate condensation gives hydrophilic character to the hydrophobic residues but shield the basic and polar residues from the solvent to reduce the solvent exposed area. This reduction in solvent exposed area of the

basic and polar residues is sensitive to anion concentration, since the anion condensation itself depends on the anion concentration. This sensitivity is reflected in the overall SASA. In absence of anion overall SASA of the peptide is around 1820 Å². The SASA shows non-monotonic changes with biphosphate ions concentration (Figure 4.11(b)) with minimum in WP20 where the anion condensation is maximum.

4.4 Model Calculation:

All-atom simulations show that the condensation of biphosphate ions around the basic residues is maximum at 20 mM concentration of anion. This observation can be rationalized based on two competitive interactions: attractive interaction between anions and basic residues and repulsion between the anions. Although the interactions are of electrostatic nature, a model system is considered with simple form of competitive interactions. Model system consists of free spherical particles in the presence of a model bead-spring chain, mimicking the anions and polypeptide respectively. The beads, having repulsion among themselves, represent sites of attraction for the free particles. The free spherical particles repel themselves. The springs connecting the beads has elastic energy cost due to stretching and bending, representing polypeptide backbone.

The details of model interactions are as follows. In the model system the non-bonded interaction energy between the beads is given by:

$$U_{bb}(r) = \infty, \text{ for } r \leq \sigma,$$

$$= -4\epsilon \left(\frac{\sigma}{r}\right)^6, \text{ otherwise.}$$

Harmonic potentials are considered for stretching and bending deformations of polymer chain associated with the bonded interactions among the beads via spring. The bonded interaction energy between beads is

$$U_{bb}(r, \theta) = \frac{1}{2} K_{bond} (r - r_0)^2 + \frac{1}{2} K_{ang} (\theta - \theta_0)^2, \text{ where}$$

r = length of the spring

θ = angle between bonds connecting three successive beads

r_0 = equilibrium spring length = 1

θ_0 = equilibrium bond angle = 0

The interactions between the free particles are hard sphere. The interactions between bead-particle are:

$$U_{bp}(r) = \infty, \text{ for } r \leq \sigma$$

$$= -4\varepsilon \left(\frac{\sigma}{r}\right)^6, \text{ otherwise}$$

The particle diameter is the length unit and energy unit $k_B T$ at room temperature ($T = 300$ K).

Here $K_{bond} \sigma^2 / k_B T = K_{ang} / k_B T = \varepsilon / k_B T = 1.0$. The beads and the particles are taken to have the same size (1 nm) which is length unit and the thermal energy at room temperature the energy unit in our calculations. The Monte Carlo (MC) simulations (see Appendix A4.1 for details) for model system are run for 500,000 MC steps. Equilibration has been checked from energy of the system and last equilibrated 400,000 configurations are considered for evaluate different quantities.

MC simulations are performed with different volume fraction of free particles (η_0) with $N (=100)$ beads in a cubic box of dimension 30 nm with the periodic boundary conditions on all three directions. η is computed by binning the distances between the beads and free particles averaged over equilibrium configurations. The data for different η_0 are shown in Figure 4.12(a). There is a prominent first peak in η indicating condensation of the free particles around the

beads. The peak intensity at the first maximum of the distribution, η_s has a maximum around $\eta_0 = 0.005$ (Figure 4.12(b)). This is qualitatively similar to the all-atom simulation result.

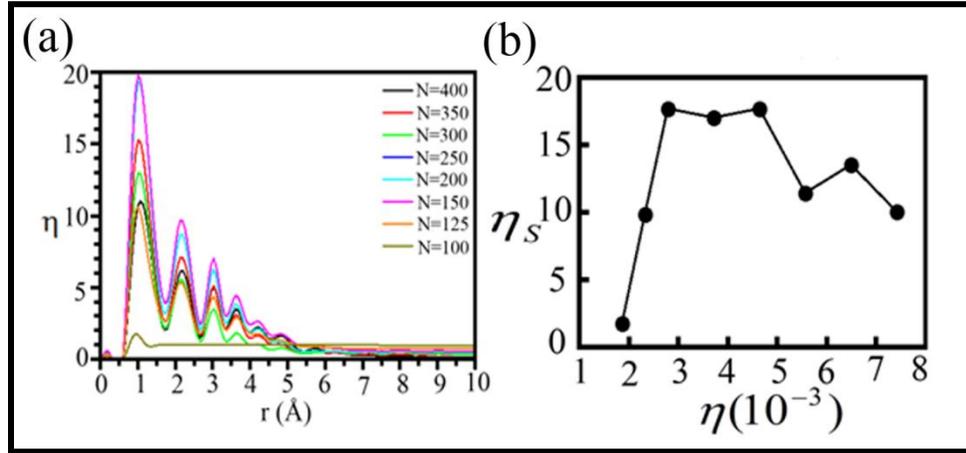


Figure 4.12: (a) Distribution of number of free particles around a polymer bead (η) for different number of free particles (N) in the system. (b) Variation of η_s with η_0 in model calculation.

The existence of maximum of η_s a function of η_0 can be understood also from free energy cost of placing the particles in the vicinity of a bead. The free energy of the particles reduces by coming in the vicinity of the bead due to attraction. However, this enhances repulsion between the particles. The free energy change can be given as

$$F/k_B T = -\frac{\varepsilon}{k_B T} \eta_s + \frac{1}{2\chi_T} \eta_s^2$$

at a temperature T . The first term account for reduction in

energy due to bead attraction of magnitude ε . The second term accounts for the repulsion between the particles via compressibility, χ_T . The repulsive interaction will hinder the particles to get compressed in order to fit themselves in the vicinity of the bead. The minimum of the free energy is given by, $\eta_s = \chi_T \varepsilon$. One can estimate χ_T by using liquid state theories⁵⁹,

$$\chi_T = \frac{\eta_0}{k_B T} \frac{1}{1 - \eta_0 C_0}$$

Here C_0 is given by the integral of liquid direct correlation function for

hard sphere particles. I find that η_s has a maximum around $\eta_0 = 0.4$ (Figure 4.13) for

$\frac{\varepsilon}{k_B T} = 1.0$ (see Appendix A4.2 for details of calculations). Since all the beads attract the

particles simultaneously, the maximum will scale down the optimum density by N , qualitatively accounting for observations of the model simulations.

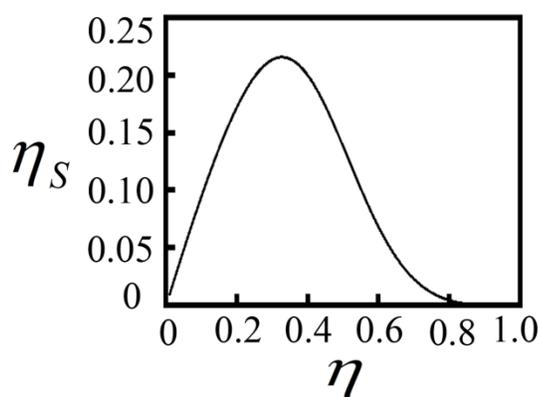


Figure 4.13: Variation of η_s with η according to analytical calculation.

4.5 Discussion:

The counter-anion condensation attains a maximum at an optimum biphosphate concentration (20 mM). Experiments show phase transfer of poly-arginine from water into chloroform in the presence of mixture of amphiphilic anions, like monomeric sodium dodecyl sulfate (SDS), egg yolk phosphatidylglycerol (EYPG), cholesterol sulfate, pyrene-butyrates, and stearate³⁰ along with phosphate ions at 10 mM concentration. The presence of the other anions would further increase the overall charge concentration leading to concentration similar to the optimum concentration in my studies. The condensation of counter-anion is also found to correlate with helical conformational property of the peptide backbone. Helical conformation has been identified in case of many trans-membrane proteins and AMPs, such as

alamethicin, magainin, melittin and ovispirin¹⁶⁻²⁰ in hydrophobic environment within a lipid bilayer. The increase in counterion condensation is concomitant with decrease in SASA. This may be interpreted as neutralization of charges at the basic residues by the counter-anions, leading to effective hydrophobicity of the polypeptide and may be a driving factor for penetration into non-polar medium, as prevailing in lipid bilayer. This may be the microscopic picture of “arginine magic”. The side chain nitrogen atoms of basic residue lysine and polar residue glutamine also behave in similar way. However, the hydrophobic residues show hydrophilic character. Anion concentration can control solvent exposure of the peptide as a whole with minimum exposure at an optimal concentration. Such control may be suitable for designing carrier for drug molecules into cellular environment¹⁹. Proposed model generalizes the existence of the maximum in condensation, observed in the all-atom simulations, in terms competition between attractive and repulsive interactions.

4.6 Conclusion:

To conclude, all-atom simulations on HIV-1 Tat polypeptide in presence of biphosphate ions in a solution show that anions condense around the side chain nitrogen atoms in different basic functional groups of the residues. The condensation is maximum at an optimum biphosphate concentration in solution. The anions shield solvent exposure of the polypeptide so that the solvent accessible area of the polypeptide is minimum when the condensation is maximum. The anion condensation changes conformational propensity of the residues. These aspects may be relevant for therapeutic applications. Simplified statistical mechanics-based model study suggests that the maximum condensation at an optimum solvent condition is generic. This study may be helpful to use such cationic polypeptides as therapeutic

Chapter 4

agents, like drug carriers, and understand competitive ligand binding quite common place in biological systems⁶⁰⁻⁶¹.

References:

1. Nordmeier, E. Phenomena of counterion binding, polyion complexation and polyion condensation. *Macromol. Symp.* **1999**, *145*, 21-37.
2. Murthy, V. L.; Rose, G. D. Is counterion delocalization responsible for collapse in RNA folding? *Biochemistry* **2000**, *39* (47), 14365-14370.
3. Locher, C. P.; Putnam, D.; Langer, R.; Witt, S. A.; Ashlock, B. M.; Levy, J. A. Enhancement of a human immunodeficiency virus env DNA vaccine using a novel polycationic nanoparticle formulation. *Immunol. Lett.* **2003**, *90* (2-3), 67-70.
4. Elmes, R. B. P.; Jolliffe, K. A. Anion recognition by cyclic peptides. *Chem. Commun.* **2015**, *51* (24), 4951-4968.
5. Gale, P. A.; Busschaert, N.; Haynes, C. J. E.; Karagiannidis, L. E.; Kirby, I. L. Anion receptor chemistry: highlights from 2011 and 2012. *Chem. Soc. Rev.* **2014**, *43* (1), 205-241.
6. Cha, M.; Ro, S.; Kim, Y. W. Rodlike Counterions near Charged Cylinders: Counterion Condensation and Intercylinder Interaction. *Phys. Rev. Lett.* **2018**, *121* (5), 058001.
7. Lo, T. S.; Khusid, B.; Koplik, J. Dynamical clustering of counterions on flexible polyelectrolytes. *Phys. Rev. Lett.* **2008**, *100* (12), 128301.
8. Tom, A. M.; Vemparala, S.; Rajesh, R.; Brilliantov, N. V. Mechanism of Chain Collapse of Strongly Charged Polyelectrolytes. *Phys. Rev. Lett.* **2016**, *117* (14), 147801.
9. Bianco, V.; Locatelli, E.; Maggaretti, P. Globulelike Conformation and Enhanced Diffusion of Active Polymers. *Phys. Rev. Lett.* **2018**, *121* (21), 217802.
10. Hayes, R. L.; Noel, J. K.; Mandic, A.; Whitford, P. C.; Sanbonmatsu, K. Y.; Mohanty, U.; Onuchic, J. N. Generalized Manning Condensation Model Captures the RNA Ion Atmosphere. *Phys. Rev. Lett.* **2015**, *114* (25), 258105.

11. Losasso, V.; Hsiao, Y. W.; Martelli, F.; Winn, M. D.; Crain, J. Modulation of Antimicrobial Peptide Potency in Stressed Lipid Bilayers. *Phys. Rev. Lett.* **2019**, *122* (20), 208103.
12. Tesei, G.; Vazdar, M.; Jensen, M. R.; Cragnell, C.; Mason, P. E.; Heyda, J.; Skepo, M.; Jungwirth, P.; Lund, M. Self-association of a highly charged arginine-rich cell-penetrating peptide. *Proc. Natl. Acad. Sci. USA* **2017**, *114* (43), 11428-11433.
13. Andrade, A. C.; Van Nistelrooy, J. G.; Peery, R. B.; Skatrud, P. L.; De Waard, M. A. The role of ABC transporters from *Aspergillus nidulans* in protection against cytotoxic agents and in antibiotic production. *Mol. Gen. Genet.* **2000**, *263* (6), 966-77.
14. Angel, C. S.; Ruzek, M.; Hostetter, M. K. Degradation of C3 by *Streptococcus pneumoniae*. *J. Infect. Dis.* **1994**, *170* (3), 600-8.
15. Herce, H. D.; Garcia, A. E. Molecular dynamics simulations suggest a mechanism for translocation of the HIV-1 TAT peptide across lipid membranes. *Proc. Natl. Acad. Sci. USA* **2007**, *104* (52), 20805-20810.
16. Chauhan, A.; Tikoo, A.; Kapur, A. K.; Singh, M. The taming of the cell penetrating domain of the HIV Tat: myths and realities. *J. Contro. Release* **2007**, *117* (2), 148-62.
17. Henriques, S. T.; Melo, M. N.; Castanho, M. A. How to address CPP and AMP translocation? Methods to detect and quantify peptide internalization in vitro and in vivo. *Mol. Membr. Biol.* **2007**, *24* (3), 173-84.
18. Pooga, M.; Langel, U. Synthesis of cell-penetrating peptides for cargo delivery. *Methods Mol. Biol.* **2005**, *298*, 77-89.
19. Wadia, J. S.; Dowdy, S. F. Transmembrane delivery of protein and peptide drugs by TAT-mediated transduction in the treatment of cancer. *Adv. Drug Deliv. Rev.* **2005**, *57* (4), 579-

96.

20. Al-Taei, S.; Penning, N. A.; Simpson, J. C.; Futaki, S.; Takeuchi, T.; Nakase, I.; Jones, A. T. Intracellular traffic and fate of protein transduction domains HIV-1 TAT peptide and octaarginine. Implications for their utilization as drug delivery vectors. *Bioconjug. Chem.* **2006**, *17* (1), 90-100.

21. Quan, X.; Sun, D.; Zhou, J. Molecular mechanism of HIV-1 TAT peptide and its conjugated gold nanoparticles translocating across lipid membranes. *Phys. Chem. Chem. Phys.* **2019**, *21* (20), 10300-10310.

22. Wang, G.; Li, X.; Wang, Z. APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res.* **2009**, *37* (Database issue), D933-7.

23. Wang, Z.; Wang, G. APD: the Antimicrobial Peptide Database. *Nucleic Acids Res.* **2004**, *32* (Database issue), D590-2.

24. Riordan, J. F. Arginyl Residues and Anion Binding-Sites in Proteins. *Mol. Cell Biochem.* **1979**, *26* (2), 71-92.

25. Sakai, N.; Sorde, N.; Das, G.; Perrottet, P.; Gerard, D.; Matile, S. Synthetic multifunctional pores: deletion and inversion of anion/cation selectivity using pM and pH. *Org. Biomol. Chem.* **2003**, *1* (7), 1226-31.

26. Lee, S.; Tanaka, T.; Anzai, K.; Kirino, Y.; Aoyagi, H.; Sugihara, G. Two mode ion channels induced by interaction of acidic amphipathic alpha-helical peptides with lipid bilayers. *Biochim. Biophys. Acta* **1994**, *1191* (1), 181-9.

27. Iwata, T.; Lee, S.; Oishi, O.; Aoyagi, H.; Ohno, M.; Anzai, K.; Kirino, Y.; Sugihara, G. Design and synthesis of amphipathic 3(10)-helical peptides and their interactions with phospholipid bilayers and ion channel formation. *J. Biol. Chem.* **1994**, *269* (7), 4928-33.

28. Agawa, Y.; Lee, S.; Ono, S.; Aoyagi, H.; Ohno, M.; Taniguchi, T.; Anzai, K.; Kirino, Y. Interaction with phospholipid bilayers, ion channel formation, and antimicrobial activity of basic amphipathic alpha-helical model peptides of various chain lengths. *J. Biol. Chem.* **1991**, *266* (30), 20218-22.
29. Anzai, K.; Hamasuna, M.; Kadono, H.; Lee, S.; Aoyagi, H.; Kirino, Y. Formation of ion channels in planar lipid bilayer membranes by synthetic basic peptides. *Biochim. Biophys. Acta* **1991**, *1064* (2), 256-66.
30. Sakai, N.; Matile, S. Anion-mediated transfer of polyarginine across liquid and bilayer membranes. *J. Am. Chem. Soc.* **2003**, *125* (47), 14348-56.
31. Schmidtchen, F. P.; Berger, M. Artificial Organic Host Molecules for Anions. *Chem. Rev.* **1997**, *97* (5), 1609-1646.
32. Snowden, T. S.; Anslyn, E. V. Anion recognition: synthetic receptors for anions and their application in sensors. *Curr. Opin. Chem. Biol.* **1999**, *3* (6), 740-6.
33. Wiskur, S. L.; Ait-Haddou, H.; Lavigne, J. J.; Anslyn, E. V. Teaching old indicators new tricks. *Acc. Chem. Res.* **2001**, *34* (12), 963-72.
34. Peczu, M. W.; Hamilton, A. D. Peptide and protein recognition by designed molecules. *Chem. Rev.* **2000**, *100* (7), 2479-94.
35. Austin, R. J.; Xia, T.; Ren, J.; Takahashi, T. T.; Roberts, R. W. Designed arginine-rich RNA-binding peptides with picomolar affinity. *J. Am. Chem. Soc.* **2002**, *124* (37), 10966-7.
36. Rooman, M.; Lievin, J.; Buisine, E.; Wintjens, R. Cation-pi/H-bond stair motifs at protein-DNA interfaces. *J. Mol. Biol.* **2002**, *319* (1), 67-76.
37. Tyagi, M.; Rusnati, M.; Presta, M.; Giacca, M. Internalization of HIV-1 tat requires cell surface heparan sulfate proteoglycans. *J. Biol. Chem.* **2001**, *276* (5), 3254-61.

38. General, S.; Thunemann, A. F. pH-sensitive nanoparticles of poly(amino acid) dodecanoate complexes. *Int. J. Pharm.* **2001**, *230* (1-2), 11-24.
39. Ariga, K.; Kunitake, T. Molecular recognition at air-water and related interfaces: Complementary hydrogen bonding and multisite interaction. *Acc. Chem. Res.* **1998**, *31* (6), 371-378.
40. Carnal, F.; Stoll, S. Explicit ions condensation around strongly charged polyelectrolytes and spherical macroions: the influence of salt concentration and chain linear charge density. Monte Carlo simulations. *J. Phys. Chem. A* **2012**, *116* (25), 6600-8.
41. Chremos, A.; Douglas, J. F. Counter-ion distribution around flexible polyelectrolytes having different molecular architecture. *Soft Matter* **2016**, *12* (11), 2932-41.
42. Muthukumar, M. Theory of counter-ion condensation on flexible polyelectrolytes: adsorption mechanism. *J. Chem. Phys.* **2004**, *120* (19), 9343-50.
43. Clavier, A.; Carnal, F.; Stoll, S. Effect of Surface and Salt Properties on the Ion Distribution around Spherical Nanoparticles: Monte Carlo Simulations. *J. Phys. Chem. B* **2016**, *120* (32), 7988-97.
44. Hocine, S.; Hartkamp, R.; Siboulet, B.; Duvail, M.; Coasne, B.; Turq, P.; Dufreche, J. F. How Ion Condensation Occurs at a Charged Surface: A Molecular Dynamics Investigation of the Stern Layer for Water-Silica Interfaces. *J. Phys. Chem. C* **2016**, *120* (2), 963-973.
45. Carrillo, J. M. Y.; Dobrynin, A. V. Molecular dynamics simulations of polyelectrolyte adsorption. *Langmuir* **2007**, *23* (5), 2472-2482.
46. Frankel, A. D.; Pabo, C. O. Cellular uptake of the tat protein from human immunodeficiency virus. *Cell* **1988**, *55* (6), 1189-93.
47. Ruben, S.; Perkins, A.; Purcell, R.; Joung, K.; Sia, R.; Burghoff, R.; Haseltine, W. A.;

Rosen, C. A. Structural and functional characterization of human immunodeficiency virus tat protein. *J. Virol.* **1989**, *63* (1), 1-8.

48. Fawell, S.; Seery, J.; Daikh, Y.; Moore, C.; Chen, L. L.; Pepinsky, B.; Barsoum, J. Tat-mediated delivery of heterologous proteins into cells. *Proc. Natl. Acad. Sci. USA* **1994**, *91* (2), 664-8.

49. Vives, E.; Brodin, P.; Lebleu, B. A truncated HIV-1 Tat protein basic domain rapidly translocates through the plasma membrane and accumulates in the cell nucleus. *J. Biol. Chem.* **1997**, *272* (25), 16010-7.

50. Futaki, S.; Suzuki, T.; Ohashi, W.; Yagami, T.; Tanaka, S.; Ueda, K.; Sugiura, Y. Arginine-rich peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery. *J. Biol. Chem.* **2001**, *276* (8), 5836-40.

51. Peloponese, J. M., Jr.; Gregoire, C.; Opi, S.; Esquieu, D.; Sturgis, J.; Lebrun, E.; Meurs, E.; Collette, Y.; Olive, D.; Aubertin, A. M.; Witvrow, M.; Pannecouque, C.; De Clercq, E.; Bailly, C.; Lebreton, J.; Loret, E. P. ¹H-¹³C nuclear magnetic resonance assignment and structural characterization of HIV-1 Tat protein. *C. R. Acad. Sci.* **2000**, *323* (10), 883-94.

52. Lehninger Albert L, D. L. N. a. M. M. C. *Lehninger Principles of Biochemistry*. 7th ed.; Worth Publisher: New York, **2000**.

53. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435-47.

54. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26* (16), 1701-18.

55. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.;

Chapter 4

Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78* (8), 1950-8.

56. Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33-8, 27-8.

57. Eisenhaber, F.; Lijnzaad, P.; Argos, P.; Sander, C.; Scharf, M. The Double Cubic Lattice Method - Efficient Approaches to Numerical-Integration of Surface-Area and Volume and to Dot Surface Contouring of Molecular Assemblies. *J. Comput. Chem.* **1995**, *16* (3), 273-284.

58. M.P. Allen, D. J. T. *Computer Simulation of Liquids* Oxford University Press: **1991**.

59. Bergersen, M. P. a. B. *Equilibrium Statistical Physics*. 3rd ed.; World Scientific: **2006**.

60. Du, X.; Li, Y.; Xia, Y. L.; Ai, S. M.; Liang, J.; Sang, P.; Ji, X. L.; Liu, S. Q. Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* **2016**, *17* (2).

61. Csermely, P.; Palotai, R.; Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem. Sci.* **2010**, *35* (10), 539-46.

Appendices of Chapter 4

A4.1 Metropolis algorithm & Monte Carlo simulation:

Monte-Carlo (MC) method allows one to evaluate the average properties of a system by generating a sequence of configurations (i.e., locations of all of the molecules in the system as well as of all the internal coordinates of these molecules) and assigning a weighting factor to these configurations⁵⁸. By introducing an especially efficient way to generate configurations that have high weighting, the MC method allows to simulate extremely complex systems that may contain millions of molecules. In MC process, one must assume that the total energy E does not include the kinetic energy of the molecules; it is only the (intramolecular and intermolecular) potential energy of the system. Usually, this energy E is expressed as a sum of intra-molecular bond-stretching and bending contributions, one for each molecule, plus a pair-wise additive intermolecular potential. However, the energy E could be computed in other ways, if appropriate because this process does not depend on how E is computed. In each “step” of the MC process, this potential energy E is evaluated for the current positions of the molecules of the system⁵⁸.

The Metropolis algorithm is generally adopted in the Monte Carlo simulation method⁵⁸. The Metropolis algorithm generates a Markov chain of states. A Markov chain satisfies the following two conditions:

1. The outcome of each trial depends only upon the preceding trial and not upon any previous trials.
2. Each trial belongs to a finite set of possible outcomes.

The equilibrium distribution for a molecular or atomic system is one in which the probabilities of each state are proportional to the Boltzmann factor. Briefly, the Metropolis algorithm

generates a trial X_{i+1} from X_i by a system-appropriate random perturbation and accepts that state if the corresponding energy is lower. If, however $E(X_{i+1}) > E(X_i)$, then the new state is accepted with probability $P = \exp(-\beta\Delta E)$, where $\Delta E = E(X_{i+1}) - E(X_i) > 0$, by comparing P to a uniformly generated number on $(0,1)$: if $P > \text{ran}$, accept X_{i+1} , and if $P \geq \text{ran}$, generate another trial X_{i+1} but recount X_i in the Markov chain. The result of this procedure is the acceptance probability at step i of:

$$\begin{aligned} P_{\text{acc, MC}} &= \min [1, \exp(-\beta\Delta E)] \\ &= \min [1, \rho(X_{i+1}) / \rho(X_i)] \end{aligned}$$

In this manner, states with lower energies are always accepted but states with higher energies have a nonzero probability of acceptance too. The perturbation by which coordinates are varied is usually chosen to make the fraction of MC steps that are accepted approximately 50% of the times of attempt. In this way, one generates a sequence of “accepted moves” that generate a series of configurations for the system of molecules in the system. This set of configurations has been shown to be properly representative of the geometries that the system will experience as it moves around at equilibrium at the specified temperature T . As the series of accepted steps is generated, one can keep track of various geometrical and energetic data for each accepted configuration. The MC procedure thus allows us to generate an ensemble of configurations and compute the equilibrium average of any property of a system.

A4.2 Calculation of η_s :

According to liquid state theory⁵⁹,

$$\frac{\chi_T}{\chi_T^0} = \frac{1}{1 - \eta C_0}$$

C_0 is given by the integral $\int C(x)dx$, where $C(x)$ is the liquid direct correlation function⁵⁹.

For hard sphere particles, $C(x)$ can be written as

$$C(x) = \lambda_1 - 6\eta\lambda_2 x^2,$$

$$\text{where } \lambda_1 = \frac{(1+2\eta)^2}{(1-\eta)^4} \text{ and } \lambda_2 = \frac{-(1+1/2\eta)^2}{(1-\eta)^4}.$$

Inserting the Percus Yevick (PY) for $C(x)$ ⁵⁹ I find,

$$\begin{aligned} C_0 &= -\lambda_1/3 - 6\eta\lambda_2/4 - \eta\lambda_1/12 \\ &= -\frac{(1+2\eta)^2}{3(1-\eta)^4} + \frac{3\eta(1+1/2\eta)^2}{2(1-\eta)^4} - \frac{\eta(1+2\eta)^2}{12(1-\eta)^4} \\ &= \frac{1}{(1-\eta)^4} \left[\frac{9\eta(2+\eta)^2 - 2(1+2\eta)(4+\eta)}{24} \right] \\ &= \frac{1}{24(1-\eta)^4} [\eta^3 + 2\eta - 4\eta^2 - 8] \end{aligned}$$

This yield,

$$\begin{aligned} 1 - \eta C_0 &= 1 - \frac{\eta}{24(1-\eta)^4} [\eta^3 + 2\eta - 4\eta^2 - 8] \\ \eta_s &= \frac{\eta}{1 - \eta C_0} = \frac{24\eta(1-\eta)^4}{24(1-\eta)^4 - (\eta^4 + 2\eta^2 - 4\eta^3 - 8\eta)} \\ &= \frac{24\eta(1-\eta)^4}{24(1-4\eta+6\eta^2-4\eta^3+\eta^4) - (\eta^4 + 2\eta^2 - 4\eta^3 - 8\eta)} \\ &= \frac{24\eta(1-\eta)^4}{(24-88\eta+142\eta^2-92\eta^3+23\eta^4)} \end{aligned}$$



CHAPTER 5

INTERACTION OF CATIONIC POLYPEPTIDE WITH LIPID BILAYER



5.1 Introduction:

Cellular uptake of biologically active molecules is the major challenge in controlled drug delivery. Cell penetrating peptides (CPPs) have potential applications for targeted drug delivery due to their capability of translocation across the cell membrane¹⁻³. The charged domains of the peptide allow interaction with hydrophilic head groups of the phospholipids, while the hydrophobic domains of the peptide interact with the hydrophobic core of the lipid bilayer, thereby driving the peptide deeper into the membrane³⁻⁴. Despite several works to date, the molecular mechanism by which the counter-anion help arginine rich CPPs to traverse the membrane barrier has not been established.

Several mechanisms have been proposed to describe translocation of CPPs⁵⁻¹³ through lipid bilayers. Among the CPPs, the arginine rich Tat peptide have been the most widely studied^{8, 10, 12-17}. Several studies suggest that the peptide can directly pass through membrane via a temporary pore^{8, 10-12}. The mechanism of direct penetration is seen plausible due to the action of related antimicrobial peptides, which are also charged, but in addition contain a large fraction of hydrophobic residues^{5-6, 18}. These peptides are found to disrupt the bilayer and form pores in membranes. Previous biophysical studies suggest that polyarginines is able to deform the membrane and induce phase transition in lipid systems^{8-9, 19}. Membrane composition with negatively charged lipids facilitates translocation of cationic peptides through lipid bilayer^{11, 20}. Tat and its derivatives are found to aggregate at phospholipid membranes and occasionally fuse vesicles^{9, 14-16}. Recent study proposes a passive cell entry mechanism which is completely different from previously suggested direct translocation mechanisms¹³. Passive entry mechanism of arginine rich CPPs is associated with the branching and layering of membranes

that leads to bifurcate the lipid bilayer structure to multilamellar structures, and allows the peptide to enter by fusion¹³.

It is found that counter-anion mediated inversion of charge and solubility of the CPPs can modulate the translocation property of the polyarginines through lipid bilayer²¹⁻²⁸. Poly- and hexa-arginine is found to phase transferred from water into chloroform in the presence of amphiphilic anions. Refined combinations of hydrophilic anions, like phosphate, trifluoroacetic acid (TFA), heparin inhibited phase transfer of 5(6)-carboxy-fluorescein (CF)-polyarginine complexes into and across lipid bilayer membranes²⁷. Therefore, charge neutralization by counter-anion scavenging can increase the lipophilicity of polyarginine and modulate the translocation behavior of oligo- and polyarginines. Previous molecular dynamics simulation studies proposed that arginine rich peptide can directly pass through the membrane via a temporary pore^{10-12,17}. However recent theoretical study considering different lipid compositions CPP is found to stabilize the pore rather than driving pore formation directly. CPPs can create lipid interface perturbation but fails to form complete pores unless external stress is applied²⁹.

