

Structural Dynamics and Computer Aided Drug Design Targeting the COT Kinase

M.Sc. Thesis

By
AHANA CHAKRABORTY



**DEPARTMENT OF BIOSCIENCES AND BIOMEDICAL
ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY
INDORE
MAY 2025**

Structural Dynamics and Computer Aided Drug Design Targeting the COT Kinase

A THESIS

*Submitted in partial fulfillment of the
requirements for the award of the degree
of*
Master of Science

by
AHANA CHAKRABORTY
2303171011



**DEPARTMENT OF BIOSCIENCES AND BIOMEDICAL
ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY
INDORE
MAY 2025**



INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **Structural Dynamics and Computer Aided Drug Design Targeting the COT Kinase** in the partial fulfillment of the requirements for the award of the degree of **MASTER OF SCIENCE** and submitted in the **DEPARTMENT OF BIOSCIENCES AND BIOMEDICAL ENGINEERING, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from July 2023 to May 2025 under the supervision of Dr. Parimal Kar, Associate Professor, Indian Institute of Technology Indore.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Ahana Chakraborty
06/05/2025
Signature of the student with date
(Ahana Chakraborty)

This is to certify that the above statement made by the candidate is correct to the best of my/our knowledge.

Parimal Kar
Signature of the Supervisor of
M.Sc. thesis (with date)
(Dr. Parimal kar)

Ahana Chakraborty has successfully given her M.Sc. Oral Examination held on 06/05/2025.

Parimal Kar
Signature(s) of Supervisor(s) of MSc thesis
Date: 06/05/2025

P.V. Kodgire
Convener, DPGC
Date: 06/05/2025

ACKNOWLEDGEMENTS

The initial journey through this research field was not at all easy for me and it took much time to understand the facts and get deeply absorbed into the project. I am well destined to have many supporting people who guided me from the beginning and cannot help but thank them from the core of my heart. First, I would like to extend my deepest gratitude to my project supervisor Dr. Parimal Kar for his continuous support and guidance throughout the journey of my M.Sc Project. It was his courses that properly introduced me to the field of computational biophysics. From then till now, his continuous guidance and helpful advice have helped me immensely. Sincere gratitude to Head of the Department, BSBE (Dr.Parimal Kar), DPGC convener (Prof. Prashant Kodgire), M.Sc. Project Coordinator (Dr. Parimal Kar) and all the faculty members for providing various support throughout the duration of my project work. I would also like to thank Department of Biotechnology, Ministry of Science and Technology, Government of India for providing studentship during my M.Sc. course.

My warmest acknowledgments and respect to Mr. Sunanda Samanta, my mentor and a very supportive plus caring person. He was the one who has taught me every minute things of Computational biology and stood by my side whenever required. I learnt making presentations and appropriately delivering it from him and he is such a person who has a pure soul and is down to earth. I also extend my gratitude to Mr. Kapil D. Ursal, Ms. Subhasmita Mahapatra, Mr. Kesarpu Uday Kumar were always there in times of struggles and extended their helping hands. Special thanks to my friend and lab mate Ms. Dibyanka Dalai, who shared much experiences with me and helped me unconditionally. I also thank my lab mate Mr. Pradeep Kumar for technical helps. It has been a wonderful experience to be a part of the computational biophysics group.

DEDICATED
TO MY
FAMILY &
MY CAT

Abstract

COT kinase is a target for drug discovery against cancer and autoimmune disorders. They are located downstream to the IKK2, unaffected by the NF – kappa B pathway. There is no such FDA approved drug against this protein and no extensive in-silico studies are being done till date. The structure is very unique as the P-loop is very flexible and a frameshift mutation exists in the α C helix. This study revolves around the structural dynamics study of COT Kinase in two forms- the apo and complex, both in unphosphorylated form and phosphorylated form. The phosphorylation is done on Threonine 290 residue. GaMD or Gaussian Accelerated Molecular Dynamics Simulation is done in triplicates for both the system. Further, the structural stability was studied for Phosphate binding loop(P-loop), Activation loop(A-loop), Catalytic loop(C-loop) and α C helix. Root Mean Square Fluctuation values and Root Mean Square Deviation values provided strong insights to the extensible flexibility of P-loop exclusively for phosphorylated and complex systems respectively. Virtual Screening is done to find out a potent inhibitor than the control molecule followed by ADMET analysis and simulation studies to check the protein stability when inhibitor binds to it. Also, the effect on structure when Mg is bound to it and Mn is bound to it, has been elucidated. It has been found that Manganese stabilizes the flexible P-loop.

TABLE OF CONTENTS

1. INTRODUCTION	1
1.1 Kinases	1
1.2 Role of kinase in humans	2
1.2.1 Kinases in Cell division and Cell death	2
1.2.2 Kinases in metabolism and hormonal signaling	3
1.3 COT Kinase and its structure	3
1.4 Signaling pathway of COT Kinase	6
1.5 Diseases associated with COT Kinase	7
1.6 Inhibitors of COT Kinase	7
1.7 Motivation for the study of this molecule	8
2. THEORETICAL ASPECTS	9
2.1 Molecular Dynamic Simulation	9
2.1 Basics	10
2.2 Force Fields	11
2.3 Periodic Boundary Conditions	12
2.4 Long-Range Interactions	13
2.5 Thermostat And Barostat	13
2.6 Solvation Models	14
2.7 Gaussian Accelerated Molecular Dynamics	15
3. MATERIALS AND METHODS	17
3.1 PDB Selection	17
3.2 System Preparation	17
3.3 Simulation Protocol	19
3.3.1 Minimization	19
3.3.2 Heating	19
3.3.3 Equilibration	19

3.3.4 Production run	20
3.4 Post-simulation trajectory analyses	20
3.4.1 Root-Mean-Square Deviation (RMSD).....	20
3.4.2 Root Mean Square Fluctuation (RMSF).....	20
3.4.3 Solvent-Accessible Surface Area (SASA).....	21
3.4.4 Radius of Gyration (RoG)	21
3.5 Essential dynamic studies.....	21
3.5.1 Principal Component Analysis (PCA).....	21
3.5.2 Dynamic Cross Correlation Matrix (DCCM).....	21
3.5.3 Dihedral Principal Component Analysis (dPCA).....	22
3.5.4 Protein Structure Network Analysis (PSN)	22
3.6 Free energy calculation by MMPB(GB)SA	22
3.7 Entropy calculation	23
3.8 K-means clustering.....	23
3.9 Computer-Aided Drug Design	24
3.10 Virtual Screening.....	24
3.11 ADMET Analysis	25
3.12 Rules for Drug Likeness.....	25
4. OBJECTIVES	27
5. RESULTS AND DISCUSSION	29
5.1: Study of the Structural Dynamics of apo and complex systems in both unphosphorylated and phosphorylated forms.	29
5.2: Identification of Novel Inhibitor(s) targeting the COT Kinase by Virtual Screening.....	52
5.3: Differential study of the effect of Mn^{2+} vs Mg^{2+} on the Structural Dynamics of COT Kinase.	54
CONCLUSION & FUTURE PROSPECTS	61
REFERENCES	63

LIST OF FIGURES

Figure no.	Figure description	Page no.
1.1	Protein phosphorylation catalyzed by a kinase enzyme	1
1.2	Structure of the catalytic domain of COT kinase, highlighting distinct functional regions	5
1.3	Schematic Representation of the COT Kinase signalling pathway	7
2.1	Periodic boundary conditions	13
5.1	The Figures portray the time evolution of backbone RMSD for all the three runs in all the systems (apo and complex)	29
5.2	The Figures portray the time evolution plot of P-loop RMSD for all the three runs in all the systems (apo and complex)	30
5.3	The Figures portray the time evolution plot of α C-helix RMSD for all the three runs in all the systems (apo and complex)	31
5.4	The Figures portray the time evolution plot of A-loop RMSD for all the three runs in all the systems (apo and complex)	32
5.5	The Figures portray the time evolution plot of C-loop RMSD for all the three runs in all the systems (apo and complex)	32
5.6	The Figure reveals residual fluctuations for all the four systems	33
5.7	Probability distribution curve of backbone RMSD for apo and complex systems respectively	34
5.8	Probability distribution curve of A-loop RMSD for apo and complex systems respectively	34
5.9	Probability distribution curve of C-loop RMSD for apo and complex systems respectively	35
5.10	Probability distribution curve of P-loop RMSD for apo and complex systems respectively	36

5.11	Probability distribution curve of α C-helix RMSD for apo and complex systems respectively	36
5.12	The Figure represents the Probability distribution curve of Radius of Gyration for all the systems (apo and complex)	37
5.13	Probability distribution curve of Solvent Accessible Surface Area for the systems	37
5.14	Probability distribution curve of ligand RMSD and Binding pocket RMSD	38
5.15	Probability distribution curve of DFG motif RMSD and HHD motif RMSD	
5.16	The Figure shows salt bridge formation between K104 and D115 in all the systems	40
5.17	Time series plot for Number of Hydrogen bonds formed by the ligand with the protein in unphosphorylated and phosphorylated forms respectively	42
5.18	Free Energy Landscape and extracted structures from the minima	43
5.19	DCCM of the four systems	44
5.20	dPCA of the A-loop to check their structural rearrangements	45
5.21	K-means clustering	46
5.22	Graphical representation of different energy values calculated by MMPB(GB)SA method for all the four systems	46
5.23	Per-residue binding energy of complexes	47
5.24	Network analysis of A- apo unphosphorylated, B- apo phosphorylated, C- complex unphosphorylated and D- complex phosphorylated	49
5.25	Figure 5.25: Per-residue analysis of P-loop of A- apo unphosphorylated, B- apo phosphorylated, C- complex unphosphorylated and D- complex phosphorylated. Energy bar is provided along with.	50

5.26	Per-residue analysis of A-loop of A- apo unphosphorylated, B- apo phosphorylated, C- complex unphosphorylated and D- complex phosphorylated. Energy bar is provided along with.	51
5.27	Time evolution graphs showing RMSD values of various regions of the molecule when control is bound to it vs inhibitor bound to it	53
5.28	Probability distribution curve of backbone RMSD for both the complexes	54
5.29	Residual fluctuation for the systems	55
5.30	Time evolution plot and KDE plot of P-loop RMSD of both the systems	56
5.31	Time evolution plot and KDE plot of A-loop RMSD of both the systems	57
5.32	Time series plot for Number of Hydrogen bonds formed by the complexes	57
5.33	Per-residue RMSD of P-loop for Mg-ATP system and Mn-ATP system	59
5.34	Per-residue RMSD of A-loop for Mg-ATP system and Mn-ATP system	60

LIST OF TABLES

Table 1: Various key structural segments of the COT kinase are listed, along with their corresponding residue numbers.	5
Table 2: The prepared systems along with their run time and number of independent replica simulation runs performed for each.	18
Table 3: The backbone RMSD, RoG and SASA are calculated for the systems by Block Average method.	38
Table 4: Hydrogen bond analysis for the complex systems.	41
Table 5: The Protein Structure Network analysis for various systems.	48
Table 6: Virtual screening and ADMET analysis of lead molecules...	52
Table 7: The values of backbone RMSD, SASA and Radius of Gyration calculated for the complexes using Block Average method.....	54
Table 8: Hydrogen bond analysis for two complexes.....	58

ACRONYMS

COT - Cancer Osaka Thyroid

FF- Force Field

AMBER- Assisted Model Building with Energy Refinement

SASA- Solvent Accessible Surface Area

RoG- Radius of Gyration

RMSD- Root Mean Square Deviation

RMSF- Root Mean Square Fluctuation

PCA- Principal Component Analysis

dPCA- dihedral Principal Component Analysis

DCCM- Dynamic Cross-Correlation Matrix

PSN- Protein Structure Network

GaMD- Gaussian accelerated Molecular Dynamics

PME- Particle Mesh Ewald

PBC- Periodic Boundary Conditions

VS- Virtual Screening

CADD- Computer Aided Drug Design

ADMET- Absorption, Distribution, Metabolism, Excretion and Toxicity.

P-loop - Phosphate binding loop

C-loop - Catalytic loop

A-loop - Activation loop

PSA - Polar Surface Area

NMA - Normal Mode Analysis

PDB- Protein Data Bank

WCSS- With Cluster Sum of Squares

OPLS- Optimized Potential for Liquid Simulations

Tpl2- Tumour Progression Locus 2

BBB- Blood Brain Barrier

NOMENCLATURE

Å	Angstrom
ns	nanosecond
μs	microsecond
C _α	Alpha Carbon atom
NVT	Number of particles, Volume and Temperature
NPT	Number of particles, Pressure and Temperature
NVE	Number of particles, Volume and Energy

CHAPTER 1

INTRODUCTION

1.1 Kinases:

Kinases are widely distributed protein molecules across prokaryotes and eukaryotes. This thesis deals with a human kinase, so the discussion is restricted only to human kinases. In humans, there are about 280 kinase proteins that have been reported and whose structures have been elucidated in detail. About 500 kinase genes are present, which make up about 2% of the human genome. It is a protein molecule, specifically an enzyme that helps in phosphorylation of a particular amino acid residue of another protein¹. It uses ATP as a cofactor and snatches the gamma phosphate to add it to another protein, so that the other protein gets phosphorylated. This leads to various intercellular processes and metabolic functions.

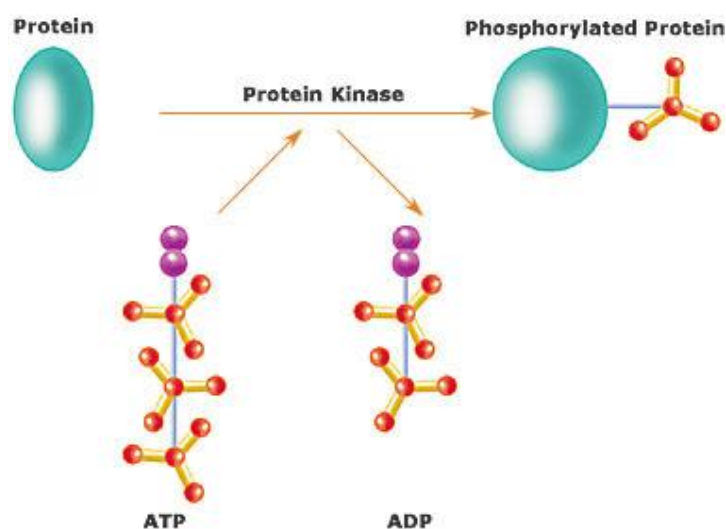


Figure 1.1: Protein phosphorylation catalyzed by a kinase enzyme.

Kinases may be protein kinases or lipid kinases, depending upon the type of biomolecule they phosphorylate. They are known as “molecular switches” as the phosphorylation induces various structural changes leading to different functional activities. In kinases, phosphorylation functions as a key regulatory modification

that activates the enzyme by inducing conformational changes essential for its catalytic activity. Kinases can be broadly grouped into serine/threonine kinases or tyrosine kinases, depending on the target substrate that they phosphorylate. Tyr kinases phosphorylate on Tyrosine residue(s) present on the protein, and Ser-Thr kinase adds a phosphate group to either Serine amino acid or Threonine amino acid or both. Well-characterized examples include tyrosine kinases like JAK and SYK, and serine/threonine kinases such as MAPK, PI3K, and Akt, each playing critical roles in signal transduction pathways. In addition to serine/threonine and tyrosine kinases, other types such as histidine kinases, dual-specificity kinases, and aspartate (Asp) kinases have also been reported, highlighting the diversity of phosphorylation-based signaling mechanisms².

1.2 Role of kinase in humans:

Kinases are central to various physiological processes in humans, including cell signaling, metabolism, and immune function. Their activities are often interdependent, with many kinases influencing one another and operating collectively within complex signaling pathways. They also play a major role in disease progression and Apoptosis³.

1.2.1 Kinases in Cell division and Cell death:

In cell growth and division, CDKs play a wide role. Their phosphorylation leads to activation of downstream molecules, thereby activating the cell cycle. CDK4, CDK6, and CDK1 are involved in cell cycle and division. The phosphorylation of CDK4/6 activates it and thereby aids in phosphorylation of Rb, which in turn detaches from E2F, a transcription factor. E2F then induces the transcription of cyclin E so that the cell can progress to S-phase. ATM and ATR kinases come into play during DNA damage. They activate the Chk1/2 molecule, which phosphorylates the p53 protein that, in turn, activates p21 so as to arrest the cell cycle and undergo Programmed Cell Death⁴.

1.2.2 Kinases in metabolism and hormonal signaling:

Glycogen metabolism through Epinephrine signaling is well controlled by kinase molecules. Epinephrine, an amine hormone, is released by the Adrenal medulla during stressful conditions. The body requires energy for fight or flight action; thus, glucose is needed for the same. Epinephrine or Nor-epinephrine binds to the GPCR, and the $G\alpha$ is activated as it is bound to GTP. The $G\alpha$ molecule binds to Adenylate Cyclase and activates it. The Adenylate Cyclase converts ATP to cAMP, which in turn activates the PKA (Protein Kinase A). This PKA gets phosphorylated and activates the Phosphorylase kinase molecule that, in turn, activates glycogen phosphorylase, which converts glycogen to glucose. This glucose is broken down by glycolysis to release energy⁵.

1.3 COT Kinase and its structure:

The COT kinase, or the Cancer Osaka Thyroid Kinase, is encoded by the cot proto-oncogene located on the 10th chromosome. Its molecular weight is ~53 kDa and belongs to the MAP3K8 family. It is also known as Tpl2 (Tumor Progression Locus 2). It takes part in cell signaling and activates the ERK pathway, which stimulates cell growth and also plays a key role in inflammatory reactions by stimulating the macrophages to produce cytokines like TNF- α and IL-1 β . It is a serine/threonine protein kinase that uses Manganese or Magnesium as a cofactor. Depending on the ion used, the affinity towards ATP differs for COT kinase. Studies suggest that Mn^{2+} lowers the K_m value, thereby increasing the affinity of the enzyme towards ATP⁶.

The structure of COT kinase is very unique, and it bears a few differences from other kinases. The broad structure of the kinase domain reveals a well-defined architecture consisting of an N-terminal lobe and a C-terminal lobe, along with key conserved elements such as the αC -helix, catalytic loop, activation loop, and the phosphate-binding loop (P-loop)/glycine-rich loop. The P-loop

is very flexible as compared to other kinases. It has a 15 amino acid insert before its GxGxxG motif, and thus it is about 20 amino acids long. The ATP binding site is located near the glycine-rich loop. Normally, the Lysine residue to $\beta 3$ strand forms a salt bridge with the Glutamate of αC -helix, and in turn, the N-atom of the lysine residue forms Hydrogen bonding with the O-atoms of α and β phosphates of ATP. COT kinase bears an exception that instead of the Glu of the αC helix, Asp residue forms the salt bridge, and this anomaly is due to the frameshift mutation present in the αC helix. The protein contains two conserved water molecules that form an H-bond with the DFG motif. DFG is essential for the active or inactive kinase conformation. While the HRD motif is typically a conserved component of the catalytic loop in kinases, this particular molecule features an HHD motif instead. Substrate binds to the Activation loop, and the Threonine 290 residue is the site for phosphorylation. Interestingly, as compared to other kinases, this molecule has a conserved SPE motif instead of an APE motif in the activation loop. The Methionine 207 residue is the gatekeeper residue of this kinase, and the P-loop has a conserved Proline residue instead of glycine. Gatekeeper residue is present in the ATP binding region that aids in accessing the hydrophobic domain. It is an important residue for inhibitor binding, and mutation in this may lead to drug resistance as it would hinder inhibitor entry to the ATP binding pocket⁷.

