

ROLE OF ELECTROSTATIC INTERACTIONS, AROMATICITY, HYDROPHOBICITY IN PROTEIN AND PROTEIN-LIGAND DYNAMICS

Thesis

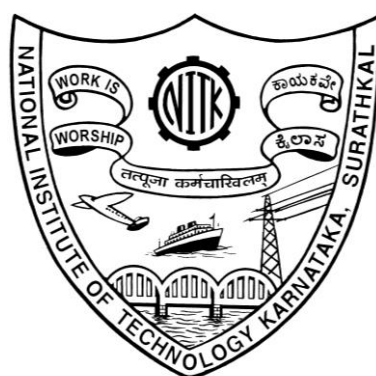
Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

PUSHYARAGA P V

(Reg. No: 177048CY005)



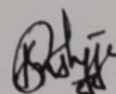
**DEPARTMENT OF CHEMISTRY
NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA
SURATHKAL, MANGALORE-575025**

APRIL, 2023

DECLARATION

By the Ph.D Research Scholar

I hereby declare that the Research Thesis entitled "**Role of Electrostatic Interactions, Aromaticity, Hydrophobicity in Protein and Protein-Ligand dynamics**" which is being submitted to the National Institute of Technology Karnataka, Surathkal in partial fulfillment of the requirements for the award of the degree of **Doctor of Philosophy in Chemistry** is a *bonafide report of the research work carried out by me*. The material contained in this Research Thesis has not been submitted to any University or Institution for the award of any degree.



PUSHYARAGA P V

Reg.No. 177048CY005

Department of Chemistry

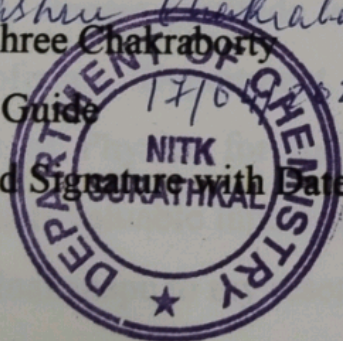
Place: NITK, Surathkal

Date: 17-04-2023

CERTIFICATE

This is to *certify* that the Research Thesis entitled “**Role of Electrostatic Interactions, Aromaticity, Hydrophobicity in Protein and Protein-Ligand dynamics**” submitted by Pushyaraga P V, (Register Number: 177048CY005) as the record of the research work carried out by her, is *accepted as the Research Thesis submission* in partial fulfilment of the requirements for the award of degree of Doctor of Philosophy.

Debashree Chakraborty
Dr Debashree Chakraborty
Research Guide 17/04/23
(Name and Signature with Date)



Wagabennadi
Chairman DRPC 17/04/23

(Signature with Date and Seal)
HEAD, DEPARTMENT OF CHEMISTRY
National Institute of Technology Karnataka
Surathkal, Srinivasnagar
MANGALORE- 575 025, D.K.

ACKNOWLEDGEMENTS

I take this opportunity to express my sincere gratitude and deep regards to all those who have helped me to complete the doctoral research work. First and foremost, I would like to express my deep gratitude to my supervisor, Dr. Debashree Chakraborty, Assistant Professor, Department of Chemistry, NITK, for her valuable guidance, suggestions, patience and encouragement. I vow my immense gratitude to her for giving me an opportunity to work under her and for providing the necessary facilities to work on this project.

I acknowledge NITK, Surathkal for providing the research fellowship, laboratory facilities and financial support necessary for the completion of my research work. Also, I'm thankful to Department of Physics and Department of Computer Science & Engineering, NITK Surathkal for additional computational facilities. I'm deeply grateful to Schrödinger Inc. for providing the trial license twice in the years 2017 and 2020.

My sincere gratitude is due towards my RPAC members, Dr. A Nityananda Shetty, Professor, Department of Chemistry and Dr. H D Shashikala, Professor, Department of Physics for their timely assessment and evaluation of my research progress. Their valuable inputs at various stages of my work have contributed greatly in giving the final shape to my research work.

I'm grateful to Dr. Uday Kumar D, Head of the Department & Associate Professor and Dr. Arun M Isloor and Dr. D Krishna Bhat, former Heads of the Department for providing the administrative facilities and infrastructure. I also thank all the faculty members of Department of Chemistry, NITK Surathkal for their support.

I sincerely thank my senior, Dilip and all other lab members Bratin, Omkar, Apoorva, Anjana, Divya, Samrat for their constant support at every stage of my research. I am thankful to all the research scholars and non-teaching staffs within and outside the department for their valuable assistance. I owe a deep gratitude to the

professionals and the well-wishers within the country and abroad who have guided and helped me during the course of research work.

I am grateful for supportive and caring family members and in-laws. It is impossible to extend enough gratitude to my beloved parents, Mr. C K Venugopalan and Mrs. P S Pushpakala, who have always been the source of inspiration at every stage of my life. I express my deep gratitude to my spouse, Mr. Omkar Singh for the constant support and encouragement he has given me all the time. I would also like to thank my extended family members for their continuous moral support.

Finally, I thank to all those who have helped me directly or indirectly and extended their support in the completion of my doctoral research work.

Thanking you.

PUSHYARAGA P V.

ABSTRACT

The process of drug discovery and development is time and resource consuming. The incorporation of computational facilities to combine chemical and biological aspects helps in drug discovery, design, development and optimization speeds up the process. Advancement in the structural biology and modern bioinformatics helps to design small molecule drug candidates with better biological activity and target specificity. Considering the fact that the protein function depends on its 3D structure, the complexity of a particular disease can be understood by deducing the dysfunctionality at molecular level. This aids the designing of efficient drug molecules. Various non-covalent interactions such as hydrogen bonding, hydrophobic interaction and aromatic interactions plays a key role in stabilizing the energetically-favored ligands at the binding site of the protein. To understand about the various interactions, a number of protein-ligand systems were studied namely, binding affinity of β -alanine derivatives with human G-protein coupled receptors, benzimidazole derivatives with poly ADP-ribose polymerase enzyme and main protease inhibitors for SARS CoV-2 main protease. The results showed that hydrophobicity, aromaticity and electrostatic interactions are crucial for stabilizing ligands at the target site. The nature and strength of non-covalent interactions can alter the binding affinity and efficacy of drug molecules towards the target. Apart from non-covalent interactions, water molecules at the binding site of protein stabilize the binding site as well as the ligand through hydrogen bonding interactions. The molecular level knowledge of protein dynamics and protein-ligand dynamics can have a positive impact on the drug development process and can accelerate the steps of drug designing with reduced cost and time.

Keywords: Molecular docking, MD simulation, Quantum chemical calculations, protein-drug interactions, GCGR, PARP enzymes, SARS CoV-2 main protease

INDEX

SI No	Contents	Page No:
CHAPTER 1- INTRODUCTION		
1.1	SIGNIFICANCE OF PROTEINS	1
1.2	DRUG DISCOVERY AND DEVELOPMENT PROCESS	3
1.2.1	Brief History of CADD	4
1.2.2	Workflow for CADD	5
1.2.3	Structure-based Drug Discovery	5
1.2.4	Ligand-based Drug Discovery	6
1.3	SIGNIFICANCE OF NON-COVALENT INTERACTIONS	6
1.3	LITERATURE SURVEY	7
1.3.1	G-protein coupled glucagon receptors	11
1.3.2	Poly (ADP-ribose) polymerase 1 & 2	12
1.3.3	Coronavirus main protease	14
1.4	OBJECTIVES OF THE WORK	16
CHAPTER 2- MATERIALS AND METHODS		
2.1	MOLECULAR DYNAMICS SIMULATION	17
2.1.1	Theory	18
2.1.2	Force Fields	20

2.2	DENSITY FUNCTIONAL THEORY	21
2.2.1	Hohenberg-Kohn Theorms	22
2.2.2	Kohn-Sham Equations	23
2.3	MOLECULAR DOCKING	24
2.3.1	Basic Requirements for Docking	25
2.3.2	Types of Molecular Docking	26
2.4	PHARMACOPHORE MODELLING	26
2.4.1	Pharmacophore Sites	27
2.5	QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP	28
2.5.1	Types of SAR Approach	29
2.5.2	Molecular Descriptors for QSAR Study	30
2.5.3	Chemometric Methods	31
2.6	ADME/TOXICITY PREDICTION	33
2.7	DATA ANALYSES FROM SIMULATION TRAJECTORY	34
2.7.1	Network Analysis	34
2.7.2	Protein Structural Parameters	35
2.7.3	Radial Distribution Function	37
2.7.4	Tetrahedral Order Parameter	37
2.7.5	Hydrogen Bond dynamics	38
2.7.6	Free Energy Calculations	38

**CHAPTER 3- EFFECT OF HYDROPHOBIC AND
HYDROGEN BONDING INTERACTIONS ON THE
POTENCY OF β -ALANINE ANALOGUES OF G-PROTEIN
COUPLED GLUCAGON RECEPTOR INHIBITORS**

3.1	BACKGROUND	41
3.2	DATASET AND PREPARATION OF 3D STRUCTURES OF LIGAND	42
3.3	3D-QSAR ANALYSIS	43
3.4	CONTOUR PLOT ANALYSIS	46
3.5	MOLECULAR DOCKING	48
3.6	FRONTIER MOLECULAR ORBITAL ANALYSIS	54
3.7	MM/GBSA CALCULATION	57
3.8	MOLECULAR DYNAMICS SIMULATION	58
3.9	ADME/ TOXICITY PREDICTION	64
3.10	CONCLUSION	65

**CHAPTER 4- THEORETICAL INSIGHTS INTO
MOLECULAR MECHANISM AND ENERGY CRITERIA
OF PARP-2 ENZYME INHIBITION BY BENZIMIDAZOLE
ANALOGUES**

4.1	BACKGROUND	67
4.2	DATASET AND PREPARATION OF 3D STRUCTURES OF LIGAND	68
4.3	MOLECULAR DOCKING	69

4.4	FRONTIER MOLECULAR ORBITAL ANALYSIS AND REACTIVITY DESCRIPTORS	72
4.5	MOLECULAR DYNAMICS SIMULATION STUDY	76
4.6	LIGAND ORIENTATION: HYDROGEN BONDING AND AROMATIC RING STACKING	85
4.7	PCA AND FEL ANALYSIS OF COMPLEX	87
4.8	DESIGN OF NEW POTENT COMPOUNDS	92
4.9	CONCLUSION	95
CHAPTER 5- SINGLE AMINO ACID SUBSTITUTION AT CATALYTIC DOMAIN OF POLY (ADP-RIBOSE) POLYMERASE 2 IDENTIFIES RESIDUES ESSENTIAL FOR ENZYME INHIBITION		
5.1	BACKGROUND	97
5.2	COMPUTATIONAL ALANINE SCANNING	98
5.3	LIGAND PREPARATION AND MOLECULAR DOCKING	102
5.4	PROTEIN STRUCTURAL STABILITY	107
5.5	IMPORTANCE OF HYDROGEN BONDS	110
5.6	IMPORTANCE OF π -STACKING INTERACTIONS	113
5.7	PRINCIPAL COMPONENT ANALYSIS	115
5.8	MM/PBSA ENERGY	123
5.9	CONCLUSION	127

**CHAPTER 6- MOLECULAR MECHANISM OF
INHIBITION OF COVID-19 MAIN PROTEASE BY β -
ADRENOCEPTOR AGONISTS AND ADENOSINE
DEAMINASE INHIBITORS USING *IN SILICO* METHODS**

6.1	BACKGROUND	129
6.2	LIGAND SELECTION	130
6.3	PHARMACOPHORE MODEL GENERATION	131
6.4	MOLECULAR DOCKING-BASED SCREENING	133
6.5	ADME/TOXICITY STUDIES	136
6.6	MOLECULAR DYNAMICS SIMULATION	137
6.7	PCA AND FEL ANALYSIS	147
6.8	MM/PBSA FREE ENERGY	149
6.9	CONCLUSION	151

**CHAPTER 7- UNDERSTANDING THE ROLE OF WATER
ON TEMPERATURE-DEPENDENT STRUCTURAL
MODIFICATIONS OF SARS COV-2 MAIN PROTEASE
BINDING SITES**

7.1	BACKGROUND	153
7.2	MOLECULAR DYNAMICS SIMULATION PROTOCOL	155
7.3	TEMPERATURE EFFECT ON M ^{Pro} ENZYME	156
7.4	TEMPERATURE EFFECT ON VARIOUS BINDING SITES	161
7.5	CONFORMATIONAL DYNAMICS OF THE PROTEIN	164

7.6	PRINCIPAL COMPONENT ANALYSIS	166
7.7	CHANGES IN DIHEDRAL ANGLES AT VARIOUS BINDING SITES	167
7.8	RADIAL DISTRIBUTION FUNCTION	168
7.9	NUMBER OF HYDROGEN-BONDED WATER MOLECULES	171
7.10	ORIENTATIONAL TETRAHEDRAL ORDER PARAMETER	172
7.11	FREE ENERGY vs. TEMPERATURE	174
7.12	HYDROGEN BOND DYNAMICS	176
7.13	STRUCTURE AND DYNAMICS OF M ^{Pro} IN mTIP3P WATER MODEL	177
7.14	CONCLUSION	180
	CHAPTER 8- SUMMARY AND CONCLUSIONS	
8.1	SUMMARY	181
8.2	CONCLUSIONS	182
	APPENDIX	185
	REFERENCES	253
	LIST OF PUBLICATIONS	287
	LIST OF CONFERENCES	289
	WORKSHOPS ATTENDED	290
	BIODATA	291

List of Figures

Figure 1.1	CADD in drug discovery/ design	4
Figure 1.2	Secondary-structure representation of human GCGR (PDB ID: 5XEZ) with a hypothetical ligand showing the trans-membrane domain (TMD), the stalk region and the extracellular domain (ECD). An enlarged view of the extrahelical ligand binding site of GCGR with the major amino acid residues at the catalytic pocket was shown next to it.	12
Figure 1.3	The cartoon diagram of superposition of the PARP enzymes. PARP-2 is shown in pink colour and PARP-1 is shown in blue colour.	13
Figure 1.4	Representation of SARS CoV-2 main protease (PDB ID: 6Y2E) showing Domain I (blue), Domain II (red) and Domain III (green).	15
Figure 2.1	Schematic representation of docking process	25
Figure 2.2	Some examples featuring the pharmacophore sites in PHASE module	28
Figure 2.3	Schematic overview of QSAR process	29
Figure 3.1	A) Skeletal structure of glucagon receptor antagonist: Region A represented in pink colour is the polar region containing R1 ring, Region B represented in red colour is the alkyl side chain which is hydrophobic in nature (H1) and Region C represented in blue colour is the hydrophobic core containing R2, R3 ring and H2 side group. B) Detailed structure of six classes of GCGR antagonists.	43
Figure 3.2	Scatter plot between the experimental activity (pK_i) vs.3D-QSAR predicted activity. A) The training dataset (unfilled red circles). B) The test dataset (unfilled blue circle). The best fitted equation for the scatter plot of test set compounds is given as $y = 0.74x + 1.64$ ($R^2 = 0.84$).	45

Figure 3.3	(A) Scatter plot showing the predicted activity vs experimental activity of external test set with best fitted line $y = 0.61x + 2.56$ ($R^2 = 0.83$) and (B) Plot of residual activity vs predicted activity of external test set	45
Figure 3.4	3D-QSAR Contour Plots visualized in the context of favourable and unfavourable positions A) hydrophobic groups, B) hydrogen bond donor groups and C) electron withdrawing groups.	48
Figure 3.5	Binding pose of lowest-energy conformation of inhibitors bound to glucagon receptor and its 2D ligand interaction diagrams are shown. A (i- ii), B (i- ii), C (i- ii), D (i- ii), E (i- ii), F (i- ii) and G (i- ii) corresponds to Compound 1, Compound 12, Compound 20, Compound 32, Compound 40, Compound 50 and Compound 58 respectively..	53
Figure 3.6	Overlay of re-docked ligand (Green) with its crystal structure conformation (Cyan) (RMSD: 0.82Å).	53
Figure 3.7	Position of A) HOMO B) LUMO regions of interacting amino acid residues at the allosteric pocket of GCGR.	57
Figure 3.8	A)Time-line representation of RMSD profile of $C\alpha$, backbone and heavy atoms of 5XEZ with respect to its initial coordinate.The RMSD of compound 20 with respect to protein backbone and its own starting structure was illustrated in red and pink colour respectively. B) RMSF profile of 5XEZ protein indicating degree of fluctuation during simulation. The green line projecting on x-axis represents ligand contact during the simulation.	60
Figure 3.9	A) Timeline representation of protein- ligand interactions, B) Histogram showing possible interaction of Compound 20 with 5XEZ residues throughout the simulation and C) 2D diagram of protein- ligand contacts.	61
Figure 3.10	Number of hydrogen bonds formed between A) protein and ligand B) ligand and water throughout 100 ns.	62

- Figure 3.11** Snapshots at A) 0 ns, B) 10 ns, C) 20 ns, D)30 ns, E) 40 ns, F) 50 ns, G) 60 ns, H) 70 ns, I) 80 ns, J) 90 ns and K) 100 ns of MD trajectory are illustrated. The interacting amino-acid residues are coloured in green, oxygen of water molecules are represented in red. The π -cation interaction and hydrogen bonding interactions are indicated by red and yellow respectively. 63
- Figure 3.12** Ligand properties: RMSD (\AA) with respect to reference conformation, radius of gyration (R_g , \AA), number of internal hydrogen bonds (intra HB), molecular surface area (\AA^2), solvent accessible surface area (\AA^2) and polar surface area (\AA^2) 64
- Figure 4.1** Binding pocket of PARP-2 enzyme bound with active, moderate and inactive ligands (a) C-1, (b) C-16 and (c) C-32. Enlarged view of binding pose is shown in 2D-ligand interaction diagram. The hydrophobic, polar, positively-charged, negatively-charged and glycine residues are represented in green, cyan, blue, red and grey colour respectively. 70
- Figure 4.2** Binding pose and ligand interaction diagram of PARP-1 enzyme with ligands (a) C-1 and (b) C-32. 71
- Figure 4.3** Overlay of re-docked ligand (Green) with its crystal structure conformation (Pink) (RMSD: 0.61\AA) 72
- Figure 4.4** Frontier molecular orbital energy values, energy gap and electron cloud density over compound1, compound 8, compound 16, compound 21, compound 27 and compound 32. 74
- Figure 4.5** The distribution of HOMO-LUMO orbitals over the cluster containing interacting amino acids and ligands (C-1 and C-32). The positive and negative lobes of HOMO orbitals are blue and red in colour respectively. The positive and negative lobes of LUMO orbitals are coloured green and pink respectively 75
- Figure 4.6** (a) RMSD profile of backbone of unbound and bound PARP enzymes over 500 ns trajectory. (b) Time evolution of RMSD of ligand with respect to protein throughout 500 ns trajectory for 78

	PARP complexes.	
Figure 4.7	Residue-based fluctuations of protein backbone of unbound enzyme and complexes over 500 ns simulation (a) PARP-2 and (b) PARP-1. (c) Radius of gyration of backbone atoms of PARP enzymes with and without ligands (C-1 and C-32).	79
Figure 4.8	Number of hydrogen bonds calculated for the evaluation of protein-ligand interaction throughout 500 ns simulation (a) PARP-2/C-1, (b) PARP-2/C-32, (c) PARP-1/C-1 and (d) PARP-1/C-32 respectively.	80
Figure 4.9	Distribution of donor-acceptor distance (nm) in hydrogen bond interaction for all the complexes.	86
Figure 4.10	(a) Plot of eigenvalue vs. first 30 eigenvector index derived from PCA over 500 ns MD trajectory for unbound and bound PARP-2 system. (b) 2-D projection of first two principal motions (PC1 and PC2) of protein in phase space for unbound and bound PARP-2 enzyme.	88
Figure 4.11	Gibbs free energy landscape for unbound and bound PARP-2 enzymes. FEL obtained from MD trajectory for all three systems using the reaction coordinates as the projection of C- α atoms onto the first two principal components. (a) unbound PARP-2 enzyme (b) PARP-2/C-1 complex and (c) PARP-2/C-32 complex.	91
Figure 4.12	Structure of newly designed PARP-2 inhibitors.	92
Figure 4.13	Time-line representation of RMSD profile of backbone atoms with respect to the initial coordinate of PARP-2 enzyme.	93
Figure 5.1	Free energy change ($\Delta\Delta G = \Delta G_{\text{wild}} - \Delta G_{\text{variant}}$) obtained from alanine substitution for PARP-2 enzyme. Negative $\Delta\Delta G$ indicate unfavourable substitution of alanine at the respective position.	99
Figure 5.2	Schematic representation of wild and four variants generated by single amino acid substitution with FMTTPBC: (a) Wild, (b) Gly429Ala, (c) Tyr462Ala, (d) Tyr473Ala and (e) Glu558Ala.	101

Figure 5.3	(a) 2D-Structure of (R)-6-fluoro-2-(5-methyl-4,5,6,7-tetrahydrothieno [2,3-c]pyridin-2-yl) 1H-benzo[d]imidazole-4-carboxamide (FMTPBC) (b) Optimized 3D-Structure from DFT.	102
Figure 5.4	2D-Ligand interaction diagram of FMTPBC with four PARP-2 enzyme variants. The non-covalent interactions: hydrogen bond interactions (pink) and π - π interactions (red) with the amino acid residues are shown. (a) Gly429Ala, (b) Tyr462Ala, (c) Tyr473Ala and (d) Glu558Ala.	104
Figure 5.5	RMSD profile of protein backbone C α atoms for last 200 ns simulation trajectory (a) unbound wild and variants (b) bound wild and variants (c) ligand with respect to protein.	108
Figure 5.6	Backbone RMSF fluctuations of (a) unbound enzymes and (b) ligand bound enzymes.	109
Figure 5.7	Occurrence of hydrogen bond interaction (a) protein-ligand (b) cavity-water and (c) ligand-water.	112
Figure 5.8	Plot of eigenvalue vs. first 20 eigenvector index derived from PCA over 250 ns trajectory for (a) unbound and (b) bound PARP-2 enzyme variants.	115
Figure 5.9	First two principal components showing the major motions of protein spread over the phase space. (a) Gly429Ala, (b) Tyr462Ala, (c) Tyr473Ala, (d) Glu558Ala and (e) Wild.	117
Figure 5.10	Porcupine plots showing prominent motions of enzyme in bound Gly429Ala variant. The protein is represented as backbone with arrow attached to C- α atoms showing the direction and magnitude of prominent motions along mode 1, mode 2 and mode 3. The overall RMSF graph of enzyme in complexes superimposed with the amount of fluctuations of amino acid residues in individual modes (below).	120
Figure 5.11	Porcupine plots showing prominent motions of enzyme in Tyr462Ala variant complex. The protein is represented as backbone with arrow attached to C- α atoms showing the	121

direction and magnitude of prominent motions along mode 1, mode 2 and mode 3. The overall RMSF graph of enzyme in complexes superimposed with the amount of fluctuations of amino acid residues in individual modes (below).

Figure 5.11 Porcupine plots showing prominent motions of enzyme in 122

Tyr473Ala variant complex. The protein is represented as backbone with arrow attached to C- α atoms showing the direction and magnitude of prominent motions along mode 1, mode 2 and mode 3. The overall RMSF graph of enzyme in complexes superimposed with the amount of fluctuations of amino acid residues in individual modes (below).

Figure 5.12 Porcupine plots showing prominent motions of enzyme in 123

Glu558Ala variant complex. The protein is represented as backbone with arrow attached to C- α atoms showing the direction and magnitude of prominent motions along mode 1, mode 2 and mode 3. The overall RMSF graph of enzyme in complexes superimposed with the amount of fluctuations of amino acid residues in individual modes (below).

Figure 6.1 (a) Structure of broad spectrum non-covalent inhibitor, derivative 132

of imidazole carboxamide (b) Pharmacophore model (AARRR) generated using Phase module inside the binding pocket of covid-19 main protease (c) Overlay of inhibitor over the generated model. The spheres indicate the excluded volume forbidden for ligands due to enzyme backbone or lipophilic interactions.

Figure 6.2 Overlay of top-scored molecules over developed pharmacophore 133
model.

Figure 6.3 RMSD profile of backbone of covid-19 main protease (6LU7) 139
complexed with Michael acceptor inhibitor N3 and top-scored hits over 100 ns trajectory.

Figure 6.4 (a) Residue-based fluctuations of 6LU7 backbone of complexes 140

	(b) Rg (nm) vs. Time (ps) for main protease (6LU7)/ inhibitor complexes.	
Figure 6.5	a) RMSD profile of backbone of main protease (6Y2G) complexed with alpha-ketoamide 13b, b) Residue-based fluctuations of protein backbone of complex and c) Rg (nm) vs. Time (ps).	141
Figure 6.6	Occupancy of secondary structure elements of covid-19 main protease (6LU7) in the inhibitor complex. A) MPro/+fenoterol, B) MPro/-fenoterol, C) MPro/FR236913 and D) MPro/FR230513. (Colour codes- Black: β -sheet, red: β -bridge, green: bend, blue: α -helix, yellow: coil, brown: 3-helix).	142
Figure 6.7	Hydrogen bond interactions between protease (6LU7) -inhibitor (red), inhibitor-water (blue) and binding site residues of protease-water (green). a) MPro/N3, b) MPro/+fenoterol, c) MPro/-fenoterol, d) MPro/FR236913 and e) MPro/FR230513.	146
Figure 6.8	Plot of eigenvalue vs. first 30 eigenvector index derived from PCA over a stable trajectory of 40 ns for protein-ligand systems.	147
Figure 6.9	FEL from a stable 40 ns trajectory using the reaction coordinates as the projection of backbone atoms of protease (6LU7) onto the first two principal components. a) MPro/N3, b) MPro/+fenoterol, c) MPro/-fenoterol, d) MPro/FR236913 and e) MPro/FR230513.	149
Figure 7.1	Global structures of SARS CoV-2 M ^{Pro} monomer (a, b) and dimer (c, d) at 278 K (black), 310 K (red), 348 K (green) and 383 K (blue) in SPC/E (a, c) and m-TIP3P (b, d) water model.	157
Figure 7.2	Global ΔG from folding/unfolding of protease vs. Temperature, calculated by eqn. 2.44.	158
Figure 7.3	Probability distribution of (a) C α -RMSD (b) radius of gyration of protein (c) SASA (d) Number of M ^{Pro} -M ^{Pro} / M ^{Pro} -water hydrogen bonds of protein (e) number of water molecules in the substrate binding site during 200 ns simulation and (f) Calculated B-factor	161

of C_{α} atoms with amino acid residues for M^{Pro} . The solid and dotted lines represent monomer and dimer-Chain A respectively in SPC/E water model.

- Figure 7.4** Probability distribution of (a) C_{α} -RMSD of binding site residues with respect to D-I, D-II, D-III of protein (b) radius of gyration of binding site residues (c) SASA of binding site residues. The binding site residues are considered as per Table 7.1. The solid and dotted lines represent monomer and dimer-Chain A respectively in SPC/E water model. 164
- Figure 7.5** Network distributions of most prominent conformations of main protease from 200 ns trajectory at different temperatures in SPC/E water model. PMF profile for SARS CoV-2 main protease as a function of RMSD (Right) (a) monomer (b) dimer-Chain A 166
- Figure 7.6** Plot of eigenvalue vs first 25 eigenvector index derived from PCA over 200 ns MD trajectory for (a) M^{Pro} monomer and (b) M^{Pro} dimer-Chain A. 2-D projection of first two principal motions (PC1 and PC2) of main protease (Right). 167
- Figure 7.7** RDF of (a) Oxygen–Oxygen and (b) Hydrogen-Oxygen of water molecules at different temperatures. 169
- Figure 7.8** Radial distribution function of C_{α} -Ow of key amino acid residues at various binding site of protease at different temperatures in SPC/E water model. Solid line and dash line represent monomer and dimer-Chain A respectively. (a) His41 (b) Cys117 (c) Glu166 (d) Arg298. 170
- Figure 7.9** Fraction of water molecules having n number of hydrogen bonds within a distance of 6.0 Å from C_{α} of amino acid residues of (a) M^{Pro} monomer and (b) dimer-Chain A for SPC/E water model at different temperatures. 172
- Figure 7.10** Orientational tetrahedral order parameter of water molecules within 6.0 Å from C_{α} of key amino acid residues at various 173

binding site of monomer (solid line) and dimer-Chain A (dash line) in SPC/E water model at different temperatures.

- Figure 7.11** Free energy function of C_{α} -Ow of amino acid residues at various binding site of protease at different temperatures for monomer protease in SPC/E water model. (a) Catalytic dyad (b) Substrate binding site (c) Dimerization site (d) Allosteric site. 175
- Figure 7.12** Probability distribution of C_{α} -RMSD; time evolution of radius of gyration and SASA in mTIP3P water model. (a) monomer (b) dimer-Chain A. 178
- Figure 7.13** Radial distribution function of C_{α} -Ow of key amino acid residues at various binding site of protease at different temperatures in mTIP3P water model. Solid line and dash line represent monomer and dimer-Chain A respectively. (a) His41 (b) Cys117 (c) Glu166 (d) Arg298. 179

List of Tables

Table 2.1	Molecular descriptors used in QSAR	30
Table 3.1	PLS regression summary of generated 3D-QSAR model	44
Table 3.2	Summary of electronic properties of selected ligands	54
Table 4.1	HOMO-LUMO energy details and total energy of selected ligands	73
Table 4.2	Global reactivity descriptors calculated from DFT calculations	76
Table 4.3	Summary of MD simulations performed	77
Table 4.4	Hydrogen bond occupancy between ligand and amino acid residues at binding pocket of enzyme	81
Table 4.5	Contribution of energy components to MM/PBSA binding free energy for PARP enzymes with C-1 and C-32 (energy in kcal/mol)	83
Table 4.6	Per residue binding energy contribution in the PARP-2/C-1 and PARP-2/C-32 complexes	84
Table 4.7	Per residue binding energy contribution in the PARP-1/C-1 and PARP-1/C-32 complexes	85
Table 4.8	Cosine content on the principal components of C- α atoms	89
Table 4.9	Predicted ADME/Toxicity properties for newly designed PARP-2 inhibitors	94
Table 5.1	Various non-covalent interactions observed between PARP-2 variants and ligand, FMTPBC	105
Table 5.2	Hydrogen bond occupancy (%) and water bridges between the enzyme variants and the ligand at the catalytic domain of PARP2	112
Table 5.3	Contribution of energy components to MM/PBSA binding free energy for PARP-2 enzyme variants with FMTPBC (energy in kcal/mol)	123
Table 5.4	Per residue binding energy contribution in the PARP-2 variant complexes	126

Table 6.1	Docking score (kcal/mol) of top-scored molecules with existing crystal structures of main protease (6LU7, 6W63, 6Y2G) and apo form (7KFI)	135
Table 6.2	Predicted ADME/Toxicity properties for top-scored molecules from β -adrenoceptor agonist and adenosine deaminase inhibitors	137
Table 6.3	Characteristics of protein (6LU7) -ligand complex obtained from 3 independent molecular dynamics simulations	138
Table 6.4	Hydrogen bond occupancy between ligand and residues at the binding site of main protease	143
Table 6.5	Contribution of energy components to MM/PBSA binding free energy for covid-19 main protease with potential hits (energy in kJ/mol)	150
Table 7.1	Important amino acid residues taken for analysis and their biological function	162
Table 7.2	T and ΔT selected for entropy and enthalpy contributions	174
Table 7.3	Hydrogen bond lifetime (ps) of water molecules around the key amino acid residues at the various sites of main protease in SPC/E water model	177

List of Appendix

Appendix I Detailed structure, experimental activity (Exp pK _i), predicted activity (Pred pK _i) and residual activity of glucagon receptor antagonists	185
Appendix II Predicted biological activity of external test set of 15 compounds	188
Appendix III QSAR contour plots visualized in terms of (a) (i)-(vii) favourable and unfavourable hydrophobic interactions on Compound 5, 12, 35, 36, 37, 40, 58 (b) (i)-(v) favourable and unfavourable hydrogen bond donor groups on Compound 13, 17, 20, 25, 50 and (c) (i)-(v) favourable and unfavourable electron withdrawing groups on Compound 41, 44, 46, 47, 50	190
Appendix IV Energy terms contributing to docking energy of protein-ligand complex (kcal/mol)	192
Appendix V The HOMO-LUMO orbital positions mapped on selected ligands at B3LYP/ 6-31G(d,p) level of calculation	194
Appendix VI Contribution of energy components to MM/ GBSA binding free energy for ligands at the binding site of 5XEZ (kcal/mol)	196
Appendix VII Characteristics of protein-ligand complex obtained from 5 independent molecular dynamics simulations	198
Appendix VIII Predicted ADME/Toxicity properties of top scored glucagon receptor antagonists	199
Appendix IX Detailed structure, experimental activity and docking energy of benzimidazole carboxamide inhibitors against PARP-1/2 enzyme	200
Appendix X The angle between the normal vectors of two rings in stacking interaction formed between ligand and aromatic amino acid residues of PARP-2 enzyme at the catalytic pocket over 500ns simulation. The geometrical shapes (circle, square and triangle) represent the interaction of amino acid residue with different rings present in the ligand. His428 and Phe463 were found to have only one interaction with the benzene ring (triangle) of benzimidazole core whereas Tyr455 found to interact with the imidazole (square) ring. Tyr462 showed interaction with benzene ring and	202

thiophene ring (circle) whereas Tyr473 showed π -stacking interaction with all the three aromatic rings present in the ligand. Average distance is given in Å

Appendix XI The angle between the normal vectors of two rings in stacking interaction formed between ligand and aromatic amino acid residues of PARP-1 enzyme at the catalytic pocket over 500ns simulation. The geometrical shapes (circle, square and triangle) represent the interaction of amino acid residue with different rings present in the ligand. His862, Tyr896 and Phe897 were found to have only one interaction with the benzene ring (triangle) of benzimidazole core whereas Tyr889 found to interact with the thiophene ring (circle). Tyr907 showed π -stacking interaction with benzene ring and imidazole ring (square) ring present in the ligand. Average distance is given in Å

Appendix XII Porcupine plots showing prominent motions of enzyme in complexes. The protein is represented as backbone with arrow attached to C- α atoms showing the direction and magnitude of prominent motions along mode 1, mode 2 and mode 3. The overall RMSF graph of enzyme in complexes superimposed with the amount of fluctuations of amino acid residues in individual modes (below) **(a)** PARP-2/C-1 complex **(b)** PARP-2/C-32 complex and **(c)** PARP-2 enzyme

Appendix XIII Hydrogen bond occupancy between the predicted drug molecules and the amino acid residues at binding pocket of enzyme

Appendix XIV Contribution of energy components to MM/PBSA binding free energy for PARP enzymes with predicted drug molecules (kcal/mol)

Appendix XV Per residue binding energy contribution in the predicted complexes

Appendix XVI Predicted free energy change ($\Delta\Delta G = \Delta G_{\text{wild}} - \Delta G_{\text{variant}}$) upon alanine substitution

Appendix XVII Free energy landscapes of binding site residues as a function of dihedral angles (ψ, ϕ in degrees). **(a)** Gly429Ala, **(b)** Tyr462Ala, **(c)** Tyr473Ala and **(d)** Glu558Ala

Appendix XVIII Probability distribution of the normal vector angles and the distance between the two rings forming π -stacking interactions. 6M: 6-membered benzene ring in the ligand, 5MN: 5-membered imidazole ring in the ligand and 5MS: 5-membered thiophene ring in the ligand	222
Appendix XIX Cosine content on the principal components of backbone C- α atoms	225
Appendix XX Crystal structure of SARS CoV-2 main protease (M ^{Pro}) complex with a) alpha-ketoamide 13b (6Y2G) and b) broad spectrum non-covalent inhibitor X77 (6W63). c) Crystal structure of SARS CoV-2 main protease apo form (7KFI)	226
Appendix XXI Molecular structure, common name, docking score and 2D ligand interaction diagrams of top-scored hit molecules with main protease (6LU7)	227
Appendix XXII Docking score and interaction of existing co-crystal ligands	231
Appendix XXIII Per-residue secondary structure analysis of main protease (6LU7) in the complex. a) MPro/+fenoterol, b) MPro/-fenoterol, c) MPro/FR236913 and d) MPro/FR230513	232
Appendix XXIV a) RMSD profile of the dimer, Chain A and Chain B at different temperatures in SPC/E water. b) Radius of gyration (rGyr) of dimer, Chain A and Chain B at different temperatures as a function of time in SPC/E water. c) Dimer Interaction energy (kJ/mol) vs. Temperature (K)	236
Appendix XXV Average occupancy of secondary structure components obtained for 2 sets of simulations each for monomer and dimer- chains using SPC/E water model	238
Appendix XXVI Secondary structure elements for the protease calculated by DSSP for the monomer and dimer (chain A) in SPC/E water model	239
Appendix XXVII (a) Correlation of inter-residue distance at different temperature for monomer. The blue colour shows the residue contact ruptures and the orange colour shows the residue contact formation (b) Change in the number of non-native contacts in monomer (circle) and dimer-chain A (diamond) as a function of time for last 100 ns MD trajectory	243

Appendix XXVIII Probability distribution of (a) C_{α} -RMSD (b) rGyr and (c) SASA for domain I, II, III for monomer protease	244
Appendix XXIX Dihedral angle (ϕ , ψ) transitions of amino acid residues at various sites of monomer during the course of simulations (a) Catalytic dyad (b) Substrate binding site (c) Dimerization site (d) Allosteric site. The colour scale represents the simulation time in ns	245
Appendix XXX Dihedral angle (ϕ , ψ) transitions of amino acid residues at various sites of dimer-chain A during the course of simulations in SPC/E water model (a) Catalytic dyad (b) Substrate binding site (c) Dimerization site (d) Allosteric site. The colour scale represents the simulation time in ns	246
Appendix XXXI Radial distribution function of C_{α} -Ow of amino acid residues at various binding site of protease at different temperatures in SPC/E water model. (a) Catalytic dyad (b) Substrate binding site (c) Dimerization site (d) Allosteric site	247
Appendix XXXII Radial distribution function of C_{α} -Ow of amino acid residues at various binding site of protease at 5 additional temperatures in SPC/E water model. (a) Catalytic dyad (b) Substrate binding site (c) Dimerization site (d) Allosteric site	248
Appendix XXXIII Enthalpy (solid lines) and entropy (dashed lines) contribution to free energy at different temperatures for monomer protease in SPC/E water model. (a) Catalytic dyad (b) Substrate binding site (c) Dimerization site (d) Allosteric site	249
Appendix XXXIV Fraction of water molecules having n number of hydrogen bonds within a distance of 6.0 Å from C_{α} of amino acid residues of (a) M^{Pro} monomer and (b) dimer-Chain A for mTIP3P water model at different temperatures	250
Appendix XXXV Orientational tetrahedral order parameter of water molecules within 6.0 Å from C_{α} of key amino acid residues at various binding site of (a) monomer and (b) dimer-chain A in mTIP3P water model at different temperatures	252

CHAPTER 1

INTRODUCTION

This chapter mainly focus on the importance of proteins and its application in the drug discovery and development process. A brief discussion regarding the drug designing protocol is included. In addition to this, a concise literature review, a brief discussion about the protein-ligand systems considered in the thesis and objectives of the present research work have been discussed.