In the previous chapter, I have considered the interaction of biphosphate anions with the Tat peptide. It is observed that anionic biphosphate ions in solution are condensed around the guanidinium head group of arginine residues and shield solvent exposure of the Tat peptide. The condensation attains maximum, solvent exposure being minimum at an optimum biphosphate concentration in the solvent. Here I study the influence of anionic biphosphate on the interaction of Tat peptide with lipid bilayer by all-atom molecular dynamics simulations. I show that condensation of biphosphate ions around the side chain of basic residues alters interaction between Tat peptide and negatively charged head group of the DPPC bilayer which induces stress and leads to uneven bilayer thickness. The chapter is organized as follows:

material & methods in section 5.2, results in section 5.3, discussion in section 5.4 and conclusion in section 5.5.

5.2 Material & methods:

5.2.1 Modeling of peptide-lipid system:

The initial coordinates of the Tat peptide are taken from the NMR structure of the HIV-1 Tat protein (PDB code 1JFW³⁰, model 1) having sequence 48GRKKRRQRRRPPQ60 as mentioned in chapter 4. The initial configuration of the DPPC bilayer with 128 lipids is generated using CHARMM-GUI³¹. I model three different systems: only lipid bilayer in water (LB), lipid bilayer along with Tat peptide but no biphosphate (LBP-WOP), and lipid bilayer along with Tat peptide and 20 mM biphosphate (HPO_4^{2-}) ions (LBP-WP) in solution (Figure 5.1).

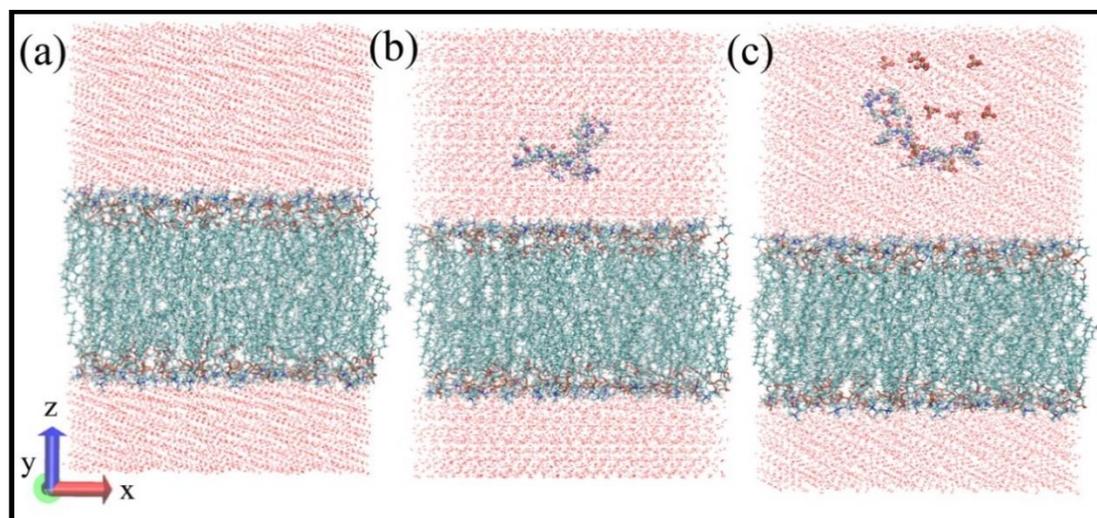


Figure 5.1: Modeled systems considered for MD simulation study: (a) lipid bilayer (LB), (b) lipid bilayer along with Tat peptide (LBP-WOP), and (c) lipid bilayer along with Tat peptide and 20 mM biphosphate (HPO_4^{2-}) ions (LBP-WP) in solution.

Amber Lipid14³² force field parameters are used to describe the DPPC lipid bilayer and the Tat peptide is described with the Amber99ILDN³³ force field. Tat peptide and biphosphate ions are placed in the upper part of the DPPC bilayer. The systems are solvated with water molecules in a 6.3 nm × 6.3 nm × 12.2 nm rectangular simulation box with periodic boundary condition in three dimensions. No water molecules are present in the hydrophobic core of the DPPC bilayer. They are present above and below the DPPC bilayer. The simple point charge (SPC)³⁴ model is used to model water. Required numbers of Na⁺ ions are added for neutralizing the charged system and 20 mM biphosphate (HPO₄²⁻) ions concentration is considered in LBP-WP case.

5.2.2 Molecular Dynamics (MD) simulation:

All-atom MD simulations of the three different systems are performed using GROMACS simulation tool³⁵⁻³⁶. Each system is relaxed through energy minimization. The long ranged electrostatic interactions are treated by the particle-mesh Ewald method. For equilibration of the systems a short NVT equilibration phase is followed by a longer NPT phase. After equilibration, each system is subjected to 1 microsecond MD run in NPT ensemble using 2 femto-second integration time step with periodic boundary conditions over the simulation box at the desired temperature and pressure (1 Bar). All simulations are performed at a temperature of 323 K which is above the main phase transition temperature of the DPPC bilayer (314 K)³⁷ at ambient pressure. Conformations in each MD run are saved every 1 picosecond from the trajectories for further analysis. The trajectories are visualized with visual molecular dynamics (VMD)³⁸. Equilibration of the studied systems is ensured from the root-mean-square displacement (RMSD) of the peptide and lipid atoms with respect to its initial

energy minimized structure and the equilibrated trajectories are used for further analysis (Figure 5.2). I consider primarily the upper leaflet of the lipid bilayer exposed to the peptide.

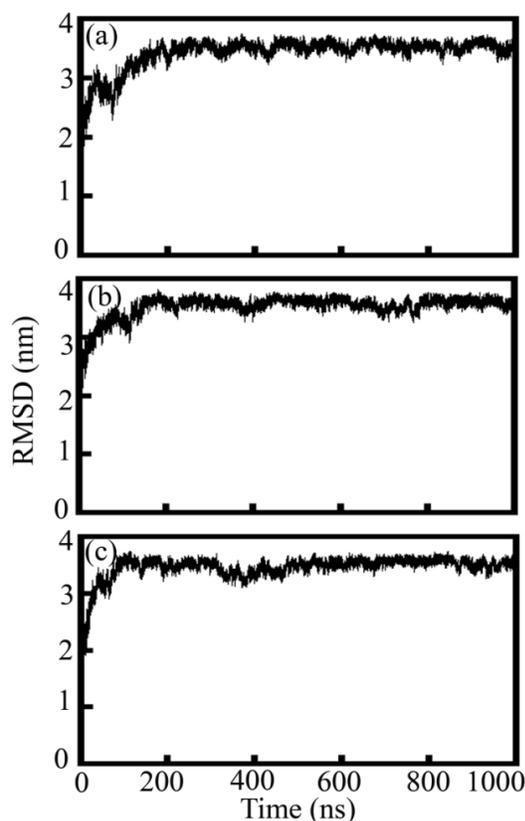


Figure 5.2: Root-mean-square deviations (RMSD) of peptide and lipid atoms with respect to the initial energy minimized structure in (a) LB (b) LBP-WOP, and (c) LBP-WP system.

5.2.3 Theoretical analysis:

Distribution function:

The distribution of anionic biphosphate around the peptide fragment is analyzed using distribution function $\rho_p^{(a)}(r)$ of P atom of biphosphate molecule with respect to the side chain N atom of α -th amino acid residue in the peptide fragment. The distribution function $\rho_p^{(a)}(r)$ is defined as: $\rho_p^{(a)}(r) = \langle N(r) \rangle / (4\pi N \rho \Delta r)$, where $\langle N(r) \rangle$ is the number of P atom of biphosphate molecules averaged over time, within a distance $r \pm \Delta r/2$ from the atom of interest, ρ is the

density of biphosphate and N is the total number of biphosphate molecules in the system. Similarly, I calculate the distributions of lipid phosphate head group and carbon atoms of lipid tail from distance r between side chain nitrogen atom of the individual peptide residues and the center of mass of each of the head groups [$\rho_{LP}^{(a)}(r)$] and that of each of the tails [$\rho_{LC}^{(a)}(r)$], respectively.

Structural properties of lipid bilayer:

I calculate the following quantities to analyze the structural properties of the DPPC bilayer using the MEMBPLUGIN³⁹ tool implemented in VMD.

- (i) *Bilayer thickness:* Bilayer thickness is defined as the distance between peaks of the phosphate atoms of lipid molecules calculated from density profile. I analyze the average thickness of the bilayer with the simulation time as well as the thickness map projected onto the xy-plane of the simulation box.
- (ii) *Lipid tilt angle:* For sn-1 tilt angle of DPPC I consider the vector connecting the atom C22 and C216, the first and the last carbon of the sn-1 chain. Similarly, for the sn-2 tilt angle of DPPC the atoms C32 and C316 of sn-2 chain are considered.
- (iii) Radial distribution of P atom of lipid head group in the upper leaflet of the bilayer is defined with the distribution function $g_p(r) = \langle N(r) \rangle / (4\pi N \rho \Delta r)$, where $\langle N(r) \rangle$ is the number of P atom in the upper leaflet of the bilayer averaged over time, within a distance $r \pm \Delta r / 2$ from the atom of interest, ρ is the density of P atom and N is the total number of P atom in the upper leaflet of the bilayer.

Principal component analysis (PCA) of lipid head group:

I perform principle component analysis (PCA) of the upper leaflet of lipid bilayer in the studied systems using Bio3D suit of program with R software package⁴⁰. Covariance matrix of positional fluctuations of P atom of the head group of lipids as derived from the equilibrated MD trajectories are considered for the upper leaflet of bilayer along the first two principal components of motion.

5.3 Results:**Lipid bilayer structure:**

I analyze the influence of Tat peptide on the structural properties of the DPPC bilayer in absence and presence of biphosphate ions. Figure 5.3(a) shows the variation of bilayer thickness with simulated time in the equilibrated trajectory in absence (LB) and in presence of Tat peptide both with biphosphate (LBP-WP) and without biphosphate (LBP-WOP) ions in solution.

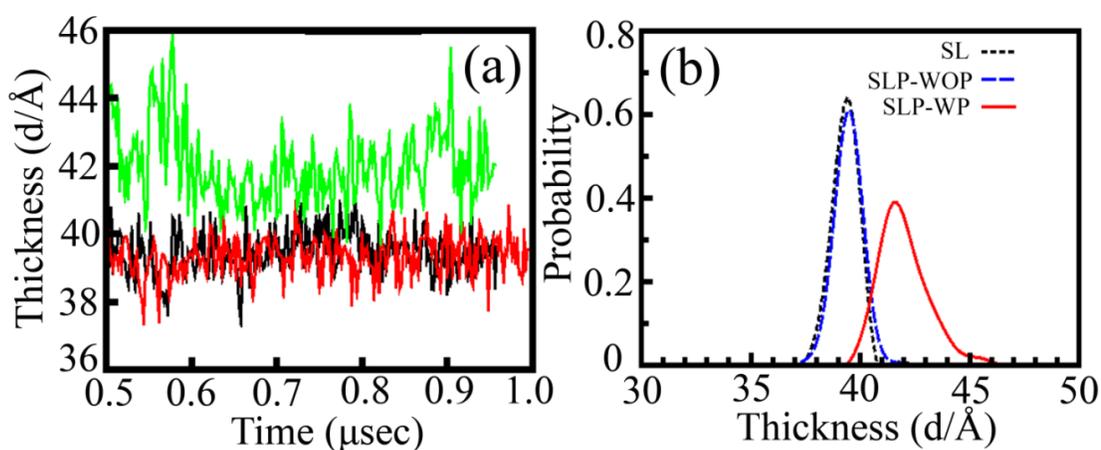


Figure 5.3: (a) Variation of bilayer thickness with simulation time in the equilibrated trajectory in absence (LB: black) and presence of Tat peptide (LBP-WOP: red and LBP-WP: green). (b) Distribution of bilayer thickness in absence and presence of Tat peptide (LB: black, LBP-WOP: blue, LBP-WP: red).

It indicates that the average bilayer thickness is similar both in absence and presence of Tat peptide without biphosphate (LBP-WOP), but higher in case of LBP-WP. Distribution of thickness of the bilayers over the equilibrated trajectories (Figure 5.3(b)) shows that average thickness value is around 39 Å with similar widths both in LB and LBP-WOP systems. In LBP-WP case, the average thickness increases to 42 Å. Moreover, the distribution is broader compared to the other two cases.

PCA analyses have been carried out for the positional fluctuation of P atom of the head group of lipids in the upper leaflet of the bilayer in presence of Tat peptide with biphosphate ions (LBP-WP) and without biphosphate ions (LBP-WOP). I consider the amplitudes of fluctuations of various head groups in two conditions for first two principal components (PCs). Let α_i , denote the amplitude of phosphate of the i -th head group in LBP-WP system and $\alpha_i^{(0)}$ that in LBP-WOP system. $\log(\frac{\alpha_i}{\alpha_i^{(0)}})$ for different head groups for two PCA components are shown in Figs. 5.4 (a) and (b).

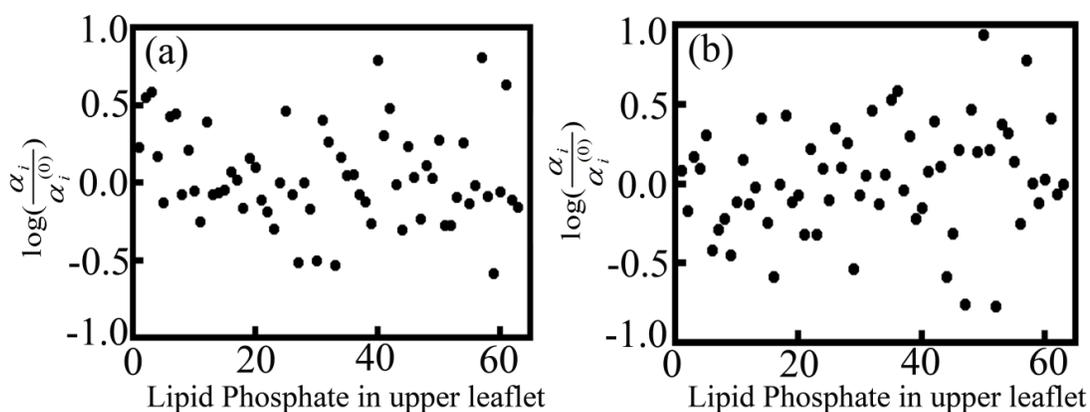


Figure 5.4: The variation of $\log(\frac{\alpha_i}{\alpha_i^{(0)}})$ for the phosphate groups of the lipids in the upper leaflet of the bilayer along (a) PC1 and (b) PC2.

Positive values of logarithm mean enhanced fluctuation along the PC and negative value means reduced fluctuations in LBP-WP case. The figure shows that the fluctuations of lipids head group are not uniform along PC1 and PC2 in LBP-WP case as compared to the LBP-WOP case. Some of the lipids head group show enhanced fluctuation while other show reduced motion in LBP-WP case.

Lipid tilt angle is defined as the angle between an arbitrary vector connecting two atoms within a lipid, and the bilayer normal (Z-axis). Lipid molecules consist of a polar or charged head group and a pair of nonpolar fatty acid tails, connected via a glycerol linkage. Two fatty acid chains, each typically having an even number of carbon atoms between 14 and 20, attach to the first and second carbons of the glycerol molecule, denoted as the sn-1 and sn-2 positions, respectively (Figure 5.5(a)).

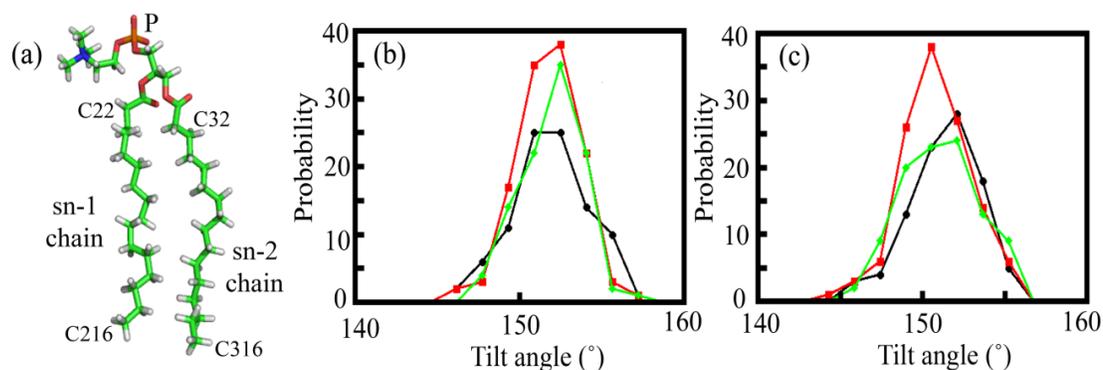


Figure 5.5: Molecular structure of DPPC lipid molecule (a) and distribution of the tilt angle of the acyl chains (b) sn-1 and (c) sn-2 of the lipid molecules at the upper leaflet of the lipid bilayer in absence (LB: black) and presence of Tat peptide (LBP-WOP: red and LBP-WP: green).

The distributions of the tilt angles of the acyl chains (sn-1 and sn-2) of the lipid molecules at the upper leaflet of the lipid bilayer are shown in Figure 5.5. Average tilt angle for sn-1 (Figure 5.5(b)) tail is similar in all the cases. However, the fluctuations are diminished in the presence

of Tat and the presence of biphosphate further reduces the fluctuations. The tilt angle for sn-2 (Figure 5.5(c)) behaves differently. Average value along with the fluctuation gets reduced in presence of Tat peptide without biphosphate and shows enhanced fluctuation with biphosphate ions.

The radial distributions of P atoms of lipid head group $g_P(r)$ in the upper leaflet of the bilayer are shown in Figure 5.6 for all the cases. This quantity shows distribution of the phosphate head groups. Figure 5.6 shows that the head groups have similar distribution in the upper leaflet both in absence and presence of Tat peptide.

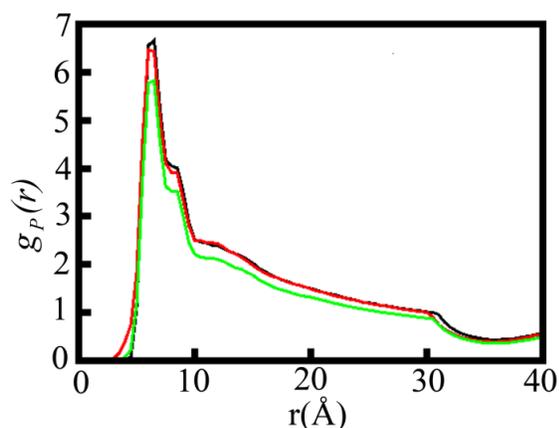


Figure 5.6: Distribution of P atom of lipid head group ($g_P(r)$) in the upper leaflet of the lipid bilayer in absence (LB: green) and in presence (LBP-WOP: red, LBP-WP: black) of Tat peptide.

Interaction of peptide with lipid bilayer:

The interaction between the lipid atoms and the polypeptide are characterized in terms of distribution of the polar head group and hydrophobic tail of the molecules. $\rho_{LP}^{(a)}(r)$ denotes the distributions of centre of mass lipid phosphate head group at radial distance r from side chain nitrogen atom of the individual peptide residues. Similarly, $\rho_{LC}^{(a)}(r)$ represents the

distribution of the centre of mass of carbon atoms hydrophobic lipid tail groups for the α -th residue. In case of arginine residue, the average is taken over both nitrogen atoms in the side chain guanidinium group.

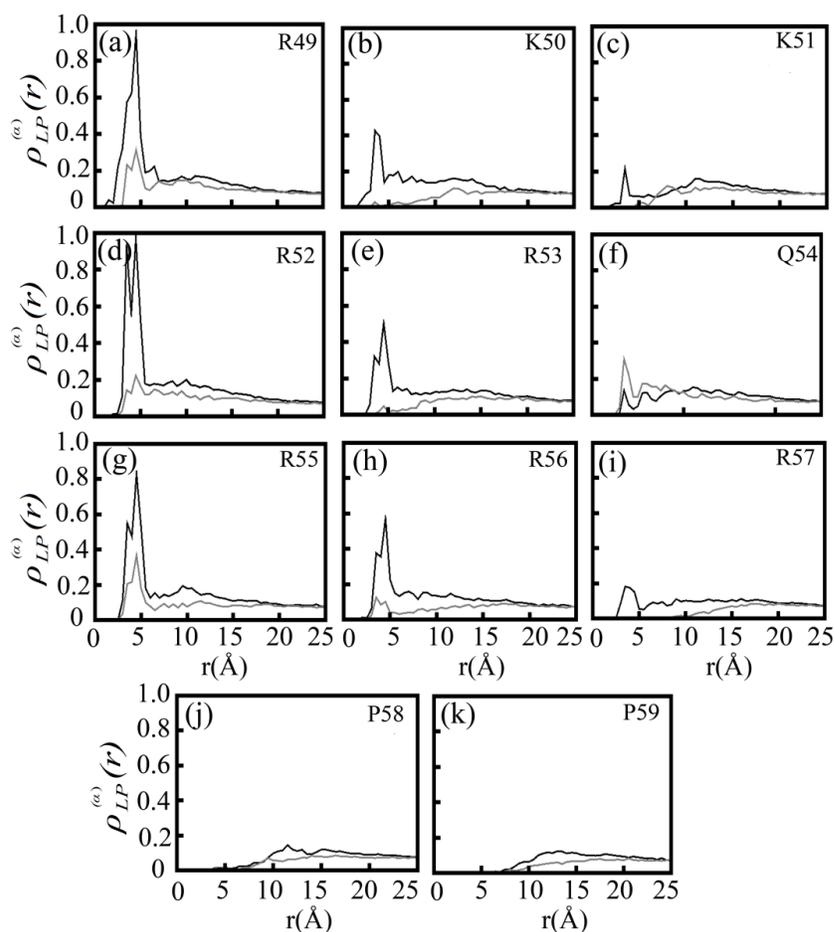


Figure 5.7: Distribution of lipid phosphate head group around side chain nitrogen atom of the peptide residues, $\rho_{LP}^{(a)}(r)$ in absence (LBP-WOP: black) and presence (LBP-WP: gray) of biphosphate ion in solution.

Figure 5.7 shows $\rho_{LP}^{(a)}(r)$ in LBP-WOP and LBP-WP systems. In LBP-WOP, $\rho_{LP}^{(a)}(r)$ has strong first peak around $r = 5 \text{ \AA}$ for R49, K50, R52, R53, R55 and R56 residues. In case of K51, Q54, R57 $\rho_{LP}^{(a)}(r)$ also shows peak around $r = 5 \text{ \AA}$ which is comparatively less intense.

$\rho_{LP}^{(P58)}(r)$ and $\rho_{LP}^{(P59)}(r)$ show broad peak quite far from the side chain nitrogen. Therefore, side

chain of basic arginine and lysine residues are in the close vicinity of the negatively charged head group of lipid molecules which are in the upper leaflet of the lipid bilayer. In LBP-WP, the peaks of $\rho_{LP}^{(a)}(r)$ are much less pronounced for all the peptide residues than LBP-WOP system, except for Q54. $\rho_{LP}^{(Q54)}(r)$ (Figure 5.7(f)) shows strong intense peak at around $r = 5 \text{ \AA}$ as compared to the LBP-WOP case. Therefore, in LBP-WP case only side-chain of Q54 makes close contact with the lipid head groups.

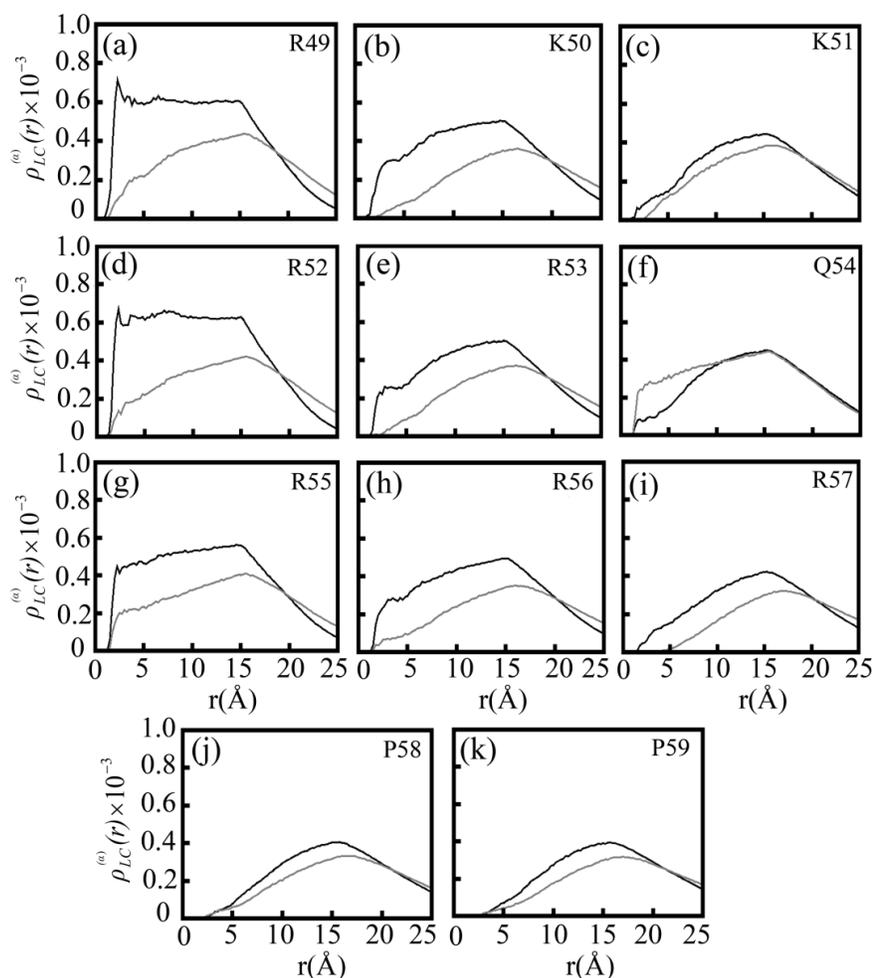


Figure 5.8: Distribution of carbon atoms of lipid tail around side chain nitrogen atom of the peptide residues, $\rho_{LP}^{(a)}(r)$ in absence (LBP-WOP: black) and presence (LBP-WP: gray) of biphosphate ions in solution.

Distribution of carbon atoms of lipid tails with respect to the side chain nitrogen of peptide residues, $\rho_{LC}^{(a)}(r)$ data are shown in Figure 5.8. The peak intensities in these distributions are generally low. However, the distributions are lower in LBP-WP than in LBP-WOP. Thus, the presence of biphosphate keeps the peptide away from the hydrocarbon tail. The only exception is observed for Q54. For this residue, the distribution of the hydrocarbon tails is stronger in LBP-WP than in LBP-WOP.

In order to understand the role of free biphosphate ions, biphosphate ion distribution $\rho_p^{(a)}(r)$ is considered in the LBP-WP system. Figure 5.9 shows $\rho_p^{(a)}(r)$ for the α -th residue. $\rho_p^{(a)}(r)$ has strong peak at around $r = 4 \text{ \AA}$ for basic arginine and lysine residues. In case of $\rho_p^{(Q54)}(r)$ the peak intensity is much lower (Figure 5.9(f)). $\rho_p^{(P58)}(r)$ and $\rho_p^{(P59)}(r)$ show broad peak quite far from the side chain nitrogen (Figures 5.9(j), (k)). This is to be contrasted to the strong peak of $\rho_{LP}^{(Q54)}(r)$ in LBP-WP system (Figure 5.9(f)). Therefore $\rho_p^{(a)}(r)$ data suggest that the presence of biphosphate ions in the vicinity of the arginine and lysine residues of the peptide shield their interaction with phosphate head groups of lipid bilayer. Only polar residue Q54 with less biphosphate concentration interacts with the head group of lipid bilayer. For this residue, $\rho_{LC}^{(a)}(r)$ data also show stronger distribution in LB-WP system than in LB-WOP system. Thus, Q54 is located in the vicinity of the bilayer, while the other basic residues are far from the bilayer.

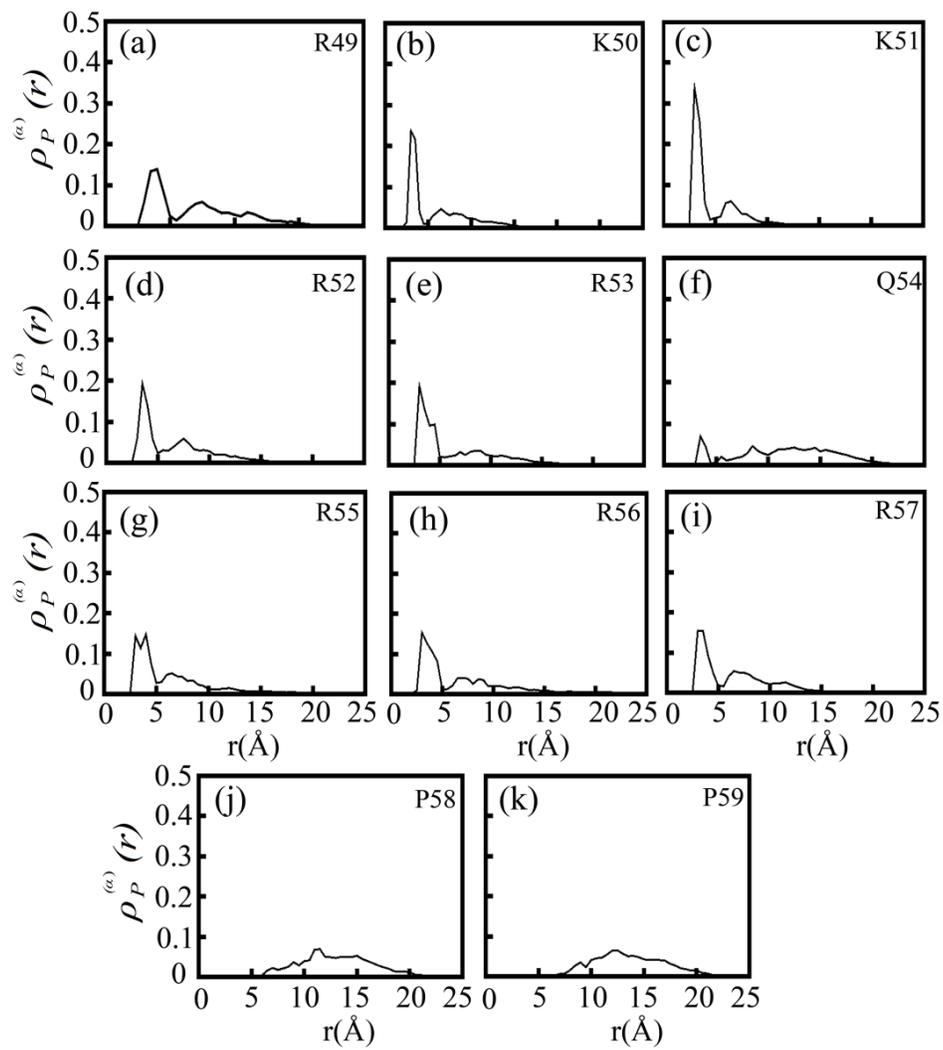


Figure 5.9: Distribution of biphosphate ions around side chain nitrogen atom of the individual peptide residues,

$\rho_P^{(\alpha)}(r)$ in presence of biphosphate ions in solution (LBP-WP).

