

**COMPUTATIONAL INVESTIGATION AND MODULATION OF
STRUCTURAL AND FUNCTIONAL PROPERTIES OF PROTEINS FOR
THERAPEUTIC PURPOSES**

A DISSERTATION SUBMITTED TO
THE GRADUATE SCHOOL OF
ENGINEERING AND NATURAL SCIENCES
OF ISTANBUL MEDIPOL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
BIOMEDICAL ENGINEERING AND BIOINFORMATICS

By

Samman Mansoor

August, 2021

COMPUTATIONAL INVESTIGATION AND MODULATION OF STRUCTURAL
AND FUNCTIONAL PROPERTIES OF PROTEINS FOR THERAPEUTIC PURPOSES

By Samman Mansoor

27 August 2021

We certify that we have read this dissertation and that in our opinion it is fully adequate,
in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Assist. Prof. Dr. Özge Şensoy (Advisor)

Assist. Prof. Dr. Kivanç Kök

Assist. Prof. Dr. Mehmet Hikmet Üçışık


Assoc. Prof. Dr. Wael M. Rabeh

Prof. Dr. Serdar Durdağı

Approved by the Graduate School of Engineering and Natural Sciences:

Prof. Dr. Yasemin Yüksel Durmaz

Director of the Graduate School of Engineering and Natural Sciences



I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Signature :

Name, Surname: SAMMAN MANSOOR

ACKNOWLEDGEMENT

The most important acknowledgment, I would like to express to my advisor Assist. Prof. Dr. Özge Şensoy for her great support, guidance and encouragement which motivated me to pursue my research throughout the tenure of my Ph.D studies. Her wisdom, patience and valuable experience has helped me to learn a lot and progress in my research. She has been my inspiration from the beginning and these four years I have spent under her supervision were not only productive but very delightful as well. Her help in both personal and professional growth has been of tremendous wealth for me. Her friendly nature has helped me to solve many hurdles in my research in a relaxed way.

I would like to extend my sincere thanks to the committee members of my Ph.D. thesis: Assist. Prof. Dr. Mehmet Hikmet Üçişik and Assist. Prof. Dr. Kivanç Kök for their constructive comments and valuable suggestions on my research works and thesis improvement. I would like to thank Prof. Dr. Serdar Durdağı and Assoc. Prof. Dr. Mustafa Güzel for their valuable collaboration with our research group.

I would like to mention my gratitude to Assoc. Prof. Dr. Wael M. Rabeh and Dr. Yasemin Yozgat for their kindness and very productive research collaboration.

I would like to thank Graduate School of Engineering and Natural Sciences and Medipol University for their administrative and financial support throughout my stay in Medipol. And Prof. Dr. Yasemin Yüksel Durmaz for her kind guidance and help in administrative work. I would like to thank TÜBİTAK (Technological Research Council of Turkey) for funding my doctoral research.

In the end I would like to appreciate the support and love I have received from my family and all of my friends to reach the end of this difficult milestone of my Ph.D studies. I have a great appreciation for all my former and current lab fellows. With their help I have learnt a lot. In the end I would like to give my Special Thanks to my “Mother and Brothers” for their true and never ending love and support.

Samman Mansoor
August, 2021

CONTENTS

	<u>Page</u>
ACKNOWLEDGEMENT	iv
CONTENTS	v
LIST OF FIGURES	vii
LIST OF SYMBOLS	x
ABBREVIATIONS	xi
ÖZET	xii
ABSTRACT	xiv
1. INTRODUCTION	1
1.1. Motivation and Thesis contribution.....	2
1.2. Literature Review.....	3
1.2.1. G protein-coupled receptors	3
1.2.2. Experimental techniques used to capture GPCR conformational states	7
1.2.3. Bivalent ligands	8
1.2.4. Heterotrimeric G protein.....	10
1.2.5. Hexokinase enzyme	12
1.2.6. Why proliferating cells metabolize glucose by aerobic glycolysis? ...	13
1.2.7. Hexokinase enzyme and cancer	14
2. THEORETICAL PART	16
2.1. Molecular Simulation.....	19
2.2. Molecular Dynamics Simulation.....	19
2.3. Accelerated Molecular Dynamics Simulation.....	20
2.4. G Protein intermediate conformational states identification.....	22
2.5. Phi-Value Analysis technique concept and MD simulation	23
3. EXPERIMENTAL PART	25
3.1. System 1 methodology workflow.....	25
3.1.1. Tetramer modeling	25
3.1.2. System preparation.....	28
3.1.3. Bivalent ligand construction	28
3.1.4. Molecular dynamics simulation	29
3.1.5. Root mean square fluctuation (RMSF).....	30
3.1.6. Cross-correlation analysis	31
3.1.7. Network construction	31
3.1.8. Binding volume calculation.....	31
3.2. System 2.....	32
3.2.1. Identification of intermediate active states from experimental phi-values	32
3.2.2. Implementation of experimental phi-values in MD simulations	33
3.2.3. Methodology workflow	33
3.2.4. Structure preparation.....	34
3.2.5. Simulation protocol	35
3.3. System 3 methodology workflow.....	35
3.3.1. Structure preparation.....	36
3.3.2. MD simulation setup	36
3.3.3. Analysis of trajectories.....	36

3.3.4.	Hexokinase 2 virtual screening	37
4.	RESULTS AND DISCUSSION	39
4.1.	Heterobivalent ligand targeting A _{2A} R dimer	39
4.1.1.	Reweighting of accelerated molecular dynamics simulation.....	39
4.1.2.	Optimum linker length is required for stable binding of bivalent ligand to A _{2A} R dimer.....	41
4.1.3.	Extra- and intracellular domains are stabilized and global dynamics is restricted by the bivalent ligand in A _{2A} R dimer	43
4.1.4.	Bivalent ligand modulates volume of ligand binding pocket and intracellular domain in A _{2A} R dimer	43
4.1.5.	Conformational preferences of microswitches and residue correlations are modulated by the bivalent ligand in A _{2A} R dimer.....	45
4.1.6.	D ₂ R dimer resembles an asymmetric unit in the presence of the linker	48
4.2.	Experimental phi-values and G protein system results	50
4.2.1.	Allosteric network mediating GDP dissociation upon receptor activation	53
4.2.2.	Comparison of our results with G protein activation determined reaction coordinates.....	54
4.2.3.	Dynamics of switch-II	55
4.3.	Hexokinase 2.....	57
4.3.1.	Principle component analysis RMSF for wild-type and mutant apo HK2	58
4.3.2.	Conformational stability comparison between wild-type and mutant .	60
4.3.3.	HK2 virtual screening results.....	60
5.	CONCLUSION AND FUTURE WORK.....	68
	BIBLIOGRAPHY.....	73
	CURRICULUM VITAE.....	88

LIST OF FIGURES

Figure 1.1: Schematic representation of GPCR oligomerization.....	4
Figure 1.2: A. depicts that ligand binding affinity of monomer (purple) increases in the presence of its partner protomer (orange); B. ligand efficacy of monomer (purple) decreases (middle) or increases (right) in the presence of its partner (orange) when they form dimer; C. functional selectivity of dimer changes in the presence of protomer; concurrently, G protein is substituted with β arrestin [23].....	6
Figure 1.3: Schematic representation of a key structural features a bivalent ligand, where the pharmacophores can be either the same or different on both sides of the ligand.....	9
Figure 1.4: Schematic representation of an agonist induced GPCR activation along G protein recruitment in step 1 followed by GDP to GTP exchange by binding and opening of α subunit in step 2. In step 3 subunits dissociate to initiate their corresponding downstream signaling.....	10
Figure 1.5: Energy production in normal vs tumor cells by Warburg's effect.....	12
Figure 1.6: The dimeric structure of Hexokinase has been shown.....	15
Figure 3.1: Methodology Workflow of first system.....	25
Figure 3.2: Schematic methodology workflow of system 2.....	33
Figure 3.3: Melting temperatures of GDP and GTP bound systems with calculated phi-values colored green are given as an example of the proposed strategy.....	34
Figure 3.4: Schematic methodology workflow of system 3.....	35
Figure 4.1: PMF profiles of the χ_2 Trp246 obtained from reweighting based on cumulant expansion to the 2 nd order (blue) and calculated from classical MD simulation (orange) for comparison.....	40
Figure 4.2: A. Schematic representation of A _{2A} R/D ₂ R tetramer with bivalent ligand. The membrane is shown in blue. A _{2A} R protomers and mini-G α_s are shown in orange, purple and brown, respectively and new cartoon representation whereas bivalent ligand is shown in green and licorice representation. D ₂ R protomers and mini-G α_i are shown in pink, cyan, and red, respectively. B. The pharmacophore groups and the linker are shown in detail together with A _{2A} R dimer.....	41
Figure 4.3: (RMSD) timeline plots. A. Tetramer. B. Agonist bound D ₂ R. C. Apo D ₂ R. D. Antagonist bound A _{2A} R. E. Agonist bound A _{2A} R.....	42
Figure 4.4: RMSF and 2D-(PCA) profiles pertaining to A. antagonist-bound A _{2A} R and B. agonist-bound A _{2A} R/mini-G α_s . The regions that showed difference between without and with linker systems are indicated with green rectangles.....	44
Figure 4.5: Changes in volumes of A. ligand binding pocket and B. G protein binding domain between without and linker systems C. Schematic representation of conformational changes adapted by A _{2A} R receptors in the presence of the linker. TM 6 and 7 positions in the absence (green) and presence (orange) of the linker is shown. Mini-G α_s bound to agonis-bound A _{2A} R is shown in new cartoon representation. Right hand side Ile274 ^{7,39} interaction with ligand emerge upon adding the linker. D. Protein-protein interaction energies measured between G protein and the receptors are also given.....	45
Figure 4.6: Dynamic cross correlation maps (DCCM) A. agonist-bound A _{2A} R/mini-G α_s and B. antagonist-bound A _{2A} R. The upper triangles in A and B correspond to DCCM of the system with linker whereas the lower triangles correspond to those without linker.....	46
Figure 4.7: Angle probability distribution of Trp246 and Tyr288. A. χ_2 of Trp246 ^{6,48} is presented for antagonist-bound A _{2A} R in the absence (black) and presence (red) of the linker, respectively. Dashed line indicate the reference value from crystal	

structure of antagonist-bound A_{2A}R (PDB ID:4EIY). **B.** The same as in A but for χ_2 of Tyr288^{7.53}. Comparison of orientation of **C.** Trp246^{6.48} in the absence (green) and presence (orange) of the bivalent ligand. Orientations of Tyr288^{7.53} in the **D.** absence of linker (dark and light green) and **E.** presence of linker (orange). The orientation of Trp246^{6.48} and Tyr288^{7.53} in the crystal structure of antagonist-bound receptor is shown in pink.....47

Figure 4.8: Allosteric interaction network pathway A_{2A}R.....48

Figure 4.9: Cholesterol binding residue shown for Agonist bound A_{2A}R.....48

Figure 4.10: Internal Water Channel formation. **A.** Agonist-bound A_{2A}R, and **B.** Antagonist-bound A_{2A}R in the absence of the linker, **C.** Agonist-bound A_{2A}R, and **D.** Antagonistbound A_{2A}R in the presence of the linker. Water occupancy maps are computed using the VolMap tool of VMD.....49

Figure 4.11: (RMSF) and 2D-PCA profiles pertaining to **A.** apo D₂R and **B.** agonist-bound D₂R/mini-G α_i . The regions that showed difference between absence and presence of the linker are indicated with green rectangles. **C.** Timeline ionic lock distance sampled by apo D₂R in the presence (red) and absence (black) of the linker. The ionic lock distance measured in the crystal structure of eticlopride-bound D₂R (PDB ID:3PBL) is indicated with blue dash line.....50

Figure 4.12: Dynamic cross correlation maps (DCCM) are presented for **A.** agonist-bound D₂R/mini-G α_i and **B.** apo D₂R. The upper diagonal in A and B correspond to DCCM of the system with linker whereas the lower diagonal correspond to those in the absence of linker.....51

Figure 4.13: The schematic representation of GPCR bound to its trimeric G protein has been shown in left side and the structure of G α subunit studied is given on right hand side.....51

Figure 4.14: Phi-values are plotted and the red arrows are given to the range of values between 0 to 1, where residue number is given on x-axis and phi-value on y-axis. To the right hand side all these values are mapped on the G α highlighted in different colors where the GDP is shown in green van der Waals representation.....52

Figure 4.15: **a).** Distance between RAS like and Helical domains, **b)** GDP to s6h5 loop distance, **c)** Distance between HG and P loop and **d)** Helix-5 time dependent change during simulation.....54

Figure 4.16: Three superimposed systems bound to GDP molecule and the switch-II region which is shown as loop, the flexibility of the region is observed trajectory wise throughout the course of simulation.....55

Figure 4.17: Comparison with crystal structures. **A.** crystal structure of G α subunit bound to RGS protein (pdb id:2ODE). **B.** crystal structure of G α bound to G $\beta\gamma$ dimer (pdb id:1GP2). **C.** the trajectory snapshots of our results depicting the loop conformation of switch-II region in the presence and absence of restraints. And the catalytic residue GLN-204 is highlighted in all the four snapshots.....56

Figure 4.18: Impact of restraints on the structure of Switch-II **a.** represents the unraveling of α -helix conformation of switch-II in different types of simulations with restraints and without restraints. **b.** This show the comparison of helical conformation lost in the presence of restraints in switch-II region which is required for the catalytical residue Glu-204 present in switch-II region to come closer to nucleotide binding pocket.....57

Figure 4.19: Interactions between switch-II and switch-II. Switch-III is shown as cyan bold color and Switch-II is shown transparent, cluster of positive residues in switch-II and cluster of negative residues of switch-III are labeled. Catalytic GLN-

204 of switch-II make interactions with negative clustered residues in switch-III encircled in red.....	58
Figure 4.20: Distance decreases between Switch-II and III colored black and GLN-204 and GDP colored red.....	59
Figure 4.21: a). RMSD of both systems. b). The N-domain is shown here. K104 and D447 residues are depicted in VDW representation the red color represents mutant-447, and orange represents WT-447.....	59
Figure 4.22: MD simulation of WT and D447A in the N-domain of HK2. (A-B) Comparison of PCA derived RMSF profiles for the first and second monomers of the N-domain of HK2 for WT (black) and D447A (red). (C) Depiction of regions of the 3D structure of the N-domain of HK2 that show higher fluctuation in the D447A mutant.....	60
Figure 4.23: Enhanced dynamics of D447A mutant in the N-domain of HK2. Comparative depiction of N-domain regions with enhanced dynamics in D447A compared to WT for the small subdomain (A), large subdomain (B), and linker helix α_{13} (C). The figures are taking as a pattern of protein movement throughout the course of 200ns simulation and the red color shows the first/initial conformation of the protein and the remaining are the frames of movement captured during the simulation.....	61
Figure 4.24: Distance and angle probability distributions of WT and D447A. (A) Distance probability distribution calculated between $C\alpha$ atoms of residues T88 and T232 displayed in a sphere representation. (B) Angle probability distribution calculated between $C\alpha$ atoms of residues T389, R468 and A839, shown as spheres.....	62
Figure 4.25: Results of Mitoxantrone molecule. A. Binding pocket distance and trajectory wise small subdomain fluctuation and binding pocket opening distance. B. Angle probability. dashed lines are reference values of mutant system.....	63
Figure 4.26: Ligand binding poses shown in van der Waals representation targeting the N-terminal domain of HK2.....	65
Figure 4.27: MD results with Vilazodone molecule. A. Binding pocket distance and trajectory wise small subdomain fluctuation and binding pocket opening distance. B. Angle probability. dashed lines are reference values of mutant system.....	65
Figure 4.28: Results of Tirofiban molecule. A. Binding pocket distance and trajectory wise small subdomain fluctuation and binding pocket opening distance. B. Angle probability. dashed lines are reference values of mutant system.....	66
Figure 4.29: Results with Idebenone molecule. A. Binding pocket distance and trajectory wise small subdomain fluctuation and binding pocket opening distance. B. Angle probability. dashed lines are reference values of mutant system.....	67

LIST OF TABLES

Table 3.1: Templates used to model the interfaces in the tetramer.....	27
Table 4.1: Screened molecules from first virtual screening results.....	63
Table 4.2: Screened molecules from second virtual screening results.....	64



LIST OF SYMBOLS

χ	:chi angle
θ	:theta
ϵ	:epsilon
ϕ	:phi
λ	:lambda
σ	:standard deviation
δ	:delta
∞	:infinity
μ	:mu



ABBREVIATIONS

GPCR	: G protein-coupled receptor
HK2	: Hexokinase 2
HK1	: Hexokinase 1
RMSD	: Root mean square deviation
RMSF	: Root mean square fluctuation
PCA	: Principle component analysis
Gα	: G protein alpha subunit
Gβ	: G protein beta subunit
Gγ	: G protein gamma subunit
H5	: G protein alpha subunit helix 5
α_{13}	: Hexokinase 2 linker helix
T_m	: Melting temperature
NADPH	: Nicotinamide adenine dinucleotide phosphate hydrogen
NAD	: Nicotinamide adenine dinucleotide
OMM	: Outer mitochondrial membrane
TS-D	: Transition and denatured protein states
NPT	: Isothermal-Isobaric ensemble
NVT	: Canonical ensemble
NTD	: N-terminal domain of HK2
CTD	: C-terminal domain of HK2

PROTEİN YAPI VE DİNAMİĞİNİN HESAPLAMALI YÖNTEMLER ARACILIĞIYLA İNCELENMESİ VE TERAPÖTİK AMAÇLAR İÇİN MODÜLASYONU

ÖZET

Samman Mansoor

Biyomedikal Mühendisliği ve Biyoinformatik, Doktora

Tez Danışmanı: Doç. Dr. Özge Şensoy

Ağustos, 2021

Protein molekülleri, hücrede gerek duyulduğunda, belli fonksiyonları yerine getirebilmek için birtakım konformasyonel değişikliğe uğrar. Proteinlerin adapte olduğu bu konformasyon topluluğunu anlayabilmek için proteinin üç-boyutlu yapısı ile fonksiyonu arasındaki ilişkiyi açığa çıkarmak gerekmektedir. Proteinlerin dinamiği proteine değişik yapıda ligand bağlanması, protein oligomerlerinin oluşması ve proteinlerin DNA veya RNA ile etkileşimi sonucu değişir. X-ray ve NMR gibi deneysel yöntemlerle elde edilen yapılar proteinlerin üç boyutlu yapısı hakkında değerli bilgiler sağlamaktadır. Öte yandan, proteinlerin belli çevresel faktörler altında adapte olduğu yapıların zamana bağlı olarak belirlenebilmesi için moleküler dinamik simülasyonlarına gereksinim duyulmaktadır. Bu tez kapsamında, her biri, farklı proteinlerin dinamiğinin incelenmesini hedef alan üç sistem ile çalışılmıştır. Bu sayede, proteinlerin yapı-fonksiyon ilişkilerinin aydınlatılması ve edinilecek bilgi birikimi ile birtakım hastalıkların tedavisine alternatif terapötik çözümlerin geliştirilebilmesi hedeflenmiştir. Tez kapsamında çalışılan ilk sistem hücresel sinyalleşmede önemli role sahip olan G protein-kenetli reseptörlerdir (GPKR). Çok sayıda fizyolojik proseste rol almasından dolayı G protein-kenetli reseptör ailesinde bulunan proteinler reçetelenen ilaçların yaklaşık olarak %40'ında hedef olarak kullanılmaktadır. Öte yandan, bu reseptör ailesini hedeflemek oldukça zordur; çünkü birden fazla reseptör bir araya geldiğinde reseptörlerin tek başlarına yerine getirdikleri fonksiyonlar değişmektedir. Bu sebeple, reseptörlerin yaptıkları oligomerin kontrol edilebilmesi bu yapıların sebep olduğu kanser, diyabet ve birtakım nörolojik hastalıkların tedavisinde alternatif yöntemlerin geliştirilmesine olanak sağlayacaktır. Bu problemi çözebilmek için, aynı anda iki reseptöre bağlanabilen heterobivalent adı verilen ligandlar geliştirilmeye başlanmıştır. Tez çalışmasının birinci bölümünde, bilgimiz dahilinde, ilk defa, heterobivalent ligandın Adenozin 2A ve Dopamin 2 reseptöründen meydana gelen ve Parkinson hastalığında rol oynadığı bilinen tetramer üzerindeki etkisini hızlandırılmış moleküler dinamik simülasyonları yardımıyla incelenmiştir. Elde edilen sonuçlar, heterobivalent ligandın A_{2A}R'nin D₂R üzerindeki antagonistik etkisini ortadan kaldırabilme potansiyeli olduğunu göstermiştir. Tek bir GPKR, G protein veya Arrestin aracılığıyla birden fazla sinyal yolağını aktive edebilmektedir. Bazı durumlarda, bu yollardan bazıları organizma için zararlı olabilecek yolların da aktive edilmesine neden olur. Bu sebeple, reseptörün spesifik olarak G protein veya Arrestin ile etkileşmesini sağlayabilecek yöntemlerin geliştirilmesi elzemdir. Bunun için, G protein ve Arrestinin aktivasyon mekanizmasını anlayabilmek

gerekmektedir. Bu gereksinimden ortaya çıkan motivasyonla G proteinin aktivasyon mekanizması tez çalışmasının ikinci bölümünde geliştirilen bir metotla açığa çıkarılmıştır. Buna göre, deneysel olarak elde edilen protein erime sıcaklıkları kullanılarak elde edilen değerler moleküler dinamik simülasyonlarında sınırlayıcı sabitler olarak kullanılmıştır. Elde edilen sonuçlar, G protein aktivasyon mekanizması hakkında önemli bilgilere ulaşılmasını sağlamıştır. Bu sayede, G proteinin aktivasyonu sırasında protein tarafından adapte olunan yapılar hedeflenerek efektör molekülünün aktivasyon durumu kontrol edilebilecek ve bu prosesin aksamasına neden olan birtakım mutasyonlardan kaynaklanan hastalıkların tedavisine alternatif çözümler önerilebilecektir.

Tez çalışmasının üçüncü bölümünde, kanser oluşumunda rol oynayan hegzokinaz enziminin aktivitesini engelleyebilen mutasyonların proteinin dinamiği üzerindeki etkileri moleküler dinamik simülasyonları kullanılarak incelenmiş ve proteinde benzer etkileri meydana getirebilecek küçük terapötik moleküllerin bulunması ile ilgili çalışmalar yapılmıştır. Bilgisayar ortamında başarılı olarak belirlenen molekül adaylarının aktiviteleri in vitro deneyler vasıtasıyla sınanacaktır.

Anahtar sözcükler: G protein coupled reseptör, Oligomerizasyon, Accelerated moleküler dinamik simülasyon, Heterobivalent ligand, Restraint moleküler dinamik simülasyonların, G protein, Phi-Value, Hegzokina enzim 2.

COMPUTATIONAL INVESTIGATION AND MODULATION OF STRUCTURAL AND FUNCTIONAL PROPERTIES OF PROTEINS FOR THERAPEUTIC PURPOSES

ABSTRACT

Samman Mansoor

Ph.D. in Biomedical Engineering and Bioinformatics

Advisor: Assist. Prof. Dr. Özge Şensoy

August, 2021

Proteins are dynamic molecules which undergo certain conformational changes to perform specific functions when needed in the cell. To have a holistic understanding of the conformational ensemble they attain one needs to explore the relation between protein 3D structure and function. The dynamics of proteins is altered upon binding of different types of ligands, formation of protein-protein oligomers and macromolecules such as DNA or RNA. The experimental structures, which are obtained by X-ray crystallography or NMR, provide a static picture of the 3D structure of proteins; however, to investigate the function of protein it is crucial to capture time-dependent conformational behavior of biological systems using computational methods such as molecular dynamics simulations.

In this thesis, three different systems, each of which focuses on understanding and modulation of dynamics of proteins, are studied in an effort to come up with alternative methodologies that can be used for therapeutic purposes.

The first system is the G protein-coupled receptor, which is a seven-transmembrane receptor involved in cellular signaling. Due to involvement in various physiological as well as pathophysiological processes, they have been used as drug targets and make up ca. 40% of prescribed drugs on the market. On the other hand, targeting this receptor family is challenging due to the oligomerization issue which modulates the character of individual protomers in the oligomers formed. Since oligomerization is very common and crucial for GPCR function, modulating this process emerges as a highly promising therapeutic strategy that can be used to treat many diseases; several compounds that are already in clinical use have been later discovered to act via modulation of protein oligomerization. To tackle this problem, (hetero)bivalent ligands, which can simultaneously bind two receptors, have been developed. In the first part of the thesis, the impact of a designed heterobivalent ligand on the GPCR tetramer, which is composed of a pair of Adenosine 2A and Dopamine 2 receptor and has been shown to be involved in the pathophysiology of Parkinson's disease, has been studied by means of atomistic accelerated molecular dynamics simulations. The results have shown that the heterobivalent ligand could inhibit the antagonistic impact of A_{2A}R on D₂R.

As it has been well-established that a single GPCR can activate more than one biological signaling pathway by coupling either to G protein or Arrestin. In some cases, one of these pathways can bring about some undesired side effects. Therefore, it is very crucial to direct

the receptor for its coupling with the specific effector. Moreover, specific targeting of a signaling pathway can be possible by having a holistic understanding of the activation mechanism of these effectors, namely G protein and Arrestin. Motivated by this fact, the activation mechanism of G protein has been studied by means of a novel method, which is developed in the course of the second part of the thesis study. According to that, experimental melting temperature values are used to calculate restraints applied in molecular dynamics simulations. The results obtained by means of these restraint values shed light into the activation mechanism of G proteins, which have been involved in the onset of various crucial genetic disorders. Therefore, the findings can be useful in targeting the effector molecule at specific regions on the protein.

In the third part of the thesis, we studied the impact of mutations that inactivate the N-terminal domain of Hexokinase 2 enzyme by means of molecular dynamics simulations.

Keywords: G protein-coupled receptors, Oligomerization, Accelerated molecular dynamics simulation, Heterobivalent ligand, Restraint molecular dynamics simulation, G protein, Phi-Value, Hexokinase 2.

CHAPTER 1

1. INTRODUCTION

Proteins are associated with a broad range of biological functions, thus constituting essential part of all living organisms. In particular, enzymes act by decreasing the activation energy needed to catalyze biochemical reactions for maintaining life processes. On the other hand, proteins such as hemoglobin helps to transport the molecular oxygen and other essential compounds to the sites of their utilization. Antibody molecules, which have become hot targets due to pandemics, are responsible for binding and neutralizing foreign materials that may be harmful to an organism. Although they serve different purposes, the common feature of proteins is their target specificity. A particular enzyme binds to specific substrate molecule and catalyzes a specific chemical reaction with that substrate. Similarly, a particular antibody molecule binds to specific antigens. Interestingly, the specificity of action is so strictly encoded in proteins such that a slight change in protein exerts a dramatic impact on its function. A number of proteins also perform regulatory roles, for instance, by binding DNA they modulate the expression of certain genes in a timely manner. As such, proteins have been involved in pathophysiology of various diseases [1]–[3].

To unveil how these complex macromolecules are involved in various functions, researchers have been putting their efforts into understanding the dynamical properties of proteins. 3D structures of proteins, which can be achieved by using X-ray crystallography or NMR, provide valuable information on how amino acid residues are packed with respect to each other. On the other hand, this corresponds only to a single snapshot of the system among many others and therefore does not provide information on the ensemble of all possible conformations which can be adapted by the system under certain conditions. From that perspective, molecular dynamics simulations, in particular, is of great

significance in studying details of dynamic behavior of biomolecular systems. As such, it serves as a bridge between theory and experiment and complements the missing piece, which is the conformational energy landscape, in experimentally determined structures. Hence, observing various conformational states of proteins and the pathways that connect them provides a dynamic trajectory of the protein studied. On the other hand, to achieve complete conformational energy landscape of proteins, in other words, to satisfy ergodic hypothesis, advanced simulation algorithms are required. With advancements in algorithms and computational power MD has started to be used in addressing various crucial biological problems such as signal transduction, allostery, cellular transport, cellular recognition, enzyme catalysis, etc. In particular, accelerated molecular dynamics simulations, which is one of the enhanced sampling techniques, has been widely used to achieve complete conformational landscape of proteins. Consequently, this provides access to conformational space and time scales which are sampled in experimental studies [4]–[8].

1.1. Motivation And Thesis Contribution

Within the scope of this thesis, three different systems are studied, all of which are key for understanding and modulating functions of proteins. Eventually, the findings gathered will provide an avenue for developing alternative therapeutic approaches to treat various crucial diseases. Below, we summarized the contributions made to the field by each project in the context of the thesis study.

- In the first system **Section 3.1** we are motivated to understand the allosteric interactions within the G protein-coupled receptor (GPCR) tetramer by means of accelerated molecular dynamics simulations, which is composed of a pair of Adenosine 2A /Dopamine 2 receptor and involved in the pathology of Parkinson's disease. Subsequently, we made use of the information gathered from molecular dynamics simulations and experiments to develop heterobivalent ligands that target A_{2A}R dimer to mask its antagonistic impact and to investigate the structural/dynamical properties of tetramer under the influence of the designed bivalent ligands which can be used as a combinational therapy in treatment of Parkinson's disease. Moreover, successful novel bivalent ligands will be tested -for the first time- for their ability to alleviate side effects of therapeutics that have been used to treat Parkinson's disease.

- Mutations in G proteins cause the onset of various genetic diseases. Understanding impact of these mutations on the function of the protein requires a holistic understanding of working principles of activation mechanism yet it has remained poorly understood. We are motivated of studying activation mechanism of G protein by means of a method developed in the course of the thesis study. The method could provide intermediate states adopted during the activation of the protein which are sampled along the activation pathway by means of phi-values and then used these values as restraints in MD simulation. Consequently, short-lived states which cannot be otherwise captured by experimental techniques, such as X-ray crystallography and NMR, will be achieved. Importantly, these intermediate states can be used to specifically target G protein at certain steps of the activation mechanism thus providing a means to modulate the function of the effector precisely **Section 3.2**.
- In the third part of the thesis study **Section 3.3**, we set out to investigate the impact of a mutation that has been shown to inactivate the N-terminal domain of Hexokinase 2, the enzyme which is upregulated in various cancer types. Subsequently, we performed virtual screening to find candidates that can mimic the impact elicited by the mutant protein.

1.2. Literature Review

1.2.1. G protein-coupled receptors

The G protein-coupled receptors (GPCRs) are the most abundant class of proteins found in all mammalian genome. This super family of proteins is comprised of structurally similar proteins divided into various families (classes). Human genome contains approximately 800 GPCRs which makes them the largest family of membrane proteins in human [9], [10]. Due to their abundance, GPCRs undertake a plethora of essential physiological functions thus serving as targets for numerous drugs on the market. Their ligands are structurally heterogeneous, including natural odorants, nucleotides, amines, peptides, proteins, and lipids. Currently, approximately 30-40% of marketed pharmaceuticals target GPCRs, such examples include angiotensin receptor and β -blockers, histamine receptor inhibitors, opioid agonists, dopamine agonists and many more [11]. Due to their abundance, there is an enormous potential of targeting these receptors for the development of new and safer

drugs [12]–[17].