1.1 SIGNIFICANCE OF PROTEINS

Proteins are versatile organic bio-macromolecules which have immense significance in all form of life including viruses. They are essential for normal cell functions and considered as building blocks of biological matter. Proteins have high nutritional value and are directly involved in the biochemical processes in living things. They are polymer of amino acid residues linked by peptide (amide) bonds that can fold into three-dimensional structures. Only 20 amino acids are known that occur naturally in the peptides and proteins. Some of the functions of protein include enzyme catalysis of chemical reactions, immunological responses, mechanical structural support, cell growth control, transportation and storage of molecules. The specificity of protein function depends on the three-dimensional structure of protein formed by the polymeric peptide chain of amino acids (Alberts et al. 2002). It is found that proteins of similar biological functions posses similar amino acid composition and sequence. The primary structure of proteins reveals the sequence and do not provide any information about it's arrangement in space. The secondary, tertiary and quaternary structure reveals the configuration of peptide chain in the proteins. The secondary structure considers only the spatial arrangement of protein backbone while the tertiary structure considers the conformation of side chains and other adjacent segments of protein backbone. The quaternary structure is formed by the arrangement of identical subunits (domains) of protein in which each domain is a separate peptide chain. Based

on the tertiary structure of proteins, they are classified into globular proteins, fibrous proteins and membrane proteins. They serve a wide range of functions including defense, transportation and organization.

Protein dysfunctionality can lead to various disease conditions in organisms which might be potentially deadly. In one way, the denaturation or misfolding of proteins results in the formation of aggregates which affects the native structure of proteins. Such changes in molecular level result in system malfunction and disease in organism as seen in Alzheimer's disease, Parkinson's disease, diabetes etc. But in case of pathogenic enzymes, the biological activity of protein can be modified by small organic molecules (agonists/antagonists) that interact either with the primary site of the protein or the other possible binding sites.

Proteins control the biological process such as cell signalling, immunological responses, signal transduction and enzyme catalysis through molecular recognition. They perform these tasks either by interacting with other proteins or small organic molecules that binds to them. The interactions of proteins with water, small organic molecules, other proteins and nucleic acids are governed by non-covalent interactions such as hydrogen bonding, π - π stacking interaction, π -cation interaction, salt bridge interaction and disulfide bridging. The stability of interactions depends on the combined effect of specific forces acting at the binding site and non-specific forces outside the binding pocket of protein. Specific force arises from chemical/ short range interactions between a unique arrangement of proteins or protein-ligand systems such as hydrogen bond interactions. In contrast, non-specific force arises from long range interactions that may occur at the interface of many atoms, molecules or even surfaces such as van der Waals, hydrophobic and steric interactions. This combined force stabilizes interactions in biomolecular collisions in solutions and adhesion between cells. The stabilization and complexity of interactions between protein and incoming moiety (water, small organic molecules, other proteins, nucleic acids) depends on the flexibility of protein binding pocket and resulting structural rearrangements upon the binding of moiety. Biophysical studies help to determine the molecular mechanism and energy requirements of biological interactions. The nature of interaction (physical or chemical) can be determined by the calculated energies of the system. This information can be used to manipulate the biological activity by modifying the

protein, the incoming moiety or both. Understanding the three-dimensional structure of proteins, dynamics of proteins and the interaction of proteins with natural ligands/potential drug molecules can aid the designing and development of potential drug candidates for disease treatment in any drug discovery process.

Traditional drug discovery and development process is time-resource consuming and also is very expensive. Biophysical studies help to limit and focus important chemical synthesis and biological testing, decreasing the need of traditional resources. The two approaches in computer-aided drug design (CADD) have been divided into ligand-based drug design (LBDD) and structure-based drug design (SBDD). The basic concepts and workflow of different computational technique in drug discovery is discussed in following section.

1.2 DRUG DISCOVERY AND DEVELOPMENT PROCESS

A wide variety of drugs are available in market for the treatment of diseases. Drugs can cure or ameliorate the symptoms of illness. Even drugs are used as preventive medicine that has future applications. Drug discovery is a fascinating area which finds treatment for diseases haven't cured yet. Computer-aided drug designing (CADD) offered the identification of target and compounds using existing knowledge minimizing the risk of rejection of drug candidates. It is regarded as the interplay between artificial and human intelligence to predict the drug candidates more efficiently which otherwise is time and resource consuming. The selection of method depends on the information about the identified target protein. The incorporation of computational techniques to combine chemical and biological aspects of target-drug interactions helps in drug discovery, design, development and optimization process. CADD facilitates hit identification, hit-to-lead selection, optimize absorption, distribution, metabolism, excretion and toxicity profile and avoid safety issues. The computer aided drug design mainly uses three basic steps: ligand-based drug design (a 3D spatial arrangement of chemical features of ligand necessary for biological activity), structure-based drug design (drug-target docking) and quantitative structure activity relationship and structure property relationship (QSAR and QSPR). Computational modelling in drug discovery is becoming popular and offers an improved efficiency for industries (Kumar et al. 2006; Stahl et al. 2006). It limit and

focus important chemical synthesis and biological testing, decreasing the need of traditional resources.

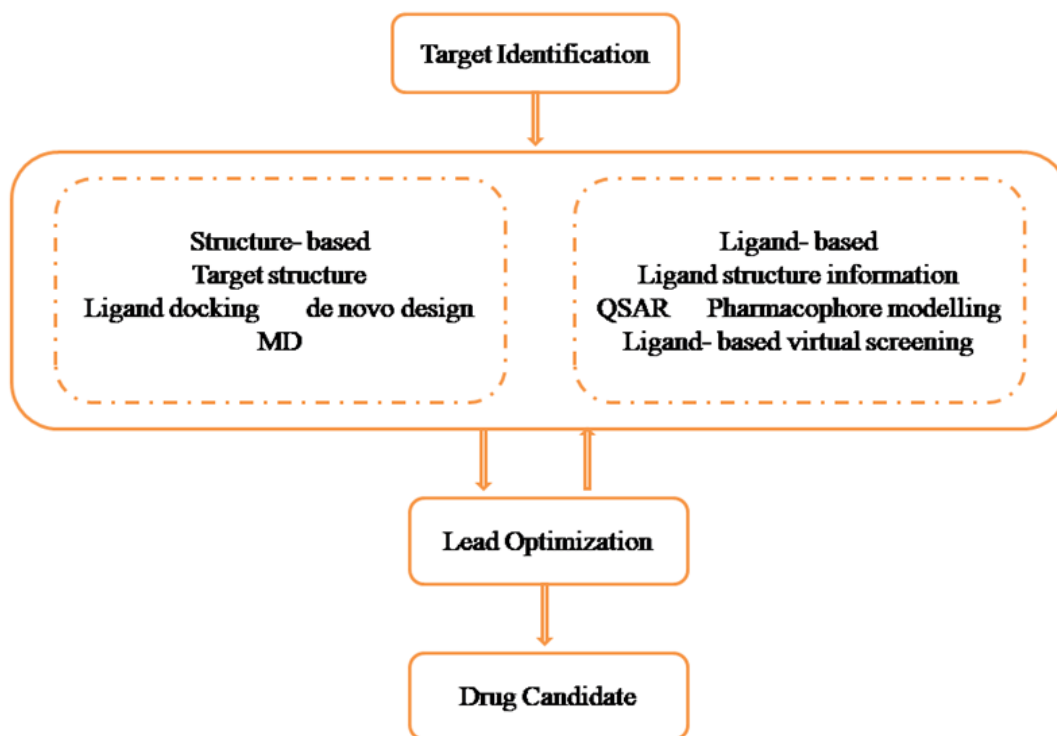


Figure 1.1: CADD in drug discovery/ design

The strategies for CADD depend on the structural information available for target and ligand (**Figure 1.1**). The main goal in drug design is to predict how strong a ligand will bind to the target. Molecular dynamics is often used to estimate the strength of intermolecular interaction between ligand and target. Other methods such as semi-empirical, *ab initio* quantum chemistry methods and density functional theory are used to obtain the optimized structures for molecular dynamics calculations and also to obtain electronic properties (HOMO-LUMO energies, electrostatic potential, polarizability, dipole moment etc) of ligand which affects the binding affinity (Lewis 2011). This can also predict the conformational changes of ligands during binding.

1.2.1 Brief History of CADD (Kore et al. 2012)

1900 : The receptor and lock-and-key concept. P Ehrlich (1909) and E Fisher (1894)

1960s : Viz- review the drug- target interaction

1970s : Quantitative structure activity relationship (QSAR)

1980s : Beginning of CADD molecular biology, X- ray crystallography, multi-dimensional NMR, Molecular modelling- docking, molecular graphics

1990s : Human genome bioinformatics, Combinatorial chemistry, High- throughput screening

2000s : Vast information handling- pharmacogenesis

1.2.2 Workflow for CADD (Kore et al. 2012)

- Target identification and validation: Identifying new targets in a particular metabolic pathway and confirming its role in diseases. The drug candidates can either enhance or inhibit the target thereby modifying the pathway.
- Hit identification: Development of ligands with affinities for a chosen target through virtual screening (structure- or ligand- based design).
- Lead optimization: The hit molecules are chemically altered to improve its affinity and selectivity to a specific target (structure- based design, QSAR).
- Biological assays: Molecular modelling and molecular graphics.
- Synthetic chemistry: Peptidomimetics, combinatorial chemistry
- Clinical trials

1.2.3 Structure-based Drug Discovery

This approach relies on the 3D structure of biological target obtained through X-ray crystallography and NMR spectroscopy (Jhoti and Leach 2007). Using the structure of the biological target, suitable drug candidate with high binding affinity and selectivity can be designed. The three main categories of structure- based drug discovery methods are virtual screening, de novo design of new ligands and optimization of known ligand by evaluating proposed analogues in binding cavity (Klebe 2000). The approach depends on the ability of molecule to interact with a specific biological target and to exert desired biological activity which depends on the favourable interaction at binding site of target (Sliwoski et al. 2013).

1.2.4 Ligand-based Drug Discovery

This approach uses a set of molecules known to interact with target molecule of interest and analyse the 2D or 3D structures. The physiochemical properties important for the interactions are retained and other non-relevant information are discarded (Sliwoski et al. 2013). It is an indirect way of drug discovery which does not focus on the target information. The most common approaches in ligand-based drug discovery include pharmacophore modelling and QSAR methods. The set of molecules are used to derive a pharmacophore that meets the minimum characteristic features responsible for biological activity (Güner 2000). Also, a correlation between calculated properties of the molecules and experimentally determined biological activity can be derived which can be used to predict the activity of new drug candidates (Roy et al. 2015a; b).

In summary, it can be concluded that CADD is suitable when minimum structural information about the target is available. Structure-based CADD is preferred for soluble proteins that can be crystallized. Ligand-based CADD is preferred for the drug candidates with high affinity towards target with minimal off-target effects and can be designed with less free energy, favourable drug metabolism and pharmacokinetic properties. Literatures supporting ligand and structure based approaches as well as the combined approaches are discussed in next section.

1.3 SIGNIFICANCE OF NON-COVALENT INTERACTIONS

Non-covalent interactions have a key role in the molecular recognition process. Understanding the interaction of bioactive/ drug molecules with the target protein helps to describe the structure-activity relationships and thus improve the efficiency of lead molecules (Anighoro 2020). It is known that protein-ligand complexes generally form non-covalent interactions such as hydrophobic contacts, classical hydrogen bonds, π -stacking interactions and salt bridges (Li et al. 2013; Patil et al. 2010; Spassov et al. 2023). Electrostatic effects play an important role in molecular recognition and have importance in virtual screening, drug design and protein-protein/ligand docking. Potency optimization of drug candidates mainly focuses on these interactions between the protein and the molecules. Apart from these

interactions, halogen atoms can form weak hydrogen bonds, variety of interactions with π -systems and steric interactions (Xu et al. 2014). Also, the importance of ‘non-classical’ hydrogen bonds where CH acts as donor group (CH --- O) in drug designing is extensively recognized (Itoh et al. 2019). Besides aromatic rings, non-aromatic π -systems can form stacking interactions. Amide stacking interactions are mostly observed in protein-ligand contacts, for instance in trypsin-like proteins. Amide stacking interactions are promising to improve the binding affinity of ligand containing heteroarenes (DeFrees et al. 2019). Halogen- π interactions, cation- π interactions and halogen bonds (CX --- O) are other promising interactions to stabilize protein folding and protein-ligand complexes.

1.4 LITERATURE SURVEY

Recent trends in drug discovery and designing focus to understand the molecular level mechanism on the stability of protein as well as the protein-protein/ protein-ligand interactions causing a certain disease. Ligand chemical similarity (i.e., the QSAR modelling) and pharmacophore modelling are the common methods used in ligand-based approaches. For example, the promising inhibitors of MARK3 protein kinase expressed in head and neck cancer were proposed based on ligand-based screening approaches such as physicochemical properties, overlapping molecular interaction fields and toxicity predictions (Almeida et al. 2014). Another study proposed the pre-clinical trials of Parvifloron compounds for the treatment of breast cancer (MCF-7 cell line). The designed compounds were subjected to QSAR and pharmacokinetic studies to test the drug-likeness (Abdulrahman et al. 2021). Structure-based approaches rely on the virtual screening techniques and dynamics of molecule at the protein binding site. For instance, various scaffolds for designing new compounds against G protein receptor 142 for type 2 diabetes has been proposed using bind and induced fit docking studies at active site of the protein. MD simulation further validated the stability and interactions (Kaushik et al. 2018). Structure-based approaches are widely used to determine the inhibitors for allosteric binding site of a protein such as G-protein coupled receptors (Congreve et al. 2017), molecular chaperones HSP70 and HSP90 (Ferraro et al. 2019), hepatitis C virus NS5B polymerase (Yan et al. 2007), dopamine receptors (Lane et al. 2013), etc. This method

allows determining the relative affinities of ligand to a particular target, for instance, the interaction of oligosaccharides with Galectin-I in promoting cell growth/inhibition (Ford et al. 2003). The combination of both the approaches has become a common tool in virtual screening process. The Pharmacophore modeling and atom-based 3D-QSAR studies were performed on HIV-1 integrase strand transfer inhibitors (dihydroxy isoindole derivatives) to design promising anti-HIV drug. QM-polarized ligand docking and MD simulations of selected compounds at the binding site provided important insights into the structural and chemical basis involved in molecular recognition (Reddy and Singh 2014). This can also be implemented for the multi-targeted drug design where single chemical/combinations can target the pathological network of a disease. An example, the combination of dasatinib (ABL/T315I inhibitor) and imatinib (tyrosine-kinase inhibitor) are proposed to target BCR-ABL fusion proteins for the treatment of chronic myelogenous leukemia (Kaiser 2011). Recently, *in silico* methods have been widely used in designing effective drug molecules as well as repurposing the existing drugs against SARS CoV-2 proteins (Aktaş et al. 2021; Brendler et al. 2021; Cusinato et al. 2021; Jang et al. 2021; Muratov et al. 2021; Yousefi et al. 2021). The insights from computational studies can drastically reduce the number of ligands for experimental assays. Several studies showed the assistance of CADD in successfully selecting the drug candidates for enzymatic assays (Daddam et al. 2020; Dawoud et al. 2022; Devi et al. 2020; Dolatkah et al. 2017). Therefore, understanding the molecular level mechanism through both the structure-based and ligand-based techniques can help to propose novel insights/hypotheses which serve as a critical step for translational medicine research. Below we are providing the literature survey of some key articles along with their significant contribution in computer-aided drug discovery process.

- ♣ DFT calculations and ADME properties of the synthesized 1,1-di-(3-carboxyphenyl)ethane derivatives revealed the flexible nature which helps the molecules to fit in the active site of as COX-1/COX-2 protein. Molecular docking studies showed that the drug-receptor interactions are favoured by both hydrogen bonds and hydrophobic interactions (Martić et al. 2004).

- ♣ Alves et al. (2007) were successful in calculating the binding energy of HIV- 1 integrase complexed with ligands using B3LYP/MM level of theory. The study revealed that there is a direct relationship between the theoretically computed property and the experimentally determined anti-HIV activity.
- ♣ Iribarne et al. (2009) examined the biological affinities of previously reported phenothiazine derivatives towards the parasite enzyme (Trypanothione reductase, TR) with respect to human counterpart (glutathione reductase, GR) using molecular docking and molecular dynamics simulation studies. The results showed that the derivatives have affinity towards TR active site, which is in agreement with the experimental information.
- ♣ In the study of cyclic imides as protoporphyrinogen oxidase inhibitor (Zhang et al. 2009), by combining DFT-based conformation analysis with QSAR, Zhang et al. developed a new approach (DFT/QSAR) to carry out bioactive conformation analyses for a series of cyclic imide derivatives. Further, the potential energy surface scan, molecular docking and molecular dynamic simulation validated that the DFT/QSAR-derived conformation is very similar to the ‘real’ bioactive conformation.
- ♣ Various non-covalent interactions between DNA and hydroxylated polycyclic aromatic hydrocarbon (HO-PAH) were explored using QSAR and molecular docking study. The non-covalent interactions such as hydrogen-bonding, π - π stacking and hydrophobic interactions were found to be the typical interactions between HO-PAHs and DNA. The generated QSAR model showed the importance of molecular size, polarizability and electrostatic potential of hydrocarbons towards DNA binding affinity (Li et al. 2011a).
- ♣ Docking and DFT Studies of 3-substituted 2,6-piperazindiones derivatives at Topoisomerase II ATP Pocket conclude that some ligand properties including the hindrance effect, hydrogen bonds, p-p interactions and stereogenic centres are important for the ligand to be recognised at the binding site (Correa-Basurto et al. 2012a).
- ♣ Docking, MD simulation, MM-PBSA energy of the drug molecules such as suramin, mol-6, sirtinol, 67, and nf675 as Human Sirtuin 2 inhibitor revealed hydrogen bonds are key interactions responsible for the inhibition and the

inhibition is driven by van der Waals/ non-polar interactions (Sakkiah et al. 2013).

- ♣ QSAR studies on *N*-aryl-monosubstituted derivatives showed that bond lengths, frontier orbital energies, molecular electrostatic potential are important in acetyl cholinesterase and butyryl cholinesterase targets. The complex formed between ChEs and best *N*-aryl compound reproduced the experimentally reported binding mode, supporting generated model in docking and simulation studies (Correa-Basurto et al. 2014).
- ♣ 3D-QSAR pharmacophore model, ADME and DFT studies of structurally diverse compounds with experimental activity values were performed to understand the structural requirements and molecular properties suitable for ketohexokinase inhibitors. The mode of KHK inhibition is found to be mainly by hydrogen bond interactions at the active site (Kavitha et al. 2015).
- ♣ DFT, Docking, free energy, MD simulation for the identification of NS2B/NS3 protease inhibitors explored the better inhibitor, which correlates well with the experimental IC₅₀ values (Balajee et al. 2016).
- ♣ MD Simulation and MM/PBSA Free Energy of Type I and Type II ligands against Cyclin-dependent kinase8- protein cyclin C (CDK8-CycC) reveals the importance of van der Waals interactions and hydrogen bonding in ligand binding (Cholko et al. 2018).
- ♣ Various analyses such as binding energy calculations, hydrogen bond occupation and alanine scanning were used to compare the stability of Adenosine deaminase-inhibitor complex using MD Simulation and quantum mechanical calculations. This study serves as a theoretical basis to design effective ADA inhibitors (Tian et al. 2018).
- ♣ Dynamical and thermodynamic properties of the specific binding between the N-PDZ of Shank3 and the extended PDZ binding motif of SAPAP (synapse-associated protein 90/postsynaptic density-95 associated protein) were performed using MD simulation and MM/GBSA approach (Piao et al. 2019).
- ♣ SARS CoV-2 proteins : Drug repurposing. Novel-phyto quinoline and existing quinoline-based drugs were screened using *in silico* methods (Docking/ Bioactivity prediction/ ADMET properties) as therapeutic agents against

SARS CoV-2 3CLPro and spike protein of virus (Alexpandi et al. 2020). High-throughput virtual screening of FDA approved drug molecules as potential inhibitors against spike glycoprotein (Awad et al. 2022). The repurposing of FDA approved drugs as SARS CoV-2 cell entry inhibitors (Choudhary et al. 2020). Virtual screening of 3000 FDA approved drug molecules were screened against spike protein and viral main proteinase (Maffucci and Contini 2020). Drug repurposing through multi-QSAR modelling (Adhikari et al. 2022). Further, *in vivo* testing to validate the SARS CoV-2 inhibition by the molecules should be performed.

- ♣ With the help of enhanced umbrella sampling techniques, the residues which are essential for enhanced inhibition potential of Aminoarylbenzosuberene (AAB) molecules against 11 β -hydroxysteroid dehydrogenase (11 β -HSD1) protein were identified (Singh et al. 2022).
- ♣ Molecular docking and Simulations of Acridinedione scaffolds identified potential molecule against dengue fever (DENV-2 capsid protein) using *in silico* strategy (Kumar et al. 2022).

The above literatures support the fact that computational studies could propose novel insights/hypotheses which aids the medicine research. With the view of this, a membrane protein (human G-protein coupled glucagon receptor) and two globular proteins (Poly (ADP-ribose) polymerase 1 & 2 and Coronavirus main protease) were selected along with the suitable small organic inhibitors for the present research work. A brief discussion about the protein-ligand systems considered in this work are provided below.

1.4.1 G-protein coupled glucagon receptors

GCGR is a 62 kDa protein which belongs to Class B family of G- protein-coupled receptor superfamily. GCGR consists of an extracellular domain (ECD) and a trans-membrane domain (TMD) with a stalk region connecting both the domains (**Figure 1.1**). The ECD consists of a common α - β - β - α fold similar to the ECD of other class B GCGRs. The TMD features the canonical seven trans-membrane helical bundles (TM1-TM7) of G-proteins (Zhang et al. 2017). Many small molecule antagonists with

varying potency and structural features have been reported as GCGR antagonists and these inhibitors range from glucagon neutralizing antibodies to small molecular antagonists (Duffy et al. 2005; Kurukulasuriya et al. 2004; Shen et al. 2011). Recently a novel allosteric pocket outside the transmembrane domains have been reported which provides a scope to design improved therapeutics or antagonists against hyperglycemia. Although several publications have been reported on highly potent small molecule antagonists with desirable selectivity, only a few have been entered to clinical trials. Therefore, there is a dire need to develop a predictive biological model comprising of structurally diverse GCGR inhibitors based on crystal structure to improve the efficacy and safety of GCGR selective inhibitors.

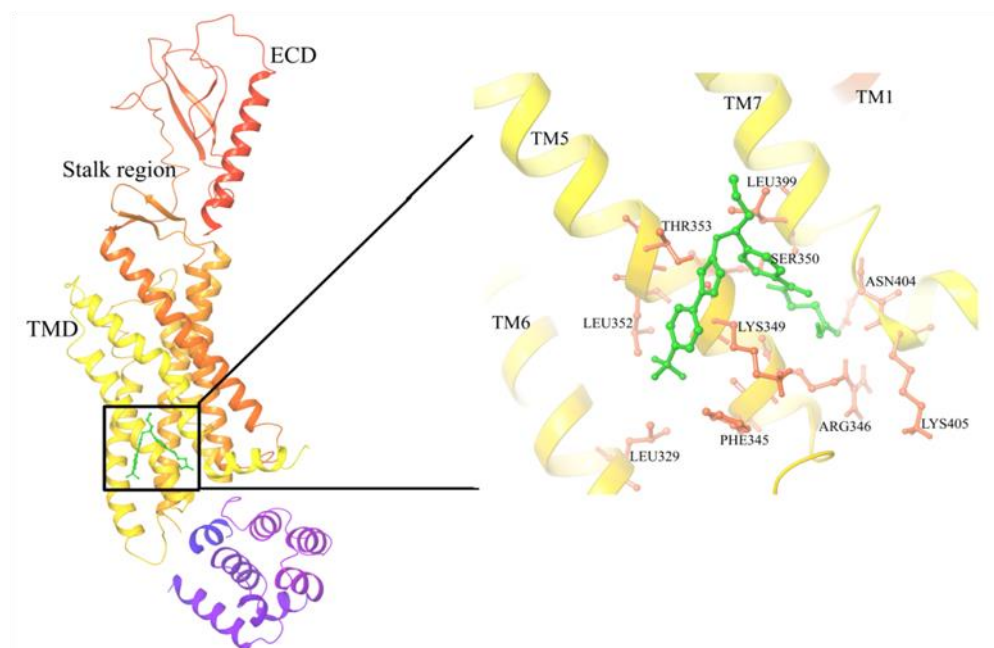


Figure 1.2: Secondary-structure representation of human GCGR (PDB ID: 5XEZ) with a hypothetical ligand showing the trans-membrane domain (TMD), the stalk region and the extracellular domain (ECD). An enlarged view of the extrahelical ligand binding site of GCGR with the major amino acid residues at the catalytic pocket was shown next to it.

1.4.2 Poly (ADP-ribose) polymerase 1 & 2

Poly ADP-ribosylation (PARylation) is an important post-translational protein modification which contributes to molecular and cellular processes including DNA

damage repair (Hottiger 2015). It has been believed that PARP-1 is responsible for all the DNA-damage dependent activities in the cell. The involvement of PARP-2 in DNA-damage dependent activity is noticed as a result of the residual DNA-dependent PARP activity in embryonic fibroblasts from PARP-1 deficient mouse (Farres et al. 2013; Shieh et al. 1998). However, it is found that PARP-2 activity is less as compared to PARP-1 towards DNA-damage response. PARP-2 bears 69% homology with PARP-1 at the catalytic domain (**Figure 1.2**) and functions by PARylation (Amé et al. 2004; Yelamos et al. 2011). Unlike PARP-1, the catalytic domain of PARP-2 contains an additional 3 amino acids in the loop connecting the β -strands of the enzyme. Even though both, PARP-1 and PARP-2, shares a similar DNA binding domain (DBD), PARP-2 is structurally different from PARP-1. PARP-2 does not contain zinc-finger motifs for DNA binding but a highly basic DNA binding domain with nuclear and nucleolar localization signals. PARP-2 has a modular structure consisting of a C-terminal domain (CAT), a central WGR (Trp-Gly-Arg) domain and an N-terminal domain (NTR). CAT domain is composed of a helical subdomain (HD) and ADP-ribosyltransferase (ART) subdomain which allows DNA-break detection (Riccio et al. 2016).

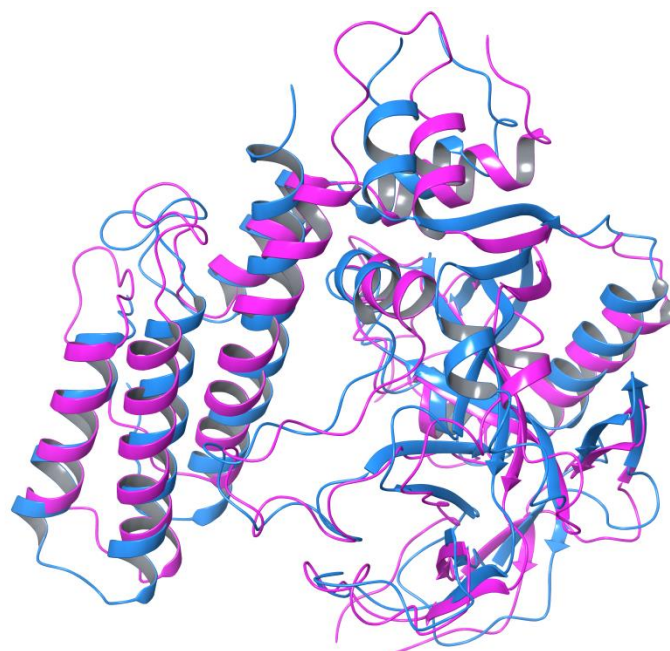


Figure 1.3: The cartoon diagram of superposition of the PARP enzymes. PARP-2 is shown in pink colour and PARP-1 is shown in blue colour.

PARP inhibition for cancer therapy works based either on the potentiation of DNA-damaging in chemo-/radiotherapy (Delaney et al. 2000; Donawho et al. 2007) or by synthetic lethality in PARP dependent cells (Ashworth 2008; Mendes-Pereira et al. 2009). The PARP inhibitors competitively inhibit the enzyme by replacing NAD⁺ substrates at nicotinamide-binding pocket. Majority of the PARP inhibitors contain carboxamide group which is similar to nicotinamide group to mimic the binding pose at catalytic site (Chen et al. 2018; Costantino et al. 2001; Reddy et al. 2018; Zhou et al. 2017). Also, the function of PARP-2 protein targets and its interaction in the cellular processes is dysregulated in carcinogenesis. For instance, selective targeting of PARP-2 enzyme inhibits androgen receptor (AR) signalling in prostate cancer by blocking the interaction with FOXA1 (forkhead box A1) proteins (Gui et al. 2019). Other studies showed that PARP-2 inhibition increased cortical cell survival in the model of post-ischaemic brain damage (Moroni et al. 2009), promoted the SIRT1 expression, mitochondrial content, protein activity (Bai et al. 2011) and regulated PPAR γ (peroxisome proliferator-activated receptor γ) in the treatment of obesity (Ke et al. 2019). From the literature it is found that Gly429Ala variant cause breast cancer (Dingerdissen et al. 2018) and Gly429 is responsible for stabilizing inhibitor through hydrogen bonding. Similarly, it is reported that Glu558Ala variant abolishes PARP-2 activity without affecting the localization to DNA damage sites. Therefore, there is a dire need of designing PARP-2 inhibitors with high isoform selectivity.

1.4.3 Coronavirus main protease

Main protease (M^{Pto}) plays a major role in the life cycle of novel coronavirus and is the key enzyme for replication and transcription process. M^{Pto} process the precursor polyproteins to form functional proteins inside the virus, mainly the post-translational processing of replicase polyprotein (Wang et al. 2016). Main protease is a 34 kDa protein (306 amino acid residues), which is composed of 3 domains (**Figure 1.3**). Domain I (8-101 residues), domain II (102-184 residues) has an antiparallel β -barrel structure and domain III (201-303 residues) has 5 α -helices arranged into an antiparallel globular structure. Domain III is connected to II by a loop region containing residues 185-200 (Jin et al. 2020). The substrate binding site and the Cys-His catalytic dyad are found in the cleft between domain I and domain II, similar to

previously reported coronavirus protease (Ren et al. 2017; Xue et al. 2008; Yang et al. 2003). The activity of protease is triggered by the binding of small organic molecules to the substrate binding site of the enzyme. However, this process can be blocked by inhibitors, which binds to the active site, thereby inhibiting the action of enzyme. Recent literatures suggest that small organic molecules such as polyphenols, alkamides, piperamides can effectively inhibit protease enzyme which can be developed as effective therapeutic drugs (Bolelli et al. 2020; Chojnacka et al. 2020; Gutierrez-Villagomez et al. 2020; Jin et al. 2020; Kumar et al. 2020a; b; Narkhede et al. 2020). Also, repurposing of existing drugs helps us to understand the probable essential molecular structure of the drugs which can be used for the inhibition of the protease (Baby et al. 2020; Riva et al. 2020; Shah et al. 2020; Zhou et al. 2020). Therefore, M^{Pro} is an ideal and attractive target for the design and development of effective drug candidates against SARS CoV-2.

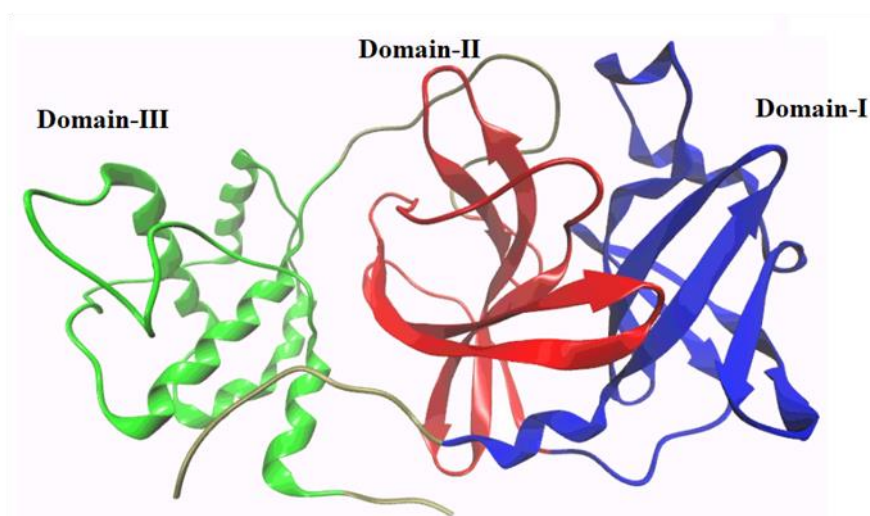


Figure 1.4: Representation of SARS CoV-2 main protease (PDB ID: 6Y2E) showing Domain I (blue), Domain II (red) and Domain III (green).

1.5 OBJECTIVES OF THE WORK

- * To determine the factors affecting the inhibition of human G-protein Coupled Glucagon Receptor (hGCGR) by β -alanine analogues.
- * To determine the mechanism and energy requirements for Poly (ADP-ribose) polymerases I and II inhibition (PARP I & II) by benzimidazole derivatives.
- * To study the effect of single amino acid substitution on ligand binding at the catalytic domain of Poly (ADP-ribose) polymerases II enzyme (PARP II).
- * To study the temperature-dependent stability of SARS CoV-2 main protease and molecular mechanism of its inhibition using small organic inhibitors.

CHAPTER 2

MATERIALS AND METHODS

In this section, a brief discussion regarding the methodology used in analysing the dynamics of various proteins and protein-ligand complexes is provided. Also, a brief background of various protein-structural analyses and water-structure analyses around the protein with and without the ligand is included.

With the development of supercomputers, software and algorithms, the computer-aided drug discovery process benefited much from various computational techniques which help in the screening of drug molecules and the prediction of binding affinity with the target. Biomolecular simulations investigate both the structural and dynamical aspects of protein-ligand complexes from the nano- to the macroscale which guides in understanding and elucidating molecular mechanisms (Tozzini 2010). Virtual screening techniques such as molecular docking, QSAR and pharmacophore modelling search the chemical database to provide the lead molecules based on the binding site properties of the target protein (Shoichet 2004; Tropsha 2010; Yang 2010). Density Functional Theory (DFT) calculations provide a detailed analysis of the electronic structure of isolated drug molecules and drug delivery systems. It can also provide deep insights into the protein-drug interactions (Hoffmann and Rychlewski 2002). A brief discussion on the background of various techniques is provided in the following sections.

2.1 MOLECULAR DYNAMICS SIMULATION

Molecular Dynamics Simulation (MD) is an extremely powerful technique which involves solving the many-particle problem in contexts relevant to the study of matter at the atomic level. The method allows the prediction of the static and dynamic properties of substances directly from the underlying interactions between the molecules. The atoms and molecules of the desired system are allowed to interact for a fixed period of time, giving a view of the dynamic evolution of the system. In

general, the trajectories of the particles are determined by numerically solving Newton's equations of motion for a system of interacting particles. The forces between the particles and their potential energies are calculated using inter-atomic potentials or molecular mechanics force fields. In 2013, Martin Karplus, Michael Levitt and Arieh Warshel were awarded the Nobel Prize in Chemistry for developing the multiscale models for complex biomolecular systems. The theory and computational methods were recognized as the direct and necessary complement to experiments (Karplus 2014).

Simulations can give ultimate details concerning the individual particle motion as a function of time. Thus, they can be used to address specific questions about the properties of a biological system (proteins, nucleic acids etc), often more efficiently than experiments of actual systems (Hospital et al. 2015).