5.4 Discussion:

Despite several works to date, mechanistic understanding of the cell penetration mechanism of CPPs is not yet clear. Previous studies suggest that antimicrobial peptides can induce positive curvature strain on lipid bilayers⁴¹⁻⁴² whereas amyloid peptides mainly induce negative curvature strain⁴³. Tat peptide is found to be capable of inducing both positive and negative curvature on a bilayer^{8, 44}. Curvature affects the structural properties of lipid bilayer and facilitates the formation of pore on lipid bilayer, which indirectly helps in peptide translocation mechanism^{8, 10, 13, 19, 44}. Formation of pore is associated with the disruption of membrane local structure due to rearrangement of lipids head group and alteration of the conformational properties of lipid molecules within the bilayer membrane. On the other hand, recent studies proposed that interaction of CPP with membrane bifurcate the lipid bilayer structure to multilamellar structures and forms vesicles that allows the peptide to enter by fusion without having to form transient pores¹³. Extra stress leading to membrane expansion has found to present in several biological systems, including bacterial cells and laboratory-prepared uni-lamellar vesicles. External stress on the bilayer surface is also found to perturb the bilayer and facilitate the penetration of CPP without forming the pore²⁹. Thus, it is plausible that membrane active CPPs utilize common mechanisms to disrupt the cell membrane.

It is worthwhile to discuss the scenario suggested by the present studies. In LBP-WOP system, the basic arginine and lysine residues of the Tat peptide are found to interact with the negatively charged head groups of lipid bilayer, as can be seen in the simulation snapshots of Figure 5.10. Due to the electrostatic interactions, the peptide is mostly stabilized parallel to the upper leaflet of the lipid bilayer, where most of the side chain of basic arginine and lysine residues point towards the lipid head groups (Figure 5.10).

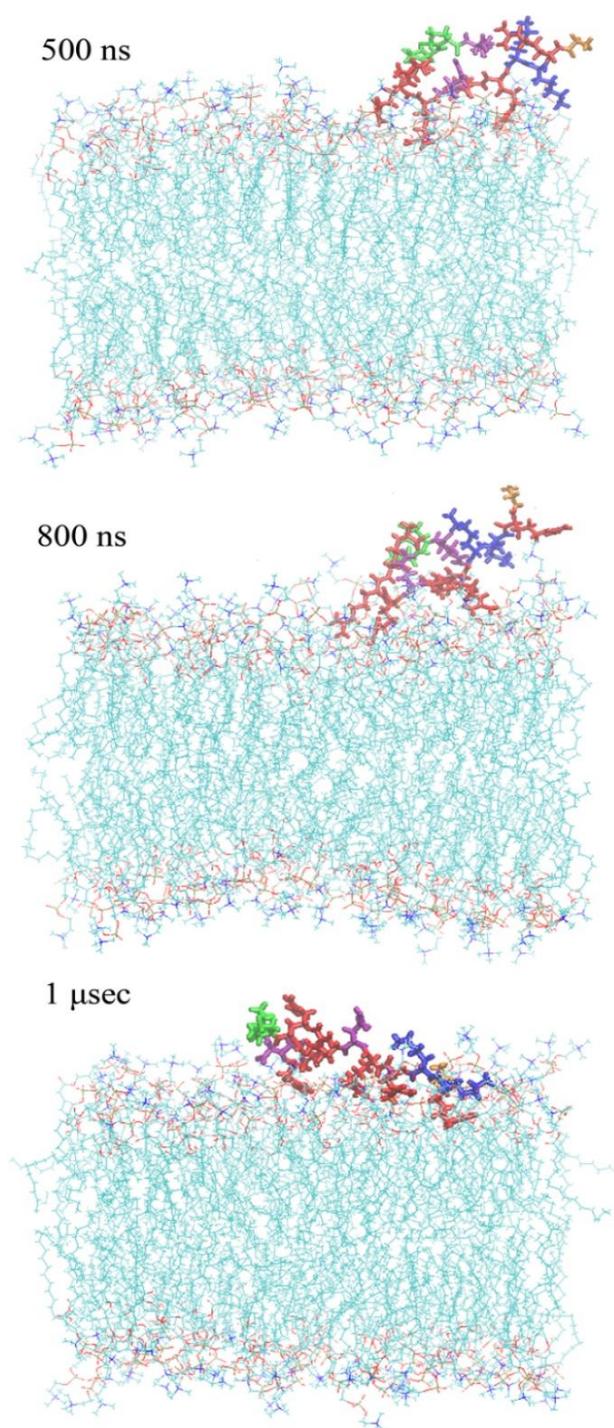


Figure 5.10: Simulation snapshots of Tat peptide interact with the upper leaflet of lipid bilayer in case of LBP-WOP system. Tat peptide is represented as stick where residues are marked with different color (Arginine: red, Lysine: blue, Glutamine: purple, Proline: green, Glycine: orange).

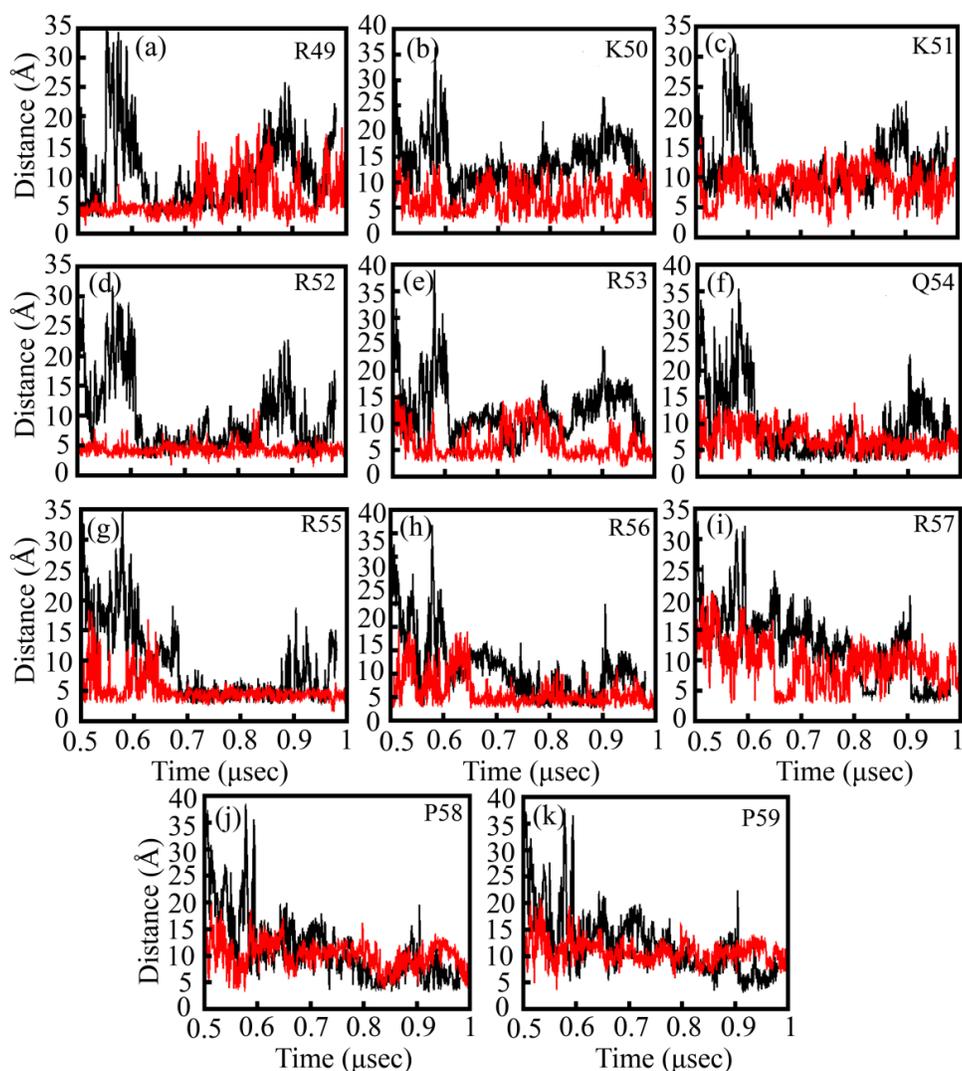


Figure 5.11: Variation of minimum distance between peptide residues and phosphate head group of lipid molecules in the upper leaflet of lipid bilayer in case of LBP-WOP (red) and LBP-WP (black) system.

Variation of minimum distance between peptide residues and phosphate head group of lipid molecules in the upper leaflet of lipid bilayer (Figure 5.11) also supports that Tat peptide is stabilized on the bilayer surface and prefers to stay within 5-10 Å from the upper leaflet of the lipid head group. This is also consistent to the $\rho_{LP}^{(a)}(r)$ data (Figure 5.7). Due to stabilization uniformly over the surface the local thickness does not show any variation as can be seen in the thickness map. The thickness map shows details of the local time-averaged thickness along

the xy-plane. Figure 5.12(a) and (b) indicate uniform thickness profile of lipid bilayer in LB and LBP-WOP case, local thickness values varying in the narrow range of 38 - 40 Å.

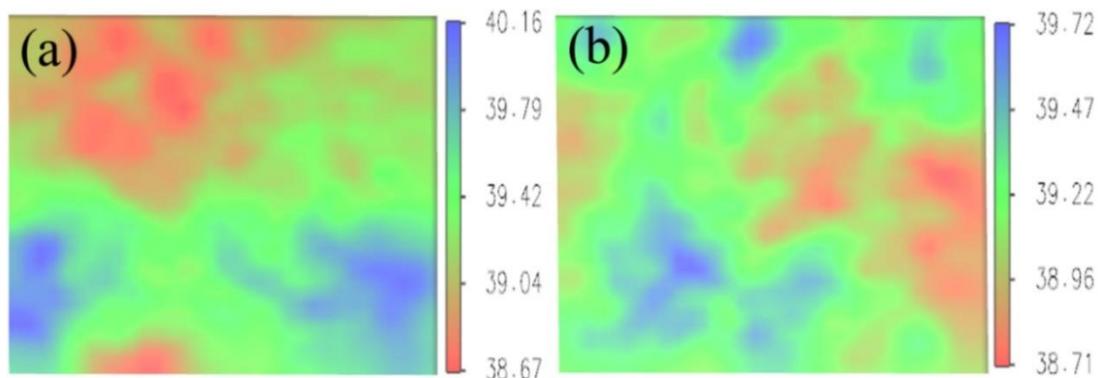


Figure 5.12: Average thickness map of lipid bilayer projected onto the xy-plane of the simulation box in case of (a) LB and (b) LBP-WOP system.

In LBP-WP system, the localization of peptide in the vicinity of the bilayer is quite different from the LBP-WOP system. Side chains of lysine and most of the arginine residues points opposite to the lipid head group due to the presence of biphosphate ions around them, as can be seen from the snapshots in Figure 5.13. Polar residue Q54 in the middle and adjacent R55 are only found to make close contacts with the phosphate head group of lipids. Hydrophobic resides P58 and P59 are also found to reside around 10 Å from the upper leaflet of the bilayer (Figure 5.11).

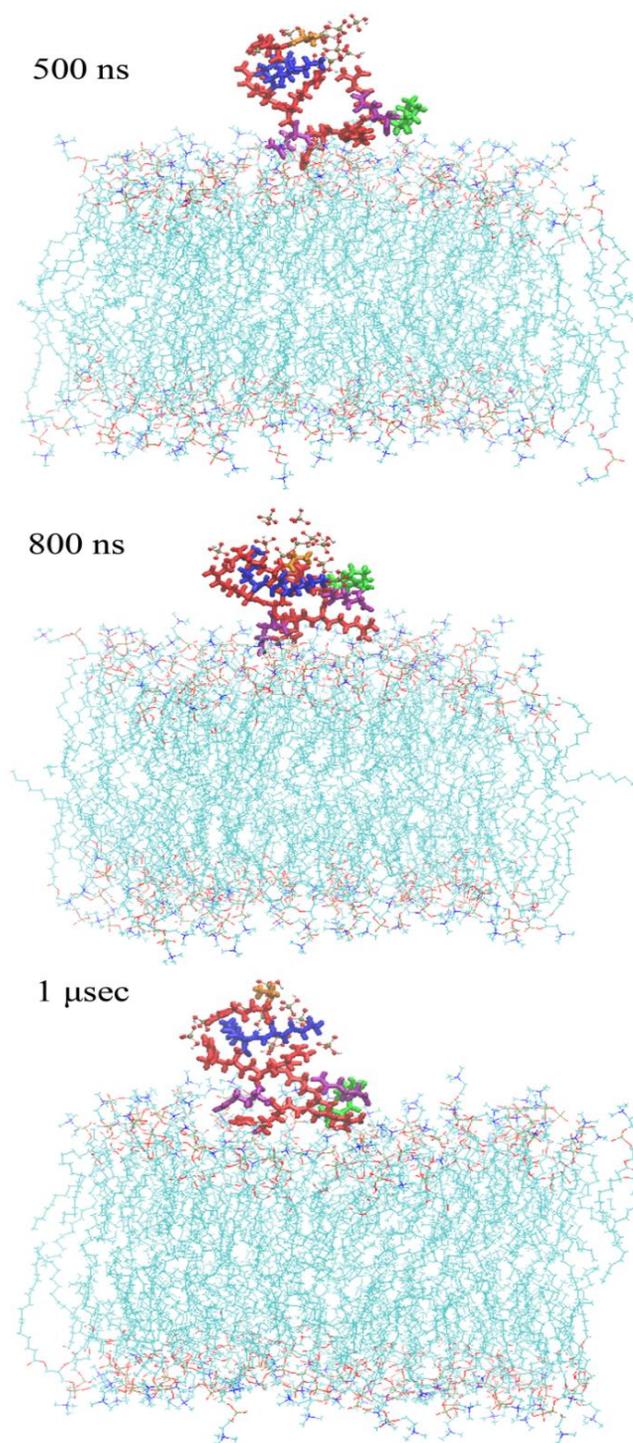


Figure 5.13: Simulation snapshots of Tat peptide interact with the upper leaflet of lipid bilayer in case of LBP-WP system. Tat peptide is represented as stick where residues are marked with different color (Arginine: red, Lysine: blue, Glutamine: purple, Proline: green, Glycine: orange). Biphosphate ions are represented with ball and stick.

However, variations of minimum distance between peptide residues and phosphate head group of lipid molecules in the upper leaflet also show higher fluctuations in LBP-WP case than LBP-WOP case (Figure 5.11). The residues typically keep large distance due to repulsion between biphosphate ions and lipid head group phosphates. Q54 and R55, having less biphosphate condensation is mostly found to stabilize within 5-10 Å from the upper leaflet of the lipids head group (Figure 5.11, Figure 5.13).

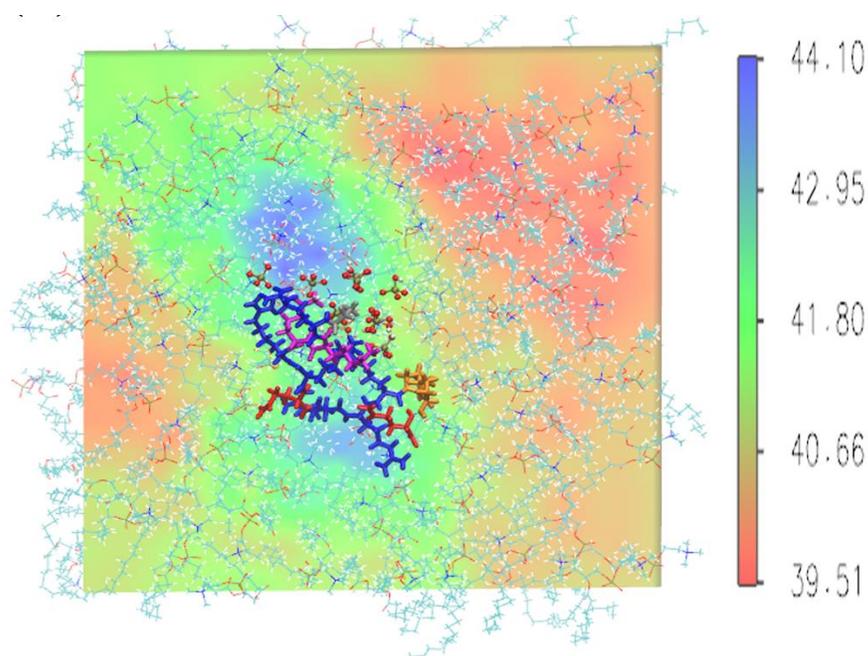


Figure 5.14: Average conformation of Tat peptide bound on the upper leaflet of lipid bilayer (top view) along with average bilayer thickness profile projected onto the xy-plane of the simulation box in presence of biphosphate ions (LBP-WP) (Peptide residues are shown in stick representation with color according to Arginine: blue, Lysine: magenta, Glutamine: red, Proline: orange, Glycine: silver; Biphosphate ions are represented with ball and stick)

The thickness map in Figure 5.14 shows uneven regions which is consistent to the larger spread in the thickness distribution in LBP-WP case. The thickness is much larger than in the normal LB system at the ends where the basic residues with biphosphate condensate around them. Not only the phosphate head groups are repelled by the anion condensate as

shown in the larger, but also the upper leaflet gets apart from the lower leaflet. This results in larger fluctuations in the tilt angle of the sn-2 tail, indicating less registry among the phospholipid tails. The larger separation between two lipid leaflets suggests that the polar Q54 which is in the vicinity of the bilayer applies stress onto the contact point which is relaxed by enhanced separation between the leaflets. The $g_p(r)$ data show that head group distribution does not change in the presence of the peptide, excluding the possibility of pore formation.

In presence of biphosphate ions interactions between polar residues and lipid bilayer induce stress on the bilayer surface. Due to induced stress on the bilayer surface, the bilayer leaflets in the vicinity of the stressed region show larger separation and associated with the loss of compactness of lipid bilayer. Therefore, this study supports the recently proposed mechanism of the stress induced perturbation of the lipid bilayer that can facilitate the penetration of CPPs without forming pores²⁹. It was found in previous biophysical study that counterion scavenging facilitate the translocation of arginine rich polypeptide through lipid bilayer²⁷. This study provides the microscopic basis of interaction of arginine rich CPP with lipid bilayer in presence and absence of counter-anion and shows that the counter-anions help the Tat peptide to induce deformation of the bilayer.

5.5 Conclusion:

Molecular dynamics simulations are performed to investigate the influence of anionic species in interaction of CPPs with lipid bilayer. It is shown that presence of anion can influence the interaction of HIV-1 Tat peptide with lipid bilayer. In absence of anion arginine rich positively charged peptide is absorbed on the negatively charged bilayer surface without changing the overall structural properties of lipid bilayer. Condensation of anion around basic arginine and lysine residues of peptide alters the absorption properties of the Tat peptide on the bilayer surface. Alteration of absorption property is associated with the induced stress on the bilayer surface that leads to uneven thickness adaptation of the lipid bilayer. Absorption of Tat peptide in presence of anion increase the average thickness of the bilayer, and fluctuation of the lipid molecules in the upper leaflet become non-uniform. However, it does not initiate the formation of pore. This study point to stress induced can facilitate the translocation of Tat in bilayer. Investigation using other peptide-lipid combinations and different anionic species will further help to understand the common underlying mechanism of translocation of peptides through membranes.

References:

1. Anderson, W. F. Human gene therapy. *Nature* **1998**, *392*, 25-30.
2. Luo, D.; Saltzman, W. M. Synthetic DNA delivery systems. *Nat. Biotechnol.* **2000**, *18* (1), 33-7.
3. Bechara, C.; Sagan, S. Cell-penetrating peptides: 20 years later, where do we stand? *FEBS Lett.* **2013**, *587* (12), 1693-702.
4. Ebenhan, T.; Gheysens, O.; Kruger, H. G.; Zeevaart, J. R.; Sathekge, M. M. Antimicrobial peptides: their role as infection-selective tracers for molecular imaging. *Biomed Res. Int.* **2014**, *2014*, 867381.
5. Nguyen, L. T.; Haney, E. F.; Vogel, H. J. The expanding scope of antimicrobial peptide structures and their modes of action. *Trends Biotechnol.* **2011**, *29* (9), 464-72.
6. Mahlapuu, M.; Hakansson, J.; Ringstad, L.; Bjorn, C. Antimicrobial Peptides: An Emerging Category of Therapeutic Agents. *Front. Cell. Infect. Microbiol.* **2016**, *6*, 194.
7. Pae, J.; Saalik, P.; Liivamagi, L.; Lubenets, D.; Arukuusk, P.; Langel, U.; Pooga, M. Translocation of cell-penetrating peptides across the plasma membrane is controlled by cholesterol and microenvironment created by membranous proteins. *J. Control Release* **2014**, *192*, 103-13.
8. Mishra, A.; Lai, G. H.; Schmidt, N. W.; Sun, V. Z.; Rodriguez, A. R.; Tong, R.; Tang, L.; Cheng, J.; Deming, T. J.; Kamei, D. T.; Wong, G. C. Translocation of HIV TAT peptide and analogues induced by multiplexed membrane and cytoskeletal interactions. *Proc. Natl. Acad. Sci. USA* **2011**, *108* (41), 16883-8.
9. Lamaziere, A.; Burlina, F.; Wolf, C.; Chassaing, G.; Trugnan, G.; Ayala-Sanmartin, J. Non-metabolic membrane tubulation and permeability induced by bioactive peptides. *PloS one*

2007, 2 (2), e201.

10. Sun, D.; Forsman, J.; Lund, M.; Woodward, C. E. Effect of arginine-rich cell penetrating peptides on membrane pore formation and life-times: a molecular simulation study. *Phys.Chem.Che .Phys.* **2014**, *16* (38), 20785-95.
11. Herce, H. D.; Garcia, A. E.; Cardoso, M. C. Fundamental molecular mechanism for the cellular uptake of guanidinium-rich molecules. *J. Am. Chem. Soc.* **2014**, *136* (50), 17459-67.
12. Ciobanasu, C.; Siebrasse, J. P.; Kubitscheck, U. Cell-penetrating HIV1 TAT peptides can generate pores in model membranes. *Biophys. J.* **2010**, *99* (1), 153-62.
13. Allolio, C.; Magarkar, A.; Jurkiewicz, P.; Baxova, K.; Javanainen, M.; Mason, P. E.; Sachl, R.; Cebecauer, M.; Hof, M.; Horinek, D.; Heinz, V.; Rachel, R.; Ziegler, C. M.; Schrofel, A.; Jungwirth, P. Arginine-rich cell-penetrating peptides induce membrane multilamellarity and subsequently enter via formation of a fusion pore. *Proc. Natl. Acad. Sci. USA* **2018**, *115* (47), 11923-11928.
14. Wadia, J. S.; Stan, R. V.; Dowdy, S. F. Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nat. Med.* **2004**, *10* (3), 310-5.
15. Papahadjopoulos, D.; Vail, W. J.; Pangborn, W. A.; Poste, G. Studies on membrane fusion. II. Induction of fusion in pure phospholipid membranes by calcium ions and other divalent metals. *Biochim. Biophys. Acta* **1976**, *448* (2), 265-83.
16. Yang, S. T.; Zaitseva, E.; Chernomordik, L. V.; Melikov, K. Cell-penetrating peptide induces leaky fusion of liposomes containing late endosome-specific anionic lipid. *Biophys. J.* **2010**, *99* (8), 2525-33.
17. Herce, H. D.; Garcia, A. E. Molecular dynamics simulations suggest a mechanism for

translocation of the HIV-1 TAT peptide across lipid membranes. *Proc. Natl. Acad. Sci. USA* **2007**, *104* (52), 20805-10.

18. Zasloff, M. Antimicrobial peptides of multicellular organisms. *Nature* **2002**, *415* (6870), 389-95.

19. Maniti, O.; Piao, H. R.; Ayala-Sanmartin, J. Basic cell penetrating peptides induce plasma membrane positive curvature, lipid domain separation and protein redistribution. *Int. J. Biochem. Cell Biol.* **2014**, *50*, 73-81.

20. Swiecicki, J. M.; Di Pisa, M.; Burlina, F.; Lecorche, P.; Mansuy, C.; Chassaing, G.; Lavielle, S. Accumulation of cell-penetrating peptides in large unilamellar vesicles: A straightforward screening assay for investigating the internalization mechanism. *Biopolymers* **2015**, *104* (5), 533-43.

21. Huang, K.; Garcia, A. E. Free energy of translocating an arginine-rich cell-penetrating peptide across a lipid bilayer suggests pore formation. *Biophys. J.* **2013**, *104* (2), 412-20.

22. Sakai, N.; Sorde, N.; Das, G.; Perrottet, P.; Gerard, D.; Matile, S. Synthetic multifunctional pores: deletion and inversion of anion/cation selectivity using pM and pH. *Org. Biomol. Chem.* **2003**, *1* (7), 1226-31.

23. Lee, S.; Tanaka, T.; Anzai, K.; Kirino, Y.; Aoyagi, H.; Sugihara, G. Two mode ion channels induced by interaction of acidic amphipathic alpha-helical peptides with lipid bilayers. *Biochim. Biophys. Acta* **1994**, *1191* (1), 181-9.

24. Iwata, T.; Lee, S.; Oishi, O.; Aoyagi, H.; Ohno, M.; Anzai, K.; Kirino, Y.; Sugihara, G. Design and synthesis of amphipathic 3(10)-helical peptides and their interactions with phospholipid bilayers and ion channel formation. *J. Biol. Chem.* **1994**, *269* (7), 4928-33.

25. Agawa, Y.; Lee, S.; Ono, S.; Aoyagi, H.; Ohno, M.; Taniguchi, T.; Anzai, K.; Kirino, Y.

Interaction with phospholipid bilayers, ion channel formation, and antimicrobial activity of basic amphipathic alpha-helical model peptides of various chain lengths. *J. Biol. Chem.* **1991**, *266* (30), 20218-22.

26. Anzai, K.; Hamasuna, M.; Kadono, H.; Lee, S.; Aoyagi, H.; Kirino, Y. Formation of ion channels in planar lipid bilayer membranes by synthetic basic peptides. *Biochim. Biophys. Acta* **1991**, *1064* (2), 256-66.

27. Sakai, N.; Matile, S. Anion-mediated transfer of polyarginine across liquid and bilayer membranes. *J. Am. Chem. Soc.* **2003**, *125* (47), 14348-56.

28. Quan, X.; Sun, D.; Zhou, J. Molecular mechanism of HIV-1 TAT peptide and its conjugated gold nanoparticles translocating across lipid membranes. *Phys. Chem. Chem. Phys.* **2019**, *21* (20), 10300-10310.

29. Losasso, V.; Hsiao, Y. W.; Martelli, F.; Winn, M. D.; Crain, J. Modulation of Antimicrobial Peptide Potency in Stressed Lipid Bilayers. *Phys. Rev. Lett.* **2019**, *122* (20), 208103.

30. Peloponese, J. M., Jr.; Gregoire, C.; Opi, S.; Esquieu, D.; Sturgis, J.; Lebrun, E.; Meurs, E.; Collette, Y.; Olive, D.; Aubertin, A. M.; Witvrow, M.; Pannecouque, C.; De Clercq, E.; Bailly, C.; Lebreton, J.; Loret, E. P. ¹H-¹³C nuclear magnetic resonance assignment and structural characterization of HIV-1 Tat protein. *C R Acad Sci. III* **2000**, *323* (10), 883-94.

31. Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29* (11), 1859-65.

32. Dickson, C. J.; Madej, B. D.; Skjevik, A. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C. Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014**, *10* (2), 865-879.

33. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78* (8), 1950-8.
34. Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., Hermans J. (1981) Interaction Models for Water in Relation to Protein Hydration. In: Pullman B. (eds) Intermolecular Forces. The Jerusalem Symposia on Quantum Chemistry and Biochemistry, vol 14. Springer, Dordrecht
35. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435-47.
36. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26* (16), 1701-18.
37. Repakova, J.; Holopainen, J. M.; Morrow, M. R.; McDonald, M. C.; Capkova, P.; Vattulainen, I. Influence of DPH on the structure and dynamics of a DPPC bilayer. *Biophys. J.* **2005**, *88* (5), 3398-410.
38. Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33-8, 27-8.
39. Guixa-Gonzalez, R.; Rodriguez-Espigares, I.; Ramirez-Angueta, J. M.; Carrio-Gaspar, P.; Martinez-Seara, H.; Giorgino, T.; Selent, J. MEMBPLUGIN: studying membrane complexity in VMD. *Bioinformatics* **2014**, *30* (10), 1478-80.
40. Grant, B. J.; Rodrigues, A. P.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **2006**, *22* (21), 2695-6.