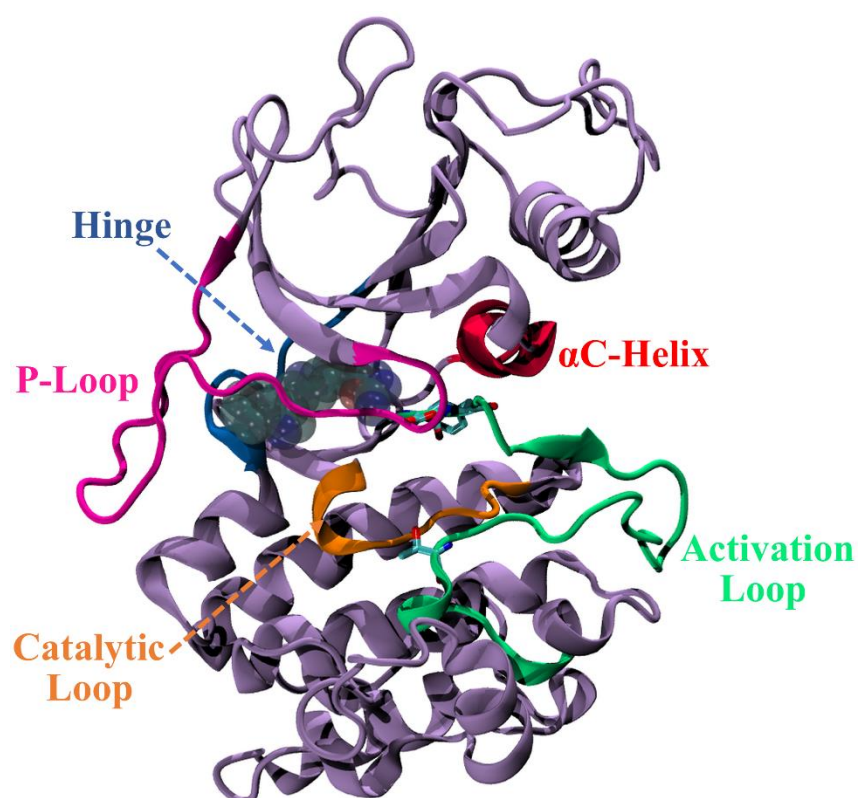


Figure 1.2: Structure of the catalytic domain of COT kinase, highlighting distinct functional regions. The visualization was created using Visual Molecular Dynamics (VMD) with structural data from PDB ID: 4Y85 (Resolution: 2.33 Å).

Table 1: Various key structural segments of the COT kinase are listed, along with their corresponding residue numbers.

Name of the region	Residue numbers
P-loop (Phosphate binding loop / Glycine-rich loop)	131-150
Activation loop (A-loop)	270-297
Catalytic loop (C-loop)	251-258
α C-helix	171-185
DFG motif	270-272
HHD motif	251-253
Gatekeeper residue	Met 207

1.4 Signaling pathway of COT Kinase:

The COTK/Tpl2 signalling begins when a ligand binds to the receptor and induces conformational changes to activate downstream molecules like IKK2. When Epidermal growth factor binds to EGFR or Interleukins bind to Interleukin receptor, the IKK2 molecule is activated, which in turn activates COTK that in turn activates MEK (MAP Kinase Kinase) by phosphorylating it on Ser/Thr residue. The MEK then phosphorylates ERK (MAP Kinase), activating gene transcription to induce cell growth and development. The IKK2 molecule mediates the NF-kappa B pathway, and since COT Kinase or Tpl2 is downstream of it, the NF-kappa B pathway is not disturbed. Tpl2 forms a complex with ABIN2 and p105. The p105 degrades to p50, and it then induces gene transcription to aid in inflammation and cell growth by inducing the JNK pathway. The signalling results in the production of IL-1 β and TNF, which promote inflammatory reactions and also induce cell growth and proliferation due to EGFR signalling. There are two key regions of interaction between Tpl-2 and p105. The kinase domain of Cot/Tpl-2 binds to the NF-kappa B-1/p105 DD (Death Domain), and the C terminus of Cot/Tpl-2 binds to the N-terminal region residues of the p105 ankyrin repeats. The interaction prevents Cot/Tpl-2 from accessing its substrate MEK, greatly inhibiting the phosphorylation/activation of MEK. The inhibitory effect of the p105 DD is amplified by the concurrent interaction of the C terminus of Cot/Tpl-2 with p105^{6,8}.

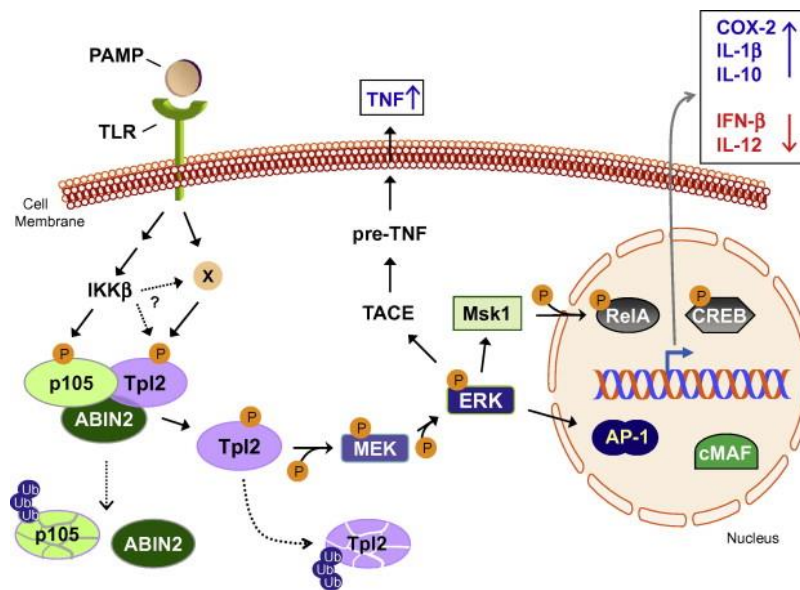


Figure 1.3: Schematic Representation of the COT Kinase Signalling Pathway.

1.5 Diseases associated with COT Kinase:

Overexpression of the protein enhances the signalling cascade and does not allow it to stop when cell signalling is not required. The excessive kinase stimulates the synthesis of huge amounts of growth factors, pro-inflammatory cytokines, and other molecules that lead to uncontrolled cell division and proliferation, thereby creating a favourable environment for cancer development. The cytokines aid in tumour progression, EMT⁹, and mediate inflammatory diseases like Diabetes Mellitus and Rheumatoid Arthritis. Papillary thyroid kinase, breast cancer, and ovarian cancers have been reported¹⁰.

1.6 Inhibitors of COT Kinase:

Targeting the catalytic kinase domain with small-molecule inhibitors has proven to be an effective therapeutic strategy against numerous disease conditions. These kinase-domain inhibitors¹¹ are generally classified into Type I, Type II, and Type III. Type I and Type II inhibitors differ in their selectivity based on the orientation of the conserved DFG (Asp-Phe-Gly) motif within the kinase domain. In the active or DFG- “in” conformation, the aspartate

residue of the DFG motif points inward toward the ATP-binding site, a positioning essential for catalytic activity. Type I inhibitors bind to this active conformation. In contrast, Type II inhibitors preferentially bind to the DFG-“*out*” conformation, where the DFG-Asp flips outward and the adjacent phenylalanine residue flips inward, a rearrangement known as the DFG-flip. This conformation represents an inactive state of the kinase, to which Type II inhibitors bind. MAP3K8 mutation is responsible for melanomas, as per studies, and it can be somehow controlled by Trametinib¹². It is not inhibited by Staurosporine. Literature tells that Imidazonaphthyridine could be a good drug candidate for the COT Kinase. As per Therapeutic Target Database, two proposed drugs are available for the same, namely: Tilpisertib and fosmecarbilo for Inflammatory Bowel Disease, and Tpl2 kinase inhibitor, which is under clinical trial¹³.

1.7 Motivation for the study of this molecule:

The main reason for studying COT Kinase lies in the fact that to date, no prominent studies have been done on structural dynamics, and no potent inhibitor has been found yet. Also, the unique structure intrigues the study to dig out any unexplained truth. An inhibitor would target this molecule and treat several disorders without even affecting the NF- κ B pathway.

CHAPTER 2

THEORETICAL ASPECTS

“Atoms” are the tiny particles present in a substance that cannot exist independently. They cannot be divided further and are capable of forming molecules and compounds. Atomic motion occurs on extremely fast timescales, typically in the order of femtoseconds. Macromolecules such as proteins and carbohydrates, which are composed of numerous atoms, exhibit dynamic behavior that is far too rapid to be detected by conventional observation or many experimental techniques. Metabolic processes in the body, governed by macromolecules, often involve structural changes arising from atomic-level fluctuations. These changes can be explored through computational simulations, commonly referred to as *in silico* studies¹⁴. Techniques such as molecular dynamics (MD) simulations allow us to observe atomistic motions by generating snapshots of structural changes at femtosecond-scale time-steps. MD simulations, along with advanced variants such as NaMD, aMD, GaMD, and Lig-GaMD, are among the widely used techniques for studying biomolecular dynamics.

2.1 MOLECULAR DYNAMIC SIMULATION:

MD simulation is a computational tool that helps in analyzing the movement of atoms and molecules over a period of time. It uses classical and molecular mechanics to study the same. Generally, the time step kept is 1-2 fs. The movement of atoms is too fast, and hence we require a small timestep to capture it. The time step of 1-2 fs allows a proper snapshot of atomic movements. It is generally chosen keeping in mind the accuracy and computational cost. There are various algorithms to calculate the velocity and position of the atoms during a course of simulation. Commonly used integration algorithms are the Verlet algorithm, the Velocity-Verlet algorithm, and the Leapfrog algorithm. They extensively use integration methods for the calculation¹⁵.

2.1 Basics:

The MD Simulation follows Newton's second law of motion to calculate the potential energy and position of atoms in a system. According to the second law of Newton, the force on a body is equal to the mass times acceleration¹⁶.

$$F = ma \quad [2.1]$$

$$a = \frac{d^2r}{dt^2} \quad [2.2]$$

$$F = m \frac{d^2r}{dt^2} \quad [2.3]$$

$$F = -\frac{dV}{dr} \quad [2.4]$$

2.1.1 Verlet Algorithm:

It is used to calculate the position of atoms. If we integrate $v(t + \Delta t)$ based on the Taylor series, the result is¹⁷:

$$r(t + \Delta t) = r(t) + \Delta t v(t) + \frac{1}{2} \Delta t^2 a(t) + \frac{1}{3!} \Delta t^3 b(t) + O(\Delta t^4) \quad [2.5]$$

$$r(t - \Delta t) = r(t) - \Delta t v(t) + \frac{1}{2} \Delta t^2 a(t) - \frac{1}{3!} \Delta t^3 b(t) + O(\Delta t^4) \quad [2.6]$$

Here, r is the position, t is the current time, and Δt is the change in time. $v(t)$ represents an acceleration of the atoms that is obtained by the first-order differentiation of position (v). $a(t)$ is the second-order differentiation value of v , and $b(t)$ is the third-order, and $O(\Delta t^4)$ is the local error.

If $r(t + \Delta t)$ and $r(t - \Delta t)$ are added,

$$r(t + \Delta t) + r(t - \Delta t) = 2r(t) + a(t)\Delta t^2 + O(\Delta t^4) \quad [2.7]$$

Therefore, velocity is:

$$v(t) = \frac{1}{2\Delta t} [r(t + \Delta t) - r(t - \Delta t)] \quad [2.8]$$

2.1.2 Velocity-Verlet Algorithm:

The algorithm helps in calculating the position, velocity and acceleration of the atoms at the same time¹⁷.

$$r(t + \Delta t) = r(t) + \Delta t v(t) + \frac{1}{2} \Delta t^2 a(t) \quad [2.9]$$

$$v(t + \Delta t) = v(t) + \frac{1}{2} \Delta t [a(t) + a(t + \Delta t)] \quad [2.10]$$

2.1.3 Leapfrog Algorithm:

It is a modified form of the Verlet algorithm where the velocities are calculated for the time step $t + \frac{1}{2} \Delta t$ and then positions are calculated at the $t + \Delta t$.

$$r(t + \Delta t) = r(t) + v(t + \frac{1}{2} \Delta t) \Delta t \quad [2.11]$$

$$v(t + \frac{1}{2} \Delta t) = v(t - \frac{1}{2} \Delta t) + a(t) \Delta t \quad [2.12]$$

2.2 FORCE FIELDS:

A force field is a functional form for calculating a system's potential energy for a set of atoms as a function of their specific coordinates. Both the bonded terms and non-bonded terms are used in calculating the potential energy. The bonded terms represent covalent interactions between atoms, whereas the non-covalent interactions are represented by non-bonded terms¹⁸.

$$V_{total} = V_{bonded} + V_{non-bonded} \quad [2.13]$$

$$V_{bonded} = V_{bond-length} + V_{bond-angle} + V_{dihedrals} + V_{improper} \quad [2.14]$$

$$V_{non-bonded} = V_{vdW} + V_{elec} \quad [2.15]$$

Here, V_{vdW} represent the van der Waals interactions (LJ potential) and V_{elec} tells the electrostatic interactions that arise due to oppositely charged ions.

$$V_{bond-length} = \sum \frac{1}{2} K (r - r_0)^2 \quad [2.16]$$

$$V_{bond-angle} = \sum \frac{1}{2} K (\theta - \theta_0)^2 \quad [2.17]$$

$$V_{dihedrals} = \sum K [1 + \cos(n\phi - \delta)] \quad [2.18]$$

Commonly used force fields are: **AMBER**¹⁹, **CHARMM**, **GROMACS**, and **OPLS**.

Commonly used force fields in molecular simulations include AMBER¹⁹, CHARMM, GROMACS, and OPLS. Selecting an appropriate force field is a critical step that significantly influences the simulation results. For proteins, widely used AMBER force fields include ff19SB²⁰, ff14SB, and ff99SB. Carbohydrate simulations typically employ GLYCAM_06j, while ligands are often parameterized using GAFF2. For nucleic acids, OL21 or OL15 is used for DNA, and OL3 is commonly applied for RNA. In this study, AMBER18¹⁶ is used and ff19SB for protein, GAFF2 for ligand.

2.3 PERIODIC BOUNDARY CONDITIONS:

It is an MD simulation condition that mimics an infinite simulation box within a finite system. PBC allow us to estimate bulk properties more accurately. In a real physiological condition, solvent molecules are also involved along with solutes. in the systems. High computational cost and negligence of the surface effect of solvent molecules, the number of atoms in the simulation system is limited. PBC²¹ is used to face these issues, which creates an illusion of a bulk environment in a simulation. The solute and solvent molecules are enclosed in a box replicated in all three Cartesian dimensions to give a periodic array completely filling the space.

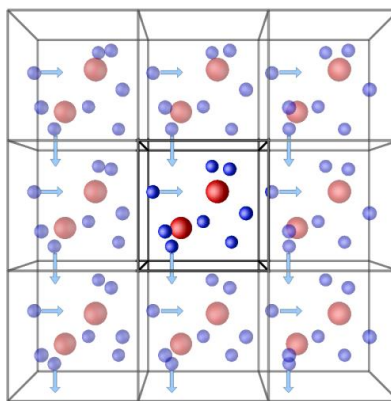


Figure 2.1: Periodic boundary conditions

2.4 LONG-RANGE INTERACTIONS:

Electrostatic interactions, van der Waals interactions, Coulomb forces, and Gravitational forces fall under long-range interactions. They imply that the atoms are at a significant distance from each other, that is beyond the cut-off distance. The main idea here is to neglect certain interaction pairs to increase the simulation efficiency and maintain the authenticity of the produced results. Both Lennard-Jones (LJ) and Coulombic interactions can be ignored beyond a certain distance, as their effect is negligible compared to the error of the simulations. For that reason, a cut-off scheme is used to ignore all long-range interactions beyond a certain distance. The intermolecular electrostatic interaction persists over a longer range than the LJ interaction, which hinders the above method's usage²¹. For that, the Particle Mesh Ewald²² (PME) scheme is most widely used for calculating the long-range electrostatics in the periodic system during MD simulation. Real time space and Reciprocal space formulas are used to calculate the long-range interactions. The $O(N^2)$ formula is obsolete and has been replaced by the Fourier Transformation method, $O(N \log N)$.

2.5 THERMOSTAT AND BAROSTAT:

To maintain the appropriate conditions of temperature and pressure throughout the course of the simulation, Thermostats and Barostats are used. Fixed temperature and pressure are required to maintain the system's kinetic energy and prevent it from blowing up. One of the

widely used Thermostat is the Langevin thermostat²³, while the Bendersen Barostat is commonly employed for pressure regulation in MD simulations. In this thesis, the Langevin thermostat has been used, which uses the formula:

$$F = F_{interaction} + F_{friction} + F_{random} \quad [2.19]$$

Here, $F_{interaction}$ is the standard interactions calculated during the simulation, $F_{friction}$ is used to tune the “viscosity” of the implicit bath and F_{random} gives how much the solvent collides randomly with each other.

The popular Berendsen barostat²⁴ couples the pressure with a pressure bath, which is shown below;

$$\lambda = 1 - k \frac{\delta t}{\tau P} (P - P_{bath}) \quad [2.20]$$

$$r_i = \lambda^{1/3} r_i \quad [2.21]$$

Here, r_i is the rescaled coordinates, τP is the time constant, and k is the isothermal compressibility that mainly governs the coupling strength to the pressure bath.

2.6 SOLVATION MODELS:

It is used to solvate the biomolecules in the system. There are two types of water models: implicit or continuum and explicit. In the implicit water models, the effects of the solvent molecules are given to the system as a whole. While in the case of the explicit water models, the solvent is attached to each residue of the biomolecule so that all the contributions are considered. An explicit solvation system is more computationally expensive and more accurate than implicit water models, which are less computationally expensive. The most used water models are the 3-site model, where each site is for the nuclei of the water molecule. We have used the OPC²⁵ water model to study the protein structural dynamics²⁶, as it is much more accurate than TIP3P²⁷ and is compatible with a protein force field. 4-site models include TIP4P²⁷ and 5-site include TIP5P. They are expensive computationally and are not used routinely.