2.1.1 Theory

In classical MD simulation, the time evolution of a group of atoms/molecules in a system of N particles can be analyzed by integrating Newton's second law of motion and is given by

$$\mathbf{F}_i = m_i \mathbf{a}_i \quad \dots\dots\dots (2.1)$$

where \mathbf{F}_i is the force acting on the atoms due to inter-atomic interactions, m_i is the mass of the atom and \mathbf{q}_i is the acceleration defined at inter-atomic distance, \mathbf{q} , by $\frac{d^2 \mathbf{q}_i}{dt^2}$. The Hamiltonian equations can be solved to obtain a continuous trajectory of a system of N particles (Tuckerman and Martyna 2000) as follows.

$$H(q_1, \dots, q_N; p_1, \dots, p_N) = \frac{1}{2} \sum_{i=1}^N \frac{p_i^2}{m_i} + U(q_1, \dots, q_N) \quad \dots\dots\dots (2.2)$$

where, m_i , q_i and p_i are masses, coordinates and momenta of particles in the system and H corresponds to the total energy of the system which is the sum of kinetic energy and potential energy. For a harmonic oscillator, kinetic energy, $K = \frac{1}{2} m_i \left(\frac{d^2 q_i}{dt^2} \right)^2$ and potential energy, $U = \frac{1}{2} k q_i^2$, where q_i is the displacement, m_i is the mass and $k > 0$.

The momentum conjugate to q_i is $p = m_i \frac{dq_i}{dt}$. Thus kinetic energy is written as $K = \frac{1}{2} \frac{p^2}{m}$. Thus the Hamiltonian of the system is given by the equation (Chew et al. 2019),

$$H = K + U = \frac{1}{2} \frac{p^2}{m} + \frac{1}{2} k q_i^2 \quad \dots\dots\dots(2.3)$$

The equations of motion can be related to the following Hamilton's equations.

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} = \frac{p_i}{m_i} \quad \dots\dots\dots (2.4)$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} = -k q_i = F_i = m_i \frac{d^2 x}{dt^2} \quad \dots\dots\dots (2.5)$$

(2.4) shows the relationship between velocity and momentum of particles in a system while (2.5) is equal to Newton's second law of motion (2.1). Therefore, the initial coordinates of particles and the initial distribution of velocity and acceleration determined by the energy gradient are the major requirements to calculate the continuous trajectories of particles in a system. The initial features of the system can be obtained from experimental techniques such as X-ray crystallography and NMR spectroscopy.

For many-particle system, numerical algorithms have been developed to integrate the equations of motion which is impossible with analytical algorithms. The Leap-frog algorithm (Hockney and Eastwood 1988) is the most commonly used algorithm in MD simulation. Other used algorithms in simulation methodology are Verlet (Verlet 1967), velocity Verlet (Swope et al. 1982) and Beeman (Beeman 1976) algorithms.

Leap-Frog Algorithm: It uses positions (r) at the time (t) and velocities (v) at time $t - \frac{1}{2} \partial t$ and updates the positions and velocity by following equations.

$$v \left(t + \frac{1}{2} \partial t \right) = v \left(t - \frac{1}{2} \partial t \right) + \frac{\partial t}{m} F(t) \quad \dots\dots\dots (2.6)$$

$$r(t + \partial t) = r(t) + \partial t v \left(t + \frac{1}{2} \partial t \right) \quad \dots\dots\dots (2.7)$$

Verlet Algorithm: It uses positions (r) and accelerations (a) of the particle at time t rather than the velocity. The new positions $t+\frac{1}{2}\partial t$ are calculated from the positions at $t-\frac{1}{2}\partial t$. The algorithm is given by

$$r(t + \partial t) = r(t) + v(t)\partial t + \frac{1}{2}a(t)\partial t^2 \quad \text{..... (2.8)}$$

$$r(t - \partial t) = r(t) - v(t)\partial t + \frac{1}{2}a(t)\partial t^2 \quad \text{..... (2.9)}$$

Summing (2.7) and (2.8),

$$r(t + \partial t) = 2r(t) - r(t - \partial t) + a(t)\partial t^2 \quad \text{..... (2.10)}$$

Velocity-Verlet Algorithm: It uses positions (r) and velocities (v) at time t to integrate the equations of motions. This algorithm gives position, velocity and acceleration at the same time.

$$r(t + \partial t) = r(t) + \partial tv(t) + \frac{1}{2}\partial t^2 a(t) \quad \text{..... (2.11)}$$

$$v(t + \partial t) = v(t) + \frac{1}{2}\partial t[a(t) + a(t + \partial t)] \quad \text{..... (2.12)}$$

2.1.2 Force Fields

In molecular modelling, force fields refer to a set of functional forms and parameters used to model potential energy of a system. The functional forms of energy include interactions by covalent bonds and long-range non-bonded interactions such as electrostatic and van der Waals forces. Many force fields are used for the simulations of biomolecular systems, among which CHARMM (Brooks et al. 1983), AMBER (Weiner et al. 1984), OPLS (Jorgensen and Tirado-Rives 1988) and GROMOS (van Gunsteren and Berendsen 1987) force fields are standard.

In CHARMM force field, potential energy (V) has the form:

$$\begin{aligned}
V = & \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} k_\phi[1 + \cos(n\phi - \delta)] \\
& + \sum_{impropers} k_\omega(\omega - \omega_0)^2 + \sum_{Urey-Bradley} k_{UB}(r_{1,3} - r_{1,3;0})^2 + \sum_{\emptyset, \phi} V_{CMAP} \\
& + \sum_{non-bonded} \varepsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^6 \right] + \sum_{non-bonded} \frac{q_i q_j}{4\pi \varepsilon_0 r_{ij}} \\
& \dots\dots\dots (2.13)
\end{aligned}$$

where, the terms corresponding to harmonic bond stretching potential (b is the bond length), harmonic bond angle potential (θ is the bond angle), proper dihedral potential (ϕ is the dihedral angle and δ is phase shift), improper dihedral potential, Urey-Bradley potential ($r_{1,3}$ is the distance between atoms 1 and 3), grid-based energy correction map, van der Waals potential (r_{ij} is the inter-particle distance, ε_{ij} is the well depth and σ is the distance where potential is zero) and electrostatic potential (q is the charge of particle and ε_0 is the permittivity of free space) respectively. k is the force constant for each interaction terms.

For AMBER force field, potential energy (V) has the form:

$$\begin{aligned}
V = & \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} \frac{1}{2} V_n [1 + \cos(n\phi - \delta)] \\
& + \sum_{impropers} k_\omega(\omega - \omega_0)^2 + \sum_{non-bonded} \varepsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^6 \right] \\
& + \sum_{non-bonded} \frac{q_i q_j}{4\pi \varepsilon_0 r_{ij}} \\
& \dots\dots\dots (2.14)
\end{aligned}$$

The potential energy in AMBER force field is similar to CHARMM force field, where the former lacks Urey-Bradley potential and CMAP correction terms.

2.2 DENSITY FUNCTIONAL THEORY

Density Functional Theory (DFT) calculations can be helpful in predicting the electronic properties of the molecules such as optimized bond angles, optimized bond distances, Mulliken charges, HOMO-LUMO energies, energy gap, dipole moment and vibrational frequency. The electronic interactions of the molecules play an

important role in its pharmacological effects. The energy of HOMO directly related to the ionization potential and the energy of LUMO is related to the electron affinity. Also, both are the indicators of possible electrophilic and nucleophilic attack sites in the molecule respectively. The energy gap value represents its chemical reactivity. The smaller energy gap indicates less stability and enables the rapid electron transfer and exchange, thereby making the molecule highly reactive. The van der Waals surface over the molecules will provide a measure of charge distribution from the viewpoint of the approaching reagent.

2.2.1 Hohenberg-Kohn Theorems

Many different methods exist for calculating the electronic structure and properties of atomic systems. In particular, a large hierarchy of methods, known collectively as quantum chemistry can ultimately provide very accurate answers (Szabo and Ostlund 2018). The principle idea of DFT calculation is to replace the complicated many body wave-function with the much simpler electron density (Cohen et al. 2008). Hohenberg and Kohn proved two theorems that enable the electron density to be used instead. The first Hohenberg-Kohn theorem is as follows:

The external potential $V_{ext}(r)$ is, to within a constant, a unique functional of the electronic density $n(r)$. Since in turn $V_{ext}(r)$ fixes \hat{H} , it is seen that the full many-particle ground state is a unique functional of $n(r)$.

So the Hamiltonian and all the properties of a system are determined uniquely by the system's electron density. The second Hohenberg-Kohn theorem is as follows:

For a trial $\tilde{n}(r)$ that satisfies the boundary conditions $\tilde{n}(r) \geq 0$ and $\int \tilde{n}(r) d^3r = N$ and is associated with a \tilde{V}_{ext} , the energy it gives is an upper bound to the true energy E_0 .

This theorem states that the energy of the system is variational with respect to the density. That is, the lowest energy is for the ground state density and any other density results in a higher energy.

2.2.2 Kohn-Sham Equations

Kohn-Sham equation is one electron Schrodinger equation of non-interacting particle that shows same density as a system of interacting particles (Kohn and Sham 1965; Parr and Yang 1994). Kohn-Sham equations are defined as a local effective external potential in which the non- interacting particles move, Kohn-Sham potential ($V_{\text{eff}}(r)$). As the particles in the system are non-interacting, the Kohn-Sham wave function is a single Slater determinant from a set of lowest energy orbitals obtained as a solution to the following equation:

$$\left(\frac{-\hbar^2}{2m}\nabla^2 + v_{\text{eff}}(r)\right)\phi_i(r) = \varepsilon_i\phi_i(r) \quad \text{..... (2.15)}$$

This is the typical Kohn-Sham equations and ε_i represents the orbital energy corresponding to Kohn-Sham orbital ϕ_i and the density for system consists of N particles are given by

$$\rho(r) = \sum_i^N |\phi_i(r)|^2 \quad \text{..... (2.16)}$$

Therefore, the energy can be expressed in the terms of density as

$$E(\rho) = T_s(\rho) + \int dr V_{\text{ext}}(r)\rho(r) + E_H(\rho) + E_{xc}(\rho) \quad \text{..... (2.17)}$$

where, T_s is the Kohn- Sham kinetic energy, V_{ext} is the external potential acting on interacting system, E_H is the Coulomb energy and E_{xc} is the exchange correlation energy.

The Kohn- Sham kinetic energy in terms of orbital is given by

$$T_s(\rho) = \sum_{i=1}^N \int dr \phi_i^*(r) \left(\frac{-\hbar^2}{2m}\nabla^2\right) \phi_i(r) \quad \text{..... (2.18)}$$

The Kohn- Sham equations are obtained by varying the total energy expression with respect to a set of orbitals that yields the Kohn- Sham potential as

$$v_{\text{eff}}(r) = v_{\text{ext}}(r) + e^2 \int \frac{\rho(r')}{|r-r'|} dr' + \frac{\delta E_{xc}(\rho)}{\delta \rho(r)} \quad \text{..... (2.19)}$$

The last term is the exchange correlation potential, $v_{xc}(r)$, the only unknown in this approach. The Kohn- Sham orbital energies ε_i have little physical significance. The total energy of the system can be given as the sum of orbital energies as

$$E = \sum_i^N \varepsilon_i - E_H(\rho) + E_{xc}(\rho) - \int \frac{\delta E_{xc}(\rho)}{\delta \rho(r)} \rho(r) d(r) \quad \dots\dots\dots (2.20)$$

2.3 MOLECULAR DOCKING

In the field of molecular modeling, molecular docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex (Lengauer and Rarey 1996). The aim of molecular docking is to achieve an optimized conformation for both the target and ligand so that the free energy of the overall system is minimized (**Figure 2.1**). Molecular recognition plays an important role in promoting fundamental biomolecular events such as enzyme-substrate, drug-protein and drug-nucleic acid interactions. Detailed understanding of the general principles that govern the nature of the interactions (van der Waals, electrostatic, hydrogen bonding, aromatic) between the ligands and their protein or nucleic acid targets may provide a framework for designing the desired potency and specificity of potential drug candidates for a given therapeutic target (Gaba et al. 2010).

Two approaches are common in docking procedure. One approach uses matching technique that describes the protein and the ligand as complementary surfaces (Meng et al. 1992; Shoichet et al. 1992). The complementary between the two surfaces amounts to the shape matching description that may helps in finding the complementary pose of docking the target and the ligand molecules. The shape complementarity based approaches are robust and fast, but they cannot usually model the movements or dynamic changes in the ligand/target conformations accurately. The second approach simulates the docking process in which the ligand-protein pair wise interaction energies are calculated (Feig et al. 2004). In this approach, the target and the ligand are separated by some physical distance, and the ligand finds its position into the active site of target after a certain number of moves in its conformational space. The moves incorporate rigid body transformations such as translations and rotations, as well as internal changes to the structure of ligand including torsion angle rotations. Each of these moves in the conformation space of the ligand induces a total energetic cost of the system, and hence after every move the total energy of the system is calculated.

The overall binding energy of the system is given by

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + (V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf}) \dots\dots\dots (2.21)$$

where, L refers to ligand, P refers to protein (target) and P-L refers to the complex.

Each pair-wise calculation (V) can be given by

$$V = W_{vdw} \sum_{i,j} \left(\frac{A_{i,j}}{r_{i,j}^{12}} - \frac{B_{i,j}}{r_{i,j}^6} \right) + W_{hbond} \sum E(t) \left(\frac{C_{i,j}}{r_{i,j}^{12}} - \frac{D_{i,j}}{r_{i,j}^{10}} \right) + E_{ele} \sum \frac{q_i q_j}{e(r_{i,j}) r_{i,j}} + W_{sol} \sum_{i,j} (S_i V_j + S_j V_i) e^{-\frac{r_{i,j}^2}{2\sigma^2}} \dots\dots\dots (2.22)$$

The optimized weighing constants, W, is to calibrate the empirical free energy based on a set of experimentally determined binding constants. The first term is a typical 6/12 potential for dispersion/repulsion interactions. The second term is for directional hydrogen bond based on 10/12 potential. The third term is Coulomb potential for electrostatics and the final term is desolvation potential.

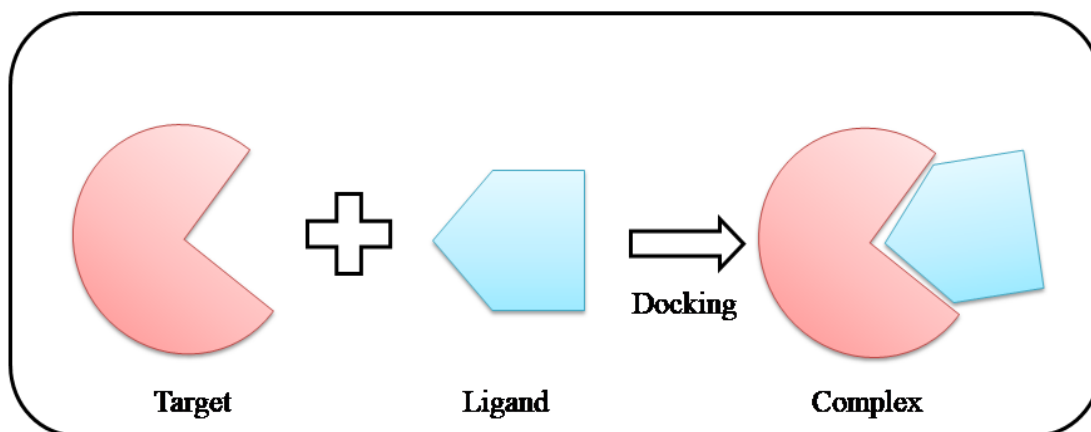


Figure 2.1: Schematic representation of docking process

2.3.1 Basic Requirements for Docking

The setup for a ligand docking approach requires a target structure, the molecules of interest or a database containing existing or virtual compounds for the docking process, and a computational framework that allows the implementation of the desired docking and scoring procedures. Most docking algorithms assume target/

protein to be rigid; the ligand is mostly regarded as flexible. Besides the conformational degrees of freedom, the binding pose of ligand in the binding pocket of target must be taken into consideration. The final binding energy of ligand to a target considers van der Waals interaction, electrostatic interactions, directional hydrogen bonds, conformational entropy, rotational entropy and desolvation terms (Meng et al. 1992, 2011).

2.3.2 *Types of Molecular Docking*

- **Rigid Body Docking:** Both the receptor and the ligand are treated as rigid bodies, limiting the search space and considering only 3 translational and 3 rotational degrees of freedom. The ligand flexibility could be addressed using a pre-computed set of ligand conformations, or by allowing a degree of atom–atom overlap between the target and the ligand (Meng et al. 2011).
- **Flexible-Ligand Docking:** Generally, the ligand is treated as flexible while the receptor is assumed to be rigid during molecular docking. It considers van der Waals interaction, directional hydrogen bonds, electrostatic interactions, conformational entropy, rotational entropy and desolvation terms to calculate the binding affinity/ docking score (Meng et al. 2011).
- **Flexible Docking:** The flexibility of protein is highly related to the behavior of ligand at the binding site (Teague 2003). Incorporating receptor flexibility is a significant challenge in the field of molecular docking. The high computational expense prevents the use of this method in the screening of large chemical/drug database.

2.4 PHARMACOPHORE MODELLING

A pharmacophore is regarded as the description of chemical features of a molecule that are necessary for the molecular recognition of a ligand by the target protein and it also points to the structurally diverse molecules which can bind to the target binding site. Ehrlich proposed the concept of pharmacophore in 1909 which is formed by the framework of atoms with active essential properties. In 1977, Gund revised the

definition of pharmacophore as the group of molecules that can recognize target protein and form structural features of biological activity. There are two main approaches in the identification of a pharmacophore model. If the target protein is available, the pharmacophore model is defined by considering the mode of action of ligand/drug molecule on the target. If the target is unknown, a series of drug molecules were analyzed based on the functional groups affecting the activity of the molecule by conformational analysis (Kaserer et al. 2015). Thus, a model can be used as a reference to identify the lead molecules that satisfy a similar biological activity from a drug database.

2.4.1 Pharmacophore Sites

For the model development, the molecule is considered as a set of points in the 3D space which represents various chemical characteristics that promotes the non-covalent interactions between the compound and the amino acid residues at the target protein. These are regarded as the pharmacophore sites which are characterized by directionality, type and location. For instance, based on the common non-covalent interactions, PHASE module in Schrodinger software provides 6 built-in pharmacophore features (**Figure 2.2**):

- A : Hydrogen bond acceptor
- D : Hydrogen bond donor
- H : Hydrophobic groups
- N : Negative ionisable groups
- P : Positive ionisable groups
- R : Aromatic rings

In addition to these features, other chemical properties can also be defined according to the properties of the molecule (Dixon et al. 2006). A hydrogen bond acceptor site is defined as the surface accessible atom which carries donatable lone pairs whereas hydrogen bond donor site is defined as the group with donatable hydrogen atoms. Hydrophobic sites include isopropyl groups, t-butyl groups, halogenated moieties, rings and chains. Positive and negative ionisable sites are considered as a point on formally charged atom or the center of a group of atoms where the charge is shared. Lastly, the aromatic rings are regarded as a separate pharmacophore site which is placed at the center of aromatic rings.

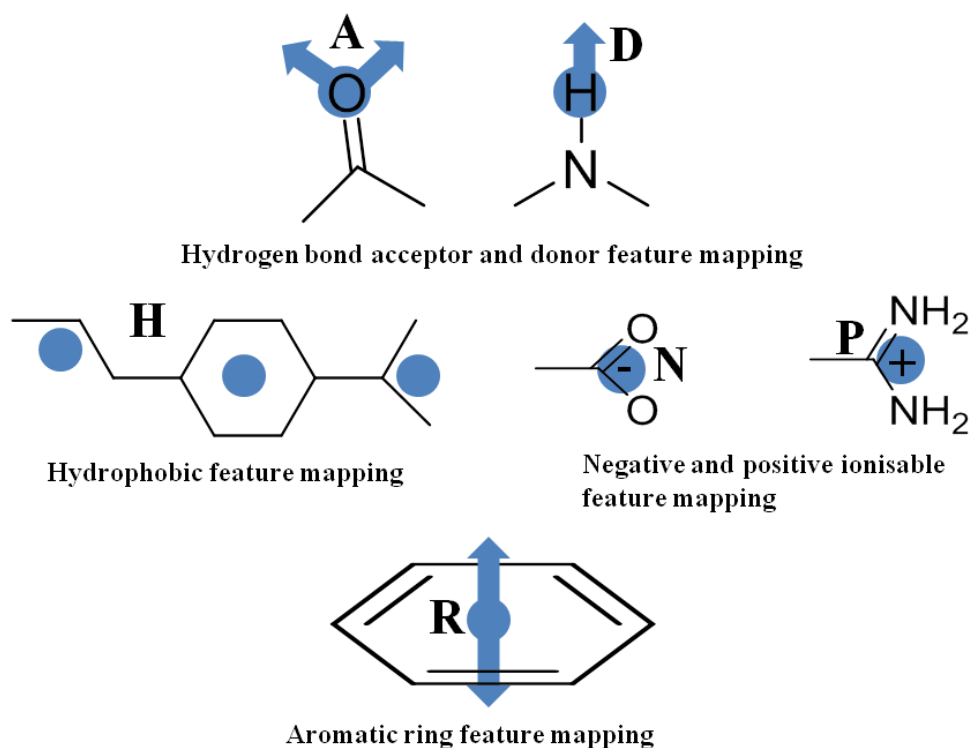


Figure 2.2: Some examples featuring the pharmacophore sites in PHASE module

2.5 QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP

Quantitative Structure-Activity Relationship (QSAR) is of great importance in the biological system, especially in the target-ligand interaction. The aim of this analysis is to transform a set of predictors (physico-chemical properties of molecule) into a mathematically quantified and computerised form (Karelson et al. 1996). It is a regression model which correlates the molecular properties to the biological activity of the molecule (**Figure 2.3**). If the target protein structure is unknown, QSAR method is regarded as the most effective and accurate technique for drug design. If the developed model is valid, it can predict the biological activity of newly designed drug candidates (Chirico and Gramatica 2012; Tropsha 2010). The four main steps in QSAR analysis (Yousefinejad and Hemmateenejad 2015) include

- ❖ Selection of data set and extraction of structural/empirical descriptors
- ❖ Variable selection
- ❖ Model construction
- ❖ Evaluation and validation

Consider a dataset of m compounds showing significant biological activity (Y) having n molecular descriptors (P), then the generalized expression for QSAR calculation can be given as:

$$Y = C_0 + C_1P_1 + \dots + C_mP_n \quad \dots\dots\dots (2.23)$$

where, C_1, C_2, \dots, C_m are the corresponding regression coefficients and C_0 is the constant term of the model. For the ideal model, the value of R^2 (multiple correlation coefficient) should be 1. As the value of R^2 deviates from 1, the fitting quality of the model deteriorates (Pirhadi et al. 2015).

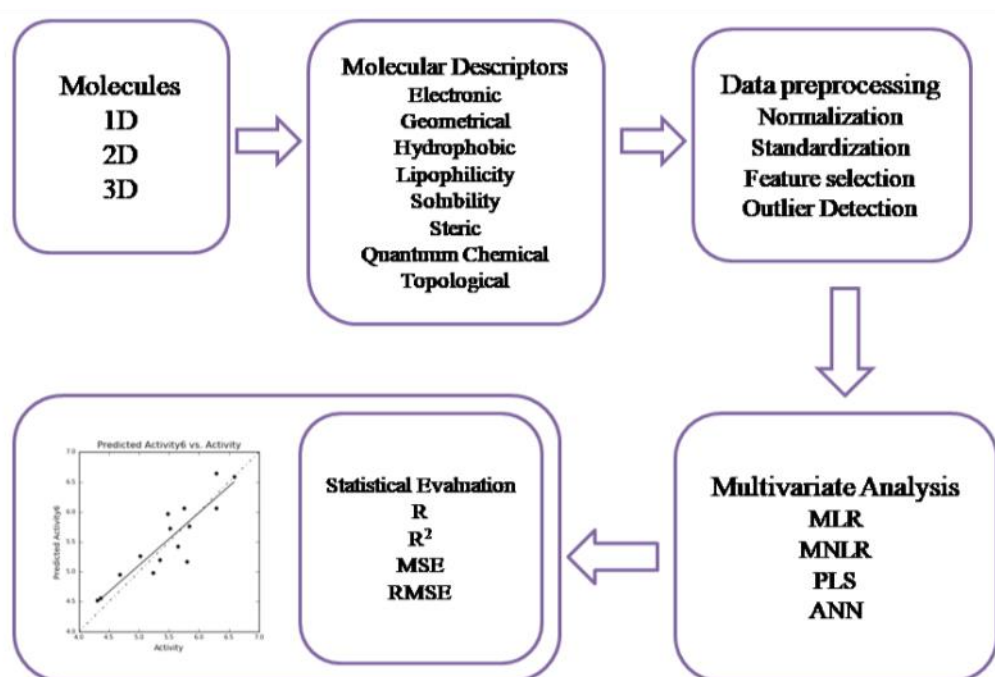


Figure 2.3: Schematic overview of QSAR process

2.5.1 Types of SAR Approach

- ❖ **Fragment- based QSAR:** In this method, each molecule in the data set is considered as a group of fragments. The fragmentation of the molecule is carried out either using template based or user defined methods. The descriptors are calculated for each fragments and the relationship between the fragment descriptors form the basis of prediction of biological activity of entire molecule (Ajmani et al. 2009).
- ❖ **3D-QSAR:** The biological activities of the ligands were predicted by exploiting their 3D properties using statistical correlation methods. It is a

promising tool for the prediction and design of new pharmaceuticals and agrochemicals (Verma et al. 2010). It includes the application of force field calculations on 3D-structures of a series of aligned small molecules with known activities. It uses Lennard-Jones potential for the entire molecule rather than the fragments. The various approaches in 3D-QSAR analysis include molecular shape analysis, molecular topology difference, comparative molecular movement analysis, hypothetical active site lattice, self organising molecular field analysis, comparative molecular field analysis and comparative molecular similarity indices.

- ❖ **Chemical descriptor based QSAR:** This approach is based on various molecular descriptors such as electronic, geometric or steric properties of a molecule. This approach is different from fragment-based approach as the chemical properties are calculated for the entire molecule rather than the fragments. It is different from 3D-QSAR approach as the properties are calculated from scalar quantities (energies, geometric parameters) than from 3D fields.

2.5.2 Molecular Descriptors for QSAR Study

Molecular descriptors are the numerical representation of chemical information obtained from the structural and electronic features in all the possible dimensions of a molecule. The various molecular descriptors such as hydrophobic parameters, electronic parameters, steric parameters, quantum chemical descriptors and spatial descriptors are given in **Table 2.1**.

Table 2.1: Molecular descriptors used in QSAR

Type	Descriptors
Hydrophobic Parameters	Partition coefficient, logP Solubility parameter, logs Distribution coefficient, logD π -substituent constant

Electronic Parameters	Hammett constant, σ Ionization constant Chemical shifts Polarizability Electronegativity Electrophilicity
Steric Parameters	Molar mass Molar volume Van der waals radius Van der waals volume Molar refractivity Surface area (approx.) Surface area (solvent accessible) Hydration energy
Quantum Chemical Descriptors	Net atomic charge HOMO- LUMO energies HOMO- LUMO gap Dipole moment
Spatial Descriptor	Radius of gyration Principle moment of inertia

2.5.3 Chemometric Methods

Multiple Linear Regression: MLR is a commonly used method in QSAR due to its simplicity, transparency, reproducibility, and easy interpretability. The generalized expression of an MLR equation is as follows:

$$Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + \dots + a_nX_n \quad \text{..... (2.24)}$$

In the above expression, Y is the response or dependent variable, X₁, X₂, ..., X_n are descriptors (features or independent variables) present in the model with the corresponding regression coefficients a₁, a₂, ..., a_n, respectively, and a₀ is the constant term of the model. For the ideal model, the sum of squared residuals being 0, the value of R² (multiple correlation coefficient) is 1. As the value of R² deviates from 1, the fitting quality of the model deteriorates. For overall significance of the regression coefficient the variance ratio (the ratio of regression mean square to deviations mean square) (F) should be high (Pirhadi et al. 2015).

Multiple Non-Linear Regression: MNLN helps improve our prediction of the structure-function relationship and to quantify the substituting effect. The coefficients R, R², and F-values were used to select the best regression performance. Some of the programmed function for MNLN is given as:

$$Y = a + (bX_1 + cX_2 + dX_3 + eX_4 + \dots) + (fX_{21} + gX_{22} + hX_{23} + iX_{24} + \dots) \dots \dots \dots (2.25)$$

where a, b, c, d represent the parameters and X₁, X₂, X₃, X₄ represent the variables (Hadaji et al. 2017).

Partial Least Squares regression: PLS is an efficient method to identify criteria based on covariance and is a generalization of MLR. It is recommended in cases where the number of variables is high and where it is likely that the explanatory variables are correlated. PLS is more realistic than MLR to construct models of the often complicated relationships between chemical structure and biological activity. In addition, PLS is able to address many collinear structure descriptor variables which make it easier to clarify the variation of compound structures (Pirhadi et al. 2015; van de Waterbeemd 1995).

Artificial Neural Network: ANN, which mimics human brain process information, is useful in detecting complex non-linear relationship between a set of inputs and outputs. The general structure of ANN has one input layer, one or more hidden layers and one output layer. Each layer has some units corresponding to neurons. The units in neighboring layers are fully interconnected with links corresponding to synapses. ANN learns an approximate nonlinear relationship by a training procedure. The

number of training cycles was selected on the basis of the Mean Squared Error (MSE) of the validation subset, which prevents the network from over-training. Neural networks for modeling in conjunction with genetic algorithms have proved very powerful for optimization (Ghanbarzadeh et al. 2015).

2.6 ADME/TOXICITY PREDICTION

Adsorption-distribution-metabolism-excretion studies of compounds are an essential part of ligand-based drug discovery. The two important concepts in ADME/Toxicity prediction include whether the molecule possesses drug-like pharmacokinetic properties and whether the chemical properties of molecule cause safety issues on human tissues (Vrbanac and Slauter 2017). The elimination of the molecules with unfavourable pharmacokinetic properties is an important step in drug development process. The permissible ranges of crucial pharmacokinetic properties used in drug discovery programs are documented below:

mol_MW	Molecular weight of the molecule	: 130.0 - 725.0 g
dipole	Dipole moment of the molecule	: 1.0- 12.5 debye
SASA	Total solvent accessible surface area	: 300.0- 1000.0 Å ²
donorHB	Estimated number of hydrogen bonds donated by the compound to water molecules in aqueous solution	: 0.0- 6.0
accptHB	Estimated number of hydrogen bonds accepted by the compound from water molecules in aqueous solution	: 2.0- 20.0
QPlogP_{o/w}	Predicted octanol/water partition coefficient	: -2.0- 6.5
QPlogS	Predicted aqueous solubility, concentration of compound in saturated solution	: -6.5- 0.5 mol dm ⁻³
rotor	Number of non-trivial, non-hindered rotatable bonds	: 0-15
PSA	van der Waals surface area of polar nitrogen and oxygen atoms, carbonyl carbon atoms	: 7.0- 200.0 Å ²

QPlogK_{hsa}	Prediction of binding to human serum albumin	: -1.5- 1.5
QPlogBB	Predicted blood/ brain partition coefficient	: -3.0- 1.2
QPPCaco	Predicted apparent Caco-2 cell permeability in nm/sec	: < 25 poor, > 500 great
QPPMDCK	Predicted apparent MDCK cell permeability in nm/sec	: < 25 poor, > 500 great

Percentage human oral absorption > 80% high and < 25% poor

RuleOfFive Based on the Lipinski's rule of 5 (Lipinski 2004). The compounds should satisfy the following rules: mol_MW < 500, QPlogPo/w < 5, donorHB <5, accptHB < 10.

RuleOfThree Based on Jorgensen's rule of three. The compounds with fewer violations of the following properties are considered to be orally available: QPlogS > -5.7, QPPCaco > 22nm/s, primary metabolites < 7 (Lionta et al. 2014).

In the next section, a brief discussion about the various structural and thermodynamical analyses carried out in the whole research work is provided.

2.7 DATA ANALYSES FROM SIMULATION TRAJECTORY

2.7.1 Network Analysis

Network analysis and clustering have been carried out to group the prominent structures where both orientation and conformation are equally essential to understand mechanism of protein function (Abramyan et al. 2016). The conformations for network analysis were obtained by in-built clustering algorithm in GROMACS software using RMSD as criterion. The selection of a proper cut-off is crucial as a lower cut off value gives too many conformations which may lead to irrelevant results; while a larger cut-off value neglects the finer details of the conformations. An open software, Gephi (Bastian et al. 2009), is used for constructing graphs for the network analysis. Edge and Node files were created by executing the following equation:

$$d(t, t') = \sqrt{\frac{1}{N} \sum_{i=1}^N (\delta_i(t))^2} \quad \dots\dots\dots (2.26)$$

where δ_i is the distance between atom i at time t' with the reference conformation of N number of heavy atoms at time t . Each different conformation can be represented by a node (circles) and the connectivity between the nodes can be represented by the edges (lines). The node and edge list for construction of the network map for each temperature is done by using above equation. The network distribution of the conformations provides an interactive visualization which explains the connectivity between the structures. The layout of the networks is determined using Fruchterman-Reingold layout, a force-directed layout algorithm in which the edges moves closer or far apart minimizing the equilibrium energy (Gajdoš et al. 2016). The parameter which characterizes each cluster is taken as modularity which measures the network structures and shows the number of communities. The network with high modularity is densely connected nodes within the community but sparse connections between the communities.

2.7.2 Protein Structural Parameters

Root mean square deviation: The root-mean-square-deviation (RMSD) of protein was calculated as a function of time in order to evaluate the degree of conformational drift of the protein from initial structure. RMSD can be calculated by least-square fitting the desired conformation to reference conformation ($t_2 = 0$) and is given by

$$RMSD(t_1, t_2) = \left[\frac{1}{M} \sum_{i=1}^N m_i ||r_i(t_1) - r_i(t_2)||^2 \right]^{1/2} \quad \dots\dots\dots (2.27)$$

where $r_i(t)$ is the position of atom i at time t . The flattening of RMSD curve is the indication of equilibrated protein system.

The energy change of the evolved conformations during a simulation as a function of RMSD can be calculated by

$$\Delta G = -k_B T \ln P(rmsd) \quad \dots\dots\dots (2.28)$$

where k_B is the Boltzmann constant (8.3145×10^{-3} kJ/mol) and $P(rmsd)$ is the probability of conformations with particular RMSD (Anandakrishnan et al. 2019).

The free energy profiles of the protein provide the information about 1-D free energy landscape which depicts the most stable conformation at a particular RMSD value.

Root mean square fluctuation: The root-mean-square-fluctuation (RMSF) calculation provides the degree of flexibility of the protein backbone when it forms complex with the ligand. If x_i is the coordinates of particle i and $\langle x_i \rangle$ is the ensemble average position of i , then RMSF can be calculated as follows:

$$\rho_i = \sqrt{\langle (x_i - \langle x_i \rangle)^2 \rangle} \quad \dots\dots\dots (2.29)$$

RMSD shows the divergence of structures from the reference over time while RMSF indicates the regions of the protein with high mobility.

The isotropic temperature factor (B-factor) was calculated from the root mean square fluctuation (RMSF) values of the protein to measure the mobility of each residue from the initial position during the simulation. The B-factor is directly related to RMSF by the following formula (Caldararu et al. 2019).

$$B - factor = \frac{8}{3} \pi^2 RMSF^2 \quad \dots\dots\dots (2.30)$$

Radius of Gyration: Radius of gyration (Rg) measures the compactness of protein-ligand complex during the simulation. Rg can be calculated using the equation,

$$Rg = \left(\frac{\sum_i |r_i|^2 m_i}{\sum_i m_i} \right)^{\frac{1}{2}} \quad \dots\dots\dots (2.31)$$

where m_i is the mass of the atom i and r_i is the position of atom i with respect to the center of mass of the molecule. It is helpful to characterize polymer solutions and biomolecules such as proteins.

Hydrogen bond occupancy: In the structural analysis, the hydrogen bond interactions exists between two non-overlapping groups of atoms, if the distance between donor heavy atom and the acceptor (HD---A), $r < 3.5 \text{ \AA}$ (Tan et al. 2021) and the angle (H-D-A), $\alpha < 30^\circ$ (Petukhov et al. 2004). The hydrogen bond occupancy is defined as the fraction of the time the molecule are hydrogen-bonded in the total simulation trajectory based on the geometric criteria.

2.7.3 Radial Distribution Function

Radial Distribution function (RDF) in a system of particles such as atoms, molecules, ions, colloids etc, describes the variation in density as a function of distance from a reference point. It can also be defined as the probability of finding a particle at a distance ‘r’ from the reference point. The equation of RDF $g(r)$ is given as

$$g(r) = 4\pi r^2 \rho dr \quad \dots\dots\dots (2.32)$$

where ρ is the average number density of the particles. In molecular dynamics simulation, the count of atom pairs in a given range of separation, r is considered to determine the RDF. The general expression to determine the RDF is given by

$$rG(r) = 4\pi r^2 (\rho(r) - \rho(0)) \quad \dots\dots\dots (2.33)$$

where $\rho(r)$ is the atomic density at distance r and $\rho(0)$ is the bulk density of atoms (Sha et al. 2011).