41. Hallock, K. J.; Lee, D. K.; Ramamoorthy, A. MSI-78, an analogue of the magainin antimicrobial peptides, disrupts lipid bilayer structure via positive curvature strain. *Biophys. J.* **2003**, *84* (5), 3052-60.
42. Matsuzaki, K.; Sugishita, K.; Ishibe, N.; Ueha, M.; Nakata, S.; Miyajima, K.; Epanand, R. M. Relationship of membrane curvature to the formation of pores by magainin 2. *Biochemistry* **1998**, *37* (34), 11856-63.
43. Smith, P. E.; Brender, J. R.; Ramamoorthy, A. Induction of negative curvature as a mechanism of cell toxicity by amyloidogenic peptides: the case of islet amyloid polypeptide. *J. Am. Chem. Soc.* **2009**, *131* (12), 4470-8.
44. Schmidt, N.; Mishra, A.; Lai, G. H.; Wong, G. C. Arginine-rich cell-penetrating peptides. *FEBS Lett.* **2010**, *584* (9), 1806-13.

Appendix of Chapter 5

A5 Molecular Dynamics simulation setup for peptide-lipid system

The coordinates of the Tat peptide are taken from the NMR structure of the HIV-1 Tat protein (PDB code 1JFW³⁰, model 1). The initial configuration of the DPPC bilayer with 128 lipids is generated using the CHARMM Membrane Builder GUI³¹ at the relevant experimental hydration level considering 30.1 waters per DPPC lipid molecule and converted to Lipid14 PDB format using the charmm lipid2amber.py script. There are 64 DPPC molecules in the upper and lower leaflet of the bilayer. During the system set up using the GROMACS simulation tool³⁵⁻³⁶, Amber Lipid 14 topology of lipid is added to the Amber99ILDN³³ force-field files. Therefore, DPPC lipid bilayer is described with Amber Lipid14³² force field parameters and the Tat peptide is described with the Amber99ILDN³³ force field. The DPPC bilayer is placed in the x - y plane of a rectangular simulation box with the normal along the z -axis having dimension $6.3 \text{ nm} \times 6.3 \text{ nm} \times 12.2 \text{ nm}$ comparable to the length and width of the DPPC bilayer. Tat peptide and bi-phosphate ions are placed little far from the upper part of the DPPC bilayer at random positions within a simulation box and the system is solvated with water molecules considering periodic boundary condition in three dimensions. No water molecules are present in the hydrophobic core of the DPPC bilayer. They are present above and below the DPPC bilayer. The simple point charge (SPC)³⁴ model is used to model water. Required numbers of Na^+ ions are added for neutralizing the charged system and 10 HPO_4^{2-} ions are considered in case if Tat peptide with biphosphate ions to maintain the 20 mM biphosphate ions concentration in solution. After the initial minimization, heat the system to production temperature. Choosing a production temperature is an important choice in the lipid bilayer simulation. Lipid bilayers have experimentally measured phase transition temperatures

Chapter 5

from highly ordered gel-like phases to liquid phases. Here I consider the production temperature of 323 K which is above the main phase transition temperature of the DPPC bilayer (314 K)³⁷ at ambient pressure.



CHAPTER 6

CONCLUSIONS & FUTURE PERSPECTIVE



Chapter 6

The primary goal of this thesis is to reveal the microscopic basis of functional ligand binding and provide *in-silico* insights to design polypeptides for drug delivery and other biomedical applications. I focus interactions of charged ligands, like anions with peptides of different functional proteins and studied the influence of anion in the peptide conformation and biological functions. The overall conclusions of my thesis are summarized below.

The molecular dynamics and quantum chemical calculations on anion binding C^oNN motif in functional proteins reveal that sequence dependent conformational preference of motif residues is crucial for anion recognition. The residues which do not undergo conformational switch due to presence of anion are the ones responsible for anion recognition. These are mostly polar and basic residues. Quantum chemical calculation considering the polarization effect due to charged species shows that the residues which do not undergo conformational switch in presence of anion coordinate with the anion at lowest energy. Therefore, the microscopic study reveals an intimate connection between coordination and conformational preferences of the motif residues which has not been established so far to the best of my knowledge. C^oNN motif containing residues are functionally important, typically part of a substrate, cofactor or protein-binding site. Therefore, this study may be useful to get the microscopic insights into their functions and design synthetic peptides for anion receptor and sensing. The changes of conformational preference of anion recognizing residues may be applied to ligand recognition in general for other functional ligand binding motifs. This study is also relevant to protein engineering and design ligand binding site of protein to control the enzymatic activity and biological functions.

Molecular dynamics studies on anion interaction with HIV-1 Tat peptide provides the atomistic basis of counterion scavenging mechanism and indicates that the condensation of

counter-anion around arginine rich Tat peptide is associated with the interaction of anionic species with the side chain guanidinium group of the basic arginine residues. There is an optimum concentration of anion in solution where the condensation is maximum, while the solvent exposed area is the minimum. This is supported by simple model calculations. Thus, hydrophilic nature and solvent exposure of charged polypeptide can be controlled by tuning the counterion concentration in solution. This aspect would be useful for therapeutic applications and design polypeptides as drug carriers into a cell.

All-atom simulations of peptide-lipid bilayer system reveal that condensation of anion around basic arginine and lysine residues of peptide alters the absorption properties of the Tat peptide on the bilayer surface. Alteration of absorption property in presence of anion is associated with the induced stress on the bilayer surface that leads to uneven thickness adaptation of the lipid bilayer. This study points to the fact that induce stress can perturb the lipid bilayer and can facilitate the translocation of Tat through bilayer. This study will provide further insight to understand the cell penetrating and antimicrobial activity of HIV-1 Tat peptide in presence of anion, which is helpful for drug delivery and other biomedical applications.

Further investigations with others anion binding peptide motifs in functional proteins would facilitate the new insight about the anion recognition mechanism in general and their biological functions. Microscopic picture of anion-condensation mechanism around the cationic HIV-1 Tat peptide demands further study based on the explicit consideration of charged functional groups for understanding the counterion condensation mechanism in case of charged biopolymers. Further studies on other peptide-lipid combinations and different

Chapter 6

anionic species would help to decipher the common underlying cell penetrating mechanism for cell penetrating peptides.



PUBLICATIONS



Anion induced conformational preference of C^αNN motif residues in functional proteins

Piya Patra¹ | Mahua Ghosh² | Raja Banerjee¹ | Jaydeb Chakrabarti^{2,3} 

¹Maulana Abul Kalam Azad University of Technology, West Bengal, (Formerly known as WBUT), BF-142, Sector-I, Saltlake, Kolkata 700 064, India

²Department of Chemical, Biological and Macro-Molecular Sciences, S.N. Bose National Centre for Basic Sciences, Sector III, Block JD, Salt Lake, Kolkata 700106, India

³The Thematic Unit of Excellence on Computational Materials Science, S. N. Bose National Centre for Basic Sciences, Sector-III, Block JD, Salt Lake, Kolkata 700106, India

Correspondence

Jaydeb Chakrabarti, S.N. Bose National Centre for Basic Sciences, Kolkata, West Bengal, India.

Email: jaydeb@bose.res.in

Raja Banerjee, Maulana Abul Kalam Azad University of Technology, West Bengal, (Formerly known as WBUT), India.

Email: banraja10@gmail.com

Mahua Ghosh, S.N. Bose National Centre for Basic Sciences, Kolkata, West Bengal, India.

Email: mahua.ghosh@gmail.com

Abstract

Among different ligand binding motifs, anion binding C^αNN motif consisting of peptide backbone atoms of three consecutive residues are observed to be important for recognition of free anions, like sulphate or biphosphate and participate in different key functions. Here we study the interaction of sulphate and biphosphate with C^αNN motif present in different proteins. Instead of total protein, a peptide fragment has been studied keeping C^αNN motif flanked in between other residues. We use classical force field based molecular dynamics simulations to understand the stability of this motif. Our data indicate fluctuations in conformational preferences of the motif residues in absence of the anion. The anion gives stability to one of these conformations. However, the anion induced conformational preferences are highly sequence dependent and specific to the type of anion. In particular, the polar residues are more favourable compared to the other residues for recognising the anion.

KEYWORDS

anion recognition, conformational thermodynamics, correlation plot, molecular dynamics, secondary structure

1 | INTRODUCTION

Proteins play important roles in biochemical processes within a living cell. Quite often functions of proteins depend on their interaction with ligands, including small molecules, ions, peptides, and so on. The molecular structures revealed by X-ray crystallography and NMR often show various ligand-recognition motifs in proteins.^{1–14} However, the microscopic basis of recognition of ligand by the protein is not well understood till date.

Among the motifs found in different proteins involved in ion binding, some are specific to cation binding,^{11,13,14} while some other recognize anions.^{1,7,10,12} Detailed bioinformatics studies,¹ with a number of proteins show that among the anion binding motifs, the C^αNN motif, comprised of backbone C_{i-1}^α (C^α), N_i (N1) and N_{i+1} (N2) atoms of three successive residues for recognising anions, like sulphate (SO₄²⁻) and phosphate (PO₄³⁻), are present in functional interface.¹ Amino acid analysis on these proteins using the fold classification based on structure-structure alignment of proteins (FSSP database)¹⁵ indicates preferences of residue type in the motif (see Supporting Information Table S1). The

occurrence of glycine (G) and serine (S) as the first residue of C^αNN motif are high (41% and 25%, respectively). There is preference for hydrophobic residue (51%) at the middle site. The third position is either a polar residue (50%) with mostly S (18%) and threonine (T, 20%) or a hydrophobic (32%) residue. The first amino acid of the three residues sequence usually adopts beta (β) conformation, while the second and the third residue adopt alpha (α) conformation. According to the crystal structure the geometry of C^αNN motif bound to sulphate,¹ out of four oxygen atoms (O) of anion (sulphate), two participate in interaction with the motif. One O atom concurrently interacts with the C^α and N1, while the other interacts only with N2.¹ It is further reported that if the first residue is serine, side chain oxygen atom at the gamma position (OG) of serine forms hydrogen bond with oxygen atom of the anion.¹ Recent experimental and computational studies on the C^αNN motif appended at the N-terminal of a model helical peptide indicate that the motif segment recognizes the anions (both sulphate or the phosphate ions) through local interaction^{16–19} via context free environment and the presence of polar residues enhances the anion recognition.^{16,18,19}

TABLE 1 Simulated polypeptides having tri residue segment motif

Serial no	PDB Id	Polypeptide	Motif	Simulation box size (nm)
1	1KQR	T124 to K145	SQT fragment: S134(C ^α)Q135(N1)T136(N2)	5.78
2	1F2D	H69 to G92	SNQ fragment: S78(C ^α)N79(N1)Q80(N2)	5.24
3	1AXI	L141 to R156	GIH fragment: G148(C ^α)I149(N1)H150(N2)	4.98
4	1E3H	I124 to A148	LYD fragment: L132(C ^α)Y133(N1)D134(N2)	5.58
5	1VDR	D54 to V82	SRS fragment: S65(C ^α)R66(N1)S67(N2)	5.01

The C^αNN motif residues in functional proteins are also having large proportion of polar residues, although the motif sequences are quite diverse. Moreover, they are impregnated in polypeptides with variety of secondary structures. Several aspects of anion recognition by the C^αNN motif in functional proteins, like the role played by the polar residues, sequence variability, type of anion and the solvent distribution around the motif atoms, are yet to be examined. In particular, there are no microscopic studies on the C^αNN motif in functional proteins.

Here we study the C^αNN motif containing polypeptides from known crystal structure of several functional proteins at neutral pH condition, using the all-atom molecular dynamic (MD) simulations. In choosing the relevant polypeptide, we retain the succeeding and the preceding structural elements. At physiological pH, the phosphate (PO₄³⁻) is mostly available in dibasic (HPO₄²⁻) form. So we consider HPO₄²⁻ instead of PO₄³⁻ in our system.² We simulate each polypeptide both by removing the anion (denoted by WOS and WOP for sulphate and biphosphate, respectively) and retaining the anion (denoted by WS and WP for sulphate and biphosphate, respectively). Our data indicate that the motif residues have conformational flexibility, while the anion gives stability to one of these conformations. The conformations both with and without anions are similar in some residues, mostly polar and basic; while there are shifts in conformational preferences in the two cases for most other residues. The anion induced conformations depend strongly on the sequence and the type of anion, but not so sensitive to the presence of water molecules in the vicinity of the motif atoms and the flanking residues containing the motif.

2 | METHODS

2.1 | System preparation

Detailed sequences of the polypeptide for different proteins having the C^αNN motif are given in Table 1. The systems without sulphate or biphosphate are prepared by deleting the anions from the fragment of polypeptides. We put the polypeptide in a cubic box (box lengths for different cases given in Table 1) and place it at least 10 Å from the box edge and fill the box with TIP3P water for solvation of the peptide. The hydrogen atoms are taken to be united atoms with C^α, N, C, O, OG atoms. We add required numbers of Na⁺ and Cl⁻ ions for

neutralizing the charged system. The resulting structure is relaxed through energy minimization.

2.2 | Molecular dynamic (MD) simulation

The energy minimized structure is subject to MD simulations. We use Gromacs 4.6.7^{20,21} tool and Amber99ILDN force field²² for simulation. The systems are equilibrated in two phases—(i) NVT equilibration and (ii) NPT equilibration. The system is equilibrated at the desired temperature (300 K) and pressure (1 Bar) using the periodic boundary conditions with 1 femto-second time step. We use the particle-mesh Ewald method for treating the long ranged electrostatic interactions. We check the equilibration from the saturation of RMSD of peptide segments considering only backbone heavy atoms. The trajectories are visualized with visual molecular dynamics (VMD).²³ We perform the following analysis from the saturated portion of the trajectories as confirmed from the RMSD of the peptide segments with respect to its initial energy minimized structure:

1. *φ-ψ correlation plot*: We generate the correlation plots for the backbone torsion angles (ϕ and ψ) for the residues of C^αNN motif from the equilibrated trajectory both in presence and absence of sulphate and biphosphate. We also calculate the relative frequency of occurrence of different secondary structural elements like β strand, right handed alpha helix (RH), left handed alpha helix (LH) and random coil (C).
2. *Distance calculation*: We calculate distances between sulphur (S) of sulphate and phosphorus (P) of biphosphate and C^α, N1, and N2 atom of C^αNN motif from simulated trajectories. Then the frequency distributions of distances between S of sulphate or P of biphosphate and C^α, N1, and N2 atoms of this motif, respectively of each polypeptide over the equilibrated trajectory are calculated. For the calculation of frequency distribution of distance we construct bins with bin size 0.5 Å.
3. *Water distribution around motif residues*: We calculate the radial distribution function $g(r)$ ²⁴ of the oxygen atom of water around the C^α, N1, and N2 atoms. The radial distribution function $g(r)$ is defined as: $g(r) = \langle \Delta N(r) \rangle / (4\pi N\rho\Delta r)$, where $\langle \Delta N(r) \rangle$ is the number of oxygen atom of water molecules averaged over time, within a distance $r \pm \Delta r/2$ from the atom of interest, ρ is the

density of water and N is the total number of water molecules within the system.

4. *Conformational thermodynamics*: We study the conformational thermodynamics of the peptide-anion interaction by the histogram based method (HBM) using the dihedral angle distributions of peptide fragment.²⁵⁻²⁷ The equilibrium conformational free energy cost associated with any peptide dihedral angle θ in the sulphate bound state as compared to the biphosphate bound state is defined as $\Delta G^{\text{conf}}(\theta) = -k_B T \ln [H_s^{\text{max}}(\theta)/H_p^{\text{max}}(\theta)]$, where $H_s(\theta)$ and $H_p(\theta)$ is the distribution of dihedral angle θ in sulphate and biphosphate bound state respectively and the 'max' superscript denotes the peak values of the histograms. On the other side the conformational entropy for a particular dihedral can be defined by the Gibbs entropy formula as $S^{\text{conf}}(\theta) = -k_B \sum_i H_i(\theta) \ln H_i(\theta)$, where the sum is taken over the histogram bins. The conformational entropy change for the dihedral can be obtained from the given expression $\Delta S^{\text{conf}}(\theta) = S_s^{\text{conf}}(\theta) - S_p^{\text{conf}}(\theta)$. The conformational free energy and entropy cost for the individual peptide residues can be obtained by taking the sum of the contribution of free energy and entropy of each dihedral angles of that residue.

3 | RESULTS AND DISCUSSION

The C^αNN motif containing sulphate and biphosphate bound polypeptide fragments used in our study are shown in Supporting Information Figures S1a-e. In the polypeptides, we retain the secondary structures both preceding and succeeding the motif residues. Table 1 shows the polypeptide sequences containing C^αNN motif of functional proteins with known X-ray crystallographic structures at neutral pH (=7.0) which have been studied in this work. The proteins are¹: (1) Rhesus rotavirus VP4, (PDB ID: 1KQR²⁸) which acts as antigenic epitope. VP4 (1KQR) has polar motif residues, S134, glutamine (Q135), T136 denoted as SQT fragment. Here the entire polypeptide containing the motif residues is a loop (Supporting Information Figure S1a). (2) Pyridoxal 5'-phosphate (PLP) dependent enzyme 1-Aminocyclopropane-1-carboxylate deaminase, (PDB ID: 1F2D²⁹) involved in transamination reactions. The motif residues are polar as well: S78, Asparagine (N79), Q80 denoted as SNQ fragment. The N-terminal flanking residues of the polypeptide in this structure form a loop, while the C-terminal flanking residues form a helix (Supporting Information Figure S1b). (3) Extracellular domain of growth hormone receptor (PDB ID: 1AXI³⁰) with the motif residues G148, Isoleucine (I149), Histidine (H150) (GIH fragment). Here none of the motif residues are polar. The entire polypeptide containing the motif is a loop (Supporting Information Figure S1c). (4) Polynucleotide phosphorylase, (PDB ID: 1E3H³¹), a phosphorolytic exonuclease involved in degrading prokaryotic mRNA. In case of polynucleotide phosphorylase (1E3H), the motif residues leucine (L132), tyrosine (Y133), aspartic acid (D134) (LYD fragment) where the middle residue is polar. In this case the N-terminal flanking residues form loop, while those in C-terminal form a helix (Supporting Information Figure S1d). (5) Dihydrofolate reductases (PDB ID: 1VDR³²) which is halophilic protein in nature. In this case (1VDR), the phosphate

binding motif residues are S65, arginine (R66), S67 (SRS fragment) belonging to a loop, the middle residue being basic and the rest two polar (Supporting Information Figure S1e).

We simulate these polypeptides both with and without the anion. Equilibrations of the systems are ensured from the root mean square deviation (RMSD) plots, shown for all the cases both with and without anions in Supporting Information Figures S2a-k. Note that the saturation of the RMSD is highly system dependent and typically faster in the presence of the anions. The equilibrated regions of trajectories for each polypeptide are considered for analysis.

3.1 | Anion induced conformational preferences of the motif residues

The correlation plots between the backbone torsion angles (ϕ and ψ) of motif residues over the simulated trajectory show their conformational preferences in terms of secondary structural elements, like right handed alpha helix (RH), left handed alpha helix (LH), β , and random coil (C) conformation. The correlation plots of the torsion angles are shown in Figures 1a-f for SQT fragment, while others in Supporting Information Figures S3-S6. Tables 2a and b show the details of percentages of different secondary structural elements of the motif residues along with their crystal structure conformations with the values of ϕ and ψ .

Let us consider the case of SQT in details. In the crystal structure (1KQR) S134 shows β conformation, whereas both Q135 and T136 have RH conformation. The preference for conformation of *S* fluctuates between RH, β , and C in WOS condition (Figure 1a). The percentages of occurrences of these structures over the entire trajectory (Table 2a) show that the major WOS conformation is RH (60%), despite having substantial conformational flexibility. However, the correlation plot shows that *S* in SQT for WS is in RH conformation (Figure 1b) over entire trajectory. The WS conformation is in agreement to the crystal structure data. Thus the major WOS conformation is stabilized by the sulphate in WS state, indicating similar conformational preferences in both cases. We find that *Q* switches between RH and β conformations with RH having slightly higher preference to β (Table 2a) in the WOS case (Figure 1c) and is in RH conformation in WS (Figure 1d). Here again there is no shift in conformational preference. On the other hand, *T* has propensities for both RH (60%) and β (40%) conformations (Figure 1e) in WOS condition and is in primarily β conformation (Figure 1f) in WS condition. In contrast to the first two motif residues, the minor conformation in WOS is stabilized by the sulphate so that there is a shift in conformational preference.

Table 2a shows that polar *S* shows slightly larger preference for β conformation than RH in WOS case in SNQ fragment. *S* adopts β conformation in WS condition which is similar to that in the crystal structure. Thus there is no shift in conformational preference induced by sulphate. The latter two residues (*N* and *Q*) do not show any shift in conformational preferences (both RH) in WOS and WS conditions as in the crystal structure (Supporting Information Figure S3). Let us now consider conformational preferences of GIH fragment in Table 2a. The hydrophobic *G* residue lacks any dominant conformational preference in WOS condition. In WS condition the conformation shifts to RH in

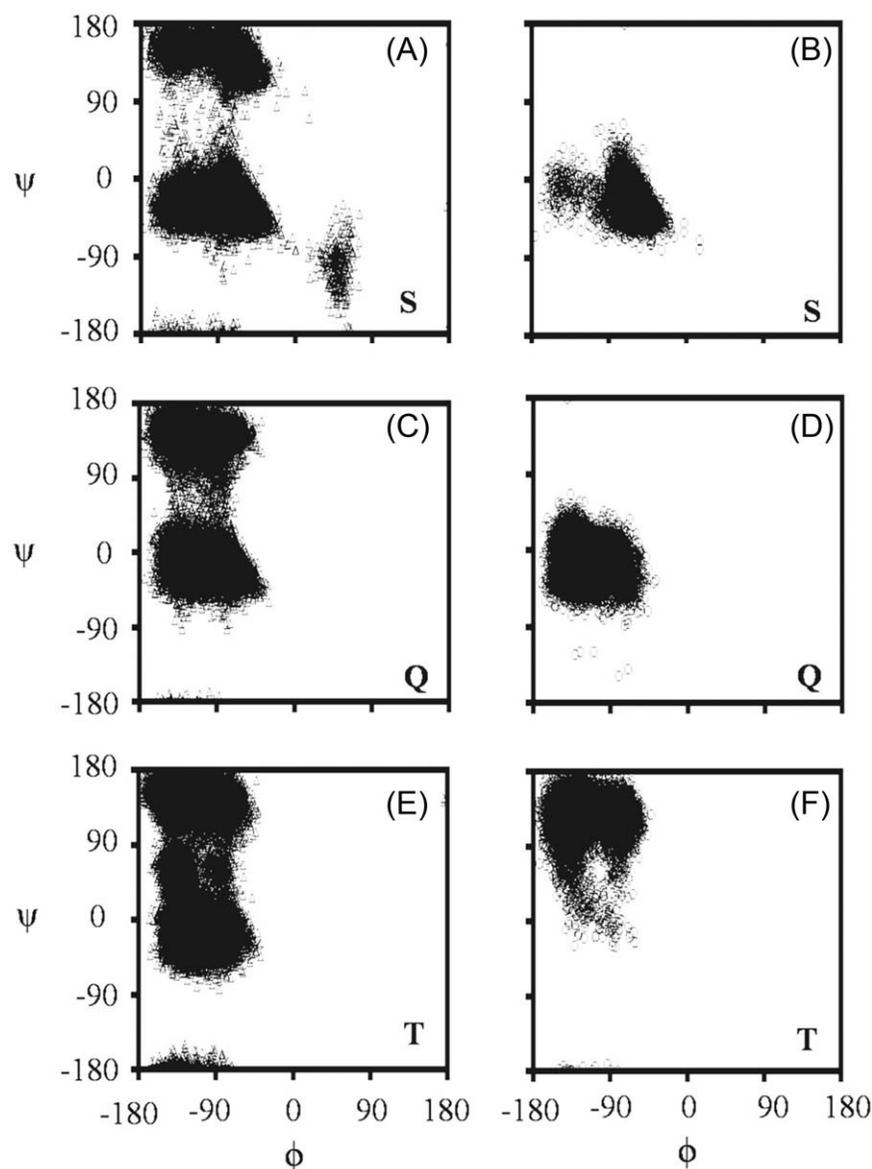


FIGURE 1 The ϕ - ψ correlation plots of the residues in SQT fragment: S134 in (a) WOS and (b) WS conditions; Q135 in (c) WOS and (d) WS conditions; and T136 in (e) WOS and (f) WS conditions

contrast to β conformation in crystal structure. This is not so surprising, for G in general lacks well defined secondary structure. Hydrophobic I in GIH fragment shows change from predominant β conformation in WOS condition to predominant RH conformation in WS condition. The basic residue H does not show any shift in conformational preferences (Supporting Information Figure S4). In case of LYD, it is evident from Table 2a that hydrophobic L shows no particular conformational preference between RH and β in WOS, while having bias to β conformation for WS. Polar Y does not change its conformational preferences, while acidic D conformation switches from RH in WOS to β conformation in WS (Supporting Information Figure S5).

Table 2b further shows that both S are predominantly in RH conformation in SRS fragment in the WOP case. Both S are mostly in β conformation in WP condition. However, the residue shows substantial fluctuations with minor contributions from other conformations. Thus, both S show shifts in conformational preferences in the WP case

compared to the WOP case. However, the basic R does not show any change in preference, being in RH conformation for both the cases (Supporting Information Figure S6). Tables 2a and b clearly show that the conformational preferences of the residues fluctuate without anion, while the anion typically stabilizes one of the residue conformations experienced in absence of the anion. The conformations of the motif residues are in good agreement to their crystal structure conformations.

3.2 | Motion of the anion

We calculate the distances between the sulphur atom and $C^\alpha(d_{C\alpha-S})$, N1 (d_{N1-S}), and N2 (d_{N2-S}) atom of the motif for different fragments. The frequencies of these distances, $f(d_{C\alpha-S})$, $f(d_{N1-S})$, and $f(d_{N2-S})$ for the SQT case are shown in Figure 2a. The distance frequencies for other cases are shown in Supporting Information Figure S7. All these distributions indicate that sulphate ion prefers to stay at large distances from the

TABLE 2 (a) The secondary structural elements of the motif residues in crystal structure and simulated structure for WOS, WS, and RP in different cases. (“–” indicates no occurrence at that particular position)

Peptide fragment (PDB Ids)	Motif Residues		RH	β	LH	C	Crystal structure (ϕ , ψ)
SQT fragment (1KQR)	S (Polar)	WOS	60	30	–	10	β (–71.9, 134.7)
		WS	100	–	–	–	
		RP	79	21	–	–	
	Q (Polar)	WOS	55	45	–	–	RH (–74.6, –29.1)
		WS	100	–	–	–	
		RP	51	49	–	–	
	T (Polar)	WOS	60	40	–	–	RH (–124.7, –16.2)
		WS	3	97	–	–	
RP		95	5	–	–		
SNQ fragment (1F2D)	S (Polar)	WOS	47	53	–	–	β (–61.5, 144.7)
		WS	–	100	–	–	
		RP	–	100	–	–	
	N (Polar)	WOS	91	9	–	–	RH (–90.9, –48.2)
		WS	100	–	–	–	
		RP	99	–	–	1	
	Q (Polar)	WOS	73	27	–	–	RH (–50.3, –47.6)
		WS	100	–	–	–	
RP		100	–	–	–		
GIH fragment (1AXI)	G (Hydrophobic)	WOS	20	15	27	38	β (–66.4, 157.7)
		WS	100	–	–	–	
		RP	21	31	19	29	
	I (Hydrophobic)	WOS	23	77	–	–	RH (–87.6, –19.9)
		WS	100	–	–	–	
		RP	27	73	–	–	
	H (Basic)	WOS	–	100	–	–	β (–117, 163.8)
		WS	–	100	–	–	
RP		21	26	19	34		
LYD fragment (1E3H)	L (Hydrophobic)	WOS	45	55	–	–	β (–107.2, 119.3)
		WS	–	100	–	–	
		RP	100	–	–	–	
	Y (Polar)	WOS	100	–	–	–	RH (–80.9, –21.9)
		WS	97	3	–	–	
		RP	48	52	–	–	
	D (Acidic)	WOS	100	–	–	–	RH (–48.6, –57.4)
		WS	17	83	–	–	
RP		100	–	–	–		
(b) The secondary structural elements of the motif residues in crystal structure and simulated structure for WOP and WP in different cases. (“–” indicates no occurrence at that particular position)							
Peptide fragment (PDB Ids)	Motif Residues		RH	β	LH	C	Crystal structure (ϕ , ψ)
SRS fragment (1VDR)	S (polar)	WOP	70	30	–	–	β (–154.8, 150.8)
		WP	11	89	–	–	
	R (basic)	WOP	89	11	–	–	RH (–63.8, –64.3)
		WP	85	15	–	–	
	S (polar)	WOP	83	17	–	–	β (–45.7, 140.9)
		WP	3	97	–	–	