2.7 GAUSSIAN ACCELERATED MOLECULAR DYNAMICS:

Gaussian Accelerated Molecular Dynamics (GaMD) refers to the technique that adds a harmonic boost potential to the initial system potential in order to lower or smooth the potential energy surface. Both the dihedral potential energy boost and the total potential energy boost can be implemented on the systems. It is a powerful augmented sampling technique that aids in lowering energy barriers, which speeds up conformational space exploration and prevents the structure from getting stuck at a point, making it able to reach its local minima easily. Unlike the Accelerated Molecular Dynamics (aMD) method, GaMD eradicates the statistical noise in large biomolecular systems during reweighting procedures. GaMD is much faster and more accurate than Conventional MD. It can sample much more of the conformational space than the conventional one. In this study, GaMD was done, and dual boost potential was given to both the apo and complex systems²⁸.

If a system's potential (V) is less than that of the energy threshold (E). In these situations, the modified potential (V^*) is obtained by adding a harmonic boost potential (ΔV):

$$\Delta V = \frac{1}{2}k(E - V)^2, \text{ if } V < E \quad [2.22]$$

$$V^* = V + \frac{1}{2}k(E - V)^2, \text{ if } V > E \quad [2.23]$$

The boost potential (ΔV) is brought to zero if the system potential (V) exceeds the threshold energy (E):

$$\Delta V = 0, \text{ if } V \geq E$$

$$V^* = V, \text{ if } V \leq E$$

CHAPTER 3

MATERIALS AND METHODS

3.1 PDB Selection

The PDB structure for COT Kinase was downloaded from the RCSB PDB². For the study, PDB ID: 4Y85 was chosen, which had a resolution of 2.33 Å. This structure is an X-ray crystallography-derived one and has a ligand with it, named as 'Imidazonaphthyridine.' The PDB has 332 amino acids, starting from G-Q, and has some missing residues too. The corresponding residues for different regions of the kinase domain of COT Kinase⁶ are represented in Figure 1.1 and listed in **Table 1**.

3.2 System Preparation

The protein structure was visualized in the UCSF Chimera²⁹ software. The non-standard residues were removed, and the crystallographic water molecules were also deleted except those within less than 5 Å distance, as this kinase has two conserved molecules of water within it. The missing residues were modelled using the Modeller plugin of UCSF Chimera³⁰. Six systems were prepared: apo unphosphorylated, apo phosphorylated, complex phosphorylated, complex unphosphorylated, Mn-ATP-protein complex, and Mg-ATP-Protein complex. The apo system has only protein, and the complex has both protein and ligand in it. Phosphorylation is manually performed at Thr290 residue for both the apo and complex. GaMD simulations and the AMBER 18 Leap module were used for the investigation. We have run a simulation of 2000 ns (2 μs) in triplicate runs for both the Apo and Apo-Phosphorylated systems. The same GaMD²⁸ simulations are run in triplicate for 2000 ns for the complex systems: COT Kinase bound to ligand Imidazonaphthyridine⁷. The complex is studied to decipher a sharp contrast between the dynamics of a complex unphosphorylated system and a complex system phosphorylated at T290. Further, a comparative study is done for both

the apo and complex systems. Imidazonaphthyridine is taken as the control molecule in this study.

The conventional Molecular Dynamics simulations were performed using AMBER¹⁶ 18's pmemd.cuda module. The Amber Leap Module was used, and the ff19SB force field was used for protein. The water model used is OPC, and the force field phosaa10 was used for the phosphorylated system. The GAFF2 force field is used for the parameterization of the ligand. Each system was solvated in an octahedral box, with a 10 Å buffer distance from the protein's surface in all directions, using the OPC water model. To neutralize the system's charge, 50 Na⁺ ions and 45 Cl⁻ ions are added for the apo systems. The complex systems were solvated, and 34 Na⁺ ions and 31 Cl⁻ ions were added to neutralize them. The SHAKE algorithm was used to satisfy bond geometrical constraints during a simulation, freeze H atoms, and cause vibrational motion of other atoms. A timestep of 2 fs was kept constant during the simulation.

Table 2: The prepared systems along with their run time and number of independent replica simulation runs performed for each.

Systems for simulation	Production run time	Number of independent replica runs
Apo unphosphorylated	2 μ s	3
Apo phosphorylated (pT ²⁹⁰)	2 μ s	3
Complex unphosphorylated (Protein + ligand)	2 μ s	3
Complex phosphorylated (pT ²⁹⁰)	2 μ s	3
Protein + ATP + Mn ²⁺ ion	2 μ s	3
Protein+ ATP + Mg ²⁺ ion	2 μ s	3

The Mn-ATP-Protein complex and Mg-ATP-Protein complex were prepared using AdmetLab3.0 software. The PDBs are downloaded, and

they are used for simulations. For both cases, the force fields used are: GAFF2 for ligand, OPC for water, and ff19SB for protein. Solvation is done using the OPC water model, keeping the box size of 10 Å. The Antechamber package of AMBER18 is used to reduce the ligand, prepare all the topology files for the ligand, and prepare it for binding to the protein so that the protein-ligand complex simulation can be run.

3.3 Simulation Protocol

GaMD is used, and simulations are run in triplicate for 2000 ns.

3.3.1 Minimization

The systems are minimized to remove bond strains and unfavourable interactions. Minimization is performed in two steps. First, the solvent molecules are minimized while restraining the solute, using 5000 cycles of the steepest descent algorithm followed by 5000 cycles of the conjugate gradient algorithm. Next, solute and solvent are minimized together without restraints, employing 100 cycles of steepest descent followed by 900 cycles of conjugate gradient. Hence, it prevents the distortions of the conformation due to the high energetic interaction between the solvent and the solute. The step is done so that the system finds a local minimum and prevents it from “blowing up.”

3.3.2 Heating

Initially, the system was at 0 K, where it had no motion, and during the heating step, it was gradually heated to 300 K using the NVT approach. The heating is done gradually so that kinetic energy increases slowly and prevents the system from becoming unstable. The system was kept under controlled temperature and pressure conditions using the Langevin thermostat and Berendsen’s barostat.

3.3.3 Equilibration

As the production run is conducted under the NPT ensemble, a buffer time is required between the heating and production run to switch the ensemble. The equilibration stage aims to evolve the system from its

initial configuration to the equilibrium state. Here, specific thermodynamic properties, such as energy, temperature, pressure, etc., are monitored. Therefore, the equilibration is performed in the NVT³¹ ensemble, allowing for the kinetic and potential energies to be equilibrated. Each system went through 20 ns of equilibration, a 3.2 ns Conventional MD simulation was run, and finally, GaMD Equilibration was done for 54.4 ns in each run.

3.3.4 Production run

After the equilibration process, the “production” phase starts, where data is collected for further analysis. The collected data is retained in this run, but it is not true for equilibration. However, parameters are the same in both the equilibration and production stages. The production simulation was performed for 2 μ s in each run.

3.4 Post-simulation trajectory analyses

3.4.1 Root-Mean-Square Deviation (RMSD)

RMSD (Root-Mean-Square Deviation) quantifies the structural difference between two molecular conformations by calculating the square root of the average squared distances between corresponding atoms. It measures how much the atomic positions have changed during a simulation. Higher RMSD values suggest a greater deviation in nuclear positions, and lower RMSD values indicate lower deviations, implying a stable conformation.

3.4.2 Root Mean Square Fluctuation (RMSF)

It tells about the residual fluctuations of a protein or biomolecule. Higher RMSF values indicate extreme movement of the residues during the simulation. Lower RMSF values depict restricted movement of the residues, leading to a stable conformation.

3.4.3 Solvent-Accessible Surface Area (SASA)

It is the area of the biomolecule that is being exposed to water. Higher SASA says that specific amino acid residues of protein are more exposed towards water than those with less SASA.

3.4.4 Radius of Gyration (RoG)

In simple terms, the Radius of Gyration is the distribution of atoms around their centre of mass. In proteins, the RoG represents the compactness. If a protein has a high RoG value, it is less compact; the atoms are less tightly packed, possibly assuming an extended conformation. Lower RoG value implies that the protein is highly compact and has a tightly folded conformation. We study the compactness to check the changes that occurred after the simulation of the biomolecule.

3.5 Essential dynamic studies

3.5.1 Principal Component Analysis (PCA)

It is a technique used to reduce the dimensionality of data by transforming it into a new set of variables called principal components. These components are ordered by the amount of variance they capture from the original dimensions, with PC1 representing the highest variance, followed by PC2, PC3, and so on. It is calculated by diagonalizing the covariance matrix averaged over all the trajectories, eradicating the rotations and translations. The eigenvalue represents the magnitude of the component, and the eigenvector depicts the direction³².

3.5.2 Dynamic Cross Correlation Matrix (DCCM)

It is the study of correlations of atoms in a system, and the output is given in the form of a matrix. The DCC produces an $N \times N$ heatmap, where N is the number of alpha carbon atoms in the system, and each matrix element corresponds to the dynamic cross-correlation between the i and j atoms. Correlation values range from -1 to $+1$, where $+1$ indicates perfect correlation, -1 indicates perfect anti-correlation, and 0

signifies no correlation. The diagonal elements of a correlation matrix typically show a value of +1, reflecting an ideal correlation of each atom with itself. A positive correlation suggests that atoms move in the same direction and phase over time, whereas a negative (anti) correlation indicates that atoms move in opposite directions or phase³³.

3.5.3 Dihedral Principal Component Analysis (dPCA)

It is used to study the structural rearrangement of loop regions. The dihedral angles are considered for the calculation. Hence, much of the overall motions can be eradicated from internal ones. The main advantage of using the backbone dihedral angle is that it will naturally undergo more changes than internal coordinates like bond length and bond angle for a molecule. Therefore, dihedral angle principal component analysis (dPCA), which is based on internal coordinates – the dihedral angle (ϕ_n, ψ_n) of the protein backbone, may help to separate the internal motions from overall dynamics³⁴.

3.5.4 Protein Structure Network Analysis (PSN)

An alternative way to analyze protein structures beyond secondary structure and fold arrangements is to represent the interactions between residues as a network. The network can be built based on $C\alpha$, $C\beta$, atom pairs, centroid networks, or interaction-strength networks. Here, the node is created based on the $C\alpha$ atom of an amino acid residue, and an unweighted edge is constructed if the paired residue $C\alpha$ - $C\alpha$ distance lies within the threshold distance (RC) of ~ 7 Å. Overall network view supports various analyses for identifying functional residue, predicting coevolving residues, understanding the mechanism of protein-protein interaction or domain-domain interaction, for understanding the communications between them³⁵.

3.6 Free energy calculation by MMPB(GB)SA

The binding free energy of a ligand can be calculated by the MM/PB(GB)SA method. The method MM/PBSA³⁶ uses the Poisson-Boltzmann Equation to calculate the binding enthalpy, and MM/GBSA

uses the Generalized Born Equation³⁶. The former method is more accurate and time-consuming as it considers all the electrostatic interactions. The GB method approximates the PB Equation and is less precise. It is also not suitable for molecules having high electrostatic interactions.

$$\Delta G = \Delta H - T\Delta S \quad [2.24]$$

$$\Delta H = \Delta E_{internal} + \Delta G_{solv} \quad [2.25]$$

$$\Delta E_{internal} = E_{covalent} + E_{non-covalent} \quad [2.26]$$

$$\Delta G_{solv} = \Delta G_{polar} + \Delta G_{non-polar} \quad [2.27]$$

$$\Delta G_{non-polar} = \alpha SASA + \beta \quad [2.28]$$

Here, α is the surface tension coefficient, and β is the offset constant. ΔH represents a change in enthalpy, ΔG is the Gibbs free energy change, and ΔS is the entropy change.

3.7 Entropy calculation

It is calculated by Normal Mode Analysis³⁷. Entropy refers to the degree of randomness of a system. Higher entropy implies greater randomness, and lower entropy resembles a stable system.

3.8 K-means clustering

It is an unsupervised ML³⁸ approach that allows grouping of datasets into different clusters based on similarity. The data points are divided into K number of clusters, which maybe done manually or by K-means³⁹ ++ methods. Then a centroid value is chosen from the data points, that corresponds to its mean. The distance is calculated between each data point and the centroid and this process is iterated by updating the centroid values, until there is no difference between the distances between centroid and data point. WCSS⁴⁰ is a method that calculates the square of the distance between data point and centroid. A lower WCSS signifies a well-defined cluster, where centroid is close to the point.

3.9 Computer-Aided Drug Design

A drug is a molecule that binds to a specific receptor present in our body and inhibits or activates some metabolic or cellular processes to treat certain diseases. The advent of the in-silico world has led to drug design methods using computational tools, reducing the pressure on wet lab experiments. AI/ML at present has outspread and extended its helping hand to drug discovery.

The Drug discovery pipeline includes:

1. Protein/target identification.
2. Target preparation.
3. Choosing a library of ligands.
4. Virtual screening.
5. Pre-clinical trials.
6. Clinical trials.
7. FDA approval of drugs.
8. Post-marketing safety analysis.

Drug development typically takes 7–10 years before a compound reaches the market. In-silico tools such as ProTox⁴¹, ADMET-AI⁴², and SwissADME⁴³ have significantly streamlined this process by enabling early assessment of drug toxicity and allergenicity. Additionally, computational tools are available for pharmacokinetic studies, and Computer-Aided Drug Design⁴⁴ (CADD) offers the potential to develop personalized medicine strategies tailored to individual patient profiles.

3.10 Virtual Screening

Virtual Screening⁴⁵ is an in-silico process that checks the binding energy or docking score when many ligands are docked with a protein. This high-throughput process requires computational tools so that the docking results are generated within less time, and many compounds can be screened simultaneously. It is an essential step in Drug Discovery. Tools like Schrodinger, AutoDock⁴⁶, and MolSoft⁴⁷ are used for this

process. We can screen a large number of compounds in a short amount of time, relieving the pressure on wet-lab experiments. The hit compounds identified are then converted to leads and optimized further. This thesis uses MolSoft software for VS and DelaDrug3.0⁴⁸ software for Lead optimization. The VS technique involves: Structure Based Drug Design⁴⁹ and Ligand Based Drug Design.

3.11 ADMET Analysis

The word stands for Absorption, Distribution, Metabolism, Excretion, and Toxicity related to drugs. It tells how toxic the drug is for human use and the extent of its absorption and metabolism. It is used to screen a drug to check whether it is fit for consumption or not. A drug should follow Lipinski's rule and toxic parameters, that is, it should not be genotoxic or cytotoxic and should not cause harm to the organs. The drug is easily absorbed into the blood and evenly distributed. It should be properly excreted as well, or else the degraded products of the drug, when accumulated in the body, may pose serious threats. ADMETLab3.0⁵⁰ is used in this thesis.

3.12 Rules for Drug Likeness

There are several rules to choose whether the drug is consumable or not. Various parameters like molecular weight, H-bond formation, absorption ability, excretion, and hydrophilicity must be characterized before choosing it as a lead compound against the target.

Lipinski's rule⁵¹ of 5 states that the molecular weight of a Drug should be less than or equal to 500 g/mol, the log P⁵² value should be less than 5, the H-bond acceptor less than 10, and the H-bond donor less than 5.

MDDR-like rule⁵³ was developed by Opera and focuses on parameters like the number of rings, rigid bonds, and rotatable bonds. As per this rule, there should be fewer than 10 rotatable bonds, as if there are a very high number of them, then the ligand becomes too flexible and might move out of the binding pocket.

Veber rule⁵⁴ focuses on polar surface area (PSA) related to oral bioavailability. It states that PSA should be less than $140 \text{ \AA}^2$ to absorb and cross the cell barrier easily. PSA is the surface of the drug exposed to solvent; in simple words, it means the sum of the surface area of polar atoms in the drug. The Ghose filter⁵⁵ is much more stringent than the Veber rule, stating that PSA should be less than $120 \text{ \AA}^2$. PSA of less than $90 \text{ \AA}^2$ is a promising drug molecule that can easily absorb and cross the BBB (Blood Brain Barrier). Log P⁵² more than 5 implies the drug is highly lipophilic, and less than zero means it is highly hydrophilic. So, it should have a value that has a good balance between lipophilicity and hydrophilicity. The highly hydrophilic molecule is unwanted as it is highly water soluble and is easily excreted, while a highly lipophilic molecule attracts fat more and is less soluble in water.

CHAPTER 4

OBJECTIVES

Objective 1

To study the structural dynamics of the protein COT kinase in both apo and complex form. Also to check the dynamics of the systems in phosphorylated form and unphosphorylated form.

Objective 2

To find a potent inhibitor against the COT Kinase using Computer Aided Drug Discovery techniques.

Objective 3

Study the differential effect of Mg^{2+} and Mn^{2+} on the dynamics of COT kinase.

CHAPTER 5

RESLUTS AND DISCUSSION

Objective1: Study of the Structural Dynamics of apo and complex systems in both unphosphorylated and phosphorylated forms.

Structural analysis- Overall stability, stability of conserved regions and ligand binding dynamics.

To check the simulation convergence, stability of the systems, we have calculated the backbone RMSD of the systems that included all the N, C, C _{α} atoms. The **Figures 5.1** (A-D) shows that all the systems have attained convergence at the end of all the runs, imparting stability to the systems.

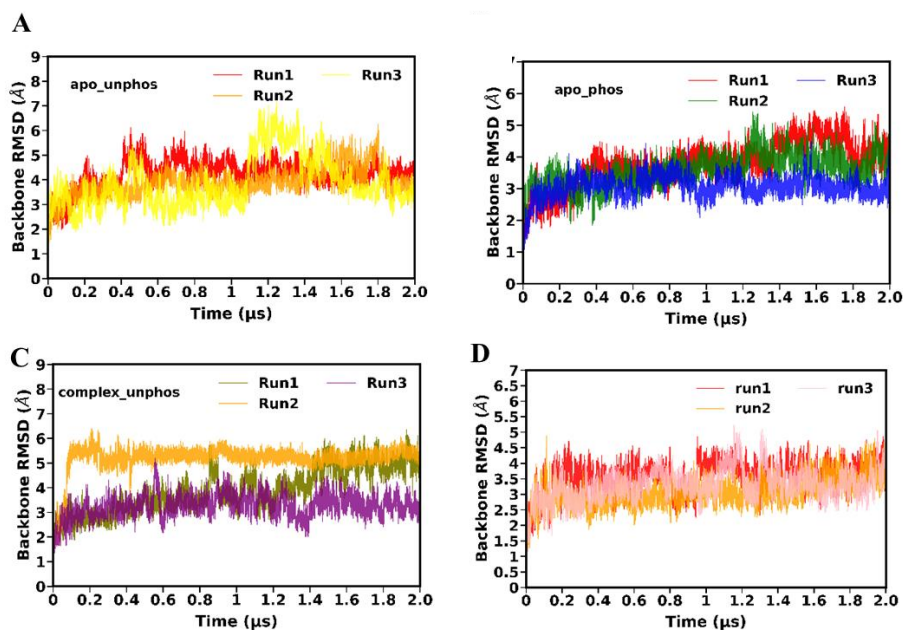


Figure 5.1: The Figures portray the time evolution for all the three runs in all the systems (apo and complex). The backbone RMSD values range from 4-6 Å for all the systems. A- Apo unphosphorylated system, B- Apo phosphorylated system, C- Complex unphosphorylated system and D- Complex phosphorylated system.