2.7.4 Tetrahedral Order Parameter

The structure of water molecules can be characterized based on the tetrahedral order parameter. The orientation order parameter is defined as the normalized sum of the squares of the differences between the cosines of the inter-bond angles and the cosine of the angles that would have been made if the bonds are arranged tetrahedrally (Chau and Hardwick 1998). Since, all the angles between bonds are the same in a tetrahedral arrangement,

$$S_g = \frac{3}{32} \sum_{j=1}^3 \sum_{k=j+1}^4 (\cos\theta_{jk} + \frac{1}{3})^2 \quad \dots\dots\dots (2.34)$$

where θ_j, k is the angle between the central oxygen atom and the j^{th} and k^{th} bonds. The factor $3/32$ normalizes the S_g values to the range $0 < S_g < 1$. The squaring ensures the contribution from each inter-bond angle is always greater than or equal to zero. If all the angles are arranged in perfect tetrahedral order, the value of S_g will be zero and 1 for extreme non-tetrahedral cases. If the arrangement is random, then S_g has a value 0.25.

2.7.5 Hydrogen Bond dynamics

The hydrogen bond population variables $h(t)$ and $H(t)$ were defined to calculate the hydrogen bond dynamics. $h(t) = 1$, when a particular pair of water molecules are hydrogen bonded at time t according to the definition of hydrogen bond. If two molecules remain continuously hydrogen bonded from $t=0$ to $t=t$, then $H(t) = 1$ or it is 0 otherwise. The two molecules are considered to be hydrogen-bonded if the interatomic distance between hydrogen of donor and acceptor atom (H---O) is $< 2.5 \text{ \AA}$. The continuous hydrogen bond autocorrelation functions $S_{HB}(t)$ are defined as,

$$S_{HB}(t) = \frac{\langle h(0)H(t) \rangle}{\langle h(0)^2 \rangle} \dots\dots\dots (2.35)$$

where $\langle \dots \rangle$ describes an average over all the pairs of a given type. The function describes the probability of hydrogen bond pair at $t=0$ continuously bonded up to time t (Luzar and Chandler 1996; Paul and Chandra 2005).

2.7.6 Free Energy Calculations

Free energy calculations can predict the energy of biomolecules (proteins, nucleic acids etc.) and protein-ligand affinities accurately. A range of computational approaches including free energy perturbation (FEP) (Kollman 1993), thermodynamic integration (TI) (Kollman et al. 2000), linear interaction energies (LIE) (Foloppe and Hubbard 2006), molecular-mechanics generalised born surface area (MM/GBSA) (Genheden and Ryde 2015), molecular-mechanics Poisson-Boltzmann surface area (MM/PBSA) (Homeyer and Gohlke 2012) have been adopted to estimate free energies.

Generally, the average binding free energy of complex was calculated using following equation,

$$\Delta G_{bind} = \Delta G_{Complex} - (\Delta G_{Protein} + \Delta G_{Ligand}) \dots\dots\dots (2.36)$$

$\Delta G_{Complex}$ is the Gibbs free energy of protein-substrate complex, whereas $\Delta G_{Protein}$ and ΔG_{Ligand} are individual Gibbs free energy of the protein and the ligand molecule respectively.

MM/PB(GB)SA Approach: The solvation effects are equally important in the simulation of biomolecular system in water. The combination of classical force fields and continuum solvation models to calculate the binding free energy of biomolecular systems are more efficient. MM/PB(GB)SA methods are based on the above combination and are computationally more efficient than other existing methods. These methods can decompose the energy into contributing energy terms.

Each energy term (X) of (2.36) can be written in the form

$$\Delta G(X) = E_{MM}(X) + \Delta G_{Sol}(X) + T\Delta S(X) \quad \text{..... (2.37)}$$

$$E_{MM}(X) = E_{Bonded} + E_{Electrostatic} + E_{vdW} \quad \text{..... (2.38)}$$

$$\Delta G_{Sol} = \Delta G_{PB(GB)} + \Delta G_{SASA} \quad \text{..... (2.39)}$$

where, $E_{MM}(X)$ is the molecular mechanics energy of system X in gas phase, $\Delta G_{Sol}(X)$ is the solvation free energy and S is the entropy. Internal energy, electrostatic energy and van der Waals energy contributes to E_{MM} while solvation free energy and non-polar energy contributes to ΔG_{Sol} of the molecular system. E_{bonded} includes bonds, angles and dihedrals, is given by

$$E_{Bonded} = E_{Bond} + E_{Angle} + E_{dihedrals} \quad \text{..... (2.40)}$$

ΔG_{SASA} is the combination of solvent accessible surface area (SASA) and two empirical parameters (γ, β).

$$\Delta G_{SASA} = \gamma.SASA + \beta \quad \text{..... (2.41)}$$

Finite-Difference Approximation: The entropy ($-\Delta S_r(T)$) contributions in free energy profile at an inter-state separation r for a temperature T is calculated by the following equation

$$-\Delta S_r(T) = \frac{\Delta G_r(T+\Delta T) - \Delta G_r(T-\Delta T)}{2\Delta T} \quad \text{..... (2.42)}$$

where the thermodynamic quantities are the functions of temperature, T and the difference is taken at constant volume and particle number. To calculate ΔG , ΔS and ΔH , three simulations are required at the temperature, $T \pm \Delta T$. The free energy ($\Delta G(r)$)

between the particular molecules can be determined with the help of pair correlated functions, $g(r)$, for each temperature (Pettitt and Rossky 1986; Smith and Haymet 1993). The equation is given by

$$\Delta G(r) = -k_B T \ln g(r) \quad \dots\dots\dots (2.43)$$

where r is the inter-atomic separations, k_B is the Boltzmann constant and T is the temperature under consideration.

Further, the enthalpic contribution ($\Delta H(r)$) to free energy for desired temperature is calculated by

$$\Delta H(r) = \Delta G(r) + T \Delta S(r) \quad \dots\dots\dots (2.44)$$

Free energy of protein folding-unfolding: The free energy of folding for protein was determined by the changes in α -helix and β -sheet elements.

$$\Delta G_{fold} = -k_B T \ln \left(\frac{F_N}{1-F_N} \right) \quad \dots\dots\dots (2.45)$$

where F_N is the fraction of protein folding (Tanford 1968). The fraction of folding (F_N) is determined by the fraction of residues with α -helix, β -sheet elements and turns present in the protein as a function of time.

CHAPTER 3

EFFECT OF HYDROPHOBIC AND HYDROGEN BONDING INTERACTIONS ON THE POTENCY OF β -ALANINE ANALOGUES OF G-PROTEIN COUPLED GLUCAGON RECEPTOR INHIBITORS

G-protein coupled glucagon receptors (GCGRs) play an important role in glucose homeostasis and pathophysiology of Type-II Diabetes Mellitus (T2DM). Hydrophobic interactions with the amino acid residues present at TM5 and hydrogen bonding interactions toward TM7 found to facilitate the favorable orientation of antagonist at GCGR allosteric pocket.

3.1 BACKGROUND

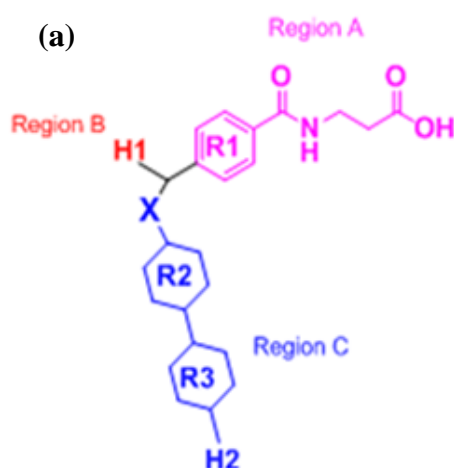
Type 2 Diabetes Mellitus (T2DM) is known to be a chronic metabolic disorder which is mainly caused due to improper lifestyle and genetics. Rapid growth of T2DM and lack of proper medication for this disease has become one of the real problems in recent days. The medical condition is characterised by elevated hepatic glucose level in the blood due to deregulated signal transduction by GCGR in hepatocytes. T2DM is also increase the risk of diabetic-related complications such as weight loss, blindness, kidney failure, amputations and cardiovascular diseases.

Glucagon is a small peptide hormone consists of 29 amino acid residues (Jazayeri et al. 2016), secreted by the α -cells of pancreatic islets and known to activate GCGR. In normal conditions, the attenuation of insulin inhibitory effect and release of glucagon from pancreas, increase the hepatic glucose level in blood by glycogenolysis during fasting. The combined activities of both insulin and glucagon have a crucial role in glucose homeostasis in the human body. Glucagon activated hepatic glucagon receptors (GCGR) transduces the activation of adenylate cyclase and initiates cAMP (cyclic adenosine monophosphate) production. This process ultimately, ends up with the expression of enzymes responsible for gluconeogenesis and glycogenolysis (Jiang and Zhang 2003). Therefore, blocking the activation of GCGR is believed to be an efficient way to control the abnormal hepatic glucose

production in T2DM patients. Despite of several medications available, the development of new improved therapeutics has been hindered due to lack of structural details of GCGR. Recently, the publications of x-ray crystal structure of GCGR provided an opportunity to design improved therapeutics by both structure and ligand based drug design for this receptor class (Jazayeri et al. 2016; Zhang et al. 2017). In this work, a series of experimentally tested pyrazole ethers and aminopyrazole compounds of β -alanine were subjected to various *in silico* methods to gain insightful information on the crucial structural features required to develop potential antagonists of GCGR.

3.2 DATASET AND PREPARATION OF 3D STRUCTURES OF LIGAND

A dataset consisting of 58 β -alanine based glucagon receptor antagonist were selected from recent experimental reports (Filipski et al. 2012; Guzman-Perez et al. 2013; Lee et al. 2014). The structural details of the inhibitors considered in this study are illustrated in **Figure 3.1**. All the drug candidates of the dataset reported to shared same assay procedure. The experimental inhibitory constant K_i was converted into pK_i (negative logarithm of K_i) for the ease of further analysis. The 3D structures of the glucagon receptor antagonists were constructed using the builder panel in Maestro Graphical User Interface (GUI). The partial charges were ascribed and possible ionization states were generated at pH 7.4 to mimic the experimental assay condition. Further, the geometry of the resulted structures was optimized by semi-empirical PM3 (Stewart 1989) and then by B3LYP/6-31G(d,p) level (Ditchfield et al. 1971; Lee et al. 1988) respectively using Gaussian09 package (Frisch et al. 2009).



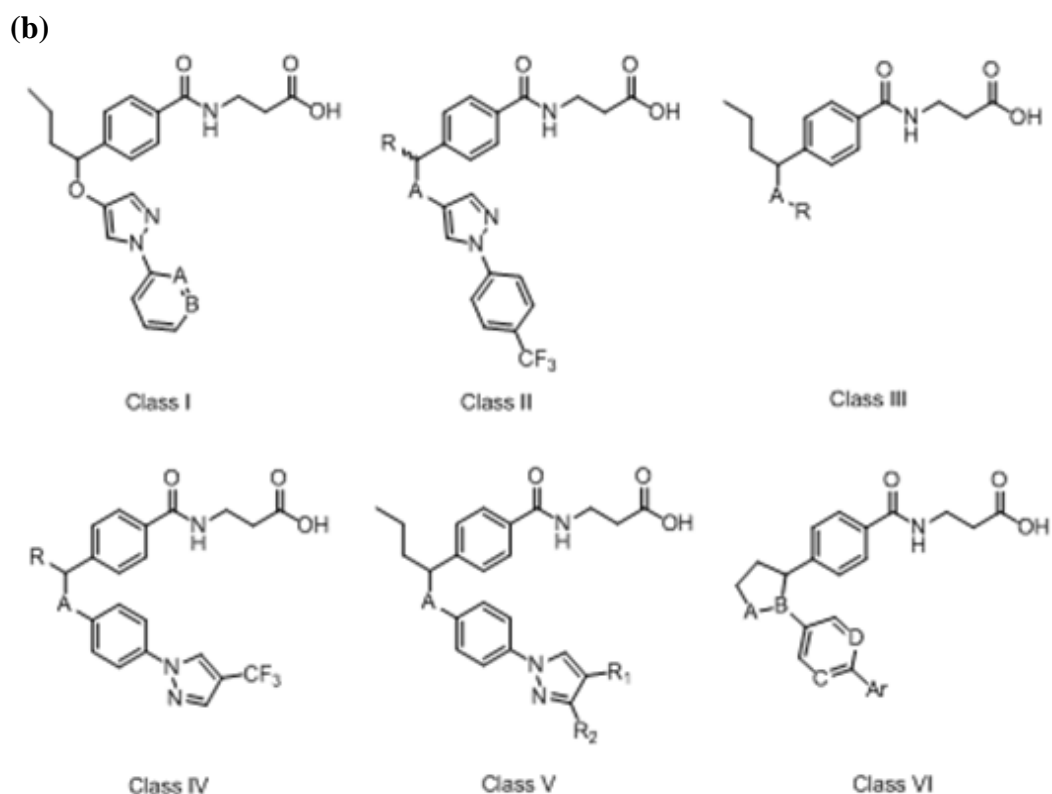


Figure 3.1: (a) Skeletal structure of glucagon receptor antagonist: Region A represented in pink colour is the polar region containing R1 ring, Region B represented in red colour is the alkyl side chain which is hydrophobic in nature (H1) and Region C represented in blue colour is the hydrophobic core containing R2, R3 ring and H2 side group. (b) Detailed structure of six classes of GCGR antagonists.

3.3 3D-QSAR ANALYSIS

In this study, Phase (Dixon et al. 2006) module of Schrödinger was employed to develop 3D-QSAR model of β -alanine based GCGR inhibitors. Prior to 3D-QSAR modeling, all the ligands were aligned using flexible shape-based alignment tool in Phase module. The entire dataset was divided into training-set and test-set constituting of 44 and 14 compounds respectively (based on standard 3:1 ratio), using “Automated Random Selection” (Rácz et al. 2015) option in Phase to yield optimum statistics in terms of correlation coefficient (R^2 , measures internal consistency) and cross-validation coefficient (Q^2 , measures internal predictability). Care was taken to include the most active and inactive molecules in the training set (Dixon et al. 2006;

Golbraikh et al. 2003). A statistically significant model was generated by using Partial Least Square regression method with a grid spacing of 1Å.

Table 3.1: PLS regression summary of generated 3D-QSAR model

PLS	SD	R ²	F	P	Stability	RMSE	Q ²	Pearson- r
1	0.5064	0.5935	61.3	9.68 x 10 ⁻¹⁰	0.667	0.38	0.6540	0.8149
2	0.2730	0.8847	157.3	5.84 x 10 ⁻²⁰	0.296	0.29	0.8009	0.9114
3	0.1613	0.9607	326.0	3.88 x 10 ⁻²⁸	0.235	0.30	0.7897	0.8973
4	0.0908	0.9879	793.2	8.93 x 10 ⁻³⁷	0.164	0.29	0.7909	0.9002
5	0.0725	0.9925	999.7	3.21 x 10 ⁻³⁹	0.154	0.28	0.8076	0.9097
6	0.0609	0.9948	1185.6	1.01 x 10 ⁻⁴⁰	0.144	0.27	0.8216	0.9172
7	0.0497	0.9966	1527.3	1.51 x 10 ⁻⁴²	0.150	0.27	0.8249	0.9189
8	0.0376	0.9981	2336.6	2.19 x 10 ⁻⁴⁵	0.152	0.27	0.8253	0.9176

SD: Standard deviation of regression; R²: Regression coefficient; F: Variance ratio; P: significance level of variance ratio; RMSE: Root mean square error; Q²: Cross validated correlation coefficient for test set; Pearson- r: square of correlation coefficient for test set.

The developed 3D-QSAR model found to have a cross- validation coefficient (Q²) of 0.8253 for test set compounds and regression co-efficient (R²) of 0.9981 for training set compounds. The model was developed with a PLS factor of 8 (N/5, where N is the number of molecules in the training set) in order to avoid the risk of over-fitting of data and to achieve a significant statistical correlation between experimental activity and predicted activity (Polański et al. 2002). PLS regression statistics of generated 3D-QSAR model shown in **Table 3.1**, exhibited good statistical stability. The greater confidence of the model is indicated from the high Pearson-r value of 0.9176 and F value of 2336.6 with smaller P value (2.19×10⁻⁴⁵). Further, an acceptably low standard deviation (SD) value of 0.0376 and root mean square error (RMSE = 0.27) indicates the predictability and reliability of the generated model. The scatter plots of experimental pK_i values vs. Phase predicted values (**Figure 3.2**) for training set and test set compounds showed a strong linear correlation and Phase predicted activities were tabulated in **Appendix I**.

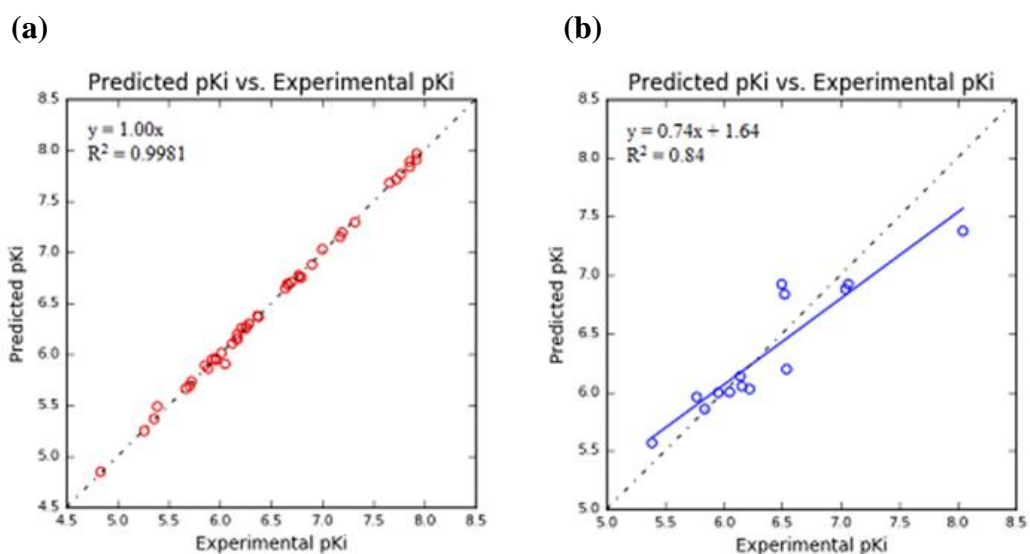


Figure 3.2: Scatter plot between the experimental activity (pK_i) vs.3D-QSAR predicted activity. **(a)** The training dataset (unfilled red circles). **(b)** The test dataset (unfilled blue circle). The best fitted equation for the scatter plot of test set compounds is given as $y = 0.74x + 1.64$ ($R^2 = 0.84$).

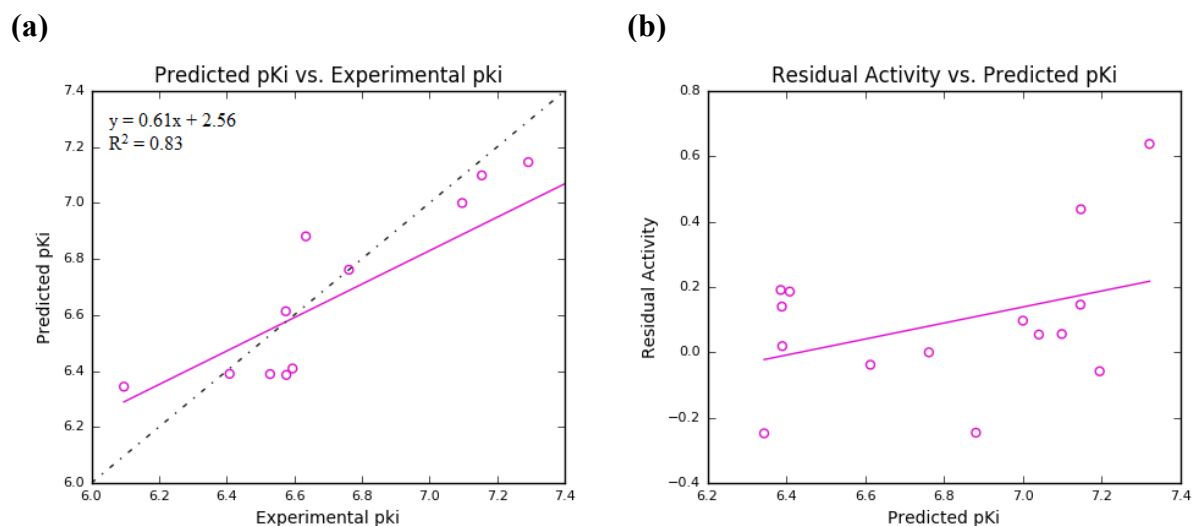


Figure 3.3: **(a)** Scatter plot showing the predicted activity vs experimental activity of external test set with best fitted line $y = 0.61x + 2.56$ ($R^2 = 0.83$) and **(b)** Plot of residual activity vs predicted activity of external test set

The efficiency of the model was validated by predicting the activity of 15 external test set compounds (Kurukulasuriya et al. 2004; Shen et al. 2011). The predicted activity of external test set is tabulated in **Appendix II**. The scatter plot of experimental activity vs. predicted activity of external test set (**Figure 3.3(a)**) showed a good linear correlation with a R^2 value of 0.83. It is believed that a 3D-QSAR model with R^2 greater than 0.5 have good predictability and reliability of the generated model (Golbraikh and Tropsha 2002). The prediction errors of developed 3D-QSAR model found to be distributed randomly near the zero line (**Figure 3.3(b)**) which denotes the absence of systematic errors due to biased calculation.

3.4 CONTOUR PLOT ANALYSIS

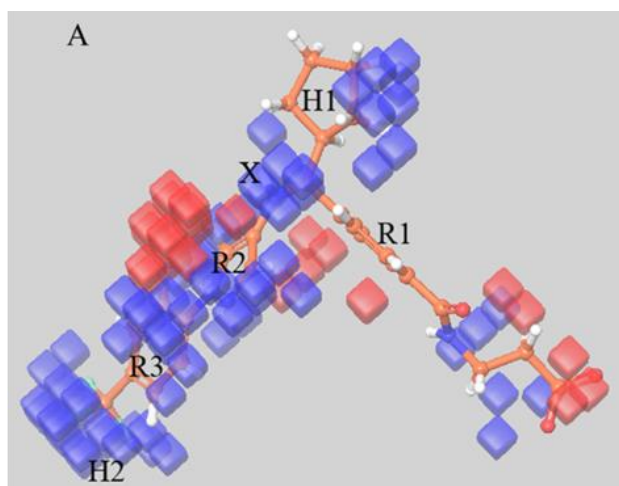
The effect of spatial arrangements of structural determinants on GCGR inhibition is analyzed by visualizing 3D-QSAR contour plots. The nature and position of substitution groups are found to have crucial role in defining the activity of chosen antagonists. It is evident from the present study that the hydrophobic/non-polar groups, hydrogen bond donor groups and electron withdrawing group have major contribution to the final 3D-QSAR model. The positive contributions are shown in blue cubes and the negative contributions are indicated by red cubes. For better visualization, the favorable and unfavorable interactions for each feature are mapped over Compound 20 and illustrated in **Figure 3.4**.

Effect of Hydrophobic core: The hydrophobic character of β -alanine analogs is found to play an important role in GCGR inhibition. **Appendix III(a) (i-vii)** illustrates the favorable positions of hydrophobic groups on R2, R3, H1 and H2 regions of the inhibitors. Among the Class I compounds, Compound 12 ($pK_i = 6.9$) and Compound 11 ($pK_i = 6.67$) are found to have high potency due to the substitution of hydrophobic $-CF_3$ group (Dalvi and Rossky 2010) and ethyl group at 4' position of R3 ring respectively. Among Class IV compounds, the acyclic hydrophobic substituents at H1 site found to increase the biological activity than their cyclic analogues. For example, Compound 40 ($pK_i = 6.769$) with tertiary-butyl group is found to be more potent than Compound 37 ($pK_i = 6.155$) having cyclo-butyl group. Similarly, Compound 36 ($pK_i = 6.143$) with iso-propyl group is found to be more potent than Compound 35 ($pK_i =$

5.886) having cyclo-propyl group. The replacement of heteroatoms with -CH and -CH₂ group at the cyclic linkage of Class VI inhibitors (Compound 58, pK_i = 7.854) enhances the hydrophobic character which contributes to its higher potency.

Effect of Hydrogen Bond Donor Group: The blue cubes indicate the favorable position of hydrogen bond donor groups (**Appendix III(b) (i-v)**). The presence of amino linkage (-NH) instead of ether group (-O-) as X at hydrophobic region found to favor the molecule to be highly active as antagonists. This explains the higher biological activity of aminopyrazole derivatives (Compound 20, pK_i = 8.046; Compound 17, pK_i = 7.721; Compound 33, pK_i = 7.921) than the pyrazole ether derivatives (Compound 13, pK_i = 6.367; Compound 31, pK_i = 7.194).

Effect of Electron Withdrawing Group: The favourable and unfavourable spatial arrangements of electron withdrawing groups are displayed in **Appendix III(c) (i-v)**. The red cubes over R3 ring indicate the unfavourable positions of electron withdrawing groups in the molecule. It is known that the presence of electron withdrawing groups alter the electron density over the ring making it more polar, thereby reducing the hydrophobicity of the molecule. The potency of Class V compounds is found to be less due to the presence of -CONMe₂, -CN, -F and -Cl groups at H2 position of R3 ring. Compound 41 (pK_i = 5.187), Compound 44 (pK_i = 5.387), Compound 46 (pK_i = 5.769) and Compound 47 (pK_i = 5.959) are found to have electron-withdrawing substituents at H2 position which make them less potent. Slight increment in the potency was observed by the replacement of electron withdrawing groups at H2 by cyclic moiety (Compound 50 (pK_i = 6.538)).



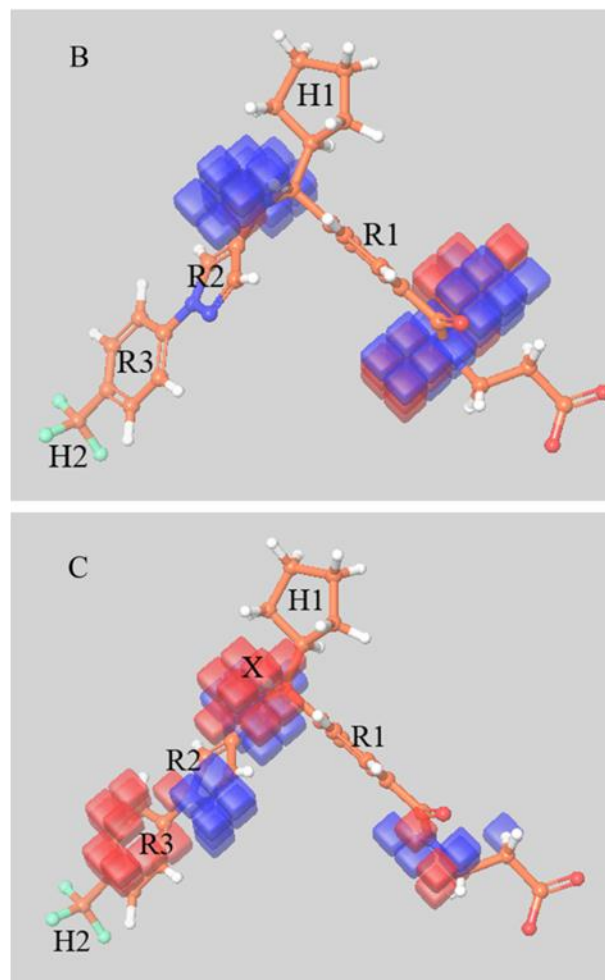


Figure 3.4: 3D-QSAR Contour Plots visualized in the context of favourable and unfavourable positions (a) hydrophobic groups, (b) hydrogen bond donor groups and (c) electron withdrawing groups over Compound 20 .

3.5 MOLECULAR DOCKING

The co-crystal structure of full-length GCGR with a negative allosteric modulator (NNC0640) (PDB ID: 5XEZ, resolution: 3Å) (Zhang et al. 2017) was retrieved from RCSB Protein Data Bank. Prior to docking the 3D structure of 5XEZ were refined with Prime (Jacobson et al. 2004) and missing atoms were added. Appropriate ionization was confirmed by adding hydrogen bond corresponding to pH of 7.4. Autodock (v4.2.6) was employed to dock pyrazole ether and aminopyrazole derivatives at the allosteric binding site of the protein. Autodock GUI (Morris et al. 2009) was used to prepare the protein coordinate suitable for docking procedure. The

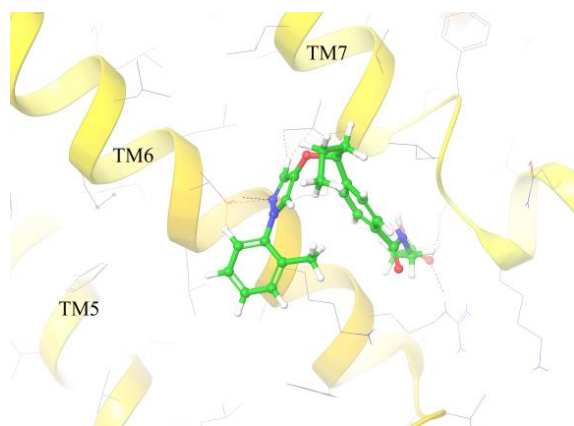
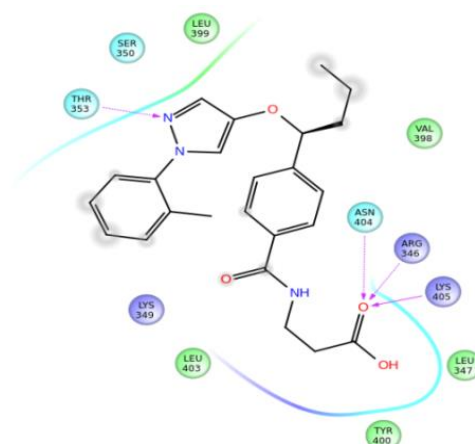
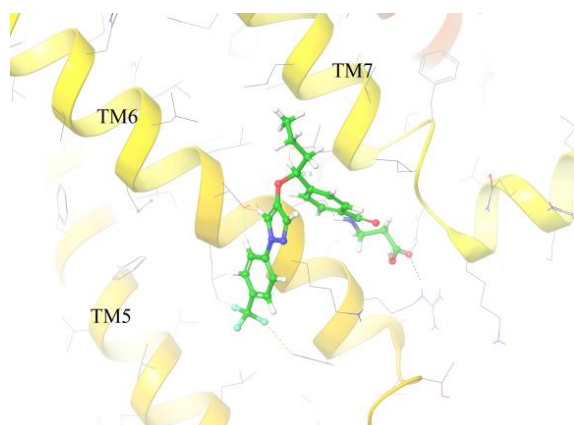
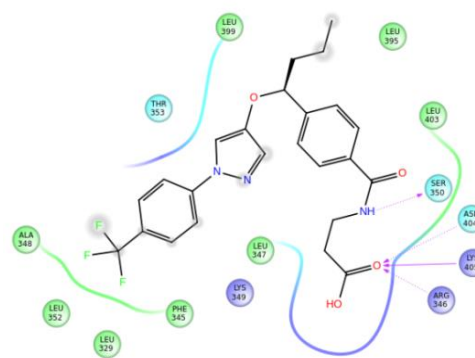
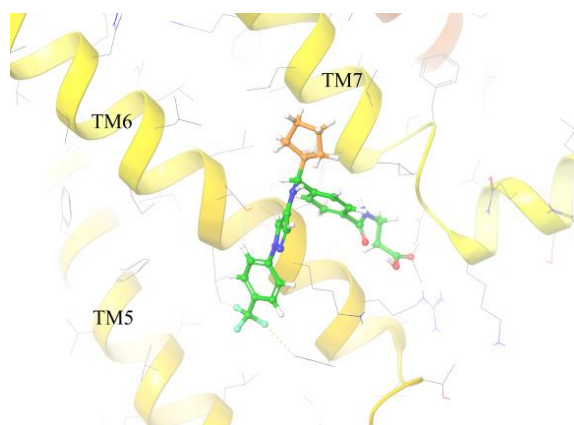
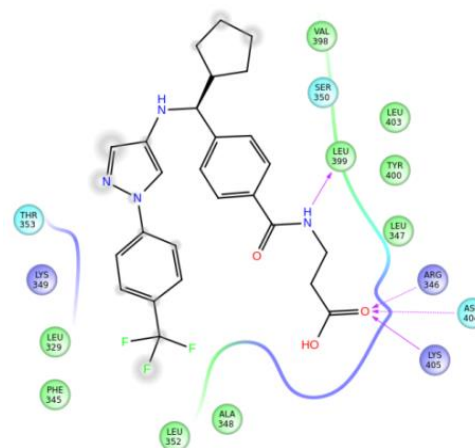
protein was prepared by removing water, membrane lipids and co-crystallized ligands. Gasteiger charges (Gasteiger and Marsili 1980) were added. A 3D grid was prepared with a dimension of 30Å x 46Å x 30Å having a spacing of 0.375Å at the allosteric pocket located to the trans-membrane domain of 5XEZ using Auto-grid program (Morris et al. 2009). Lamarckian Genetic Algorithm (Morris et al.) was used and the runs were set to 100 in order to search all the possible ligand binding conformations at the allosteric pocket. Molecular docking was performed with an initial population of 150 and a number of 2.5×10^6 energy evaluations were carried out.

The molecules were divided into 6 classes (**Figure 3.1**) for the better analysis of binding interaction of ligand with the protein. The results were tabulated (**Appendix IV**) in the increasing order of pK_i value for each class of ligands. The score obtained from docking results showed a good correlation with the experimental biological activity. The energy terms contributing to the docking energy of the molecules were given in **Appendix IV**. From the study, it was found that ligand interaction takes place in the region of amino acid residues from Arg346 to Lys405 which covers the binding site of the protein. The inhibitor orients in such a way that the polar region is aligned towards TM7 and the hydrophobic part is aligned towards TM5 as mentioned in the literature (Jazayeri et al. 2016). All the complexes were found to have the same binding pocket, with different stabilization energy. Compound 20 showed the highest potency with a least docking score of -8.25 kcal/mol and Compound 1 showed the least potency with a docking score of -5.33 kcal/mol. Hydrogen bonding and hydrophobic interactions between the ligand and the protein plays a crucial role in the stabilization of protein-ligand complex. The interaction of all the ligands mainly comprises hydrogen bonding network with residues Arg346, Ser350, Leu399, Asn404 and Lys405.

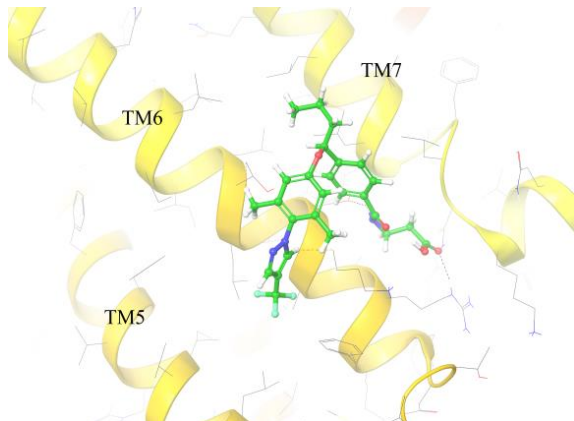
Among class I ligands, Compound 12 was found to be the most potent inhibitor with lowest docking energy of -7.79 kcal/mol. The polar region found to stabilize the complex by forming hydrogen bonds between -C=O and the residues Arg346, Ser350, Asn404 and Lys405 (**Figure 3.5b (i-ii)**). The introduction of electronegative CF_3 group at *para* position of R3 ring resulted in hydrophobic interaction (Dalvi and Rosicky 2010) with the amino acid residues which stabilizes the

hydrophobic core of ligand. The least potency of Compound 1 was attributed due to the presence of large number of steric clashes than the attractive non-bonding interactions. The presence of methyl group at *ortho* position does not favour the orientation of hydrophobic core towards TM5 (**Figure 3.5a (i-ii)**).