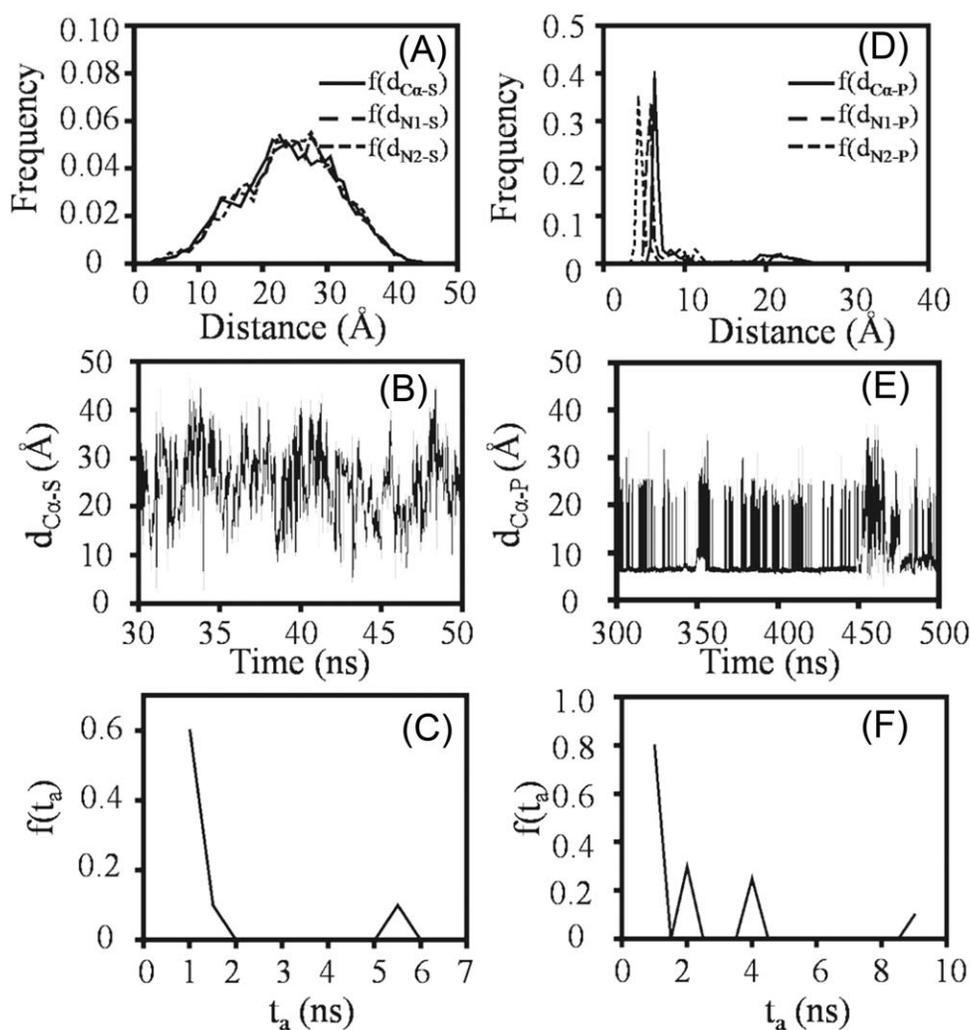


FIGURE 2 (a) The frequency distributions of distances; $f(d_{C\alpha-S})$ (solid line), $f(d_{N1-S})$ (long dashed line), and $f(d_{N2-S})$ (short dashed line) between sulphur atom of sulphate and C^α , N1, and N2 atoms, respectively, of $C^\alpha NN$ motif in SQT fragment. (b) $d_{C\alpha-S}$ as a function of time in SQT fragment. (c) The distribution $f(t_a)$ of time interval, t_a between two successive approaches by sulphate to the vicinity of C^α atom ($d_{C\alpha-S} \leq 5 \text{ \AA}$) in SQT fragment. (d) The frequency distributions of distances; $f(d_{C\alpha-P})$ (solid line), $f(d_{N1-P})$ (long dashed line), and $f(d_{N2-P})$ (short dashed line) between phosphorous atom of biphosphate and C^α , N1, and N2 atoms, respectively, of $C^\alpha NN$ motif in SRS fragment. (e) $d_{C\alpha-P}$ as a function of time in SRS fragment. (f) The distribution $f(t_a)$ of time interval, t_a between two successive approaches by biphosphate to the vicinity of C^α atom ($d_{C\alpha-P} \leq 5 \text{ \AA}$) in SRS fragment

$C^\alpha NN$ motif atoms. Figure 2b shows $d_{C\alpha-S}$ as a function of time over the simulated trajectory of SQT. The plot shows that sulphate ion moves back and forth in the neighbourhood of the motif ($\leq 5 \text{ \AA}$) to larger distances ($\geq 10 \text{ \AA}$). Similar picture holds for other sulphate bound fragments as well, shown in Supporting Information Figure S8. We show in Figure 2c the distribution ($f(t_a)$) of time interval, t_a between any two successive approaches of sulphate to the vicinity of C^α atom ($\leq 5 \text{ \AA}$). The distribution is sharply peaked for low t_a with mean time $\langle t_a \rangle$ ($\sim 0.5 \text{ ns}$) which is much smaller compared to the time scale of secondary structural relaxation.^{33,34} Thus, the sulphate induced structure cannot relax due to multiple approach of the ion to the vicinity of the motif in very short time scale. Figure 2d shows the frequencies of distances C^α ($f(d_{C\alpha-P})$), N1 ($f(d_{N1-P})$), and N2 ($f(d_{N2-P})$) from phosphorus atom of the biphosphate ion for the SRS case. The peak of $f(d_{C\alpha-P})$, $f(d_{N1-P})$, and $f(d_{N2-P})$ is around 6, 5, and 4 \AA , respectively. This indicates unlike sulphate, biphosphate ion

stabilizes in the vicinity of the motif. However $d_{C\alpha-P}$ as a function of time over the simulated trajectory (Figure 2e) indicates that biphosphate ion also moves back and forth in the neighbourhood of the motif, while it prefers to stay mostly $\leq 10 \text{ \AA}$ from the motif. The distribution of time interval ($f(t_a)$) (Figure 2f) is also sharply peaked for low t_a with mean time $\langle t_a \rangle$ ($\sim 0.5 \text{ ns}$), as in the sulphate cases.

3.3 | Basis of conformational preferences

Now we turn our focus to different factors which may be responsible for conformational preferences of different motif residues.

3.3.1 | Effect of sequence

The shifts in conformational preferences due to sulphate or biphosphate are sensitive to the sequence of the motif residues. It is

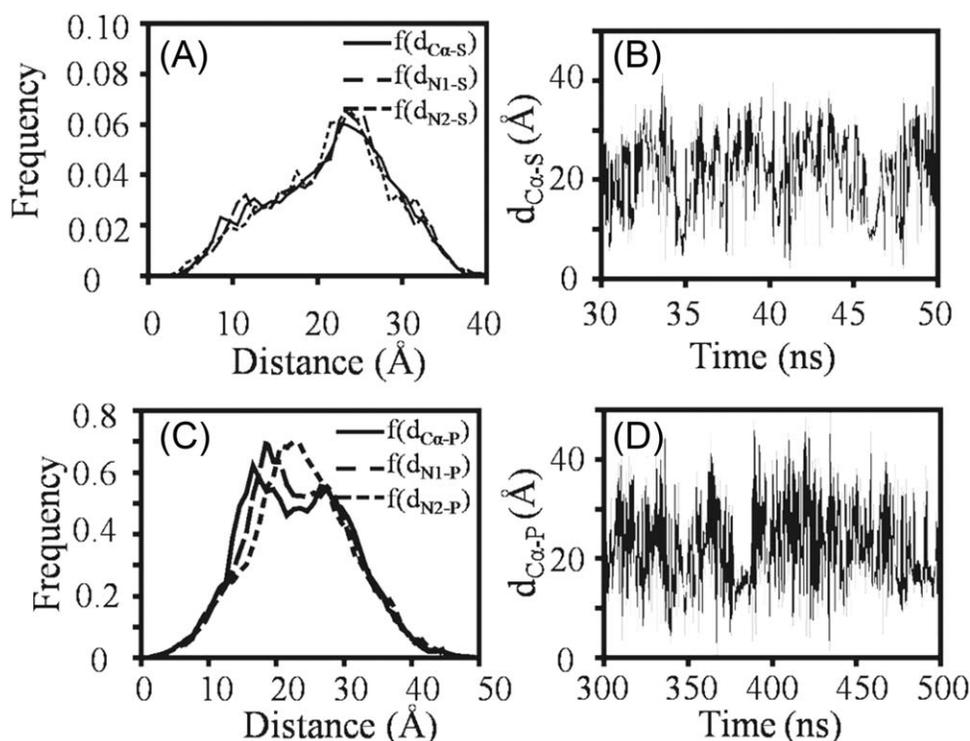


FIGURE 3 (a) The frequency distributions of distances; $f(d_{C\alpha-S})$ (solid line), $f(d_{N1-S})$ (long dashed line), and $f(d_{N2-S})$ (short dashed line) between sulphur atom of sulphate and C^α, N1, and N2 atoms, respectively of C^αNN motif in LYD fragment. (b) The distances ($d_{C\alpha-S}$) between C^α and sulphate as functions of time for LYD fragment. (c) The frequency distributions of distances; $f(d_{C\alpha-P})$ (solid line), $f(d_{N1-P})$ (long dashed line), and $f(d_{N2-P})$ (short dashed line) between phosphorus atom of biphosphate and C^α, N1, and N2 atoms, respectively of C^αNN motif in LYD fragment. (d) The distances ($d_{C\alpha-P}$) between C^α and biphosphate as functions of time for LYD fragment

interesting to note that the same residue shows different conformational preferences in WS condition depending on the sequence. For instance, the conformational preference of S in SQT fragment (RH) is quite in contrast to that of S in SNQ fragment (β). The polar S behaves distinctly in the presence of different anions. In WS condition S in SQT and SNQ fragments shows no shift in conformational preference with respect to the WOS condition. But, in case of SRS fragment both S show biphosphate induced shifts in conformational preferences.

3.3.2 | Effect of anion

The residue conformations of a given motif are sensitive to the type of anion. We replace sulphate by biphosphate in different fragments to see the effect of anions on conformational preference of C^αNN motif. We illustrate in details the case of LYD fragment. We choose this fragment, for the fragment has hydrophobic, polar, and charged residues. We observe that secondary structural preference of the motif residues is different in the sulphate replaced by biphosphate (RP) case with respect to WS case (Table 2a). Without any anion L132 fluctuates between RH (45%) and β (55%) conformations. In WS case L132 is primarily β conformation but it prefers RH conformation in RP case selecting the minor WOS conformation. Thus, L132 shows shifts in conformation from almost equally probable WOS conformations. However, the shift is dependent on the type of anion. Y133 primarily prefers RH conformation both in WS and WOS conditions. In RP case Y133 fluctuates between RH (48%) and β (52%) conformations (Table 2a). Thus, Y133 lacks

particular conformational preference in RP case, unlike that of WS. D134 prefers RH in WOS, β conformation in WS and RH conformation in RP condition, indicating shift in conformation in WS but no shift in RP.

The distance frequencies, $f(d_{C\alpha-S})$, $f(d_{N1-S})$, and $f(d_{N2-S})$ in Figure 3a and $d_{C\alpha-S}$ as a function of time in Figure 3b for LYD fragment in WS condition indicate that sulphate ion prefers to stay at large distances from the C^αNN motif atoms with back and forth motion in the neighbourhood of the motif. When sulphate is replaced by biphosphate, the frequencies $f(d_{C\alpha-P})$, $f(d_{N1-P})$, and $f(d_{N2-P})$ in Figure 3c and $d_{C\alpha-P}$ as a function of time in Figure 3d show similar behaviour of the biphosphate as the sulphate in WS condition. This behavior is in striking contrast with that of biphosphate in the SRS fragment in WP condition (Figure 2d). This may be due to the enhanced flexibility experienced by Y133 due to replacement with biphosphate.

The differences in fluctuations of the backbone dihedrals are shown in Figure 4 for LYD segment both in WS and replacement by biphosphate conditions. Distribution of backbone dihedral ϕ for the motif residues L, Y, and D are defined as $H_L(\phi)$, $H_Y(\phi)$, and $H_D(\phi)$, and that of ψ are defined as $H_L(\psi)$, $H_Y(\psi)$, and $H_D(\psi)$, respectively. $H_L(\phi)$ shows single peak distribution both in WS and RP condition and not alter much due to replacement of sulphate by biphosphate (Figure 4a). $H_L(\psi)$ also shows single peak distribution both in WS and RP condition (Figure 4b). Depending on the anion the peak positions are different, $\psi \sim 140^\circ$ in WS and $\psi \sim -20^\circ$ in RP condition. $H_Y(\phi)$ shows single broad peak in WS, but becomes bimodal distribution in RP condition

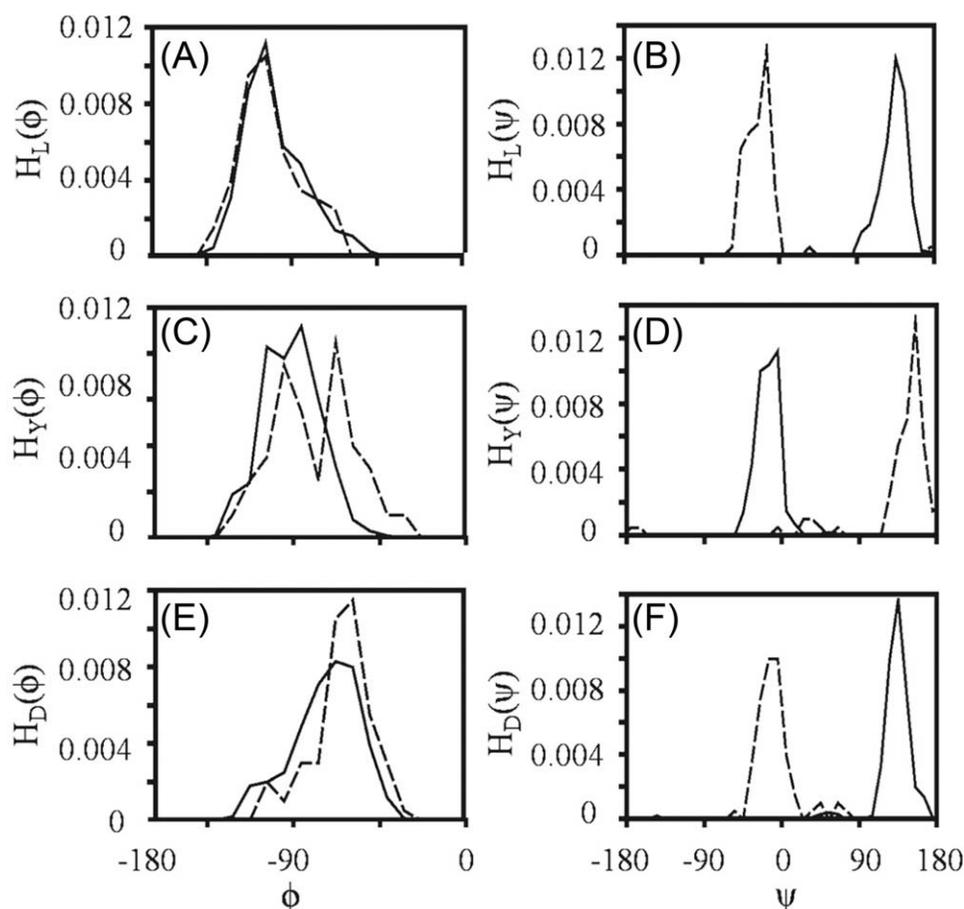


FIGURE 4 Distributions of backbone dihedral angle in WS (solid line) and RP (dashed line) conditions for motif residues in LYD fragment: (a) $H_L(\phi)$, (b) $H_L(\psi)$, (c) $H_Y(\phi)$, (d) $H_Y(\psi)$, (e) $H_D(\phi)$, and (f) $H_D(\psi)$

(Figure 4c). $H_Y(\psi)$ shows single peak distribution both in WS and RP condition (Figure 4d). Here again the peak positions are different, $\psi \sim -10^\circ$ in WS and $\psi \sim 170^\circ$ in RP condition. The splitting of $H_Y(\phi)$ peak is consistent with fluctuations of conformation of Y133 in RP condition. $H_D(\phi)$ and $H_D(\psi)$ show single peak distribution both in WS and RP condition (Figures 4e and f). While the $H_D(\phi)$ peaks nearly coincide, the $H_D(\psi)$ peaks are $\psi \sim 135^\circ$ in WS and $\psi \sim -10^\circ$ in RP condition.

Similar sensitivities of conformational preferences depending on the type of anion have been observed for other fragments as well (Table 2a). In SQT fragment, S does not show any shift in conformation both in WS and RP cases compared to WOS. However, this residue shows substantial fluctuations in conformation in RP condition. Q shifts to RH conformation in WS, but lacks conformational preference in RP case. T shifts to β conformation in WS condition, but in RP condition there is no shift in conformation. In SNQ fragment S shifts to β conformation both in WS and RP cases as compared to WOS. N and Q do not show any shift in conformation both in WS and RP cases compared to WOS. In GIH fragment G shifts to RH conformation in WS, but lacks conformational preference in RP case. I shifts to RH conformation in WS condition, but in RP condition there is no shift in conformation. H does not show any shift in conformation in WS but lacks conformational preference in RP case.

Upon ligand binding the changes in fluctuations in dihedral angles cost conformational free energy and entropy.²⁵ Earlier experimental and theoretical works show that the conformational thermodynamics govern the stability of ligand bound proteins. We study the conformational free energy and entropy changes of the motif residues due to replacement of sulphate by biphosphate from the backbone dihedral distributions of motif residues, $H_\alpha(\phi)$ and $H_\alpha(\psi)$, using conformational thermodynamics.²⁵⁻²⁷ Conformational thermodynamics data for the backbone dihedrals of the biphosphate bound motif residues with respect to the sulphate bound case for different peptide fragment are shown in Table 3. ΔG_i^{conf} and $T\Delta S_i^{\text{conf}}$ indicates the change in conformational free energy and entropy of the *i*th residue, obtained by summing over the backbone dihedrals. $\Delta G_i^{\text{conf}}(\phi)$, $\Delta G_i^{\text{conf}}(\psi)$, and $T\Delta S_i^{\text{conf}}(\phi)$, $T\Delta S_i^{\text{conf}}(\psi)$ indicates the change in conformational free energy and entropy of the *i*th motif residues due to backbone dihedral ϕ and ψ , respectively. The total change in conformational free energy (ΔG^{conf}) and entropy ($T\Delta S^{\text{conf}}$) of the motif is obtained by sum of the changes of individual residues of the motif. The negative change in conformational free energy and entropy indicates the conformational stability due to replacement by biphosphate as well as ordering, whereas the positive change indicates the instability and disorder due to biphosphate replacement.

The free energy changes are not significant in none of the residues in case of SQT. There is a large amount of disorder due to both the

TABLE 3 Conformational thermodynamics change of motif residues of peptide fragment and motif in presence of biphosphate with respect to presence of sulphate

Peptide fragment	Motif residue	Change in conformational free energy			Change in conformational entropy			(a) Total change in conformational free energy (ΔG^{conf}) (b) Total change in conformational entropy ($T\Delta S^{\text{conf}}$)
		ΔG_i^{conf} in kJ/mol	$\Delta G_i^{\text{conf}}(\phi)$ in kJ/mol	$\Delta G_i^{\text{conf}}(\psi)$ in kJ/mol	$T\Delta S_i^{\text{conf}}$ in kJ/mol	$T\Delta S_i^{\text{conf}}(\phi)$ in kJ/mol	$T\Delta S_i^{\text{conf}}(\psi)$ in kJ/mol	
SQT	S134	0.46	0.10	0.36	2.76	1.55	1.21	(a) 0.48
	Q135	0.04	0.02	0.02	0.88	0.73	0.15	(b) 3.45
	T136	-0.02	0.36	-0.38	-0.19	0.36	-0.55	
SNQ	S78	-2.09	1.41	-3.5	2.19	0.81	1.38	(a) -1.389
	N79	-0.002	-0.019	0.017	-0.57	-0.034	-0.536	(b) 4.29
	Q80	0.703	0.08	0.623	2.67	0.64	3.31	
GIH	G148	3.41	3.41	0	3.03	0.84	2.19	(a) 13.38
	I149	14.6	0.71	13.89	3.19	1.26	1.93	(b) 5.01
	H150	-4.63	8.54	-13.17	-1.21	-1.27	0.06	
LYD	L132	3.46	6.30	-2.84	1.51	0.38	1.13	(a) 13.98
	Y133	7.65	7.47	0.18	-1.63	1.06	-2.69	(b) 1.98
	D134	2.87	2.73	0.14	2.10	2.73	-0.63	

backbone dihedrals of S. Overall, the motif shows slight increase in free energy and large amount of disorder by the anion replacement. In SNQ fragment free energy change is negative for S and positive for Q due to backbone dihedral ψ . However backbone dihedral ψ introduce disorder in both S and Q. Due to anion replacement this motif gets the conformational stability along with large amount of disorder. In GIH fragment G and I destabilized and disordered due to backbone dihedral ϕ and ψ . However, H is stabilized and ordered due to backbone dihedral ψ and ϕ , respectively. Overall large positive change in free energy and entropy indicates instability and enhanced disorder of this motif due to anion replacement. In LYD fragment all the motif residues destabilized due to backbone dihedral distribution, particularly ϕ (Figure 4). There is also disorder of L and D due to the backbone dihedral ψ and ϕ , respectively. Overall due to anion replacement in this case also this motif gets conformational instability and disorder. Overall the thermodynamic data indicate that the free energy costs are typically very large and positive due to the presence of biphosphate instead of sulphate in case of LYD and GIH fragment. This means that the anion replacement is not conformationally favourable for these fragments, while SQT and SNQ are largely unaffected due to replacement. The anion replacement also introduces disorder in all the cases.

3.3.3 | Solvent distribution

We study the radial distribution of the oxygen (O) atoms of solvent water molecules around C^α , N1, and N2 atoms of the motif in presence and absence of sulphate and biphosphate. For SQT fragment, a prominent first peak in distribution $g_{C^\alpha}^{\text{WS}}(r)$ is observed around $r = 4 \text{ \AA}$ with respect to C^α atom in WS case (Figure 5a). In absence of sulphate the distribution $g_{C^\alpha}^{\text{WOS}}(r)$ with respect to C^α atom is also similar as $g_{C^\alpha}^{\text{WS}}(r)$ (Figure 5b). There is no shift in conformational preference for S residue corresponding to this atom. There is no peak in radial distributions $g_{N1}^{\text{WS}}(r)$ and $g_{N2}^{\text{WS}}(r)$ with respect to N1 and N2 atom, respectively, in presence of sulphate. In WOS case also there is no peak in distributions

$g_{N1}^{\text{WOS}}(r)$ and $g_{N2}^{\text{WOS}}(r)$. However, conformational shift is observed in T containing N2 but no shift is observed for Q having N1. In case of SNQ fragment, first peak of $g_{C^\alpha}^{\text{WS}}(r)$ is around $r = 4 \text{ \AA}$ (Figure 5c). In Figure 5d, there is also a peak in distribution of $g_{C^\alpha}^{\text{WOS}}(r)$ around $r = 4 \text{ \AA}$. Conformational shift is not observed at this position for the motif residue S. There is a peak of $g_{N1}^{\text{WS}}(r)$ at $r = 3 \text{ \AA}$ but there is no peak in distribution of $g_{N1}^{\text{WOS}}(r)$. Conformational shift is also not observed for N residue at this position. There is no peak in distributions of $g_{N2}^{\text{WS}}(r)$ and $g_{N2}^{\text{WOS}}(r)$. Here also no conformational shift takes place Q residue. The distributions of water molecules in the presence and absence of the anion does not show difference in GIH and LYD fragments as shown in Supporting Information Figure S9. However conformational shift is observed for I with N1 in case of GIH fragment and for D with N2 in case of LYD fragment.

In case of SRS fragment, a prominent first peak in distribution $g_{C^\alpha}^{\text{WP}}(r)$ is observed around $r = 4 \text{ \AA}$ with respect to C^α atom in WP case (Figure 5e). In WOP case the distribution $g_{C^\alpha}^{\text{WOP}}(r)$ has peak around $r = 3 \text{ \AA}$ (Figure 5f). Conformational shift is observed for the residue S at this position. There is no peak in distribution $g_{N1}^{\text{WP}}(r)$ and $g_{N2}^{\text{WP}}(r)$ with respect to N1 and N2, respectively, in presence of biphosphate. In WOP case the distributions, $g_{N1}^{\text{WOP}}(r)$ and $g_{N2}^{\text{WOP}}(r)$ are peaked around $r = 3 \text{ \AA}$ (Figure 5f). Conformational shift is not observed in case of R residue at N1 position, but observed for S residue at N2 position. Thus, changes in water distributions around the motif C^α , N1, and N2 atoms are not correlated to the changes in conformational preferences of the motif residues.

3.3.4 | Flanking residues

We examine the conformational preferences of a couple of flanking residues (Supporting Information Table S2a and Supporting Information Table S2b) in both N and C-terminal in the vicinity of the motif residues. The data in these tables show that the flanking residues show conformational fluctuations without the anion. However, the anions

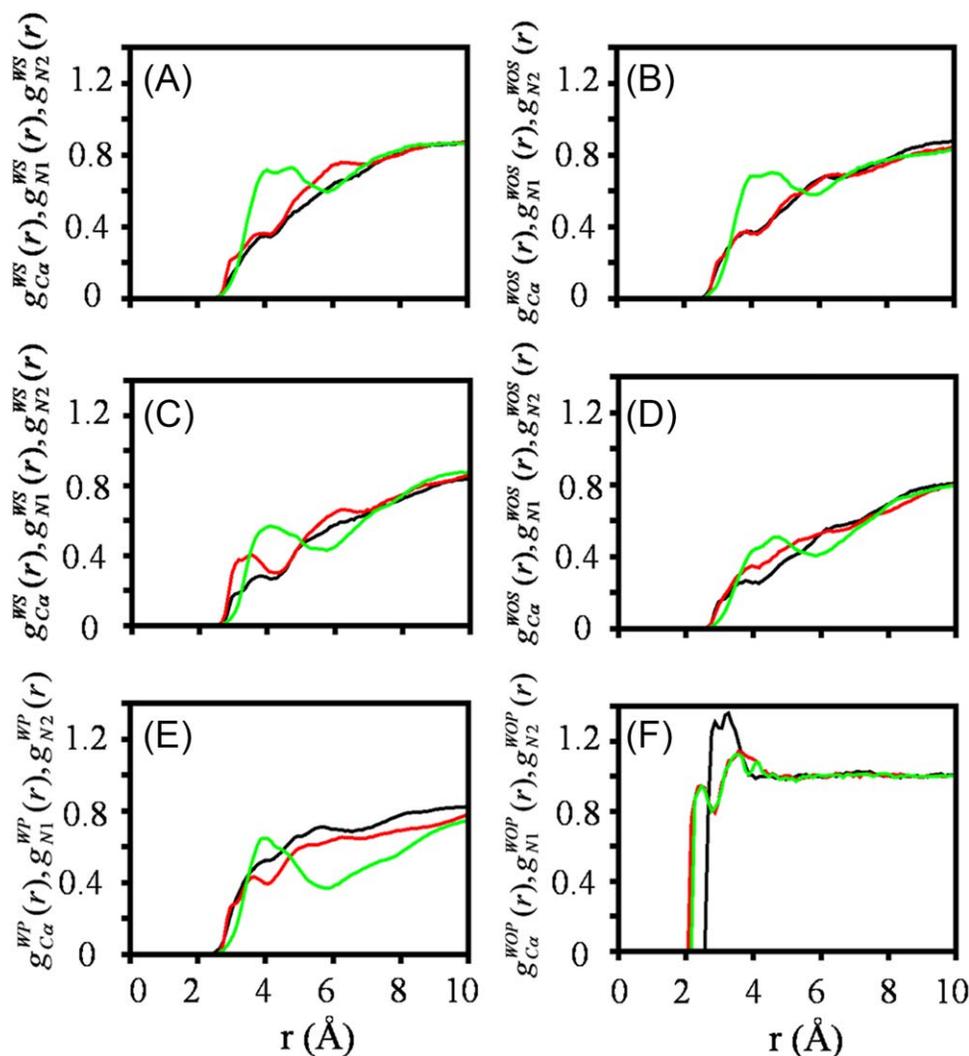


FIGURE 5 Radial distributions of the oxygen (O) atom of solvent water molecule around C^α (green), N1 (red), and N2 (black) atoms of the motif: for SQT fragment (a) in presence [$g_{C^\alpha}^{WS}(r)$, $g_{N1}^{WS}(r)$, $g_{N2}^{WS}(r)$] and (b) in absence [$g_{C^\alpha}^{WOS}(r)$, $g_{N1}^{WOS}(r)$, $g_{N2}^{WOS}(r)$] of sulphate; (c) and (d) similar quantities for SNQ fragment; (e) for SRS fragment in presence [$g_{C^\alpha}^{WP}(r)$, $g_{N1}^{WP}(r)$, $g_{N2}^{WP}(r)$] and (f) in absence [$g_{C^\alpha}^{WOP}(r)$, $g_{N1}^{WOP}(r)$, $g_{N2}^{WOP}(r)$] of biphosphate

give stability to one of the conformations without anion. The anion induced conformations are in general agreement to the crystal structure data. In SQT fragment N132 and alanine (A133) are N-terminal and Q137 and tryptophan (W138) are C-terminal residues, all of which have loop conformation in sulphate bound crystal structure. Supporting Information Table S2a shows shift in conformational preferences in all the residues except A133. In SNQ fragment, the N-terminal residues are R76 and Q77, both being in loop conformation and the C-terminal residues are T81 and R82 both having helix conformation in the crystal structure. The simulated trajectories show no shift in conformational preferences in all these residues except R82 (Supporting Information Table S2a). The N-terminal flanking residues of GIH (L146 and T147), and C-terminal residues A151 and D152 are all having loop conformation in crystal structure. Out of these residues only A151 does not show any shift in conformation (Supporting Information Table S2a). None of the LYD N-terminal residues (D130 and H131, loop conformation in crystal structure) and C-terminal residues of these fragments

(valine, V135, and V136, helix in crystal structure), shows shift in conformational preferences (Supporting Information Table S2b). In SRS fragment, V63 and methionine, M64 are the N-terminal residues and glutamine, E68 and R69 are the C-terminal residues, all being loop in the crystal structure (Supporting Information Table S2b). Here, except E68, no other residue shows shift in conformational preference between WP and WOP conditions. Overall, the conformational preferences of the flanking residues do not seem to be well correlated to those of motif residues.

The shift of conformation in the presence of anion may be interpreted as the conformation without the anion is not favourable and the residue needs to undergo conformational rearrangement to accommodate the anion. On the other hand, no shift in conformational preference indicates that the anion can be readily accommodated. This way one can identify the residues primarily responsible for anion recognition. In case of SNQ fragment where all the residues are polar, the major conformational state in WOS condition has been stabilized by

the sulfate ion in WS condition. This implies that the WOS conformation of all three residues is favourable for sulphate binding. In case of SQT fragment the first two residues do not show shift in conformational preferences in WS condition as compared to the WOS condition. Similarly basic *H* in GIH fragment, hydrophobic *L*, and polar *Y* in LYD and basic *R* in SRS do not show shifts in conformation. All these results point to the fact that the polar residues are favourable for sulphate recognition as found for the synthetic peptides.^{16,18,19} Interestingly, the basic residues also adopt favourable conformation for anion recognition. The conformational preferences for majority of the motif residues show shift in conformation in the presence of anions. This indicates that the anions can act as conformational switches in these residues. More importantly, the switching of the conformations can be controlled by the residue sequences and the type of anion for a given residue.