The GPCRs have a conserved structure of seven transmembrane domains (TMD) of successive amino acid residues that express moderately high levels of hydrophobicity and are characterised by α -helices spanning the plasma membrane and an eighth helix at the C-terminus [18]–[20] as depicted in **Figure 1.1**. In general the transduction of extracellular stimuli into intracellular signals is the primary function of GPCRs.

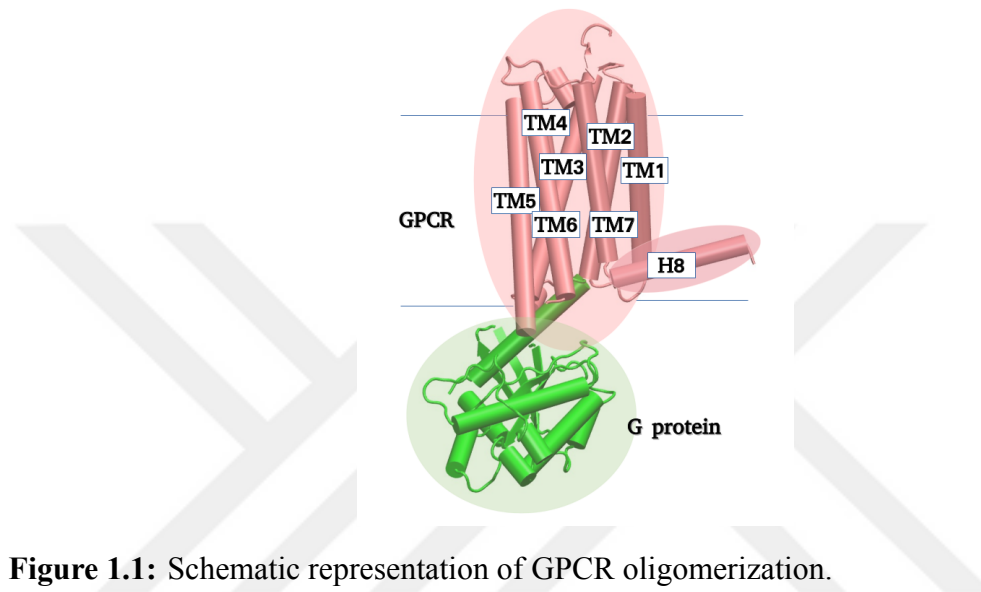


Figure 1.1: Schematic representation of GPCR oligomerization.

The diversity of GPCRs has caused a hindrance in developing a comprehensive classification system and thus they are widely classified on the basis of their physiological and structural features. The first introduced system of classification was the A–F system that delineate the order of GPCRs on the basis of their functional and sequence similarities into six different classifications, termed as: class A (rhodopsin-like receptors), class B (secretin receptor family), class C (metabotropic glutamate receptors), class D (parasitic mating pheromone receptors), class E (cyclic AMP receptors) and class F (frizzled/smoothened receptors) [11]. The phylogenetic classification of GPCRs classified them as “GRAFS” which includes main five families; G (Glutamate), R (Rhodopsin), A (Adhesion), F (Frizzled/taste2) and S (Secretin). Where the major difference among both classification schemes is related to family B which is additionally divided into secretin and adhesion families in GRAFS [21].

Ligand binding to a GPCR brings a alteration in the dynamics of the receptor in several

ways, generally by modifying the fraction of time spent by a receptor in its certain conformational state [22]. Historically GPCRs were considered as two states entities that can only adopt two conformations either active or inactive. This model served as a basis for the well established trio-complex model of GPCR-driven signalling, which proposes a uniform active conformation favored by G protein-coupled receptor kinases (GRKs), arrestins and G proteins that bind to their corresponding receptors. Nevertheless, the biophysical experiments with $\beta 2$ adrenergic receptor have opened new dimensions in the conventional understanding of GPCR conformational states demonstrating that a receptor can adopt numerous conformations and the equilibrium of these conformations is dependent on the type of bound ligand and its proximity towards its respective G protein binding. Ligand binding can influence the receptor as an example, binding of certain agonist may enhance the active conformational state population, or certain ligand binding may induce such a conformational change which has not been observed previously as shown in **Figure 1.2** [23]. In addition the rate of transitions between conformational states can also be impacted by the binding of certain ligand to the receptor. Moreover, other events such as binding of intracellular effector molecule i.e; arrestin or G protein, post-translational modification, pH change or membrane composition of GPCR and dimerization of GPCR bring about same impact on GPCR dynamics [24], [25]. Therefore, the dynamics of a GPCR reflects as a combinatorial effect of structural changes due to external perturbations and intrinsic dynamics of the receptor, the events emerge upon the binding or dissociation of ligand from the receptor.

For transmitting signals across the membrane the GPCR transmembrane helices undergo certain rearrangements and such conformational changes exhibit a critical role and are particularly important for the function of the protein [26], [27]. As a GPCR undergoes transition from an inactive (a state where intracellular effector molecules cannot bind to GPCR) to active state, helices underwent a relative rearrangements to one another by means of position shifting, twisting and tilting. Some reorganizations such as bending and kink in the structure occur usually at a position where glycine or proline disrupted the hydrogen bonding of the backbone in the helical sequence. The intracellular site of the receptor tends to get experience relatively bigger conformational changes compared to the extracellular half which also undergoes certain conformational changes [28]. In particular, the subtle changes occur in the ligand binding site of the receptor which is then transmitted

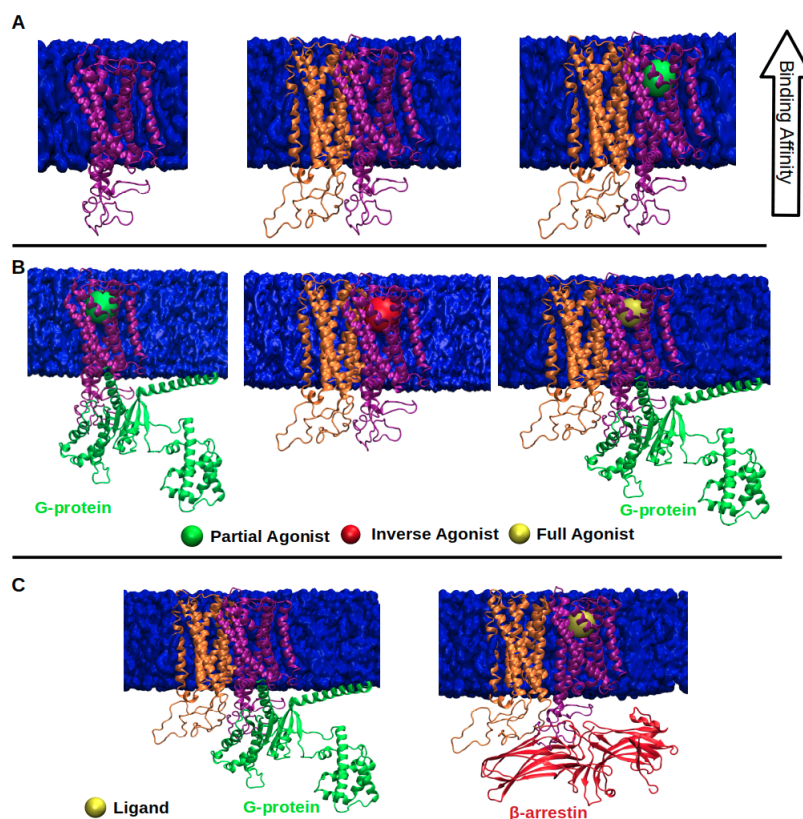


Figure 1.2: **A.** depicts that ligand binding affinity of monomer (purple) increases in the presence of its partner protomer (orange); **B.** ligand efficacy of monomer (purple) decreases (middle) or increases (right) in the presence of its partner (orange) when they form dimer; **C.** functional selectivity of dimer changes in the presence of protomer; concurrently, G protein is substituted with β -arrestin [23].

to the later event of conformational changes of helices at the intracellular site of the receptor. Post-translational modifications such as palmitoylation and phosphorylation also contribute to the function of receptor, for instance phosphorylation of certain residues on C terminus and intracellular loop regions cause structural difference which can drastically impact the coupling propensity of their respective intracellular partners [29]. GPCRs are sensitive to the flexibility of the intracellular/extracellular loops, and the binding kinetics of a ligand is also impacted by the flexibility of the extracellular loops of the receptor. Other than the orthosteric ligand binding pocket, these loops can form alternative regions for ligand binding known as allosteric binding sites. Hence these regions not only alter but also modulate the properties of binding cavity in terms of binding affinity and efficacy of the ligands. Several lines of evidence indicate that there are multiple distinct conformational states adopted by a GPCR transmembrane helices, contradicting the conventional

idea of just a single inactive vs active state and emerging the concept of GPCR oligomerization [23], [30], [31].

In drug discovery GPCRs stand as hot targets and currently computational biology methods are being employed to understand GPCRs as such drug targets. Recent breakthrough in GPCRs crystallization have benefited the novel drug discovery by means of suggesting better off-target rationalization and virtual screening methods [32]–[34]. Recently, a research group have developed a computational protocol to explore the geometrical and physicochemical properties of the conformational space and flexibility of the bioamine receptor family, by combining the concepts of cheminformatics and statistical mechanics. Other examples of computational studies in the field of GPCR include the multiple microsecond molecular dynamics simulations, GPCR-ModSim is a web based portal for performing homology modeling and carrying out MD simulations studies of GPCRs and GPCR structure based drug designing method known as molecular docking [35]–[37].

1.2.2. Experimental techniques used to capture GPCR conformational states

Over time, the GPCR can adopt a fundamentally infinite number of conformations which dictate the function of the protein [38]. Crystallography and NMR (Nuclear magnetic resonance) spectroscopy provide a wealth of information on variations of conformational states of GPCRs. Concurrently, the GPCR crystallization in its various conformational states has remained challenging. The crystal structures of a GPCR, represent the only conformational state under which the receptor has been crystallized, and within the structure the positions of atoms represent the averages of existing conformations in that crystallized conformational state. On the other hand, protein NMR spectroscopy provides a useful way to detect subtle conformational changes, in this technique NMR spectrum adopts the physiological environment in the presence of a solvent to describe the relative position along the shape of a peak on the spectrum of NMR [39], [40]. To determine the distance among two different probes double electron-electron resonance (DEER) spectroscopy histogram have been used [41].

These spectroscopy experiments have not produced a detailed structural description of such conformational states, besides the one resolved structure by NMR; nevertheless these experiments generate information about the localized position of used chemical probes. From that perspective, MD simulations complement because they provide a complete

atomic level description of the intrinsic structural changes occurring over time, capturing the transition and dynamics of conformational states of the structure. On the other hand, these simulation studies are constrained and demanding in high computational resources due to the reduced accuracy and precision of their underlying physical models, yet in last years there have been significant improvement and advancement in both fronts [42]–[46].

1.2.3. Bivalent ligands

G protein-coupled receptors (GPCRs) mediate most of our physiological responses to stimulants by coherent action of various modulators such as orthosteric/allosteric ligands, membrane and receptor partners, collective interaction of which leads to formation of GPCR oligomers. In these supramolecular structures, continuous information flow among the protomers transforms GPCR homo/heteromers into allosteric hubs, thus altering functional behavior of individual receptors. Over the last decades, it has been thought that the supramolecular structure, which is composed of a GPCR homodimer and G protein, constitutes the main functional unit of GPCR signalosomes [12]–[14]. Indeed, a recent experimental study has shown that a GPCR heterotetramer is comprised of homodimers of $A_{2A}R$ and D_2R along with their cognate G proteins [47]. GPCRs are involved in pathologies of many crucial diseases such as Parkinson's disease (PD), Alzheimer's disease (AD), diabetes, and cancer, thus making them hot targets in drug discovery studies [48]–[50]. On the other hand, allosteric interactions present in GPCR oligomers make targeting this class of proteins challenging in the field of GPCR pharmacology [51]. As a notable example, PD can be considered, which is caused by the loss of neurons that produce dopamine in substantia nigra [52]. As an effective therapy, dopaminergic agonists have been considered to increase dopamine level but remained insufficient [53]–[55]. It was after the discovery of antagonistic impact of $A_{2A}R$ on D_2R , where $A_{2A}R$ decreases affinity and intrinsic efficacy of dopamine at D_2R protomer [56], $A_{2A}R$ antagonists have been used as combination therapy along with D_2R agonists in treatment of PD [57], [58]. Recently, it has been shown that $A_{2A}R$ - D_2R assembles into a tetrameric structure which is composed of a pair of this dimer. Moreover, it has been also demonstrated that simultaneous occupation of both $A_{2A}R$ protomers by an agonist and an antagonist in the tetramer does not induce an allosteric modulation of D_2R agonist binding and intrinsic efficacy [47].

Therefore, this suggests that effective modulation of GPCR oligomers requires simultaneous targeting of individual receptors. Hence, bivalent ligands have emerged as useful tools to simultaneously target receptor dimers within the oligomer [59], [60]. A bivalent ligand is composed of two pharmacophores, which are covalently linked by a spacer group and can be categorized into two groups: homo- and heterobivalent ligands which consist of same and different pharmacophore groups, respectively as shown in **Figure 1.3**.

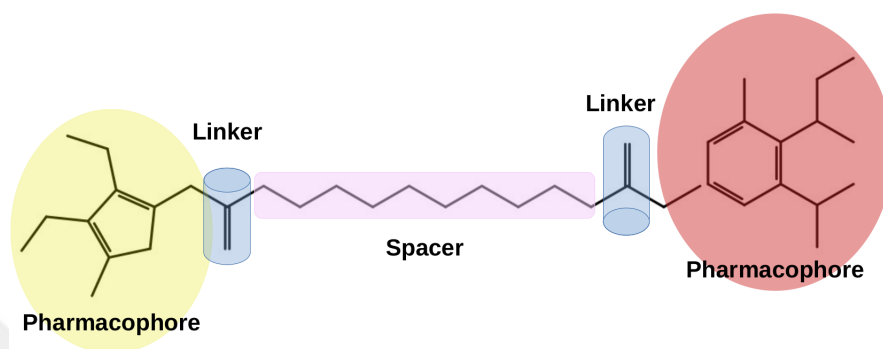


Figure 1.3: Schematic representation of a key structural features a bivalent ligand, where the pharmacophores can be either the same or different on both sides of the ligand.

There has been a couple of experimental studies where the optimum length of the spacer and stability of the bivalent ligand were determined using computational tools, and their biological activities were also tested by in vitro studies [61], [62]; however, none of them provide a mechanistic insight into the impact of these special class of ligands on structure and dynamics of receptors which is crucial for development of effective ligands with precise targeting capabilities. In this part of the thesis study, we investigated the impact of a designed bivalent ligand on dynamics of a tetramer, which is composed of $A_{2A}R$ and D_2R homodimers and cognate mini- $G_{\alpha s}$ and mini- $G_{\alpha i}$, respectively. The bivalent ligand was modeled to consist of an $A_{2A}R$ agonist (CGS-21680) and an antagonist (istradefylline), which are linked via an affinity-generating biphenyl triazole-moiety and a spacer collectively called as linker [63]. The tetramer was modeled based on a recent experimental study where the dimerization interfaces of homodimers and heterodimer were shown to include transmembrane TM6 and TM4/TM5, respectively [64]. Thereafter, we performed accelerated molecular dynamics (aMD) simulations using the tetramer model with and without the linker-including only pharmacophore groups instead. Comparative analyses of trajectories showed that the bivalent ligand not only linked pharmacophores but also has an impact on the dynamics of protomers in the oligomer. Notably, it reduces fluctuations in antagonist-bound $A_{2A}R$ at extracellular loops (ECLs) 1 and 3 as well as

intracellular loop (ICL) 2, TM 6 and 7, both of which are involved in the homodimer interface. Moreover, this rearrangement led to alterations in the number of contacts, volume of ligand binding pocket and intracellular domain, and allosteric interaction network in agonist-bound $A_{2A}R$ mini- $G\alpha s$ protomer. Interestingly, in spite of not binding directly to D_2R , the bivalent ligand also impacts dynamics of apo D_2R in D_2R dimer. In particular, apo D_2R resembles conformational properties of an inactive receptor in the presence of the linker, thus maintaining an asymmetrical activated dimer when the maximum signaling is achieved [65].

1.2.4. Heterotrimeric G protein

Cells communicate with each other through the exchange of chemical signals which are sensed by specific receptors present across the cell membranes. GPCRs are the largest and diverse group of membrane proteins that are responsible for turning extracellular chemical signals into intracellular responses by activating relevant downstream signaling pathways. Agonist binding to the extracellular region of the GPCR causes a conformational change which leads to the activation of the receptor and binding of subsequent heterotrimeric G protein to the receptor. The activated receptor causes to release of guanosine diphosphate nucleotide, GDP, from the $G\alpha$ subunit of the G protein and replace it with guanosine triphosphate nucleotide, GTP as shown in the **Figure 1.4**. As such, G proteins play an important role in signaling as they act as molecular switches in signal transduction.

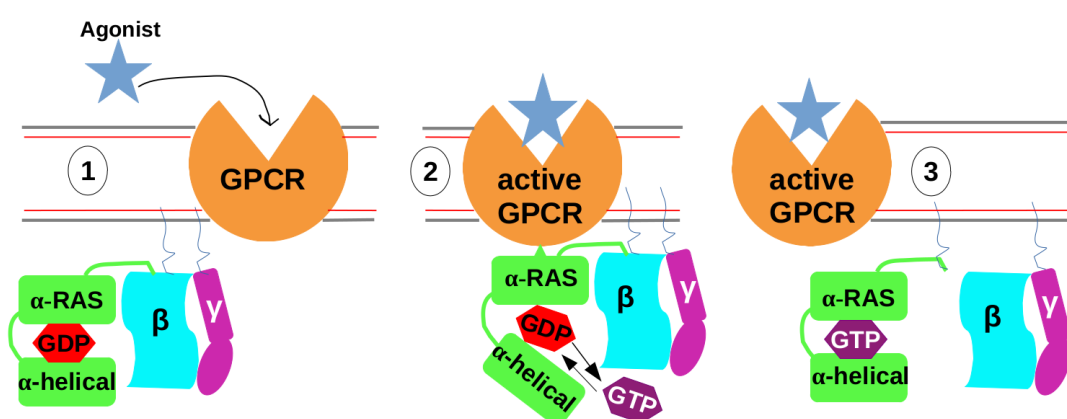


Figure 1.4: Schematic representation of an agonist induced GPCR activation along G protein recruitment in step 1 followed by GDP to GTP exchange by binding and opening of α subunit in step 2. In step 3 subunits dissociate to initiate their corresponding downstream signaling.

GPCRs are the most diverse class of proteins; however, despite such diversity they act

relatively with a smaller number of types of G proteins. These heterotrimeric proteins consist of 3 subunits namely α , β and γ , and there are 21 forms of α subunit while we have only 6 β and 12 γ subunits. $G\alpha$ subunit is categorized into four classes based on their primary sequence similarity such as $G\alpha$ -stimulatory, $G\alpha$ -inhibitory, $G\alpha_q$ and $G\alpha_{12}$. $G\alpha$ subunit has the guanine binding site and structurally the protein is composed of two main domains named as GTPase domain, which are also called as RAS-like domain, and a helical domain [66]–[71].

There have been various studies showing that mutant G proteins are responsible for the onset of various heart and genetic diseases [72]–[75]. There has been a vast number of studies that provide knowledge on the structural constituents of G protein while very little is known about the activation mechanism of the effector. Hence, there is an urgent need to study the detailed mechanism of G protein activation at the atomic level by using available crystal structures that the protein adopts going from inactive to the active state. In this part of the study, we set out to establish such a comprehensive and detailed understanding of G protein activation by means of a modified version of well-known experimental protein engineering technique, Phi-value analysis. Conformational phi-value analysis is an experimental method that has been traditionally used to obtain structural information of transition states adopted during protein folding, which is difficult to determine by traditional methods such as NMR or X-ray crystallography. According to the method, single point mutations to alanine are introduced in the protein and their effect on stability is measured in denaturation experiments [76], [77].

These measurements provide experimental data on the contribution of a given residue to the stabilization of the transition state. Changes in stability caused by a mutation reflect changes in the local environment at the mutation site in the transition state. Thus, this data can be translated to residue-residue interactions at the mutated position in the transition state, which provide information about its structure. To do this, we applied experimentally determined phi values as restraints in molecular dynamics simulations, which characterized the influence of each residue with its surrounding by means of the number of contacts made by that residue with its surrounding in molecular dynamics simulations to investigate mechanism of G protein activation. Our results showed an increase in the flexibility of the nucleotide binding region, domain opening and helix 5 tilting. Additionally, the conformational state of Switch II emerged as a major component responsible for transition from

inactive to the active state of the protein, in other words, the activation mechanism of the protein. The findings might provide a better understanding for developing drugs which could inhibit the G protein in a precise manner.

1.2.5. Hexokinase enzyme

Normal differentiated cells primarily depend on mitochondrial oxidative phosphorylation to generate required energy needs for cellular processes, in contrast, the cancer cells obtain cellular energy by means of aerobic glycolysis known as “the Warburg’s effect”. In the 1920s Otto Warburg pioneered this striking discovery, which states; in principle, aerobic glycolysis is an inefficient way to generate ATP even in the presence of sufficient O₂ levels in cancer cells shown in **Figure 1.5**.

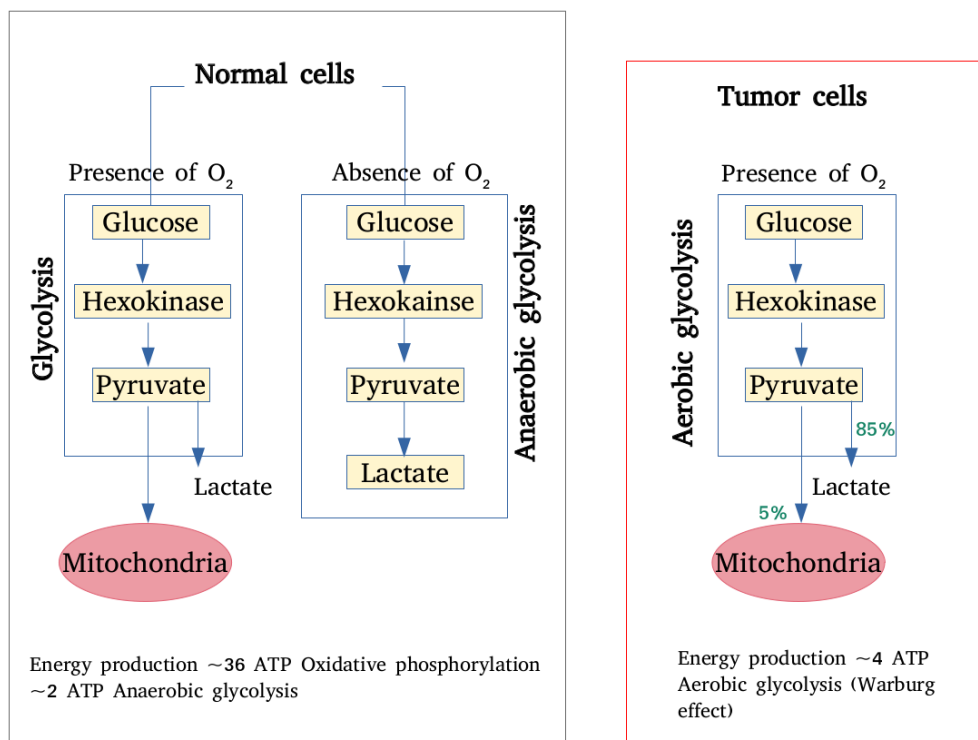


Figure 1.5: Energy production in normal vs tumor cells by Warburg’s effect.

Along this cancer propagates by the malfunctioning of significant metabolic pathways that facilitate the incorporation and uptake of nutrients (glucose) to generate biomass for highly proliferating cancer cells. Another hallmark of cancer cells progression is an altered mechanism of cell apoptosis, perhaps the metabolic impairment which favors tumor microenvironment is not properly understood but in principle the metabolic dependencies of tumor cells can be exploited for cancer treatment [78]–[84]. Glucose transport in

the cell is the first step in glucose metabolism and overexpression of glucose transporters (GLUT) have often been observed in malignant cells [85]. Upon entering the cell, glucose is being phosphorylated to glucose-6-phosphate G6P, this product will be catabolized in glycolysis while anabolically used in glycogen and lipid synthesis via pentose phosphate shunt. In both cases the first step is to catalyze glucose by Hexokinases (HKs), which is a family of four isozymes namely HK1, HK2, HK3 and HK4 (Glucokinase). Among them HK1 is ubiquitously expressed in all mammalian cells and HK2 expressed majorly in skeletal, adipocytes and cardiac cells. Among all isozymes only HK1 exhibit different isoforms due to alternate splicing of HKX1 hexokinase gene-1, the term coined by Mori in 1996 [86].

1.2.6. Why proliferating cells metabolize glucose by aerobic glycolysis?

In the presence of oxygen, aerobic glycolysis is regulated in such a manner which leads to overproduction of lactate. This phenomenon relies on the ability of differentiated cells to metabolize glucose as a primary source into carbon dioxide by oxidation of glycolytic pyruvate in tricarboxylic acid (TCA) cycle occurring in mitochondria. Products of the reaction include NADPH (reduced nicotinamide adenine dinucleotide), which is a major contributor fuel for maximum ATP production in oxidative phosphorylation, with minimal lactate production. In contrast, cancer cells fuel themselves by producing higher lactate under oxidative conditions, which refers to their cell energy mechanism as aerobic glycolysis. Lactate overproduction is a serious metabolic impairment and improper utilization of cellular resources, in which every lactate molecule provides three carbon molecules which are routed either to ATP production or macromolecular synthesis [87].

In oxidative phosphorylation, each glucose molecule produces 36 ATPs, while in contrast with lactate metabolism in cancer cells only 2 ATP molecules are generated per glucose molecule as shown in **Figure 1.5**. The reason behind cancer cells adaptation of this less efficient energy producing lactate mechanism relies on the fact, that during mitosis, dividing cells need to divide all of its cellular components into two viable cells, this puts pressure on dividing cells to have a proper energy supply along enough macromolecules such as lipids, amino acids and nucleotides. Here the significant role of glucose prompts up to resolve these issues of energy supply for proliferating cells to generate ATP and as well as biomass production via ATP hydrolysis. Even glucose helps in providing further

cell required components, for instance, it provides carbon atoms required for fatty acid synthesis, by producing acetyl-CoA (coenzyme-A) which is a major precursor of fatty acids synthesis [88]. In mitochondrial matrix glucose is converted to acetyl-CoA this is the first step in TCA cycle which produces citrate and in highly proliferating, this citrate is excreted out to cytosol which is an outcome of high NADPH/NAD⁺ and ATP/ADP production. Citrate will further produce cytosolic acetyl-CoA in the presence of ATP citrate lyase (ACL) enzyme, hence this will serve as carbon fuel for the production of acyl chains, impairment of this highly significant ACL enzyme harms cancer cells. Additionally palmitate synthesis which is a major cellular membrane component, requires 7 ATP molecules, 8 acetyl-CoA molecules from 16 carbon atoms and 14 NADPH molecules to have 28 electrons, all these requirements are filled by glucose. Henceforth, we know the reasons behind glucose to be routed in aerobic glycolytic pathway of energy metabolism along high lactate production and this phenomenon is beneficial for proliferating cells to escape energy generating pressure with a well routed continuous energy and nutrient supply [87], [89], [90].

1.2.7. Hexokinase enzyme and cancer

Upregulation of HK1 and 2 has been widely observed in various cancer types and co-localization of these enzymes on the outer mitochondrial membrane OMM promotes apoptotic resistance to tumor cells [91]–[96]. Structurally HKs are similar and are dimeric proteins sharing highly conserved glucose binding active sites among four isozymes HK1, 2, 3 and 4, their structures comprising two domains N & C identical in both halves (monomers) of the dimer, dimeric structure shown in **Figure 1.6**. The N-domain in general is responsible for regulating the overall enzyme's stability, the connecting helix region between N and C domain is known as linker-helix α_{13} . Despite having active site similarity in all HK isozymes, only HK2 has catalytically active N-domain as compared to HK1 and 3 [97]–[102].

Targeting aerobic glycolysis has been reported as a promising strategy to reduce cancer cells proliferation and growth which is dependent on aerobic glycolysis, this leads to the discovery of potential glycolytic inhibitors [103], [104]. Currently the most common therapeutic approach against cancer is chemotherapy, a procedure with high side effects due to its low target specificity [105]. To date, many inhibitors of HK2 such as 2-deoxyglucose,

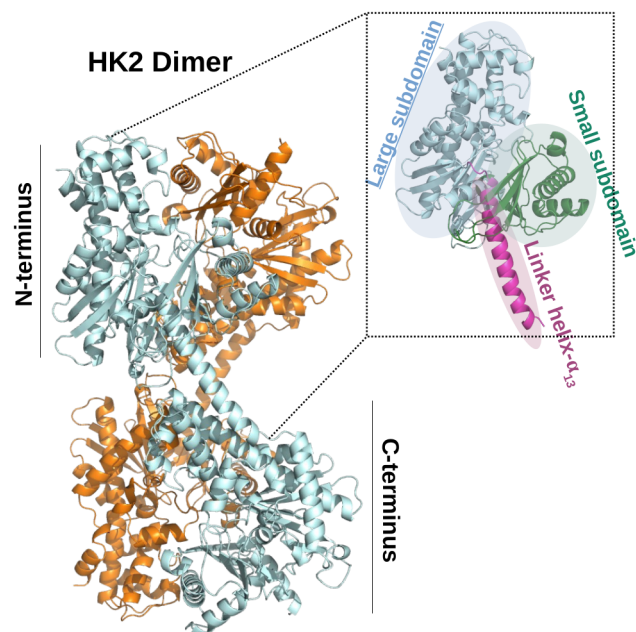


Figure 1.6: The dimeric structure of Hexokinase has been shown.

Ionidamine, 3-bromopyruvate, benserazide and GEN-27 have been reported against various cancers, mostly targeting the active site of the enzyme [106]–[110]. Keeping in view the active site similarities for all HK isozymes the greatest challenge lies in developing such anticancer drugs which specifically binds to one type of HK without interfering with other normal expressing types of HK in the body.

To this extent a research study has shown the inactivation of HK2 N-domain by mutation resulting in overall enzyme’s activity loss, which suggests a unique off-site target of inhibitors for specific types of HK. In particular, mutation of Aspartic acid at 447 position to alanine showed full inactivation of the enzyme. We are motivated to investigate the impact of this mutation on the protein, the knowledge of which can be used to design inhibitors that can inactivate the function of the enzyme.