Apo unphosphorylated system shows highest fluctuation in P-loop as the RMSD ranges from 8-9 Å in **Figure 5.2**. Phosphorylation of apo system lowered the deviation to some extent. Complexation strongly stabilized the P-loop as the RMSD ranges from 4-5 Å.

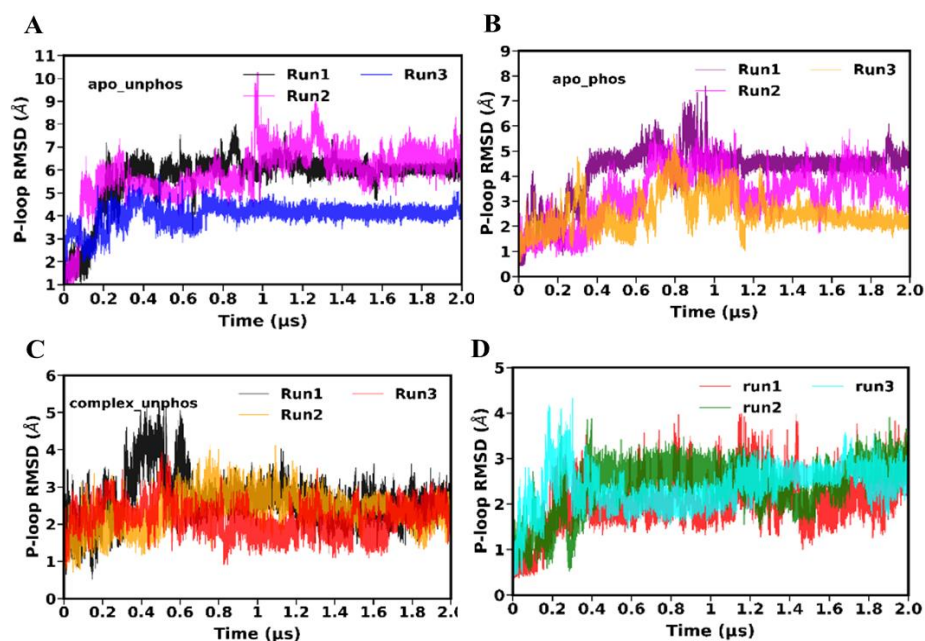


Figure 5.2: The Figures portray the time evolution plot of P-loop RMSD for all the three runs in all the systems (apo and complex). A- Apo unphosphorylated system, B- Apo phosphorylated system, C- Complex unphosphorylated system and D- Complex phosphorylated system.

From **Figure 5.3**, it can be stated that α C-helix for the systems is stable throughout the simulation, indicating no evident structural deviation. The RMSD ranges around 1 Å.

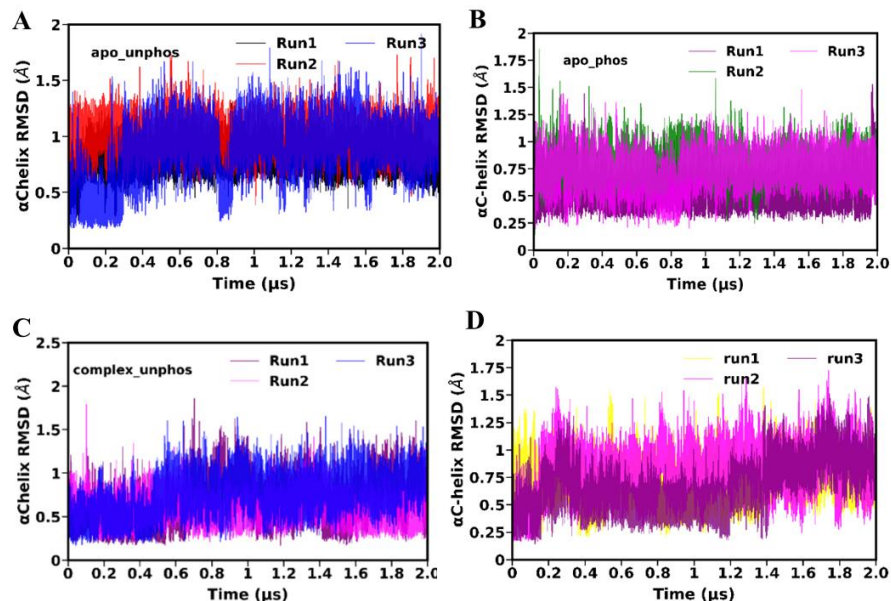


Figure 5.3: The Figures portray the time evolution plot of α C-helix RMSD for all the three runs in all the systems (apo and complex). A- Apo unphosphorylated system, B- Apo phosphorylated system, C- Complex unphosphorylated system and D- Complex phosphorylated system.

Activation loop is more or less stable for the four systems as seen in **Figure 5.4**. Slight deviations are observed in complex systems and it is due to the fact that on ligand binding, the Activation loop have adopted a new conformation for allowing or disallowing the substrate binding. For apo, the RMSD is around 2 Å and for complex it increases to 4 Å.

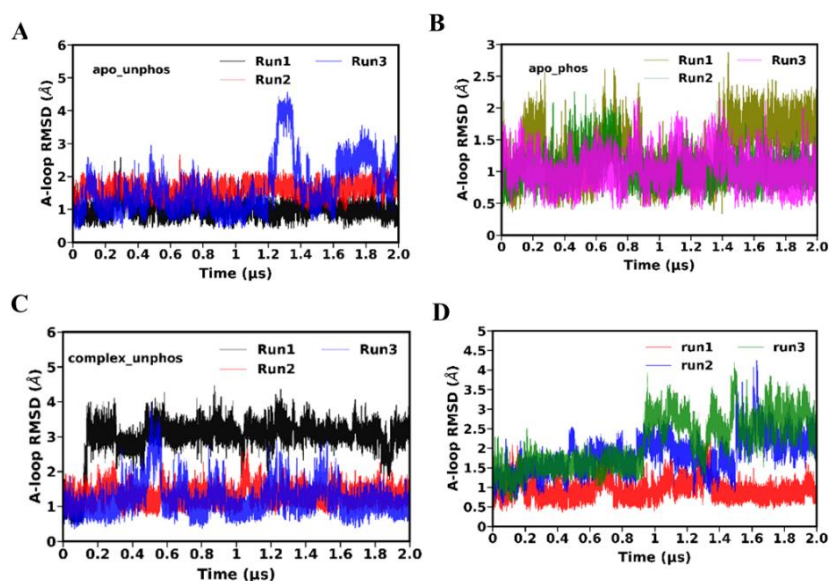


Figure 5.4: The Figures portray the time evolution plot of A-loop RMSD for all the three runs in all the systems (apo and complex). A- Apo unphosphorylated system, B- Apo phosphorylated system, C- Complex unphosphorylated system and D- Complex phosphorylated system.

The Catalytic loop is rigid for the systems. Their RMSD ranges within the range of 1-1.25 Å. Runs are converged at the end of the simulations in **Figure 5.5**.

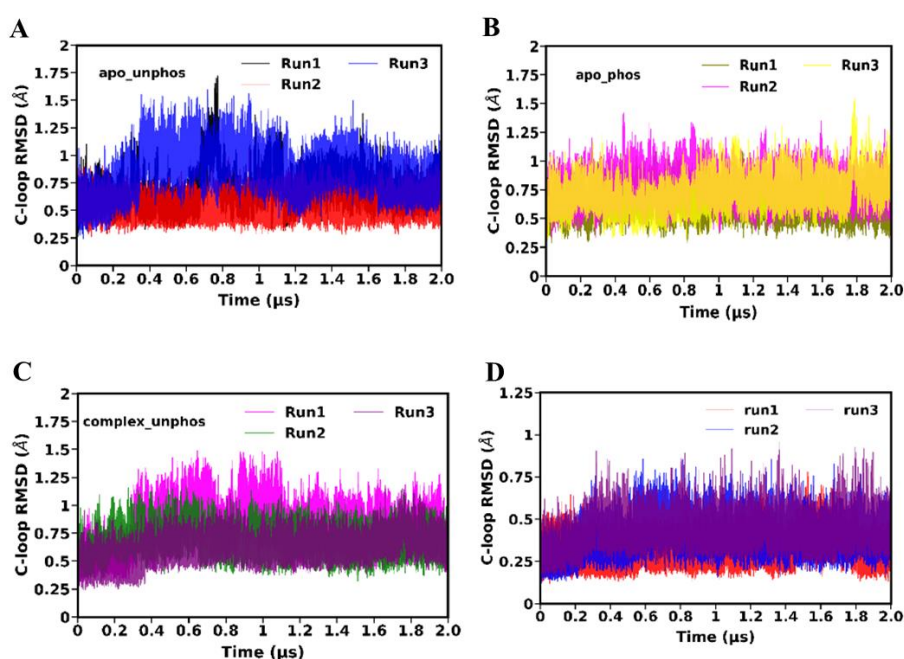


Figure 5.5: The Figures portray the time evolution plot of C-loop RMSD for all the three runs in all the systems (apo and complex). A- Apo unphosphorylated system, B- Apo phosphorylated system, C- Complex unphosphorylated system and D- Complex phosphorylated system.

Residual fluctuation analysis:

The protein residues may be flexible or rigid. High RMSF values indicate higher flexibility, like P-loop of the molecule shows very high

residual flexibility especially for the apo unphosphorylated system. The α C-helix and C-loop are not showing fluctuations, rather they are rigid for the systems. Phosphorylation and complex formation is stabilizing the P-loop and restricting its extensive movements. Complex unphosphorylated system shows high fluctuation of the Activation loop, but the phosphorylation is making the loop stable. C-terminal end of the protein has evident flexibility may be due to structural changes on ligand binding and phosphorylation as in case of **Figure 5.6**.

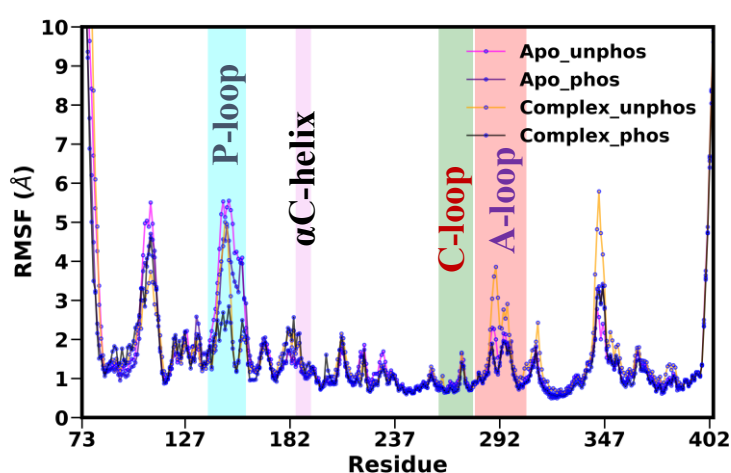


Figure 5.6: The Figure reveals residual fluctuations for all the four systems.

KDE plot is a graphical representation of the probability distribution of non-parametric or undistributed data sets. The **Figure 5.7** shows the probability density of backbone RMSD for all the four systems. It showcases which RMSD value of the system has the highest probability. The Figure illustrates that both the complex phosphorylated system and the apo phosphorylated system undergo similar structural changes during simulation, hence their peaks are almost identical and in the same RMSD range, ~ 3.5 Å. An unphosphorylated system has slightly more structural deviation than the phosphorylated one, and its RMSD is around 4 Å. The unphosphorylated system initially shows a similar structural conformation to the phosphorylated one, but it adopts a new conformation whose RMSD is around 5.5 Å.

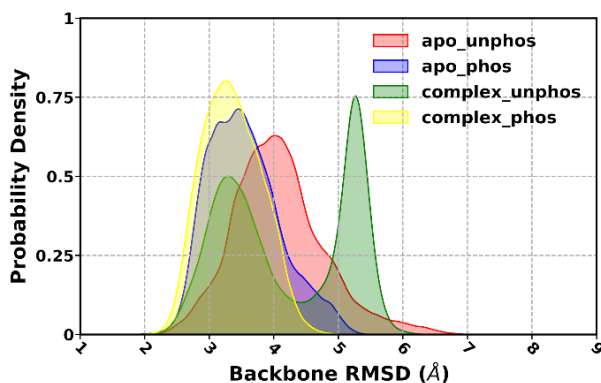


Figure 5.7: Probability distribution curve of backbone RMSD for apo and complex systems respectively.

The **Figure 5.8** shows the probability density of A-loop RMSD for all four systems. Initially, all the systems have undergone similar structural changes during simulation, hence their peaks are almost identical and in the same RMSD range, ~ 1 - 1.25 Å. Apo phosphorylated system shows a single sharp peak and suggests a rigid activation loop. Two peaks are observed for the complex phosphorylated system and the apo unphosphorylated system, one at around 1 Å and the other at 1.75 Å. The complex unphosphorylated system shows bimodal distribution, and it has adopted a new structure with an RMSD of around 3.25 Å.

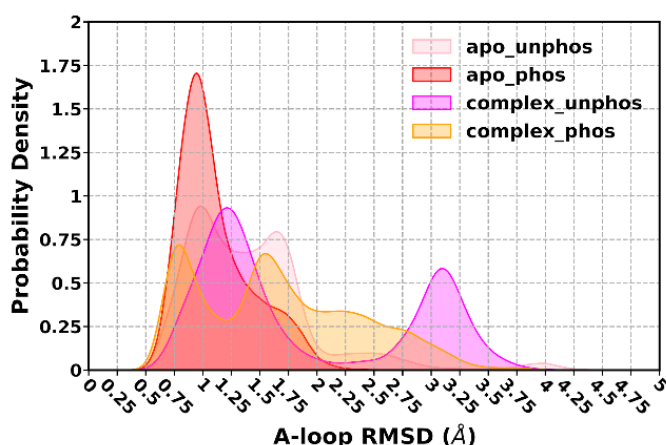


Figure 5.8: Probability distribution curve of A-loop RMSD for apo and complex systems respectively.

The probability distribution curve suggests that overall, the Catalytic loop is rigid for all the systems. **Figure 5.9** shows the probability density of C-loop RMSD for all four systems. All the peaks show unimodal distribution and a sharp peak. The complex phosphorylated system shows a sharp peak at around 0.3 Å. The C-loop is rigid for the same system. Apo phosphorylated system and Complex unphosphorylated system underwent similar conformational changes.

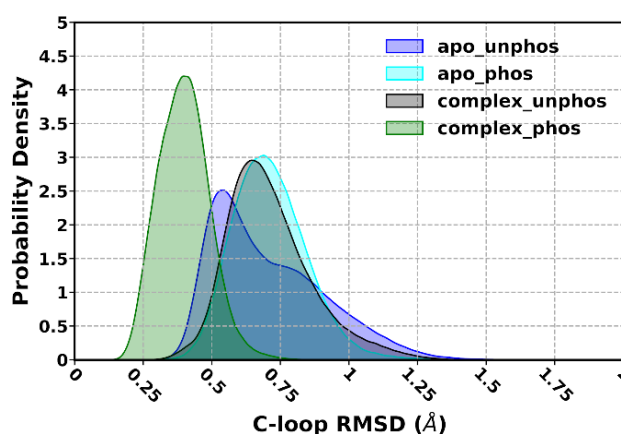


Figure 5.9: Probability distribution curve of C-loop RMSD for apo and complex systems respectively.

The **Figure 5.10** shows the probability distribution plot of P-loop RMSD for all four systems. Apo unphosphorylated system shows highest structural deviation. Complex formation stabilizes the P-loop. So, it can be inferred that the ligand, when bound to the pocket, imparts rigidity to the P-loop, restricting its movement. An extra peak is observed in the case of a complex phosphorylated system, but the RMSD ranged from 2-3 Å. A wide range of distribution is observed in the case of the apo phosphorylated system.

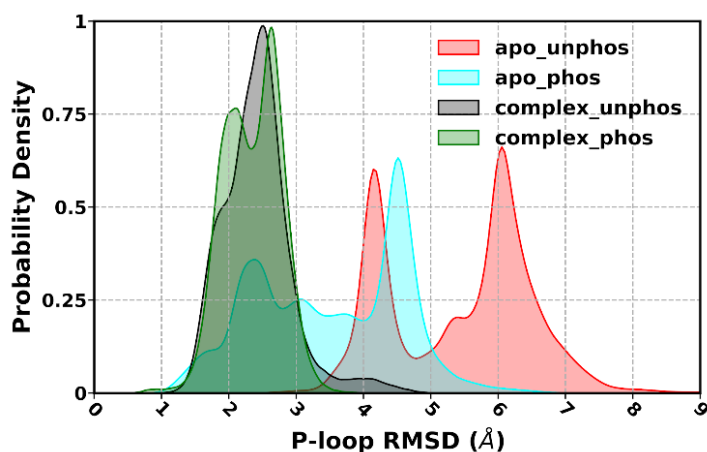


Figure 5.10: Probability distribution curve of P-loop RMSD for apo and complex systems respectively.

The **Figure 5.11** shows the probability distribution plot of α C-helix RMSD for all four systems. Not extreme structural deviations are observed and two peaks are formed in apo phosphorylated system. Broad peaks are present for both the complex phosphorylated system and complex unphosphorylated systems, indicating the presence of many conformations within the range of that RMSD.

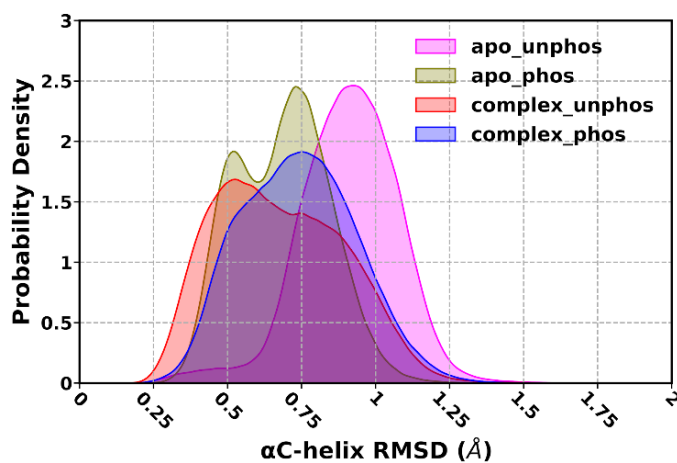


Figure 5.11: Probability distribution curve of α C-helix RMSD for apo and complex systems respectively.