4 | CONCLUSIONS

To summarize, our classical force field based simulations indicate that the motif residues fluctuate between different conformations without anion, one of which is stabilized by the anion. The anion induced conformational preference in C^αNN motif residues changes are sequence specific. Although the sulphate ion is not stabilized in the vicinity of the motif atoms, the back and forth motion of the ion with very short time scale between successive approaches aids the structural changes. On the other hand the biphosphate ion is stabilized in the vicinity of the motif residues, leading to their conformational preferences. For a given sequence, the conformational preferences of the residues are dependent on the nature of the anion as well. However, the preference of the conformation is not sensitive to the water distribution around the motif residues and the flanking residues. The residues showing no shift in conformational preferences by the anion are likely to be the ones responsible for anion recognition. These are mostly polar and basic residues. The microscopic method of identifying the anion recognizing residues may be useful to understand ligand recognition in general. The possibility of tuning conformational preferences by different anions may be useful for device applications based on conformational tunability of biomacromolecules.^{35,36}

ACKNOWLEDGEMENTS

RB and PP thank the TEQIP-II, MAKAUT, WB and DBT-BIF for funding. PP acknowledges Mr. Suvankar Ghosh for helping force-field parameterization. PP is also thankful to SNBNCBS, Kolkata for giving the computational facility. MG and JC thank DST for financial support.

ORCID

Jaydeb Chakrabarti  <http://orcid.org/0000-0003-0753-3259>

REFERENCES

- [1] Denessiouk KA, Johnson MS, Denesyuk AI. Novel C^αNN structural motif for protein recognition of phosphate ions. *J Mol Biol.* 2005; 345:611–629.
- [2] Chakrabarti P. Anion binding sites in protein structures. *J Mol Biol.* 1993;234:463–482.
- [3] Kobayashi N, Go N. ATP binding proteins with different folds share a common ATP-binding structural motif. *Nat Struct Biol.* 1997;4:6–7.
- [4] Kinoshita K, Sadanami K, Kidera A, Go N. Structural motif of phosphate-binding site common to various protein superfamilies: all-against-all structural comparison of protein-mono-nucleotide complexes. *Protein Eng.* 1999;12:11–14.
- [5] Denessiouk KA, Johnson MS. When fold is not important: a common structural framework for adenine and AMP binding in 12 unrelated protein families. *Proteins.* 2000;38:310–326.
- [6] Denessiouk KA, Rantanen VV, Johnson MS. Adenine recognition: a motif present in ATP-, CoA-, NAD-, NADP-, and FAD-dependent proteins. *Proteins.* 2001;44:282–291.
- [7] Watson JD, Milner-White EJ. A novel main-chain anion-binding site in proteins: the nest. A particular combination of phi, psi values in successive residues gives rise to anion-binding sites that occur commonly and are found often at functionally important regions. *J Mol Biol.* 2002;315:171–182.
- [8] Denesyuk AI, Denessiouk KA, Korpela T, Johnson MS. Functional attributes of the phosphate group binding cup of pyridoxal phosphate-dependent enzymes. *J Mol Biol.* 2002;316:155–172.
- [9] Denessiouk KA, Johnson MS. "Acceptor-donor-acceptor" motifs recognize the Watson-Crick, Hoogsteen and Sugar "donor-acceptor-donor" edges of adenine and adenosine-containing ligands. *J Mol Biol.* 2003;333:1025–1043.
- [10] Milner-White EJ, Nissink JW, Allen FH, Duddy WJ. Recurring main-chain anion-binding motifs in short polypeptides: nests. *Acta Crystallogr D Biol Crystallogr.* 2004;60:1935–1942.
- [11] Torrance GM, Leader DP, Gilbert DR, Milner-White EJ. A novel main chain motif in proteins bridged by cationic groups: the niche. *J Mol Biol.* 2009;385:1076–1086.
- [12] Wintjens R, Rooman M. Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. *J Mol Biol.* 1996;262:294–313.
- [13] Struhl K. Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. *Trends Biochem Sci.* 1989;14:137–140.
- [14] Berg JM. Zinc fingers and other metal-binding domains. Elements for interactions between macromolecules. *J Biol Chem.* 1990;265: 6513–6516.
- [15] Holm L, Sander C. The FSSP database: fold classification based on structure-structure alignment of proteins. *Nucleic Acids Res.* 1996; 24:206–209.
- [16] Sheet T, Supakar S, Banerjee R. Conformational preference of 'CalphaNN' short peptide motif towards recognition of anions. *PLoS One.* 2013;8:e57366.
- [17] Sheet T, Banerjee R. Sulfate ion interaction with 'anion recognition' short peptide motif at the N-terminus of an isolated helix: a conformational landscape. *J Struct Biol.* 2010;171:345–352.
- [18] Sheet T, Banerjee R. The '(CNN)-N-alpha' motif: an intrinsic lover of sulfate and phosphate ions. *Rsc Adv.* 2016;6:54129–54141.
- [19] Sheet T, Ghosh S, Pal D, Banerjee R. Computational design of model scaffold for anion recognition based on the 'Calpha NN' motif. *Biopolymers.* 2017;108:1–15.
- [20] Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput.* 2008;4:435–447.

- [21] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem*. 2005;26:1701–1718.
- [22] Lindorff-Larsen K, Piana S, Palmo K, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*. 2010;78:1950–1958.
- [23] Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996;14:33–38; 27–38.
- [24] Allen DJT. *Computer Simulation of Liquids*. Oxford: Oxford University Press; 1991:400.
- [25] Das A, Chakrabarti J, Ghosh M. Conformational contribution to thermodynamics of binding in protein-peptide complexes through microscopic simulation. *Biophys J*. 2013;104:1274–1284.
- [26] Das A, Chakrabarti J, Ghosh M. Thermodynamics of interfacial changes in a protein-protein complex. *Mol Biosyst*. 2014;10:437–445.
- [27] Sikdar S, Chakrabarti J, Ghosh M. A microscopic insight from conformational thermodynamics to functional ligand binding in proteins. *Mol Biosyst*. 2014;10:3280–3289.
- [28] Dormitzer PR, Sun ZY, Wagner G, Harrison SC. The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. *Embo J*. 2002;21:885–897.
- [29] Yao M, Ose T, Sugimoto H, et al. Crystal structure of 1-aminocyclopropane-1-carboxylate deaminase from *Hansenula saturnus*. *J Biol Chem*. 2000;275:34557–34565.
- [30] Atwell S, Ultsch M, De Vos AM, Wells JA. Structural plasticity in a remodeled protein-protein interface. *Science*. 1997;278:1125–1128.
- [31] Symmons MF, Jones GH, Luisi BF. A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation. *Structure*. 2000;8:1215–1226.
- [32] Pieper U, Kapadia G, Mevarech M, Herzberg O. Structural features of halophilicity derived from the crystal structure of dihydrofolate reductase from the dead sea halophilic archaeon, *Haloferax volcanii*. *Structure*. 1998;6:75–88.
- [33] Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*. 2007;450:913–916.
- [34] Charlier C, Khan SN, Marquardsen T, et al. Nanosecond time scale motions in proteins revealed by high-resolution NMR relaxometry. *J Am Chem Soc*. 2013;135:18665–18672.
- [35] Lee TS, Radak BK, Harris ME, York DM. A two-metal-ion-mediated conformational switching pathway for HDV ribozyme activation. *Acs Catal*. 2016;6:1853–1869.
- [36] Knipe PC, Thompson S, Hamilton AD. Ion-mediated conformational switches. *Chem Sci*. 2015;6:1630–1639.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Patra P, Ghosh M, Banerjee R, Chakrabarti J. Anion induced conformational preference of C^αNN motif residues in functional proteins. *Proteins*. 2017;85:2179–2190. <https://doi.org/10.1002/prot.25382>

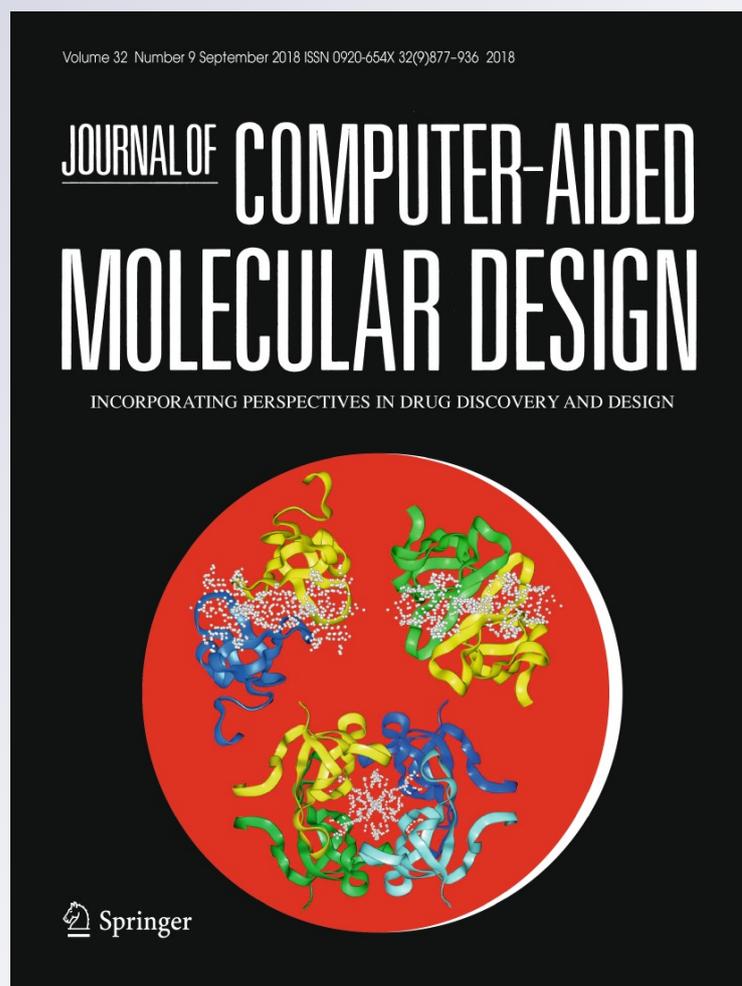
Quantum chemical studies on anion specificity of C α NN motif in functional proteins

Piya Patra, Mahua Ghosh, Raja Banerjee & Jaydeb Chakrabarti

**Journal of Computer-Aided
Molecular Design**
Incorporating Perspectives in Drug
Discovery and Design

ISSN 0920-654X
Volume 32
Number 9

J Comput Aided Mol Des (2018)
32:929-936
DOI 10.1007/s10822-018-0157-3



Your article is protected by copyright and all rights are held exclusively by Springer Nature Switzerland AG. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



Quantum chemical studies on anion specificity of C^αNN motif in functional proteins

Piya Patra¹ · Mahua Ghosh² · Raja Banerjee¹ · Jaydeb Chakrabarti^{2,3} Received: 16 May 2018 / Accepted: 29 August 2018 / Published online: 4 September 2018
© Springer Nature Switzerland AG 2018

Abstract

Anion binding C^αNN motif is found in functionally important regions of protein structures. This motif based only on backbone atoms from three adjacent residues, recognizes free sulphate or phosphate ion as well as phosphate groups in nucleotides and in a variety of cofactors. The mode of anion recognition and microscopic picture of binding interaction remains unclear. Here we perform self-consistent quantum chemical calculations considering sulphate and phosphate bound C^αNN motif fragments from crystal structures of functional proteins in order to figure out microscopic basis of anion recognition. Our calculations indicate that stability and preference of the anion in the motif depends on the sequence of the motif. The stabilization energy is larger in case of polar residue containing motif fragment. Nitrogen atom of the polar residue of motif mainly participates in the coordination at the lowest energy levels. Anion replacement decreases stabilization energy along with coordination between motif atoms and oxygen atoms of anion shifted to higher energies, suggesting preference of the motif residues to specific anion. Our analysis may be helpful to understand microscopic basis of interaction between proteins and ionic species.

Keywords Anion recognition · Hydrogen bond · Coordination · Partial density of state · Polar residue

Introduction

Binding of anion to active sites of protein triggers biochemical processes, like enzymatic reaction, substance transportation and signal transduction [1–7]. Based on the type of

interaction and the binding site structure revealed by X-ray crystallography and NMR, there are different kind of anion recognition motifs [8–12]. Among them sulphate (SO₄²⁻) and phosphate (PO₄³⁻) binding motifs receive attention due to the importance of sulphate and phosphate bound proteins in biochemical activities [8, 9, 13]. Recognition mechanism and interactions of sulphate or phosphate with the binding motif is not yet understood well.

Bioinformatics studies with a number of functional proteins [8] and recent biophysical studies with model peptide fragments [14–16] show that three residue C^αNN motif can recognize sulphate and phosphate and form functional interfaces. This motif is constituted of backbone C_{i-1}^α (C^α), N_i (N1) and N_{i+1} (N2) atoms of three successive residues for recognizing anion [8, 13]. According to the crystal structure, out of four oxygen atoms (O) of anion (sulphate or phosphate), two interact with the motif atoms via hydrogen bonding [8, 13]. Though backbone atoms interact with anion, bioinformatics studies indicate preference of amino acid residue in the motif [8]. Recent experimental studies based on synthetic peptides show that presence of polar residues enhances anion recognition [14, 15]. Our recent molecular dynamics (MD) simulation studies on C^αNN

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10822-018-0157-3>) contains supplementary material, which is available to authorized users.

✉ Piya Patra
pp9875@gmail.com

✉ Jaydeb Chakrabarti
jaydeb@bose.res.in

¹ Maulana Abul Kalam Azad University of Technology, West Bengal (Formerly known as WBUT), BF-142, Sector-I, Salt Lake, Kolkata 700064, India

² Department of Chemical, Biological and Macro-Molecular Sciences, S.N. Bose National Centre for Basic Sciences, Sector III, Block JD, Salt Lake, Kolkata 700106, India

³ The Thematic Unit of Excellence on Computational Materials Science, S.N. Bose National Centre for Basic Sciences, Sector-III, Block JD, Salt Lake, Kolkata 700106, India

motif containing peptide fragments from functional proteins show that anion induces conformational changes in the motif residues in sequence dependent manners and also depends on the type of anion [17].

Since the backbone atoms do not possess any distinct electrical character, the stability of the anion is likely to involve polarization of the backbone atoms due to anion. All the previous theoretical studies are based on classical force field calculations which cannot capture polarization effects. The polarization effects can be best described by quantum chemical (QC) calculations considering electrons of different atomic species in the system [18–22]. QC calculations for large bio-molecules are challenging due to the involvement of a large number of electrons. Such calculations are usually done by truncating the system to only active parts [20, 21], while the solvent treated as a continuum [18–22]. In this study we perform QC calculations on anion (sulphate or phosphate) coordinated C^αNN motif fragments along with two flanking residues on both terminals, taken from the crystal structures of functional proteins. We perform the calculation in vacuum, as earlier studies show that metal ion coordination to protein fragments are well described by vacuum QC calculation [21]. Although the proteins structures are experimentally known, we optimize the structures in QC calculation, since we consider only a fragment of the protein in our calculations.

Different systems are as follows: (1) SNQ fragment (Pyridoxal 5'-phosphate (PLP) dependent enzyme 1-Aminocyclopropane-1-carboxylate deaminase [23]; PDB ID: 1F2D) with motif residues Serine(S)78, Asparagine(N)79, Glutamine(Q)80; (2) SQT fragment (Rhesus rotavirus VP4 [24]; PDB ID: 1KQR) with motif residues, Serine(S)134, Glutamine(Q)135, Threonine(T)136 denoted as SQT fragment; (3) LYD fragment (Polynucleotide phosphorylase [25]; PDB ID: 1E3H), with motif residues are Leucine(L)132, Tyrosine(Y)133, Aspartic acid(D)134 denoted as LYD fragment; (4) GIH fragment (Extracellular domain of growth hormone receptor [26]; PDB ID: 1AXI) with motif residues Glycine(G)148, Isoleucine(I)149, Histidine(H)150; and (5) SRS fragment (Dihydrofolate reductases [27]; PDB ID: 1VDR) with motif residues, Serine(S)65, Arginine(R)66 and Serine(S)67. Here the first four fragments bind to sulphate and the last fragment is phosphate binding motif. We have reported classical force field calculation on these systems [17]. This gives us opportunity to directly compare the QC results with those from force field calculations.

We consider all the fragments both with (denoted by WS and WP as with sulphate and bi-phosphate, respectively) and without (denoted by WOS and WOP) anion. At physiological pH, the phosphate (PO₄³⁻) is mostly available in dibasic (HPO₄²⁻) form. So we consider HPO₄²⁻ instead of PO₄³⁻ in our system as done in our previous study [17]. We also study cases where SO₄²⁻ is

replaced by HPO₄²⁻ (SRP) and vice versa (PRS). The systems are shown in Table 1. We have added neutral cap, N-terminus by acetyl (–COCH₃) group and the C-terminus by N-methylamide (–NHCH₃) group to the flanking residues as done earlier [28–31]. We have added the missing hydrogen atoms using visual molecular dynamics (VMD) based on CHARMM27 topologies [32]. The peptide segments are optimized in vacuum using density functional form of electronic ground state [33–35] in the Vienna Ab initio Simulation Package (VASP) with plane wave (PW) [34, 36] basis and projector augmented-wave (PAW) [33, 35] potentials. DFT-PW calculations have been used to describe systems, like bio-molecules in vacuum [37, 38] and water clusters [39], where non-bonded interactions are better described. We show that the stability of C^αNN motif is governed by the coordination of motif atoms with oxygen atoms of anion. The sequence of the motif as well as the type of anion influences the coordination pattern and consequently, the stabilization energy.

Theoretical methods

We use the plane wave basis with cutoff energy of 400 eV and periodic box size of 30 × 30 × 30 Å³ in the VASP [34, 35, 40, 41]. We have used PAW potential with PBE as exchange-correlational functional and a gamma centered κ mesh of 1 × 1 × 1. We perform ionic relaxation for all the peptide fragment atoms followed by self consistent calculation for relaxing the electronic degrees of freedom to determine the electronic ground state energy.

Root mean square deviation (RMSD) calculation

We calculate the RMSD between the crystal structure of motif fragment and optimized structure of motif fragment in presence and absence of anion by superposing the two structures using the Pymol [42] software and evaluating root mean square displacement between the atoms of the two motif fragments.

Table 1 Systems taken for studies

System	PDB IDs	Sequences
1	1F2D	(77)Gln- Ser(C^α) - Asn(N) -Gln(N)-Thr(81)
2	1KQR	(133)Ala- Ser(C^α) - Gln(N) - Thr(N) -Gln(137)
3	1E3H	(131)Hse- Leu(C^α) - Tyr(N) - Asp(N) -Val(135)
4	1AXI	(147)Thr- Gly(C^α) - Ile(N) - Hse(N) -Ala(151)
5	1VDR	(64)Met- Ser(C^α) - Arg(N) - Ser(N) -Glu(68)

The residues shown in bold are motif residues

Calculation of hydrogen bond

We consider motif atoms as a hydrogen bond donor and oxygen atoms of anion as a hydrogen bond acceptor. We use the criteria of donor (D) to acceptor (A) atom distance $< 3.5 \text{ \AA}$ and D-H-A angle $> 120^\circ$ for consideration of hydrogen bonding interaction.

Calculation of partial density of state (PDOS)

We calculate the density of states (DOS), representing the number of electronic states per unit of energy (in eV), in the optimized geometry. The total DOS is calculated from the energy levels of each individual state by projecting the DOS on all the orbitals. We then compute the different contribution of the different orbitals for each and individual atoms and calculate the partial density of state (PDOS) of the atom.

Calculation of stabilization energy

We calculate the single point energy of the ion–electron system in the optimized structure of peptide fragments both in presence and absence of anion. The stabilization energy (SE), ΔE_{SE} is defined as the energy difference between the anion bound motif and the free components. We also calculate the energy difference by replacing the anion.

Results

Optimized structures

The optimized geometries of SNQ motif fragment in WOS, WS and SRP conditions are shown in Fig. 1a–c. With respect to the crystal structure the RMSD of optimized structures of SNQ motif fragment in WOS (Fig. 1a), WS (Fig. 1b) and SRP (Fig. 1c) is 0.585 \AA , 0.310 \AA and 1.237 \AA , respectively. Oxygen atoms of the anion are numbered according to the crystal structure of the corresponding peptide fragments [23–27]. Here in WS condition C^α and N1 atoms of the motif forms hydrogen bonds with O3 and N2 atom of the motif forms hydrogen bond with O4 atom (Fig. 1b). In SRP the

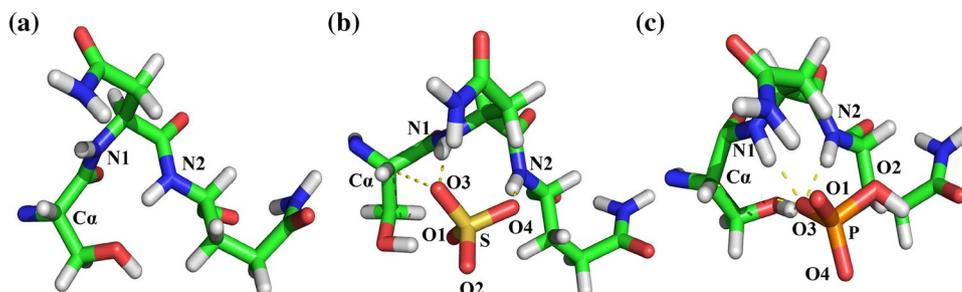
hydrogen bond pattern is different; N1 and N2 atoms of the motif form hydrogen bonds with O3 atom of bi-phosphate and hydrogen bonding interaction between C^α atom of the motif and oxygen atom of anion is disrupted (Fig. 1c).

Similar patterns are observed in other motif fragments. In the optimized geometry of WS and WP fragments, most of the hydrogen bonds between anion and motif atoms are improved in terms of hydrogen bond distance and angle (Table S1). C^α and N1 atoms of the motif form hydrogen bonds with one oxygen atom of the anion concurrently and N2 atom of the motif forms hydrogen bond with another oxygen atom of anion. In SRP—SQT fragment C^α and N1 atoms of the motif form hydrogen bonds with O1 atom of bi-phosphate, while hydrogen bond between N2 atom of the motif and oxygen atom of anion is disrupted (Fig. S1). In SRP—LYD case C^α and N1 atoms of the motif form hydrogen bonds with O3 atom and N2 atom of the motif forms hydrogen bond with O2 atom of bi-phosphate (Fig. S1). In SRP—GIH case C^α atom of the motif forms hydrogen bond with O2 atom and N1 and N2 atoms of the motif form hydrogen bond with O3 atom of bi-phosphate (Fig. S1). Due to replacement of bi-phosphate by sulphate (PRS—SRS) hydrogen bonding interaction between C^α atom of the motif and oxygen atoms of anion is disrupted. N1 and N2 atom of the motif forms hydrogen bond with O2 and O4 atom of sulphate, respectively (Fig. S1). Therefore, hydrogen bonding patterns between motif atoms and anion change due to replacement of anion.

Coordination of anion

We calculate PDOS contribution of the peptide and the capping atoms of the peptide fragment across the energy spectra. The energy values are given with respect to the Fermi energy. Representative cases are shown for SNQ for WOS (Fig. 2a), WS (Fig. 2b) and SRP (Fig. 2c) conditions. Similarly data for SRS fragment in WOP, WP and PRS are shown in Fig. 2d–f respectively. PDOS of all capping atoms (black line) is found to be negligible compared to all peptide atoms (red line) in the entire energy range. Additional data are shown in Supporting Information. Thus our QC results are not affected by the capping. The PDOS of different atoms of

Fig. 1 Optimized geometry of C^α NN motif residues in SNQ fragment: **a** in absence of sulphate (WOS), **b** in presence of sulphate (WS) and **c** in presence of bi-phosphate replacing the sulphate (SRP). The hydrogen bonds between motif atoms and oxygen atoms of anion are shown in yellow dotted line



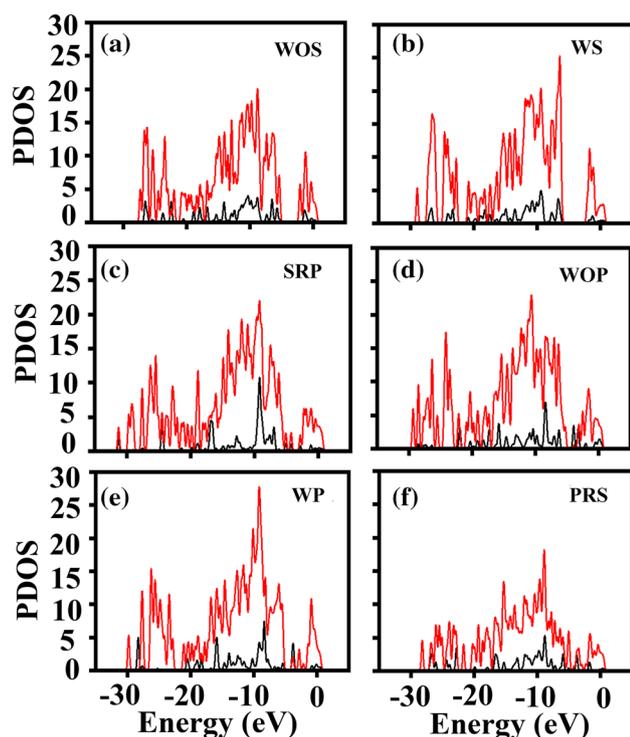


Fig. 2 PDOS of all peptide atoms barring the capping atoms (red) and the capping atoms (black) for different energy for SNQ fragment: **a** in absence of sulphate (WOS), **b** in presence of sulphate (WS) and **c** in presence of bi-phosphate replacing the sulphate (SRP) and for SRS fragment: **d** in absence of bi-phosphate (WOP), **e** in presence of bi-phosphate (WP) and **f** in presence of sulphate replacing the bi-phosphate (PRS)

sulphate and bi-phosphate are shown in Fig. 3a, b, respectively. The PDOS peaks of oxygen atoms of anion are mainly distributed in the energy range -10 eV to -30 eV. PDOS of individual atoms of the motif residues S, N and Q in WS—SNQ fragment are shown in Fig. S2a–c, respectively as representative case. Here also we find major PDOS contributions of the atoms come from the energy range -5 eV to -30 eV.

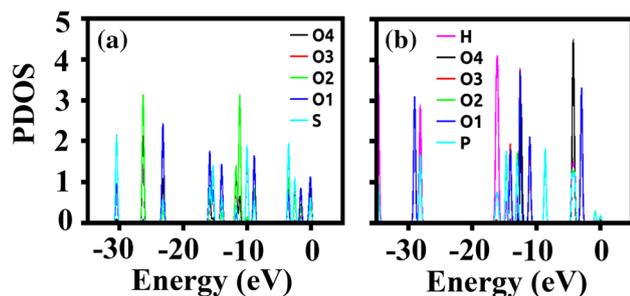


Fig. 3 PDOS of **a** sulphate and **b** bi-phosphate for different energy

We analyze the simultaneous PDOS contributions of motif atoms of the peptide fragment and the atoms of the anionic species in WS and WP cases. If the peaks of PDOS of a motif atom and an atom of the anion lie within 0.04 eV (thermal energy at room temperature), the atom pairs are taken to coordinate at an energy, given by the mean of the peak values. Let us now consider the motif atoms in different sequences. Figure 4a–c show the PDOS of the motif atoms of SNQ fragment along with the sulphur and oxygen atoms of sulphate. It shows simultaneous PDOS contribution of the motif atoms and oxygen atoms of the sulphate take place within the energy range -10 eV to -16 eV. The lowest energy coordination takes place between N2 atom of the motif and O4 atom of sulphate at $E = -15.30$ eV. Both C^α and N1 atoms of the motif coordinate with O3 atom of sulphate at a slightly higher energy (-14.27 eV). These oxygen atoms of anion are also found to form hydrogen bond with the motif atoms C^α , N1 and N2 in SNQ fragment.

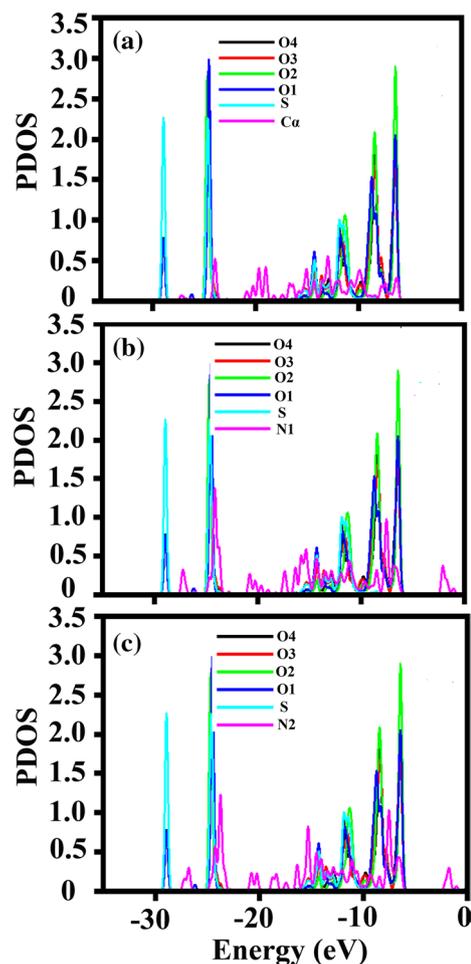


Fig. 4 PDOS of S (cyan), O1 (blue), O2 (green), O3 (red) and O4 (black) atoms of bound sulphate and motif atoms of peptide (magenta): **a** C^α of S78, **b** N1 of N79 and **c** N2 of Q80 in case of SNQ fragment in WS condition as function of energy

The other cases are shown in Figs. S3–S6. In case of SQT fragment PDOS of the motif atoms along with the sulphur and oxygen atoms of sulphate are shown in Fig. S3a–c. The lowest energy coordination takes place between C $^{\alpha}$ and N1 atom of the motif and O2 atom of sulphate both at E = -11.70 eV. N2 atom of the motif coordinates with O1 atom of sulphate at slightly higher energy, E = -11.04 eV. In LYD fragment N1 and C $^{\alpha}$ atom of the motif coordinates with O4 atom of sulphate at lowest energies E = -14.90 eV and -14.85 eV, respectively (Fig. S4a–c). At E = -13.48 eV, N2 atom of the motif coordinates with O2. N1 atom of the motif also coordinates with O4 atom of sulphate at higher energy -13.44 eV and -11.28 eV. In GIH fragment lowest energy coordination takes place between N2 atom of the motif and O3 atom of sulphate at E = -16.52 (Fig. S5a–c). At E = -15.01 eV, N2 atom of the motif also coordinates with O3 atom of sulphate. N1 atom of the motif coordinates with O4 atom of sulphate at higher energy, E = -10.06 eV. Fig. S6a–c show the PDOS of the motif atoms of bi-phosphate bound SRS fragment along with the phosphorus and oxygen atoms of bi-phosphate. Here, the lowest energy coordination takes place between N2 atom of the motif and O3 atom of bi-phosphate at E = -11.95 eV. Both C $^{\alpha}$ and N1 atoms of the motif coordinate with O2 atom of bi-phosphate at a slightly higher energy (-11.12 eV).