To this end, we performed all-atomistic molecular dynamics simulations using wild-type apo HK2 and mutant D447A system. From these results further we have performed Virtual Screening by targeting the identified potential non- active site D447 for HK2 inhibition and FDA approved drugs were screened to find out the best candidate. The results of docking studies have shown that interaction of a couple of ligands with HK2 at the target position could elicit similar dynamic behavior observed in the D447A mutant system. These preliminary findings could serve as a new dimension for the treatment of various cancer types.

CHAPTER 2

2. THEORETICAL PART

Particular interactions between specific residues of the protein with a substrate molecule are required for substrate recognition and binding [111]. Non-covalent and covalent both interactions are critical and significant for maintaining the structural integrity of the protein. These interactions can be accurately characterized at both the quantum scale and macroscopic scale by utilizing the classical semi-empirical methods which are based on approximations. It is critical to quantify and comprehend all the significant and potential interactions that maintain protein structural integrity and drive its function in order to fully understand protein behavior.

Protein's functional diversity would lead one to assume a commensurate level of complexity in the detailed structure of these biological molecules. X-ray diffraction studies, which have resolved the crystal structures of numerous proteins, have corroborated this expectation. For every molecule in a globular protein their polypeptide chain is compactly folded into its three-dimensional geometry. Despite the complexity of the resulting structures, it has been generally seen that the packing density of protein components is driven in such way to screen out those residues which are having propensity to make favourable free energy interactions with water molecule and stays near to the protein surface. Although, X-ray diffraction studies of globular proteins with bound ligands have been demonstrated in several circumstances. These investigations have revealed that crucial amino acids with chemically active groups are positioned strategically in their well-defined active sites such that they can interact with the ligand in a coordinated manner [112]–[116].

Proteins are not the static entities thus the dynamic properties of protein structure and function have received a huge attention in recent years. It has been shown from experimental

studies that proteins exhibit significant structural fluctuations which are substantially fundamental to the protein function, yet the exact phenomenal nature of these fluctuations associated with structure has remained elusive [117]. They constitute an ensemble of conformations at room temperature, thus among these different states the rate of transition occur on a variety of time scales ranging from nanoseconds to seconds, connecting these structural ensembles to functionally important phenomenon like enzyme catalysis and allosteric signaling. Due to the dynamic nature of proteins, changes in their structure reinforce the interplay between various types of interactions that preserve their structural integrity and function. Theoretical studies that have provided a precise picture of dynamical atomic motions within a protein have driven the recent interest in protein dynamics [118]–[120].

The concepts derived from protein structure theory and theoretical chemical physics provides a theoretical understanding of proteins. The strategies adapted from chemical physics embrace methods and techniques that have been implemented successfully in the past to examine the atomic motion of liquids and solids in dense materials. In view of large size and high density globular proteins the significance of these methods are considered appropriate. With the advancement of these theoretical developments, new experimental techniques are being developed and available which provide a holistic insights in understanding the protein dynamics. Therefore, this field has evolved robustly as a result of the need in understanding the interplay between experimental and theoretical work. A variety of fundamental properties, including the average magnitude of atomic thermal displacements with their fluctuation magnitude throughout the protein at a particular time scale has been successfully predicted by such theoretical studies [121].

The necessary and stabilizing interactions within the structure of a globular protein can be described quantitatively in the form of potential energy function of the atomic positions that made up the molecule [122]. With this approach the potential energy function is specified as a sum of concepts commonly employed in the description of simpler systems. A typical model potential energy function has the following contributions;

$$U_{\text{total}} = U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedral}} + U_{\text{vdW}} + U_{\text{Coulomb}} \quad (2.1)$$

$$U_{\text{bond}} = \frac{1}{2} \sum_{\text{bonds}} K_b (r - r_o)^2 \quad (2.2)$$

$$U_{angle} = \frac{1}{2} \sum_{bondangles} K_{\theta}(\theta - \theta_o)^2 \quad (2.3)$$

$$U_{dihedral} = \frac{1}{2} \sum_{dihedralangles} K_{\chi}(1 + \cos(n\chi - \gamma)) \quad (2.4)$$

$$U_{impropers} = \sum_{impropers} K_{\phi}(\phi - \phi_0)^2 \quad (2.5)$$

$$U_{nonbonded} = \sum_{nonbond} \left\{ \epsilon_{ij} \left[\left(\frac{R_{ij}^{min}}{r_{ij}} \right)^{12} - \left(\frac{R_{ij}^{min}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon_l r_{ij}} \right\} \quad (2.6)$$

force constants described above are derived from quantum-mechanical and experimental studies, the first term U_{bond} equation 2.2 which is the bond count has been given to every covalent bond in the protein. The stretching and bond length fluctuations are approximated by using simple Hookean spring for maintaining a good approximation at normal physiological temperatures. The second term U_{angle} equation 2.3 refers the deformation energy of angle formed between the two covalent bonds sharing a common atom at the vertex. The third and fourth term $U_{dihedral}$ $U_{impropers}$ equation 2.4 & 2.5 describes dihedral bond which represents the intrinsic deformation energy of atom pairs separated by three covalent bonds known as torsional angle ϕ and an improper dihedral for the geometry of four planar covalent bonds. As a result of summation of these three terms the variation of covalent bonding energy of a protein can be calculated. Electrostatic and van der Waals interactions which are the determinant interactions of the protein structure, are calculated by the summation of terms mentioned in the fifth term $U_{nonbonded}$ equation 2.6.

The above described model for potential energy function have slight different advantages over the calculations made from the function of a full quantum mechanical studies. These include a rapid computation of energy for a given configuration of atoms, and simplifying the analytical expression for computing the spatial energy functions derivatives.

Generally, it is essential to incorporate the solvent effects by adjusting the energy function parameters to study the structural changes in the interior of the globular proteins. This can be achieved by switching the potential energy function with potential mean force (providing average possible configuration of surrounding water molecules respective to the protein) which is dependent on temperature [123], [124].

2.1. Molecular Simulation

Dynamics is such a term which refers to any of the intrinsic motion within proteins. These intrinsic motions are the collective motions at the atomic level within the protein subunits as well as protein domains such as their movement respect to one another. The atomic motions are measured at a few nanoseconds time scale, while the collective motions such as domain motions are a measure of microsecond or millisecond time scale. Force fields are considered as the potentials of molecular simulation methods, and with the help of these potentials the net energy of a biological system (physics-based and knowledge-based) are being accessed. As explained in the previous section, the physics based potential which are derived from quantum mechanical (QM) calculations are used to model interatomic interactions, which include bonded and non-bonded interactions such as bond lengths, bond angles, dihedral angles, van der Waal and electrostatic interactions. However to calculate the interactions within the protein structure the statistical and knowledge-based potentials are used. By means of these approaches the contacts formed between amino acids pairs are determined by differentiating the favorable type of these interactions from the unfavorable ones. The broadly used methods of molecular simulations are categorized on the basis of how the method is carried out to represent the underlying potential of the protein structure [2], [125].

2.2. Molecular Dynamics Simulation

To investigate the dynamics of a biomolecular systems, molecular dynamics (MD) is one of the primary as well as theoretical tool developed by Fermi in 1955 has been widely used. Levitt and Warshel were the first who performed molecular dynamics simulations on biological systems, later McCammon and Karplus carried out MD simulation on a specific molecule bovine pancreatic trypsin inhibitor. Typically all atomic forces i.e; bonded and non-bonded are considered in MD simulation, as the first step of the method is to minimize the molecule of interest with respect to the underlying potential. A standard MD simulation protocol involves solving Newton's equations of motion for the system of interacting molecules. An initial velocity and acceleration is assigned to every particle in the system then after a short time interval usually few femtoseconds the new position (coordinates) of every particle is being calculated by integrating the Newton's equation.

The force on each particle is then calculated from the new position of every particle's potential energy slope and then the acceleration is obtained. The next positions of particles are determined by this information and the procedure is repeated for a defined time length, as a result output trajectory which contains the dynamics of the all atoms of the system have been accumulated. Besides investigating protein dynamics, MD simulations have also been used to study protein folding [126]–[129].

2.3. Accelerated Molecular Dynamics Simulation

By the integration of Newtonian equations of motion to a molecular system the time dependent behavior and its evolution in its conformational space can be determined. And to capture the local and global conformational changes of a protein, accurate potential energy landscape of the conformational space along kinetic and thermodynamics properties should be determined. While in actuality due to the nonergodic nature of large biological systems where these thermodynamics properties are hard to calculate due to large potential energy barriers the conventional methodology of MD simulations can't simulate large systems. Hence to capture the dynamics of real molecular systems there is a need to be able to simulate these potential energy minima which is a series of transition events. This emerged as a need to introduce an alternative methodological technique to address this problem and thus accelerated molecular dynamics simulation emerged as a very useful enhanced sampling technique, which is used to investigate the dynamic properties of proteins which require larger time scales to accomplish. As the simulation time for most systems is limited to nanoseconds in conventional molecular dynamics, and also the energy landscape of the overall system to be separated by high barriers. The system remained under pressure of various potential energy minima barriers to be overcome to capture actual/transition long time scale events. The system thus is in need to have a high computational power to overcome all these energy minima barriers, to compensate with the problem accelerated MD serves as a potential technique. In this technique a robust bias potential function is used to surpass the energy barriers which are encountered by normal molecular systems to cover the series of rare events between different potential energy basins [130]–[133].

In practice the technique underlies by adding a bias potential to actual potential of the

system to overcome or surpass the energy barrier, this in result lowers the energy surface barrier to go from lower to higher energy state. While the potential energy near to the saddle points are left unaffected and only the surface near the minima are raised to jump the barrier. This means the potential energy surface area has been modified in a way to invest less computational energy and thus capturing higher energy transition states events. In theory the negative potential is given to the system to modify the actual potential, where $V(r)$ is the actual potential and $\Delta V(r)$ is the negative bias boost potential. Hence the modified potential with respect to the boost energy value E is given as;

$$V^*(r) = V(r), \quad V(r) \geq E \quad (2.7)$$

and if the simulation is carried out in the normal mode or true potential then the potential surface is sampled above the boost energy value E and it is sampled around the local minimum potential energy surface. And the complete modified potential is based on unmodified (true) potential, bias (modified) potential and the boost energy, such as;

$$V^*(r) = V(r) + \Delta V(r), \quad V(r) < E \quad (2.8)$$

The accelerated molecular dynamics simulation technique should be able to converge to the canonical distribution. The potential boost is designed in such a way that the implemented bias potential should rise from the true potential but remain under the threshold boost energy E value and importantly should be able to reproduce the same shape of minima as with true potential energy minima. In this way when the modified potential will be reverted back to the true potential then the original surface minima will remain intact. Thus the modified potential factor α is specified and the boost potential will be given as;

$$\Delta V(r) = \frac{(E - V(r))^2}{\alpha + E - V(r)} \quad (2.9)$$

The terms E and α are of importance as these will determine the rate of acceleration given to the system. The term α should not be made as $\alpha=0$ in comparison to E as it will make the energy surface discontinuous and to maintain the actual shape of potential energy the α should be set as a very low range from zero.

Among other enhanced sampling techniques one of the advantages of accelerated molecular dynamics simulation is to sample and simulate the conformational space of a molecular system without prior knowledge of predefined parameters such as potential energy wells (barriers) [134].

2.4. G Protein Intermediate Conformational States Identification

Activating the required cellular response through one of the four major G protein families and their respective intracellular effectors (such as phospholipase C and adenylyl cyclase) or through β -arrestin-dependent kinase activation and others, is crucial for a favourable physiological response. Recent discoveries of biased agonists which can preferentially trigger one of these pathways, have the potential to offer new approach for reducing side effects. This leads to the idea of definite conformational states of a receptor which are stabilized by different ligands, and this will lead to activation of specific signalling and regulatory proteins. Biased signaling alternatively defined as functional selectivity, is of great interest in the pharmaceutical industry, as it raises the possibility of developing drugs stimulating the desirable signaling pathways without triggering the harmful ones, thus dramatically reducing the side effects of such drugs. The structural basis for biased signaling is not well understood, however it is thought to be dependent on the GPCR's intracellular interface of coupling to assume more than just a single inactive and a single active conformational state [66], [135]–[137].

Due to long standing pharmacological importance of GPCRs, different therapeutic approaches are available to target these receptors. As it has been known a single receptor can activate more than one biological signaling pathway and biased ligands have the potential to stabilize a certain receptor conformation that can preferentially stimulate only a specific signaling pathway. That is to say, classic ligands can initiate both G protein and arrestin-mediated signaling pathways while biased ligands can specifically initiate one of these pathways [138]–[141]. Specific targeting of a signaling pathway is crucial since either G protein or arrestin-mediated pathway can bring about some undesirable side effects. For instance, opioid receptors (ORs) have been widely used as targets for analgesics; however, OR-targeted drugs cause some side effects such as respiratory depression, and constipation [142], [143]. According to reported studies it has shown that analgesic effect

of μ -OR is mediated by G_i signaling, whereas β -arrestin-mediated signaling causes these undesirable outcomes [144].

This phenomenon is very critical for cell signaling due to this reason there is a need to study such protein complexes of GPCR-ligand-effector (g-protein/arrestin) in detail, while there is a limited number of available crystal structures of these complexes. In this regard we have proposed a strategy to study such complexes, and to map their intermediate states linked to the conformational changes which are contributed by biased ligands by means of phi-values. Phi-value analysis is a popular protein engineering method used to map transition/intermediary states of protein folding, protein binding, catalysis as well as conformational transition at individual residue level of proteins. This method was introduced by Alan Fersht, the main experimental methodology of this analysis is based on mutational studies of equilibrium constant and folding states, where alanine mutation is introduced in amino-acid and thus the protein stability change upon mutation is determined between native and denatured state of that protein [145]–[147].

2.5. Phi-Value Analysis Technique Concept And MD Simulation

Protein folding reaction is an ensemble of unfolding states having the propensity to adapt many different conformations enroute to attain its final tertiary structure. The formation of stable tertiary structure is a crucial step followed by a strict hierarchy from early secondary to tertiary conformation which is determined by various transition states of folding either by nucleation-condensation or hydrophobic-collapse mechanisms. Determination of these transition states is well illustrated by phi-value analysis, a technique of protein engineering introduced by Fersht in 1995 [77], [148]–[151]. The mutational phi-value analysis invokes the linear free energy relation to measure the impact of mutation on protein folding kinetics and equilibrium, quantitatively expressed as:

$$\phi = \frac{(\Delta G_W^{\text{TS-D}} - \Delta G_M^{\text{TS-D}})}{(\Delta G_W^{\text{N-D}} - \Delta G_M^{\text{N-D}})} \quad (2.10)$$

$$\phi = \frac{(\Delta \Delta G^{\text{TS-D}})}{(\Delta \Delta G^{\text{N-D}})} \quad (2.11)$$

where $\Delta G^{\text{TS-D}}$ represents the energy difference between transition and denatured state W

denotes wildtype and M mutant, and ΔG^{N-D} difference between native and denatured state. ϕ -values range from 0-1 where value close to 0 indicates distorted transition state at the site of perturbation (mutation) and range 1 indicates the transition state retains native like conformation around perturbation [152]. To date ϕ -value is the only technique giving insight to the local and non-local residue contacts formed between transition states of protein folding events, but there is a need to define these transition states on the structure of proteins.

Molecular dynamics simulation has become an integral part to the solution, Dagget in 1994 introduced the term ϕ_{MD} (phi-molecular dynamics) the ratio of number of contacts a particular residue attains in its transition state relative to the native state being calculated by the structural change observed around that residue in chymotrypsin inhibitor 2 [153]. Methodological advances have now enabled the structural determination of transition states by means of MD to obtain atomic level details of individual protein folding, but remained limited to a smaller number of residues to capture millisecond time scale. Such as reported studies on different proteins where very long all-atom MD simulations of 20 μ s-1ms were performed on special machine Anton to reveal the full folding pathway of various proteins [154], [155], but computationally it is cumbersome to carry out such long simulations and with a higher number of atoms.

The limitation of such long length simulations is a very short number of atoms to be considered for study, not more than 100 residues. Likewise, there is another approach where they have used experimentally derived ϕ -values as restraints to determine the transition state ensembles by means of ensemble-averaged molecular dynamics simulation [156].

CHAPTER 3

3. EXPERIMENTAL PART

3.1. System 1 Methodology Workflow

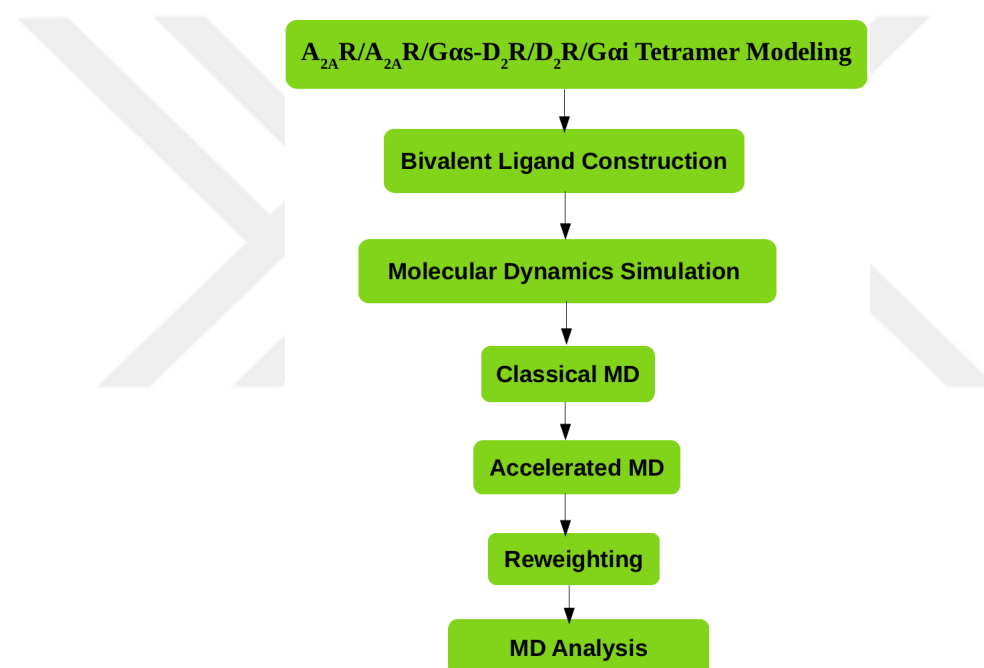


Figure 3.1: Methodology Workflow of first system.

3.1.1. Tetramer modeling

The crystal structure of the tetramer is not available but interaction interfaces in homo- and heterodimers were revealed in a recent experimental study [64]. According to that, i) homodimer and heterodimer interfaces were built up of TM6 and TM4/TM5, respectively, and ii) agonist-bound protomers complex with effectors were located at the periphery of the tetramer to accommodate G proteins. Before modeling the dimers, first receptor monomers were modeled. In particular, CGS-21680-bound A_{2A}R mini Gα_s complex, we have used crystal structures of NECA-bound A_{2A}R mini Gα_s complex (with RAS-like

domain)(PDB ID:5G53) [157] as a template since NECA shares similar chemical groups with CGS-21680. In fact, the crystal structure of CGS-21680-bound A_{2A} (PDB ID:4UHR) [57] is available but it cannot accommodate G protein because of the narrower opening of the intracellular site of the receptor. Therefore, we aligned the two structures and observed that residues that interact with the ligand were aligned well. Consequently, CGS-21680 was transferred from 4UHR to 5G53. For modeling istradefylline-bound $A_{2A}R$, we used crystal structure of $A_{2A}R$ bound with ZM241385 (PDB ID:4EIY) [158] since it shares similar chemical structure with the target ligand. The Na^+ ion which is present in allosteric site of antagonist bound $A_{2A}R$ was kept in our construct from the original crystal structure of 4EIY. We performed induced fit docking, which is implemented in Schrodinger's Glide docking software [159], [160] to find best possible pose for istradefylline within the binding pocket using the hydrogen bond as a restraint with ASN253^{6.55} (superscripts denote Ballesteros-Weinstein numbering) [161] residue of the receptor.

As to the D_2R dimer, we used crystal structure of D_2R bound with risperodine (PDB ID:6CM4) [162] for modeling *apo* D_2R by removing T4 lysozyme which is fused to ICL3 of the receptor. The crystal structure of μ -opioid receptor complex with mini $G\alpha_i$ (PDB ID: 6DDF) [163] was used for modeling quinpirole-bound D_2R - $G\alpha_i$ complex since the crystal structure of D_2R - G_i complex was not available at the time of simulations were done. The sequence similarity between μ -opioid and D_2R was 35%. After it's release, we also compared backbone root mean square deviation (RMSD) between 6DDF and 6VMD (D_2R complex with G_i) and found to be 1Å. To find the best possible pose of the ligand we performed induced fit docking by using the hydrogen bond as a restraint with ASP114^{3.32} residue of the receptor. Herein, it is important to emphasize that D_2R has a long ICL3 which is not resolved in crystal structures. We modeled the missing regions as a loop by using the first 60 amino acids of ICL3 to prevent any restraint that may be introduced upon smaller length of loops.

For modeling dimers, we downloaded all the available crystal structures of GPCRs from G protein-coupled receptors database (GPCRdb) (245 structures were present as of April, 2018) to use as templates for generating corresponding homo- and heterodimer interfaces in the tetramer. First, the crystal mates were generated in Maestro, and then, receptor monomers were aligned to corresponding monomers of the templates. In particular, we

could get 38 templates with TM4/TM5 and 2 templates with TM6/TM6 dimerization interfaces, the list of which can be found in **Table 3.1**. In fact, various different orientations of the monomers in the dimer can satisfy those interfaces but we made use of the experimental data [64] to further pick up the biologically relevant interfaces. Accordingly, the crystal structures with PDB IDs: 5WIU [163] and 5O9H [164] were used to model TM6/TM6 and TM4/TM5 interfaces, respectively. Slight steric clashes in the models were alleviated by using Prime module of Maestro program.

Table 3.1: Templates used to model the interfaces in the tetramer.

Receptor	PDB (IDs)	Interface
A _{2A} R	5IU4	TM4-5
A _{2A} R	5IU7	TM4-5
A _{2A} R	5IU8	TM4-5
A _{2A} R	5IUA	TM4-5
A _{2A} R	5IUB	TM4-5
A _{2A} R	5JTB	TM4-5
A _{2A} R	5K2A	TM4-5
A _{2A} R	5K2B	TM4-5
A _{2A} R	5K2C	TM4-5
A _{2A} R	5K2D	TM4-5
A _{2A} R	5MZJ	TM4-5
A _{2A} R	5MZP	TM4-5
A _{2A} R	5N2R	TM4-5
A _{2A} R	5NLX	TM4-5
A _{2A} R	5NM2	TM4-5
A _{2A} R	5NM4	TM4-5
A _{2A} R	5OLG	TM4-5
A _{2A} R	5OLH	TM4-5
A _{2A} R	5OLO	TM4-5
A _{2A} R	5OLV	TM4-5
A _{2A} R	5OLZ	TM4-5
A _{2A} R	5OM1	TM4-5
A _{2A} R	5OM4	TM4-5
A _{2A} R	5UVI	TM4-5
A _{2A} R	5VRA	TM4-5
A _{2A} R	6AQF	TM4-5
A _{2A} R	4EIY	TM4-5
ADRB1	4GPO	TM4-5
ADRB1	5F8U	TM4-5
ADRB2	3D4S	TM4-5
C5AR1	5O9H	TM4-5
OPSD	2Z73	TM4-5
OPSD	3AYM	TM4-5

OPSD	3AYN	TM4-5
OPSD	4WW3	TM4-5
P2Y12	4NTJ	TM4-5
SMO	4JKV	TM4-5
SMO	4QIN	TM4-5
D4R	5WIU	TM6-6
D4R	5WIV	TM6-6

3.1.2. System preparation

Protein structure preparation was performed using “Protein Preparation Wizard” implemented in Maestro molecular modeling package [165]. Non-protein entities like BRIL, which is included in the crystal structure of ZM241385-bound A_{2A}R (PDB ID:4EIY), was removed and the missing region was filled by original residues of the receptor using “Cross-Link Protein” module implemented under Bioluminate [166] panel of Schrodinger software. Also, missing regions present either in receptors or in mini G proteins were filled similarly. Missing hydrogen atoms were added using PROPKA [167] to maintain protonation state of residues at pH:7.0. H-atoms were optimized and constrained energy minimization was carried out on the final tetramer model was obtained. The tetramer was aligned in the membrane using Orientations of Proteins in Membrane (OPM) server [168], [169] and then prepared using Membrane Bilayer builder program [170] implemented in CHARMM-GUI server [171]. The membrane was prepared to include 30% cholesterol and 70% sphingomyelin molecules to mimic lipid rafts, thus resulting with a total number of *ca* 500000 atoms. TIP3P [172] and CHARMM36m force fields [173] were used to model water molecules and protein in the system. Ligand parametrization was done using CGenFF, implemented in CHARMM-GUI server [174]. The output files were examined to check if penalty values associated with ligands exceeded the threshold level but they didn’t exceed.

3.1.3. Bivalent ligand construction

The bivalent ligand is composed of two pharmacophore groups, namely, istradefylline and CGS-21680 along with the linker. The linker was composed of an affinity-generating biphenyltriazole-moiety along with a spacer group. After modeling the tetramer, it was examined to determine possible regions on the pharmacophore groups to which the linker

was attached. Consequently, methoxy groups of istradefylline and carboxyl group of CGS-21680 were found to be accessible from the extracellular domain of the receptors and the linker was added to these regions using “build” panel of Schrodinger software. The final chemical composition of the linker was determined by considering possible synthesis opportunities of the bivalent ligand. In addition, to determine the optimum length of the linker we performed MD simulations on the structures having different linker lengths for 250 ns. Analysis of these trajectories showed that linker lengths that were shorter than 54Å were unstable and left the ligand binding pocket during the simulation. Therefore, the bivalent ligand having a linker length of 55Å was stable and thus used in further studies.

3.1.4. Molecular dynamics simulation

In this study, we performed both classical (cMD) and accelerated molecular dynamics (aMD) simulations using Nanoscale Molecular dynamics (NAMD) [175] package. For aMD, “dual-boost” was chosen to enhance sampling of the conformational space and performed by getting average potential energy values from cMD simulations. Specifically, the systems studied were heated from 0K to 310K within 1ns and pre-equilibrated using NVT (canonical) ensemble at 310K for 50 ns.

This was followed by 500ns cMD simulations which were performed using NPT (isothermal–isobaric) ensemble at 1atm and 310K to get average potential energies, from which the acceleration parameters were calculated. The last snapshots of these simulations was used as starting conformations in aMD simulations and aMD simulations were performed for 1 μ s where the backbone root mean square deviation (RMSD) of the protein fluctuated no more than 2Å as shown in the **Figure 4.2**. Nose-Hoover [176] and Parrinello-Rahman [177] coupling algorithms were used to maintain constant temperature and pressure, respectively throughout simulations. The integration step of 2fs was used in all simulations performed. In the dual-boost aMD, a dihedral potential and a potential boost were added to all the atoms in the systems. The dihedral and potential boost

acceleration parameters were calculated using following equations:

$$E_{dihed} = V_{dihed-avg} + \lambda * V_{dihed-avg} \quad (3.1)$$

$$\alpha_{dihed} = \lambda * V_{dihed-avg} / 5 \quad (3.2)$$

$$E_{total} = V_{total-avg} + 0.2 * N_{atoms}, \quad (3.3)$$

$$\alpha_{total} = 0.2 * N_{atoms}, \quad (3.4)$$

where N_{atoms} is the total number of atoms, and $V_{dihed-avg}$ and $V_{total-avg}$ are the average dihedral and potential energies obtained from cMD simulations. λ is an adjustable acceleration parameter. We used 0.3 for λ which was shown to perform effectively for GPCR systems [178]. The systems were simulated twice, each of which was simulated starting with a different initial velocity distribution.

The reweighted PMF profiles were obtained using cumulant expansion to the 2nd order; however, reweighting caused a large energetic noise as shown in **Figure 4.1**. Considering that the overall shape of free energy profiles is maintained in aMD simulations, unweighted data should provide an estimate of the free energy differences. Therefore, unweighted free energy profiles are presented in this study. [134], [179].

3.1.5. Root mean square fluctuation (RMSF)

Residue mean square fluctuation is described as the degree of flexibility index of a residue and was calculated by considering backbone atoms of each residue in the protein using the following RMSF formula:

$$RMSF = \sqrt{(1/N) \sum_j^N (x_i(j) - \bar{x}_i)^2} \quad (3.5)$$

$x_i(j)$ shows the i^{th} coordinates of $C\alpha$ atom in the structure of j^{th} model, and (\bar{x}_i) is the averaged position of i_{th} $C\alpha$ atom in all models.

3.1.6. Cross-correlation analysis

To calculate the extent of correlation among residues, dynamic cross-correlation maps were obtained. According to that, normalized co-variance values of residues were computed by considering C α atoms [180] and using the following equation:

$$DCC(i, j) = \frac{\langle \Delta r_i(t) \cdot \Delta r_j(t) \rangle_t}{\sqrt{\langle \|\Delta r_i(t)\|^2 \rangle_t} \sqrt{\langle \|\Delta r_j(t)\|^2 \rangle_t}} \quad (3.6)$$

where $r_i(t)$ and $r_j(t)$ represents the coordinates of the i^{th} and j^{th} atoms as a function of time t , $\langle \rangle$ indicates the time ensemble average, $\Delta r_i(t) = r_i(t) - \langle r_i(t) \rangle_t$ and $\Delta r_j(t) = r_j(t) - \langle r_j(t) \rangle_t$. The correlation range of residues lies within -1.0 to 1.0, where positive correlated displacement falls between 0-1.0 whereas anti-correlation displacement falls between -1.0-0.

3.1.7. Network construction

Network analysis was carried out using PSN-Ensemble Software [181] by considering side chain atoms of the residues in A_{2A}R receptor. As an input, two residues were provided as the source and the sink Glu169^{4.90} and Ile292^{7.57}, which were distant from each other on the protein and between which allosteric communication signal was generated. Herein, cross correlation matrices were used as inputs to calculate and weigh the shortest communication pathways used by the receptor in the presence and absence of the bivalent ligand.

3.1.8. Binding volume calculation

Ligand binding and G protein binding volumes were calculated using KvFinder plugin of pymol, which is based on geometrical grid-based method combined with space segmentation capabilities [182]. The space defined between molecular surfaces of the two probes is known as the cavity. The probe-in was set to 0.8Å and probe-out was set to 8.0Å with volume of cavity as 200Å³. For ligand binding pocket Ile16^{1.42}, Ala59^{2.57}, Thr68^{2.66}, His75^{3.23}, Val86^{3.34}, Phe183^{5.44}, Trp246^{6.48}, Phe258^{6.60} and His278^{7.43} were selected and Ser35^{1.61}, Gln38^{1.64}, Pro109^{3.57}, Ile200^{5.61}, Ala204^{5.65}, Gln226^{6.27} and Arg293^{7.58} residues were selected for calculating the volume of G-protein binding pocket.