Protein compactness study:

Overall, the compactness of protein remained similar for the systems. Based on **Figure 5.12**, we can't strongly comment on the fact that phosphorylation or complexation had any effect on the tight or loose binding of protein (compactness). The RoG values ranged from 21-21.5 Å.

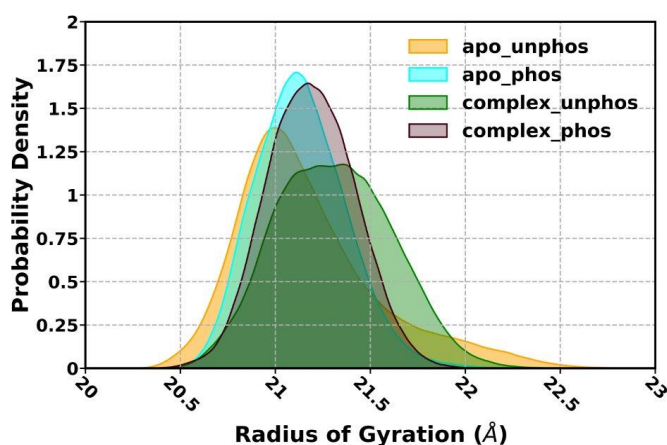


Figure 5.12: The Figure represents the Probability distribution curve of Radius of Gyration for all the systems (apo and complex).

Study of Solvent Accessibility:

The **Figure 5.13** portrays that the solvent accessibility is similar for the systems, there is no significant changes in them. The similar residues are equally exposed to the solvent and there is no significant change in the residues that got exposed due to complexation or phosphorylation. Thus, it had no effect on solvent accessibility.

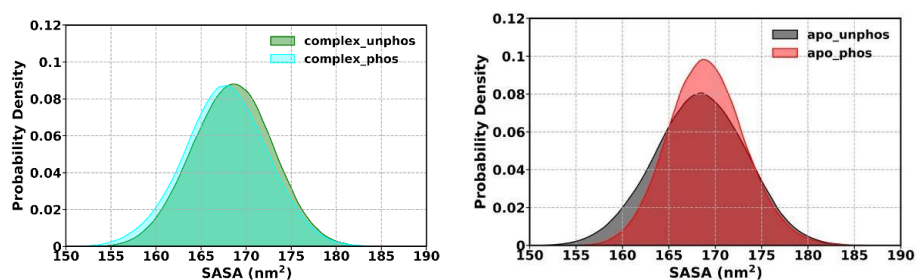


Figure 5.13: Probability distribution curve of Solvent Accessible Surface Area for the systems.

Table 3: The backbone RMSD, RoG and SASA are calculated for the systems by Block Average method. The standard deviations are given in parentheses.

<i>Systems</i>	<i>Backbone RMSD (in Å)</i>	<i>Radius of Gyration (in Å)</i>	<i>SASA (in nm²)</i>
<i>Apo_unphos</i>	$3.97 \pm (0.68)$	$21.17 \pm (0.32)$	$168.49 \pm (3.6)$
<i>Apo_phos</i>	$3.45 \pm (0.55)$	$21.14 \pm (0.18)$	$169.3 \pm (2.47)$
<i>Complex_unphos</i>	$4.10 \pm (0.99)$	$21.28 \pm (0.25)$	$168.5 \pm (2.82)$
<i>Complex_phos</i>	$3.29 \pm (0.41)$	$21.18 \pm (0.16)$	$167.72 \pm (3.15)$

Ligand and Pocket stability:

The **Figure 5.14 A** implies that phosphorylation stabilizes the ligand. The ligand is slightly more flexible in case of unphosphorylated system. **Figure 5.14 B** tells that complexation stabilizes the binding pocket. The pocket fluctuates too much when it is free, that is during the absence of a ligand (apo form), the RMSD reaches upto 5 Å.

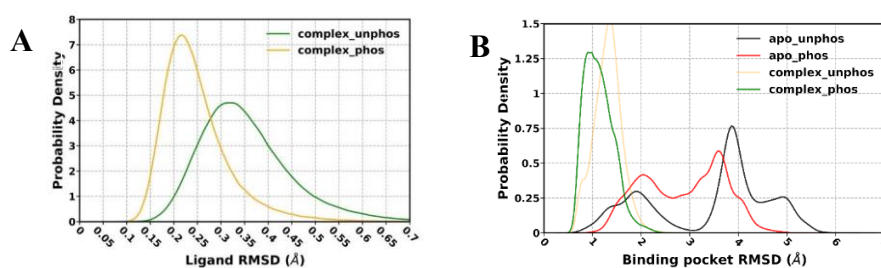


Figure 5.14: Probability distribution curve of ligand RMSD and Binding pocket RMSD.

Stability of motifs:

Overall, the HHD motif is stable. Complex unphosphorylated system shows more flexibility of the motif as it has a higher RMSD as compared to other systems. DFG motif is also stable and on complex formation, it showed similar structural variations as highlighted in **Figure 5.15**.

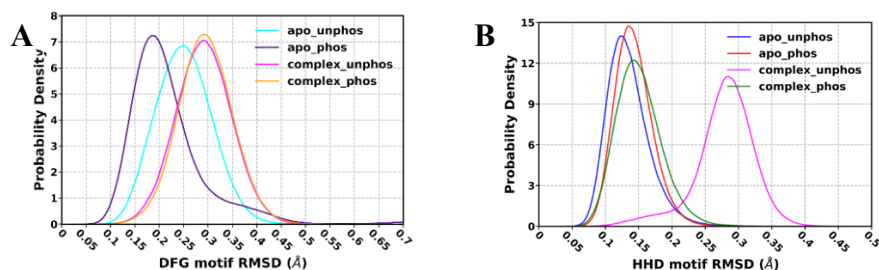


Figure 5.15: Probability distribution curve of DFG motif RMSD and HHD motif RMSD.

Salt bridge analysis:

Normally, salt bridge is formed between the Lys of $\beta 3$ helix and Glu of αC -helix. But, in this kinase there is a frameshift mutation that leads to salt bridge formation between Lys of $\beta 3$ helix and Asp of αC -helix. This is essential for proper kinase functioning and placement of ATP in the binding pocket to ensure proper catalysis. Generally, salt bridge is formed within 4 Å from one residue to other. The **Figure 5.16** shows salt bridge formation in various systems. A sharp peak is observed in case of complex phosphorylated system, indicating formation of strong salt bridge.

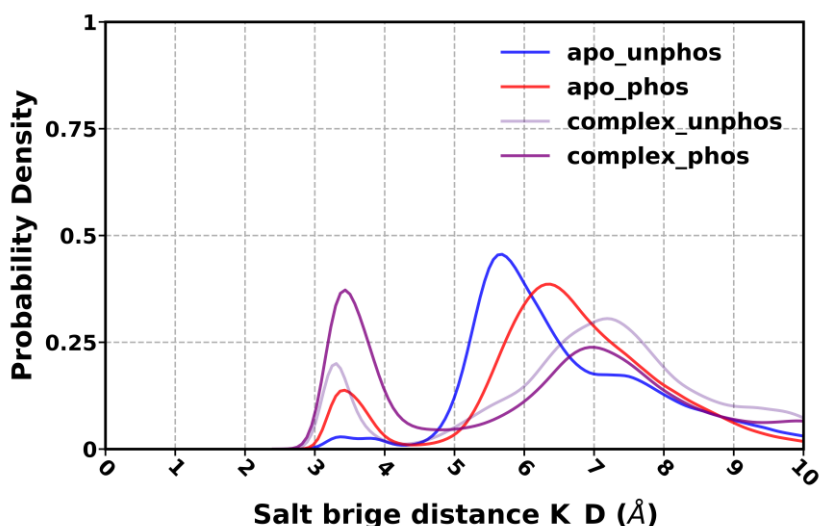


Figure 5.16: The Figure shows salt bridge formation between K104 and D115 in all the systems.

Hydrogen bond analysis:

H bond analysis is a very important study used to analyse ligand binding strength to the protein. According to Lipinski rule, the H bond donor should be less than or equal to 6 and H bond acceptor less than 10. It is a very crucial parameter to design a drug and choose a potential drug candidate. It also portrays which residues of protein and ligand are participating in the bonding, thus modifying it would make a better drug.

The **Figure 5.17** shows time evolution of number of H bonds formed by complex in both phosphorylated and unphosphorylated form. Both the system's H bond range from 1-2, thus the phosphorylation of protein showed no significant difference in the number of H bonds formed on ligand binding.

Table 4: Hydrogen bond analysis for the complex systems.

System	Run	Hydrogen bond		Occu- pancy	Ave- rage dist ance	Average angle
		Acceptor	Donor			
Complex unnphos	1	UNK_333 @N4	GLY_147@ N	32.58 %	2.91 Å	159.2°
		GLU_145 @O	UNK_333 @N3	73%	2.85 Å	158.7°
		ASP_207 @OD1	UNK_333 @N	20.55 %	2.83 Å	164°
		ASP_207 @OD2	UNK_333 @N	13.55 %	2.83 4 Å	163.25°
	2	UNK_333 @N4	GLY_147@ N	28.98 %	2.91 Å	153.6°
		GLU_145 @O	UNK_333 @N3	72.14 %	2.84 Å	162°
		ASP_207 @OD1	UNK_333 @N	11.57 %	2.83 6 Å	163°
		ASP_207 @OD2	UNK_333 @N	13.13 %	2.83 Å	163°
	3	UNK_333 @N4	GLY_147@ N	25.6%	2.92 Å	153.6°
		GLU_145 @O	UNK_333 @N3	72.85 %	2.84 Å	162.18°
Complex phos	1	UNK_333 @N4	GLY_147@ N	34.83 %	2.91 Å	153.48°
		GLU_145 @O	UNK_333 @N3	77.69 %	2.84 Å	162.5°

		GLU_154 @OE2	UNK_333 @N5	26.05 %	2.82 Å	156.84°
		GLU_154 @OE1	UNK_333 @N5	22.28 %	2.82 Å	155.39°
	2	UNK_333 @N4	GLY_147@ N	23.87 %	2.92 Å	160.2°
		GLU_145 @O	UNK_333 @N3	73.10 %	2.84 Å	161.5°
		ASP_207 @OD2	UNK_333 @N	12.33 %	2.83 Å	163.59°
		ASP_207 @OD1	UNK_333 @N	14.88 %	2.83 Å	163.63°
	3	UNK_333 @N4	GLY_147@ N	26.36 %	2.92 Å	152.67°
		GLU_145 @O	UNK_333 @N3	76.62 %	2.84 Å	161.38°
		ASP_207 @OD2	UNK_333 @N	12.75 %	2.83 Å	162.36°
		ASP_207 @OD1	UNK_333 @N	12.17 %	2.83 Å	163.3°

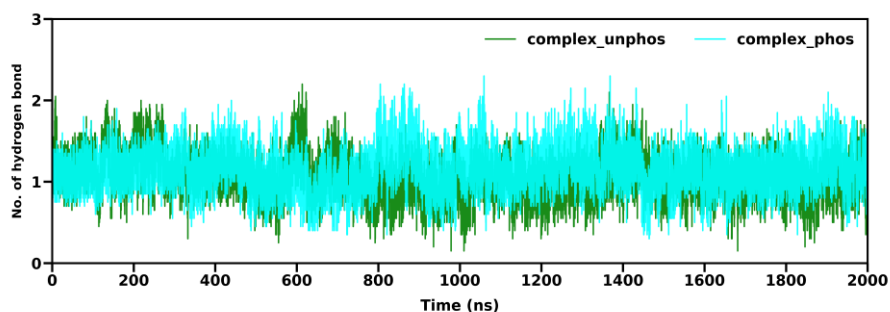


Figure 5.17: Time series plot for Number of Hydrogen bonds formed by the ligand with the protein in unphosphorylated and phosphorylated forms respectively.

Principal Component Analysis:

The **Figure 5.18** suggests that apo unphosphorylated system has three global minima, separated by low energy barriers. Thus, the three conformations are easily interconvertible and the conformations are displayed in the same figure. Apo phosphorylated system has three conformations. Complex unphosphorylated is most stable as it has only one minimum.

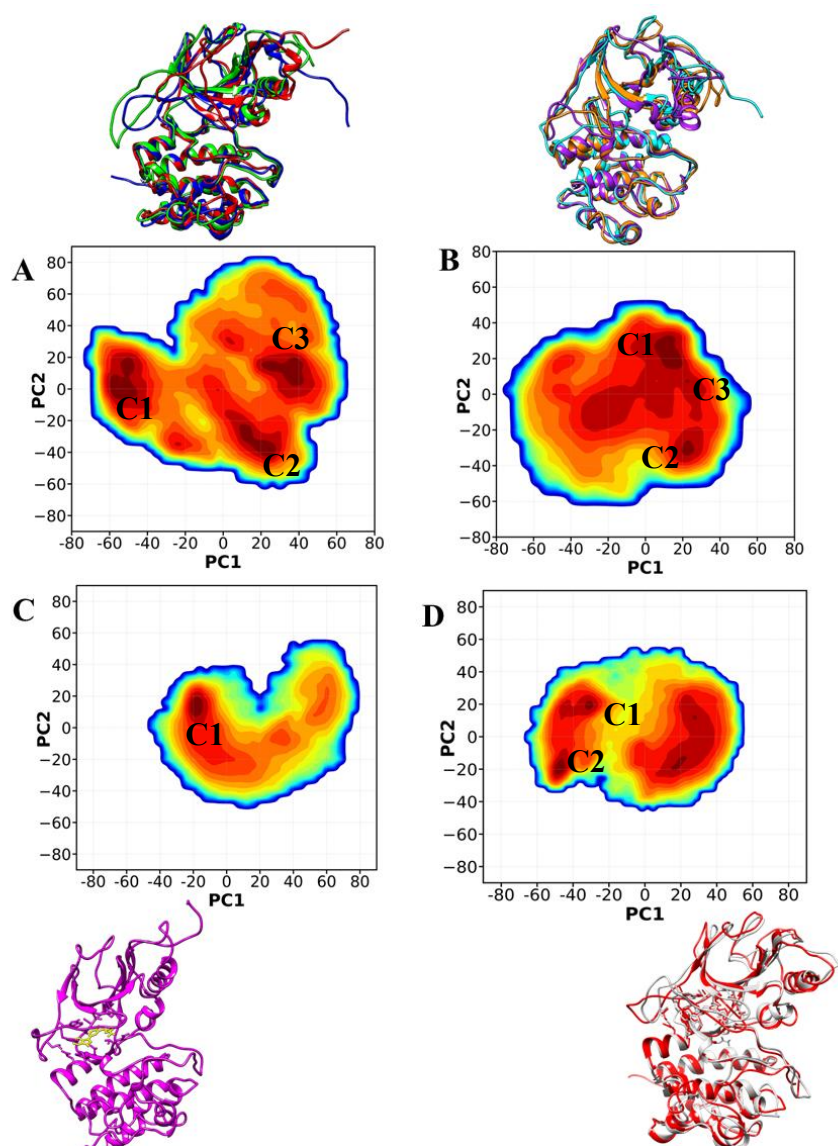


Figure 5.18: Free Energy Landscape and extracted structures from the minima. A- Apo unphosphorylated system, B- Apo phosphorylated system, C- Complex unphosphorylated system and D- Complex phosphorylated system.

Dynamic Cross-Correlation Matrix:

The highlighted region R1 in **Figure 5.19** shows more anti-correlation in apo unphosphorylated system, which is significantly reduced in the other three systems. However, slightly more anti-correlation is also observed in complex phosphorylated system. The region shows correlation or anti-correlation between the C-terminus of the protein and the α C-helix.

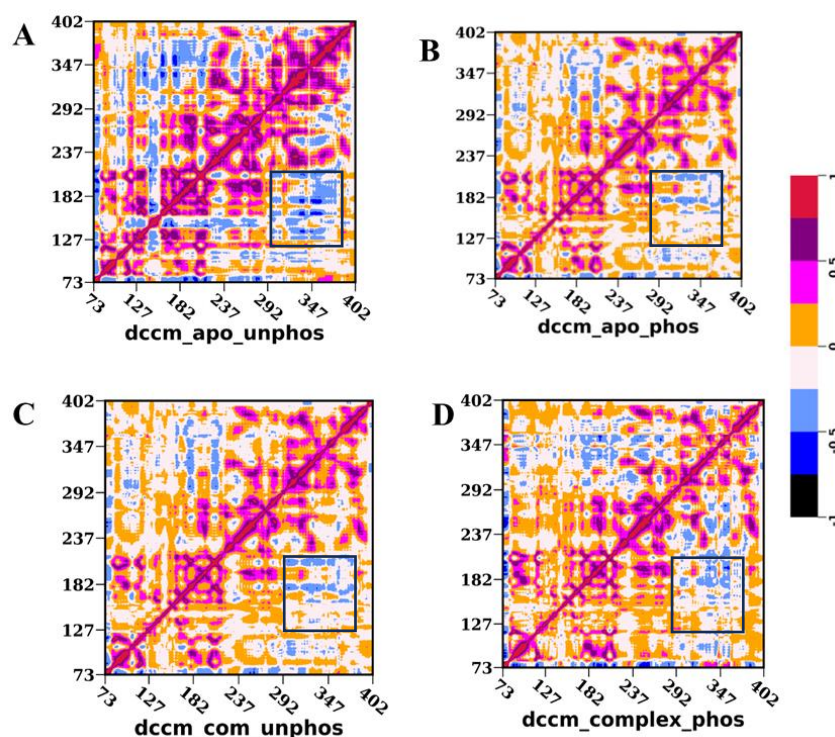


Figure 5.19: DCCM of the four systems.

Dihedral Principal Component Analysis:

The **Figure 5.20** shows that apo unphosphorylated system, apo phosphorylated system and complex unphosphorylated system have two global minima; rendering a stable A-loop. Complex phosphorylated system has five global minima, thus it is not that stable.