Overall in case of SNQ, GIH and SRS fragments the lowest energy coordination takes place between oxygen atom of anion and N2 atom of the motif. For SQT and LYD fragments C $^{\alpha}$ and N1 atoms of motif coordinate with one oxygen atom of the anion simultaneously at lowest energy. In all these motif fragments, the coordinated oxygen atoms of anion are also found to be hydrogen bonded with the corresponding motif atoms (Table S1). We also study the simultaneous PDOS contributions of side chain atoms of the motif residues and the oxygen atoms of the anion. In case of S of SNQ fragment, other than C $^{\alpha}$ atom of motif residue side chain atoms C $^{\beta}$ and O $^{\gamma}$ coordinate with O1 atom at E = -14.5 eV (Fig. 5a). Side chain N $^{\delta 2}$ atom of motif residue N also coordinates with O3 atom of sulphate at E = -13.8 eV (Fig. 5b). Thus the side chain atoms of S and N in SNQ fragment participate in coordination with sulphate. We do not observe side chain atom coordination in other motif sequences.

PDOS of C $^{\alpha}$ and N atoms of terminal flanking residues Q77 and T81 in case of SNQ fragment along with atoms of sulphate are shown in Fig. 5c, d. Simultaneous PDOS contributions of oxygen atoms of sulphate and any of these atoms of flanking residue Q77 and T81 are not observed (Fig. 5c, d). In the other cases also terminal flanking residues do not contribute simultaneously with the PDOS of oxygen atoms of anion in the ground state energy range. Therefore, the terminal flanking residues do not participate to coordinate with the oxygen atoms of anion.

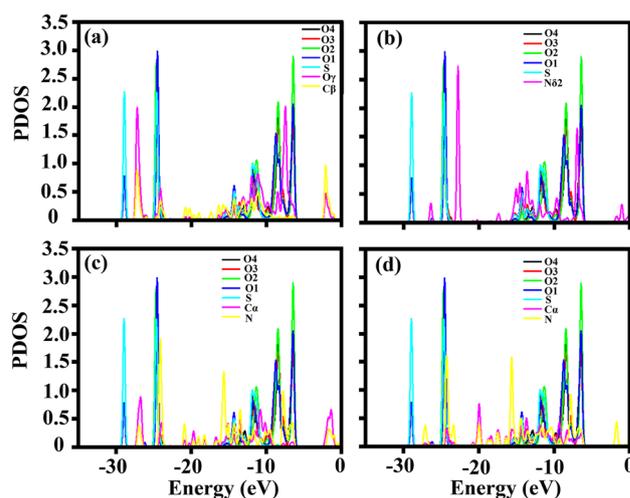


Fig. 5 PDOS of S (cyan), O1 (blue), O2 (green), O3 (red) and O4 (black) atoms of bound sulphate and **a** side chain atoms of S, **b** side chain atom of N, **c** C $^{\alpha}$ and N atoms of the flanking residue Q77 and **d** C $^{\alpha}$ and N atoms of the flanking residue T81 in case of SNQ fragment in WS condition as function of energy

Table 2 Stabilization energy of C $^{\alpha}$ NN motif containing fragments in presence and absence of the anion

Fragment containing the C $^{\alpha}$ NN motif residues	Stabilization energy in kcal/mol $\Delta E_{SE} = E_{WS/WP} - (E_{WOS} + E_{SO4}/E_{HPO4})$
SNQ fragment (WS)	-439.4
SQT fragment (WS)	-436.6
LYD fragment (WS)	-428.8
GIH fragment (WS)	-422.1
SRS fragment (WP)	-257.3

Stabilization energy

We show in Table 2 the stabilization energy ΔE_{SE} of the systems due to presence of anion. Table 2 indicates that in all the cases the stability of the system enhances in presence of anion. However, ΔE_{SE} is sequence dependent: SNQ > SQT > LYD > GIH. For both the SNQ and SQT fragments all the coordinating residues are polar. The additional stability of SNQ compared to SQT may be due to coordination of side chain atoms of S and N as well. LYD fragment gives more stabilization energy as compared to GIH fragment. In LYD fragment L is non-polar, Y is polar and D is acidic in nature. In GIH fragment G and I are non-polar and H is also non-polar at physiological pH. Presence of polar residue in LYD fragment gives larger stabilization energy as compared to GIH fragment. In bi-phosphate bound SRS fragment stabilization energy is less than sulphate bound cases, though there are two polar and one basic residue.

Replacement of anion

We also study the coordination of the motif atoms with the oxygen atoms of anion by replacement of anion. The coordination scenario changes due to anion replacement. In case of SRP—SNQ fragment PDOS of N atoms of the motif are found to appear simultaneously with the PDOS of O3 atom of bi-phosphate (Fig. S7). The lowest energy coordination takes place between N1 atom of the motif and O3 atom of bi-phosphate at $E = -7.526$ eV. N2 atom of the motif coordinates with O3 atom of bi-phosphate at higher energy, $E = -6.982$ eV (Fig. S7). On the other hand in presence of sulphate (WS—SNQ) motif atoms coordinate with oxygen atoms of sulphate at the energy range -16 eV to -14 eV. Thus, in contrast to the sulphate bound case, here the coordination takes place at higher energy.

We find similar trends in all other cases. In SRP—SQT fragment energetically the deepest coordination takes place between N1 atom of the motif and O1 atom of bi-phosphate at $E = -7.142$ eV. N2 atom of the motif coordinates with O3 atom of bi-phosphate at higher energy, $E = -6.738$ eV (Fig. S8a–c). In SRP—LYD fragment the lowest energy coordination takes place between N1 atom of the motif and O3 atom of bi-phosphate at $E = -7.939$ eV. N2 atom of the motif coordinates with O2 atom of bi-phosphate at slightly higher energy (-7.883 eV) (Fig. S9a–c). The lowest energy coordination takes place between N1 atom of the motif and O3 atom of bi-phosphate at $E = -11.123$ eV in case of SRP—GIH fragment (Fig. S10a–c). In case of PRS—SRS fragment lowest energy coordination takes place between N2 atom of the motif and O4 atom of sulphate at $E = -8.427$ eV. N1 atom of the motif coordinates with O2 atom of sulphate at higher energy (-7.603 eV) (Fig. S11a–c). Thus due to replacement of anion lowest energy coordination between motif atoms and oxygen atoms of anion take place at higher energy for all the fragments. Due to anion replacement in the motif fragment stabilization energy decreases and relative stabilization energy also becomes positive (Table 3) in all the cases, indicating energetically unfavorable conditions.

Discussion

We also perform benchmark calculation for SQT fragment using Gaussian 03 [43] considering B3LYP functional with 6-31G (2d, 2p) basis set which is normally used for biomolecules systems [20, 44–46]. We find that the dihedral angles (φ , ψ , χ_1) in the VASP optimized and crystal structures are similar (Table S2). Consequently we take the crystal structure and perform partial optimization of SQT peptide fragment both in presence and absence of sulphate by fixing all the non-hydrogen atoms. In the optimized geometry we analyze the orbital contributions of motif atoms and anion

Table 3 Stabilization energy of C $^\alpha$ NN motif containing fragments on replacement of anion

Fragment containing the C $^\alpha$ NN motif residues	Relative stabilization energy due to replacement of anion ($E_{\text{SRP/PRS}} + E_{\text{SO}_4/\text{HPO}_4}$) - ($E_{\text{WS/WP}} + E_{\text{HPO}_4/\text{SO}_4}$) in kcal/mol
SNQ fragment (SRP)	441.2
SQT fragment (SRP)	280.0
LYD fragment (SRP)	308.9
GIH fragment (SRP)	531.2
SRS fragment (PRS)	104.5

over the molecular orbitals of the system. We also study the charge distribution in the optimized geometry both in presence and absence of sulphate using natural population analysis (NPA). We compute the occupancies of all the atomic orbitals in the molecular orbitals. Our calculations indicate that the simultaneous contribution of motif atoms and oxygen atoms of sulphate takes place within the energy range -14 eV to -18 eV in case of SQT fragment (Fig. S12). N2 atom of the motif coordinates with O1 atom of sulphate at $E = -16.12$ eV. C $^\alpha$ and N1 atom of the motif coordinates with O2 atom of sulphate at $E = -15.52$ eV and -15.98 eV, respectively. The coordination is comparable to the VASP results, although the co-ordinations take place at lower energies. We also calculate the changes in natural charges on the atoms of the motif residues due to presence of sulphate. We find that the amount of charges lost by sulphate (0.218) in bound state is distributed over the peptide atoms as shown in Table S3.

Since the same systems have been studied by us earlier by classical force field simulations [17], it is worth to compare the present results to those from simulations. Our classical force field based molecular dynamics (MD) simulation studies show anion induced conformational preference in the motif residues [17]. A conformational preference of a residue has been defined as percentage of right handed helix (RH), beta (β), left handed helix (LH) and coil (C) over simulated conformation in equilibrium. It is observed that conformational preferences of motif residues fluctuate without anion and anion typically stabilizes one of the residue conformations observed in absence of the anion [17]. If the conformational preference in the presence of anion is one as the minor population in the absence of anion, the residue is said to undergo conformational switching induced by the anion, else the residue does not undergo conformational switching. Tables S4 and S5 show in details the conformational preference of motif residues and the coordination pattern in these residues. The motif residue in each peptide fragment which does not show conformational switching due to presence of anion coordinates with the oxygen atom of the anion at lowest energy.

These residues are mostly polar in nature. The residues undergoing conformational switching have coordinated at higher energy levels in general. Thus the residues having no conformational switching are the ones in energetically favorable condition for coordination.

The connection between conformational switch and coordination energy is revealed by anion replacement as well, although the coordination energies are shifted to higher values. Due to anion replacement conformational switching of the motif residues is not observed in case of SNQ, SQT and GIH fragment (Table S5) where we observe coordination with anion oxygen, albeit shifted at higher energy than WS/WP cases. Only L in LYD fragment undergoes conformational switching in SRP condition. This residue does not participate in coordination (Table S5). Similarly, S in SRS fragment undergoes conformational switching in PRS condition and it also does not participate in coordination.

According to the bioinformatics study [8] based on the crystal structure database analysis of the sulphate/phosphate binding proteins, the anion-binding C^αNN motif interacts with the anionic oxygen via backbone C^α and N atoms of the consecutive motif residues in a specific pattern, where both C^α and N1 atoms of the motif interact with one oxygen atom of the anion concurrently and N2 atom of the motif interacts with another oxygen atom of anion. In QC study we find that this stable interaction pattern mainly stabilizes the anion in the motif and gives the anion selectivity. Our studies also support the observation from previous biophysical studies [14, 15] based on the C^αNN motif containing synthetic peptide fragments that polar residues are the primary ones to give stability to the motif. Our studies suggest that the connection between conformational change and coordination of anion with the polar residue atoms is universal.

Conclusion

Our quantum chemical calculations show that the stability of C^αNN motif is governed by the coordination of motif atoms with oxygen atoms of anion. The sequence of the motif influences the coordination energy and stabilization energy. We find larger stabilization energy in case of motif having more polar residues. Stability of the motif also depends on the type of anion. The most nontrivial aspect of our results is the intimate connection between coordination and conformational preferences of the residues which has not been established so far to the best of our knowledge. We consider the motif residues from different proteins which have different functional sites, and, therefore, this study may be useful to get the microscopic insights into their functions. Similar analysis can be useful for understanding stability of other ligand binding motifs as well.

Acknowledgements PP thanks to Dr. Manas Mondal for helping in VASP calculation. RB and PP thank the TEQIP-III, MAKAUT, WB and DBT-BIF for funding. PP is also thankful to SNBNCBS, Kolkata for giving the computational facility. JC thanks DST for financial support.

References

- Gale PA, Perez-Tomas R, Quesada R (2013) Anion transporters and biological systems. *Acc Chem Res* 46(12):2801–2813
- Gale PA, Busschaert N, Haynes CJ, Karagiannidis LE, Kirby IL (2014) Anion receptor chemistry: highlights from 2011 and 2012. *Chem Soc Rev* 43(1):205–241
- Caballero A, Zapata F, Beer PD (2013) Interlocked host molecules for anion recognition and sensing. *Coord Chem Rev* 257(17–18):2434–2455
- Gale PA (2011) From anion receptors to transporters. *Acc Chem Res* 44(3):216–226
- Duke RM, Veale EB, Pfeffer FM, Kruger PE, Gunnlaugsson T (2010) Colorimetric and fluorescent anion sensors: an overview of recent developments in the use of 1,8-naphthalimide-based chemosensors. *Chem Soc Rev* 39(10):3936–3953
- Elmes RB, Jolliffe KA (2015) Anion recognition by cyclic peptides. *Chem Comm* 51(24):4951–4968
- Kral V, Rusin O, Shishkanova T, Volf R, Matejka P, Volka K (1999) Anion binding: from supramolecules to sensors. *Chem Listy* 93(9):546–553
- Denessiouk KA, Johnson MS, Denesyuk AI (2005) Novel C^αNN structural motif for protein recognition of phosphate ions. *J Mol Biol* 345(3):611–629
- Chakrabarti P (1993) Anion binding sites in protein structures. *J Mol Biol* 234(2):463–482
- Watson JD, Milner-White EJ (2002) A novel main-chain anion-binding site in proteins: the nest. A particular combination of phi, psi values in successive residues gives rise to anion-binding sites that occur commonly and are found often at functionally important regions. *J Mol Biol* 315(2):171–182
- Milner-White EJ, Nissink JW, Allen FH, Duddy WJ (2004) Recurring main-chain anion-binding motifs in short polypeptides: nests. *Acta Crystallogr* 60:1935–1942
- Wintjens R, Rooman M (1996) Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. *J Mol Biol* 262(2):294–313
- Sheet T, Supakar S, Banerjee R (2013) Conformational preference of ‘C^αNN’ short peptide motif towards recognition of anions. *PLoS ONE* 8(3):e57366
- Sheet T, Banerjee R (2016) The ‘C^αNN’ motif: an intrinsic lover of sulfate and phosphate ions. *RSC Adv* 6(59):54129–54141
- Sheet T, Ghosh S, Pal D, Banerjee R (2017) Computational design of model scaffold for anion recognition based on the ‘C^αNN’ motif. *Biopolymers* 108(1):e22921
- Sheet T, Banerjee R (2010) Sulfate ion interaction with ‘anion recognition’ short peptide motif at the N-terminus of an isolated helix: a conformational landscape. *J Struct Biol* 171(3):345–352
- Patra P, Ghosh M, Banerjee R, Chakrabarti J (2017) Anion induced conformational preference of C^αNN motif residues in functional proteins. *Proteins* 85(12):2179–2190
- Kagawa H, Mori K (1999) Molecular orbital study of the interaction between MgATP and the myosin motor domain: the highest occupied molecular orbitals indicate the reaction site of ATP hydrolysis. *J Phys Chem B* 103(34):7346–7352
- Blomberg MRA, Siegbahn PEM (2001) A quantum chemical approach to the study of reaction mechanisms of redox-active metalloenzymes. *J Phys Chem* 105(39):9375–9386

20. Sikdar S, Ghosh M, De Raychaudhury M, Chakrabarti J (2015) Quantum chemical studies on stability and chemical activities in calcium ion bound calmodulin loops. *J Phys Chem B* 119(46):14652–14659
21. Sikdar S, Ghosh M, De Raychaudhury M, Chakrabarti J (2014) Quantum chemical studies on the role of residues in calcium ion binding to calmodulin. *Chem Phys Lett* 605:103–107
22. Fukui K (1982) Role of frontier orbitals in chemical reactions. *Science* 218(4574):747–754
23. Yao M, Ose T, Sugimoto H, Horiuchi A, Nakagawa A, Wakatsuki S, Yokoi D, Murakami T, Honma M, Tanaka I (2000) Crystal structure of 1-aminocyclopropane-1-carboxylate deaminase from *Hansenula saturnus*. *J Biol Chem* 275(44):34557–34565
24. Dormitzer PR, Sun ZY, Wagner G, Harrison SC (2002) The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. *EMBO J* 21(5):885–897
25. Symmons MF, Jones GH, Luisi BF (2000) A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation. *Structure* 8(11):1215–1226
26. Atwell S, Ultsch M, De Vos AM, Wells JA (1997) Structural plasticity in a remodeled protein-protein interface. *Science* 278(5340):1125–1128
27. Pieper U, Kapadia G, Mevarech M, Herzberg O (1998) Structural features of halophilicity derived from the crystal structure of dihydrofolate reductase from the Dead Sea halophilic archaeon, *Haloferax volcanii*. *Structure* 6(1):75–88
28. Lepsik M, Field MJ (2007) Binding of calcium and other metal ions to the EF-hand loops of calmodulin studied by quantum chemical calculations and molecular dynamics simulations. *J Phys Chem B* 111(33):10012–10022
29. Wu EL, Mei Y, Han K, Zhang JZ (2007) Quantum and molecular dynamics study for binding of macrocyclic inhibitors to human alpha-thrombin. *Biophys J* 92(12):4244–4253
30. He X, Zhang JZ (2006) The generalized molecular fractionation with conjugate caps/molecular mechanics method for direct calculation of protein energy. *J Chem Phys* 124(18):184703
31. Zhang DW, Chen XH, Zhang JZ (2003) Molecular caps for full quantum mechanical computation of peptide-water interaction energy. *J Comput Chem* 24(15):1846–1852
32. Brooks BR, Brooks CL III, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caffisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30(10):1545–1614
33. Blochl PE (1994) Projector augmented-wave method. *Phys Rev B* 50(24):17953–17979
34. Kresse G, Hafner J (1993) Ab-initio molecular-dynamics for open-shell transition-metals. *Phys Rev B* 48(17):13115–13118
35. Kresse G, Joubert D (1999) From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys Rev B* 59(3):1758–1775
36. Car R, Parrinello M (1985) Unified approach for molecular dynamics and density-functional theory. *Phys Rev Lett* 55(22):2471–2474
37. Henkelman G, LaBute MX, Tung CS, Fenimore PW, McMahon BH (2005) Conformational dependence of a protein kinase phosphate transfer reaction. *Proc Natl Acad Sci USA* 102(43):15347–15351
38. Adhikari P, Wen AM, French RH, Parsegian VA, Steinmetz NF, Podgornik R, Ching WY (2014) Electronic structure, dielectric response, and surface charge distribution of RGD (1FUV) peptide. *Sci Rep* 4:5605
39. Silvestrelli PL, Parrinello M (1999) Structural, electronic, and bonding properties of liquid water from first principles. *J Chem Phys* 111(8):3572–3580
40. Kresse G, Hafner J (1993) Abinitio molecular-dynamics for liquid-metals. *Phys Rev B* 47(1):558–561
41. Kresse G, Furthmuller J (1996) Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys Rev B* 54(16):11169–11186
42. The PyMOL Molecular Graphics System. Version 2.0 edn. Schrodinger, LLC
43. Gaussian 03, Revision C, Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian, Inc., Wallingford CT
44. Trado-Rives J, Jorgensen WL (2008) Performance of B3LYP density functional methods for a large set of organic molecules. *J Chem Theory Comput* 4(2):297–306
45. Siegbahn PE, Borowski T (2006) Modeling enzymatic reactions involving transition metals. *Acc Chem Res* 39:729–738
46. Blomberg MR, Siegbahn PE (2006) Quantum chemistry applied to the mechanisms of transition metal containing enzymes-cytochrome c oxidase, a particularly challenging case. *J Comput Chem* 27:1373–1384

FULL PAPER

Self-assembly in amphiphilic macromolecules with solvent exposed hydrophobic moieties

Sutapa Dutta¹  | Piya Patra² | Jaydeb Chakrabarti^{1,3}

¹Department of Chemical, Biological and Macro-Molecular Sciences, S. N. Bose National Centre for Basic Sciences, Sector III, Block JD, Salt Lake, Kolkata, India

²Maulana Abul Kalam Azad University of Technology, West Bengal, Haringhata, Nadia, West Bengal, India

³Unit of Nanoscience and Technology-II and The Thematic Unit of Excellence on Computational Materials Science, S. N. Bose National Centre for Basic Sciences, Sector III, Block JD, Salt Lake, Kolkata, India

Correspondence

Sutapa Dutta, Department of Chemical, Biological and Macro-Molecular Sciences, S. N. Bose National Centre for Basic Sciences, Sector III, Block JD, Salt Lake, Kolkata 700106, India.

Email: sutapa.dutta@bose.res.in

Funding information

DST, Govt. of India

Abstract

Self-assembly by amphiphilic molecules with solvent exposed hydrophobic groups are relevant in biomolecular systems as well as in technological applications. Here we study such self-assembly in these systems using a model system of spherical particles having charge at core but solvent repelling surface, using Monte-Carlo simulations and mean field treatment. We find that solvophobicity mediated attraction leads aggregation, while electrostatic repulsions control stability of finite clusters. The aggregation threshold relates the parameters of two interactions through an algebraic dependence. The study also qualitatively explains experimental observations on aggregation of misfolded proteins and can be useful guide to tune stability of nm sized self-assembly in systems with exposed hydrophobic groups.

KEYWORDS

amphiphilic assembly, long-ranged repulsion, misfolded protein, Monte Carlo simulation, short-ranged attraction

1 | INTRODUCTION

Self-assembled structures at nanometer (nm) length scale are often formed by amphiphilic macromolecules with hydrophobic parts exposed to water,^[1–3] just opposite to their normal tendencies. Such self-assembled structures have applications in material science,^[4–6] and biomolecular systems as well.^[6–9] Although self-assembly is a manifestation of balance of competing attractive and repulsive forces,^[10,11] microscopic picture of stabilizing self-assembled structures in macromolecules with solvent exposed hydrophobic parts, is yet far from understood.

Amphiphilic biomacromolecules,^[12] like proteins have natural tendencies to expose their hydrophilic moieties to water and bury hydrophobic moieties to avoid water in native state. Clustering in native proteins, like lysozyme in aqueous medium^[13] has been explained using colloidal self-assembly where balance is achieved between short-ranged attraction and long-ranged repulsion (SALR).^[10,11] The short-ranged attraction results from osmotic pressure difference due to solvent depletion^[10,11] and the long-ranged repulsion are electrostatic. Sometimes due to mutation, physiological stress and changes in

physiochemical conditions proteins undergo deviations from native structures, exposing the hydrophobic parts as well to water.^[6,14] Extended fibril structure of peptide aggregations with exposed hydrophobic groups, known as amyloids, has been widely studied due to their relevance in neurodegenerative diseases.^[6–9] Neurotoxicity of amyloids depends on their sizes: Large amyloids are not neurotoxic.^[6–9] Experimental studies indicate that hydrophobic parts of the peptides are largely responsible for amyloid formation.^[15] However, the size of amyloids also depends on salt concentration and pH of the medium.^[15,16] These experiments suggest that the competing forces in these systems are attraction between hydrophobic moieties, known to lead to hydrophobic collapse^[17,18] and electrostatic repulsion. The competition is quite complicated in nature as has been revealed from recent experiments.^[19] These experiments report solubility of amyloids as a function sodium chloride salt concentration has a minimum at a given temperature, the self-assembled structures being stable above the solubility line. Colloids^[20–24] with attractive patches on particle surface has been used to model self-assembled structures of macromolecules with exposed hydrophobic moieties. However, there has been yet no attempt to the best of our knowledge to

understand the experimental observations on such systems, like amyloids, as balance of forces between hydrophobic and hydrophilic sites. It is not even clear if such structures are formed in equilibrium or driven by kinetics.^[25]

Solvophobic particles much larger than the solvent molecules have been considered in the past.^[26–30] These particles repel solvent molecules stabilizing vapor phase of low density around them. The vapor phase grows thicker as the liquid-gas phase coexistence is approached in subcritical condition.^[31] The solvent mediated effective interaction between two such solutes arises due to difference in Laplace pressure at the liquid-vapor interface.^[32] The attraction can be represented by a harmonic potential with spring constant^[32] α . Physically, harmonic dependence originates from stability of configuration where solvent molecules are pushed away by both the solvophobic surfaces. Here we consider model particles with spherical surface repelling solvent molecules which mimics exposed hydrophobic groups and a charge placed at the center to mimic the hydrophilic (charged) sites. In this qualitative study, we do not explicitly take solvent, rather the solvent effects are taken implicitly through effective pair potentials between the model solute particles, routinely done for soft matter systems.^[33,34] The effective harmonic interaction is taken between solvophobic surfaces. The screened electrostatic repulsion is taken between the core charges with screening length given by κ^{-1} , used for charge-stabilized colloids.^[35–37] Our Monte Carlo simulations studies coupled with mean field analysis show that finite nm sized clusters are stable below a threshold value of κ , while the system forms aggregate above the threshold for a given α . The threshold shows algebraic dependence between α and κ with exponent, independent of system parameters. The experimental observations are qualitatively understood from the model.

2 | METHODS

2.1 | Monte-Carlo (MC) simulations of model system

The interaction potential between two solvophobic particles of radius R , surrounded by vapor of radius Λ_0 , has been derived in Chakrabarti and Dutta.^[32] Excluding the linear terms, the effective interaction, $V_A(s) = \alpha s^2/2R$, so far as the surface-to-surface distance^[32] $s \leq 2\Lambda_0$. Here $\alpha = 8\pi\gamma[\Lambda_0/R - 1]$ is an effective spring constant where γ is surface tension at the gas-liquid interface of width^[32] Λ_0 . We take screened electrostatic repulsion between the core charges, known as Derjaguin-Landau-Verwey-Overbeek (DLVO) potential^[38] with the form, $V_C(s) = Xe^{-\kappa s}/s$. The prefactor of the electrostatic repulsion is considered as $X = (Z^2e^2e^{2\kappa R}/[1 + \kappa R]^2)1/4\pi\epsilon_0\epsilon$ and $\kappa = (\sum_i z_i^2 c_i e^2 N_A / \epsilon_0 \epsilon k_B T)^{1/2}$ is the inverse of Debye screening length. Here Z is the number of charges present in a model particle, e the electronic charge, ϵ dielectric constant of solvent, ϵ_0 the electric permittivity in vacuum, z_i the valence of each type (i) of ion present in the solution, including contributions from salt as well as from the model particles, k_B the Boltzmann constant, c_i the ionic concentration in mol/L and N_A the Avogadro's Number. $\sum_i z_i^2 c_i$, excluding the particle

charge, gives ionic strength(I). The electrostatic repulsion gets exponentially damped with a length scale κ^{-1} .

We consider particle radius as the length unit and energy unit $k_B T_R$ at room temperature ($T_R = 300$ K). The MC simulations have been performed in a cubic box of dimension $l = 14.0$ nm with the periodic boundary conditions in all three directions. There are $N=1000$ model particles of diameter ($d = 2R$) 1.0 nm and charge 28, typical, for instance, in proteins in neutral solvent at room temperature. The volume fraction (volume of a particle times the number density) of the model particles is 0.15. The particle positions are updated according to the Metropolis algorithm.^[39] The simulations are run for 100 000 MC steps out of which first ~ 30 000 MC steps are discarded for equilibration, as judged from the potential energy of the system and different quantities are averaged for configurations^[39] of last 70 000 steps. Initially we perform simulation for a fixed value of α and κ . Then, we gradually enhance magnitude of κ . That certain value of κ , for which all the particles come together to form aggregate is considered threshold value of κ , (κ_{th}) for that fixed α .

We identify clusters formed by the particles by arbitrarily choosing i th particle and then calculate distance r_{ij} for all of j th particles ($i \neq j$) with respect to i th particle over equilibrium trajectories, as detailed in Allen and Tildesley.^[39] If r_{ij} is less than a certain distance ($r_{cl} = 1.2d$) then these j th particles are considered to belong to the same cluster as the i th^[39] particle and total number of particles belong to that particular cluster gives cluster size, C_s . The process is repeated for other particles. We calculate number of clusters of different sizes in equilibrium configurations to yield distribution of cluster size $P(C_s)$ and compute the mean value ($\overline{C_s}$). We further compute the structural pair correlation functions^[40] ($g(r)$), which gives the probability of finding a neighbor atom around a central atom within a spherical shell of radii r and $r + \Delta r$. $g(r)$ is computed by binning the separation between different pair of particles averaged over equilibrium configurations.