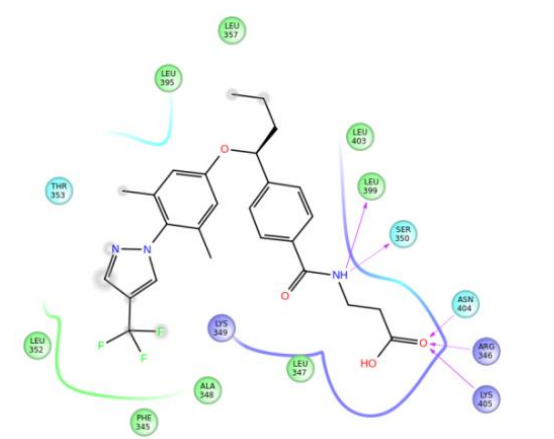
Among Class II ligands, Compound 20 is found to be highly potent with lowest docking energy of -8.25 kcal/mol. The introduction of cyclopentyl group as H1 stabilizes the protein-ligand complex (**Figure 3.5c (i-ii)**). Varying the hydrophobic core of Class III inhibitors changes the potency and stability of protein-ligand complexes. Compound 32 with a pK_i value of 7.854 showed an additional hydrogen bonding interaction between -NH and the residues Ser350 and Leu399 (**Figure 3.5d (i-ii)**). The introduction of different alkyl side chains to Class IV ether series showed less difference in its binding energy. The docking studies showed that the side chains favours the orientation of hydrophobic and polar cleft of inhibitors towards the TM5 and TM7 membranes respectively. The alkyl chain at H1 position of Compound 40 favours the formation of a stable protein-ligand complex than their cyclic analogues (**Figure 3.5e (i-ii)**). The complexity of the substituents on the heterocyclic ring of Class V inhibitors has an impact on the orientation and binding energy of the complexes. The Compound 49 and Compound 50 contains cyclic substituents which support the orientation of hydrophobic cleft towards TM5 and polar cleft towards TM7 favouring hydrogen bond formation (**Figure 3.5f (i-ii)**). The heteroatoms at cyclic linkage (H1 region) of Class VI inhibitors do not favour the formation of stable protein- ligand complex. The replacement of heteroatoms with -CH₂ and -CH groups in H1 region favours the hydrophobicity of the molecules and forms a protein-ligand complex with least binding energy (Compound 56, Compound 57, Compound 58). Additionally, Compound 58 showed hydrogen bonding between -NH and Leu399 and π -cation interaction with Lys349 (**Figure 3.5g (i-ii)**). Docking studies showed that the proper orientation of hydrophobic region of the molecule towards TM5 is essential for the formation of hydrogen bonds with Arg346, Asn404 and Lys405 amino acids to stabilize protein-ligand complex.

a(i)**a(ii)****b(i)****b(ii)****c(i)****c(ii)**

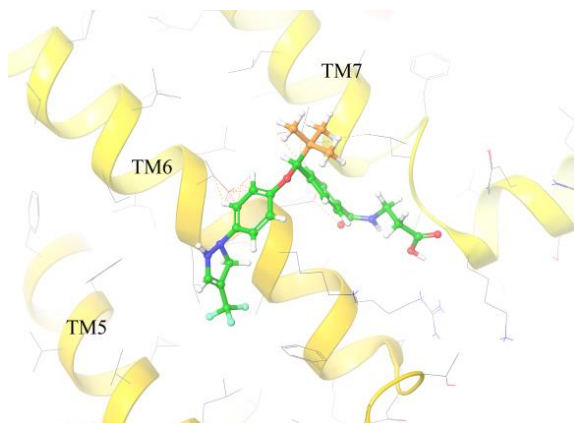
d(i)



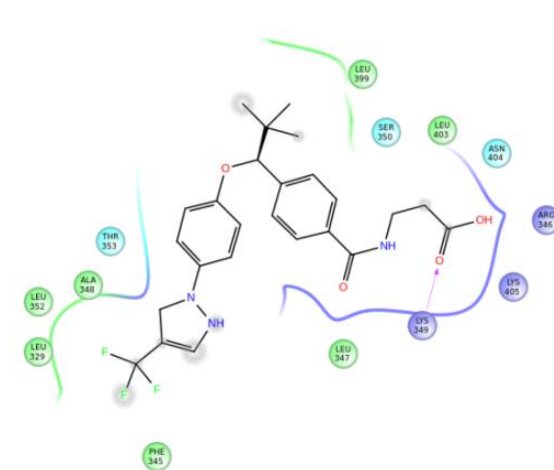
d(ii)



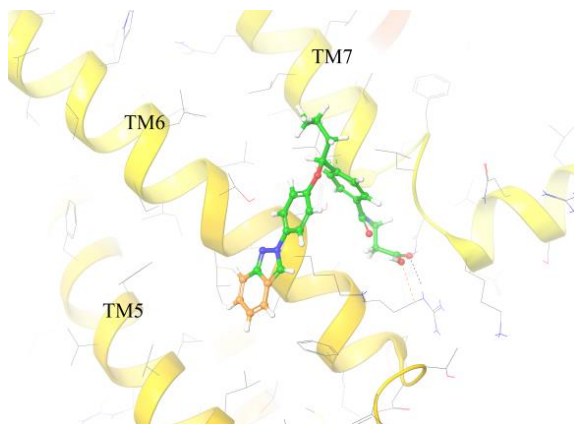
e(i)



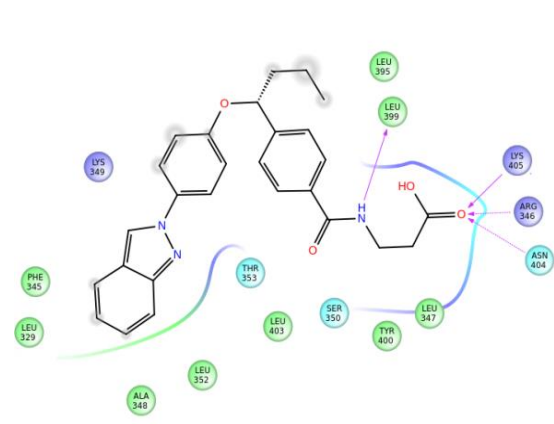
e(ii)



f(i)



f(ii)



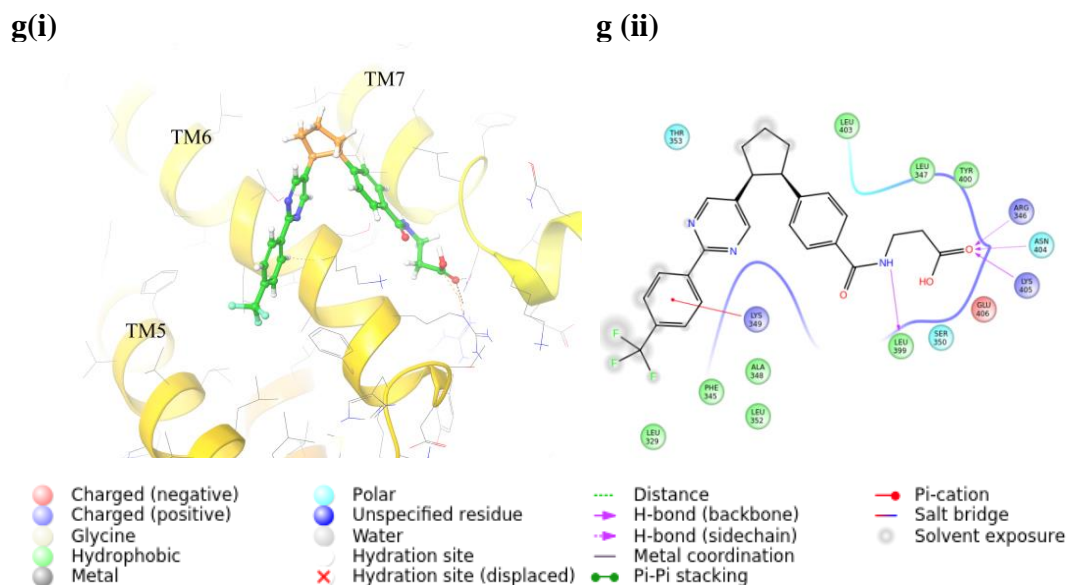


Figure 3.5: Binding pose of lowest-energy conformation of inhibitors bound to glucagon receptor and its 2D ligand interaction diagrams are shown. **a (i- ii)**, **b (i- ii)**, **c (i- ii)**, **d (i- ii)**, **e (i- ii)**, **f (i- ii)** and **g (i- ii)** corresponds to Compound 1, Compound 12, Compound 20, Compound 32, Compound 40, Compound 50 and Compound 58 respectively.

The docking results were validated by re-docking the co-crystallized ligand NNC-0640 to the binding site of glucagon class B Gprotein-coupled receptor (PDB ID: 5XEZ). The Root Mean Square Deviation (RMSD) value of re-docked and X-ray crystal structure of ligand was calculated. The docking pose and the interactions obtained after re-docking showed good agreement with the literature (Zhang et al. 2017) with an RMSD value of 0.82Å (Kramer et al. 1999) (**Figure 3.6**).

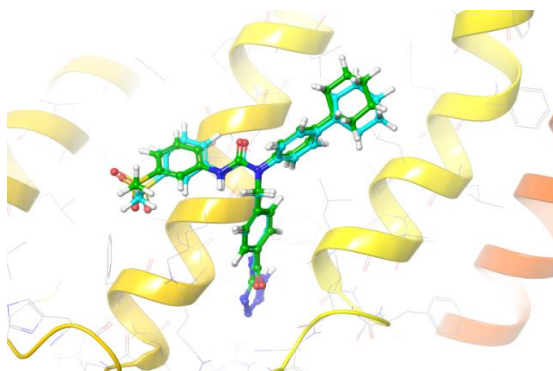


Figure 3.6: Overlay of re-docked ligand (Green) with its crystal structure conformation (Cyan) (RMSD: 0.82Å).

3.6 FRONTIER MOLECULAR ORBITAL ANALYSIS

Single point energy calculations (SPE) using Self Consistent Field (SCF) approach (Blinder 1965) were performed using Gaussian09 package (Frisch et al. 2009) to explain antagonist bound receptor in electronic level. The structures were optimized using B3LYP level (Becke 1988, 1993), 6-31G(d,p) basis set (Lee et al. 1988) and then energies were determined using SPE calculations. Three ligands (active, moderately active and inactive) from each class were chosen for energy calculations. The electronic interactions of the molecules play an important role in its pharmacological effects. The position of HOMO-LUMO orbital is responsible for the electron transfer in a chemical reaction and the energy gap value represents its chemical reactivity. The electronic properties of active, moderately active and inactive compounds from each class were shown in **Table 3.2**. The value of HOMO ranges from -5.495 to -6.620eV, LUMO ranges from -1.045 to -1.720eV and the energy gap ranges from 4.076 to 5.368eV. The dipole moment of the ligands ranges from 1.7147 to 9.5281 Debye. The low energy gap of 4.076eV for the potent ligand (Compound 20) among all the classes indicates the high chemical reactivity and low kinetic stability. The lowest potent ligand (Compound 1) among all the classes has a higher energy gap of 5.368eV, indicates low reactivity and comparatively high stability among inhibitors.

Table 3.2: Summary of electronic properties of selected ligands

Compound		E _{HOMO} (eV)	E _{LUMO} (eV)	E _{Gap} (eV)	Dipole (Debye)
Class I	1	-6.428	-1.060	5.368	3.3852
	8	-6.074	-1.045	5.029	3.4788
	12	-6.478	-1.381	5.097	3.3376
Class II	13	-6.287	-1.332	4.955	4.9411
	17	-5.495	-1.187	4.308	4.6355
	20	-5.684	-1.608	4.076	7.1049
Class III	21	-6.592	-1.687	4.905	3.5415
	27	-6.215	-1.395	4.820	7.7252
	33	-5.776	-1.259	4.517	5.6430
Class IV	34	-6.092	-1.263	4.829	1.7147
	36	-6.037	-1.280	4.757	2.8992
	40	-6.158	-1.457	4.701	9.5281

Class V	41	-6.423	-1.130	5.293	3.9758
	49	-6.594	-1.237	5.357	5.9272
	50	-6.166	-1.659	4.507	5.3732
Class VI	54	-6.309	-1.068	5.241	6.3752
	55	-6.525	-1.378	5.147	7.5089
	58	-6.620	-1.720	4.900	8.6880

The figure given in **Appendix V** shows the position of HOMO-LUMO orbitals of selected ligands in each class based on its potency. It was observed that the substitution of various groups at the hydrophobic part of the inhibitors alters the topology of HOMO-LUMO orbitals in molecules. Majority of the ligand showed a well-defined separation in the position of HOMO-LUMO orbitals indicating that the energies are localized on different parts of the molecule. Among Class I inhibitors, HOMO covers mainly the polar part while LUMO covers R2, R3 ring of Compound 12. We found HOMO electron density on the H1, R2 and R3 region and LUMO density over R1 and polar end for inactive Compound 1. This can be attributed due to the presence of electronegative *p*-CF₃ group at R3 ring of Compound 12 which shifts the electron density towards the polar region of molecule. The presence of *o*-CH₃ group at the R3 ring of Compound 1 increases the electron density over hydrophobic region. A low energy gap of 5.097 eV shows that Compound 12 is chemically reactive among Class I compound. The HOMO orbitals were located mainly on the polar R1 region and cyclobutyl ring (H1) of Compound 20 while LUMO orbitals were located on R2, R3 ring of hydrophobic region. The compound has a low energy gap value of 4.076 eV. Inactive Compound 13 found to have HOMO over hydrophobic (R2 and R3) region and LUMO over polar R1 region. The active inhibitors among Class III (Compound 33) and Class IV (Compound 40) compounds found to have LUMO orbital over polar R1 region and HOMO orbital mainly over hydrophobic R2 and R3 region of the molecule. The HOMO-LUMO energy gap of highly active Compound 33 and Compound 40 was found to be 4.517 and 4.701eV respectively. The active inhibitor among Class V compounds (Compound 50) showed HOMO electron density near hydrophobic R3 and H2 region and LUMO electron density over R2, R3 and H2 region. The active inhibitor of Class VI compounds (Compound 58) showed HOMO in the region of R1, R2, R3 and H1; while LUMO electron densities at the R2 and R3

region only. The energy gap value for Compound 50 in Class V and Compound 58 in Class VI were found to be 4.507 and 4.900eV respectively indicating high chemical reactivity among their respective classes. Therefore, from the above discussion it is clear that presence of HOMO electron density near the polar part, mainly over the R1 region and the hydrophobic H1 region contributes to higher potency of the inhibitor (Compound 12, Compound 20 and Compound 58). The LUMO electron density was found near R2 and R3 region for such cases. This can be further clarified from the HOMO-LUMO analysis of the interacting amino acid of GCGR allosteric site.

Finally, the HOMO-LUMO electron density over the interacting amino acid residues at the allosteric pocket was determined to predict the mechanism of ligand binding (**Figure 3.7**). The N-terminals and C-terminals of the amino acid residues were capped with N-acetyl group and N-methyl amide group respectively using Protein Preparation Wizard: Maestro (Madhavi Sastry et al. 2013). The SPE calculation for the amino acid residues Leu329, Phe345, Arg346, Lys349, Ser350, Leu352, Thr353, Leu399, Asn404 and Lys405 were performed using B3LYP level and 6-31G(d,p) basis set. From the literature it is known that, HOMO orbitals of ligand interacts with the LUMO orbitals of amino acid residues at the binding site (Pang et al. 2012). Similarly, HOMO orbitals of binding site residues interact with the LUMO orbitals of ligand during complex formation (Correa-Basurto et al. 2012b). The LUMO density over hydrophobic part of ligands are responsible for its interactions with the amino acid residues Leu329, Phe345 and Ser350 having HOMO density. The ligands having HOMO density over the polar end of the molecules are highly potent due to its ability to form stable hydrogen bonding interactions with Lys405, Leu399, Ser350 and Arg346 having LUMO density. A well-separated HOMO and LUMO electron density over the hydrophobic part and polar end respectively of the ligands help in stabilizing the formed complex. The HOMO over the cyclopentyl side chain and the polar part of most potent Compound 20 interacts well with LUMO of amino acid residues Leu399 and Ser350. This interaction helps to orient the ligand to form a stable hydrogen bonding with the carbonyl oxygen of Leu399 (C=O---H-N). The amino acid residues Arg346, Asn404, Lys405 form hydrogen bond between the carbonyl oxygen of ligand (N-H---C=O) and Leu399

forms hydrogen bond with the amino group of ligand (C=O---H-N). This study confirms the importance of presence of HOMO orbital near the polar part and LUMO near the hydrophobic region of the inhibitor to form stable protein- ligand complex.

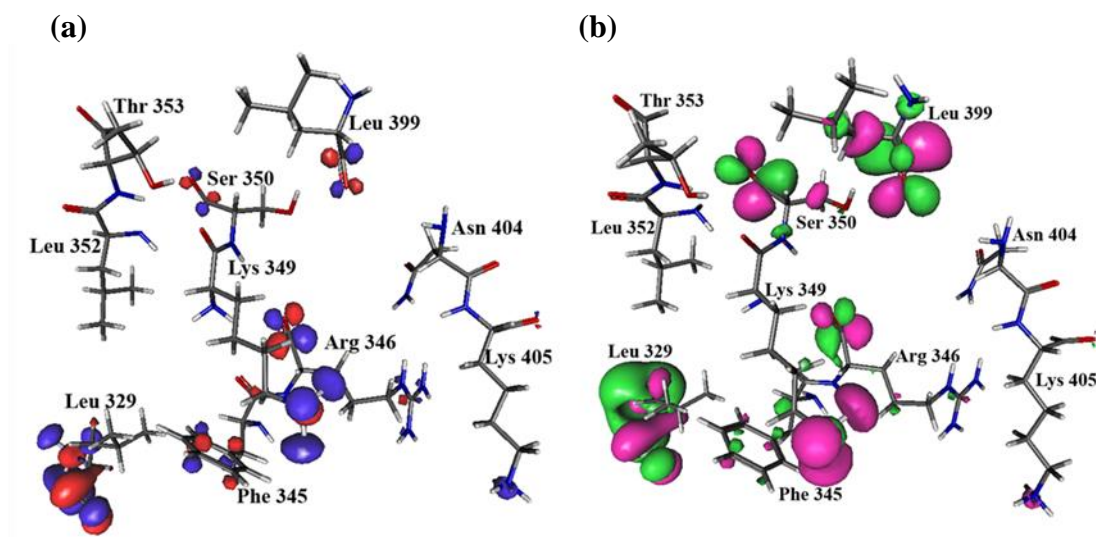


Figure 3.7: Position of (a) HOMO (b) LUMO regions of interacting amino acid residues at the allosteric pocket of GCGR.

3.7 MM/GBSA CALCULATION

The binding free energy of the docked ligands inside the trans-membrane binding pocket of the complexes was calculated by employing Molecular-Mechanics/Generalized Born Surface Area (MM/GBSA) approach (Hou et al. 2011), incorporating OPLS_2005 force field (Jorgensen et al. 1996). An implicit membrane was placed surrounding the transmembrane domain of GCGR using default options of Prime and local optimization sampling algorithm. Simulations were carried out using VSGB solvation model (Li et al. 2011b) (dielectric constant, $\epsilon=80$) with input ligand partial charges.

The average free energy of binding and the corresponding energy components of the bound antagonists towards the trans-membrane allosteric pocket of GCGR were calculated from the simulated trajectories performed in implicit solvent. From the table given in **Appendix VI**, It is found that the binding free energy of all the antagonists considered in present study ranges from -34.675 kcal/mol to -64.18 kcal/mol. Compound 20 ($pK_i=8.046$) displayed higher binding energy of -63.475

kcal/mol towards GCGR whereas Compound 1 ($pK_i=4.833$) found to have binding energy of -41.097 kcal/mol. The binding free energy decomposition of individual inhibitors was carried out according to equation 2.37, 2.38 and 2.39 and illustrated in **Appendix VI**. The van der Waals energy terms ($\Delta G_{bind-vdw}$) and non-polar solvation energy ($\Delta G_{bind-SA}$) term for highest potent Compound 20 is found to be -50.516 kcal/mol and -16.853 kcal/mol respectively. Similarly, the $\Delta G_{bind-vdw}$ and $\Delta G_{bind-SA}$ energy terms found to favour the strong binding of all active compounds. The major contribution of hydrophobic stabilization energy indicates the importance of benzene rings and hydrophobic residues located at region C of the inhibitors. The least potent Compound 1 displayed remarkable decrease in van der Waals energy component ($\Delta G_{bind-vdw} = -36.623$ kcal/mol) in comparison to other highly active inhibitors. Since the antagonists are buried inside the membrane bilayer the van der Waals and hydrophobic solvation energy terms are found to be dominating rather than electrostatic solvation energy terms. Similarly, the covalent energy terms are also disfavoring the binding of inhibitors towards GCGR. The energy terms due to H-bond formation of all the inhibitors ranges from -0.002 kcal/mol to -1.979 kcal/mol which indicates the small contribution of electrostatic interactions for stable inhibitor binding at 5XEZ allosteric pocket. Therefore, the van der Waals ($\Delta G_{bind-vdw}$) and non-polar solvation energy ($\Delta G_{bind-SA}$) terms seems to be key contributing factor for thermodynamically stable binding of active inhibitors at the 5XEZ allosteric pocket.

3.8 MOLECULAR DYNAMICS SIMULATION

Atomistic molecular dynamics simulation was employed to confirm the stability of highly active antagonist (Compound 20) at the allosteric pocket located at the trans-membrane domain of GCGR. The MD simulation of complex 20 with GCGR (PDB ID: 5XEZ) was carried out with OPLS_2005 force field (Jorgensen et al. 1996) in explicit solvent SPC (simple point-charge) water model (Jorgensen et al. 1983) using Desmond software. POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) membranes was properly placed by defining trans-membrane residues from M137 to W418 (Zhang et al. 2017). The entire system was solvated in a periodic orthorhombic box with 20\AA buffer region between protein atoms and box sides to fill with water. The system was neutralized by adding 29 Cl^- ions. The system was minimized to a

gradient threshold of $25 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ using Steepest Descent algorithm and the iteration steps during minimization were kept as 2000 until a convergence threshold of 1.0 kcal/mol/\AA was attained. For long-range electrostatic interactions, smooth Particle Mesh Ewald (PME) method (Essmann et al. 1995) was used with a tolerance of 1×10^{-9} and for short-range electrostatic interactions a cut-off radius of 9.0 \AA was applied. RESPA (Reversible Reference System Propagator Algorithm) integrator algorithm (Tuckerman et al. 1992, 1991) was applied with time steps of 2 fs for bonded, 2 fs for ‘near’ non-bonded and 6 fs for ‘far’ non-bonded interactions. A 5 ns MD run with NVT ensemble was carried out to equilibrate the system at 300K with a time step of 2 fs. Noose-Hoover thermostat (Martyna et al. 1992) was chosen to maintain the system temperature and thermostat relaxation time was kept at the interval of 200 ps, with a time step of 2 fs. Next, NPT equilibration was performed for 5 ns with a time step of 2 fs at 300K and 1 atm using Noose-Hoover thermostat (Martyna et al. 1992) (thermostat relaxation time =200 ps) and Martyna-Tobias-Klein barostat (Martyna et al. 1994) (barostat relaxation time =200 ps). Lastly, a 100 ns production MD (removing restrain on solute heavy atom) run was carried out using NPT ensemble releasing the restrain. The simulations were replicated 5 times to ensure the reliability of the results.

The root-mean-square-deviation (RMSD) profiles of the C_{α} , backbone, side chain and heavy atoms of 5XEZ for simulation with most stable ligand binding are illustrated in **Figure 3.8(a)**. The RMSD value of the protein C_{α} was found to increase up to a value of 6.2 \AA with respect to its starting coordinate ($t=0$) for first 10 ns and stabilize around an average value of 5.748 \AA for rest of the MD trajectories. The average RMSD of backbone, side chains and heavy atoms are found to be 5.745 \AA , 6.657 \AA and 6.179 \AA respectively which indicate significant change in protein backbone compared to its crystal structure. It is evident from the RMSD of Compound 20 that the movement of ligand copes well with the movement of amino acid residues at protein allosteric site (**Figure 3.8(a)**). The average RMSD values of C_{α} , backbone, side chains and heavy atoms for 5 independent MD simulations are found to be $6.403 \text{ \AA} \pm 0.65$, $6.393 \text{ \AA} \pm 0.64$, $7.211 \text{ \AA} \pm 0.58$ and $6.785 \text{ \AA} \pm 0.62$ respectively from the table in **Appendix VII**. Further, the root-mean-square-

fluctuation (RMSF) of the backbone at the allosteric site of 5XEZ is found to be in the range of 2.548- 4.561Å (**Figure 3.8(b)**) which indicates lower degree of flexibility in that region. It is clear from the above discussion that ligand movement was stable during the simulation. It is evident from **Figure 3.8(b)** that residues stretches including His44-Arg60, Asp70-Ala77, Lys98-Gly112, Trp415-Arg419, Arg444-Pro454 have high fluctuations and reside away from the trans-membrane allosteric site.

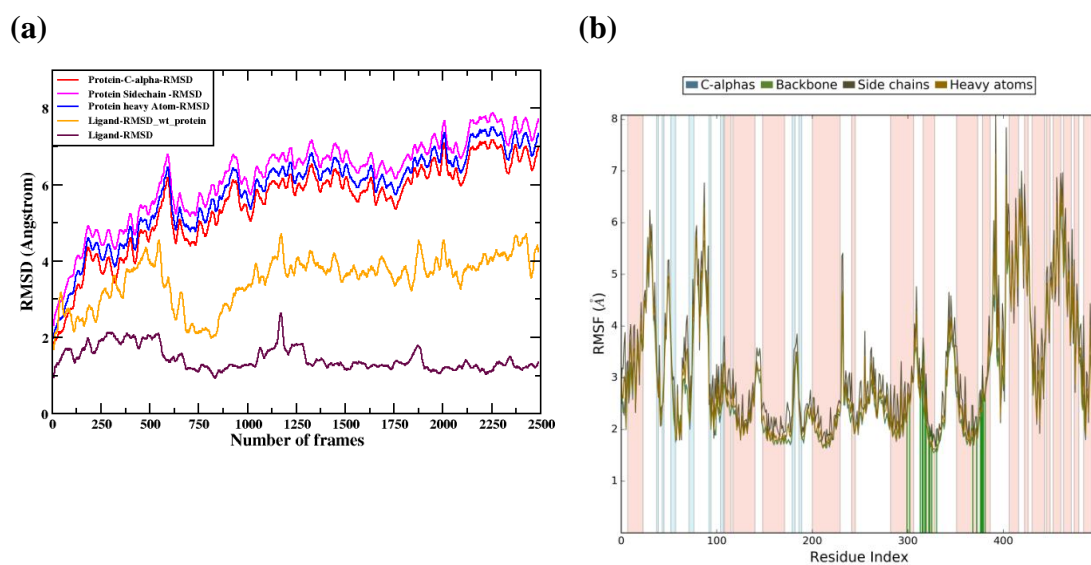


Figure 3.8: (a) Time-line representation of RMSD profile of C_{α} , backbone and heavy atoms of 5XEZ with respect to its initial co-ordinate. The RMSD of compound 20 with respect to protein backbone and its own starting structure was illustrated in red and pink colour respectively. (b) RMSF profile of 5XEZ protein indicating degree of fluctuation during simulation. The green line projecting on x-axis represents ligand contact during the simulation.

The key non-bonded interactions between Compound 20 and 5XEZ during 100 ns MD simulation are illustrated in **Figure 3.9**. It is clear from **Figure 3.9 (a & b)** that non bonded interactions are mainly present in the region of Arg346-Leu352 and Leu399-Glu406, whereas the region Val363-Lys381 found to be more fluctuating (**Figure 3.8**). Hence, no interactions are found between these regions. It is evident from **Figure 3.9 (a & b)** that Lys349 and Leu399 have major contact with the ligand throughout the simulation and are probably responsible for stabilization of Compound

20 at 5XEZ catalytic pocket. However, Arg346, Leu352, Leu395 and Leu403 residues were found to have less interaction with the ligand throughout the MD trajectory.

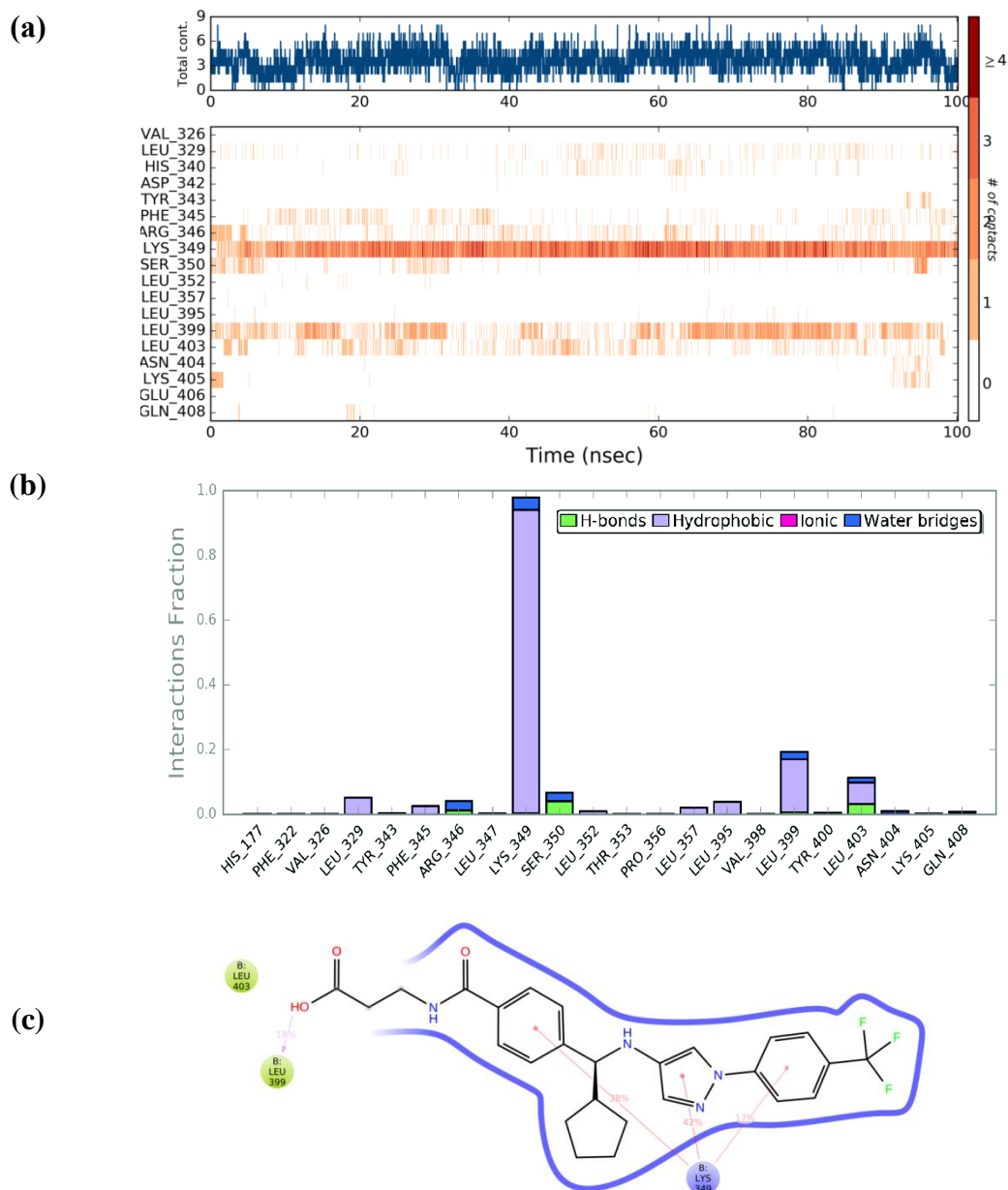


Figure 3.9: (a) Timeline representation of protein- ligand interactions, (b) Histogram showing possible interaction of Compound 20 with 5XEZ residues throughout the simulation and (c) 2D diagram of protein- ligand contacts.

It is found that hydrophobic and hydrogen bonding interactions are major contributing factor for stabilizing Compound 20 at the trans-membrane allosteric pocket of 5XEZ which is in accordance with our MM/GBSA result. Lys 349 found to

exhibit π -cation interaction with the ligand benzene ring and pyrazole ring for 97% of the MD trajectory (**Figure 3.9(c)**). Among the four hydrogen bond predicted by Autodock, only one is found to be preserved during MD simulation. The carbonyl oxygen of Leu399 accepts a hydrogen bond with hydroxyl hydrogen (region A) of the ligand for 18% of the MD trajectory. The number of hydrogen bonds between Compound 20 and 5XEZ throughout the trajectory is found to be 1 (**Figure 3.10(a)**). The average number of water mediated hydrogen bond with Compound 20 is found to be 1.5 (**Figure 3.10(b)**). This further adds to the stability of the Compound 20.

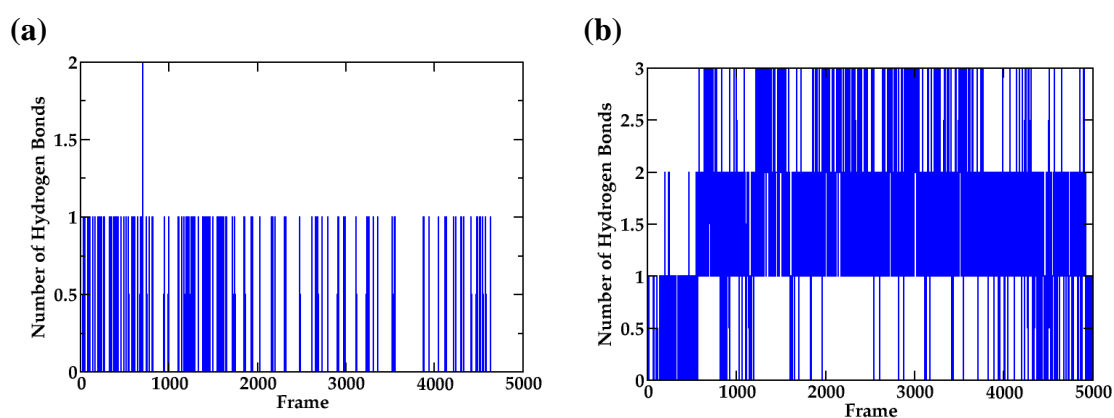


Figure 3.10: Number of hydrogen bonds formed between (a) protein and ligand (b) ligand and water throughout 100 ns.

The snapshots of the simulation at each 5 ns interval are illustrated in **Figure 3.11** to further confirm ligand stability. Region C of Compound 20 found to move away from its initial position around 5 ns of the simulation and further stabilized by π -cation interaction formed by Lys349. The π -cation interaction between Compound 20 and Lys349 found to be present in each snapshot displayed in **Figure 3.11** which confirms its key role in anchoring the inhibitor at 5XEZ catalytic pocket. A well-defined, water mediated hydrogen bond network is observed between the hydrophilic part of Compound 20 and TM7 amino acid residues during the last phase of the simulation. Those water molecules are probably responsible for stabilizing the hydrophilic part of Compound 20.

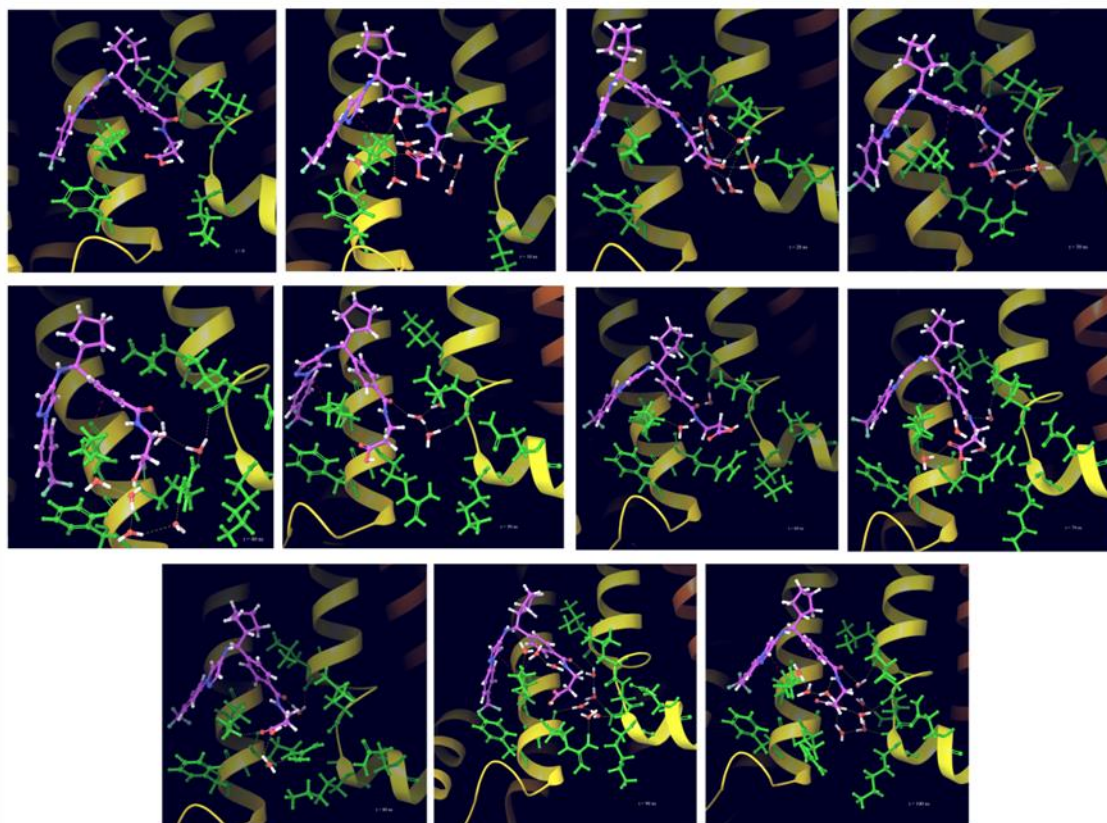


Figure 3.11: Snapshots at 0 ns, 10 ns, 20 ns, 30 ns, 40 ns, 50 ns, 60 ns, 70 ns, 80 ns, 90 ns and 100 ns of MD trajectory are illustrated. The interacting amino-acid residues are coloured in green, oxygen of water molecules are represented in red. The π -cation interaction and hydrogen bonding interactions are indicated by red and yellow respectively.

A low RMSD value ($1.571\text{\AA} \pm 0.57$) of ligand indicates a less conformational change with respect to the initial conformation. The gyration radius (R_g , measures the extendedness of a ligand) found to stabilize after 5 ns of the simulation with an average value of $5.612\text{\AA} \pm 0.44$. The solvent accessible surface area (SASA), polar surface area (PSA) and molecular surface area (MolSA) of ligand for most stable ligand binding are found to be in the range of $277.069\text{-}449.461\text{\AA}^2$, $140.24\text{-}180.367\text{\AA}^2$ and $439.852\text{-}462.211\text{\AA}^2$ respectively which supports the stabilization of ligand in binding site of protein during simulation (**Figure 3.12**).