3.2. System 2

3.2.1. Identification of intermediate active states from experimental phi-values

Phi-values are measured by temperature denaturation experiments in the absence and presence of a panel of ligands of different efficacies (partial agonists, full agonists and biased agonists). Ligands induce changes in the local environment of the mutated residues that modify their contribution to the overall thermal stability of the protein. Thus, ϕ -values measure how close the structure around the mutated residue in the intermediate active state resembles the inactive state. Each mutated position is then effectively behaving as a probe that reads out local conformational changes in the activation intermediate states. By comparing the results obtained with different ligands, we can detect the differences in the intermediate active states that they stabilize.

The ϕ -value for each mutant can be expressed as:

$$\phi = \frac{\Delta G_{(M,L)} - \Delta G_{(W,L)}}{\Delta G_{(M)} - \Delta G_{(W)}} \quad (3.7)$$

$$\phi = \frac{\Delta\Delta G_{(M-W)(L)}}{\Delta\Delta G_{(M-W)}} \approx \frac{\Delta Tm_{(L)}}{\Delta Tm} \quad (3.8)$$

where ΔG is the unfolding energy for the mutant (M) or wild type (W) in the absence or presence (L) of ligand, and ΔTm is the difference in apparent melting temperatures.

To calculate the phi-values of intermediate conformational states of G protein from GDP to GTP, melting temperatures of GDP bound and GTP bound G proteins were divided for each residue and a single value was obtained which gives the exact phi-value. For the calculation of ϕ -value for G protein system the equation was derived from the above mentioned equation and thus it is given as:

$$\phi_{\text{GTP-GDP}} = \frac{(\Delta Tm)_{\text{GTP}}}{(\Delta Tm)_{\text{GDP}}} \quad (3.9)$$

Therefore, in practice, ϕ -values (i.e. changes in stability upon ligand binding) can be approximated as the ratio between the difference in apparent melting temperatures between the mutant and the wild type, in the presence or absence of ligand.

3.2.2. Implementation of experimental ϕ -values in MD simulations

The experimental ϕ -values were used as restraints in MD simulations to generate models of the intermediate activation states. Such combination of experimental data and computational techniques has proved very powerful to characterize the heterogeneous structural ensembles in protein folding. To translate ϕ -values to structural parameters, the experimentally measured change in stability caused by a mutation was converted to a change in the number of contacts that the residue side chain at the mutated position makes with its surroundings. This way, in the MD simulations, ϕ -values were estimated at each time step simply from residue-residue contacts. Such definition based on side chains is appropriate, since experimental ϕ -values are primarily a measure of the loss of side chain contacts at the transition state, relative to the native state.

3.2.3. Methodology workflow

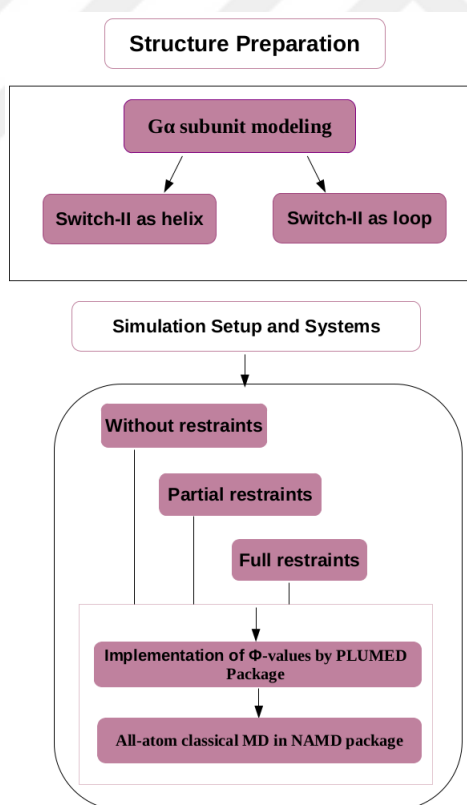


Figure 3.2: Schematic methodology workflow of system 2.

3.2.4. Structure preparation

GDP bound $G\alpha$ subunit structure was obtained from 3UMR pdb the only available structure which contains the GDP in its physiological form. [183], it is known for alpha subunit it is responsible for catalyzing the hydrolysis of GDP, therefore only the $G\alpha$ subunit was included in our study. All available crystal structures of G proteins have missing switch-II region, thus we model this missing region by using MODELLER and we have modeled it in two different conformations i.e; coiled-helix and loop and performed simulations on both systems individually. The all-atom classical MD simulations were performed with NAMD [175] using the PLUMED plugin [184], where the experimental ϕ -values were implemented as collective variables that measure the number of contacts between two residues.

Three different sets of simulations were taken in consideration (i) in the absence of restraints (ii) in the presence of partial ϕ -values as restraints applied to N-terminal region of $G\alpha$ subunit (iii) in the presence of restraints applied to whole protein. Here it is important to mention that the simulations performed with partially applied restraints was used as trail to test the strength of our proposed methodology, and the results were promising meaning that the restraints simulations could capture different dynamics as compared to classical MD without restraints. Considering these findings, we could confidently test the phi-value approach on the whole protein i.e; $G\alpha$ subunit.

To obtain the values of restraints from experimental ϕ -values, the melting temperature ΔT_m of GDP bound residue was divided by the melting temperature of GTP bound state residue and then the value was used as a restraint for MD simulation.

Domain name	CGN numbering	Residue number	Alanine Mutants	ΔT_m °C (GDP)	ΔT_m °C (GTP)	Residue type	Phi-Value
GTPase	G.HN.8	1	M1A	-0.2	-0.3	A12G	0.75
GTPase	G.HN.9	2	G2A	-0.3	-0.3	V13A	0.6667
GTPase	G.HN.10	3	C3A	-0.2	-0.2	E14A	0.666
GTPase	G.HN.11	4	T4A	-0.2	-0.3	R15A	0.75
GTPase	G.HN.12	5	L5A	-1	-1.1	S16A	0
GTPase	G.HN.22	6	S6A	-0.2	-0.2	K17A	0.5
GTPase	G.HN.23	7	A7G	-0.2	-0.1	M18A	0.333
GTPase	G.HN.24	8	E8A	-0.3	-0.6	I19A	0.25
GTPase	G.HN.25	9	D9A	-0.4	-0.5	D20A	-0.5
GTPase	G.HN.34	10	K10A	-0.3	-0.3	R21A	0.166
GTPase	G.HN.35	11	A11G	-0.2	-0.3	N22A	0.25
GTPase	G.HN.36	12	A12G	0.4	0.3	L23A	1
GTPase	G.HN.37	13	V13A	-0.3	-0.2	R24A	1
GTPase	G.HN.38	14	E14A	-0.3	-0.2	E25A	0
GTPase	G.HN.39	15	R15A	-0.4	-0.3	D26A	0

Figure 3.3: Melting temperatures of GDP and GTP bound systems with calculated phi-values colored green are given as an example of the proposed strategy.

3.2.5. Simulation protocol

3.2.5.1. Classical MD

The input files were prepared by CHARMM-GUI web server [171], the protein was solvated in the presence of water as a solvent, further the system was neutralized by 0.15M KCL concentration. Temperature was set as 310K and long range electrostatic and van der Waals interactions were considered within cutoff of 9Å, two sets/replica for each system were run separately.

3.2.5.2. Plumed MD (Restraint applied MD)

PLUMED is a simulation package for multipurpose, the phi-values which were near to zero 0 were considered as restraints to be applied in PLUMED package. NAMD simulation package can be synchronized with PLUMED so the restraints applied PLUMED output file was given as an input to run simulation in NAMD package.

3.3. System 3 Methodology Workflow

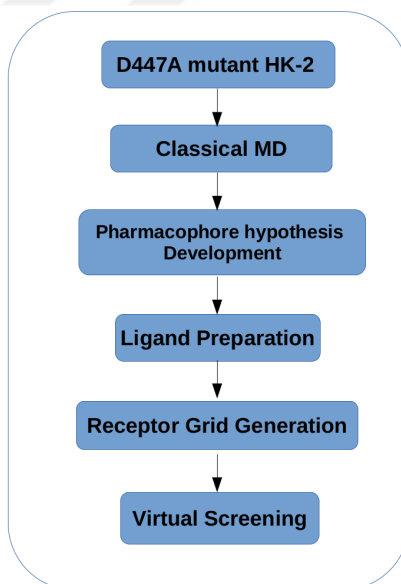


Figure 3.4: Schematic methodology workflow of system 3.

3.3.1. Structure preparation

The crystal structure of human HK2 in the dimeric form (PDB-ID: 2NZZ) was used in the MD simulations [185]. Prior to structure preparation, glucose and G6P were removed from the crystal structure to acquire the apo-state of HK2. The missing parts of the protein were modeled using “Crosslink Protein” panel of BioLuminate, which is implemented in Schrodinger suite [166]. The hydrogen atoms were added using “Protein preparation” wizard of Schrodinger suite [165]. The protonation states of amino acids were determined using PROPKA at pH 7.0 [167]. The D447A mutant of the NTD (N terminal domain) was introduced using the “Build” panel of Schrodinger suite [165].

3.3.2. MD simulation setup

The MD simulation files were prepared using CHARMM-GUI web server [171]. The protein was modeled by CHARMM36 force field whereas water molecules were modeled using TIP3P model [186], [187]. The systems were neutralized with salt concentration of 0.15 M KCl, and simulations were carried out using GROMACS software [188]. Periodic boundary conditions were applied in all directions and simulation time-step was set to 2fs. The temperature and pressure were maintained at 310 K and 1 atm using Nose- Hoover and Parrinello-Rahman coupling algorithms, respectively to achieve canonical ensemble [176], [177], [189]. Particle mesh Ewald (PME) method was used to compute long-range electrostatic forces and Van der Waals forces were treated with 9Å cut-off [190]. The systems were simulated for 200 ns, where the backbone RMSD of the protein fluctuated no more than 2Å. All coordinates were saved at 10ps intervals for further analysis. Two separate simulations were performed, where each was started with different initial velocity distribution.

3.3.3. Analysis of trajectories

Principal component analysis was carried out on the apo-state of the wild-type and mutant proteins to explore dominant motions in the systems. First, the resultant trajectories of each system were aligned with respect to backbone $C\alpha$ atom of the corresponding initial structure. Computation and diagonalization of covariance matrices were done by using “gmx covar” module of GROMACS, and the “gmx anaeig” module of GROMACS was

used to obtain eigenvectors and eigenvalues from diagonalized covariance matrices [188]. The first three vectors, which collectively captured more than 50% of the overall global motion, were presented for each monomer separately. The distance measurement was done using the $C\alpha$ atoms of T88 and T232, which are located in the small- and large-subdomains of the NTD, respectively. On the other hand, the angle measurement was done using the $C\alpha$ atoms of T389, R468 and A839 of the NTD, linker helix- α_{13} , and CTD (C terminal domain), respectively.

3.3.4. Hexokinase 2 virtual screening

3.3.4.1. Ligand preparation

A list of 2446 FDA approved Drugs was downloaded from DrugBank, 3D coordinates of all the ligands were generated by LigPrep module implemented in Schrodinger. Ligand minimization was first carried out to assign the correct bond orders and bond angles with OPLS3e force field. Next their biological ionization states were determined by Epik option in the same panel of LigPrep [191].

3.3.4.2. Ligand filtering

In this step the selected set of ligands were filtered on the basis of MW molecular weight and rotatable bonds, MW was set at 500KDa and 15 rotatable bonds.

3.3.4.3. Pharmacophore hypothesis

Pharmacophore hypothesis was generated for HK2 to specify the ligand binding site positioned at 447 and surrounding residues. This task was performed by Develop Pharmacophore hypothesis panel in Schrodinger. Six pharmacological features are taken in consideration with the pharmacophore hypothesis, listed as [A] hydrogen bond acceptor, [D] hydrogen bond donor, [H] hydrophobic group, [N] negatively ionisable, [P] positively ionisable and [R] aromatic ring.

3.3.4.4. Receptor grid generation

In the next step of receptor grid generation the software searched for a favorable interaction between ligands and the protein/receptor. The grid generation represents the van der Waals

and electrostatic properties of selected region of the protein.

3.3.4.5. Virtual screening

All the selected ligands were subjected to flexible docking studies in Glide [160], through (SP) standard precision implemented in Schrodinger. The filtering options for ligands were selected on the basis of QikProp to retain the correct ionization state of ligands and 3D geometry option was used to obtain stereochemical information of ligands.



CHAPTER 4

4. RESULTS AND DISCUSSION

4.1. Heterobivalent Ligand Targeting A_{2A}R Dimer

4.1.1. Reweighting of accelerated molecular dynamics simulation

Accelerated MD improves the sampling efficiency of biomolecular conformations, but it is necessary to recover the original free energy landscape of the system. For this the boost potential of each frame the probability of reaction coordinates can be reweighted to recover the canonical ensemble distribution of the system here shown for χ_2 angle of Trp246^{6,48} in **Figure 4.1**. To reduce the noise generated by huge data points the cumulant expansion approximation is better considered to calculate the ensemble-averaged reweighting [134], the cumulant expansion can be obtained by the given equation;

$$\langle e^{\beta\Delta V} \rangle = \exp \left\{ \sum_{k=1}^{\infty} \frac{\beta^k}{k!} C_k \right\} \quad (4.1)$$

$\langle e^{\beta\Delta V} \rangle$ ensemble-averaged reweighting factor, C_2 is the cumulant expansion to the second order where $(\sigma_{\Delta V}^2)$ in equation 2 represents the standard deviation of boost potential ΔV ;

$$C_2 = \langle \Delta V^2 \rangle - \langle \Delta V \rangle^2 = \sigma_{\Delta V}^2 \quad (4.2)$$

The free energy can then be derived from cumulant expansion as:

$$F(A_j) = F^*(A_j) - \frac{1}{\beta} \sum_{k=1}^{\infty} \frac{\beta^k}{k!} C_k + F_c \quad (4.3)$$

where $F^*(A_j)$ is the modified free energy surface sampled in the aMD simulation and the

constant $F_c = (1/\beta) \ln \sum_{j=1}^M \langle e^{\beta \Delta V(r)} \rangle_j$ where M is the number of bins and $\beta = 1/k_B T$ for simulation found in j_{th} bin.

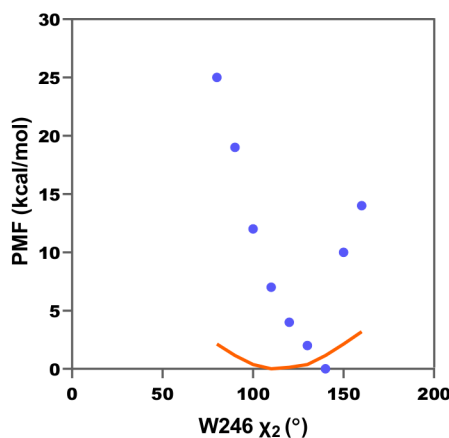


Figure 4.1: PMF profiles of the χ_2 Trp246^{6,48} obtained from reweighting based on cumulant expansion to the 2nd order (blue) and calculated from classical MD simulation (orange) for comparison.

We checked if unweighted data can be used to discuss changes by comparing unweighted PMF (potential of mean force) profiles to those obtained from cMD. Herein, it is important to point out that cMD simulations might not be relevant for the comparison since these simulations were short and performed to get average dihedral and potential energy values required for aMD simulations. So, it is likely that some energy minima might not have been sampled. We started comparison with χ_2 angle of Tyr288^{7,53} of antagonist-bound A_{2A}R (**Figure 4.7B**) as it displayed two peaks, thus presenting a challenging reaction coordinate. Interestingly, the two minima sampled in cMD and aMD simulations were similar in spite of energy difference between them-being higher in cMD as shown below. This is -in fact- in correspondence with the theory of aMD which states that the barrier that separates energy minima is decreased in aMD simulations. However, it is still true that although the shape of the energy profile is conserved the probabilities of these regions of energy minima might be different. For our purposes, the values of energy minima are more important as they correspond to most possible conformations of the target residues as indicated in the results. Also, the minima are different between without linker and linker systems which make it possible to compare them on the plots. Considering that the barriers that separate energy minima decrease while the overall shape of the energy profiles is conserved in aMD simulations, unweighted data can be considered as an estimate of the original free energy profile.

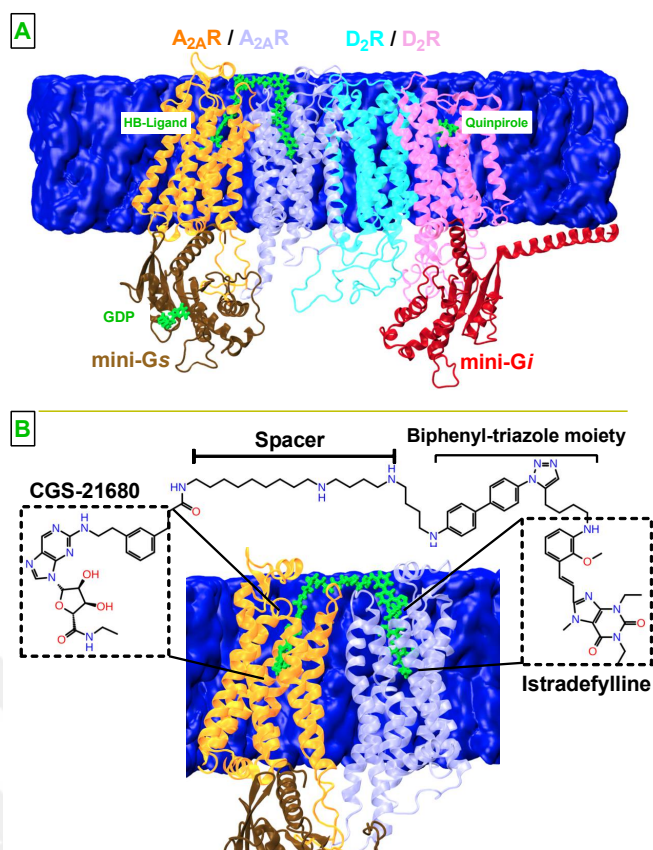


Figure 4.2: **A.** Schematic representation of A_{2A}R/D₂R tetramer with bivalent ligand. The membrane is shown in blue. A_{2A}R protomers and mini-G α_s are shown in orange, purple and brown, respectively and new cartoon representation whereas bivalent ligand is shown in green and licorice representation. D₂R protomers and mini-G α_i are shown in pink, cyan, and red, respectively. **B.** The pharmacophore groups and the linker are shown in detail together with A_{2A}R dimer.

4.1.2. Optimum linker length is required for stable binding of bivalent ligand to A_{2A}R dimer

The root mean square deviation (RMSD) of whole tetramer as well as individual protomer was calculated to show that the system have reached to an equilibrium as depicted in 4.3. As explained in the Methods section tetramer was modeled to consist of dimers of A_{2A}R and D₂R along with their corresponding effectors, mini-G α_s and mini-G α_i , respectively, to be in line with experimental data [47].

In a follow-up study of Navarro *et.al* [64], homodimer interfaces were shown to consist of TM6 whereas heterodimer interface was shown to consist of TM4/TM5. According to the same study, it was also suggested that protomers which were bound to G protein should be located on the periphery of the quaternary structure to prevent clashes between two

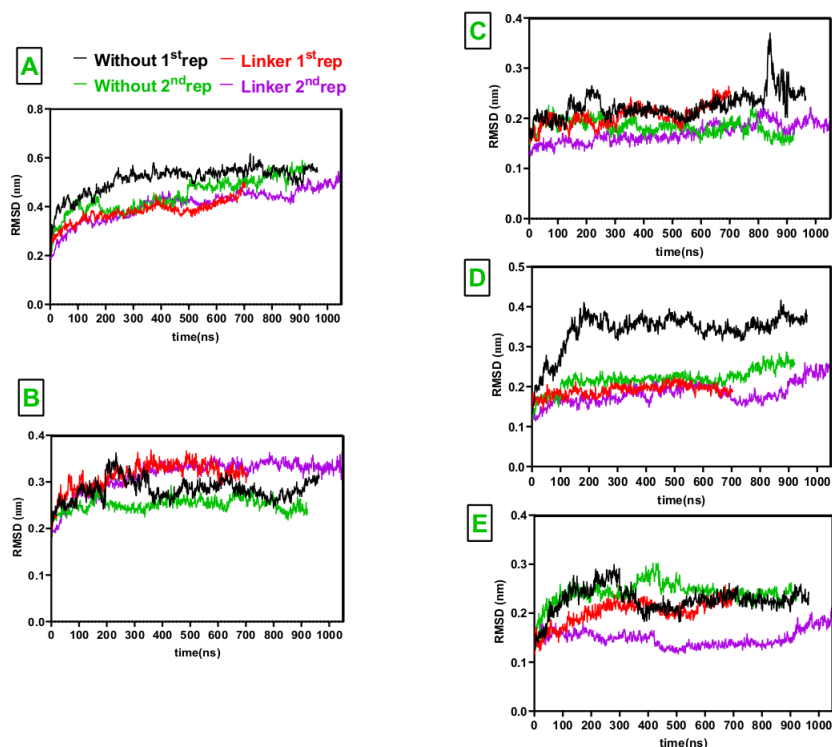


Figure 4.3: (RMSD) timeline plots. **A.** Tetramer. **B.** Agonist bound D₂R. **C.** Apo D₂R. **D.** Antagonist bound A_{2A}R. **E.** Agonist bound A_{2A}R.

G proteins whereas *apo* D₂R and antagonist-bound A_{2A}R were found in the middle. To identify possible attachment points of pharmacophores with the linker, crystal structures of CGS-21680 (PDB ID:4UHR) [57] and ZM241385-bound A_{2A}R (PDB ID:4EIY) [158] were used. Examination of these structures revealed that methoxy groups of istradefylline and carboxyl group of CGS-21680 were accessible from the extracellular domain of the receptors.

Therefore, these two points were linked to each other using an affinity generating biphenyltriazole-moiety and spacer as shown in **Figure 4.2**. To determine the optimum linker length we performed classical molecular dynamics (cMD) simulations with bivalent ligands having different number of spacer units. Our results showed that bivalent ligand which was shorter than 54Å could not stably bind to the dimer which was in correspondence with a study of Hubner *et.al* [63]. Here, it is important to emphasize that total distance between the two attachment points was -in fact- smaller than 54Å; however, the path from one attachment point to the other was blocked by ECLs so that longer linker was preferred.

4.1.3. Extra- and intracellular domains are stabilized and global dynamics is restricted by the bivalent ligand in A_{2A}R dimer

The bivalent ligand was built to composed of an A_{2A}R antagonist (istradefylline) and an agonist (CGS-21680), where affinity-generating biphenyltriazole-moiety was connected to istradefylline. Comparative analyses of the trajectories showed that the linker made contacts with extracellular loop (ECL) 1 and 3, thus decreasing fluctuation at these regions in antagonist-bound A_{2A}R as shown in **Figure 4.4**. Specifically, side chains of Thr68^{2.56} and Asp261^{6.63}/Ser263^{6.65}, which are located on ECL1 and ECL3, respectively made hydrogen bonds with nitrogen atoms located on the linker. Moreover, the linker also decreased conformational fluctuation at intracellular loop (ICL) 2, TM1, TM6 and TM7, both of which were included at homodimer interface. On the other hand the linker did not impact fluctuations in agonist-bound A_{2A}R/mini-Gα_s complex. Here, it is important to emphasize that the impact of the linker on agonist-bound A_{2A}R/mini-Gα_s was reflected in the number of contacts made between the ligand and certain residues of ligand binding pocket of receptor. It was increased by 50% in agonist-bound A_{2A}R/mini-Gα_s while remained the same in antagonist-bound receptor. Specifically, contacts between Trp246^{6.48} and Ile274^{7.39}, which have been implicated in agonist binding, and CGS-21680 increased in the presence of the linker. Having observed that the linker modulated fluctuation pattern of residues **Figure 4.4** we also set out to investigate its impact on the global dynamics of A_{2A}R protomers. Results showed that global motion of the receptors was confined in the presence of the ligand as they sampled narrower region on the conformational space, which was revealed by the first and the second eigenvectors of the systems shown in 2D-PCA (Principle component analysis) plots (**See Figure 4.4**).

4.1.4. Bivalent ligand modulates volume of ligand binding pocket and intracellular domain in A_{2A}R dimer

We set out to investigate possible impact of the linker on structure of ligand binding pocket as the number of contacts made between CGS-21680 and receptor increased in the presence of the linker. Interestingly, we observed that volume of the ligand binding pocket was decreased by ~15% in CGS-21680- bound A_{2A}R while this change was not significant for istradefylline-bound A_{2A}R protomer. Moreover, we also measured the volume of the G protein binding site of A_{2A}R protomers to investigate if the change observed in the ligand

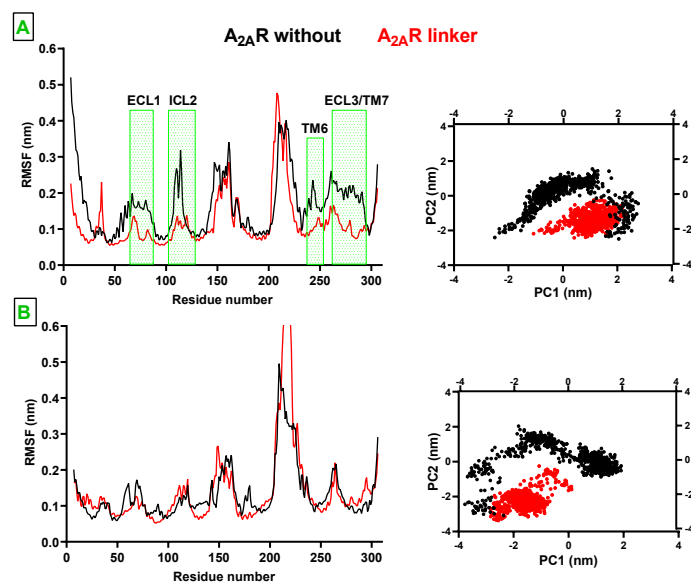


Figure 4.4: RMSF and 2D-(PCA) profiles pertaining to **A.** antagonist-bound A_{2A}R and **B.** agonist-bound A_{2A}R/mini-Gα_s. The regions that showed difference between without and with linker systems are indicated with green rectangles.

binding pocket was transmitted to the intracellular domain of the receptors. Indeed, the volume of intracellular domain was increased by ~10% in CGS-21680-bound A_{2A}R/mini-Gα_s. With that, agonist-bound A_{2A}R could achieve such volume range, which was shown to be sampled by active receptors, [192] in the presence of the linker. On the other hand, the volume of the intracellular domain in the absence of the linker was closer to the value which was sampled by the receptor found in an intermediate state between the inactive and active receptor [192]. This change was also reflected in the ionic lock distances as well such that longer distance was observed in agonist-bound A_{2A}R in the presence of the linker, thus having wider intracellular domain. Here, it is important to emphasize that the volume of G protein binding site in istradefylline-bound A_{2A}R decreased despite the absence of a significant change in the volume of ligand binding pocket (See **Figure 4.5**). Since istradefylline-bound A_{2A}R was placed between CGS-21680-bound A_{2A}R/mini-Gα_s and *apo* D₂R in the tetramer, expansion of the intracellular domain of neighboring A_{2A}R might confine the space available for antagonist-bound A_{2A}R protomer. Here in, we need to elaborate that expansion of intracellular domain of agonist-bound A_{2A}R also impacted interactions between the receptor and mini-Gα_s: Coulomb energy decreased (became more favorable) by 10% in the presence of the linker, whereas the change in Lennard Jones energy was not significant between without and linker systems.

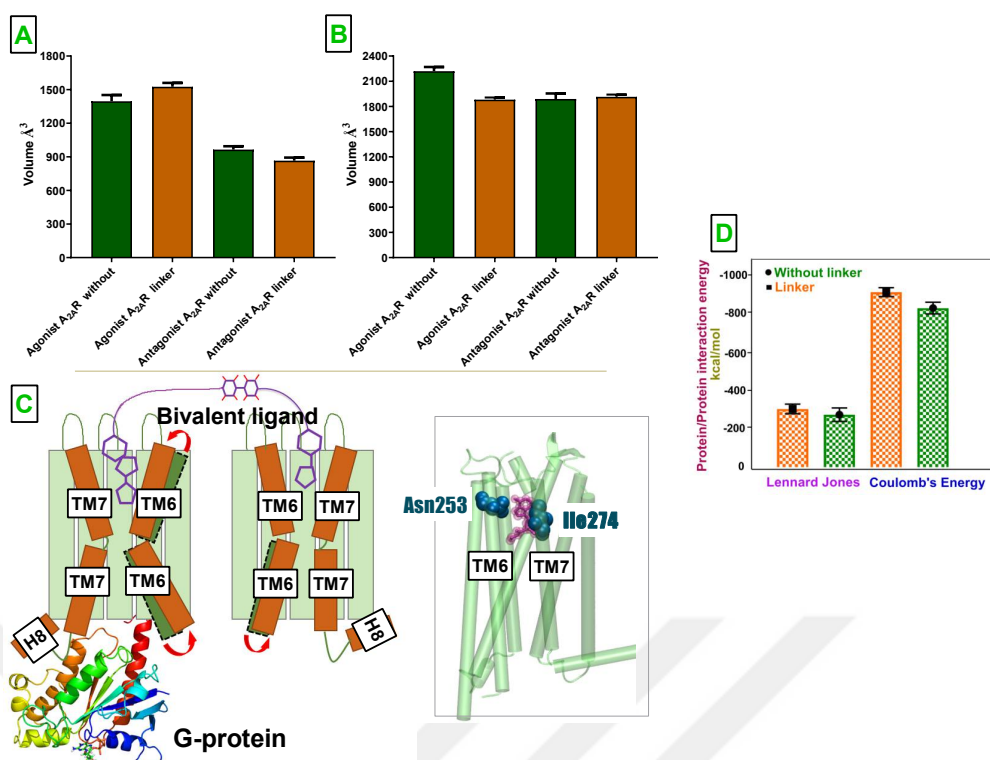


Figure 4.5: Changes in volumes of **A.** ligand binding pocket and **B.** G protein binding domain between without and linker systems **C.** Schematic representation of conformational changes adapted by A_{2A}R receptors in the presence of the linker. TM 6 and 7 positions in the absence (green) and presence (orange) of the linker is shown. Mini-G α_s bound to agonist-bound A_{2A}R is shown in new cartoon representation. Right hand side Ile274^{7,39} interaction with ligand emerge upon adding the linker. **D.** Protein-protein interaction energies measured between G protein and the receptors are also given.