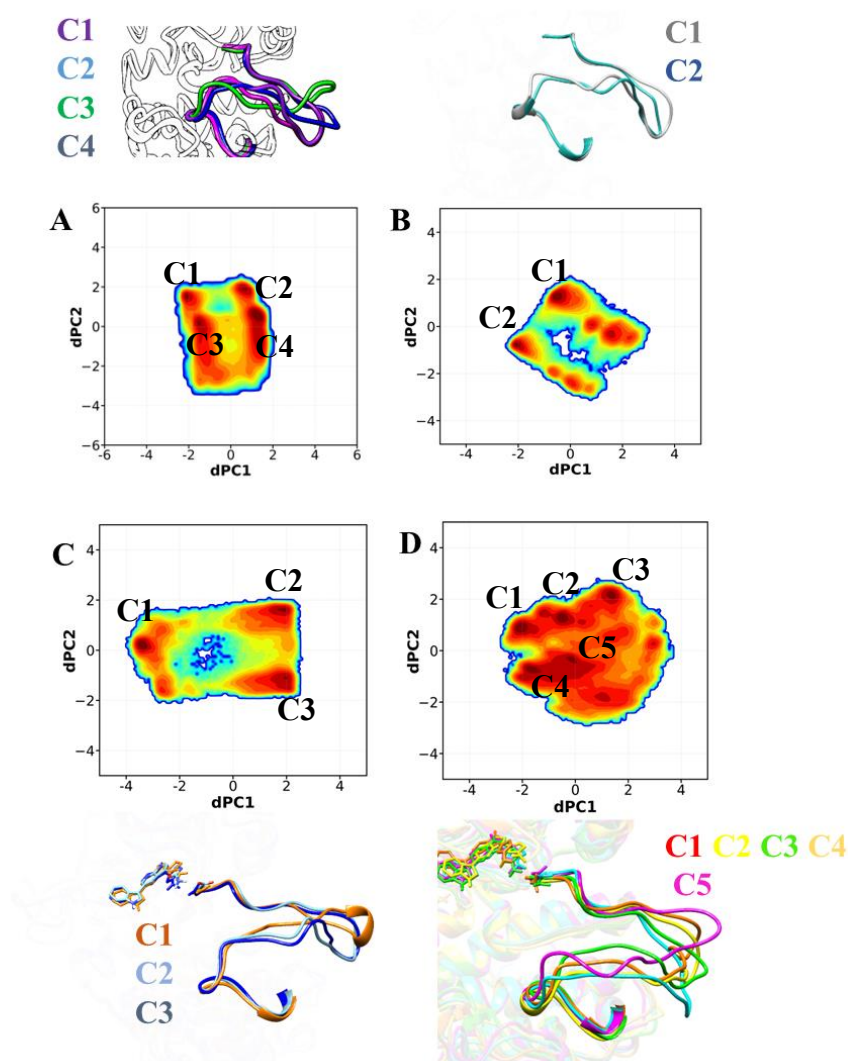


Figure 5.20: dPCA of the A-loop to check their structural rearrangements. A- apo unphosphorylated system, B- apo phosphorylated system, C- complex unphosphorylated system and D- complex phosphorylated system. The P-loop structures are provided along with the FES.

K-means clustering:

The data points of the systems are divided into clusters based on WCCS method. Here, the elbow plot depicts that all the systems' data points can be divided into 3 clusters, $K=3$. Then, the clusters are depicted in the graphs for individual systems, along the percentage occupancy. The percentage tells how much prominent the conformational state is

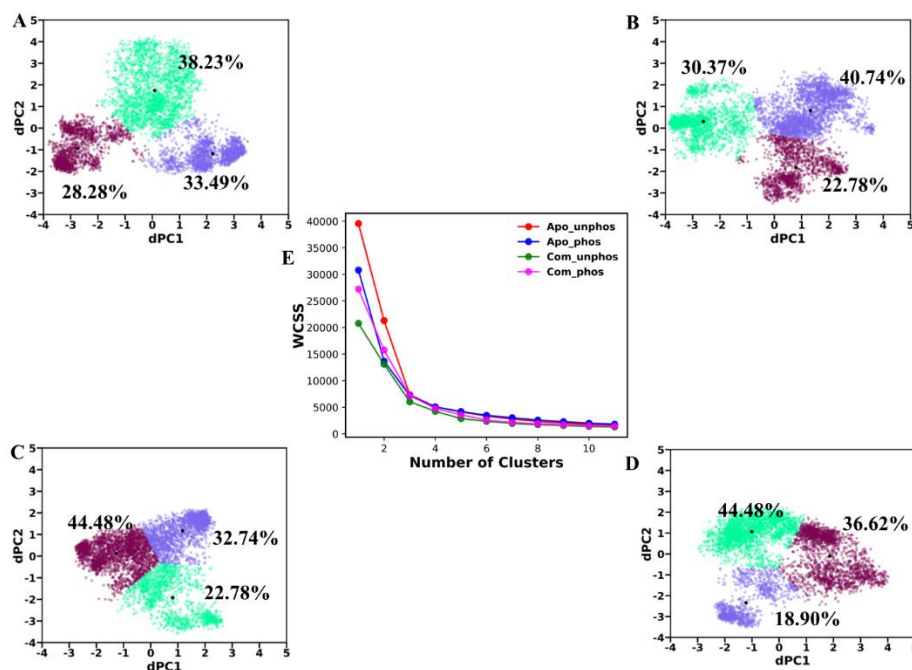


Figure 5.21: K-means clustering. E- Elbow plot, A- Apo unphosphorylated system, B- Apo phosphorylated system, C- Complex unphosphorylated system and D- Complex phosphorylated system.

Binding free energy calculation by MMPBSA:

The binding free energies of complex unphosphorylated system and complex phosphorylated system were checked. van der Waals interaction and Electrostatic interaction favours binding in both the system. The phosphorylated system had greater binding enthalpy than unphosphorylated one as evident from the table. Polar solvation enthalpy discourages binding due to positive enthalpy.

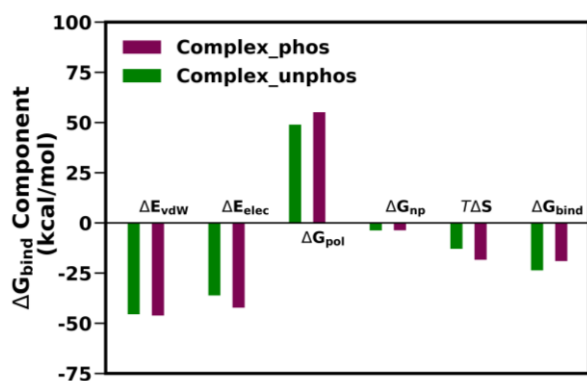


Figure 5.22: Graphical representation of different energy values calculated by MMPB(GB)SA method for all the four systems.

Per-residue enthalpy:

Trp 132, Pro 145 and Val 269 shows negative binding enthalpy for both the systems and thus these residues favour binding as observed in **Figure 5.23**.

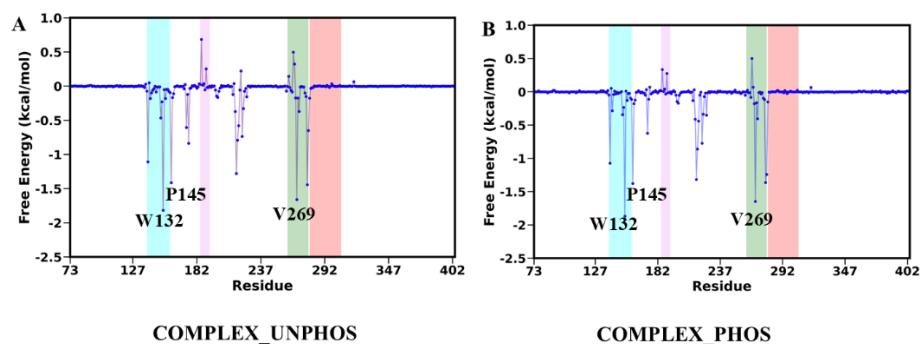


Figure 5.23: Per-residue binding energy of complexes.

Protein Structure Network analysis:

In PSN, the nodes represent the amino acid residues of a protein, secondary structure or atoms present. The edges depict the interactions present between the atoms or residues. Hubs are the entities or the residues that are connected to four residues and interacts with each other. Communities are group of proteins that are densely connected to one another than that of others.

Table 5: The Protein Structure Network analysis for various systems.

Network Properties	Apo unphos	Apo phos	Complex unphos	Complex Phos
I_{min}	2.62	2.91	1.71	2.85
Number of Linked Nodes	301	298	302	310
Number of Links	379	351	371	348
Number of Hubs	68	49	60	43
Number of Links mediated by Hubs	230	170	213	163
Number of Communities	9	11	12	11
Number of Nodes involved in Communities	83	60	73	48
Number of Links involved in Communities	143	96	115	76

The **Table 5** reflects that apo unphosphorylated system has highest number of links and hubs, followed by complex unphosphorylated system. Complex unphosphorylated system shows highest number of communities, portraying a stable structure. Hence, it is densely interconnected as compared to other nodes.

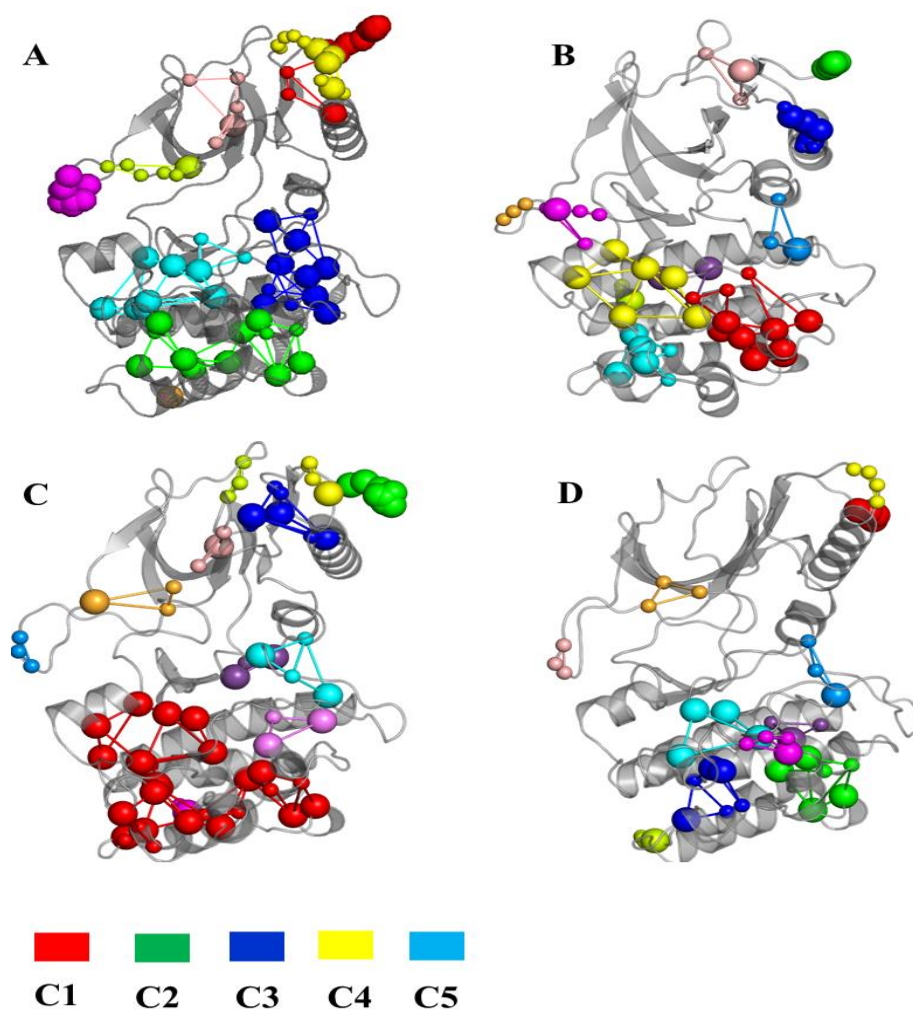


Figure 5.24: Network analysis of A- apo unphosphorylated, B- apo phosphorylated, C- complex unphosphorylated and D- complex phosphorylated. The red represents highest number of communities followed by green, blue, yellow, cyan and pink.

Per-residue RMSD of P-loop:

The structure of P-loop is very intriguing and hence we have studied the per residue RMSD of the heavy atoms present in the P-loop to check which residues are contributing to structural deviations. The **Figure 5.25** shows that the loop is too much flexible in case of apo unphosphorylated system and on complexation the higher RMSD values (marked in orange-yellow) significantly reduces. Asparagine 75, Aspartate 79, Isoleucine 76 are significantly contributing to high RMSD. Ser 78 shows fluctuations in all the system except complex phosphorylated one.

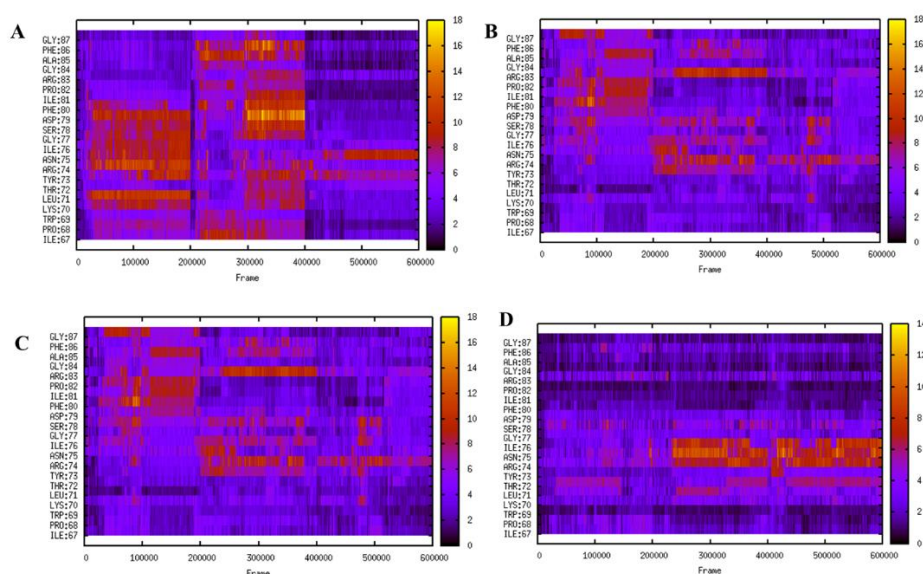


Figure 5.25: Per-residue analysis of P-loop of A- apo unphosphorylated, B- apo phosphorylated, C- complex unphosphorylated and D- complex phosphorylated. Energy bar is provided along with.

Per-residue RMSD of A-loop:

The Activation loop is generally stable, except in the case of the complex unphosphorylated system. Specifically, the residues from 217-222 in this system contribute to significant structural deviations, which are mitigated by phosphorylation at the Thr290 residue. In contrast, no notable changes are seen in the A-loop of apo systems (from **Figure 5.26**). Therefore, the process of complexation facilitates structural movements and conformational alterations in the A-loop, resulting in increased flexibility.

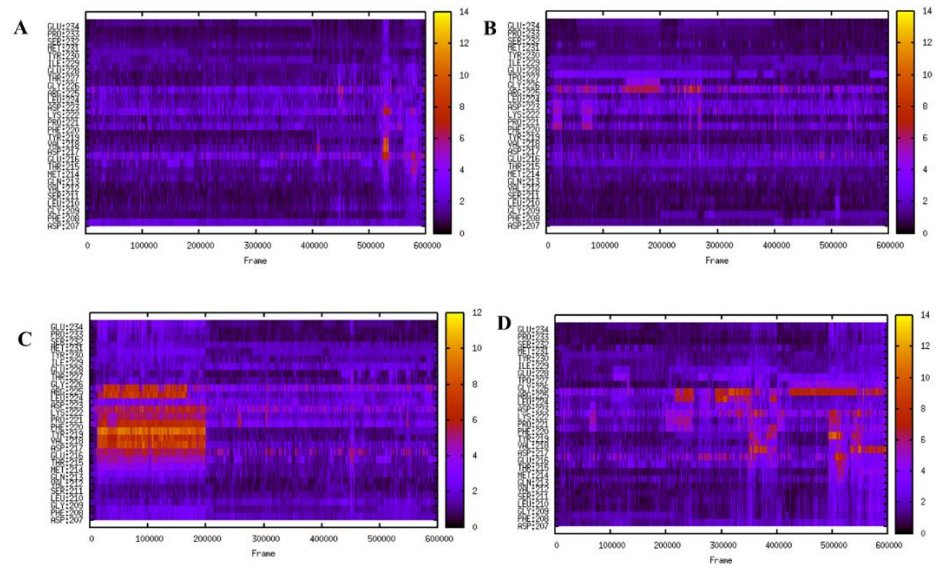


Figure 5.26: Per-residue analysis of A-loop of A- apo unphosphorylated, B- apo phosphorylated, C- complex unphosphorylated and D- complex phosphorylated. Energy bar is provided along with.

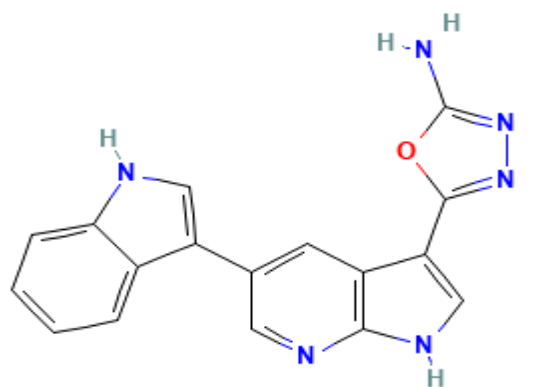
Objective 2: Identification of Novel Inhibitor(s) targeting the COT Kinase by Virtual Screening.

Table 6: Virtual screening and ADMET analysis of lead molecules.

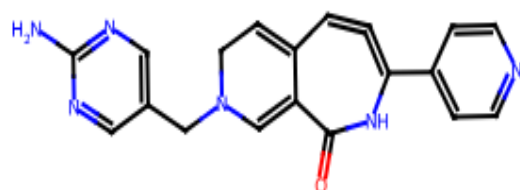
Drug	Molsoft score	Mol.wt	HBA	HBD	Log P	Toxicity
Control	-53.6	316.11	7	4	2.9	Eye irritant and hepatotoxic
Inhibitor	-56.02	347.11	8	3	-0.2	Non corrosive and non-cytotoxic.

Inhibitor SMILE:

[H]C1=C=C(c2c([H])c([H])nc([H])c2[H])N([H])C(C2C1=C([H])C([H]))([H])N(C=2[H])C([H])([H])c1c([H])nc(nc1[H])N([H])[H])=O



control



Proposed inhibitor

The inhibitor was docked with COT Kinase protein and simulation was run for 400 ns to check the overall stability.

Structural stability of ‘Control’ vs ‘Inhibitor’:

The time evolution plot in **Figure 5.27** shows that the inhibitor is stabilizing the kinase in a way much better than that of the control. The P-loop is showing high fluctuations in case of control, but it is highly stable when inhibitor is bound and RMSD is strictly 1.5 Å. For control, the Activation loop fluctuates from 150 ns, but the overall RMSD is maintained at 1 Å in case of inhibitor.