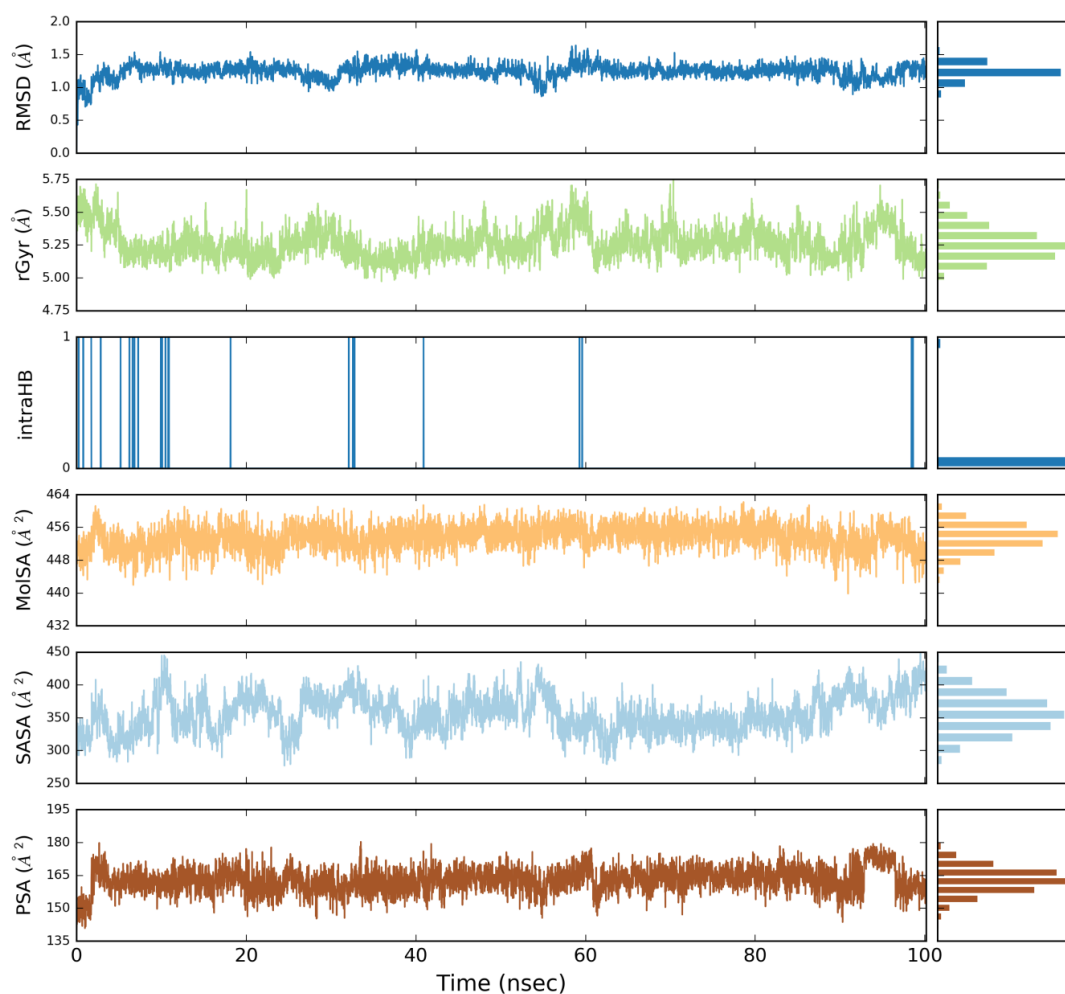


Figure 3.12: Ligand properties: RMSD (\AA) with respect to reference conformation, radius of gyration (Rg, \AA), number of internal hydrogen bonds (intra HB), molecular surface area (\AA^2), solvent accessible surface area (\AA^2) and polar surface area (\AA^2).

3.9 ADME/ TOXICITY PREDICTION

The absorption, distribution, metabolism, excretion, toxicity properties and different physically significant descriptors of the top scored inhibitors of GCGR were predicted using Qikprop module (Schrödinger Release 2018-4: QikProp, Schrödinger, LLC, NY, 2018). The reliability of such predictions has already been reported and benchmarked (Ioakimidis et al. 2008). Qikprop program employs the method of Jorgensen to predict the pharmacokinetics properties of drug-like molecules (Jorgensen and Duffy 2002). The drug-likeness of the top scored GCGR antagonists was also evaluated using Lipinski's rule of 5 (Lipinski 2004). The permissible ranges

of pharmacokinetic properties and the predicted ADME/tox properties of 14 top-scored ligands are tabulated as **Appendix VIII**. All the calculated properties were found to be in their permissible range and hence confirming their drug like properties.

The bioavailability of top scored molecules was determined through polar surface area analysis (PSA) which measures the cell wall permeability or transport through membranes. The glucagon receptor antagonists possess PSA value in the range of 81.913 to 124.622Å² which shows good oral availability. The number of non-hindered rotatable bonds was found to vary from 4 to 9. The calculated molecular weights of top scored hits were found to be in the range of 474.482 to 528.453 which is acceptable for orally consumable drugs. The value of QPlogP_{o/w} was found to be in the range of 3.387 to 7.403 and most of the top scored compounds are in the permissible range of -2.0 to 6.5. Compound 32 and Compound 57 exhibited QPlogP_{o/w} value of 7.073 and 7.403 respectively. The high value of logP_{o/w} indicates higher lipophilic character of those molecules. The aqueous solubility (QPlogS) was found to be in range of -9.2 to -5.5 for all the top scored hits. The highest active Compound 20 was found to exhibit the QPlogS value of -6.130 which is under the permissible zone. The number of hydrogen bond acceptors (4.5-7.5) and hydrogen bond donors (1-2) were also found to vary in their permissible value. QPlogBB values found to be in acceptable range of -1.344 to -0.3. The binding of molecule to human serum albumin (QPlogK_{hsa}), and the present of human oral absorption were found to be in threshold limit. It is believed that orally active compound should not have more than two violations of Lipinski's rule which is in accordance with our result. Some of the molecules showed deviations of two parameters of Lipinski's rule of 5 due to its higher hydrophobic character. From the above discussion it is evident that top scored compounds obtained from the present study have good oral bioavailability.

3.10 CONCLUSION

3D-QSAR, molecular docking, DFT calculation, MM/GBSA, molecular dynamics simulation and ADME/Toxicity studies of β-alanine analogues were performed to investigate the effect of structural determinants responsible for GCGR antagonism. The developed 3D-QSAR model gave R² value of 0.9981 and Q² value of 0.8253,

indicating excellent consistency and internal predictability of the model. Contour plots obtained from 3D-QSAR model revealed the position of hydrophobic/non-polar substituents contributing to increase the inhibitory activity. Electron withdrawing groups present at R2 and R3 rings are found to have unfavorable contribution to the potency of inhibitors. Further, docking study predicted the binding pose of antagonist at the binding site of GCGR. The docking study suggested that the polar region of the ligand forms hydrogen bonding network with Arg346, Ser350, Leu399, Asn404 and Lys405 amino acid residues. The presence of $-CF_3$ group at R3 ring and cyclopentyl group at H2 position stabilize Compound 20 at hydrophobic region of TM5 which helps to make hydrogen bonds between the polar part of the ligand and TM7 region. The presence of *o*-methyl group at R3 ring of Compound 1 found to destabilize the ligand toward TM5. It is evident from the study that the alignment of hydrophobic region towards TM5 facilitates the proper orientation required for GCGR allosteric inhibition. HOMO-LUMO orbital analysis described the interaction mechanism of ligand with the protein at quantum level. The presence of HOMO near hydrophobic H1 region and polar R1 region gives favourable interactions with amino acids Lys405, Leu399, Ser350, Arg346 having LUMO density. Similarly, the presence of LUMO near hydrophobic R2 and R3 region of the ligand gives favourable interactions with amino acids Leu329, Phe345, Ser350 having HOMO density. MM/GBSA calculation displayed that van der Waals and non-polar solvation energy terms contribute mostly for stabilizing the antagonist binding to GCGR. The binding energy of highly active Compound 20 was found to be -63.475 kcal/mol. Further, stability of Compound 20 at 5XEZ allosteric pocket was confirmed by 100 ns atomistic MD simulation. Simulation revealed that π -cation interaction of Lys349 and hydrogen bonding of Leu399 have crucial role in stabilizing Compound 20 under motion. Water molecules near the hydrophilic part of the ligand found to have hydrogen bonding with the ligand, thereby stabilizing the protein-ligand complex. Lastly, ADME/tox calculation of top scored compounds obtained from present study assured their safe administration in human body. The outcomes of the present study provide insightful information regarding the design of novel glucagon receptor antagonists to treat Type II Diabetes Mellitus.

CHAPTER 4

THEORETICAL INSIGHTS INTO MOLECULAR MECHANISM AND ENERGY CRITERIA OF PARP-2 ENZYME INHIBITION BY BENZIMIDAZOLE ANALOGUES

The emergence of PARP inhibitors targeting a class of PARP enzymes has gained a great interest in cancer therapy. Majority of the PARP inhibitors are not isoform-selective which may cause unwanted off-target effects. In the present study, we explore the molecular mechanism and energy requirements for PARP-2 inhibition using various in silico methods.

4.1 BACKGROUND

The genomic stability of an organism is vulnerable to both biotic and abiotic stresses for its entire biological life cycle (Ciccia and Elledge 2010). It requires an appropriate and rapid response to external environmental stimuli for maintaining genomic stability and cell survival. The genomic integrity is managed by a series of sophisticated mechanisms such as DNA repair signalling pathways, chromatin reorganisation and protein modification (Wei and Yu 2016). The defects in these processes result in genomic instability, which is a significant factor in tumorigenesis. In humans, this modification is catalyzed by poly (ADP-ribose) polymerase (PARP) family of enzyme which consists of 17 members (Hottiger 2015). PARP-1, PARP-2 and PARP-3 are the only members which are known for DNA damage dependent response, DNA metabolism and chromatin architecture. PARP homologues have been found in viruses, prokaryotes (bacteria) and eukaryotes (fungi, plants and animals) (Perina et al. 2014).

The enzymes of PARP family use nicotinamide adenine dinucleotide (NAD⁺) molecules as the substrate to produce and transfer ADP-ribose polymers onto the target proteins (Schreiber et al. 2006), involved in various processes such as repairing of DNA damage sites, chromatin structure modification, control of transcription and cell division (Kraus 2015). The PARP inhibitors competitively inhibit the enzyme by

replacing NAD⁺ substrates at nicotinamide-binding pocket. In co-crystal structures, it is found that the inhibitors bind to the catalytic pocket of PARPs by hydrogen bonding and π - π interactions (Aoyagi-Scharber et al. 2014). Majority of the PARP inhibitors contain carboxamide group which is similar to nicotinamide group to mimic the binding pose at catalytic site (Chen et al. 2018; Costantino et al. 2001; Reddy et al. 2018; Zhou et al. 2017). Since PARP-inhibitors mimic the binding mode of nicotinamide, benzimidazole moiety which are planar and aromatic can be a potent competitive inhibitor to nicotinamide. Benzimidazole is a heterocyclic aromatic compound found in natural compounds. The derivatives of benzimidazole were found to have various therapeutic applications in cardiovascular disease, ophthalmology and neurology. It is also used as antitumor, antihistamine, antiparasitic, analgesics, antiviral and antifungal medications (Yerragunta et al. 2014). Recently, the crystal structure of murine PARP-2 was solved (Oliver 2004), which revealed that the binding mode of inhibitors at the catalytic domain of PARP-2 is similar to that of PARP-1 (Ruf et al. 1996). Even though the structural differences of these enzymes at DNA binding domain and catalytic domain serve as an essential basis to design isoform-selective inhibitors.

4.2 DATASET AND PREPARATION OF 3D STRUCTURES OF LIGAND

A dataset consisting of 32 benzimidazole carboxamide derivatives were selected which shared same enzymatic and cellular assay procedure from recent experimental reports (Chen et al. 2018). The biological activity of the compounds was given by IC₅₀ (half maximal inhibitory concentration) values and reported to have an inhibitory activity spanned from 14 to 382 μ M. The 3D structures of the ligands were constructed using the builder panel in Maestro (Schrödinger Release 2018: Maestro). The geometry of the compounds was optimized by semi-empirical PM3 (Stewart 1989), followed by B3LYP/6-31G(d,p) level (Ditchfield et al. 1971; Lee et al. 1988) using Gaussian09 package (Frisch et al. 2009). The resulted structures were used for further studies. HOMO-LUMO energy values and electronic properties of selected ligands were determined to understand kinetic stability and chemical reactivity of ligands (Bharathi et al. 2016).

4.3 MOLECULAR DOCKING

The co-crystal structure of PARP-1 (PDB ID: 5A00, resolution 2.75 Å) and PARP-2 (PDB ID: 4ZZY, resolution 2.2 Å) enzymes bound with isoindolinone inhibitor was obtained from Protein Data Bank. Prior to the study, the protein structures were prepared by removing co-crystallized ligands and water molecules. The molecular docking was performed using Autodock (v4.2.6) software which considers protein to be rigid during docking (Morris et al. 2009). A 3D grid was created at the binding site of PARP-2 having a dimension of 38 Å × 30 Å × 30 Å with a default spacing of 0.375 Å, using an Auto-Grid algorithm. The grid includes six active site residues namely His428, Gly429, Tyr455, Tyr463, Ser470 and Tyr473. The molecular docking simulations were performed using Lamarckian Genetic Algorithm (Morris et al. 1998). The population size and the number of energy evaluation steps were fixed to 150 and 2.5×10^6 respectively for docking simulations.

The stable binding pose of PARP-2 inhibitors along with their binding energy was predicted using molecular docking simulations. From table shown in **Appendix IX**, C-1 is found to be the most stable PARP 2-inhibitor complex with binding energy of -7.94 kcal/mol and C-32 is least stable with binding energy of -6.09 kcal/mol. Docking studies suggest that the interactions were mainly influenced by hydrogen bonding, π - π stacking interactions and hydrophobic interactions. The ligand interactions mainly resided in the region from Gly429 to Tyr473 which covers the catalytic site of the protein. The interaction of all the ligands with the enzyme mainly comprises of hydrogen bonding with Gly429 and π - π stacking with Tyr473. There are three possibilities for π -stacking interactions between the ligand rings and aromatic amino acid residues. Two rings of the benzimidazole moiety and one ring from the substituted group forms stacking interactions with the tyrosine residue depend on the orientation of ligand at the binding pocket. Detailed analysis of ligand orientation is discussed in the following sections.

The protein-ligand complex of most potent compound, C-1, is stabilized by hydrogen bonding, π - π interaction and hydrophobic interactions (**Figure 4.1(a)**). It is found that tetrahydrothienopyridinyl moiety of ligand makes hydrophobic interaction

with the amino acid residues Ile331, Tyr455 and Met456. This stable hydrophobic interaction can be attributed due to the presence of methyl group on the substituted moiety. The ligand forms two hydrogen bonding interactions. One between the carbonyl oxygen of the ligand and amino group of Gly429 ($C=O_{\text{lig}} \cdots NH_{\text{Gly}}$; 1.70 Å) and the other between amino group of the carboxamide core and carbonyl oxygen of Gly429 ($C=O_{\text{Gly}} \cdots NH_{\text{lig}}$; 2.10 Å). For this complex, the imidazole ring of the ligand shows π - π interaction with Tyr473.

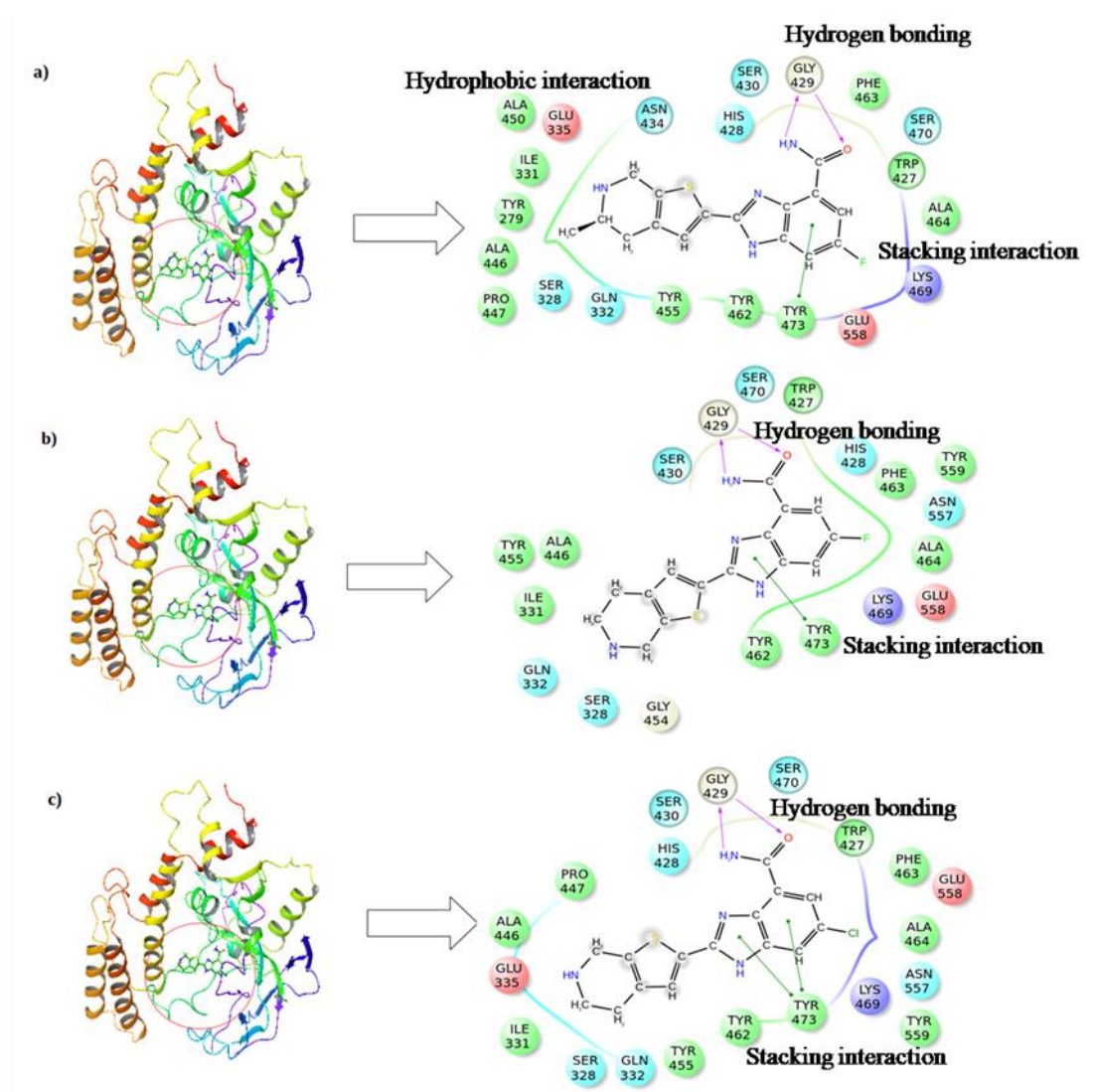


Figure 4.1: Binding pocket of PARP-2 enzyme bound with active, moderate and inactive ligands (a) C-1, (b) C-16 and (c) C-32. Enlarged view of binding pose is shown in 2D-ligand interaction diagram. The hydrophobic, polar, positively-charged, negatively-charged and glycine residues are represented in green, cyan, blue, red and grey colour respectively.

The moderately potent compound, C-16, found to have one strong hydrogen bond interaction ($C=O_{\text{lig}} \cdots NH_{\text{Gly}}$; 1.76 Å) and one π - π stacking interaction from the imidazole ring of ligand with Tyr473 (**Figure 4.1(b)**). Additionally, a weak hydrogen bonding interaction is found between amino group of the carboxamide core and carbonyl oxygen of Gly429 (2.05 Å). The least potent compound, C-32, shows two weak hydrogen bonding interactions: $C=O_{\text{Gly}} \cdots NH_{\text{lig}}$ and $C=O_{\text{lig}} \cdots NH_{\text{Gly}}$ (bond length > 2.25 Å). Also, both the rings of benzimidazole moiety of the ligand form π - π stacking interactions with Tyr473 (**Figure 4.1(c)**). It can be noted here that, in the case of complex of C-16 and C-32, hydrophobic interactions are absent between the substituted group of ligand and amino acid residues at the binding site.

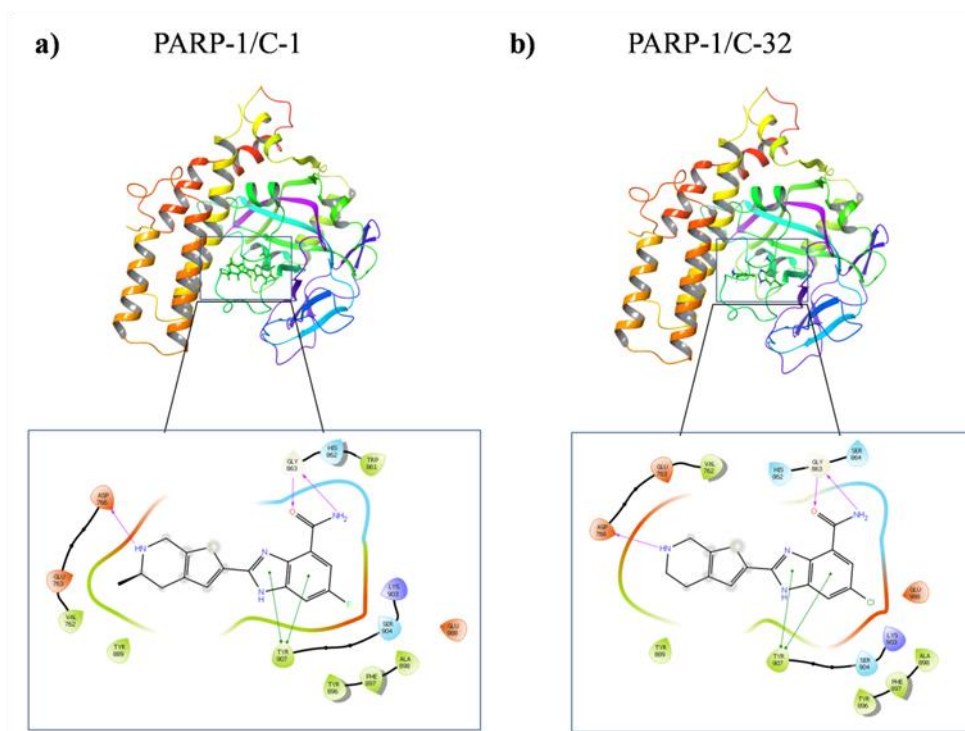


Figure 4.2: Binding pose and ligand interaction diagram of PARP-1 enzyme with ligands (a) C-1 and (b) C-32.

The docking results obtained for PARP-1 enzyme is given in **Figure 4.2**. The docking of selected compounds showed hydrogen bonding (Gly863 and Asp766) and π - π interaction (Tyr907) with similar residues as in case of PARP-2. The docking results were validated by re-docking the co-crystallized ligand isoindolinone to the catalytic domain of PARP-2 protein (PDB ID: 4ZZY). The Root Mean Square

Deviation (RMSD) of re-docked and X-ray crystal structure of ligand was calculated using the same procedure. The docking pose and the interactions obtained after re-docking showed good correlation with the literature with an RMSD value of 0.061 nm (Kramer et al. 1999) (**Figure 4.3**).

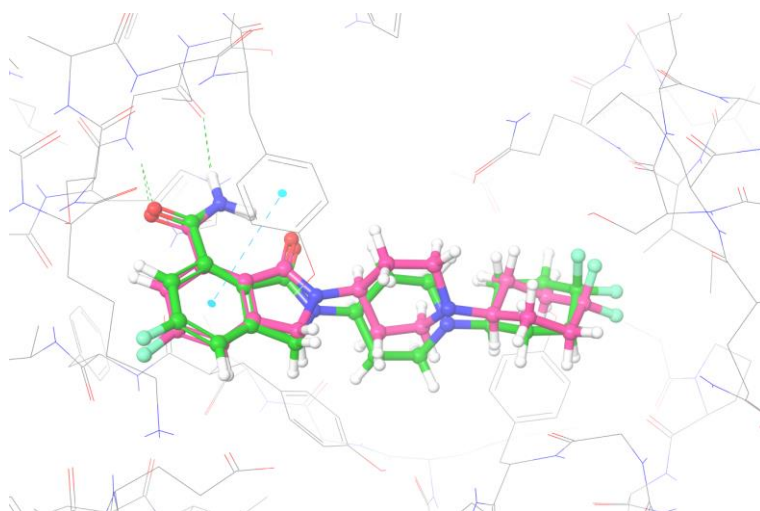


Figure 4.3: Overlay of re-docked ligand (Green) with its crystal structure conformation (Pink) (RMSD: 0.61Å).

4.4 FRONTIER MOLECULAR ORBITAL ANALYSIS AND REACTIVITY DESCRIPTORS

HOMO-LUMO Analysis: Statistically significant parameters such as HOMO-LUMO energy (eV), energy gap (E_g , eV), total energy (SCF, au) and dipole moment (debye) for selected compounds were calculated (**Table 4.1**). The HOMO energy (E_{HOMO}) ranges from -6.164 to -5.677 eV and the LUMO energy (E_{LUMO}) ranges from -2.097 to -1.741 eV. The higher value of E_{HOMO} indicates the tendency of compounds to donate electrons to lower empty orbital of the amino acid residues at the catalytic domain of the protein (Malkhasian and Howlin 2016). The results suggested that C-1 have the highest (-5.677 eV) and C-32 have the lowest (-6.164 eV) values of E_{HOMO} (**Figure 4.4**). Also, the active compound, C-1 possesses the highest value of E_{LUMO} and is found to be -1.741 eV. The HOMO-LUMO energy gap of 3.936 eV for C-1 (active) and 4.408 eV for C-32 (inactive) indicates the comparative high chemical reactivity and kinetic stability for compound 1. This trend of energy values is in good

correlation with the experimental IC₅₀ values. Also, HOMO-LUMO energy gap of selected benzimidazole derivatives are compared with newly designed PARP-2 inhibitors in the following section. The HOMO-LUMO regions are plotted onto the selected molecules to analyze the electron donor and acceptor sites of the molecule (**Figure 4.4**). The HOMO-LUMO energy gap of selected molecules (C-1 < C-8 ~ C-16 < C-21 ~ C-27 < C-32) was found to be in correlation with the biological activity (C-1 > C-8 > C-16 > C-21 > C-27 > C-32). It is found that LUMO electron density is delocalized over the benzimidazole and thiophene rings of the compound. This region is involved in crucial π - π stacking interaction with Tyr473. The HOMO orbitals are localized over the molecule which has a major role in hydrogen bonding interactions with Gly429 at the catalytic domain of the enzyme.

Table 4.1: HOMO-LUMO energy details and total energy of selected ligands

Compound	E _{HOMO} (eV)	E _{LUMO} (eV)	E _g (eV)	Total Energy (kcal/mol)
C-1	-5.677	-1.741	3.936	-885453.323
C-8	-5.721	-1.755	3.966	-860778.543
C-16	-5.770	-1.805	3.965	-860778.543
C-21	-5.993	-1.697	4.296	-870851.388
C-27	-6.032	-1.746	4.286	-872068.984
C-32	-6.164	-1.756	4.408	-1086928.700

Energy Calculation

Terms	Energy (kcal/mol)	ΔG_{bind} (kcal/mol)
PARP-2/C-1		
G _{complex}	-3017852.41	-122.19
G _{C-1}	-885453.32	
G _{residues}	-2132276.90	
PARP-2/C-32		
G _{complex}	-2261960.00	-41.30
G _{C-1}	-1086928.70	
G _{residues}	-1174990.00	

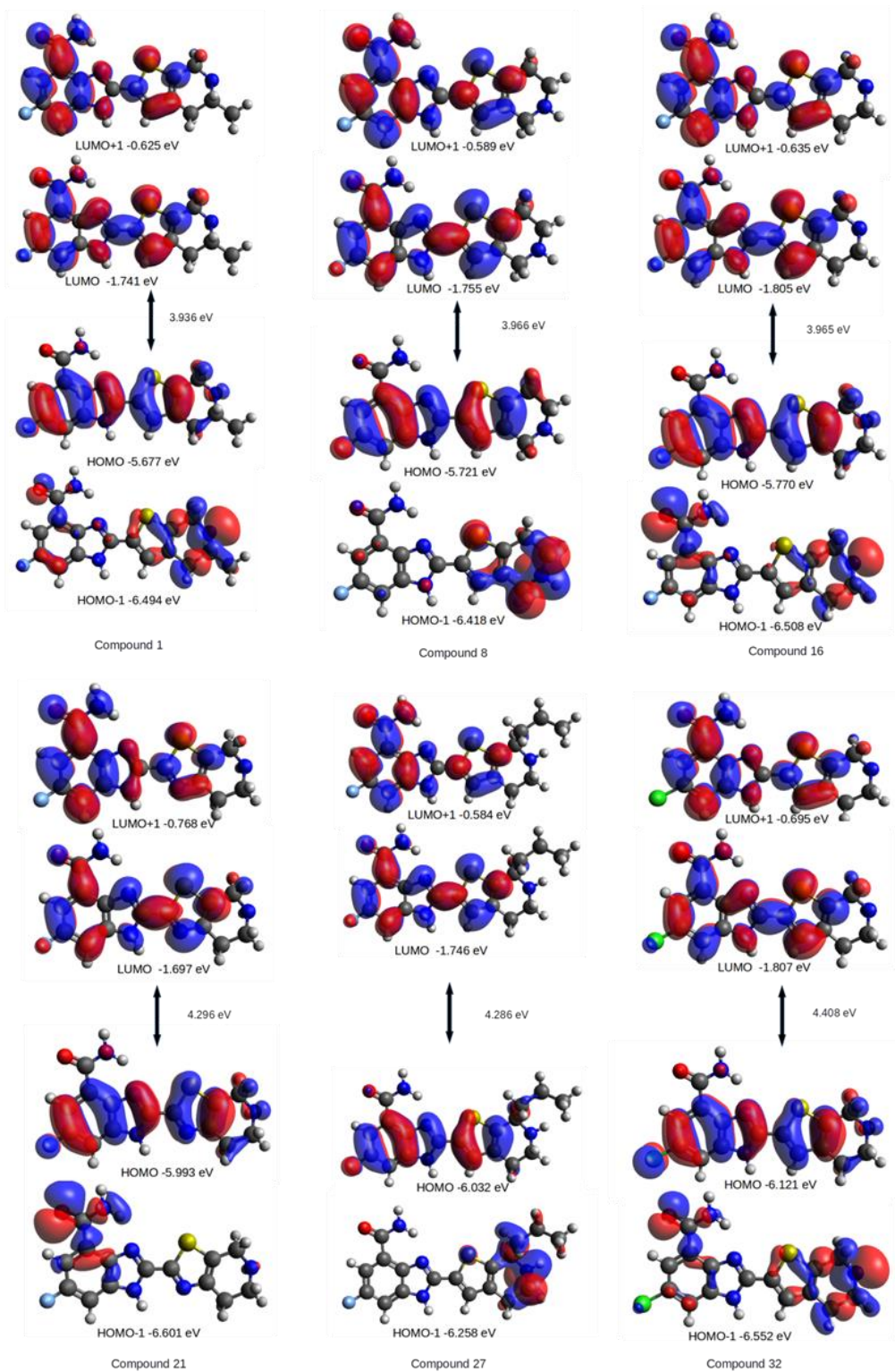


Figure 4.4: Frontier molecular orbital energy values, energy gap and electron cloud density over compound1, compound 8, compound 16, compound 21, compound 27 and compound 32.

Since the HOMO-LUMO positions and the total energy of the ligand changes when bound to the protein; single point energy calculation for the complexes was carried out by selecting a cluster containing interacting amino acids and the ligand using SCF (Self Consistent Field) approach. The total energy of the cluster with C-1 (active) and C-32 (inactive) was found to be -3017852.41 kcal/mol and -2261960.00 kcal/mol respectively which indicate the stability of C-1 at the binding site of enzyme. The value of ΔG_{bind} for C-1 and C-32 is found to be -122.19 kcal/mol and -41.3 kcal/mol respectively which suggest that the compound C-1 binds strongly at the enzyme binding site. The electron cloud over clusters of ligands showed that HOMO density over Tyr473 interacts with LUMO density over the ligands (C-1 and C-32) to form complexes. The presence of HOMO electron density over Ile331, Tyr455, Met456 amino residues and tetrahydrothienopyridinyl moiety of C-1 in cluster indicated the possibility of favourable hydrophobic interactions. These interactions are absent in cluster C-32, which resulted in the lower stability of complex (**Figure 4.5**).

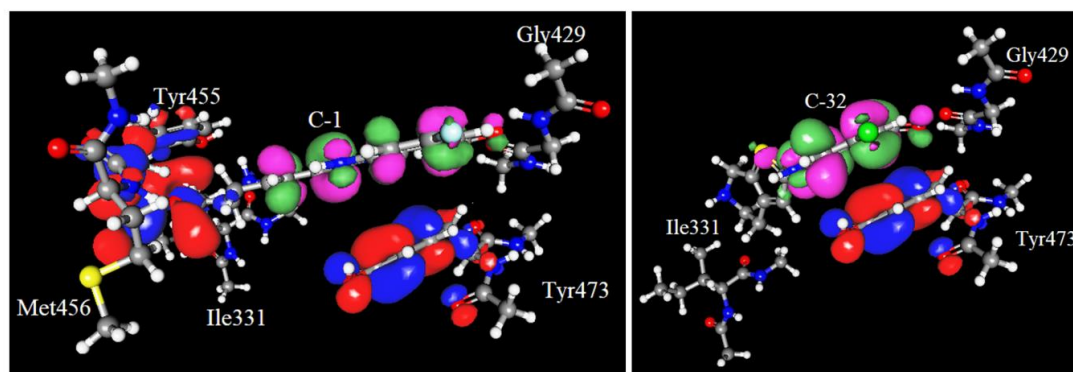


Figure 4.5: The distribution of HOMO-LUMO orbitals over the cluster containing interacting amino acids and ligands (C-1 and C-32). The positive and negative lobes of HOMO orbitals are blue and red in colour respectively. The positive and negative lobes of LUMO orbitals are coloured green and pink respectively.

Reactivity Descriptors: The stabilization and interactions of the compounds with the catalytic site of enzyme can explain by global reactivity descriptors. The descriptors such as dipole moment (debye), electronegativity (χ), chemical potential (μ), hardness (η), softness (s) and electrophilicity index (ω) were calculated (**Table 4.2**). The dipole

moment of the compound ranges from 6.449 to 9.993 Debye which lies in the accepted values for drug molecules (Lien et al. 1982). The dipole moment of C-1 is 9.050 Debye while C-32 shows lowest dipole moment (6.449 Debye). Higher the dipole moment, higher the possibilities of hydrogen bond formation and other non-bonded interactions due to the polar nature of molecule. It is found that C-1 with low value (1.968 eV) of hardness is biologically active than other compounds. The high value (2.204 eV) of hardness for C-32 indicates the stability of the molecule thereby decreasing its reactivity. C-1 showed a low electrophilicity index of 0.927 eV as compared to other molecules which indicates the electron-rich nature or nucleophilicity.

Table 4.2: Global reactivity descriptors calculated from DFT calculations

Cpd	Electronegativity (χ , eV)*	Chemical Potential (μ , eV)**	Hardness (η , eV)#	Softness (s, eV ⁻¹)##	Electrophilicity index (ω , eV)\$	Dipole moment (debye)\$§
C-1	3.709	-3.709	1.968	0.508	0.927	9.050
C-8	3.738	-3.738	1.983	0.504	0.935	9.562
C-16	3.788	-3.788	1.983	0.504	0.947	8.605
C-21	3.845	-3.845	2.148	0.466	0.961	7.623
C-27	3.889	-3.889	2.143	0.467	0.972	9.993
C-32	3.960	-3.960	2.204	0.454	0.990	6.449

* $\chi = -\frac{1}{2}(E_{HOMO} + E_{LUMO})$

** $\mu = \frac{1}{2}(E_{HOMO} + E_{LUMO})$

$\eta = \frac{1}{2}(E_{LUMO} + E_{HOMO})$

$s = \frac{1}{2\eta}$

\$ $\omega = \frac{-(HOMO+LUMO)}{8}$

§§ Permissible range of dipole moment: 1.0 – 12.5

4.5 MOLECULAR DYNAMICS SIMULATION STUDY

Molecular dynamics simulation was performed by Gromacs (Abraham et al. 2015) 2018.4 using SPC/E water model (Hess and van der Vegt 2006) for six biological systems (Table 4.3). The topology was generated using AMBER99SB force field (Hornak et al. 2006). The complex was solvated in a periodic cubic box with distance

between protein and box boundary as 1.0 nm in all directions. The charge of the system was neutralized by adding one Na⁺ ion. The energy minimization was done by steepest descent algorithm with a maximum of 50000 steps until a convergence tolerance of 1000 kJmol⁻¹nm⁻¹. A 10 ns NVT equilibration was carried out at 300 K using V-rescale temperature coupling method (Bussi et al. 2007) with a time step of 2 fs. Next, a 10 ns NPT equilibration was carried out with a time step of 2 fs at 300 K and 1 atm using Nose-Hoover thermostat (Martyna et al. 1992) and Parrinello-Rahman pressure coupling scheme (Parrinello and Rahman 1981) respectively. The long range electrostatic interactions were calculated by particle mesh Ewald method (Essmann et al. 1995) and Fourier spacing of 0.16 nm. The short range van der Waals cut-off was fixed to 1.2 nm. In both cases, LINCS constraints (Hess et al. 1998) were used to restrain the bond involving hydrogen atoms. Finally, a production run of 500 ns was performed by removing restrain to relax the system.