4.1.5. Conformational preferences of microswitches and residue correlations are modulated by the bivalent ligand in A_{2A}R dimer

Reciprocal changes observed in the volumes of both ligand binding pocket and G protein binding site of CGS-21680-bound A_{2A}R in the presence of the linker suggested alterations in correlation of interactions at certain TM domains. To corroborate such impact, we calculated dynamics cross correlation matrices for the systems studied (See Figure 4.6) and showed that strength of correlated motion between TM1 and TM3, TM1 and TM7, TM5 and TM6 as well as TM6 and TM7 was increased in the presence of the linker. However, the strength of anti-correlated motion was increased at TM3/TM6, all of which were occurred upon receptor activation. That is to say, increment in strength of anti-correlated and correlated motion in TM3/TM6 and TM5/TM6 pairs, respectively, triggered wider

opening of the intracellular domain of the receptor, which was also reflected in the volume of G protein binding site at CGS-21680-bound A_{2A}R. This observation was also in line with a recent study where shrinking of the volume of orthosteric ligand binding site and expansion of the intracellular domain of the receptor were shown to be associated with similar changes occurring in TM3, TM6 and TM7 [193].

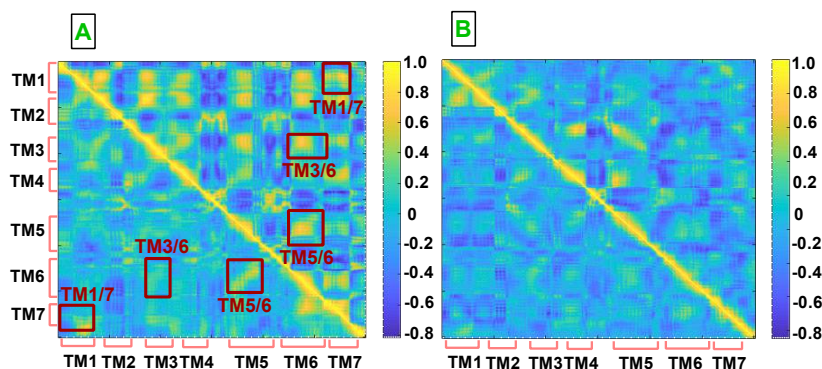


Figure 4.6: Dynamic cross correlation maps (DCCM) **A.** agonist-bound A_{2A}R/mini-G α_s and **B.** antagonist-bound A_{2A}R. The upper triangles in A and B correspond to DCCM of the system with linker whereas the lower triangles correspond to those without linker.

Moreover, these rearrangements also led to stabilization of side chains of certain residues at istradefylline-bound A_{2A}R in the presence of the linker. Importantly, Trp246^{6.48} adopted χ_2 angle values which provided close packing of the residue against the sodium ion as seen in the crystal structure of inactive A_{2A}R (PDB ID:4EIY). Also, hydroxyl group of Tyr288^{7.53}, which is involved in G protein binding, is positioned farther from the G protein binding site (See Fig 4.7) than in without linker system and crystal structure of inactive A_{2A}R (PDB ID: 4UHR) [194].

Besides residue correlations, the bivalent ligand also impacted allosteric interaction network in agonist-bound protomer in A_{2A}R dimer.

Specifically, TM5 emerged as one of the most significant domain that participated in the interaction network in the presence of the linker. Recalling coupling between orthosteric ligand binding pocket and intracellular domain in agonist-bound A_{2A}R, it can be said that the result was in correspondence with a study where allosteric coupling between extra- and intracellular domain of the receptor was shown to be mediated through TM5 [195]. Moreover, Asn253^{6.55} residue, which is one of the key residues of the ligand binding pocket that anchors the exocyclic amine of the ligand's central core, participated to allosteric interaction network in the presence of the linker and interestingly, TM3 didn't contribute

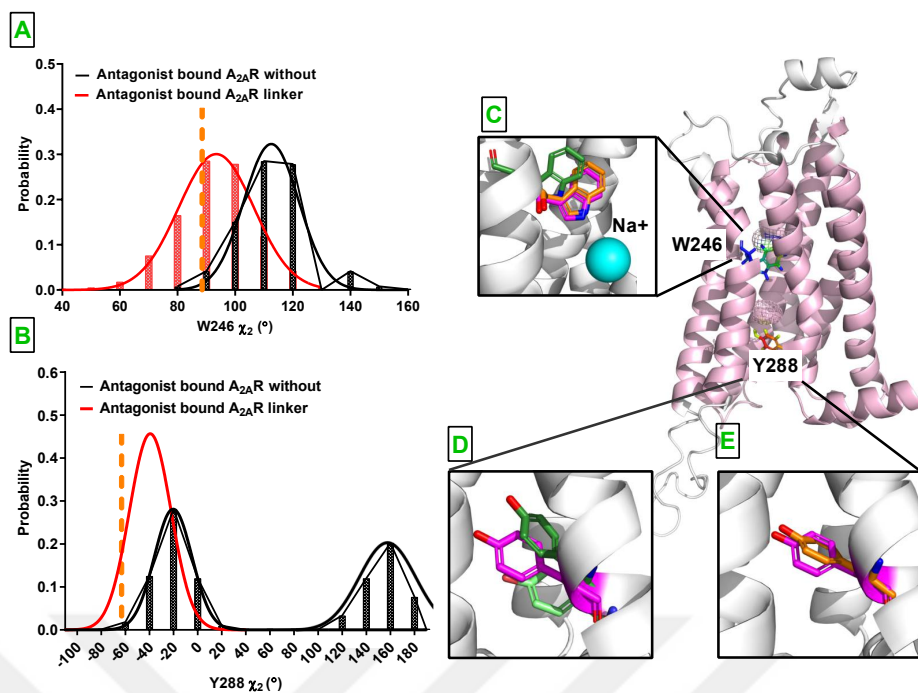


Figure 4.7: Angle probability distribution of Trp246 and Tyr288. **A.** χ_2 of Trp246^{6,48} is presented for antagonist-bound $A_{2A}R$ in the absence (black) and presence (red) of the linker, respectively. Dashed line indicate the reference value from crystal structure of antagonist-bound $A_{2A}R$ (PDB ID:4EIY). **B.** The same as in A but for χ_2 of Tyr288^{7,53}. Comparison of orientation of **C.** Trp246^{6,48} in the absence (green) and presence (orange) of the bivalent ligand. Orientations of Tyr288^{7,53} in the **D** absence of linker (dark and light green) and **E.** presence of linker (orange). The orientation of Trp^{6,48} and Tyr^{7,53} in the crystal structure of antagonist-bound receptor is shown in pink.

in the presence of the linker (See Figure 4.8). We also investigated interactions between $A_{2A}R$ dimer and membrane and showed that Trp129^{4,50}, which was shown to be part of the cholesterol consensus motif, of agonist-bound $A_{2A}R$ made close contacts with a cholesterol molecule in the presence of the linker in one of the replicates (See Figure 4.9) [196]. Therein, the cholesterol moved by almost 5Å throughout the trajectory from its original position and eventually contacted to the receptor, whereas no such interaction was observed for the antagonist-bound $A_{2A}R$ neither in the presence nor the absence of the linker. Lastly, we also compared pattern of water channels formed within the receptors between without and linker systems using the VolMap plugin of VMD. In accordance with a study which showed a correlation between receptor activation and continuous water channel formation [192], we observed a continuous water channel in agonist-bound $A_{2A}R$ whereas it was disrupted in antagonist-bound $A_{2A}R$ (See Figure 4.10). Interestingly, continuous water channel could be also maintained in the presence of the linker

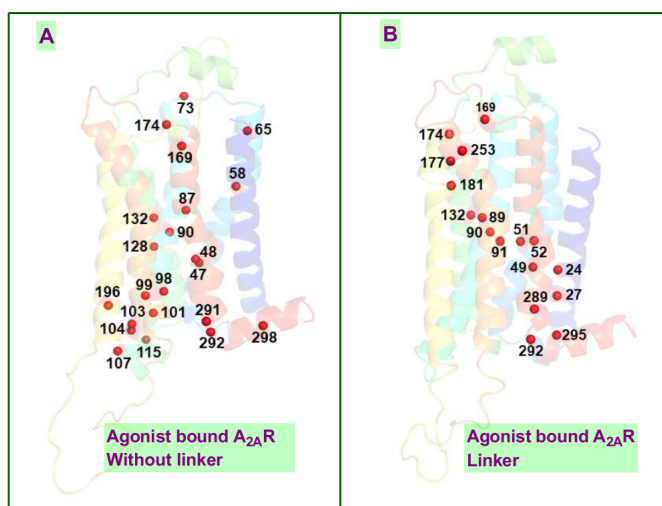


Figure 4.8: Allosteric interaction network pathway A_{2A}R.

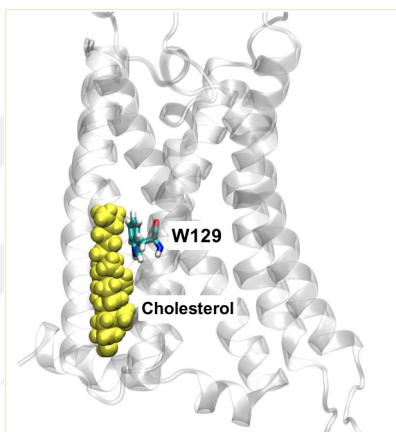


Figure 4.9: Cholesterol binding residue shown for Agonist bound A_{2A}R.

in agonist-bound receptor; however, water density was decreased in the region between the extracellular domain and orthosteric ligand binding pocket as observed in antagonist-bound A_{2A}R. Analysis of the trajectories showed that the linker precluded water flow from extracellular side to the inside of the receptor. Consequently, water molecules could be clustered only around Na⁺ ion in antagonist-bound receptor.

4.1.6. D₂R dimer resembles an asymmetric unit in the presence of the linker

The tetramer is composed of pairs of A_{2A}R and D₂R, where the homodimer interfaces are formed by TM6 and the heterodimer interface is formed by TM4 and TM5. In accordance with experimental data, one of the D₂R protomers was modeled as *apo* and the other protomer was modeled to bound with D₂R agonist, quinpirole and mini-Gα₁. The results showed that ICL2, albeit to a lesser extent, and ICL3 were stabilized in *apo* D₂R in the

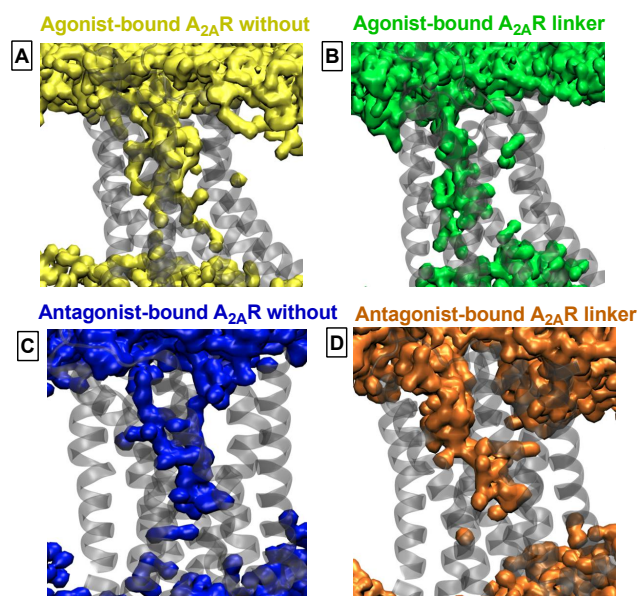


Figure 4.10: Internal Water Channel formation. **A.** Agonist-bound $A_{2A}R$, and **B.** Antagonist-bound $A_{2A}R$ in the absence of the linker, **C.** Agonist-bound $A_{2A}R$, and **D.** Antagonist-bound $A_{2A}R$ in the presence of the linker. Water occupancy maps are computed using the VolMap tool of VMD.

presence of the linker while only a small part of the ICL3 could be stabilized in agonist-bound D_2R /mini- $G\alpha_i$. Moreover, similar to $A_{2A}R$ dimer, the linker also confined global dynamics of the protomers as revealed by principal component analysis (PCA) as shown in **Figure 4.11**. Interestingly, the linker also modulated residue correlations at *apo* D_2R despite not being directly bound to the receptor. Specifically, an increase was observed in the strength of correlated motion in residues located on TM3 and TM4, two of which construct the heterodimer interface with antagonist-bound $A_{2A}R$. Moreover, an increase was also observed in the correlated motion of TM3 and TM6 as well as TM3 and TM4 in the presence of the linker in *apo* D_2R indicating that they tend to get closer to each other as shown in **(Figure 4.12)**.

Herein, it is also important to emphasize that the intracellular domain of *apo* D_2R was modulated by the linker as the receptor could sample a wider range of ionic lock distances which are calculated between Arg132^{3.50} ($C\gamma$) and Glu368^{6.30} ($C\delta$) atoms [197] in the absence of the linker whereas it could only sample shorter distances in the presence of the linker. In light of these results, it can be concluded that *apo* D_2R resembles dynamics of an inactive receptor within the D_2R dimer.

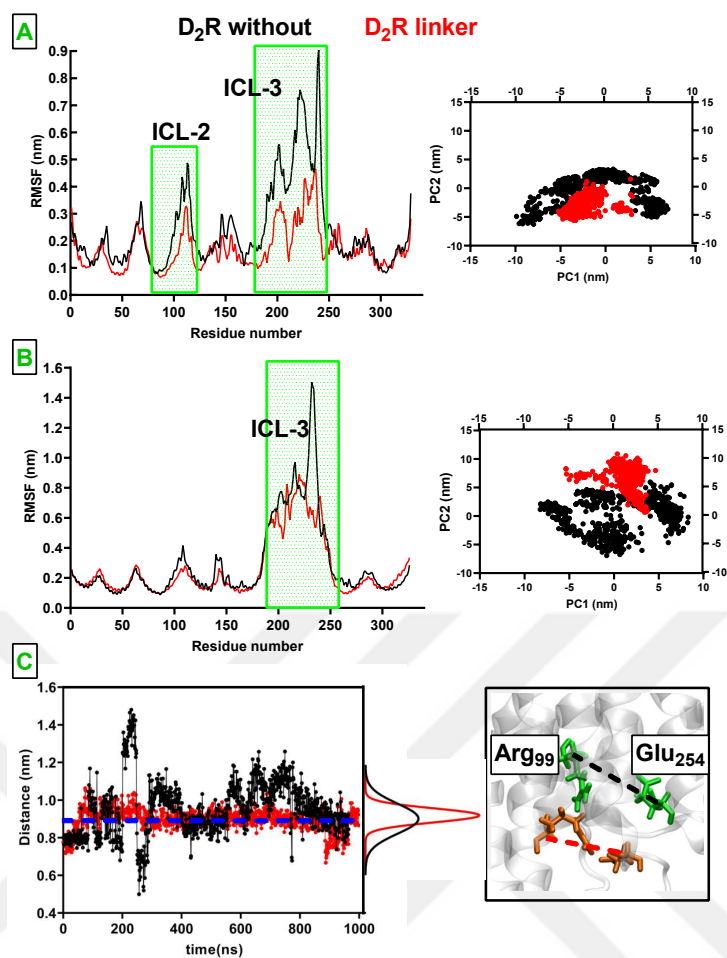


Figure 4.11: (RMSF) and 2D-PCA profiles pertaining to **A.** *apo* D₂R and **B.** agonist-bound D₂R/mini-Gα₁. The regions that showed difference between absence and presence of the linker are indicated with green rectangles. **C.** Time-line ionic lock distance sampled by *apo* D₂R in the presence (red) and absence (black) of the linker. The ionic lock distance measured in the crystal structure of eticlopride-bound D₂R (PDB ID:3PBL) is indicated with blue dash line.

4.2. Experimental Phi-values And G Protein System Results

G protein subtype Gα_i has been studied. The G protein is a complex of three subunits named as alpha, beta and gamma as shown in **4.13**. Among them Gα subunit is responsible for the hydrolysis of GDP along this it served as the point of attachment to the receptor which makes this subunit as a key player in G protein catalysis. Structurally the Gα has two domains RAS-like and Helical domain and the nucleotide binding site resides in the cleft of these two domains, for the nucleotide exchange the two domains open and releases the GDP and exchanged it with GTP shown in **Figure 1.4**. But the exact mechanism of this event is elusive, the Gα subunit contains helix-5 (H5) responsible for binding to receptor

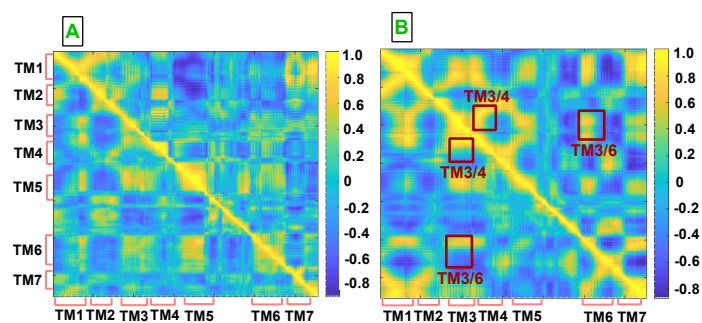


Figure 4.12: Dynamic cross correlation maps (DCCM) are presented for **A.** agonist-bound $D_2R/mini-G\alpha_i$ and **B.** *apo* D_2R . The upper diagonal in A and B correspond to DCCM of the system with linker whereas the lower diagonal correspond to those in the absence of linker.

and it makes interaction with partner β and γ subunits.

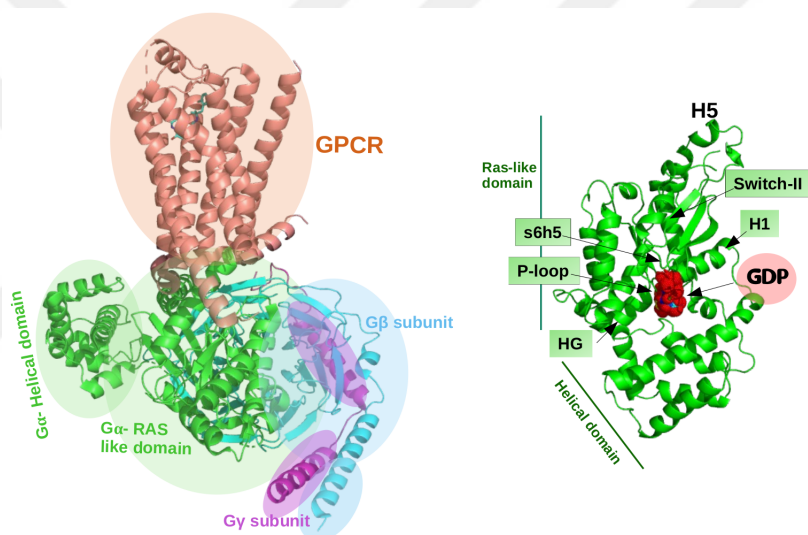


Figure 4.13: The schematic representation of GPCR bound to its trimeric G protein has been shown in left side and the structure of $G\alpha$ subunit studied is given on right hand side.

The experimentally derived phi-values were used as restraints in the MD simulations which has been explained in the method **section 3.2**, the values which ranges between 0-1 were retrieved and when mapped on the structure by residue number it has been seen the major regions were clustered around the nucleotide binding site depicted in **Figure 4.14**. This gives us an insight to the significance of these phi-values along with the results of simulations.

We aimed to examine the dynamics of G protein activation along the GDP dissociation which is a process with a very slow kinetics ($\sim 10^{-3} \text{min}^{-1}$), for better understanding of GPCR mediated signal transduction. The binding of an agonist to the receptor accelerates

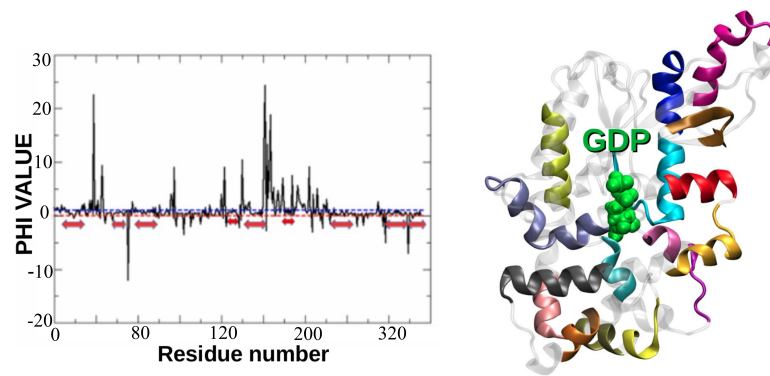


Figure 4.14: Phi-values are plotted and the red arrows are given to the range of values between 0 to 1, where residue number is given on x-axis and phi-value on y-axis. To the right hand side all these values are mapped on the $G\alpha$ highlighted in different colors where the GDP is shown in green van der Waals representation.

the process of GDP release from the G protein permitting the binding of GTP instead to turn on the receptor, then turning off the receptor by hydrolyzing the bound GTP, the whole process is kinetically controlled. The biochemical and structural data have shown the interaction of $G\alpha$ to the activated receptor is mainly promoted by the α -H5 of C-terminal, the rotation and translational movement of α -H5 towards the receptor poses unstability to the surrounding residues. The movement of α -H5 is towards the β 6 strand within the $G\alpha$ subunit which causes conformational changes in the regions of switch-I and II, thus distorting the nucleotide binding pocket, this might give an insight how signal from receptor extracellular site is conveyed to the 30Å far nucleotide binding site intracellular site, it is distorted and release the GDP [198].

Historically, most attention in this process has focused on conformational rearrangements pertaining to GPCR-binding helix-5 (H5) and displacement of the helical domain relative to the Ras-like domain. On the other hand, comparison of crystal structures which are representative of inactive, intermediate, and active state of G protein, revealed remarkable differences in positioning of Switch II domain as well. Specifically, side chain of catalytically important glutamine residue, which is located near the N-terminus of Switch-II, is positioned 7Å closer to $P\beta$ of the nucleotide in GTP-bound $G\alpha$ (active state) than GDP-bound (inactive state) and receptor-bound heterotrimeric G protein (intermediate state). Moreover, various experimental studies showed that kinetics of Switch-II impacted prior to GDP release, which is an indication of increased conformational disorder. Interestingly, Switch-II adopts α -helical conformation in these crystal structures displaying no

conformational heterogeneity.

In the light of computational studies new interactions and interacting regions of $G\alpha$ have been identified which are equally important in the whole mechanism, such of these findings highlighted the significance of switch-II. For activation the RAS-like domain and helical domains tend to separate by permitting disruption of multiple inter-domain interactions, one such crucial interactions are a salt bridge between Glu49 of P-loop and Arg-183 of switch-I.

In a reported studies the major factor that trigger GDP release are listed as [199].

1. RAS-like and helical domain gets apart
2. Distance between GDP and s6h5 loop (loop between β -strand 6 and helix-5 according to CGN numbering scheme) tends to increase
3. HG helix-G and P-loop distance gets enlarge
4. Tilting and translation the upward movement of H5 helix-5 occur
5. Switch-II shifts toward GDP binding site and causes a disruption in the surrounding site

4.2.1. Allosteric network mediating GDP dissociation upon receptor activation

Upon receptor mediated activation the prominent event occurred is conformational changes in helix-5 (H5) but there are evidences which state that there must be some other factors involved. The tilting of H5 brings some conformational changes to the s6h5 loop which is in contact with nucleobase of GDP and hence contacts between TCAT motif and Ala331 are broken. Further this distortion affects the H1 (helix-1) which is strongly coupled with s6h5 loop moves apart, and this results in disorderness in P-loop and this loop moves apart from H1. Such that the nearby residues including Glu-49, Ser-50, Gly-51, Lys-52 and Ser-53 are impacted and all this leads to the final breakage of salt bridge Glu-49 & Arg-183 of GTPase and helical domain. In this way the inter-domain separation occurs and both helical and GTPase domains moves apart under the influence of receptor activation starting from extracellular site of H5 all the way down to inter-domain opening and GDP-release event [199].

4.2.2. Comparison of our results with G protein activation determined reaction coordinates

To test our proposed methodology we have first compared all the above mentioned significant factors reported in literature [199] involved in activation of G-protein with our simulations, and our results were found to be in line as depicted in **Figure 4.15**. Suggesting that in the presence of restraints we could able to attain such conformation of $G\alpha$ subunit which resembles to all the events necessary for activation process.

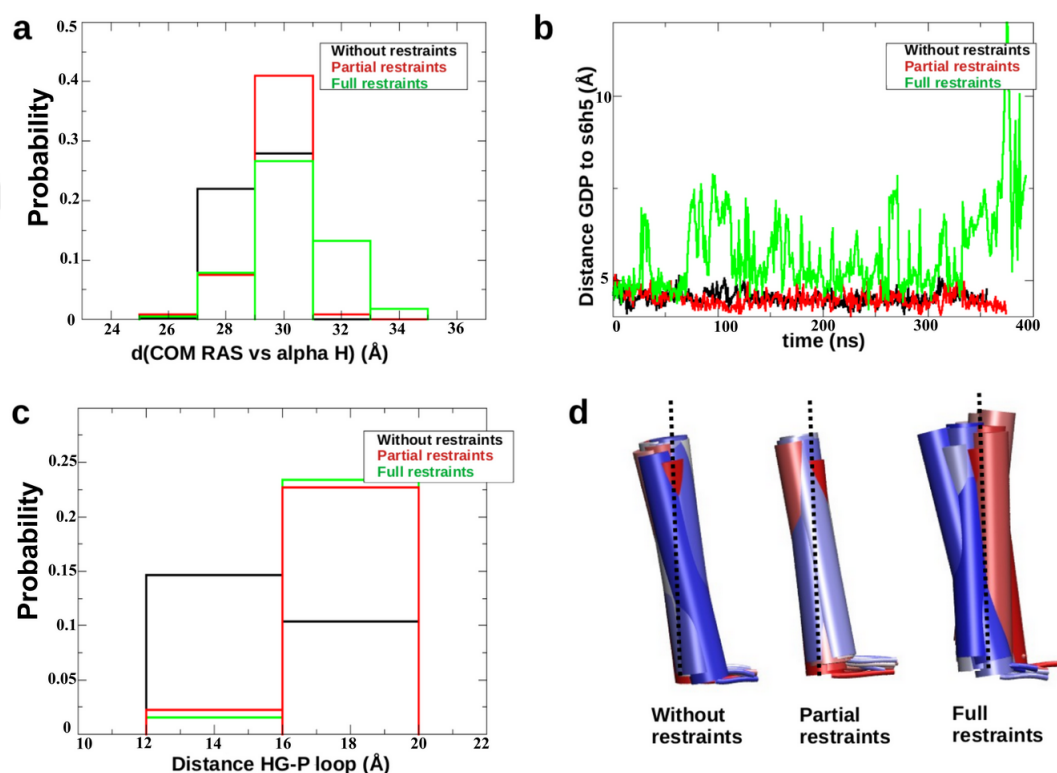


Figure 4.15: **a)** Distance between RAS like and Helical domains. , **b)** GDP to s6h5 loop distance, **c)** Distance between HG and P loop and **d)** Helix-5 time dependent change during simulation.

As reported the two domains (RAS-like and helical) are separated due to interdomain residues distortion, the same effect has been observed in the simulation applied with restraints to the whole structure. In the absence of restraints or by applying partial restraints on N-terminal the distance between domains is smaller than observed with full restraints simulation shown in **Figure 4.15 (a)**. In the simulations performed with partial restraints or without restraints have shown no such difference in distance between GDP and s6h5 loop shown in **Figure 4.15 (b)**. Thus, after applying restraints the distance start to become wider towards the course of 400ns simulation time. Distance between HG and P loop (the

regions surround the nucleotide binding pocket) tends to increase as well shown in **Figure 4.15 (c)**. In the presence of restraints, the separation of both regions become wider which is not the case in the absence of restraints. Translation of H5 was also captured in the presence of restraints as shown in **Figure 4.15 (d)** where the starting conformation of H5 is colored red and ending as blue perpendicular to the receptor's axis, throughout the course of simulation.

4.2.3. Dynamics of switch-II

Significantly the high dynamics of switch-II region was observed and detailed investigation has shown that it has an impact in the overall $G\alpha$ subunit stability. It has been observed that in the presence of restraints the switch-II has been shifted towards the nucleotide binding region which is not the case partial or without applied restraints as shown in **Figure 4.16**.

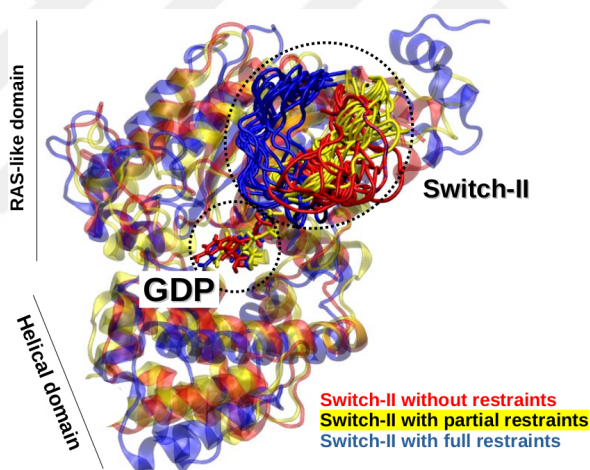


Figure 4.16: Three superimposed systems bound to GDP molecule and the switch-II region which is shown as loop, the flexibility of the region is observed trajectory wise throughout the course of simulation.

And the catalytic residue GLN-204 of switch-II was placed nearer to the nucleotide binding pocket this suggests it might come close to the pocket to aid the catalysis of GTP. All these events were captured for switch-II in the simulation where switch-II was in loop conformation. Switch-II has been crystallized as helix in the available two crystal structures of G protein colored as blue and yellow in the **Figure 4.17**. In contrast to the crystal structure conformation our results have shown that switch-II unravels during the course of simulation in both cases e.g; if the starting structure is in helix or loop conformation.

Interestingly, with starting conformation as loop with and without restraints, the residue GLN-204 come closer to the nucleotide binding pocket only when the restraints are applied in loop conformation.

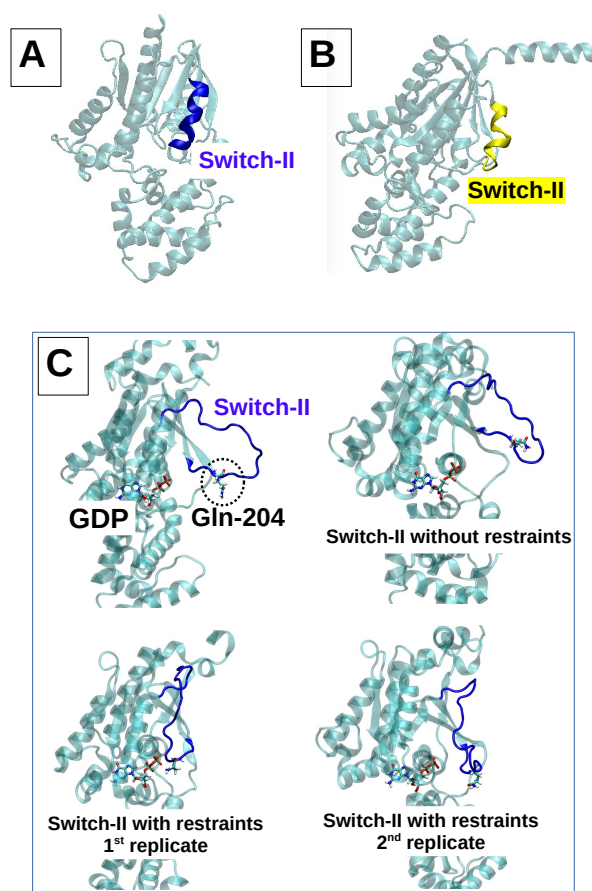


Figure 4.17: Comparison with crystal structures. A. crystal structure of $G\alpha$ subunit bound to RGS protein (pdb id:2ODE). B. crystal structure of $G\alpha$ bound to $G\beta\gamma$ dimer (pdb id:1GP2). C. the trajectory snapshots of our results depicting the loop conformation of switch-II region in the presence and absence of restraints. And the catalytic residue GLN-204 is highlighted in all the four snapshots.