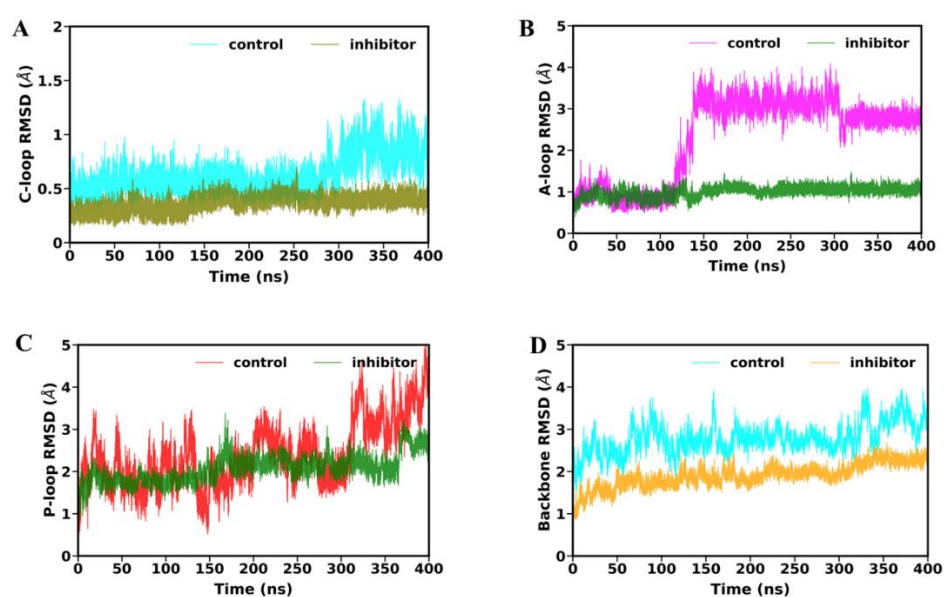


Figure 5.27: Time evolution graphs showing RMSD values of various regions of the molecule when control is bound to it vs inhibitor bound to it.

Objective 3: Differential study of the effect of Mn^{2+} vs Mg^{2+} on the Structural Dynamics of COT Kinase.

Literature studies suggest that when Mn is bound to protein, it lowers the K_m value of the molecule and hence it has higher affinity for ATP as compared to when Mg is bound.

We have studied the effect of both the cofactors on the structural stability of the complex.

The **Figure 5.28** shows that on Mn-ATP complexation, two RMSD are possible indicating the presence of two different conformations. Mg-ATP complex shows a broader peak in the range of 4 Å.

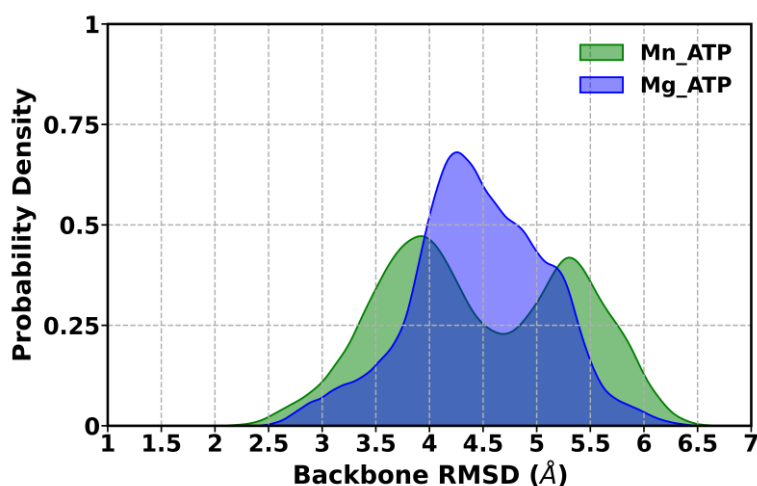


Figure 5.28: Probability distribution curve of backbone RMSD for both the complexes.

Table 7: The values of backbone RMSD, SASA and Radius of Gyration calculated for the complexes using Block Average method. SD is given in parentheses.

System	Backbone RMSD (in Å)	RoG (in Å)	SASA (in nm ²)
Mn-ATP	3.97 ± (0.68)	21.17 ± (0.32)	168.49 ± (3.6)
Mg-ATP	3.45 ± (0.55)	21.14 ± (0.18)	169.3 ± (2.47)

No such significant changes are observed in case of protein compactness and solvent accessibility.

The **Figure 5.29** shows that P-loop is too much flexible in case of Mg-ATP system, but it becomes much stabler when Mn is bound. The Catalytic loop and A-loop are not too flexible for the systems. C-terminus region shows flexibility in Mn-ATP system.

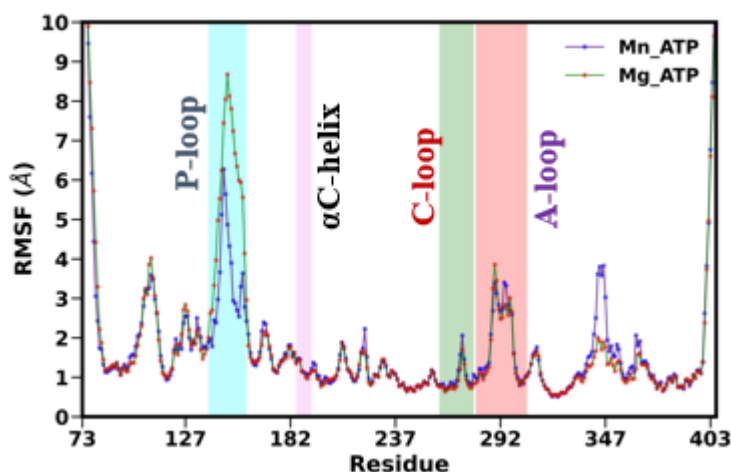


Figure 5.29: Residual fluctuation for the systems.

Too many wide peaks are observed especially in P-loop. Mn-ATP stabilizes it. It is due to conformational changes. Three peaks are seen for A-loop in both the systems, indicating the presence of three different structures that might have formed during simulation.

Phosphate binding loop stability:

The P-loop is extremely flexible for the Mg-ATP system and on Mn-ATP complex formation, the loop's flexibility lessens. Multiple peaks are observed for Mg-ATP system, indicating the possibility of too many conformations as evident from **Figure 5.30**.

The time evolution plot unveils that Mg-ATP shows high deviations, especially in the run2 and all the runs are converged for the other system.

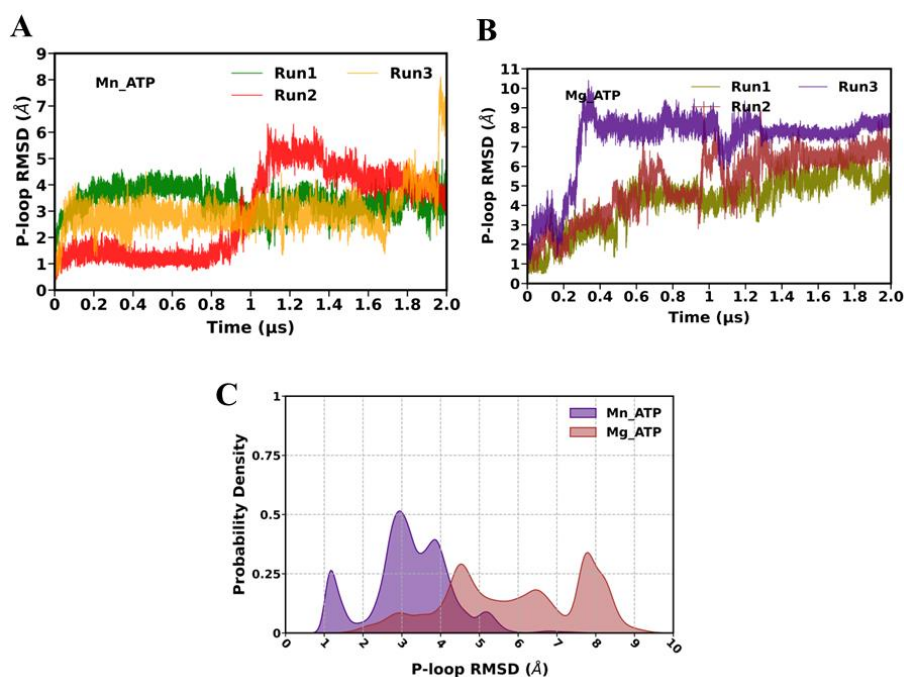


Figure 5.30: Time evolution plot and KDE plot of P-loop RMSD of both the systems.

Activation loop stability:

A-loop or the Activation loop is stabilized at around 4 Å for both the systems. Three sharp peaks are observed in case of Mg-ATP system, indicating the presence of three different conformations having varying RMSD values.

The **Figure 5.31** exhibits two sharp peaks in case of Mn-ATP system, pointing to the fact that there are two possible structures and they are less deviating than Mg-ATP system.

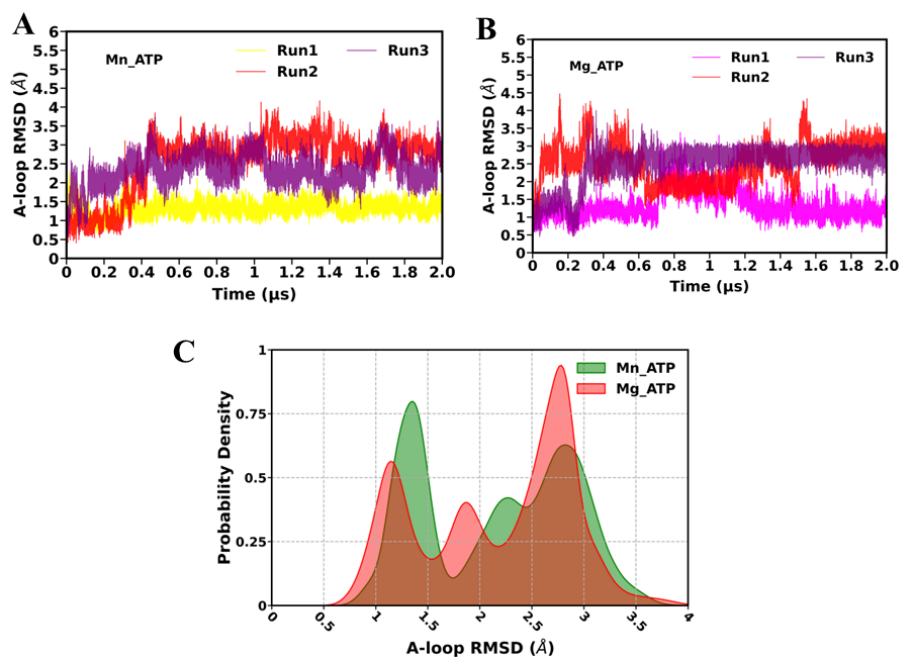


Figure 5.31: Time evolution plot and KDE plot of A-loop RMSD of both the systems.

Hydrogen bond analysis:

H bond analysis is a very important study used to analyse ligand binding strength to the protein. The **Figure 5.32** shows time series plot for number of hydrogen bonds formed by Mn-ATP complex and Mg-ATP complex. The Mg-ATP complex forms around 2 H-bonds and Mn-ATP forms about 4 H-bonds.

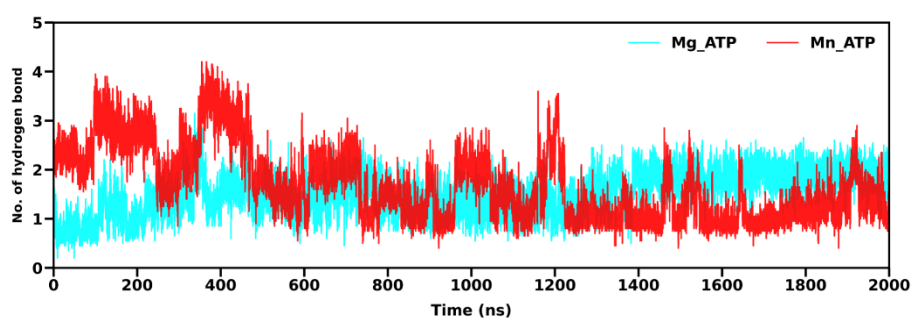


Figure 5.32: Time series plot for Number of Hydrogen bonds formed by the complexes.

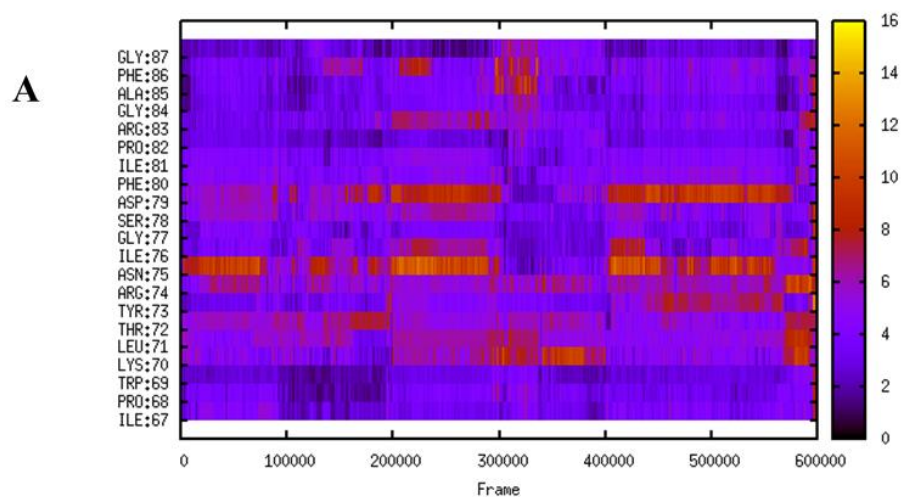
Table 8: Hydrogen bond analysis for two complexes.

Binding couple		Molecular dynamics	
Acceptor	Donor	Occupancy (%)	Average Distance (Å)
Mg-ATP complex			
ASN_195@OD1	ATP_333@O3'	14.41%	2.70 Å
GLU_154@OE1	ATP_333@N6	10.28%	2.76 Å
Mn-ATP complex			
ATP333@O2A	SER_194@OG	44.72%	2.69 Å
ATP333@O1A	SER_194@OG	29.43%	2.69 Å
SER_194@O	ATP_333@O3'	25.84%	2.67 Å
SER_194@O	ATP_333@O2'	13.70%	2.77 Å

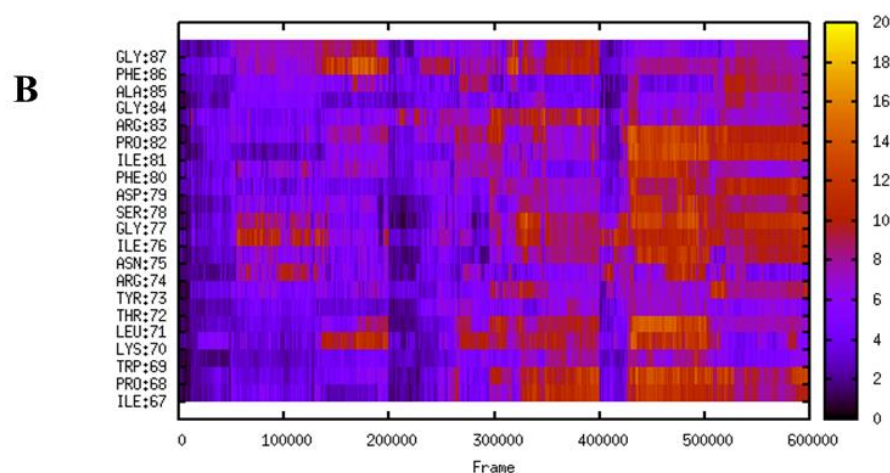
The table shows that for Mn-ATP system, Ser 194 contributes maximum to the bonding and for Mg-ATP system, contribution of Asn 195 is maximum.

Per-residue RMSD of P-loop:

The per-residue RMSD for the P-loop of both Mn and Mg systems are calculated. It can be observed that Mn-ATP shows lower structural deviations as compared to that of Mg-ATP system. The **Figure 5.33** shows that the residues like Asp 79, Asn 75, Arg 74 contributes to maximum deviation in both the systems. Various residues like Pro 82, Arg 83, Ser 78, Leu 71, Pro 68 of Mg-ATP system contributed to maximum flexibility and it is being diminished on complexation with Mn-ATP.



Mn-ATP complex

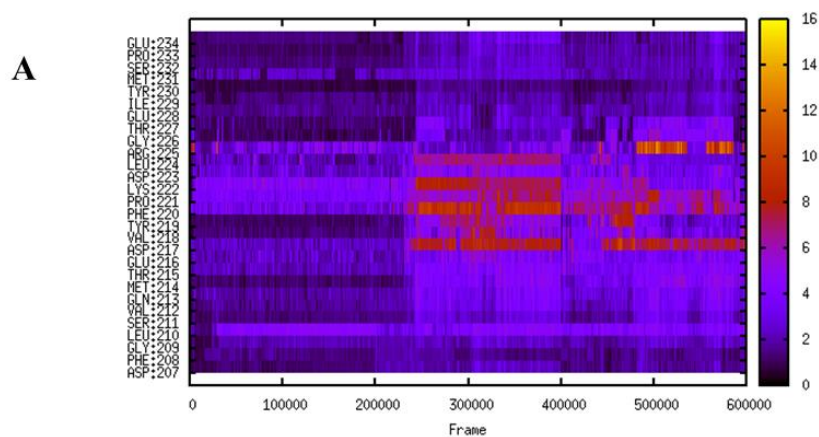


Mg-ATP complex

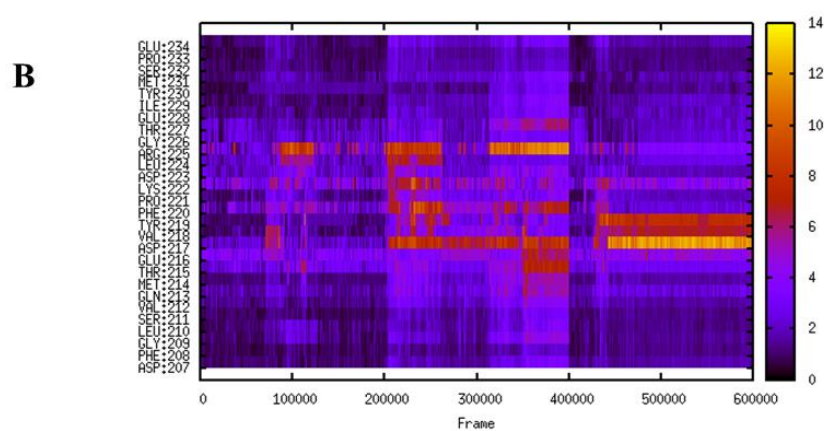
Figure 5.33: Per-residue RMSD of P-loop for Mg-ATP system and Mn-ATP system.