Table 4.3: Summary of MD simulations performed

Protein	Ligand	Number of Water	Number of Atoms	Time (ns) (Production run)
PARP-2	-	22559	73284	500
PARP-2	Active (C-1)	22543	73274	500
PARP-2	Inactive (C-32)	22545	73277	500
PARP-1	-	20379	66542	500
PARP-1	Active (C-1)	20369	66550	500
PARP-1	Inactive (C-32)	20370	66550	500

Stability of Protein-Ligand Complex: The RMSD (root mean square deviation) profile for the complexes is illustrated in **Figure 4.6**. The average values for ligand and backbone of protein are found to be 0.12 nm and 0.21 nm respectively. The RMSD for PARP-2/C-1 is found to be stable after 25 ns of simulation and showed a fluctuation around 150-200 nm. The ligand RMSD converges to 0.12 nm indicating that movement of ligand copes well with the binding site residues of protein. PARP-2 complexes and PARP-1/C-1 are found to be stable throughout the simulation whereas; PARP-1/C-32 complex showed high degree of instability after 150 ns.

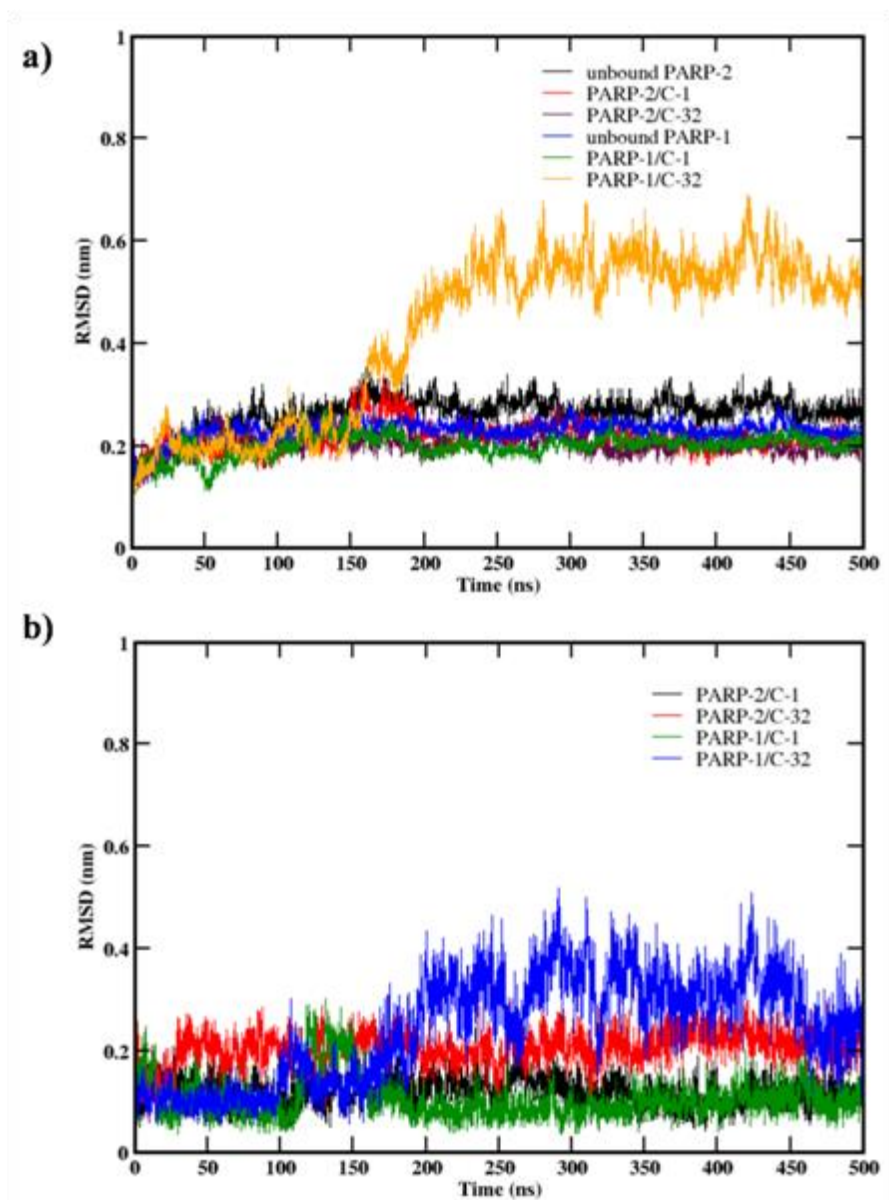


Figure 4.6: (a) RMSD profile of backbone of unbound and bound PARP enzymes over 500 ns trajectory. (b) Time evolution of RMSD of ligand with respect to protein throughout 500 ns trajectory for PARP complexes.

RMSF (root mean square fluctuation) values of protein backbone measures the mobility of amino acid residues. The average RMSF of the binding pocket covering residues from Gly429 to Tyr473 for unbound enzyme, PARP-2/C-1 and PARP-2/C-32 complexes are found to be 0.072 nm, 0.09 nm and 0.10 nm respectively. The RMSF profile of PARP-2 complexes indicates that the binding pocket has fewer fluctuations compared to other regions. The highest value for fluctuations were found

in residues Phe313-Thr317, Glu350-Glu355, Pro392-Ser395 and Leu547-Thr553 which resides away from the catalytic domain of PARP-2 enzyme (**Figure 4.7(a)**). Even though fewer fluctuations were found in binding site residues Gly429-Tyr473, it helps the ligand to form a stable complex with the enzyme. The protein backbones of PARP-1/C-1 complex found to have almost similar fluctuations as that of unbound PARP-1 enzyme. Whereas, the protein backbones of PARP-1/C-32 complex have higher RMSF compared to the unbound enzyme indicating larger fluctuations at binding site of the protein (**Figure 4.7(b)**).

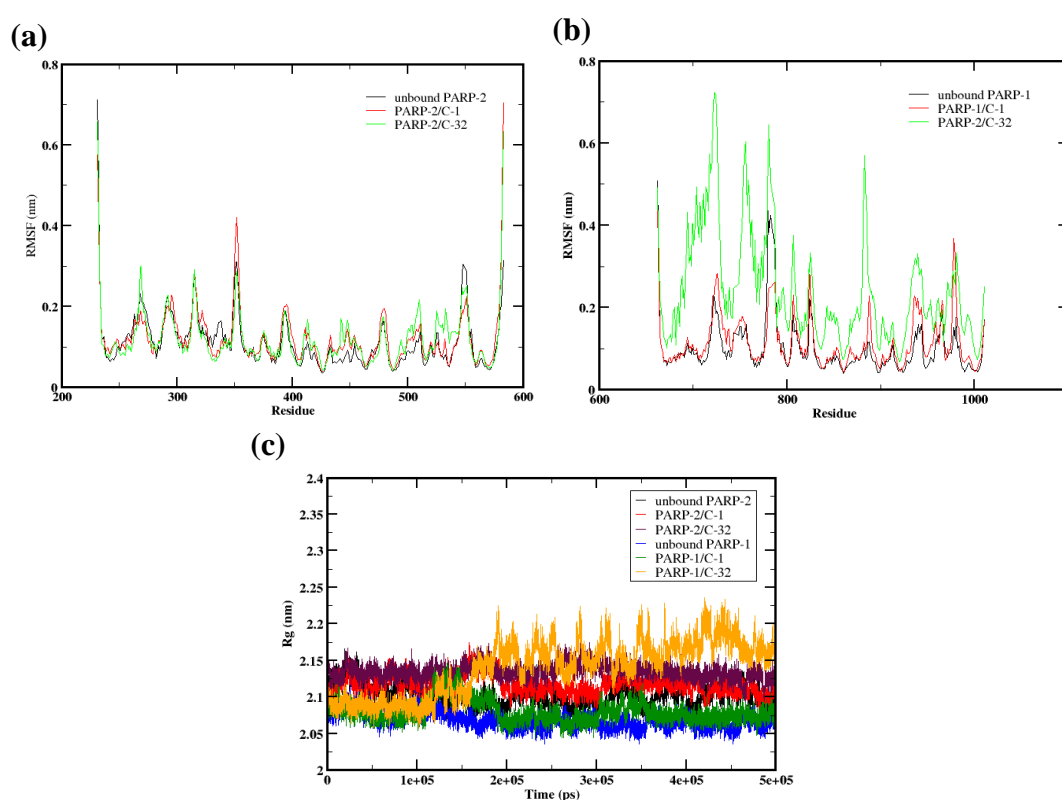


Figure 4.7: Residue-based fluctuations of protein backbone of unbound enzyme and complexes over 500 ns simulation (a) PARP-2 and (b) PARP-1. (c) Radius of gyration of backbone atoms of PARP enzymes with and without ligands (C-1 and C-32).

Radius of gyration (Rg) measures the conformational variation and compactness of protein-ligand complex during the simulation. Rg plots of the unbound enzymes and the four complexes are shown in **Figure 4.7(c)**. The average Rg value for PARP-2 enzyme, PARP-2/C-1 and PARP-2/C-32 complexes are found

to be around 2.1 nm, 2.1 nm and 2.13 nm respectively. The lesser Rg value for PARP-2/C-1 complex indicates the compactness of complex causing the interactions between protein and ligand stronger. Similarly, the average Rg value for PARP-1 enzyme, PARP-1/C-1 and PARP-1/C-32 complexes are found to be around 2.06 nm, 2.08 nm and 2.14 nm respectively. It can be seen that for PARP-1 complexes there is an increase in average Rg values compared to unbound PARP-1 enzyme which clearly shows the unstable nature of PARP-1 complexes, out of which PARP-1/C-32 complex is more unstable.

Hydrogen Bonding Interactions with PARP Enzymes: Hydrogen bonding interactions are found to be a stabilizing factor, apart from π - π stacking interactions. The average number of hydrogen bonds formed throughout the simulation of C-1 and C-32 with PARP enzymes is found to be 3-4 (**Figure 4.8**).

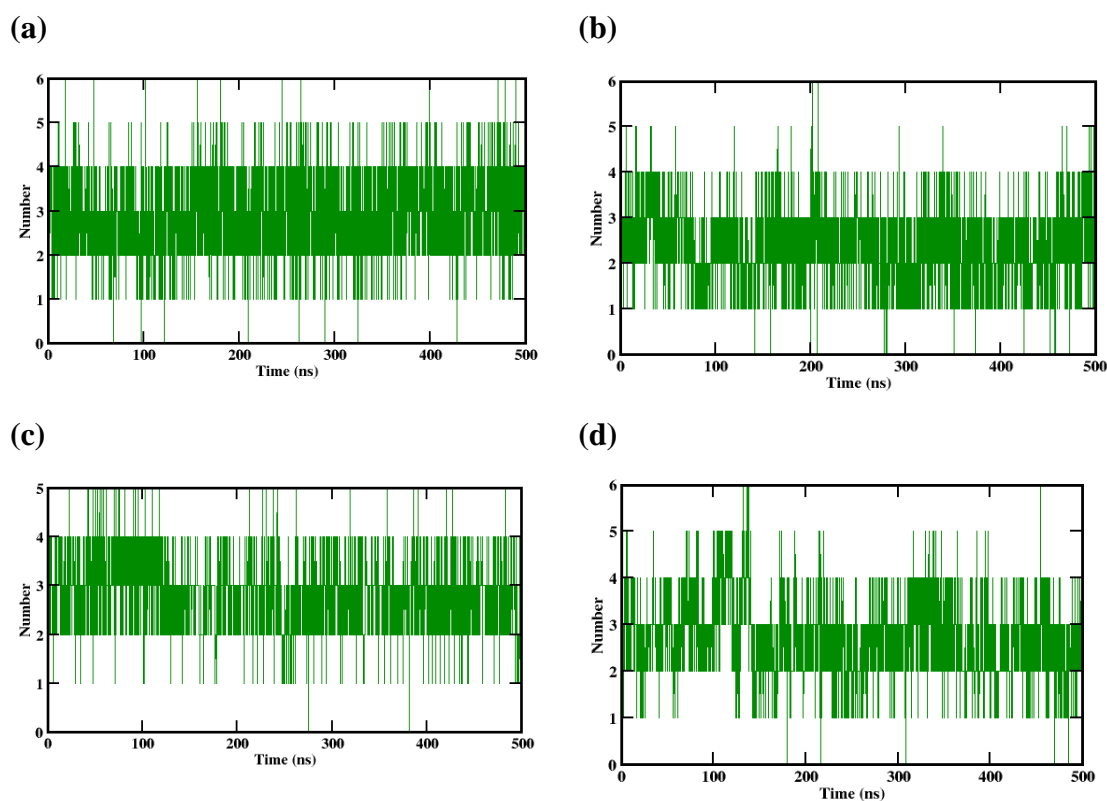


Figure 4.8: Number of hydrogen bonds calculated for the evaluation of protein-ligand interaction throughout 500 ns simulation (a) PARP-2/C-1, (b) PARP-2/C-32, (c) PARP-1/C-1 and (d) PARP-1/C-32 respectively.

Molecular docking showed only one hydrogen bond with glycine in the complexes. Molecular dynamics studies revealed that glycine contribute to major part of hydrogen bond interaction in all the complexes (**Table 4.4**). Other residues such as tyrosine (-OH) and serine also contributes to hydrogen bonding in the complex. Additionally, the simulation of PARP-2/C-1 complex showed hydrogen bond with glutamine residue. This can be attributed due to the possibility of ligand to adopt favourable conformation in the binding site during simulation. Also, water mediated hydrogen bond with C-1 and glutamatic acid provides extra stability to the complex. Even though similar hydrogen bond interactions are observed for PARP-2/C-32, hydrogen bond occupancy is decreased in this complex. Since the catalytic pocket of PARP-1 is similar to that of PARP-2, similar hydrogen bonding interactions are observed with lesser interactions and greater occupancy, suggesting that the ligand is bound intact within the site in case of PARP-2. The hydrogen bond interactions with aspartic acid observed in docking is found to be missing during the simulation. Water-mediated hydrogen bonds are absent in PARP-2/C-32 and PARP-1 complexes.

Table 4.4: Hydrogen bond occupancy between ligand and amino acid residues at binding pocket of enzyme

System	Hydrogen Bond Occupancy		Water-mediated Hydrogen Bonds
PARP-2/ C-1	Lig (H) --- Gly429 (O)	83.8%	Lig (H) --- water --- 10.85% Glu558
	Lig (H) --- Tyr455 (OH)	25.6%	
	Ser470 (H) --- Lig (O)	95.3%	
	Gly429 (H) --- Lig (O)	60.2%	
	Gly429 (H) --- Lig (N)	16.0%	
	Gln332 (H) --- Lig (N)	10.6%	
PARP-2/ C-32	Lig (H) --- Gly429 (O)	85.4%	Nil
	Lig (H) --- Tyr455 (OH)	15.8%	
	Lig (H) --- Ser328 (O)	13.7%	
	Ser470 (H) --- Lig (O)	24.7%	
	Gly429 (H) --- Lig (O)	90.6%	

PARP-1/ C-1	Lig (H) --- Gly863 (O)	89.4%	Nil
	Lig (H) --- Tyr889 (OH)	12.5%	
	Ser904 (H) --- Lig (O)	88.8%	
	Gly863 (H) --- Lig (O)	90.8%	
PARP-1/ C-32	Lig (H) --- Gly863 (O)	89.4%	Nil
	Lig (H) --- Gly888 (O)	12.4%	
	Ser904 (H) --- Lig (O)	51.8%	
	Gly863 (H) --- Lig (O)	97.6%	
	Gln759 (H) --- Lig (N)	15.8%	

MM/PBSA Free Energy: To estimate the strength of protein-ligand interactions, binding energy analysis was carried out by MM/PBSA method for the whole 500 ns MD trajectory. The binding free energy of PARP-2/C-1 complex is found to be -25.29 ± 2.9 kcal/mol. The most important contribution for PARP-2/C-1 complex is from van der Waals energy (-45.67 ± 2.9 kcal/mol) compared to the electrostatic contribution (-11.72 ± 2.0 kcal/mol). A significant amount of SASA energy (-4.14 ± 0.2 kcal/mol) also contributes to the stability of complex. The high positive value for polar solvation energy (36.24 ± 2.9 kcal/mol) disfavors the ligand binding at the catalytic pocket. MM/PBSA binding free energy correlates well and explains the importance of hydrophobic groups. The free energy decomposition for all the system is given in **Table 4.5**. PARP-1/C-1 is found to have binding energy of -27.54 ± 3.2 kcal/mol due to the intact conformation of ligand which favours hydrogen bonding interaction, thereby increases the contribution of electrostatic energy. PARP-2/C-32 and PARP-1/C-32 complex showed higher energy indicating the unstable nature of complex. Overall, the energy contribution to the final binding energy of the complex can be written in the order: van der Waals energy > electrostatic energy > SASA energy.

Table 4.5: Contribution of energy components to MM/PBSA binding free energy for PARP enzymes with C-1 and C-32 (energy in kcal/mol)

System	van der Waals Energy	Electrostatic Energy	Polar Solvation Energy	SASA	Binding Energy
PARP-2/C-1	-45.67 ± 2.9	-11.72 ± 2.0	36.24 ± 2.9	-4.14 ± 0.2	-25.29 ± 2.9
PARP-2/C-32	-43.02 ± 3.1	-7.95 ± 2.3	31.56 ± 2.9	-4.08 ± 0.2	-23.65 ± 3.1
PARP-1/C-1	-42.37 ± 3.2	-15.78 ± 2.3	34.61 ± 2.9	-4.01 ± 0.2	-27.54 ± 3.2
PARP-1/C-32	-37.73 ± 3.2	-12.47 ± 3.2	32.17 ± 4.8	-3.82 ± 0.3	-21.85 ± 3.0

From the protein-ligand complex analysis based on the protein structural changes, it is found that PARP-2/C-1 complex is more stable than other complexes (PARP-2/C-32 and PARP-1/C-32) under consideration. The stability of PARP-2/C-1 complex is further confirmed by water-mediated hydrogen bond interactions. C-1 complexes of both the enzymes showed lower binding free energy.

Per Residue Energy Analysis of PARP-2 complexes: Per residue binding energy contributions to the enzyme-ligand complex were determined. It suggested that Ser328, Ile331, Gln332, Glu335, His428, Gly454, Tyr455, Tyr462, Lys469, Ser470, Tyr473 and Glu558 contribute to non-polar solvation (SASA) energy in ligand complex. Similarly, Ser328, His428, Gly429, Gly454, Tyr455, Tyr462, Phe463, Ala464, Lys469, Ser470, Tyr473 and Glu558 contribute to the molecular-mechanics (electrostatic and van der Waals) energy. Among these residues, average binding energy contribution of Tyr473, His428, Tyr462, Ser470, Glu558 and Gly429 are found to be significantly high compared to other residues at the binding site. The comparison of average per residue binding energy contribution of amino acid residues to each energy term for the PARP-2 complexes is given in **Table 4.6**. Some residues such as Ser328, Gln332, His428, Gly429, Gly454, Tyr455, Lys469, Ser470 and Tyr473 showed higher positive energy towards polar solvation, though it is negligible compared to their contribution towards favourable energy component. The residue Glu558 which is responsible for water-mediated hydrogen bonding with C-1, showed extremely higher polar solvation energy (4.84 kcal/mol). Overall, the most favourable residues contributing to stable ligand binding are Tyr473> Tyr462> His428> Ser470> Tyr455> Glu558> Gly429> Gln332. This result suggests the significance of Tyrosine

and Histidine amino acid residues for the stable ligand interaction. The stabilization of the complex is well-balanced by hydrophobic and hydrogen bonding interactions between ligand and protein.

Table 4.6: Per residue binding energy contribution in the PARP-2/C-1 and PARP-2/C-32 complexes

Energy Term	Contributing Amino acids	Major Contribution (Energy in kcal/mol)	
		PARP-2/C-1	PARP-2/C-32
SASA	Ser328, Ile331, Gln332,	Tyr473 ~ -0.34	Tyr473 ~ -0.36
	Glu335, His428, Gly429,	Tyr462 ~ -0.24	Tyr462 ~ -0.24
	Gly454, Tyr455, Met456,	Tyr455 ~ -0.20	Tyr455 ~ -0.21
	Tyr462, Phe463, Ala464,	Gln332 ~ -0.14	Gln332 ~ -0.11
	Lys469, Ser470, Tyr473, Glu558	Ser328 ~ -0.11	Ser328 ~ -0.10
van der Waals Energy	Ser328, Ile331, Gln332,	Tyr473 ~ -4.44	Tyr473 ~ -4.29
	His428, Gly429, Gly454,	Tyr462 ~ -3.00	Tyr462 ~ -2.61
	Tyr455, Tyr462, Phe463,	His428 ~ -2.84	His428 ~ -2.92
	Ala464, Lys469, Ser470,	Tyr455 ~ -2.78	Tyr455 ~ -2.57
	Tyr473, Glu558	Phe463 ~ -1.14	Phe463 ~ -0.79
Electrostatic Energy	Tyr473, His428, Tyr462,	Glu558 ~ -3.17	Glu558 ~ -1.38
	Ser470, Glu558, Gly429	Ser470 ~ -2.70	Ser470 ~ -1.78
		Gly429 ~ -1.69	Gly429 ~ -1.75
Polar Solvation Energy	Ser328, Gln332, His428,	Glu558 ~ 4.84	Glu558 ~ 2.04
	Gly429, Gly454, Tyr455,	Tyr473 ~ 2.14	Tyr473 ~ 1.82
	Lys469, Ser470, Tyr473,	Ser470 ~ 1.61	Ser470 ~ 1.20
	Glu558		

The per residue energy distribution analysis of PARP-1 complexes proves that the binding of ligand at the catalytic site is less stable compared to PARP-2 complexes. The amino acid residues contribute to the total energy of PARP-1

complexes is found to be similar as that of PARP-2 complexes (**Table 4.7**). But the individual contribution from the residues are higher in case of PARP-2 complexes, indicates a better binding efficiency of inhibitor towards PARP-2 enzyme.

Table 4.7: Per residue binding energy contribution in the PARP-1/C-1 and PARP-1/C-32 complexes

Energy Term	Contributing Amino acids	Major Contribution (Energy in kcal/mol)	
		PARP-1/C-1	PARP-1/C-32
SASA	Gln759, Glu763, His862, Gly863, Gly888, Tyr889, Tyr896, Phe897, Lys903, Ser904, Tyr907, Glu988	Tyr907 ~ -0.32	Tyr907 ~ -0.35
		Tyr896 ~ -0.29	Tyr896 ~ -0.17
		Tyr889 ~ -0.14	Tyr889 ~ -0.29
		Glu763 ~ -0.16	Glu763 ~ -0.03
van der Waals Energy	Gln759, His862, Gly863, Gly888, Tyr889, Tyr896, Phe897, Ala898, Lys903, Ser904, Tyr907, Glu988	Tyr907 ~ -3.87	Tyr907 ~ -4.29
		Tyr896 ~ -3.12	Tyr896 ~ -2.18
		His862 ~ -2.67	His862 ~ -2.75
		Tyr889 ~ -1.59	Tyr889 ~ -3.08
		Phe897 ~ -1.03	Phe897 ~ -0.66
Electrostatic Energy	Tyr907, Arg858, His862, Tyr896, Ser904, Glu988, Gly863	Glu988 ~ -2.70	Glu988 ~ -1.11
		Ser904 ~ -2.53	Ser904 ~ -1.84
		Gly863 ~ -1.82	Gly863 ~ -1.88
Polar Solvation Energy	Gln759, Arg858, His862, Gly863, Gly888, Tyr889, Lys903, Ser904, Tyr907, Glu988	Glu988 ~ 2.87	Glu988 ~ 2.39
		Tyr907 ~ 1.72	Tyr907 ~ 1.60
		Ser904 ~ 1.54	Ser904 ~ 1.22

4.6 LIGAND ORIENTATION: HYDROGEN BONDING AND AROMATIC RING STACKING

As obtained from molecular docking and molecular dynamics simulation, hydrogen bonding and aromatic ring stacking plays a major role in the formation of stable

PARP-2/C-1 complex. The favourable ligand orientation at the binding pocket is a key factor in stable protein-ligand interactions. The hydrogen bond distance (D---A) criteria is 2.5-3.0 Å and 3.0-4.0 Å for strong and weak interactions respectively (Desiraju and Steiner 2001). Detailed analysis of PARP-2 complexes showed that nearly all the hydrogen bond distance between the residues and C-1 are 2.8 Å while for other complexes the distance is around 3 Å (**Figure 4.9**). The inter atomic distance of donor-acceptor atoms are used as a measure of hydrogen bond strength. Thus, the interactions in PARP-2/C-1 complex are stronger compared to other complexes.

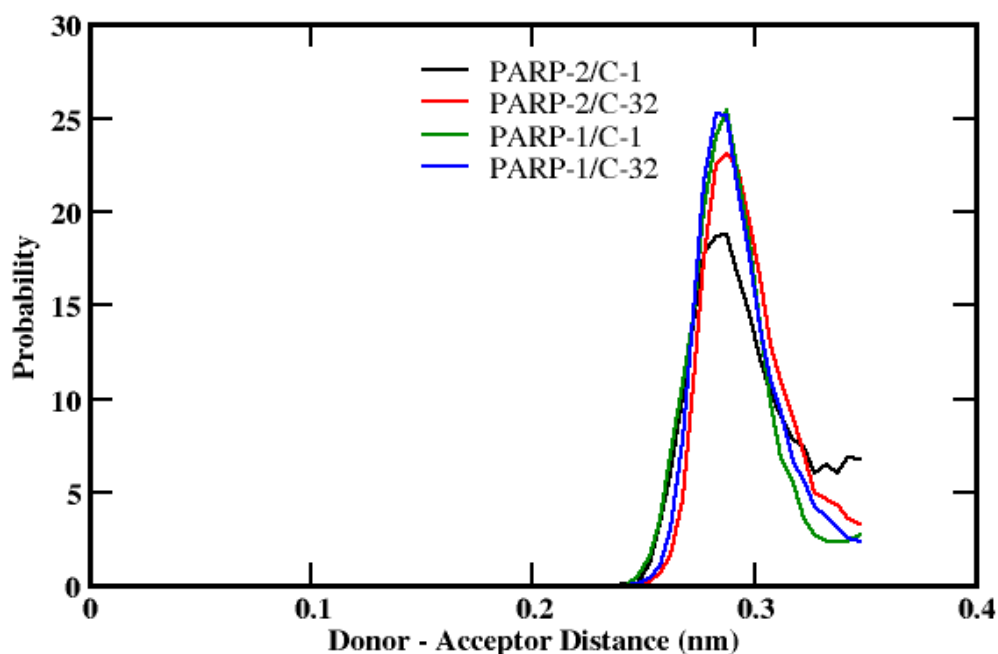


Figure 4.9: Distribution of donor-acceptor distance (nm) in hydrogen bond interaction for all the complexes.

In addition to hydrogen bonding interactions, π - π stacking interactions are commonly observed in PARP complexes. This interaction is characterized by the angle between normal vectors of two rings from the ligand and the residues. The geometric parameters for perfect π - π stacking interactions are 4.5-5.5 Å for distance and the angle between normal vectors of two rings should be in range of 0-90° (Brylinski 2018). PARP-2/C-1 complex showed stronger interaction with the residues based on angle criteria (**Appendix X**). In the simulations, the ligand formed π -

stacking interactions with His428, Tyr455, Tyr462, Phe463 and Tyr473 in both the complexes. Small angle (0-15°) indicates face-to-face π -stacking and larger angle (near-right angles) indicates edge-to-face stacking interactions. The ligand C-1 preferred to form small stacking angle (face-to-face stacking interactions) with all the aromatic residues at the binding pocket namely, His428, Tyr455, Tyr462 and Tyr473 residues of PARP-2 enzyme except Phe463 (edge-to-face interactions).

PARP-1 complexes form unstable aromatic interaction based on angle criteria and distance criteria compared to PARP-2 complexes (**Appendix XI**). The amino acid residues His862, Tyr889, Tyr896, Phe897 and Tyr907 at the binding site of PARP-1 enzyme analogous to PARP-2 enzyme, showed less favourable stacking interactions with C-1 and C-32. Phe463 in PARP-2 residues plays a role in stabilizing both the complex by forming edge-to-face interactions (7.34 Å and 7.54 Å) while Phe897 in PARP-1 fluctuates between lower and higher angles. Tyr473 showed π -stacking interaction with all the three aromatic rings present in the ligand during the course of simulation whereas Tyr907 showed only two π -stacking interactions after 250 ns. This shows the greater contribution of hydrophobic interactions in PARP-2 complexes and correlates well with van der Waals and SASA energy components.

4.7 PCA AND FEL ANALYSIS OF COMPLEX

Principal Component Analysis (PCA) is used to determine the most important functional motions of biomolecules (Hayward et al. 2008). PCA was performed for C- α atoms on 500 ns MD trajectory using Gromacs 2018.4. Using in-house scripts, the first principal components and corresponding eigenvectors of the system were calculated. Since, this study is mainly focussed on developing a promising inhibitor for PARP-2 enzyme, principal component analysis (PCA) and FEL analysis were carried out for only PARP-2 complexes. As evident from the **Figure 4.10(a)**, the eigenvalue decreases rapidly whereas first three eigenvectors contribute significantly to the conformational changes in the protein during simulation. In this study, first 25 eigenvectors are selected for the determination of significant motions along 500 ns trajectory. The first nine eigenvectors account for 67.77%, 69.64% and 64.77% motions for PARP-2 enzyme, PARP-2/C-1 and PARP-2/C-32 respectively,

suggesting that ligands induce different conformational changes to the protein in its bound state. The first three eigenvectors contribute >50% and >40% of the possible movements during ligand binding in PARP-2/C-1 and PARP-2/C-32 complexes respectively. In both the complexes, the enzyme is showing lesser correlated motions with adjacent amino acid residues. PCA analysis suggested the flexibility of amino acid residues at the binding pocket of PARP-2/C-1 complex, agreeing with the per residue energy analysis.

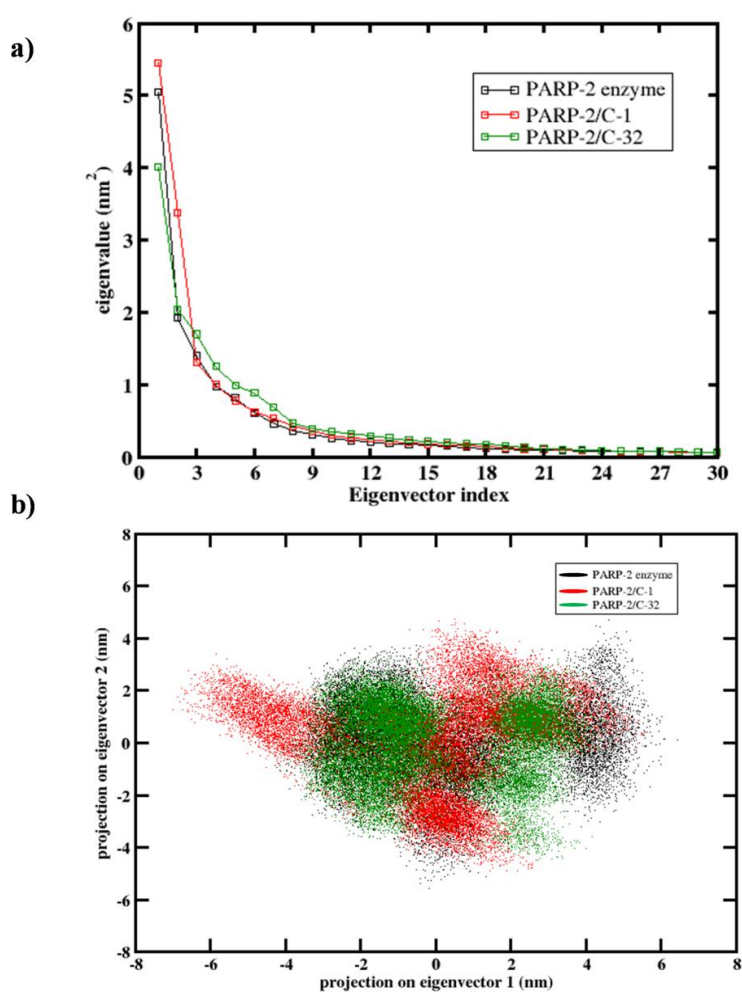


Figure 4.10: (a) Plot of eigenvalue vs. first 30 eigenvector index derived from PCA over 500 ns MD trajectory for unbound and bound PARP-2 system. (b) 2-D projection of first two principal motions (PC1 and PC2) of protein in phase space for unbound and bound PARP-2 enzyme.

For better representation, first two eigenvectors are plotted in 2D projection plot (**Figure 4.10b**) for bound and unbound enzymes. PARP-2/C-1 complex spread over the phase space more than unbound PARP-2 and PARP-2/C-32 complex, indicating that flexibility of protein enhances the stability of complex. PARP-2/C-1 complex forms stable clusters compared to PARP-2/C-32 complex. Thus, it is found that PARP-2/C-1 complex is stable having lower binding energy.

The cosine content of principal components provides fair estimate of conformational changes in the protein during the course of simulation (Khan et al. 2017). The cosine value from 0 to 1 indicates whether the MD trajectory has sampled free energy landscape sufficiently for convergence. The low cosine value indicates fewer fluctuations of backbone C- α atoms (**Table 4.8**) and well-sampled. The value of first two principal components (0.360 and 0.031 respectively) for PARP-2/C-1 complex suggests that 500 ns trajectory is strong enough to explain the stable movements of complex. The value of cosine content of first principal component close to 1 for PARP-2 enzyme and PARP-2/C-32 complex indicates the random diffusion of protein chain which mask the features of free energy landscape.

Table 4.8: Cosine content on the principal components of C- α atoms

Systems	PC1	PC2	PC3
PARP-2	0.712	0.005	0.017
PARP-2/C-1	0.360	0.031	0.016
PARP-2/C-32	0.789	0.004	0.012

Porcupine plots show the magnitude and direction of particular eigenvectors associated with proteins. The first three principal motions from PCA of complexes are shown in **Appendix XII**. The length of arrow gives the magnitude and the arrow-head shows the direction of significant motion. The first three modes of unbound protein showed larger fluctuations at the loop region and binding site (**Appendix XII**). In the case of complex, the reduction of these fluctuations can be attributed due to the binding of ligand at the catalytic domain of the enzyme. In PARP-2/C-1 complex, the system is found to be stable with fewer fluctuations at the binding pocket. The fluctuations are found higher in N- and C-terminal regions of enzymes. Depends on

the flexibility of binding site and the ligands, the protein-ligand complex can vary the structural robustness from very loose to very tight which have functional implications based on the nature of inhibition (Majewski et al. 2019). From the 2D-projection plot, the cosine values and the different modes of protein motion, it is evident that lower structural robustness favours the potency of C-1 while robustness decreases the potency of C-32 in PARP-2 complexes. The movement of loop region of enzyme gives greater contribution to the stability of ligand in PARP-2/C-1 complex. Overall, the motion of protein is reduced in the case of PARP-2/C-32 complex, indicating less residue interaction with the ligand and lower stability of the complex.

RMSF can also be used as a feature to validate the correctness of principal components by overlaying the overall RMSF of protein with the amino acid residue movements of each component from PCA calculation (Antunes et al. 2019). In PARP-2/C-1 complex, the residue contribution to PC1 is mostly from the binding site region (400-500 amino acid residues) which is enclosed by the loops. A large contribution to PC2 from region 350-360 residues can be observed, which reside away from the binding sites. PC3 is found to have negligible contribution towards the ligand binding. Overall, the first and second principal components have major contribution in the stable ligand binding. In PARP-2/C-32 complex, all the components form half the value of overall RMSF and have negligible contribution to ligand binding. Even though, the residues from 500-580 contributes to PC1, these amino acids resides away from the binding pocket. The higher fluctuations found at the end points in both complexes is due to the C-and N-terminal residues. Even though both the complex showed similar overall backbone RMSF, the mode 1 contributes higher in PARP-2/C1 complex while there is a negligible contribution from all the modes in PARP-2/C-32 complex.

Free Energy Landscapes (FEL) can elucidate the dynamic behaviour of biological system (Frauenfelder et al. 1991). PC1 and PC2, which form the major protein motions, were taken as the reaction coordinates to construct the FEL. Free energy landscapes were constructed to relate the structural properties with the thermodynamic information obtained for unbound and ligand-bound enzyme (**Figure 4.11**). The energy minima on free energy landscape for complexes indicate the stable

ligand conformation at the binding pocket of enzyme over estimated time scale while the energy barriers indicate the presence of transient states. Three well-defined minima are found for PARP-2/C-1 complex indicates that the complex undergoes transitions to achieve most stable configuration. PARP-2/C-32 complex showed two well-separated minima at slightly higher energy indicates its less stability and unbound PARP-2 enzyme is found to be stable. The free energy plot for C-1 complex gets shifted towards the left side of the graph in **Figure 4.11(b)** compared to **Figure 4.11(a) & (c)** indicating the stable nature of C-1 complex. PARP-2/C-1 complex has higher blue region (minimum energy) on the energy plot suggesting the formation of complex is thermodynamically favourable.