4.2.3.1. Unraveling of switch-II helical conformation

The results showed unraveling of switch-II helix in the presence of restraints the α -helical conformation for switch-II was lost, but not in the absence of restraints. When the structural impact on restraints on the switch-II were observed, we found that conformations of Turn and 3-10 helix which represent less compact packing of helix were sampled higher in the presence of restraints than in their absence. By applying restraints, the distance between GLU-204 and nucleotide binding pocket reduced as switch-II adopts loop conformation and as shown in red color with reference range of GSP bound which resembles

close to intermediate state conformation and GTP bound structure which represents active conformation colored as black straight lines **Figure 4.18 (b)**.

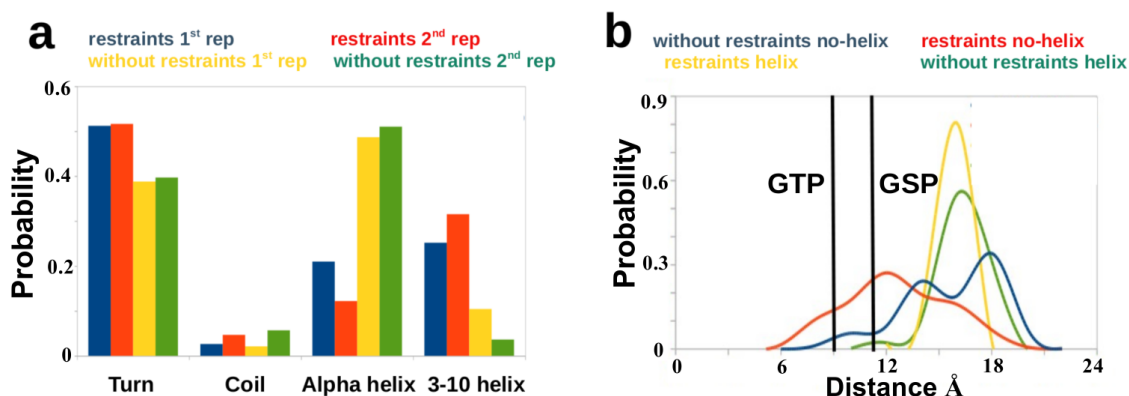


Figure 4.18: Impact of restraints on the structure of Switch-II **a.** represents the unraveling of α -helix conformation of switch-II in different types of simulations with restraints and without restraints. **b.** This show the comparison of helical conformation lost in the presence of restraints in switch-II region which is required for the catalytical residue Glu-204 present in switch-II region to come closer to nucleotide binding pocket.

The results have shown that switch-II has two positive residues GLN-204 and ARG-205, and switch-III has clustering of negatively charge residues GLU-236, ASP-237, GLU-238 and GLU-239. When restraints are applied to switch-II in loop conformations there has formed an interaction between negative clustered residues of switch-III with positive residues of switch-II, contrarily which is not observed when switch-II is in helix conformation as given in **Figure 4.19**.

When the distance between switch-II and switch-III was calculated it was seen, the distance decreased in the similar manner as it has been measured for the distance between GLN-204 and GDP in the presence of restraints and switch-II in loop conformation. Both events have occurred simultaneously, suggesting that in the loop conformation switch-II will come closer to switch-III by the formation of oppositively charged residues interactions and thus interacting with GDP molecule shown in **Figure 4.20**. These interactions formed between switches might be responsible of dragging the GLN-204 towards GDP and thus providing a chance for it to interact with β -phosphate of GDP.

4.3. Hexokinase 2

The impact of mutation on wild-type apo HK2 has been investigated in this section. The apo systems of HK2 were prepared by removing the ligands from HK2 crystal structure

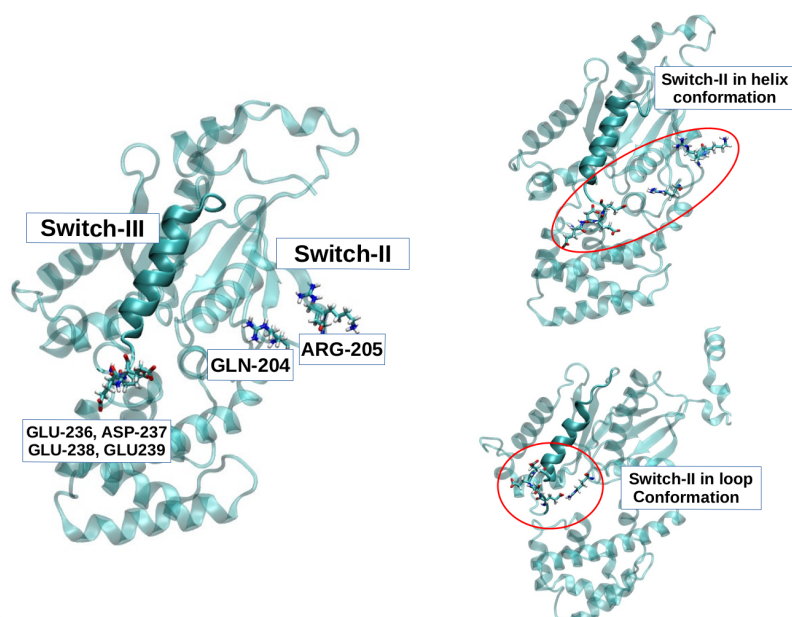


Figure 4.19: Interactions between switch-II and switch-III. Switch-III is shown as cyan bold color and Switch-II is shown transparent, cluster of positive residues in switch-II and cluster of negative residues of switch-III are labeled. Catalytic GLN-204 of switch-II make interactions with negative clustered residues in switch-III encircled in red.

with pdb id 2NZT and the aspartic acid residue at position 447 was mutated to alanine (See methods for detail). First the RMSD of the systems was calculated to ensure that the system has reached to an equilibrium towards the end of simulation time. The simulations were carried out on the whole dimeric protein for both WT and Mutant systems but the major contributor to activity loss is N-Domain and the results for only this domain are represented here. In the results of simulations it has been observed that the electrostatic interaction between Lysine K104 and aspartic acid D447 was broken as a result of mutation depicted in **Figure 4.21**. Upon mutation the interaction between linker helix and large subdomain is lost and this could exert fluctuations to the small subdomain and causing a wider opening of ligand binding pocket.

4.3.1. Principle component analysis RMSF for wild-type and mutant apo HK2

Principle component analysis (PCA) was carried out to understand the overall global motion of the protein system and vectors covering more than 50% of overall motion that depicts the dominant motion of the system. For analysing the results the alignment was done with respect to the static large subdomain of the enzyme. The major contributing regions which shows high fluctuations are small sub- domain and linker helix also two

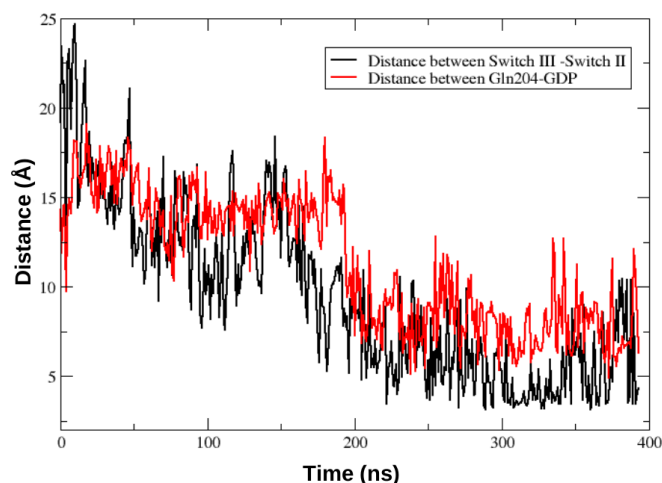


Figure 4.20: Distance decreases between Switch-II and III colored black and GLN-204 and GDP colored red.

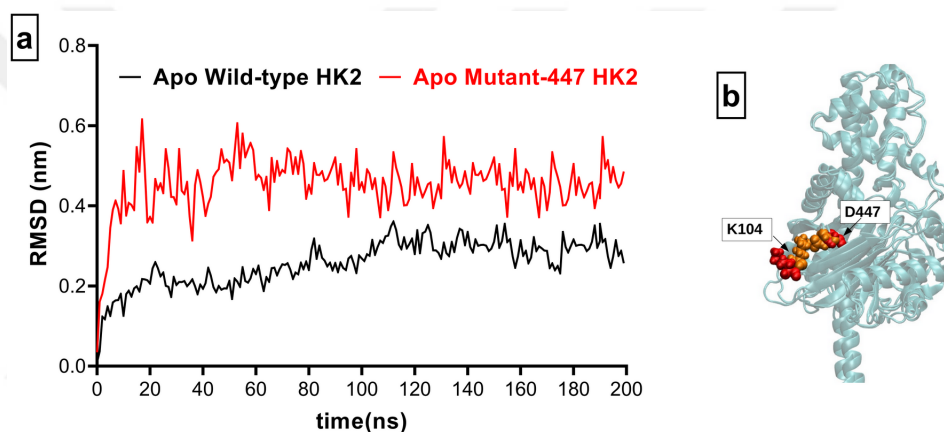


Figure 4.21: a). RMSD of both systems. b). The N-domain is shown here. K104 and D447 residues are depicted in VDW representation the red color represents mutant-447, and orange represents WT-447.

regions in large sub-domain in the mutant system which causes a wider opening of the small sub-domain **Figure 4.22**. D447 is present in the beginning of linker helix.

From the PCA analysis major regions were highlighted and compared within WT and mutant to understand the overall structural relation, which has provided us an insight that the mutation causes a disruption in electrostatic interaction of linker helix and large subdomain and this causes displacement of small and large subdomain and also the linker helix from their original position depicted in **Figure 4.23**.

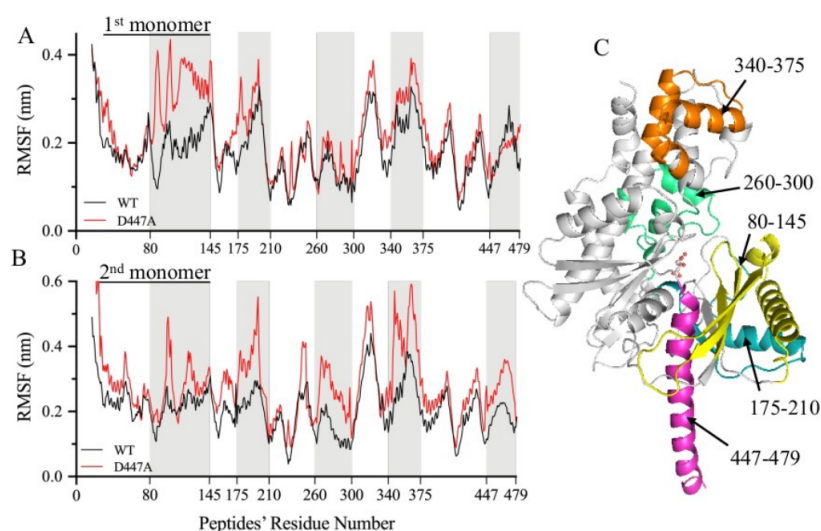


Figure 4.22: MD simulation of WT and D447A in the N-domain of HK2. (A-B) Comparison of PCA-derived RMSF profiles for the first and second monomers of the N-domain of HK2 for WT (black) and D447A (red). (C) Depiction of regions of the 3D structure of the N-domain of HK2 that show higher fluctuation in the D447A mutant.

4.3.2. Conformational stability comparison between wild-type and mutant

Since the active site and linker helix- α_{13} lies in a cleft between the small and large subdomains, the enhanced conformational dynamics induced by the D447A mutation caused wider opening of the N-domain active site. The wider opening of the active site facilitated by D447A is indicated by the large distance between the $C\alpha$ atoms of T88 and T232 of the small and large subdomains of the N-domain, respectively. The distance between T88 and T232 increased from a range of 5Å–14Å for the WT to 7Å–20Å for D447A. In addition, the displacement of the linker helix α_{13} impacted the relative orientation of the N-domain and C-domain in the D447A mutant, as represented by the wider angle between the two domains (**Figure 4.24**). The angle was measured at the $C\alpha$ atoms of T389, R468 and A839 on the N-domain, linker helix- α_{13} , and C-domain, respectively. The angle distribution was wider for D447A, with values of 160°–190°, than for the WT, with values of 150°–180°.

4.3.3. HK2 virtual screening results

After identifying the potential of the targeted residue 447 for HK2 enzyme activation, we set to find suitable small molecules that can potentially bind to the site and inhibit the enzyme. A library of 2466 FDA approved drugs were retrieved from Drug DataBank

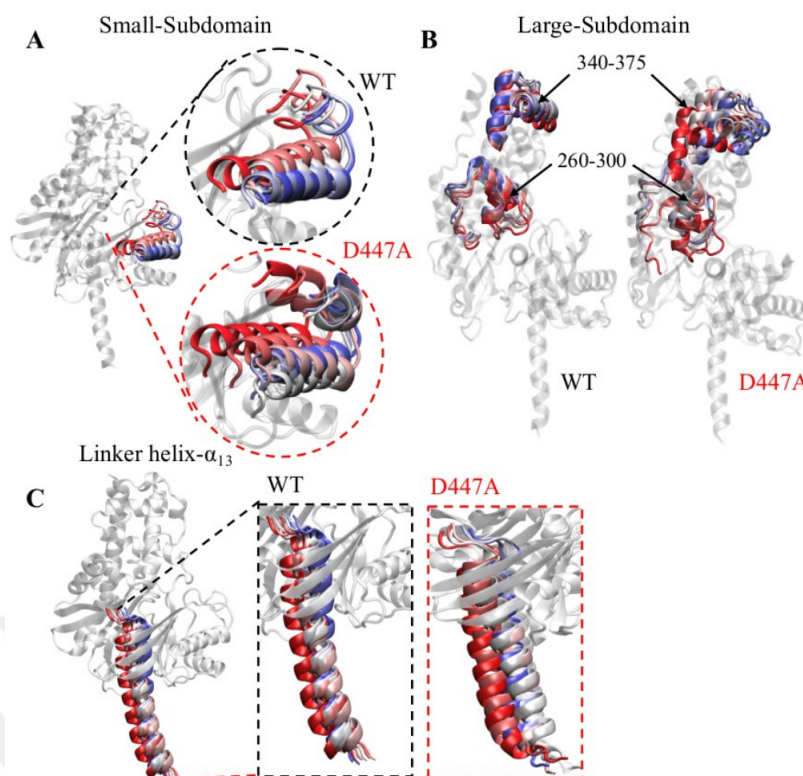


Figure 4.23: Enhanced dynamics of D447A mutant in the N-domain of HK2. Comparative depiction of N-domain regions with enhanced dynamics in D447A compared to WT for the small subdomain (A), large subdomain (B), and linker helix- α_{13} (C). The figures are taken as a pattern of protein movement throughout the course of 200ns simulation and the red color shows the first/initial conformation of the protein and the remaining are the frames of movement captured during the simulation.

database, among them 99 compounds were screened by standard precision (SP) flexible docking carried out by ligand Virtual screening. Different strategies were adopted to select the best candidates, to do so we set 4 different definitions of pharmacophore with different combination of residues. Due to the fact that D447 residue is resided towards the exposed site of enzyme in space, causing ligand detachment easier and faster in the first cycle of ligands screening. Thus we designed a new pharmacophore pocket by selecting more residues from the interior cavity of enzyme to make the shallower and exposed screening pocket more deeper. Here it is important to emphasize the selection criterion of screened ligands, which was based not only on lower binding energy but also the best pose facing the interior of enzyme with higher number of interactions with the pocket residues.

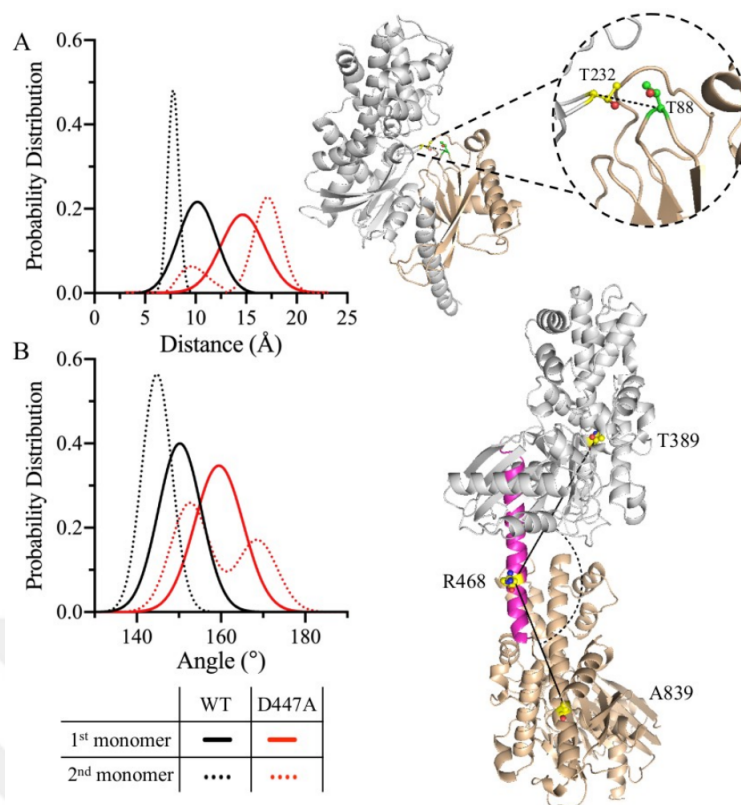


Figure 4.24: Distance and angle probability distributions of WT and D447A. (A) Distance probability distribution calculated between $C\alpha$ atoms of residues T88 and T232 displayed in a sphere representation. (B) Angle probability distribution calculated between $C\alpha$ atoms of residues T389, R468 and A839, shown as spheres.

4.3.3.1. Results of first pharmacophore pocket

The first pocket has D447, K451 and I114 residues, and the ligands were screened on both monomers individually. In the first strategy FDA approved ligands from 2 databases i.e; Drug DataBank and e-Drug3D (250 screened ligands from the library of 1993 compounds) were screened and among them the common screened molecules including Zileuton, Mitoxantrone and Propylthiouracil were selected for further MD studies. Two replicates of each system were simulated to 200ns. The screened ligands are given in the table below with their generic name, Glide binding score (Gscore) and therapeutic uses.

Results of MD simulations have shown despite of good binding energy scores and interactions Zileuton and Propylthiouracil have very short residency time in their pockets and they left from both monomers within 5-10ns in both replicas of the simulation. Structurally these two molecules are smaller than Mitoxantrone suggesting that smaller molecules are not suitable for our proposed binding site and thus both of these ligands were dropped.

Table 4.1: Screened molecules from first virtual screening results.

Generic name	Gscore (kcal/mol)	Therapeutic uses
Zileuton	-5.2	to prevent wheezing, shortness of breath, coughing and chest tightness due to asthma
Mitoxantrone	-5.1	to treat leukemia and other cancers
Propylthiouracil	-5.0	to treat hyperthyroidism

The simulation results of mitoxantrone have shown in both replicates the ligand is stably bound to its pocket in the 1st monomer; while it left from its pocket within 5-10ns for the 2nd monomer, with a consistent behavior in both replicas. Thus the impact of ligand on the structure of enzyme was measured by the same reaction coordinates (of T88-T232 residue distance and T389-R468-A839 residue angle) used to describe the impact of mutation in the previous section. The results shown that the binding pocket distance in both monomers has been affected and sampling the range nearer to mutant like system in both replicates and thus causing fluctuations to the small sub-domain. The angle range is similar to what observed for mutant system, while the impact on linker helix fluctuation was not observed to be significant as shown in **Figure 4.25**.

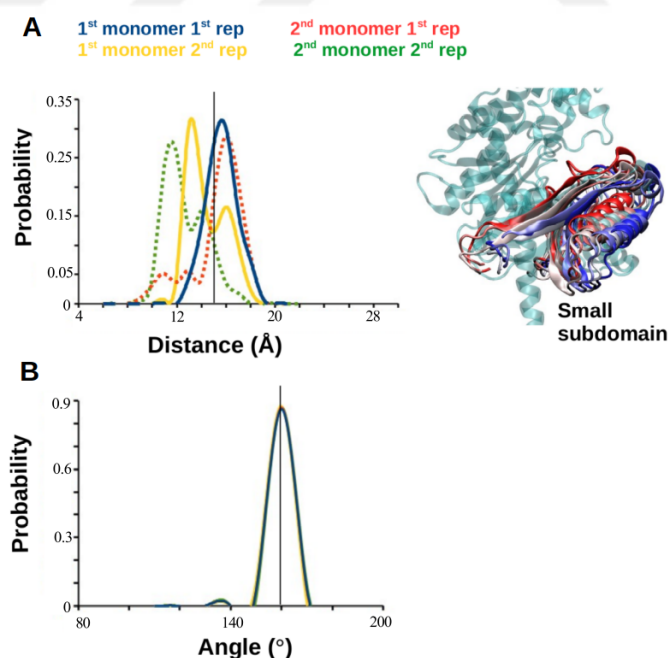


Figure 4.25: Results of Mitoxantrone molecule. A. Binding pocket distance and trajectory wise small sub-domain fluctuation and binding pocket opening distance. B. Angle probability. dashed lines are reference values of mutant system.

4.3.3.2. Results with second pharmacophore pocket

From the results of previous virtual screening, we observed all the ligands on 2nd monomer were not stable in their pockets and they tend to leave in the beginning of the simulation. Thus in view of these findings we set out to screen another round of ligands by targeting only 1st monomer with a deeper binding pocket. The selected residues for pharmacophore pocket were T88, N89, R91, T232, D413 and D447.

Drug Databank database was used, with this time including H-bond constraint to D447 residue to ensure the filtering of maximum ligands forming interaction with residue D447. The results with this new pocket has shown to be more promising where all molecules are having interaction with the residue of interest. But the selection criteria was based on low energy as well as number of higher interacting residues within the pharmacophore pocket but most importantly the deeper binding pose was taken in consideration. On the basis of these selection parameters three ligands were selected listed in **Table 2.**, and 200ns simulations were performed with these ligands to test their *insilico* stability. The good reason to test these ligands was their deeper occupation of the proposed binding pharmacophore pocket similar to Mitoxantrone screened pocket depicted in **Figure 4.26.**

Table 4.2: Screened molecules from second virtual screening results.

Generic name	Gscore (kcal/mol)	Therapeutic uses
Vilazodone	-4.1	treating depression
Idebenone	-3.7	treating Alzheimer's disease, liver disease, and heart disease, Leber's disease, mitochondrial encephalomyopathies
Tirofiban	-3.6	to prevent blood clots or heart attack

These interacting residues of these three ligands were; residue D447 and T88 interacting with Vilazodone, Idebenone interacts with S415, T232, T88 and D447 and Tirofiban have interaction with D447, N109, N89 and T88.

4.3.3.3. Vilazodone results

The simulation results with Vilazodone have shown stable residency in the binding pocket throughout the course of simulation in both replicates. Thus the impact of small subdomain fluctuation, distance of binding pocket and angle was measured which is given in

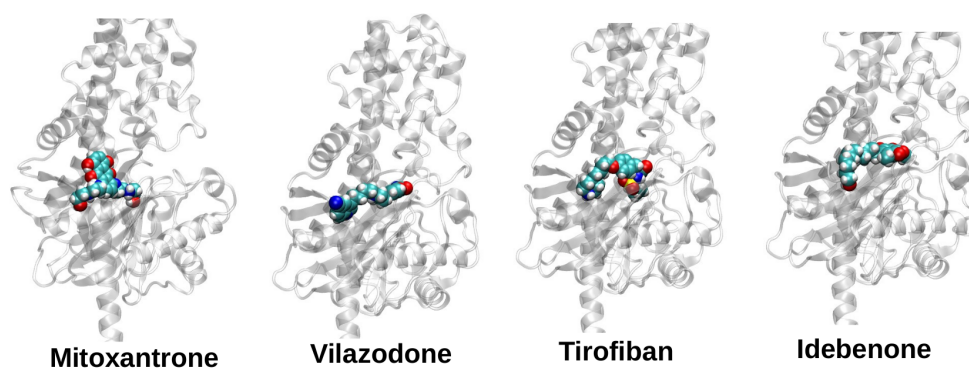


Figure 4.26: Ligand binding poses shown in van der Waals representation targeting the N terminal domain of HK2.

probability distribution graphs. Interestingly the results shown the impact of vilazodone on the dynamics of enzyme is similar to what has been observed for mutant system. Except the high fluctuation of the linker helix which is not similar to mutant system. The significant impact of this ligand is on the binding pocket distance, which became more wider as compared to angle.

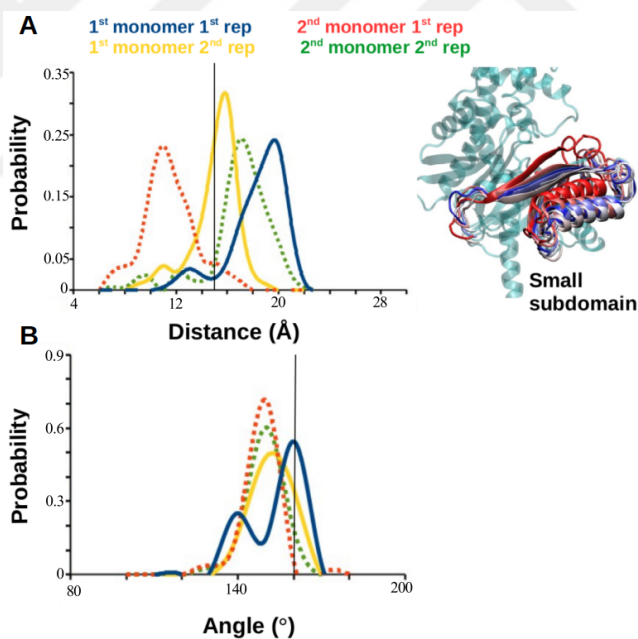


Figure 4.27: MD results with Vilazodone molecule. A. Binding pocket distance and trajectory wise small sub-domain fluctuation and binding pocket opening distance. B. Angle probability. dashed lines are reference values of mutant system.

4.3.3.4. Tirofiban results

The results of next ligand tirofiban are also interesting which resembles the findings observed with vilazodone, among both replicates the ligand is bound stably in one of the

replicate. While in the second replicate ligand loses interaction after a short while still resulting in the wider opening of the binding pocket in the 1st monomer but interestingly has no effect on the 2nd monomer. The impact of ligand was calculated from probability distribution graphs and found to be significant on binding pocket distance and is also similar to D447 mutant, while the impact on linker helix and angle is not significant in this case as well as shown in **Figure 4.28**.

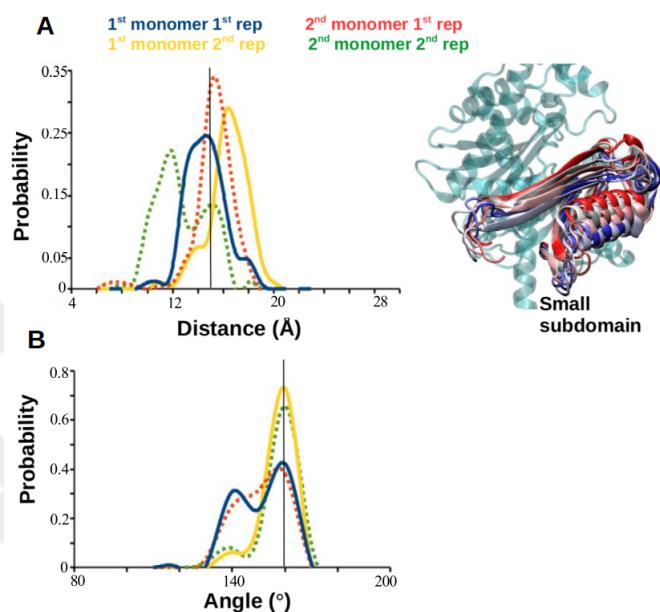


Figure 4.28: Results of Tirofiban molecule. A. Binding pocket distance and trajectory wise small sub-domain fluctuation and binding pocket opening distance. B. Angle probability. dashed lines are reference values of mutant system.

4.3.3.5. Idebenone results

The interesting findings for this ligand is the impact on linker-helix fluctuation showing high fluctuation as compared to the previous three ligands. Ligand remained attached to the pocket with different interaction compared to the starting positions until the end of simulation. While for 2nd replicate the ligand residency time was shorter than previous replicate. Results are obtained from probability distribution graphs, giving similar results with mutant system on both reaction coordinates of binding pocket distance and angle with their corresponding structural fluctuations as shown in **Figure 4.29**.

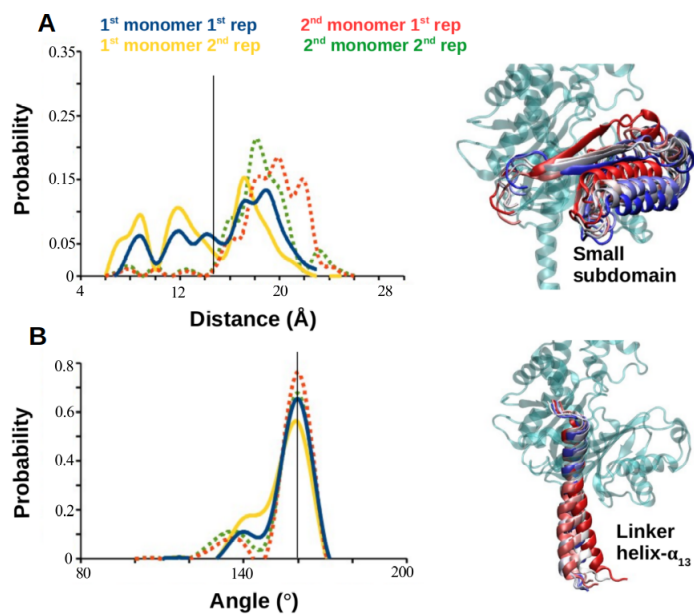


Figure 4.29: Results with Idebenone molecule. A. Binding pocket distance and trajectory wise small sub-domain fluctuation and binding pocket opening distance. B. Angle probability. dashed lines are reference values of mutant system.