Per-residue RMSD of A-loop:

The Activation loop of Mg-ATP system is slightly more flexible than that of Mn-ATP. The Gly 226, Val 218 shows highly RMSD values in case of Mg-ATP, that is diminished in case of Mn-ATP system as shown in **Figure 5.34**.



Mn-ATP complex



Mg-ATP complex

Figure 5.34: Per-residue RMSD of A-loop for Mg-ATP system and Mn-ATP system.

CONCLUSION & FUTURE PROSPECTS

The thesis dealt with the structural dynamics of a unique kinase, the COT Kinase. It unravels the effect of phosphorylation and complexation. Another aspect is studied: the use of cofactors, Mg vs Mn. On complexation, the P-loop is becoming stable and when Mn is used as a cofactor, the same phenomenon is observed. Catalytic loop is remaining stable overall. The parameters like SASA, RoG remained similar for the apo and complex systems respectively. Very high fluctuations are observed for the P-loop and similar pattern is seen for the C-terminal region of the kinase. Few amino acid residues are contributing to high RMSD values for the systems and it is evident from the graphs. The principal component analysis revealed the presence of global minima for the systems, from which the structures can be extracted. The conformations are interconvertible due to low energy barrier between them. Complex formation reduced the anti-correlated motions for the systems.

Further, the study revolved around finding a potential drug candidate against the Kinase. The control molecule against the protein is “Imidazonaphthyridine” and through Virtual Screening, an inhibitor is found and studied further. The inhibitor is stabilizing the molecule and extensive studies are required to prove its efficacy. The simulation results portrayed that the inhibitor is highly capable of preventing P-loop fluctuations and also it stabilized the overall movements of the loop regions.

No change in protein compactness is found when protein has Mg cofactor versus Mn cofactor. It also reflected no alteration in the solvent accessibility. However, Mn stabilized the flexible P-loop and on per-residue RMSD analysis, similar trend is observed. However, more studies are required on the same.

The study has opened many future paths like the mutational aspect, discovery of drugs that would selectively bind to COT Kinase and not others, role of phytochemicals and natural products to treat

diseases caused due to overexpression of this molecule. Further, the torsion angle analysis, binding free calculation of the inhibitor, essential dynamic studies need to be done to support the claim that inhibitor is better than our control. Moreover, genomics can be included to give another dimension to the study and find out how the drugs would influence a person's health based on his genetic profile.

REFERENCES

- (1) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The Protein Kinase Complement of the Human Genome. *Science* **2002**, *298* (5600), 1912–1934. <https://doi.org/10.1126/science.1075762>.
- (2) Laufkötter, O.; Hu, H.; Miljković, F.; Bajorath, J. Structure- and Similarity-Based Survey of Allosteric Kinase Inhibitors, Activators, and Closely Related Compounds. *J. Med. Chem.* **2022**, *65* (2), 922–934. <https://doi.org/10.1021/acs.jmedchem.0c02076>.
- (3) Andreotti, A. H.; Dötsch, V. Allosteric Regulation of Kinase Activity. *eLife* **2024**, *13*, e97084. <https://doi.org/10.7554/eLife.97084>.
- (4) Aslam, M.; Ladilov, Y. Emerging Role of cAMP/AMPK Signaling. *Cells* **2022**, *11* (2), 308. <https://doi.org/10.3390/cells11020308>.
- (5) Sophocleous, G.; Owen, D.; Mott, H. R. The Structure and Function of Protein Kinase C-Related Kinases (PRKs). *Biochemical Society Transactions* **2021**, *49* (1), 217–235. <https://doi.org/10.1042/BST20200466>.
- (6) George, D.; Salmeron, A. Cot/Tpl-2 Protein Kinase as a Target for the Treatment of Inflammatory Disease. *CTMC* **2009**, *9* (7), 611–622. <https://doi.org/10.2174/156802609789007345>.
- (7) Glatthar, R.; Stojanovic, A.; Troxler, T.; Mattes, H.; Möbitz, H.; Beerli, R.; Blanz, J.; Gassmann, E.; Drückes, P.; Fendrich, G.; Gutmann, S.; Martiny-Baron, G.; Spence, F.; Hornfeld, J.; Peel, J. E.; Sparrer, H. Discovery of Imidazoquinolines as a Novel Class of Potent, Selective, and in Vivo Efficacious Cancer Osaka Thyroid (COT) Kinase Inhibitors. *J. Med. Chem.* **2016**, *59* (16), 7544–7560. <https://doi.org/10.1021/acs.jmedchem.6b00598>.
- (8) Nakae, S.; Kitamura, M.; Fujiwara, D.; Sawa, M.; Shirai, T.; Fujii, I.; Tada, T. Structure of Mitogen-Activated Protein Kinase Kinase 1 in the DFG-out Conformation. *Acta Crystallogr F Struct Biol Commun* **2021**, *77* (12), 459–464. <https://doi.org/10.1107/S2053230X21011687>.

- (9) Babaei, G.; Aziz, S. G.-G.; Jaghi, N. Z. Z. EMT, Cancer Stem Cells and Autophagy; The Three Main Axes of Metastasis. *Biomedicine & Pharmacotherapy* **2021**, *133*, 110909. <https://doi.org/10.1016/j.biopha.2020.110909>.
- (10) Vougioukalaki, M.; Kanellis, D. C.; Gkouskou, K.; Eliopoulos, A. G. Tpl2 Kinase Signal Transduction in Inflammation and Cancer. *Cancer Letters* **2011**, *304* (2), 80–89. <https://doi.org/10.1016/j.canlet.2011.02.004>.
- (11) Roskoski, R. Combination Immune Checkpoint and Targeted Protein Kinase Inhibitors for the Treatment of Renal Cell Carcinomas. *Pharmacological Research* **2024**, *203*, 107181. <https://doi.org/10.1016/j.phrs.2024.107181>.
- (12) Galons, H. Introduction for the Special Issue on Kinase Inhibitors. *European Journal of Medicinal Chemistry* **2013**, *61*, 1. <https://doi.org/10.1016/j.ejmech.2012.12.001>.
- (13) Xu, D.; Matsumoto, M. L.; McKenzie, B. S.; Zarrin, A. A. TPL2 Kinase Action and Control of Inflammation. *Pharmacological Research* **2018**, *129*, 188–193. <https://doi.org/10.1016/j.phrs.2017.11.031>.
- (14) Sk, M. F.; Jonniya, N. A.; Roy, R.; Kar, P. Unraveling the Molecular Mechanism of Recognition of Selected Next-Generation Antirheumatoid Arthritis Inhibitors by Janus Kinase 1. *ACS Omega* **2022**, *7* (7), 6195–6209. <https://doi.org/10.1021/acsomega.1c06715>.
- (15) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J Comput Chem* **2011**, *32* (10), 2319–2327. <https://doi.org/10.1002/jcc.21787>.
- (16) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J Comput Chem* **2005**, *26* (16), 1668–1688. <https://doi.org/10.1002/jcc.20290>.
- (17) Rozmanov, D.; Kusalik, P. G. Robust Rotational-Velocity-Verlet Integration Methods. *Phys. Rev. E* **2010**, *81* (5), 056706. <https://doi.org/10.1103/PhysRevE.81.056706>.

- (18) González, M. A. Force Fields and Molecular Dynamics Simulations. *JDN* **2011**, *12*, 169–200. <https://doi.org/10.1051/sfn/201112009>.
- (19) Loschwitz, J.; Jäckering, A.; Keutmann, M.; Olagunju, M.; Olubiyi, O. O.; Strodel, B. Dataset of AMBER Force Field Parameters of Drugs, Natural Products and Steroids for Simulations Using GROMACS. *Data in Brief* **2021**, *35*, 106948. <https://doi.org/10.1016/j.dib.2021.106948>.
- (20) Raguette, L. E.; Cuomo, A. E.; Belfon, K. A. A.; Tian, C.; Hazoglou, V.; Witek, G.; Telehany, S. M.; Wu, Q.; Simmerling, C. phosaa14SB and phosaa19SB: Updated Amber Force Field Parameters for Phosphorylated Amino Acids. *J. Chem. Theory Comput.* **2024**, *acs.jctc.4c00732*. <https://doi.org/10.1021/acs.jctc.4c00732>.
- (21) Dobson, M. Periodic Boundary Conditions for Long-Time Nonequilibrium Molecular Dynamics Simulations of Incompressible Flows. *The Journal of Chemical Physics* **2014**, *141* (18), 184103. <https://doi.org/10.1063/1.4901276>.
- (22) Simmonett, A. C.; Brooks, B. R. Analytical Hessians for Ewald and Particle Mesh Ewald Electrostatics. *The Journal of Chemical Physics* **2021**, *154* (10), 104101. <https://doi.org/10.1063/5.0044166>.
- (23) Davidchack, R. L.; Handel, R.; Tretyakov, M. V. Langevin Thermostat for Rigid Body Dynamics. *The Journal of Chemical Physics* **2009**, *130* (23), 234101. <https://doi.org/10.1063/1.3149788>.
- (24) Janek, J.; Kolafa, J. Novel Barostat Implementation for Molecular Dynamics. *The Journal of Chemical Physics* **2024**, *160* (18), 184111. <https://doi.org/10.1063/5.0193281>.
- (25) Ando, T. Shear Viscosity of OPC and OPC3 Water Models. *The Journal of Chemical Physics* **2023**, *159* (10), 101102. <https://doi.org/10.1063/5.0161476>.
- (26) Khuntawee, W.; Kunaseth, M.; Rungnim, C.; Intagorn, S.; Wolschann, P.; Kungwan, N.; Rungrotmongkol, T.; Hannongbua, S. Comparison of Implicit and Explicit Solvation Models for *Iota* - Cyclodextrin Conformation Analysis from Replica Exchange

- Molecular Dynamics. *J. Chem. Inf. Model.* **2017**, *57* (4), 778–786. <https://doi.org/10.1021/acs.jcim.6b00595>.
- (27) Linse, J.-B.; Hub, J. S. Three- and Four-Site Models for Heavy Water: SPC/E-HW, TIP3P-HW, and TIP4P/2005-HW. *The Journal of Chemical Physics* **2021**, *154* (19), 194501. <https://doi.org/10.1063/5.0050841>.
- (28) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11* (8), 3584–3595. <https://doi.org/10.1021/acs.jctc.5b00436>.
- (29) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J Comput Chem* **2004**, *25* (13), 1605–1612. <https://doi.org/10.1002/jcc.20084>.
- (30) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J Comput Chem* **2004**, *25* (13), 1605–1612. <https://doi.org/10.1002/jcc.20084>.
- (31) Krief, M.; Ashkenazy, Y. Calculation of Elastic Constants of Embedded-Atom-Model Potentials in the N V T Ensemble. *Phys. Rev. E* **2021**, *103* (6), 063307. <https://doi.org/10.1103/PhysRevE.103.063307>.
- (32) Ringnér, M. What Is Principal Component Analysis? *Nat Biotechnol* **2008**, *26* (3), 303–304. <https://doi.org/10.1038/nbt0308-303>.
- (33) Karplus, M.; Petsko, G. A. Molecular Dynamics Simulations in Biology. *Nature* **1990**, *347* (6294), 631–639. <https://doi.org/10.1038/347631a0>.
- (34) Moulick, A. G.; Chakrabarti, J. Conformational Fluctuations in the Molten Globule State of α -Lactalbumin. *Phys. Chem. Chem. Phys.* **2022**, *24* (35), 21348–21357. <https://doi.org/10.1039/D2CP02168D>.

- (35) Wardah, W.; Khan, M. G. M.; Sharma, A.; Rashid, M. A. Protein Secondary Structure Prediction Using Neural Networks and Deep Learning: A Review. *Computational Biology and Chemistry* **2019**, *81*, 1–8. <https://doi.org/10.1016/j.compbiolchem.2019.107093>.
- (36) Akkus, E.; Tayfuroglu, O.; Yildiz, M.; Kocak, A. Revisiting MMPBSA by Adoption of MC-Based Surface Area/Volume, ANI-ML Potentials, and Two-Valued Interior Dielectric Constant. *J. Phys. Chem. B* **2023**, *127* (20), 4415–4429. <https://doi.org/10.1021/acs.jpcb.3c00834>.
- (37) Dykeman, E. C.; Sankey, O. F. Normal Mode Analysis and Applications in Biological Physics. *J. Phys.: Condens. Matter* **2010**, *22* (42), 423202. <https://doi.org/10.1088/0953-8984/22/42/423202>.
- (38) Jordan, M. I.; Mitchell, T. M. Machine Learning: Trends, Perspectives, and Prospects. *Science* **2015**, *349* (6245), 255–260. <https://doi.org/10.1126/science.aaa8415>.
- (39) Sinaga, K. P.; Yang, M.-S. Unsupervised K-Means Clustering Algorithm. *IEEE Access* **2020**, *8*, 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>.
- (40) Kuraria, A.; Jharbade, N.; Soni, M. Centroid Selection Process Using WCSS and Elbow Method for K-Mean Clustering Algorithm in Data Mining. *IJSRSET* **2018**, 190–195. <https://doi.org/10.32628/IJSRSET21841122>.
- (41) Banerjee, P.; Kemmler, E.; Dunkel, M.; Preissner, R. ProTox 3.0: A Webserver for the Prediction of Toxicity of Chemicals. *Nucleic Acids Research* **2024**, *52* (W1), W513–W520. <https://doi.org/10.1093/nar/gkae303>.
- (42) Swanson, K.; Walther, P.; Leitz, J.; Mukherjee, S.; Wu, J. C.; Shivnaraine, R. V.; Zou, J. ADMET-AI: A Machine Learning ADMET Platform for Evaluation of Large-Scale Chemical Libraries. *Bioinformatics* **2024**, *40* (7), btae416. <https://doi.org/10.1093/bioinformatics/btae416>.
- (43) Daina, A.; Michielin, O.; Zoete, V. SwissADME: A Free Web Tool to Evaluate Pharmacokinetics, Drug-Likeness and Medicinal

- Chemistry Friendliness of Small Molecules. *Sci Rep* **2017**, 7 (1), 42717. <https://doi.org/10.1038/srep42717>.
- (44) Nascimento, I. J. D. S.; De Aquino, T. M.; Da Silva-Júnior, E. F. The New Era of Drug Discovery: The Power of Computer-Aided DrugDesign (CADD). *LDDD* **2022**, 19 (11), 951–955. <https://doi.org/10.2174/1570180819666220405225817>.
- (45) Zhang, X.; Shen, C.; Zhang, H.; Kang, Y.; Hsieh, C.-Y.; Hou, T. Advancing Ligand Docking through Deep Learning: Challenges and Prospects in Virtual Screening. *Acc. Chem. Res.* **2024**, 57 (10), 1500–1509. <https://doi.org/10.1021/acs.accounts.4c00093>.
- (46) Tang, S.; Ding, J.; Zhu, X.; Wang, Z.; Zhao, H.; Wu, J. Vina-GPU 2.1: Towards Further Optimizing Docking Speed and Precision of AutoDock Vina and Its Derivatives. *IEEE/ACM Trans. Comput. Biol. and Bioinf.* **2024**, 21 (6), 2382–2393. <https://doi.org/10.1109/TCBB.2024.3467127>.
- (47) Sarkar, A.; Concilio, S.; Sessa, L.; Marrafino, F.; Piotto, S. Advancements and Novel Approaches in Modified AutoDock Vina Algorithms for Enhanced Molecular Docking. *Results in Chemistry* **2024**, 7, 101319. <https://doi.org/10.1016/j.rechem.2024.101319>.
- (48) Alberga, D.; Lamanna, G.; Graziano, G.; Delre, P.; Lomuscio, M. C.; Corriero, N.; Ligresti, A.; Siliqi, D.; Saviano, M.; Contino, M.; Stefanachi, A.; Mangiatordi, G. F. DeLA-DrugSelf: Empowering Multi-Objective de Novo Design through SELFIES Molecular Representation. *Computers in Biology and Medicine* **2024**, 175, 108486. <https://doi.org/10.1016/j.compbio.2024.108486>.
- (49) Schneuing, A.; Harris, C.; Du, Y.; Didi, K.; Jamasb, A.; Igashov, I.; Du, W.; Gomes, C.; Blundell, T. L.; Lio, P.; Welling, M.; Bronstein, M.; Correia, B. Structure-Based Drug Design with Equivariant Diffusion Models. *Nat Comput Sci* **2024**, 4 (12), 899–909. <https://doi.org/10.1038/s43588-024-00737-x>.
- (50) Han, H.; Shaker, B.; Lee, J. H.; Choi, S.; Yoon, S.; Singh, M.; Basith, S.; Cui, M.; Ahn, S.; An, J.; Kang, S.; Yeom, M. S.; Choi, S. Employing Automated Machine Learning (AutoML) Methods to

- Facilitate the In Silico ADMET Properties Prediction. *J. Chem. Inf. Model.* **2025**, *65* (7), 3215–3225. <https://doi.org/10.1021/acs.jcim.4c02122>.
- (51) *2-Deoxy-D-Glucose: Chemistry and Biology*; Singh, R., Kumar, A., Singh, K., Eds.; BENTHAM SCIENCE PUBLISHERS, 2024. <https://doi.org/10.2174/97898153051591240101>.
- (52) Akar, I.; Foster, J. C.; Leng, X.; Pearce, A. K.; Mathers, R. T.; O'Reilly, R. K. Log Poct/SA Predicts the Thermoresponsive Behavior of P(DMA-Co-RA) Statistical Copolymers. *ACS Macro Lett.* **2022**, *11* (4), 498–503. <https://doi.org/10.1021/acsmacrolett.1c00776>.
- (53) Jung, W.; Goo, S.; Hwang, T.; Lee, H.; Kim, Y.-K.; Chae, J.; Yun, H.; Jung, S. Absorption Distribution Metabolism Excretion and Toxicity Property Prediction Utilizing a Pre-Trained Natural Language Processing Model and Its Applications in Early-Stage Drug Development. *Pharmaceuticals* **2024**, *17* (3), 382. <https://doi.org/10.3390/ph17030382>.
- (54) Raghavendra, N. M.; Kumar, B. R. P.; Sasmal, P.; Teli, G.; Pal, R.; Gurubasavaraja Swamy, P. M.; Sajeew Kumar, B. Designing Studies in Pharmaceutical and Medicinal Chemistry. In *The Quintessence of Basic and Clinical Research and Scientific Publishing*; Jagadeesh, G., Balakumar, P., Senatore, F., Eds.; Springer Nature Singapore: Singapore, 2023; pp 125–148. https://doi.org/10.1007/978-981-99-1284-1_9.
- (55) Jocys, Z.; Grundy, J.; Farrahi, K. DrugPose: Benchmarking 3D Generative Methods for Early Stage Drug Discovery. *Digital Discovery* **2024**, *3* (7), 1308–1318. <https://doi.org/10.1039/D4DD00076E>.