3 | RESULTS AND DISCUSSION

Since we do not consider explicitly the solvent molecules, we tune parameters of the effective potential to mimic effects of changing solvent conditions. Let us consider the case of room temperature. Water at room temperature is in subcritical condition.^[41] We take $\Lambda_0/R \approx 1.27$ observed for Lennard-Jones systems in earlier studies in a subcritical liquid near phase-existence.^[32] Using experimental value^[42,43] of γ at room temperature, we obtain that $\alpha^* = \alpha R^2/k_B T_R = 3.0$. We observe that in presence of solvophobicity-mediated attraction only, particles tend to form aggregated structure, along with $\overline{C_s} \sim N$ spanning the system. In order to ensure equilibrium of our model system, we generate $\overline{C_s}$ for different (14) windows, each consisting of 5000 configurations. We estimate error in the mean cluster size from the standard deviations of the average cluster sizes over the windows. We find that for all of the cases, $\overline{C_s}$ shows similar values within an error of 1%, indicating equilibrium aggregation in the system. This aggregation of the order of system size is analogous to hydrophobic

collapse^[17,18] known in the literature. Similar system spanning clusters have been reported in SALR colloidal models as well.^[44,45]

Next, we consider effect of the electrostatic repulsion which competes with the solvophobic attraction. The harmonic potential will bring particles close together, but the electrostatic repulsion will take them apart. The self-assembly is due to balance between these forces. We increase κ^* for $\alpha^* (=3.0)$ fixed at the room temperature. Experimentally, κ^* is increased by increasing salt concentration for a fixed temperature, solvent dielectric constant, and charge at the center of the solute. Some representative cases are shown in Figure 1A-C. $P(C_s)$ has sharply peaked structures, showing uniformity in size distribution of

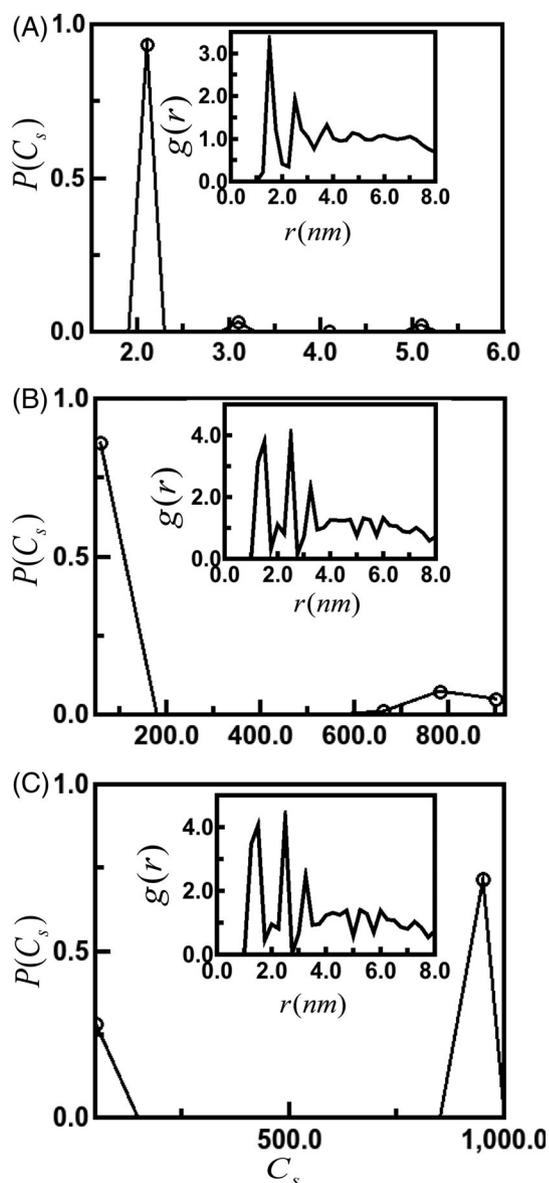


FIGURE 1 Probability distribution of cluster size $P(C_s)$ of model system, in presence of a fixed solvophobicity mediated attraction ($\alpha^* = 3.0$) and different κ^* . A, Finite size cluster for $\kappa^* = 2.0$. Inset: Radial distribution function $g(r)$ vs r plot. B, Coexisting small clusters and aggregation for $\kappa^* = 6.0$. Inset: $g(r)$ vs r plot. C, Large aggregation at $\kappa^* = 13.0$. Inset: $g(r)$ vs r plot

clusters. We find that at $\alpha^* = 3.0$, for low κ^* ($= 2.0$), $P(C_s)$ is unimodal (Figure 1A) and $\overline{C_s} \sim 2.0$, indicating formation of small clusters. The error in the mean cluster size is $\sim 1\%$. If we further increase screening ($\kappa^* = 6.0$), $P(C_s)$ (Figure 1B) is bimodal with a strong peak around $\overline{C_s} \approx 100$ and a much lower peak around $\overline{C_s} \approx 800 \sim N$. This is indicative of predominant presence of finite but large clusters at this screening. The bimodal cluster size distribution gives a larger error ($\sim 10\%$) in the mean cluster size. Finally, at $\kappa^* = 13.0$, $P(C_s)$ is primarily unimodal (Figure 1C) with $\overline{C_s} \sim N$. Thus, there is a threshold value of κ^* , κ_{th}^* , below which finite clusters are stable, while above this the peak corresponding to system spanning clusters is stronger than that for finite clusters.

The pair correlation functions $g(r)$ for different cases are shown in insets of Figure 1A-C. The $g(r)$ data in inset (a) shows a single strong peak characteristic of a liquid. However, the $g(r)$ data in insets (b) and (c) show second peak of almost equal magnitude as the first one in the presence of clusters, indicating strong correlations persisting up to second coordination shell in the clusters. However, the correlation peak decay within a short distance (~ 4 nm), typical for amorphous structures. Considering spherical amorphous clusters at mean scaled density as in our simulation, typical radius of a cluster of 100 particles is approximately 5 nm. This length is comparable to the decay length of the correlation.

Earlier^[31] studies show that at a given temperature in the subcritical region, the gas-liquid interface decreases as one moves away from the gas-liquid phase coexistence line by increasing liquid density. Thus decrease in α^* corresponds to increase in liquid solvent density away from the gas-liquid phase coexistence at a given subcritical temperature. We vary α^* to mimic the effect of changing solvent density at room temperature. We show in $\ln \alpha^*$ vs $\ln \kappa^*$ plot (Figure 2A), the regions of stability of finite clusters vis-a-vis aggregations spanning the system. The κ_{th}^* values for different values of α^* are shown by the symbols. The curves joining the lines are the best fits which we call the aggregation lines. The aggregation line is the boundary between two different self-assembled structures. Above the lines the particles form aggregates, while below it the particles form finite clusters. We find two different regimes of aggregation lines: For lower κ^* , we find from the slope of the log-log plot that $\alpha^* \sim (\kappa_{th}^*)^{-1.5}$ giving the equation of the aggregation line. For larger κ^* , the dependence is much weaker, the aggregation curve being $\alpha^* \sim (\kappa_{th}^*)^{-0.2}$.

We also consider cluster size variation by tuning the electrostatic repulsion at room temperature upon changing dielectric constant ϵ and charge of the particles, Z at a given salt concentration, while keeping the hydrophobic part unchanged. Dielectric constant of solvent is known^[46,47] to control self-assembly of charged colloids. Here, we model different solvent implicitly by changing dielectric constant, ϵ . We show $V_C(r)$ for different values of ϵ and Z in Figure 2B. We observed that $V_C(r)$ is longer ranged for $\epsilon=60$ than that for $\epsilon=40$ for the same value of $Z(=28)$. The prefactor of the electrostatic repulsion, X , decreases with increasing ϵ . However, inverse Debye screening length κ^* also decreases with increasing ϵ which renders the range of

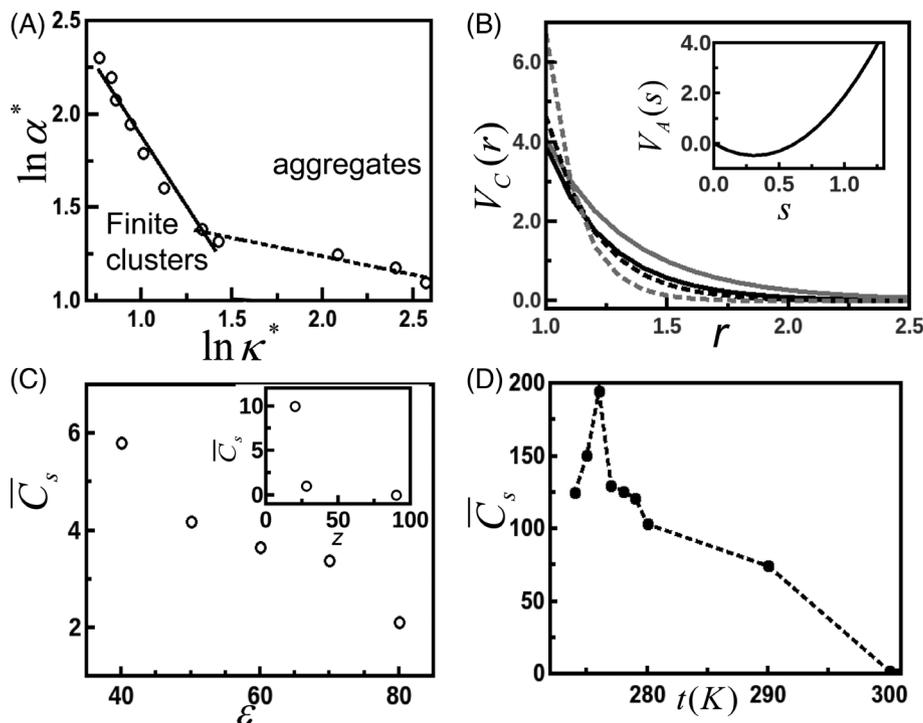


FIGURE 2 A, Phase diagram in α^* and κ^* in plane showing finite size cluster and large aggregated structure. The solid line shows $\alpha^* \sim (\kappa_{th}^*)^{-1.5}$ dependence, while dashed line represents $\alpha^* \sim (\kappa_{th}^*)^{-0.2}$ for aggregation threshold. B, $V_C(r)$ vs r for different values of ϵ and Z . Black solid: $\epsilon=40$, $Z=28$, gray solid: $\epsilon=60$ and $Z=28$. Black dashed: $\epsilon=80$, $Z=50$ and gray dashed: $\epsilon=80$, $Z=90$. Inset: $V_A(s)$ vs s plot for a fixed value of α^* ($=3.0$). C, Increase of \bar{C}_s with decrease of polarity (ϵ) of the solvent. Inset: Decrease in cluster size (\bar{C}_s) with increasing charge of the core (Z). D, Decrease of \bar{C}_s with increase of temperature (T)

the repulsion longer and outweighs the decrease in X . We observe (Figure 2C) that for $\alpha^* = 3.0$ the cluster size decreases as we increase ϵ . The enhancement of electrostatic repulsion, compared to the hydrophobicity mediated attraction, $V_A(s)$ (inset of Figure 2B) leads to decrease in cluster size. Similar enhancement in range of electrostatic repulsion by decreasing charge at the center (Z) (Figure 2B) leads to decrease in cluster size (inset of Figure 2C). Here also, the increase in X , is outweighed by increase in the Debye screening length with Z , leading to decrease in electrostatic potential compared to the hydrophobic attraction.

We now turn our attention to temperature dependence of self-assembly in our model. Since the dielectric constant of a normal liquid varies inversely with temperature^[40] the electrostatic repulsion is not sensitive to temperature. However, the hydrophobic potential is sensitive to temperature.^[32] The width of liquid-vapor interface around a solvophobic particle increases with temperature in a subcritical solvent due to increasing compressibility.^[43,48] On the other hand, the surface tension decreases with temperature.^[42] These two effects compete so that α^* has nonmonotonic dependence on temperature as shown earlier.^[30,32,49] We use known experimental values of surface tension of water at different temperatures.^[42] The experimental values of isothermal compressibility at different temperatures^[43] of water have been employed to calculate the interfacial width using relation derived in Meyer *et al.*^[27] Meyer *et al.*^[27] considers Lennard-Jones fluid, not polar water. However, compressibility is related to long wavelength fluctuations of density^[40] primarily governed by the positions of the oxygen atoms. Hence, orientation of water molecules in the solvent would not be important for compressibility. The data in Figure 2D indicates that for small value of κ^* ($=2.0$), the cluster size shows nonmonotonic dependence over temperature with maximum

at an intermediate temperature. We also observe that at physiological temperature (37°C) the cluster size is ~ 2.0 . In the low and high temperature region, the error in mean cluster size is around 1%, while that in the intermediate region close to maximum cluster size is $\sim 10\%$.

3.1 | Case of misfolded protein clusters

Since misfolded proteins form an important class of self-assembled structures with exposed hydrophobic parts,^[1-3] it is interesting to compare our model to the self-assembly of misfolded proteins. The most important feature of the model is the presence of gas bubble surrounding the solvophobic particles. Formation of gas bubble^[50,51] has been observed at interface of hydrophobic solid and water. The solvation behaviors of proteins are complex due to presence of hydrophilic residues in the vicinity of hydrophobic moieties. We perform full atomistic molecular dynamics (MD) simulations of polypeptides in several functional proteins^[52-56] (details of the peptide, simulation protocol, and the secondary structures along with root mean squared deviation plots of the peptides are in Table S1, Supplemental material and in Figures S1(a)-(j), respectively) having both hydrophobic and hydrophilic amino acid residues in water in physiological condition to calculate solvent distribution around different residues. The simulations have been carried out using the Amber99ILDN^[57] force field parameters in the Gromacs package.^[58,59]

We compute water distribution around different types of residues ($\rho(r)$) by binning the separation r between oxygen atom of water and backbone carbon atom of the polypeptide over equilibrated trajectories. It may be noted that the water distribution about different residues are resultant of different interactions between water molecules and the atoms in a residue. Figure 3A shows the distributions around

acidic, basic, polar, and hydrophobic residues. The figure indicates that water distribution is largest around acidic residues and then follows the basic residues, while polar and nonpolar residues are having water distribution weaker by about 30% compared to the acidic ones. We assume that picture of stabilization of low-density water vapor holds near the hydrophobic residues of polypeptide so that our model can be applied at least qualitatively. The hydrophobic effects are represented by a smooth solvent repelling spherical surface. The hydrophilic charged sites are given by charge at the center of the sphere.

The phase diagram in Figure 2A shows that the aggregates are stable with increasing α^* for a given κ^* . This is qualitatively consistent with observations that the stretch of hydrophobic parts of the peptides favors amyloid formation.^[15] Fluorescence correlation spectroscopic (FCS) studies^[15] indicate that size of clusters increases with increasing ionic strength of the solution by adding NaCl. We find from our simulation that the cluster size increases monotonically for a fixed value of α^* below κ_{th}^* . The $\ln \bar{C}_s$ vs $\ln I^*$ plot in Figure 3B for $\alpha^* = 3.0$ shows $\bar{C}_s \sim (I^*)^{2.3}$ dependence below $\kappa_{th}^* (=13.0)$ which is in qualitative agreement with FCS studies.^[15] Scattering measurement shows that salts of divalent metal ions at micromolar concentration^[16] leads β -amyloid to precipitate, forming large sized assembly. This is consistent with our results, for increase in valance of metal ion increases κ^* which would result in larger clusters. The variation of pH of the

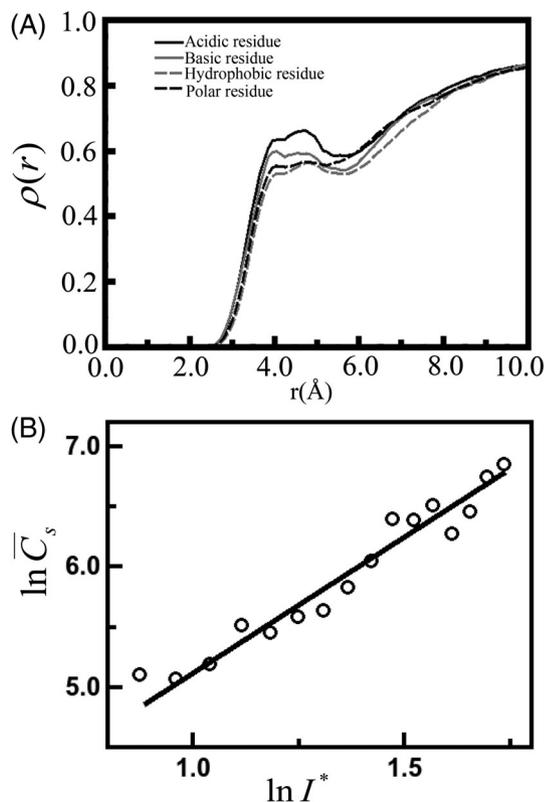


FIGURE 3 A, Distribution functions ($\rho(r)$) of water around hydrophilic and hydrophobic residues of protein, black solid line; acidic, gray solid line; basic, gray dashed; hydrophobic and black dashed; polar residues. B, $\ln \bar{C}_s$ vs $\ln I^*$ plot to show cluster size variation with ionic strength

medium leads to changes in charge state of the residues. FCS studies^[15] also show that increase in residue charges due to decrease in solution pH leads to smaller clusters. The reduction in cluster size with increasing Z , observed in our model system, is qualitatively consistent with these experimental results. The system spanning aggregates will be insoluble in aqueous phase. The minimum of solubility line in the salt concentration at a given temperature, reported in Adachi *et al.*^[19] can be understood from the algebraic dependence of α^* on κ_{th}^* . The experimental solubility curve is parabolic in salt concentration at a temperature. Since α^* has a maximum as a function of temperature, the algebraic dependence implies that there is a maximum in κ_{th}^* as well.

3.2 | Mean field theory

We account for the aggregation line from a simple theoretical framework. The structural information of thermodynamically stable system is given by the scattering function^[40] $s(q)$. Clustering in a system is represented by peak at low wave vector ($q_0 = 2\pi/R$) of $s(q)$, while for aggregation the peak shifts to $q \rightarrow 0$ limit. $s(q)$ is connected to of liquid correlation function,^[40] $c(q)$ via $s(q) = 1/[1 - \rho c(q)]$ for a system of density ρ . The peak of $s(q)$ corresponds to that of $c(q)$ as well. We examine the low q peak of $c(q)$ from a mean field treatment. At low density, the mean field approximation^[60] $c(q) \approx -\beta u(q)$, where $\beta = 1/k_B T$ and $u(q)$ is the Fourier Transform of the interaction potential. In the mean field approximation for our system, $c(q) = C_{electro}(q) + C_{solvo}(q)$. Here $C_{electro}(q)$ is Fourier Transform of direct correlation function for DLVO potential and $C_{solvo}(q)$ is that for solvophobic term. This is valid for long-ranged electrostatic potential, namely low κ . Since, the solvophobicity mediated term operates till $s \leq 2\Lambda_0$, we use mean field treatment for this term also. As $q \rightarrow 0$ (details in Supplemental Material), $C(q) \approx C_0 + C_2 q^2$. Here $C_0 = \sqrt{2\pi} (-\beta\alpha/R) \{c'\} - 4\beta X \sqrt{2\pi}/\kappa^2$, where c' is function of Λ_0 . $C_2 = (\rho/\kappa^4 - \alpha g)$, p depends on prefactor of DLVO potential (X), for a fixed temperature g depends on R and Λ_0 .

We find that $C(q)$ has a minimum at $q = 0$, if $C_2 > 0$ and a maximum at $q = 0$, if $C_2 < 0$. Hence, the condition for aggregation is $C_2 = 0$, which yields that $\alpha \sim \kappa^{-4.0}$. The mean field analysis reveals that the stability of the aggregated phase is in qualitative agreement to that obtained from our numerical simulation for low κ^* . This is not surprising for the mean field treatment is valid for longer-ranged potential. This qualitative agreement suggests that the stability of the structures is achieved in thermodynamic sense. However, in this analysis we overestimate stability of the aggregated phase which could be due to mean field nature of the analysis, ignoring fluctuations.

System spanning clusters, similar to our model, have been observed in SALR models as well.^[44,45] The contrasting feature of our model compared to colloidal SALR models lies in attractive part of interaction. In SALR models, the short-ranged attraction is due to depletion of solvent molecules, also known as the Asakura-Oosawa model depletion.^[10,11] To the contrary, the range of attraction in our model is given by radius of vapor around the particles, and is not necessarily short-ranged, depending on the thermodynamic condition of the solvent. For instance, the interfacial width of the gas-liquid

interface increases as the gas-liquid phase coexistence is approached, the width being divergent near gas-liquid critical point.^[27] Our model captures qualitatively the competition between attraction between hydrophobic parts and electrostatic repulsion in stabilizing self-assembled structures in macromolecules with exposed hydrophobic groups. We predict the existence of aggregation line, demarking the system spanning clusters with finite clusters. System spanning clusters will be insoluble in aqueous phase. Our prediction can be verified by extracting cluster size distributions on misfolded proteins, as in dynamic light scattering experiments.^[15]

However, it is important to note the caveats: Most common end product of misfolded protein aggregation is extended β -amyloid fibril^[61,62] having β sheets parallel to fibril axis. The model ignores shape anisotropy and inherent roughness of the protein surface. The charge at the center mimics the charged residues over the protein surface. The distribution of charged sites over the molecular surface has not been accounted for. Since the protein is represented by a rigid sphere, the internal motions including those of the side chains have not been included as well. The solvent effects are taken through the effective potentials, not explicitly. Moreover, due to spherical symmetry of the effective interactions, the clusters are also spherical. These limitations bound us to compare our results with realistic data only qualitatively. Quantitative comparison will require including the subtle details in the model. For instance, ignoring the distribution of charged residues amounts to neglecting multipole moments of the charge distribution, which may be important for anisotropic self-assembled structures.

The simple forms of interaction in the model are likely to make it robust, holding for a large class of systems with exposed hydrophobic sites. There is algebraic dependence between effective parameters for the hydrophobic and electrostatic parts, namely, α^* and κ_{th}^* respectively over the aggregation line. The exponent is independent of system details, determined by gross properties, like surface tension, compressibility, dielectric constant, the Debye screening length and overall charge on the macromolecule and should hold in general.

4 | CONCLUSIONS

To summarize, our studies describe a model to capture stability of nm-sized clusters over aggregation in macromolecules with solvent exposed hydrophobic moieties. The finite clusters are stable below a threshold value of the inverse Debye screening length and aggregates are stable above the threshold. The experimental observations on stability of amyloid clusters are qualitatively explained by this model. Our model may provide useful guideline to control self-assembled structures formed by macromolecules with exposed hydrophobic sites.

ACKNOWLEDGMENTS

S.D. thanks DST for INSPIRE Fellowship and JC thanks DST for funding. We acknowledge Suman Chakrabarti for critically reading the manuscript.

ORCID

Sutapa Dutta  <https://orcid.org/0000-0003-1517-6239>

REFERENCES

- [1] A. Dehsorkhi, V. Castelletto, I. W. Hamley, *J. Pept. Sci.* **2014**, *20*, 453.
- [2] Y. Chen, H. X. Gan, Y. W. Tong, *Macromolecules* **2015**, *48*, 2647.
- [3] H. G. Cui, M. J. Webber, S. I. Stupp, *Biopolymers* **2010**, *94*, 1.
- [4] I. Moreno-Gonzalez, C. Soto, *Semin. Cell Dev. Biol.* **2011**, *22*, 482.
- [5] M. Saiki, K. Shiba, M. Okumura, *FEBS Lett.* **2015**, *589*, 3541.
- [6] V. Vetri, V. Fodera, *FEBS Lett.* **2015**, *589*, 2448.
- [7] T. K. Karamanos, A. P. Kalverda, G. S. Thompson, S. E. Radford, *Prog. Nucl. Magn. Reson. Spectrosc.* **2015**, *88-89*, 86.
- [8] U. Sengupta, A. N. Nilson, R. Kayed, *EBioMedicine* **2016**, *6*, 42.
- [9] P. Cizas, R. Budvytyte, R. Morkuniene, R. Moldovan, M. Broccio, M. Losche, G. Niaura, G. Valincius, V. Borutaite, *Arch. Biochem. Biophys.* **2010**, *496*, 84.
- [10] M. M. van Schooneveld, V. W. de Villeneuve, R. P. Dullens, D. G. Aarts, M. E. Leunissen, W. K. Kegels, *J. Phys. Chem. B* **2009**, *113*, 4560.
- [11] Z. Varga, J. Swan, *Soft Matter* **2016**, *12*, 7670.
- [12] C. Wang, Z. Wang, X. Zhang, *Acc. Chem. Res.* **2012**, *45*, 608.
- [13] A. Stradner, H. Sedgwick, F. Cardinaux, W. C. Poon, S. U. Egelhaaf, P. Schurtenberger, *Nature* **2004**, *432*, 492.
- [14] J. Beerten, J. Schymkowitz, F. Rousseau, *Curr. Top. Med. Chem.* **2012**, *12*, 2470.
- [15] B. Sahoo, S. Nag, P. Sengupta, S. Maiti, *Biophys. J.* **2009**, *97*, 1454.
- [16] K. Garai, P. Sengupta, B. Sahoo, S. Maiti, *Biochem. Biophys. Res. Commun.* **2006**, *345*, 210.
- [17] I. T. Li, G. C. Walker, *Acc. Chem. Res.* **2012**, *45*, 2011.
- [18] J. B. Udgaonkar, *Arch. Biochem. Biophys.* **2013**, *531*, 24.
- [19] M. Adachi, M. Noji, M. So, K. Sasahara, J. Kardos, H. Naiki, Y. Goto, *J. Biol. Chem.* **2018**, *293*, 14775.
- [20] E. Bianchi, R. Blaak, C. N. Likos, *Phys. Chem. Chem. Phys.* **2011**, *13*, 6397.
- [21] E. Bianchi, B. Capone, I. Coluzza, L. Rovigatti, P. D. J. van Oostrum, *Phys. Chem. Chem. Phys.* **2017**, *19*, 19847.
- [22] G. R. Yi, D. J. Pine, S. Sacanna, *J. Phys. Condens. Matter* **2013**, *25*, 193101.
- [23] S. James, M. K. Quinn, J. J. McManus, *Phys. Chem. Chem. Phys.* **2015**, *17*, 5413.
- [24] R. Vacha, D. Frenkel, *Biophys. J.* **2011**, *101*, 1432.
- [25] A. Sorrenti, J. Leira-Iglesias, A. J. Markvoort, T. F. A. de Greef, T. M. Hermans, *Chem. Soc. Rev.* **2017**, *46*, 5476.
- [26] B. J. Berne, J. D. Weeks, R. H. Zhou, *Annu. Rev. Phys. Chem.* **2009**, *60*, 85.
- [27] E. E. Meyer, K. J. Rosenberg, J. Israelachvili, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 15739.
- [28] P. Attard, *Adv. Colloid Interface Sci.* **2003**, *104*, 75.
- [29] D. M. Huang, D. Chandler, *J. Phys. Chem. B* **2002**, *106*, 2047.
- [30] K. Lum, D. Chandler, J. D. Weeks, *J. Phys. Chem. B* **1999**, *103*, 4570.
- [31] J. Chakrabarti, S. Chakrabarti, H. Lowen, *J. Phys. Condens. Matter* **2006**, *18*, L81.
- [32] J. Chakrabarti, S. Dutta, *Chem. Phys. Lett.* **2015**, *620*, 109.
- [33] G. Reddy, A. Yethiraj, *Macromolecules* **2006**, *39*, 8536.
- [34] E. Brini, E. A. Algaer, P. Ganguly, C. L. Li, F. Rodriguez-Roperro, N. F. A. van der Vegt, *Soft Matter* **2013**, *9*, 2108.
- [35] K. S. Schmitz, *Langmuir* **2000**, *16*, 2115.
- [36] J. B. Caballero, A. M. Puertas, A. Fernandez-Barbero, F. J. de las Nieves, *J. Chem. Phys.* **2004**, *121*, 2428.
- [37] A. Engelbrecht, H. J. Schope, *Soft Matter* **2012**, *8*, 11034.
- [38] E. J. W. Verwey, J. T. G. Overbeek, *Theory of the Stability of Lyophobic Colloids*, Elsevier, New York-Amsterdam **1948**.

- [39] M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*, 2nd ed., Oxford Science Publications, New York **1989**.
- [40] J.-P. Hansen, I. R. McDonald, *Theory of Simple Liquids, Fourth Edition*, Academic Press, Elsevier Science Publishing Co Inc, United States **2013**.
- [41] M. Moller, P. Nilges, F. Harnisch, U. Schroder, *ChemSusChem* **2011**, 4, 566.
- [42] J. Pellicer, V. Garcia-Morales, L. Guanter, M. J. Hernandez, M. Dolz, *Am. J. Phys.* **2002**, 70, 705.
- [43] G. S. Kell, *J. Chem. Eng. Data* **1975**, 20, 97.
- [44] P. D. Godfrin, N. E. Valadez-Perez, R. Castaneda-Priego, N. J. Wagner, Y. Liu, *Soft Matter* **2014**, 10, 5061.
- [45] N. E. Valadez-Perez, R. Castaneda-Priego, Y. Liu, *RSC Adv.* **2013**, 3, 25110.
- [46] K. Barros, E. Luijten, *Phys. Rev. Lett.* **2014**, 113, 017801.
- [47] M. M. Gudarzi, *Langmuir* **2016**, 32, 5058.
- [48] J. Dzubielia, J. Chakrabarti, H. Lowen, *J. Chem. Phys.* **2009**, 131, 044513.
- [49] D. M. Huang, D. Chandler, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, 97, 8324.
- [50] N. Ishida, T. Inoue, M. Miyahara, K. Higashitani, *Langmuir* **2000**, 16, 6377.
- [51] F. Lugli, S. Hofinger, F. Zerbetto, *J. Am. Chem. Soc.* **2005**, 127, 8020.
- [52] P. R. Dormitzer, Z. Y. J. Sun, G. Wagner, S. C. Harrison, *EMBO J.* **2002**, 21, 885.
- [53] M. Yao, T. Ose, H. Sugimoto, A. Horiuchi, A. Nakagawa, S. Wakatsuki, D. Yokoi, T. Murakami, M. Honma, I. Tanaka, *J. Biol. Chem.* **2000**, 275, 34557.
- [54] S. Atwell, M. Ultsch, A. M. De Vos, J. A. Wells, *Science* **1997**, 278, 1125.
- [55] M. F. Symmons, G. H. Jones, B. F. Luisi, *Structure* **2000**, 8, 1215.
- [56] U. Pieper, G. Kapadia, M. Mevarech, O. Herzberg, *Structure* **1998**, 6, 75.
- [57] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, *Proteins* **2010**, 78, 1950.
- [58] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. Berendsen, *J. Comput. Chem.* **2005**, 26, 1701.
- [59] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, *J. Chem. Theory Comput.* **2008**, 4, 435.
- [60] J. L. Lebowitz, J. K. Percus, *Phys. Rev.* **1966**, 144, 251.
- [61] A. Morriss-Andrews, J. E. Shea, *Annu. Rev. Phys. Chem.* **2015**, 66, 643.
- [62] M. Landreh, M. R. Sawaya, M. S. Hipp, D. S. Eisenberg, K. Wuthrich, F. U. Hartl, *J. Intern. Med.* **2016**, 280, 164.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Dutta S, Patra P, Chakrabarti J. Self-assembly in amphiphilic macromolecules with solvent exposed hydrophobic moieties. *Biopolymers*. 2019;e23330. <https://doi.org/10.1002/bip.23330>