CHAPTER 5

5. CONCLUSION AND FUTURE WORK

Proteins are made up of network of amino acid residues which can communicate even if they are located far from each other. The information flow among different regions of the protein can be maintained by means of ligand binding, interaction with another protein or with membrane, etc. Experimentally determined structures do provide structural information which is not enough to have a complete mechanistic insight into the working principles of proteins. To complement this missing piece, molecular dynamics simulations stand as indispensable tools to study time-dependent structural and dynamical behavior of proteins as long as the ergodic hypothesis is satisfied. Consequently, this provides calculation of the partition function, hence the thermodynamic properties of the systems studied. On the other hand, proteins have a rugged energy surface which makes it challenging to access all the local minima available to the systems. Under such conditions, enhanced sampling techniques can be used to overcome this problem to access longer time and larger length scales. In this thesis, three systems were studied, in each of which we focused on understanding dynamics of a protein, which paves the way for modulating its function for either therapeutic or biotechnological purposes. Enhanced sampling techniques were used when needed and a method was developed which can be used for studying conformational changes that are linked to function of proteins. As a future work, the hypotheses made throughout the thesis along with successful ligand candidates discovered will be tested in vitro experiments.

Specifically, in the first part of the thesis **Section 4.1**, we showed that dynamics of G protein-coupled receptor oligomers could be modulated by means of bivalent ligands. Bivalent ligands have long been thought to serve as linkers that connect two pharmacophore groups together. In this study, to the best of our knowledge for the first time, we showed

that they also modulated dynamics of protomers within the oligomer. The model system used in this study was a tetramer which was shown to consist of pairs of A_{2A}R and D₂R. The bivalent ligand was designed to consist of A_{2A}R agonist/antagonist which was in line with an experimental study. We showed that the linker made interactions with residues located at the extracellular domain of A_{2A}R protomers.

Interestingly, this led to stabilization of antagonist-bound A_{2A}R while the impact was not significant for agonist-bound A_{2A}R/mini- Gα_s suggesting that the latter was stabilized by the effector and no further stabilization was observed by the bivalent ligand. On the other hand, the linker caused an increase in the number of interactions formed with CGS-21680 while it was not observed for istradefylline-bound A_{2A}R. This showed that stabilization of the intracellular domain of CGS-21680-bound A_{2A}R by mini-Gα_s could help modulation of the extracellular domain of the receptor. In accordance with that, volume of the ligand binding pocket was shrunk while the intracellular domain was expanded in agonist CGS-21680-bound A_{2A}R/mini-Gα_s. In spite of having no remarkable change at the ligand binding site, the volume of the intracellular domain was decreased in istradefylline-bound A_{2A}R. Antagonist-bound A_{2A}R in the tetramer was embedded between CGS-21680-bound A_{2A}R/mini-Gα_s and apo D₂R. Therefore, expansion of the intracellular domain in agonist-bound A_{2A}R/mini-Gα_s might constraint the conformational space available for neighboring antagonist-bound A_{2A}R.

Herein, it is also important to emphasize that conformational preferences of some of the microswitches were modulated by the linker. Specifically, conserved Trp246^{6,48} adopted side chain dihedral angles which provided residue to be tightly packed against the sodium ion in antagonist-bound A_{2A}R. Also, another conserved residue, Tyr288^{7,53}, was oriented far from G protein binding site therein. On the other hand, no such alterations were observed in conformational preferences of these key residues in agonist-bound A_{2A}R-mini Gα_s. We also showed that above mentioned conformational changes led to alterations in residue correlation patterns and allosteric interaction network in the A_{2A}R dimer.

Specifically, we observed that conformational rearrangements which occurred during receptor activation were strengthened in the agonist-bound A_{2A}R-mini Gα_s by the presence of linker. Moreover, TM5 was emerged as the dominant participant of the allosteric interaction network in agonist-bound A_{2A}R- mini Gα_s in the presence of the linker. A similar

trend was observed in $A_{2A}R$ upon complex formation with mini- $G\alpha_s$ suggesting that impact of mini- $G\alpha_s$ was accentuated by the presence of linker in CGS-21680-bound $A_{2A}R$. Besides $A_{2A}R$, the bivalent ligand also affected dynamics of the D_2R dimer in the tetramer. Notably, apo D_2R could sample wider range of ionic lock distances in the absence of the linker, whereas the distance was confined to the value which was sampled by the inactive receptor in the presence of the linker. In a study of Han et.al, it was shown that inverse agonist binding to one of the protomers in D_2R dimer enhanced D_2R signaling from agonist-bound D_2R whereas agonist binding blunted signaling. In accordance with that we observed similar pattern for both dimers in the tetramer. That is to say, $A_{2A}R$ and D_2R dimers resembled asymmetric activation units in the presence of the linker.

In a recent study, where the stoichiometry and the identity of the tetramer (the model used here) were revealed, it was shown that simultaneous occupation of both $A_{2A}R$ s in the tetramer by an agonist and antagonist did not induce an allosteric modulation of D_2R agonist binding and intrinsic efficacy. Therefore, these results suggest that the bivalent ligand has the potential to prevent/reduce antagonistic impact of $A_{2A}R$ on D_2R . Lastly, $A_{2A}R$ - D_2R oligomers have been used as targets in treatment of Parkinson's disease; however, findings of this study can help understand impact of any bivalent ligand on any GPCR oligomer, thus paving a way for improving design strategies of such ligands and effective modulation of GPCR oligomers.

In **Section 4.2** the second part of the thesis, we developed a method to study the transition states adopted by G protein by applying experimentally derived ϕ -values in MD simulation. The results were first compared with determined reaction coordinates reported by Sun et al., 2018 [15] and our results were in line with their findings, the unknown reason for how a 30Å far receptor binding site could have an impact on GDP dissociation, was studied along this we have observed the dynamics of switch-II region in GDP release in our results. For detailed insight we have compared switch-II modeled region obtained from our simulations with available crystal structures of $G\alpha$ subunit bound to RGS effector-protein and $G\beta\gamma$ dimer, and interestingly we have observed the switch-II in the crystal structures was in helical conformation while during the course of our simulations we have observed loop-conformation of switch-II. This helical conformation was observed even

when the region was modeled as coil helix, suggesting that the switch-II is in loop conformation in the transition state of GDP to GTP catalysis.

Moreover the catalytic residue Glutamine-204 is present near the N-terminus of switch-II is closer to nucleotide binding region which suggests that in the transition state when switch-II is in loop conformation the rearrangement has been already achieved for Glutamine that can govern GDP binding pocket distortion. And once bound to $\beta\gamma$ dimer unit the GDP not only dissociate but at the same time the dimer also fall apart.

In this study we have tried to set out an example that supports the use of ϕ -values for the determination of such conformational states which are not possible to capture via classical MD simulation, and thus it has an advantage over enhanced sampling method techniques which require predetermined reaction coordinates to be used to govern conformational change studies. There are some reported studies on very long MD simulations of millisecond timescale which studied the unfolding of various different proteins on a special machine ANTON, but there still remained a challenge to perform such simulations with proteins having more than 100 number of residues. These limitations of computational cost and limited number of residues can be overcome by combining the phi-value data with all-atom classical MD simulations for solving protein folding and protein activation pathway problems.

In **Section 4.3** the third part of the thesis, is section we have studied a major type of Hexokinase 2 enzyme involved in pathologies of various different types of cancers. We have investigated the impact of an experimentally identified potential mutation on aspartic acid at position 447 on the linker helix region of the enzyme by means of MD. The findings from HK2 simulations are concluded as; D447 residue of Apo HK2 when mutated to Alanine lost complete enzyme's activity, the findings of HDX experimental technique could further differentiate highly dynamic regions of Apo D447A mutant from Apo wild-type. We then set out simulation experiment on this mutant and observed from the RMSF fluctuation results the highly fluctuating residues were overlapping with those obtained from HDX experimental findings. These regions include (i) linker helix which is a connection between two domains and (ii) small subdomain which is responsible for the opening and closing of the Glucose and ATP binding pocket.

The mutation is potent because it breaks the connection between linker helix and large

subdomain, the impact surpasses to the small subdomain region causing a wider opening to the binding pocket on both domains N and C in this way it disrupted the overall structural organization of the enzyme.

These findings are very fruitful in the identification of this non-active target site of the enzyme which could serve as the potential off-site target for HK2 inhibition, without interfering with normally expressing other HK sub-types.

After identifying the target site of HK2 we have performed ligand screening with FDA approved drugs. The results have provided few candidates including Mitoxantrone, Vilazodone, Tirofiban and Idebenone might have the potential to be repurposed for targeting HK2. To suggest a final candidate to be served as a potential cancer inhibitor of HK2 further these ligands need to be tested in-vitro to show their efficacy in the proposed way of HK2 inhibition.

BIBLIOGRAPHY

- [1] P. Bernadó and M. Blackledge, “Proteins in dynamic equilibrium,” *Nature*, vol. 468, no. 7327, pp. 1046–1048, 2010.
- [2] K. A. Henzler-Wildman, M. Lei, V. Thai, S. J. Kerns, M. Karplus, and D. Kern, “A hierarchy of timescales in protein dynamics is linked to enzyme catalysis,” *Nature*, vol. 450, no. 7171, pp. 913–916, 2007.
- [3] B. Smit, “Molecular-dynamics simulations of amphiphilic molecules at a liquid-liquid interface,” *Physical Review A*, vol. 37, no. 9, p. 3431, 1988.
- [4] T. Hansson, C. Oostenbrink, and W. van Gunsteren, “Molecular dynamics simulations,” *Current Opinion in Structural Biology*, vol. 12, no. 2, pp. 190–196, 2002.
- [5] M. B. Kubitzki, B. L. de Groot, and D. Seeliger, “Protein dynamics: from structure to function,” in *From Protein Structure to Function with Bioinformatics*. Springer, 2017, pp. 393–425.
- [6] M. Karplus and J. Kuriyan, “Molecular dynamics and protein function,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 19, pp. 6679–6685, 2005.
- [7] M. C. Zwier and L. T. Chong, “Reaching biological timescales with all-atom molecular dynamics simulations,” *Current Opinion in Pharmacology*, vol. 10, no. 6, pp. 745–752, 2010.
- [8] J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw, “Long-timescale molecular dynamics simulations of protein structure and function,” *Current Opinion in Structural Biology*, vol. 19, no. 2, pp. 120–127, 2009.
- [9] W. K. Kroeze, D. J. Sheffler, and B. L. Roth, “G-protein-coupled receptors at a glance,” *Journal of Cell Science*, vol. 116, no. 24, pp. 4867–4869, 2003.
- [10] R. Fredriksson, M. C. Lagerström, L.-G. Lundin, and H. B. Schiöth, “The g-protein-coupled receptors in the human genome form five main families. phylogenetic analysis, paralogon groups, and fingerprints,” *Molecular Pharmacology*, vol. 63, no. 6, pp. 1256–1272, 2003.
- [11] E. Ghosh, P. Kumari, D. Jaiman, and A. K. Shukla, “Methodological advances: the unsung heroes of the gpcr structural revolution,” *Nature Reviews Molecular Cell Biology*, vol. 16, no. 2, pp. 69–81, 2015.
- [12] M. Bouvier, “Oligomerization of g-protein-coupled transmitter receptors,” *Nat. Rev. Neurosci.*, vol. 2, no. 4, p. 274, 2001.
- [13] A. Salahpour, S. Angers, and M. Bouvier, “Functional significance of oligomerization of g-protein-coupled receptors,” *Trends Endocrinol. Metab.*, vol. 11, no. 5, pp. 163–168, 2000.
- [14] B. A. Jordan and L. A. Devi, “G-protein-coupled receptor heterodimerization modulates receptor function,” *Nature*, vol. 399, no. 6737, p. 697, 1999.
- [15] S. R. George, B. F. O’Dowd, and S. P. Lee, “G-protein-coupled receptor oligomerization and its potential for drug discovery,” *Nat. Rev. Drug Discovery*, vol. 1, no. 10, p. 808, 2002.

- [16] J. Ellis, J. D. Padiani, M. Canals, S. Milasta, and G. Milligan, "Orexin-1 receptor-cannabinoid cb1 receptor heterodimerization results in both ligand-dependent and-independent coordinated alterations of receptor localization and function," *J. Biol. Chem.*, vol. 281, no. 50, pp. 38 812–38 824, 2006.
- [17] D. Mesnier and J.-L. Banères, "Cooperative conformational changes in a g-protein-coupled receptor dimer, the leukotriene b4 receptor blt1," *J. Biol. Chem.*, vol. 279, no. 48, pp. 49 664–49 670, 2004.
- [18] A. Tobin, "G-protein-coupled receptor phosphorylation: Where, when and by whom," *British Journal of Pharmacology*, vol. 153, no. S1, pp. S167–S176, 2008.
- [19] D. M. Rosenbaum, S. G. Rasmussen, and B. K. Kobilka, "The structure and function of g-protein-coupled receptors," *Nature*, vol. 459, no. 7245, pp. 356–363, 2009.
- [20] B. K. Kobilka and X. Deupi, "Conformational complexity of g-protein-coupled receptors," *Trends in Pharmacological Sciences*, vol. 28, no. 8, pp. 397–406, 2007.
- [21] H. B. Schiöth and R. Fredriksson, "The grafts classification system of g-protein coupled receptors in comparative perspective," *General and Comparative Endocrinology*, vol. 142, no. 1-2, pp. 94–101, 2005.
- [22] L. Xue, X. Rovira, P. Scholler, H. Zhao, J. Liu, J.-P. Pin, and P. Rondard, "Major ligand-induced rearrangement of the heptahelical domain interface in a gpcr dimer," *Nature Chemical Biology*, vol. 11, no. 2, pp. 134–140, 2015.
- [23] M. Ilter, S. Mansoor, and O. Sensoy, "Utilization of biased g protein-coupled receptor signaling towards development of safer and personalized therapeutics," *Molecules*, vol. 24, no. 11, p. 2052, 2019.
- [24] A. Ranganathan, R. O. Dror, and J. Carlsson, "Insights into the role of asp792.50 in β 2 adrenergic receptor activation from molecular dynamics simulations," *Biochemistry*, vol. 53, no. 46, pp. 7283–7296, 2014.
- [25] R. Dawaliby, C. Trubbia, C. Delporte, M. Masureel, P. Van Antwerpen, B. K. Kobilka, and C. Govaerts, "Allosteric regulation of g protein-coupled receptor activity by phospholipids," *Nature Chemical Biology*, vol. 12, no. 1, pp. 35–39, 2016.
- [26] R. O. Dror, D. H. Arlow, P. Maragakis, T. J. Mildorf, A. C. Pan, H. Xu, D. W. Borhani, and D. E. Shaw, "Activation mechanism of the β 2-adrenergic receptor," *Proceedings of the National Academy of Sciences*, vol. 108, no. 46, pp. 18 684–18 689, 2011.
- [27] W. Huang, A. Manglik, A. Venkatakrisnan, T. Laeremans, E. N. Feinberg, A. L. Sanborn, H. E. Kato, K. E. Livingston, T. S. Thorsen, R. C. Kling *et al.*, "Structural insights into μ -opioid receptor activation," *Nature*, vol. 524, no. 7565, pp. 315–321, 2015.
- [28] P. M. Dijkman, O. K. Castell, A. D. Goddard, J. C. Munoz-Garcia, C. De Graaf, M. I. Wallace, and A. Watts, "Dynamic tuneable g protein-coupled receptor monomer-dimer populations," *Nature Communications*, vol. 9, no. 1, pp. 1–14, 2018.

- [29] H. Dwivedi, M. Chaturvedi, M. Baidya, T. M. Stepniewski, S. Pandey, J. Maharana, A. Srivastava, N. Caengprasath, A. C. Hanyaloglu, J. Selent *et al.*, “Distinct phosphorylation sites in a prototypical gpcr differently orchestrate β -arrestin interaction, trafficking, and signaling,” *Sci Adv.* 2020; 6 (37): eabb8368, 2020.
- [30] P. S-H Park, “Ensemble of g protein-coupled receptor active states,” *Current Medicinal Chemistry*, vol. 19, no. 8, pp. 1146–1154, 2012.
- [31] D. Provasi, M. C. Artacho, A. Negri, J. C. Mobarec, and M. Filizola, “Ligand-induced modulation of the free-energy landscape of g protein-coupled receptors explored by adaptive biasing techniques,” *PLoS Computational Biology*, vol. 7, no. 10, p. e1002193, 2011.
- [32] A. S. Hauser, M. M. Attwood, M. Rask-Andersen, H. B. Schiöth, and D. E. Gloriam, “Trends in gpcr drug discovery: new agents, targets and indications,” *Nature Reviews Drug Discovery*, vol. 16, no. 12, pp. 829–842, 2017.
- [33] K. A. Jacobson, S. Costanzi, and S. Paoletta, “Computational studies to predict or explain g protein coupled receptor polypharmacology,” *Trends in Pharmacological Sciences*, vol. 35, no. 12, pp. 658–663, 2014.
- [34] C. S. Odoemelam, B. Percival, H. Wallis, M.-W. Chang, Z. Ahmad, D. Scholey, E. Burton, I. H. Williams, C. L. Kamerlin, and P. B. Wilson, “G-protein coupled receptors: structure and function in drug discovery,” *RSC Advances*, vol. 10, no. 60, pp. 36 337–36 348, 2020.
- [35] A. Heifetz, G. F. X. Schertler, R. Seifert, C. G. Tate, P. M. Sexton, V. V. Gurevich, D. Fourmy, V. Cherezov, F. H. Marshall, R. I. Storer *et al.*, “Gpcr structure, function, drug discovery and crystallography: report from Academia-Industry International Conference (UK Royal Society) Chicheley Hall, 1–2 september 2014,” *Naunyn-Schmiedeberg’s Archives of Pharmacology*, vol. 388, 2015.
- [36] M. Esguerra, A. Siretskiy, X. Bello, J. Sallander, and H. Gutiérrez-de Terán, “Gpcr-modsim: A comprehensive web based solution for modeling g-protein coupled receptors,” *Nucleic Acids Research*, vol. 44, no. W1, pp. W455–W462, 2016.
- [37] V. V. Gurevich and E. V. Gurevich, “Molecular mechanisms of gpcr signaling: a structural perspective,” *International Journal of Molecular Sciences*, vol. 18, no. 12, p. 2519, 2017.
- [38] R. Nygaard, Y. Zou, R. O. Dror, T. J. Mildorf, D. H. Arlow, A. Manglik, A. C. Pan, C. W. Liu, J. J. Fung, M. P. Bokoch *et al.*, “The dynamic process of β 2-adrenergic receptor activation,” *Cell*, vol. 152, no. 3, pp. 532–542, 2013.
- [39] S. Granier and B. Kobilka, “A new era of gpcr structural and chemical biology,” *Nature Chemical Biology*, vol. 8, no. 8, pp. 670–673, 2012.
- [40] D. S. Wishart, “Interpreting protein chemical shift data.” *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 58, no. 1-2, pp. 62–87, 2010.
- [41] G. Jeschke, “Deer distance measurements on proteins,” *Annual Review of Physical Chemistry*, vol. 63, pp. 419–446, 2012.

- [42] M. Lohse, P. Hein, C. Hoffmann, V. Nikolaev, J.-P. Villardaga, and M. Büne-
mann, “Kinetics of g-protein-coupled receptor signals in intact cells,” *British
Journal of Pharmacology*, vol. 153, no. S1, pp. S125–S132, 2008.
- [43] M. J. Lohse, S. Nuber, and C. Hoffmann, “Fluorescence/bioluminescence reso-
nance energy transfer techniques to study g-protein-coupled receptor activation
and signaling,” *Pharmacological Reviews*, vol. 64, no. 2, pp. 299–336, 2012.
- [44] S. E. Mansoor, M. A. DeWitt, and D. L. Farrens, “Distance mapping in
proteins using fluorescence spectroscopy: the tryptophan-induced quenching
(triq) method,” *Biochemistry*, vol. 49, no. 45, pp. 9722–9731, 2010.
- [45] S. H. Park, B. B. Das, F. Casagrande, Y. Tian, H. J. Nothnagel, M. Chu,
H. Kiefer, K. Maier, A. A. De Angelis, F. M. Marassi *et al.*, “Structure of
the chemokine receptor cxcr1 in phospholipid bilayers,” *Nature*, vol. 491, no.
7426, pp. 779–783, 2012.
- [46] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O.
Dror, and D. E. Shaw, “Improved side-chain torsion potentials for the amber
ff99sb protein force field,” *Proteins: Structure, Function, and Bioinformatics*,
vol. 78, no. 8, pp. 1950–1958, 2010.
- [47] J. Bonaventura, G. Navarro, V. Casadó-Anguera, K. Azdad, W. Rea,
E. Moreno, M. Brugarolas, J. Mallol, E. I. Canela, C. Lluís *et al.*, “Allosteric in-
teractions between agonists and antagonists within the adenosine a2a receptor-
dopamine d2 receptor heterotetramer,” *Proceedings of the National Academy
of Sciences*, vol. 112, no. 27, pp. E3609–E3618, 2015.
- [48] B. G. Jenkins, A. Zhu, P. Poutiainen, J.-K. Choi, K.-E. Kil, Z. Zhang, D. Ku-
ruppu, N. Aytan, A. Dedeoglu, and A.-L. Brownell, “Functional modulation of
g-protein coupled receptors during parkinson disease-like neurodegeneration,”
Neuropharmacology, vol. 108, pp. 462–473, 2016.
- [49] D. M. Ridddy, P. Delerive, R. J. Summers, P. M. Sexton, and C. J. Langmead,
“G protein-coupled receptors targeting insulin resistance, obesity, and type 2
diabetes mellitus,” *Pharmacol. Rev.*, vol. 70, no. 1, pp. 39–67, 2018.
- [50] A. K. Arakaki, W.-A. Pan, and J. Trejo, “Gpcrs in cancer: protease-activated
receptors, endocytic adaptors and signaling,” *Int. J. Mol. Sci.*, vol. 19, no. 7, p.
1886, 2018.
- [51] S. Ferré, “The gpcr heterotetramer: challenging classical pharmacology,”
Trends Pharmacol. Sci., vol. 36, no. 3, pp. 145–152, 2015.
- [52] M. Niethammer, C. C. Tang, Y. Ma, P. J. Mattis, J. H. Ko, V. Dhawan, and
D. Eidelberg, “Parkinson’s disease cognitive network correlates with caudate
dopamine,” *Neuroimage*, vol. 78, pp. 204–209, 2013.
- [53] G. C. Cotzias, P. S. Papavasiliou, and R. Gellene, “Modification of
parkinsonism—chronic treatment with l-dopa,” *N. Engl. J. Med.*, vol. 280,
no. 7, pp. 337–345, 1969.
- [54] P. A. LeWitt, “Levodopa therapy for parkinson’s disease: pharmacokinetics
and pharmacodynamics,” *Mov. Disord.*, vol. 30, no. 1, pp. 64–72, 2015.
- [55] A. J. Loonen and S. A. Ivanova, “New insights into the mechanism of drug-
induced dyskinesia,” *CNS Spectr.*, vol. 18, no. 1, pp. 15–20, 2013.

- [56] M. Cieślak, M. Komoszyński, and A. Wojtczak, “Adenosine a2a receptors in parkinson’s disease treatment,” *Purinergic Signal.*, vol. 4, no. 4, p. 305, 2008.
- [57] G. Lebon, P. C. Edwards, A. G. Leslie, and C. G. Tate, “Molecular determinants of cgs21680 binding to the human adenosine a2a receptor,” *Mol. Pharmacol.*, vol. 87, no. 6, pp. 907–915, 2015.
- [58] S. Mansoor, G. Kayik, E. Alshahaby, M. Guzel, S. Durdagi, and O. Sensoy, “Development of novel therapeutic that can target multiple receptors for treatment of parkinson’s disease.” *3rd Annual Meeting, MuTaLig, COST-Action CA15135, Valletta, Malta*, 2018.
- [59] A. Soriano, R. Ventura, A. Molero, R. Hoen, V. Casadó, A. Cortés, F. Fanelli, F. Albericio, C. Lluís, R. Franco, and M. Royo, “Adenosine a2a receptor-antagonist/dopamine d2 receptor-agonist bivalent ligands as pharmacological tools to detect a2a-d2 receptor heteromers,” *J. Med. Chem.*, vol. 52, no. 18, pp. 5590–5602, 2009.
- [60] B. Z. Stanton, E. J. Chory, and G. R. Crabtree, “Chemically induced proximity in biology and medicine,” *Science*, vol. 359, no. 6380, 2018.
- [61] Y. Xu, S. Duggineni, S. Espitia, D. D. Richman, J. An, and Z. Huang, “A synthetic bivalent ligand of cxcr4 inhibits hiv infection,” *Biochem. Biophys. Res. Commun.*, vol. 435, no. 4, pp. 646–650, 2013.
- [62] N. Krall, F. Pretto, and D. Neri, “A bivalent small molecule-drug conjugate directed against carbonic anhydrase ix can elicit complete tumour regression in mice,” *Chem. Sci.*, vol. 5, no. 9, pp. 3640–3644, 2014.
- [63] H. Hübner, T. Schellhorn, M. Gienger, C. Schaab, J. Kaindl, L. Leeb, T. Clark, D. Möller, and P. Gmeiner, “Structure-guided development of heterodimer-selective gpcr ligands,” *Nat. Commun.*, vol. 7, p. 12298, 2016.
- [64] G. Navarro, A. Cordoní, V. Casadó-Anguera, E. Moreno, N.-S. Cai, A. Cortés, E. I. Canela, C. W. Dessauer, V. Casadó, L. Pardo, C. Lluís, and S. Ferre, “Evidence for functional pre-coupled complexes of receptor heteromers and adenylyl cyclase,” *Nat. Commun.*, vol. 9, no. 1, pp. 1–12, 2018.
- [65] Y. Han, I. S. Moreira, E. Urizar, H. Weinstein, and J. A. Javitch, “Allosteric communication between protomers of dopamine class a gpcr dimers modulates activation,” *Nat. Chem. Biol.*, vol. 5, no. 9, p. 688, 2009.
- [66] A. G. Gilman, “G proteins: transducers of receptor-generated signals,” *Annual Review of Biochemistry*, vol. 56, no. 1, pp. 615–649, 1987.
- [67] M. A. Wall, D. E. Coleman, E. Lee, J. A. Iñiguez-Lluhi, B. A. Posner, A. G. Gilman, and S. R. Sprang, “The structure of the g protein heterotrimer $g_{i\alpha 1\beta 1\gamma 2}$,” *Cell*, vol. 83, no. 6, pp. 1047–1058, 1995.
- [68] S. R. Sprang, “G proteins, effectors and gaps: structure and mechanism,” *Current Opinion in Structural Biology*, vol. 7, no. 6, pp. 849–856, 1997.
- [69] W. M. Oldham, N. Van Eps, A. M. Preininger, W. L. Hubbell, and H. E. Hamm, “Mechanism of the receptor-catalyzed activation of heterotrimeric g proteins,” *Nature Structural & Molecular Biology*, vol. 13, no. 9, pp. 772–777, 2006.

- [70] T. K. Bjarnadóttir, D. E. Gloriam, S. H. Hellstrand, H. Kristiansson, R. Fredriksson, and H. B. Schiöth, “Comprehensive repertoire and phylogenetic analysis of the g protein-coupled receptors in human and mouse,” *Genomics*, vol. 88, no. 3, pp. 263–273, 2006.
- [71] A. L. Hopkins and C. R. Groom, “The druggable genome,” *Nature Reviews Drug Discovery*, vol. 1, no. 9, pp. 727–730, 2002.
- [72] L. S. Weinstein, M. Chen, T. Xie, and J. Liu, “Genetic diseases associated with heterotrimeric g proteins,” *Trends in Pharmacological Sciences*, vol. 27, no. 5, pp. 260–266, 2006.
- [73] J. Wang, C. Gareri, and H. A. Rockman, “G-protein-coupled receptors in heart disease,” *Circulation Research*, vol. 123, no. 6, pp. 716–735, 2018.
- [74] P. R. B. Evora and F. Nobre, “The role of g-proteins in the pathophysiology of the cardiovascular diseases,” *Arquivos Brasileiros de Cardiologia*, vol. 72, no. 2, pp. 209–229, 1999.
- [75] A. M. Spiegel and L. S. Weinstein, “Inherited diseases involving g proteins and g protein-coupled receptors,” *Annu. Rev. Med.*, vol. 55, pp. 27–39, 2004.
- [76] T. Ternström, U. Mayor, M. Akke, and M. Oliveberg, “From snapshot to movie: ϕ analysis of protein folding transition states taken one step further,” *Proceedings of the National Academy of Sciences*, vol. 96, no. 26, pp. 14 854–14 859, 1999.
- [77] A. R. Fersht, A. Matouschek, and L. Serrano, “The folding of an enzyme: I. theory of protein engineering analysis of stability and pathway of protein folding,” *Journal of Molecular Biology*, vol. 224, no. 3, pp. 771–782, 1992.
- [78] R. A. Gatenby and R. J. Gillies, “Why do cancers have high aerobic glycolysis?” *Nature Reviews Cancer*, vol. 4, no. 11, pp. 891–899, 2004.
- [79] L. Gay, A.-M. Baker, and T. A. Graham, “Tumour cell heterogeneity,” *F1000Research*, vol. 5, 2016.
- [80] O. Warburg, “On the origin of cancer cells,” *Science*, vol. 123, no. 3191, pp. 309–314, 1956.
- [81] M. G. Vander Heiden, L. C. Cantley, and C. B. Thompson, “Understanding the warburg effect: the metabolic requirements of cell proliferation,” *Science*, vol. 324, no. 5930, pp. 1029–1033, 2009.
- [82] P. P. Hsu and D. M. Sabatini, “Cancer cell metabolism: Warburg and beyond,” *Cell*, vol. 134, no. 5, pp. 703–707, 2008.
- [83] W. H. Koppenol, P. L. Bounds, and C. V. Dang, “Otto warburg’s contributions to current concepts of cancer metabolism,” *Nature Reviews Cancer*, vol. 11, no. 5, pp. 325–337, 2011.
- [84] L. Schwartz, T. Seyfried, K. O. Alfarouk, J. D. V. Moreira, and S. Fais, “Out of warburg effect: An effective cancer treatment targeting the tumor specific metabolism and dysregulated ph,” in *Seminars in Cancer Biology*, vol. 43. Elsevier, 2017, pp. 134–138.
- [85] M. L. Macheda, S. Rogers, and J. D. Best, “Molecular and cellular regulation of glucose transporter (glut) proteins in cancer,” *Journal of Cellular Physiology*, vol. 202, no. 3, pp. 654–662, 2005.

- [86] C. Mori, N. Nakamura, J. Welch, K. Shiota, and E. Eddy, "Testis-specific expression of mrnas for a unique human type 1 hexokinase lacking the porin-binding domain," *Molecular Reproduction and Development: Incorporating Gamete Research*, vol. 44, no. 1, pp. 14–22, 1996.
- [87] A. Lehninger, "Principles of biochemistry, worth., new york, 1982. r. miller, s. treppo, a. voigt, w. zingg and aw neumann," *Colloids Surfaces*, vol. 69, p. 203, 1993.
- [88] G. Hatzivassiliou, F. Zhao, D. E. Bauer, C. Andreadis, A. N. Shaw, D. Dhanak, S. R. Hingorani, D. A. Tuveson, and C. B. Thompson, "Atp citrate lyase inhibition can suppress tumor cell growth," *Cancer Cell*, vol. 8, no. 4, pp. 311–321, 2005.
- [89] C. Corbet and O. Feron, "Cancer cell metabolism and mitochondria: nutrient plasticity for tca cycle fueling," *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, vol. 1868, no. 1, pp. 7–15, 2017.
- [90] M. G. Vander Heiden, L. C. Cantley, and C. B. Thompson, "Understanding the warburg effect: the metabolic requirements of cell proliferation," *Science*, vol. 324, no. 5930, pp. 1029–1033, 2009.
- [91] D. Hanahan and R. A. Weinberg, "The hallmarks of cancer," *Cell*, vol. 100, no. 1, pp. 57–70, 2000.
- [92] Q. Duan, D. Li, L. Xiong, Z. Chang, and G. Xu, "Silac quantitative proteomics and biochemical analyses reveal a novel molecular mechanism by which adam12s promotes the proliferation, migration, and invasion of small cell lung cancer cells through upregulating hexokinase 1," *Journal of Proteome Research*, vol. 18, no. 7, pp. 2903–2914, 2019.
- [93] R. A. Nakashima, P. S. Mangan, M. Colombini, and P. L. Pedersen, "Hexokinase receptor complex in hepatoma mitochondria: evidence from n, n²-dicyclohexylcarbodiimide-labeling studies for the involvement of the pore-forming protein vdac," *Biochemistry*, vol. 25, no. 5, pp. 1015–1021, 1986.
- [94] K. F. Ferri and G. Kroemer, "Mitochondria—the suicide organelles," *Bioessays*, vol. 23, no. 2, pp. 111–115, 2001.
- [95] E. Blachly-Dyson and M. Forte, "Vdac channels," *IUBMB life*, vol. 52, no. 3-5, pp. 113–118, 2001.
- [96] V. Shoshan-Barmatz, A. Israelson, D. Brdiczka, , and S. Sheu, "The voltage-dependent anion channel (vdac): function in intracellular signalling, cell life and cell death," *Current Pharmaceutical Design*, vol. 12, no. 18, pp. 2249–2270, 2006.
- [97] X. Liu, C. S. Kim, F. T. Kurbanov, R. B. Honzatko, and H. J. Fromm, "Dual mechanisms for glucose 6-phosphate inhibition of human brain hexokinase," *Journal of Biological Chemistry*, vol. 274, no. 44, pp. 31 155–31 159, 1999.
- [98] J. Wilson, "Hexokinases," *Reviews of Physiology, Biochemistry and Pharmacology, Volume 126*, pp. 65–198, 1995.
- [99] A. E. Aleshin, C. Zeng, H. D. Bartunik, H. J. Fromm, and R. B. Honzatko, "Regulation of hexokinase i: crystal structure of recombinant human brain

- hexokinase complexed with glucose and phosphate,” *Journal of Molecular Biology*, vol. 282, no. 2, pp. 345–357, 1998.
- [100] H. M. Katzen and R. T. Schimke, “Multiple forms of hexokinase in the rat: tissue distribution, age dependency, and properties.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 54, no. 4, p. 1218, 1965.
- [101] L. Grossbard and R. T. Schimke, “Multiple hexokinases of rat tissues: purification and comparison of soluble forms,” *Journal of Biological Chemistry*, vol. 241, no. 15, pp. 3546–3560, 1966.
- [102] G. Xie and J. E. Wilson, “Tetrameric structure of mitochondrially bound rat brain hexokinase: a crosslinking study,” *Archives of Biochemistry and Biophysics*, vol. 276, no. 1, pp. 285–293, 1990.
- [103] Y. Zhao, E. B. Butler, and M. Tan, “Targeting cellular metabolism to improve cancer therapeutics,” *Cell Death & Disease*, vol. 4, no. 3, pp. e532–e532, 2013.
- [104] B. M. Madhok, S. Yeluri, S. L. Perry, T. A. Hughes, and D. G. Jayne, “Targeting glucose metabolism: an emerging concept for anticancer therapy,” *American Journal of Clinical Oncology*, vol. 34, no. 6, pp. 628–635, 2011.
- [105] S. N. Garcia, R. C. Guedes, and M. M. Marques, “Unlocking the potential of hk2 in cancer metabolism and therapeutics,” *Current Medicinal Chemistry*, vol. 26, no. 41, pp. 7285–7322, 2019.
- [106] S. Ganapathy-Kanniappan, M. Vali, R. Kunjithapatham, M. Buijs, L. Syed, P. Rao, S. Ota, B. Kwak, R. Loffroy, and J. Geschwind, “3-bromopyruvate: a new targeted antiglycolytic agent and a promise for cancer therapy,” *Current Pharmaceutical Biotechnology*, vol. 11, no. 5, pp. 510–517, 2010.
- [107] H. Pelicano, D. Martin, R. Xu, , and P. Huang, “Glycolysis inhibition for anti-cancer treatment,” *Oncogene*, vol. 25, no. 34, pp. 4633–4646, 2006.
- [108] W. Kim, J.-H. Yoon, J.-M. Jeong, G.-J. Cheon, T.-S. Lee, J.-I. Yang, S.-C. Park, and H.-S. Lee, “Apoptosis-inducing antitumor efficacy of hexokinase ii inhibitor in hepatocellular carcinoma,” *Molecular Cancer Therapeutics*, vol. 6, no. 9, pp. 2554–2562, 2007.
- [109] Q. Du, Y. Wang, C. Liu, H. Wang, H. Fan, Y. Li, J. Wang, X. Zhang, J. Lu, H. Ji *et al.*, “Chemopreventive activity of gen-27, a genistein derivative, in colitis-associated cancer is mediated by p65-cdx2- β -catenin axis,” *Oncotarget*, vol. 7, no. 14, p. 17870, 2016.
- [110] W. Li, M. Zheng, S. Wu, S. Gao, M. Yang, Z. Li, Q. Min, W. Sun, L. Chen, G. Xiang *et al.*, “Benserazide, a dopadecarboxylase inhibitor, suppresses tumor growth by targeting hexokinase 2,” *Journal of Experimental & Clinical Cancer Research*, vol. 36, no. 1, pp. 1–12, 2017.
- [111] J. M. Berg, J. L. Tymoczko, and L. Stryer, “Protein structure and function,” *Biochemistry*, vol. 262, pp. 159–173, 2002.
- [112] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, “The energy landscapes and motions of proteins,” *Science*, vol. 254, no. 5038, pp. 1598–1603, 1991.

- [113] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, “Funnel, pathways, and the energy landscape of protein folding: a synthesis,” *Proteins: Structure, Function, and Bioinformatics*, vol. 21, no. 3, pp. 167–195, 1995.
- [114] H. A. Carlson, “Protein flexibility and drug design: how to hit a moving target,” *Current Opinion in Chemical Biology*, vol. 6, no. 4, pp. 447–452, 2002.
- [115] D. D. Boehr, R. Nussinov, and P. E. Wright, “The role of dynamic conformational ensembles in biomolecular recognition,” *Nature Chemical Biology*, vol. 5, no. 11, pp. 789–796, 2009.
- [116] M. Parker, “Protein structure from x-ray diffraction,” *Journal of Biological Physics*, vol. 29, no. 4, pp. 341–362, 2003.
- [117] A. Ansari, J. Berendzen, S. Bowne, and H. Frauenfelder, “men, iet, sauke, tb, shyamsunder, e., young, rd protein states and proteinquakes,” *Proc. Natl. Acad. Sci. USA*, vol. 82, pp. 5000–5004, 1985.
- [118] R. Nussinov and C.-J. Tsai, “Free energy diagrams for protein function,” *Chemistry & Biology*, vol. 21, no. 3, pp. 311–318, 2014.
- [119] K. Lindorff-Larsen, R. B. Best, M. A. DePristo, C. M. Dobson, and M. Vendruscolo, “Simultaneous determination of protein structure and dynamics,” *Nature*, vol. 433, no. 7022, pp. 128–132, 2005.
- [120] E. J. Levin, D. A. Kondrashov, G. E. Wesenberg, and G. N. Phillips Jr, “Ensemble refinement of protein crystal structures: validation and application,” *Structure*, vol. 15, no. 9, pp. 1040–1052, 2007.
- [121] R. W. Pastor and M. Karplus, “Parametrization of the friction constant for stochastic simulations of polymers,” *The Journal of Physical Chemistry*, vol. 92, no. 9, pp. 2636–2641, 1988.
- [122] M. J. Sippl, “Calculation of conformational ensembles from potentials of mean force: an approach to the knowledge-based prediction of local structures in globular proteins,” *Journal of Molecular Biology*, vol. 213, no. 4, pp. 859–883, 1990.
- [123] J. McCammon, “Protein dynamics,” *Reports on Progress in Physics*, vol. 47, no. 1, p. 1, 1984.
- [124] M. Karplus and J. A. McCammon, “Molecular dynamics simulations of biomolecules,” *Nature Structural Biology*, vol. 9, no. 9, pp. 646–652, 2002.
- [125] K. Teilum, J. G. Olsen, and B. B. Kragelund, “Functional aspects of protein flexibility,” *Cellular and Molecular Life Sciences*, vol. 66, no. 14, pp. 2231–2247, 2009.
- [126] M. Levitt and A. Warshel, “Computer simulation of protein folding,” *Nature*, vol. 253, no. 5494, pp. 694–698, 1975.
- [127] J. A. McCammon, B. R. Gelin, and M. Karplus, “Dynamics of folded proteins,” *Nature*, vol. 267, no. 5612, pp. 585–590, 1977.
- [128] H. A. Scheraga, M. Khalili, and A. Liwo, “Protein-folding dynamics: overview of molecular simulation techniques,” *Annu. Rev. Phys. Chem.*, vol. 58, pp. 57–83, 2007.

- [129] J. D. Durrant and J. A. McCammon, “Molecular dynamics simulations and drug discovery,” *BMC Biology*, vol. 9, no. 1, pp. 1–9, 2011.
- [130] D. Hamelberg, J. Mongan, and J. A. McCammon, “Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules,” *The Journal of Chemical Physics*, vol. 120, no. 24, pp. 11 919–11 929, 2004.
- [131] P. R. Markwick, L. C. Pierce, D. B. Goodin, and J. A. McCammon, “Adaptive accelerated molecular dynamics (ad-amd) revealing the molecular plasticity of p450cam,” *The Journal of Physical Chemistry Letters*, vol. 2, no. 3, pp. 158–164, 2011.
- [132] D. Hamelberg, C. A. F. de Oliveira, and J. A. McCammon, “Sampling of slow diffusive conformational transitions with accelerated molecular dynamics,” *The Journal of Chemical Physics*, vol. 127, no. 15, p. 10B614, 2007.
- [133] L. C. Pierce, R. Salomon-Ferrer, C. Augusto F. de Oliveira, J. A. McCammon, and R. C. Walker, “Routine access to millisecond time scale events with accelerated molecular dynamics,” *Journal of Chemical Theory and Computation*, vol. 8, no. 9, pp. 2997–3002, 2012.
- [134] Y. Miao, W. Sinko, L. Pierce, D. Bucher, R. C. Walker, and J. A. McCammon, “Improved reweighting of accelerated molecular dynamics simulations for free energy calculation,” *J. Chem. Theory Comput.*, vol. 10, no. 7, pp. 2677–2689, 2014.
- [135] G. Downes and N. Gautam, “The g protein subunit gene families,” *Genomics*, vol. 62, no. 3, pp. 544–552, 1999.
- [136] J. Garcia-Nafria, R. Nehme, P. C. Edwards, and C. G. Tate, “Cryo-em structure of the serotonin 5-ht 1b receptor coupled to heterotrimeric g o,” *Nature*, vol. 558, no. 7711, pp. 620–623, 2018.
- [137] Y. I. Wolf, S. E. Brenner, P. A. Bash, and E. V. Koonin, “Distribution of protein folds in the three superkingdoms of life,” *Genome Research*, vol. 9, no. 1, pp. 17–26, 1999.
- [138] M. C. Michel and S. J. Charlton, “Biased agonism in drug discovery—is it too soon to choose a path?” *Molecular Pharmacology*, vol. 93, no. 4, pp. 259–265, 2018.
- [139] J. D. Violin, A. L. Crombie, D. G. Soergel, and M. W. Lark, “Biased ligands at g-protein-coupled receptors: promise and progress,” *Trends in Pharmacological Sciences*, vol. 35, no. 7, pp. 308–316, 2014.
- [140] J. D. McCorvy, K. V. Butler, B. Kelly, K. Rechsteiner, J. Karpiak, R. M. Betz, B. L. Kormos, B. K. Shoichet, R. O. Dror, J. Jin *et al.*, “Structure-inspired design of β -arrestin-biased ligands for aminergic gpcrs,” *Nature Chemical Biology*, vol. 14, no. 2, pp. 126–134, 2018.
- [141] J. Gundry, R. Glenn, P. Alagesan, and S. Rajagopal, “A practical guide to approaching biased agonism at g protein coupled receptors,” *Frontiers in Neuroscience*, vol. 11, p. 17, 2017.
- [142] J.-x. Cheng, T. Cheng, W.-h. Li, G.-x. Liu, W.-l. Zhu, and Y. Tang, “Computational insights into the g-protein-biased activation and inactivation mechanisms of the μ opioid receptor,” *Acta Pharmacologica Sinica*, vol. 39, no. 1, pp. 154–164, 2018.

- [143] M. R. Bruchas and B. L. Roth, “New technologies for elucidating opioid receptor function,” *Trends in Pharmacological Sciences*, vol. 37, no. 4, pp. 279–289, 2016.
- [144] L. M. Bohn, R. R. Gainetdinov, F.-T. Lin, R. J. Lefkowitz, and M. G. Caron, “ μ -opioid receptor desensitization by β -arrestin-2 determines morphine tolerance but not dependence,” *Nature*, vol. 408, no. 6813, pp. 720–723, 2000.
- [145] A. R. Fersht and S. Sato, “ ϕ -value analysis and the nature of protein-folding transition states,” *Proceedings of the National Academy of Sciences*, vol. 101, no. 21, pp. 7976–7981, 2004.
- [146] P. Neudecker, A. Zarrine-Afsar, A. R. Davidson, and L. E. Kay, “ ϕ -value analysis of a three-state protein folding pathway by nmr relaxation dispersion spectroscopy,” *Proceedings of the National Academy of Sciences*, vol. 104, no. 40, pp. 15 717–15 722, 2007.
- [147] H. Feng, N.-D. Vu, Z. Zhou, and Y. Bai, “Structural examination of ϕ -value analysis in protein folding,” *Biochemistry*, vol. 43, no. 45, pp. 14 325–14 331, 2004.
- [148] V. Daggett and A. R. Fersht, “Is there a unifying mechanism for protein folding?” *Trends in Biochemical Sciences*, vol. 28, no. 1, pp. 18–25, 2003.
- [149] S. Gianni, N. R. Guydosh, F. Khan, T. D. Caldas, U. Mayor, G. W. White, M. L. DeMarco, V. Daggett, and A. R. Fersht, “Unifying features in protein-folding mechanisms,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 23, pp. 13 286–13 291, 2003.
- [150] A. Matouschek, J. T. Kellis, L. Serrano, and A. R. Fersht, “Mapping the transition state and pathway of protein folding by protein engineering,” *Nature*, vol. 340, no. 6229, pp. 122–126, 1989.
- [151] W. A. Eaton, E. R. Henry, and J. Hofrichter, “Application of linear free energy relations to protein conformational changes: the quaternary structural change of hemoglobin,” *Proceedings of the National Academy of Sciences*, vol. 88, no. 10, pp. 4472–4475, 1991.
- [152] A. Matouschek, D. E. Otzen, L. S. Itzhaki, S. E. Jackson, and A. R. Fersht, “Movement of the position of the transition state in protein folding,” *Biochemistry*, vol. 34, no. 41, pp. 13 656–13 662, 1995.
- [153] A. Li and V. Daggett, “Characterization of the transition state of protein unfolding by use of molecular dynamics: chymotrypsin inhibitor 2,” *Proceedings of the National Academy of Sciences*, vol. 91, no. 22, pp. 10 430–10 434, 1994.
- [154] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan *et al.*, “Atomic-level characterization of the structural dynamics of proteins,” *Science*, vol. 330, no. 6002, pp. 341–346, 2010.
- [155] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, “How fast-folding proteins fold,” *Science*, vol. 334, no. 6055, pp. 517–520, 2011.
- [156] C. D. Geierhaas, R. B. Best, E. Paci, M. Vendruscolo, and J. Clarke, “Structural comparison of the two alternative transition states for folding of ti i27,” *Biophysical Journal*, vol. 91, no. 1, pp. 263–275, 2006.

- [157] B. Carpenter, R. Nehmé, T. Warne, A. G. Leslie, and C. G. Tate, “Structure of the adenosine a2a receptor bound to an engineered g protein,” *Nature*, vol. 536, no. 7614, pp. 104–107, 2016.
- [158] W. Liu, E. Chun, A. A. Thompson, P. Chubukov, F. Xu, V. Katritch, G. W. Han, C. B. Roth, L. H. Heitman, A. P. IJzerman, V. S. Cherezov, and R. C., “Structural basis for allosteric regulation of gpcrs by sodium ions,” *Publ. Am. Assoc. Adv. Sci.*, vol. 337, no. 6091, pp. 232–236, 2012.
- [159] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, P. F. David E. Shaw, and P. S. Shenkin, “Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy,” *J. Med. Chem.*, vol. 47, no. 7, pp. 1739–1749, 2004.
- [160] T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard, and J. L. Banks, “Glide: a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening,” *J. Med. Chem.*, vol. 47, no. 7, pp. 1750–1759, 2004.
- [161] J. A. Ballesteros and H. Weinstein, “[19] integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in g protein-coupled receptors,” in *Methods Neurosci.*, 1995, vol. 25, pp. 366–428.
- [162] S. Wang, T. Che, A. Levit, B. K. Shoichet, D. Wacker, and B. L. Roth, “Structure of the d2 dopamine receptor bound to the atypical antipsychotic drug risperidone,” *Nature*, vol. 555, no. 7695, pp. 269–273, 2018.
- [163] Y. Kang, O. Kuybeda, P. W. de Waal, S. Mukherjee, N. Van Eps, P. Dutka, X. E. Zhou, A. Bartesaghi, S. Erramilli, T. Morizumi *et al.*, “Cryo-em structure of human rhodopsin bound to an inhibitory g protein,” *Nature*, vol. 558, no. 7711, pp. 553–558, 2018.
- [164] N. Robertson, M. Rappas, A. S. Doré, J. Brown, G. Bottegoni, M. Koglin, J. Cansfield, A. Jazayeri, R. M. Cooke, and F. H. Marshall, “Structure of the complement c5a receptor bound to the extra-helical antagonist ndt9513727,” *Nature*, vol. 553, no. 7686, pp. 111–114, 2018.
- [165] G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman, “Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments,” *J. Comput.-Aided Mol. Des.*, vol. 27, no. 3, pp. 221–234, 2013.
- [166] K. Zhu, T. Day, D. Warshaviak, C. Murrett, R. Friesner, and D. Pearlman, “Antibody structure determination using a combination of homology modeling, energy-based refinement, and loop prediction,” *Proteins: Struct., Funct., Bioinf.*, vol. 82, no. 8, pp. 1646–1655, 2014.
- [167] C. R. Søndergaard, M. H. Olsson, M. Rostkowski, and J. H. Jensen, “Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pka values,” *J. Chem. Theory Comput.*, vol. 7, no. 7, pp. 2284–2295, 2011.
- [168] A. L. Lomize, I. D. Pogozheva, and H. I. Mosberg, “Anisotropic solvent model of the lipid bilayer. 2. energetics of insertion of small molecules, peptides, and

- proteins in membranes,” *J. Chem. Inf. Model.*, vol. 51, no. 4, pp. 930–946, 2011.
- [169] M. A. Lomize, I. D. Pogozheva, H. Joo, H. I. Mosberg, and A. L. Lomize, “Opm database and ppm web server: resources for positioning of proteins in membranes,” *Nucleic Acids Res.*, vol. 40, no. D1, pp. D370–D376, 2011.
- [170] J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O’Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell Jr, and R. W. Pastor, “Update of the charmm all-atom additive force field for lipids: validation on six lipid types,” *J. Phys. Chem. B.*, vol. 114, no. 23, pp. 7830–7843, 2010.
- [171] S. Jo, T. Kim, V. G. Iyer, and W. Im, “Charmm-gui: a web-based graphical user interface for charmm,” *J. Comput. Chem.*, vol. 29, no. 11, pp. 1859–1865, 2008.
- [172] P. Mark and L. Nilsson, “Structure and dynamics of the tip3p, spc, and spc/e water models at 298 k,” *J. Phys. Chem. A*, vol. 105, no. 43, pp. 9954–9960, 2001.
- [173] J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, V. S. P. Sunhwan Jo, C. L. B. David A. Case, J. B. K. Alexander D. MacKerell, and W. Im, “Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field,” *J. Chem. Theory Comput.*, vol. 12, no. 1, pp. 405–413, 2015.
- [174] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, V. I. Lopes, P, and M. A. Jr, “Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields,” *J. Comput. Chem.*, vol. 31, no. 4, pp. 671–690, 2010.
- [175] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, “Scalable molecular dynamics with namd,” *J. Comput. Chem.*, vol. 26, no. 16, pp. 1781–1802, 2005.
- [176] D. J. Evans and B. L. Holian, “The nose–hoover thermostat,” *J. Chem. Phys.*, vol. 83, no. 8, pp. 4069–4074, 1985.
- [177] M. Parrinello and A. Rahman, “Polymorphic transitions in single crystals: A new molecular dynamics method,” *J. Appl. Phys.*, vol. 52, no. 12, pp. 7182–7190, 1981.
- [178] Y. Miao, S. E. Nichols, P. M. Gasper, V. T. Metzger, and J. A. McCammon, “Activation and dynamic network of the m2 muscarinic receptor,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 27, pp. 10 982–10 987, 2013.
- [179] I. G. Tikhonova, B. Selvam, A. Ivetac, J. Wereszczynski, and J. A. McCammon, “Simulations of biased agonists in the β 2 adrenergic receptor with accelerated molecular dynamics,” *Biochemistry*, vol. 52, no. 33, pp. 5593–5603, 2013.
- [180] K. Kasahara, I. Fukuda, and H. Nakamura, “A novel approach of dynamic cross correlation analysis on molecular dynamics simulations and its application to ets1 dimer–dna complex,” *PloS One*, vol. 9, no. 11, p. e112419, 2014.

- [181] M. Bhattacharyya, C. R. Bhat, and S. Vishveshwara, “An automated approach to network features of protein structure ensembles,” *Protein. Sci.*, vol. 22, no. 10, pp. 1399–1416, 2013.
- [182] S. H. Oliveira, F. A. Ferraz, R. V. Honorato, J. Xavier-Neto, T. J. Sobreira, and P. S. de Oliveira, “Kvfinder: steered identification of protein cavities as a pymol plugin,” *BMC Bioinf.*, vol. 15, no. 1, p. 197, 2014.
- [183] C. A. Johnston, K. Afshar, J. T. Snyder, G. G. Tall, P. Gönczy, D. P. Siderovski, and F. S. Willard, “Structural determinants underlying the temperature-sensitive nature of a $g\alpha$ mutant in asymmetric cell division of *caenorhabditis elegans*,” *Journal of Biological Chemistry*, vol. 283, no. 31, pp. 21 550–21 558, 2008.
- [184] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia *et al.*, “Plumed: A portable plugin for free-energy calculations with molecular dynamics,” *Computer Physics Communications*, vol. 180, no. 10, pp. 1961–1972, 2009.
- [185] M. H. Nawaz, J. C. Ferreira, L. Nedyalkova, H. Zhu, C. Carrasco-López, S. Kirmizialtin, and W. M. Rabeh, “The catalytic inactivation of the n-half of human hexokinase 2 and structural and biochemical characterization of its mitochondrial conformation,” *Bioscience Reports*, vol. 38, no. 1, 2018.
- [186] J. Huang and A. D. MacKerell Jr, “Charmm36 all-atom additive protein force field: Validation based on comparison to nmr data,” *Journal of Computational Chemistry*, vol. 34, no. 25, pp. 2135–2145, 2013.
- [187] P. T. Kiss and A. Baranyai, “A systematic development of a polarizable potential of water,” *The Journal of Chemical Physics*, vol. 138, no. 20, p. 204507, 2013.
- [188] H. J. Berendsen, D. van der Spoel, and R. van Drunen, “Gromacs: a message-passing parallel molecular dynamics implementation,” *Comput. Phys. Commun.*, no. 1-3, pp. 43–56, 1995.
- [189] S. Nosé and M. Klein, “Constant pressure molecular dynamics for molecular systems,” *Molecular Physics*, vol. 50, no. 5, pp. 1055–1076, 1983.
- [190] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, “A smooth particle mesh ewald method,” *The Journal of Chemical Physics*, vol. 103, no. 19, pp. 8577–8593, 1995.
- [191] S. LigPrep, “2: Ligprep, schrödinger, llc, new york, ny, 2018,” *New York, NY*, 2018.
- [192] S. Yuan, S. Filipek, K. Palczewski, and H. Vogel, “Activation of g-protein-coupled receptors correlates with the formation of a continuous internal water pathway,” *Nat. Commun.*, vol. 5, no. 1, pp. 1–10, 2014.
- [193] Y. Miao and J. A. McCammon, “Mechanism of the g-protein mimetic nanobody binding to a muscarinic g-protein-coupled receptor,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 115, no. 12, pp. 3036–3041, 2018.
- [194] A. S. Doré, N. Robertson, J. C. Errey, I. Ng, K. Hollenstein, B. Tehan, E. Hurrell, K. Bennett, M. Congreve, F. Magnani, C. G. Tate, M. Weir, and F. H. Marshall, “Structure of the adenosine a2a receptor in complex with zm241385

- and the xanthines xac and caffeine,” *Structure*, vol. 19, no. 9, pp. 1283–1293, 2011.
- [195] F. Xu, H. Wu, V. Katritch, G. W. Han, K. A. Jacobson, Z.-G. Gao, V. Cherezov, and R. C. Stevens, “Structure of an agonist-bound human a2a adenosine receptor,” *Publ. Am. Assoc. Adv. Sci.*, vol. 332, no. 6027, pp. 322–327, 2011.
- [196] C. McGraw, L. Yang, I. Levental, E. Lyman, and A. S. Robinson, “Membrane cholesterol depletion reduces downstream signaling activity of the adenosine a2a receptor,” *Biochim. Biophys. Acta, Biomembr.*, vol. 1861, no. 4, pp. 760–767, 2019.
- [197] R. C. Kling, T. Clark, and P. Gmeiner, “Comparative md simulations indicate a dual role for arg1323. 50 in dopamine-dependent d2r activation,” *PloS One*, vol. 11, no. 1, p. e0146612, 2016.
- [198] W. M. Oldham and H. E. Hamm, “Heterotrimeric g protein activation by g-protein-coupled receptors,” *Nature Reviews Molecular Cell Biology*, vol. 9, no. 1, pp. 60–71, 2008.
- [199] X. Sun, S. Singh, K. J. Blumer, and G. R. Bowman, “Simulation of spontaneous g protein activation reveals a new intermediate driving gdp unbinding,” *Elife*, vol. 7, p. e38465, 2018.

CURRICULUM VITAE

Name Surname : Samman Mansoor

E-Mail :

EDUCATION:

B.Sc. : 2015, University of Gujrat Pakistan, Biochemistry and Molecular Biology

M.Sc. : 2017, University of Gujrat Pakistan, Biochemistry and Biotechnology

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2017-2021, Research Assistant
- BAP-Graduate Education Incentive Grant (Istanbul Medipol University 2017-2021).
- Travel reimbursement for MuTaLig COST Action Meeting in Malta (2018)
- Travel reimbursement for BlmBS 2019 Training School (Lugano 2019)

PUBLICATIONS, PRESENTATIONS, AND PATENTS ON THE THESIS:

- M. Ilter, S. Mansoor and O. Sensoy, “ Utilization of biased G protein-coupled receptor signaling towards development of safer and personalized therapeutics,” *Molecules*, 24(11), 2052, May. 2019, doi: 10.3390/molecules24112052.
- J.C. Ferreira, A.R. Khrbtli, C.L. Shetler, S. Mansoor, L. Ali, O. Sensoy and W.M. Rabeh, “Linker residues regulate the activity and stability of hexokinase 2, a promising anticancer target,” *Int. J. Biol. Chem.*, vol. 296, 100071, Jan. 2021, doi: doi.org/10.1074/jbc.RA120.015293.
- S. Mansoor, G. Kayik, S. Durdagi and O. Sensoy, “Understanding the molecular basis of bivalent ligand-mediated affinity and efficacy modulation in G protein-coupled receptor oligomer,” *Comput. Struct. Biotechnol. J.* (Manuscript under revision)
- S. Mansoor, G. Kayik, E. Alshahaby, S. Durdagi, M. Guzel and O. Sensoy, “Development of novel therapeutic that can target multiple receptors for treatment of Parkinson’s disease,” in *3rd Annual meeting MuTaLig Cost Action CA-15135, Valletta (Malta)*, 2018.
- S. Mansoor, G. Kayik, S. Durdagi and O. Sensoy, “Understanding Allosteric Interaction Network within $A_{2A}R/A_{2A}R/G\alpha s-D_2R/D_2R-G\alpha i$ complex bound to Heterobivalent Ligand,” in *6th International BAU Drug Design Congress, Turkey* 2018.
- Application of modified phi-values for capturing intermediate states in G protein activation mechanism. (Manuscript under preparation)

- U.S accepted International Patent Application “Heterobivalent ligands suitable for use in the treatment of Parkinson’s Disease” *Patent number: WO2019004970*



COMPUTATIONAL INVESTIGATION AND MODULATION OF STRUCTURAL AND FUNCTIONAL PROPERTIES OF PROTEINS FOR THERAPEUTIC PURPOSES

ORIGINALITY REPORT

13%

SIMILARITY INDEX

9%

INTERNET SOURCES

9%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	1%
2	J A McCammon. Reports on Progress in Physics, 01/1984 Publication	1%
3	pubs.rsc.org Internet Source	1%
4	www.nature.com Internet Source	<1%
5	Submitted to Eastern Mediterranean University Student Paper	<1%
6	pubs.acs.org Internet Source	<1%
7	repository.bilkent.edu.tr Internet Source	<1%
8	d-nb.info Internet Source	<1%