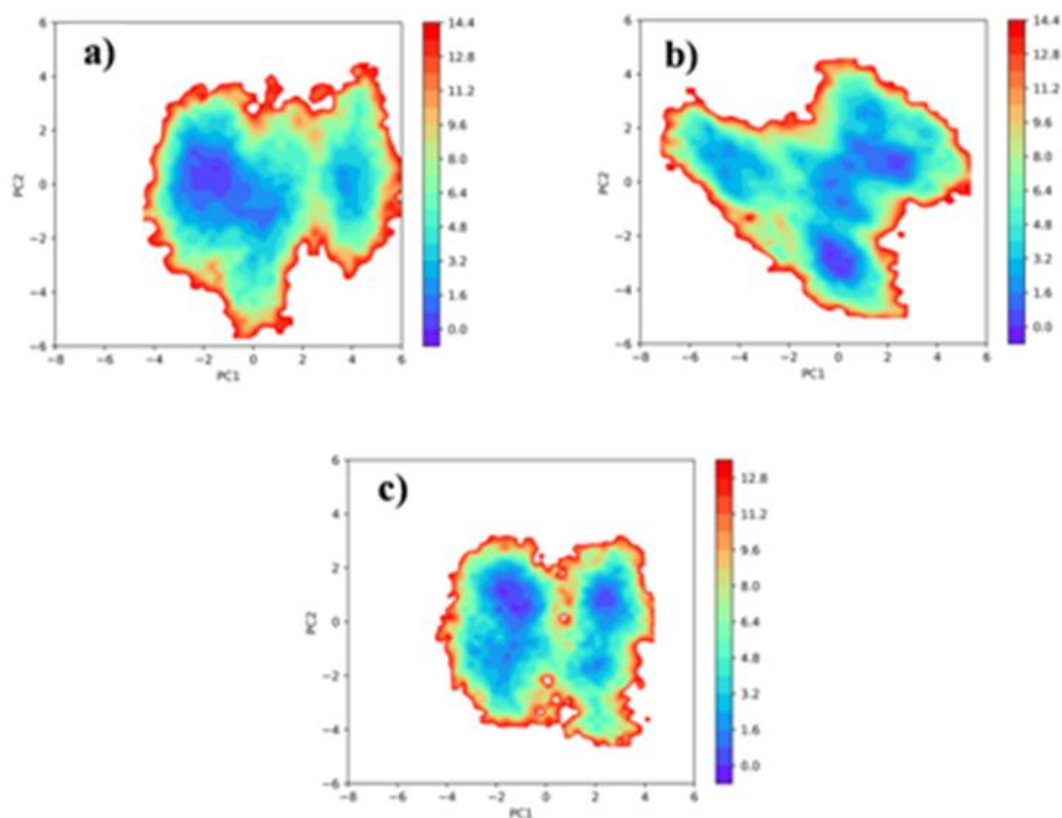


Figure 4.11: Gibbs free energy landscape for unbound and bound PARP-2 enzymes. FEL obtained from MD trajectory for all three systems using the reaction coordinates as the projection of C- α atoms onto the first two principal components. **(a)** unbound PARP-2 enzyme **(b)** PARP-2/C-1 complex and **(c)** PARP-2/C-32 complex.

4.8 DESIGN OF NEW POTENT COMPOUNDS

From molecular dynamics and binding free energy analysis, it is found that hydrogen bond donor-acceptor groups and aromatic ring structures are important for effective ligand-binding at the binding pocket of the enzyme. It is found from PCA analysis that lower structural robustness of the ligand favors the efficiency of inhibition, which should be accounted in designing the drug molecule. For designing new effective PARP-2 inhibitors, the benzimidazole core is changed to quinoxaline core, which is more aromatic in nature, for three molecules (P2, P3, and P4) and benzofuranone, which is less aromatic, in P1. Based on the above results, heteroaromatic and aliphatic rings were added to increase the hydrophobicity of the ligands.

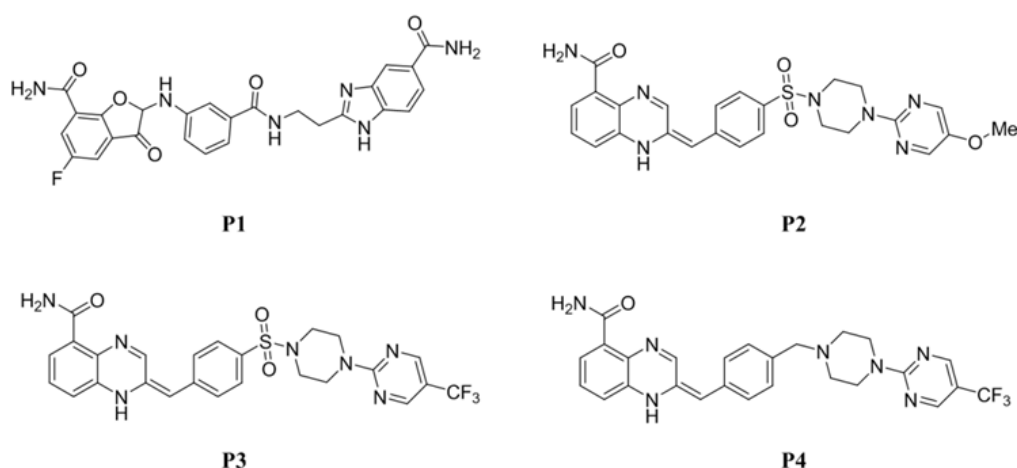


Figure 4.12: Structure of newly designed PARP-2 inhibitors.

Four molecules are predicted which showed stable binding energy compared to the dataset under consideration (**Figure 4.12**). The predicted molecules were optimized using B3LYP/6-31G(d,p) level. The HOMO-LUMO energy gap indicates that all the predicted molecules are chemically reactive compared to the benzimidazole derivatives. Among the predicted molecules, P1 is found to have higher energy gap of 3.64 eV, which indicates the low reactivity of molecule. Further, the activity of newly designed molecules was analyzed by 25 ns MD and binding free energy analysis (**Table 4.9**). The RMSD of protein backbone for all the protein-ligand complexes are stabilized after 5 ns of the simulation (**Figure 4.13**). The compound P1

found to have a lower binding free energy than the other molecules. It is evident from molecular docking, RMSD and binding free energy, the structural modifications in aromaticity and hydrophobicity of the benzimidazole derivatives resulted in an increase in activity of molecules. The number of hydrogen bond interactions, the binding energy decomposition and per residue analysis for each complex shows the importance of ligand flexibility at the binding site (**Appendix XIII-XV**).

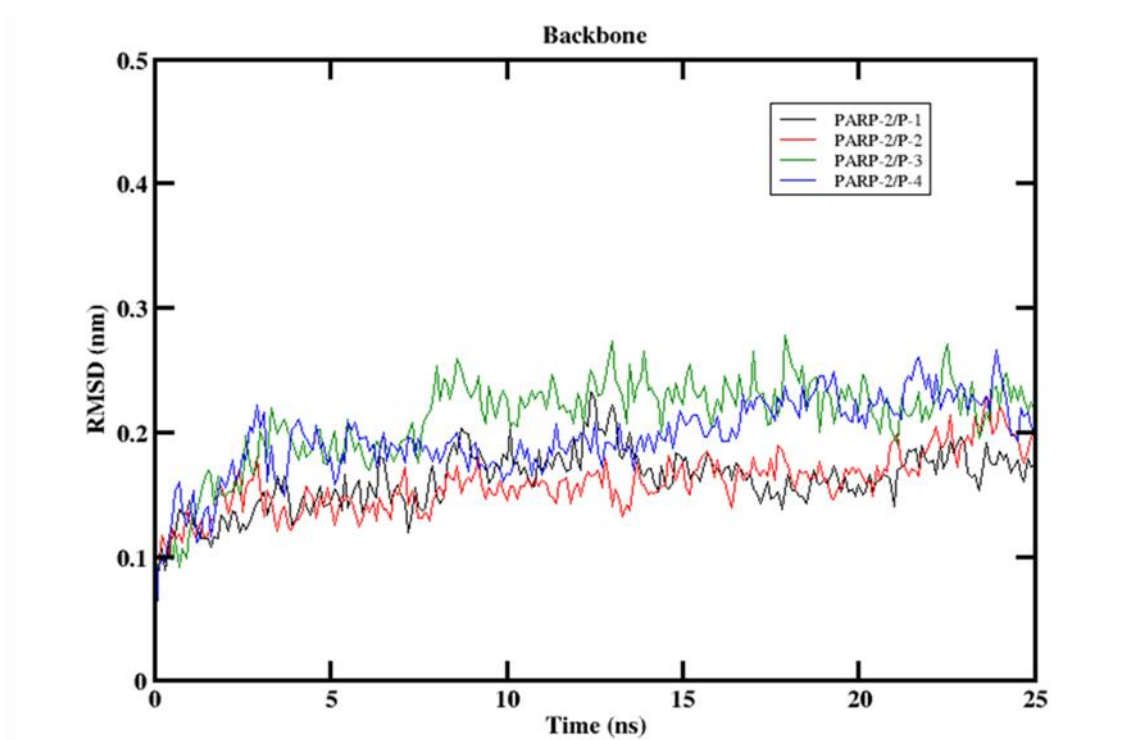


Figure 4.13: Time-line representation of RMSD profile of backbone atoms with respect to the initial coordinate of PARP-2 enzyme.

The pharmacokinetic (PK) properties (drug-likeness) of the predicted drugs were determined by Qikprop module (Schrödinger Release 2020-1). The PK properties like absorption, distribution, metabolism, excretion, and toxicity (ADMET) were determined in order to confirm the effectiveness and bioavailability of the four predicted molecules (**Table 4.9**). The ADME/toxicity prediction showed that P2, P3, and P4 are suitable drug candidates for PARP-2 inhibition whereas P1 failed to be a suitable inhibitor. The lipophilicity of the molecules measured by QPlogPo/w was found to be optimum except for P1 (0.528). The dipole moment of each molecule lies in the permissible range (1.0-12.5 Debye). The number of hydrogen bond donors (3)

and acceptors (8.5-11.75) were found to be in acceptable range for P2, P3, and P4. The cell permeability value (polar surface area [PSA]) for the predicted molecules lies in the range 104.85 to 148.14 Å² which shows good oral availability except for P1 (222.23 Å²). The blood/brain barrier values (QPlogBB) and binding of molecule to human serum albumin (QPlogK_{hsa}) were found to be in threshold limit except for the molecule P1. The molecule P1 showed poor percentage of oral absorption whereas other molecules showed higher oral absorption percentage. The three molecules (P2, P3, and P4) obeyed Lipinski's rule of five with at most two deviations whereas P1 showed violation of Lipinski's rule of 5. The results suggest that the predicted molecules, except P1, can serve as drug candidates for effective PARP-2 inhibition.

Table 4.9: Predicted ADME/Toxicity properties for newly designed PARP-2 inhibitors

*Properties	Predicted drugs			
	P1	P2	P3	P4
Molecular Weight	516.5	519.6	573.5	507.5
Dipole Moment	6.605	12.268	9.528	11.472
SASA	841.085	838.753	860.09	848.273
Donor Hydrogen Bond	7	3	3	3
Acceptor Hydrogen Bond	12.75	11.75	11	8.5
QPlogP _{o/w}	0.528	2.512	3.523	4.101
QPlogS	-5.577	-6.208	-7.315	-6.854
Rotor	8	6	6	5
PSA	222.23	148.14	147.68	104.85
QPlogK _{hsa}	-0.348	0.13	0.366	0.699
QPlogBB	-4.154	-2.428	-2.157	-0.953
QPPC _{aco}	3.173	66.297	67.5	74.882
QPPMDCK	1.766	26.672	122.96	148.097
Percent Human Oral Absorption	0.139	48.336	54.395	71.547
Rule of 5	3	2	2	1

*Permissible ranges of physico-chemical properties are given in Section 2.6

4.9 CONCLUSION

PARP inhibition is a promising therapeutic strategy in cancer treatment. However, the clinically trialled drug candidates have no selectivity for PARP enzymes. Here, we aimed to analyze the molecular mechanism and energy requirements for the inhibition of PARP-2 enzyme by benzimidazole analogues using structure-based and DFT methods. Utilizing molecular docking procedure, a set of 32 benzimidazole derivatives was screened to determine the binding mode and its affinity. Docking studies revealed that the ligand forms hydrogen bonding with Gly429 and π - π stacking interaction with Tyr473 at the binding pocket. Also, binding pose of most potent compound, C-1, suggested that non-polar groups such as methyl group stabilize the complex by hydrophobic interactions with Ile331, Tyr455 and Met456. HOMO-LUMO analysis and the reactivity descriptors described the characteristics of ligand at the quantum level. It is found that presence of LUMO over benzimidazole ring of ligand favours π - π stacking interactions and HOMO orbital favours hydrogen bonding interactions with the amino acid residues. The stability of complex formed confirmed by 500 ns molecular dynamics simulation. The simulation revealed that the protein-ligand systems are stabilized by hydrogen bonding and π - π interactions. MD results showed the complex is stabilized by additional hydrogen bond interactions with Gln332, Tyr455 and Ser470. Face-to-face π -stacking interaction is formed between aromatic amino acid residues and C-1 for stable protein-ligand interaction. MM/PBSA free energy results for PARP-2/C-1 complex suggested that van der Waals energy and non-polar solvation energy (SASA) are crucial for the ligand binding at catalytic pocket. Per residue binding energy contribution showed that tyrosine, phenyl alanine and histidine residues make the binding pocket hydrophobic in nature, facilitating the ligand to bind effectively. PCA analysis showed the importance of low structural robustness of amino acid residues for the efficient ligand-binding at the catalytic pocket of the enzyme. It is evident from the RMSF of first and second principal component (mode 1 & 2) that flexibility of residues at binding site and loop regions favours the potency of ligand C-1. Analysis was also done to analyze the stability of PARP-1 complexes with the screened ligands by docking, molecular dynamics simulation and MM/PBSA free energy calculations. PARP-1/C-32 complex found to be less stable compared to PARP-2 complexes and PARP-1/C-1 complex,

structurally and thermodynamically. Overall, the study suggests that the presence of aromatic rings in the ligand favours the effective inhibition of PARP-2 enzyme by forming π - π stacking interactions with tyrosine residues at the binding pocket. From the outcomes of this study, four new inhibitors (P1, P2, P3, and P4) were proposed with higher activity out of which the three molecules appeared to be promising candidates for PARP-2 inhibition except P1. These insights are useful for the designing target-specific and isoform-selective PARP-2 inhibitors.

CHAPTER 5

SINGLE AMINO ACID SUBSTITUTION AT CATALYTIC DOMAIN OF POLY (ADP-RIBOSE) POLYMERASE 2 IDENTIFIES RESIDUES ESSENTIAL FOR ENZYME INHIBITION

In co-crystal structure, it is found that the binding of substrate, nicotinamide, at PARP-2 catalytic domain are stabilized by hydrogen bonding and π - π stacking interactions. In this study, single amino acid substitution is employed to identify the essential amino acid residues responsible for PARP-2 inhibition.

5.1 BACKGROUND

Molecular recognition between the biomacromolecules (proteins, polysaccharides, nucleic acids, lipids) and the small organic molecules (natural substrates/ inhibitors) is necessary for the normal cell function and metabolism. Among the biomacromolecules, enzymes act as biological catalysts which accelerate the metabolism and chemical reactions inside the living cell (Robinson 2015). Poly ADP-ribosylation (PARylation) is an important post-translational protein modification which contributes to molecular and cellular processes including DNA damage repair (Hottiger 2015). In humans, this modification is catalyzed by poly (ADP-ribose) polymerase (PARP) family of enzyme which consists of 17 members (Hottiger 2015). The enzymes of PARP family use nicotinamide adenine dinucleotide (NAD⁺) molecules as the substrate to produce and transfer ADP-ribose polymers onto the target proteins (Schreiber et al. 2006), involved in various processes such as repairing of DNA damage sites, chromatin structure modification, control of transcription and cell division (Kraus 2015). Polymorphisms and mutations in PARP-2 enzyme are found to associate with different cancer which suggests the importance of PARP-2 tumorigenesis.

The biological function of PARP-2 is achieved by physical interactions with other proteins or PARylation of target proteins. This is implicated in various biological processes such as genomic maintenance, cell cycle regulation, cell death,

transcription, cell signalling and metabolism. The role of PARP-2 in DNA-repair can be a potential utility for selective PARP-2 inhibition in chemo-/radiotherapy or synthetic lethality in cancer treatment. The PARP inhibitors competitively inhibit the enzyme by replacing NAD⁺ substrates at the nicotinamide-binding pocket. In co-crystal structures, it is found that the inhibitors bind to the catalytic pocket of PARP enzymes by hydrogen bonding and π - π interactions (Aoyagi-Scharber et al. 2014). Majority of the PARP inhibitors contain carboxamide group which is similar to nicotinamide group to mimic the binding pose of substrate at the catalytic site (Chen et al. 2018; Costantino et al. 2001; Reddy et al. 2018; Zhou et al. 2017). Thus, benzimidazole carboxamide derivatives can be an ideal ligand to study the importance of polar residues and aromatic residues at the binding pocket of PARP-2 enzyme.

5.2 COMPUTATIONAL ALANINE SCANNING

Computational alanine scanning (CAS) helps to predict the effect of site-directed mutation on the stability of protein. In this approach, each residue of the target protein is substituted by alanine so that the neutral methyl group replaces the reactive group of the residues on the protein side chain. Alanine substitution is widely used to predict the functional impact of amino acid residues on the protein as well as the changes in protein structural stability caused by mutation. CAS has been performed on the whole residues of PARP-2 enzyme in order to determine the protein stability upon single amino acid substitution. The amino acid substitution is carried out by Site Directed Mutator (SDM) which predicts the stability score based on the free energy difference between wild-type and variant protein (Worth et al. 2011). The stability score, $\Delta\Delta G$ (kcal/mol), were evaluated to determine the role of specific amino acid in stabilizing the enzyme structure. The free energy difference, ΔG^{U-F} between the folded (F) and unfolded (U) states determines the protein stability. If a residue, a , in the wild protein is mutated to the residue, b , then the free energy change in the reversible folding-unfolding process is given by $\Delta\Delta G = \Delta G_b^{U-F} - \Delta G_a^{U-F}$ (Topham et al. 1997). $\Delta\Delta G < 0$ indicates the reduced stability of protein whereas $\Delta\Delta G \geq 0$ indicates that the resulted protein structure gained stability upon mutation (**Figure 5.1**).

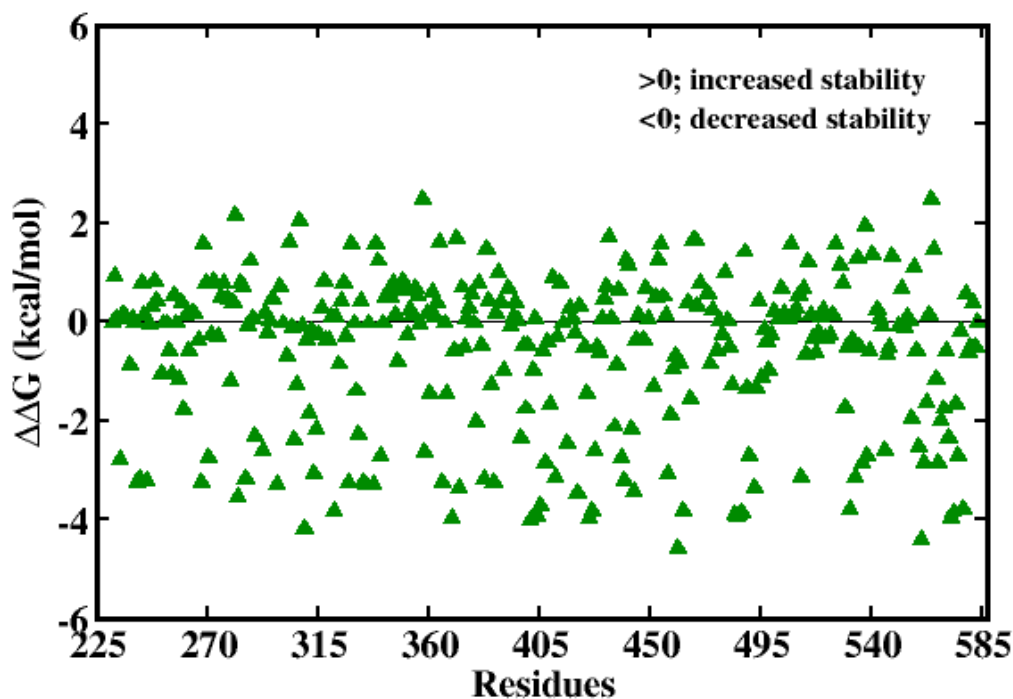
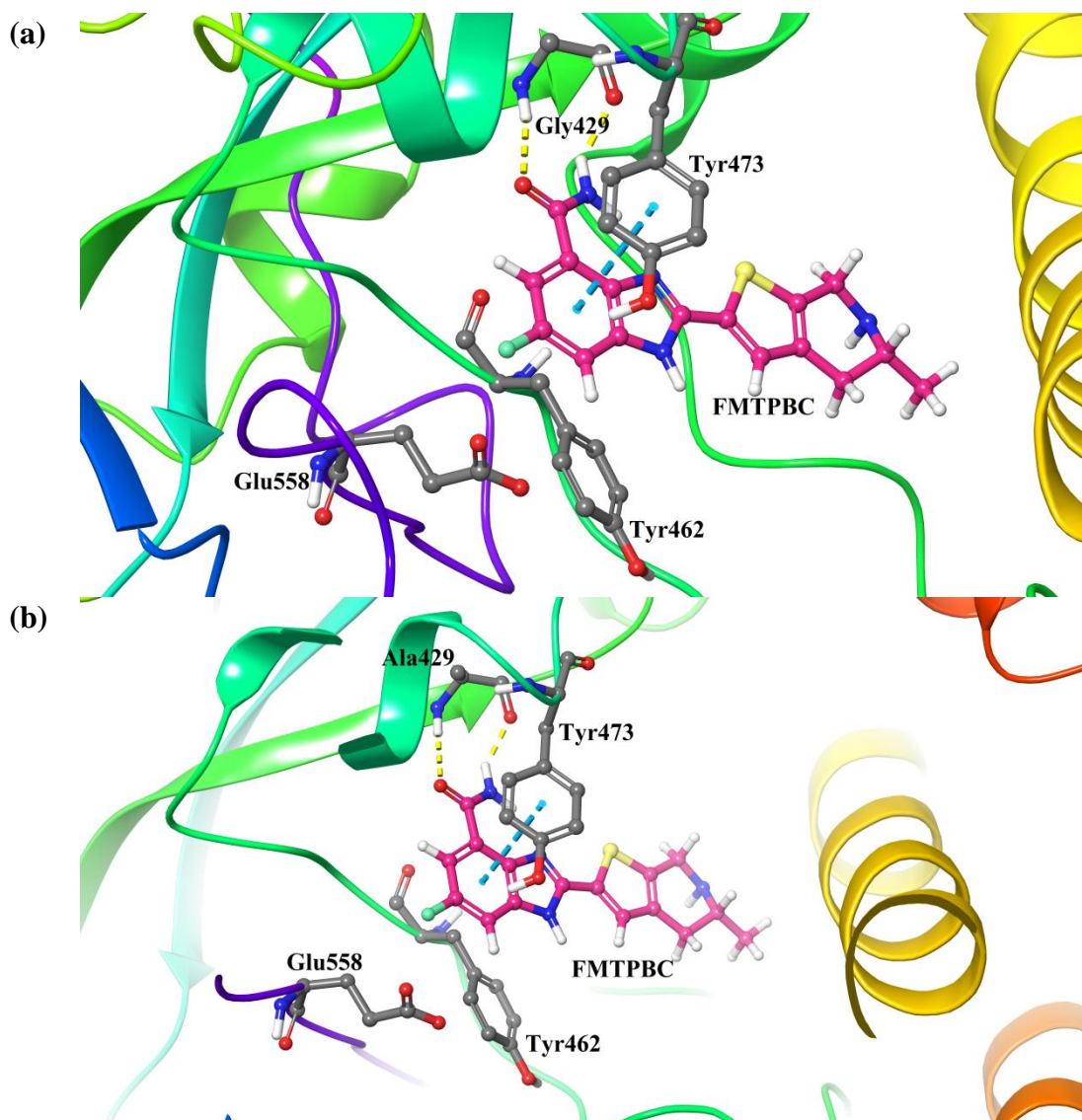


Figure 5.1: Free energy change ($\Delta\Delta G = \Delta G_{\text{wild}} - \Delta G_{\text{variant}}$) obtained from alanine substitution for PARP-2 enzyme. Negative $\Delta\Delta G$ indicate unfavourable substitution of alanine at the respective position.

CAS studies on PARP-2 enzyme shows that the alanine substitution at each residue site mostly decreases the stability of enzyme variant. The key amino acid residues contributing to the binding energy for effective PARP-2 inhibition by drug molecules are Tyr473 > Tyr462 > His428 > Ser470 > Tyr455 > Glu558 > Gly429 > Gln332 (Venugopal et al. 2021; Wang et al. 2020). Among these residues, the alanine substitution of Gln332, Ser470 and Tyr473 has increased the variant stability (0.42, 0.79 and 0.56 kcal/mol respectively); whereas the alanine substitution of His428, Gly429, Tyr462 and Glu558 has decreased the variant stability (-0.51, -0.6, -0.82 and -0.59 kcal/mol respectively) (**Appendix XVI**). Since we focus on the π stacking interactions formed by aromatic residues, Tyr462 and Tyr473 were selected for further analysis. From the literature it is found that Gly429Ala variant cause breast cancer (Dingerdissen et al. 2018) and Gly429 is responsible for stabilizing inhibitor through hydrogen bonding. Similarly, it is reported that Glu558Ala variant abolishes PARP-2 activity without affecting the localization to DNA damage sites (Riccio et al.

2016) and Glu558 is responsible for water-mediated hydrogen bond with the inhibitor (Venugopal et al. 2021). Therefore, Gly429 and Glu558 residues which contribute for electrostatic interactions were also considered in the mutation study. Therefore, four PARP-2 variants were generated by substituting alanine at the binding site residues individually. Thus the binding site of the four variants differs by a single amino acid change (**Figure 5.2**).



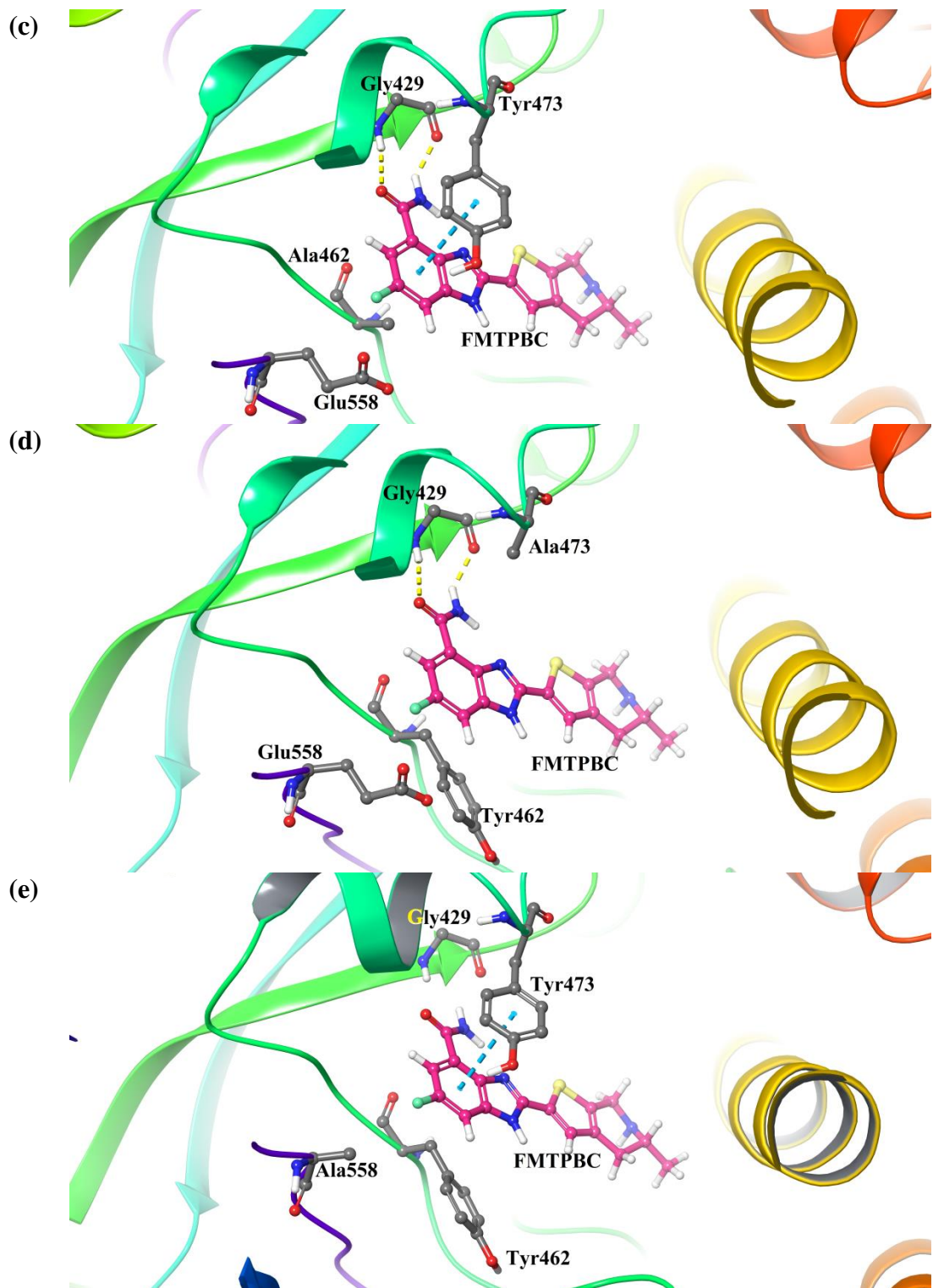


Figure 5.2: Schematic representation of wild enzyme and four variants generated by single amino acid substitution with FMTPBC: (a) Wild, (b) Gly429Ala, (c) Tyr462Ala, (d) Tyr473Ala and (e) Glu558Ala.

5.3 LIGAND PREPARATION AND MOLECULAR DOCKING

The compound, (R)-6-fluoro-2-(5-methyl-4,5,6,7-tetrahydrothieno[2,3-c]pyridin-2-yl) 1H-benzo[d]imidazole-4-carboxamide (FMTTPBC) (**Figure 5.3**), with IC₅₀ value 14 μM from 32 experimentally tested benzimidazole carboxamide derivatives (Chen et al. 2018) was chosen for the present study. The structure of the ligand was constructed using builder panel in Maestro (Schrödinger Release 2018). Partial charges were assigned and the possible ionization states were generated at pH of 7.4 to mimic the experimental conditions. Further, the geometry of the compound was optimized by semi-empirical PM3 (Stewart 1989) and then by B3LYP/6-31G(d,p) level (Ditchfield et al. 1971; Lee et al. 1988) respectively using Gaussian09 package (Frisch et al. 2009). The optimized structure was used for docking studies.

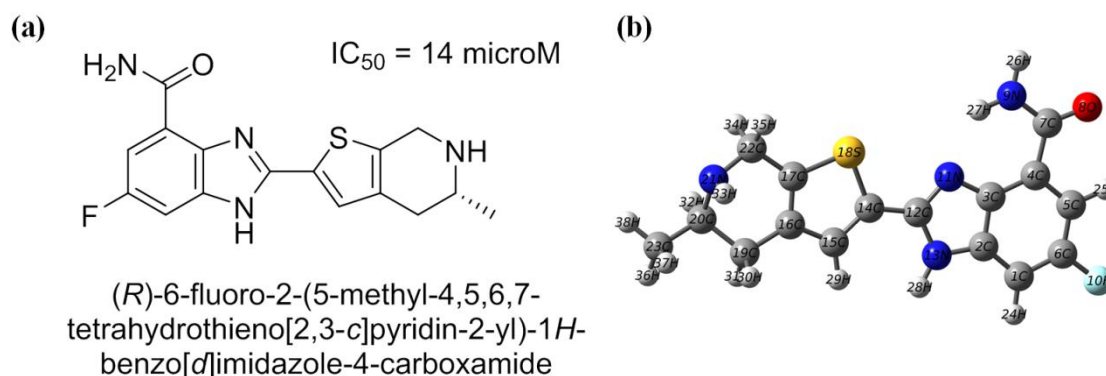
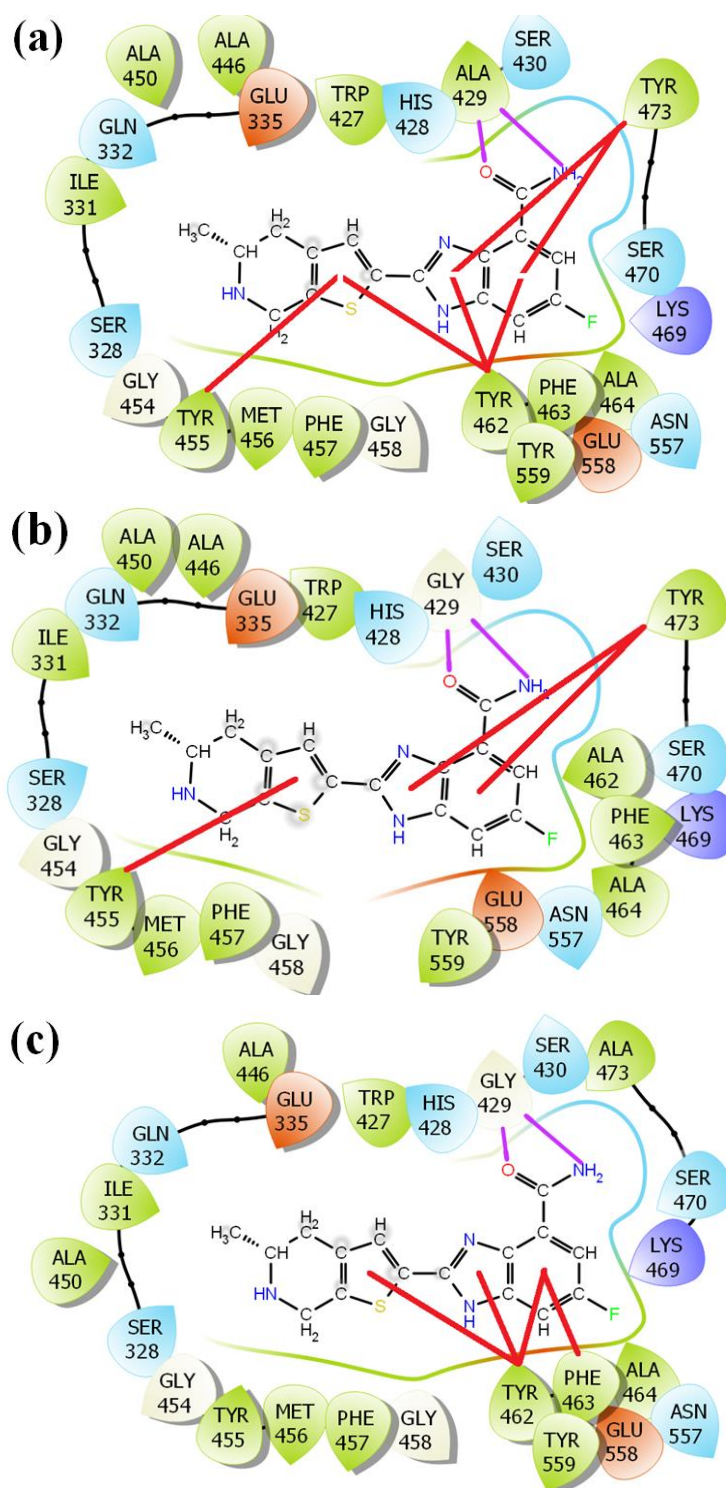


Figure 5.3: (a) 2D-Structure of (R)-6-fluoro-2-(5-methyl-4,5,6,7-tetrahydrothieno [2,3-c]pyridin-2-yl) 1H-benzo[d]imidazole-4-carboxamide (FMTTPBC) (b) Optimized 3D-Structure from DFT.

The co-crystal structure of PARP-2 (PDB ID: 4ZZY, resolution 2.2 Å) (Pape et al. 2015) was retrieved from Protein Data Bank and the docking of FMTTPBC was carried out using Autodock (v4.2.6) (Morris et al. 2009). Four PARP-2 variants were generated namely, Gly429Ala, Tyr462Ala, Tyr473Ala and Glu558Ala, using Pymol Mutagenesis Wizard (Schrödinger, LLC 2015). Gasteiger charges (Gasteiger and Marsili 1980) and polar hydrogen were added to the resulted protein structure using Autodock graphical user interface. A 3D grid was created at the catalytic domain (aa 204-559) (Amé et al. 1999 p. 2) of the variants having a size of 38 Å × 30 Å × 30 Å with a default spacing of 0.375 Å, using Auto-Grid algorithm. The docking

simulations were performed using Lamarckian Genetic Algorithm (Morris et al. 1998). The protein is considered to be rigid during docking. The initial population, the number of energy evaluations and the number of GA runs were fixed to 150, 2.5×10^6 and 100 respectively for all docking simulations.



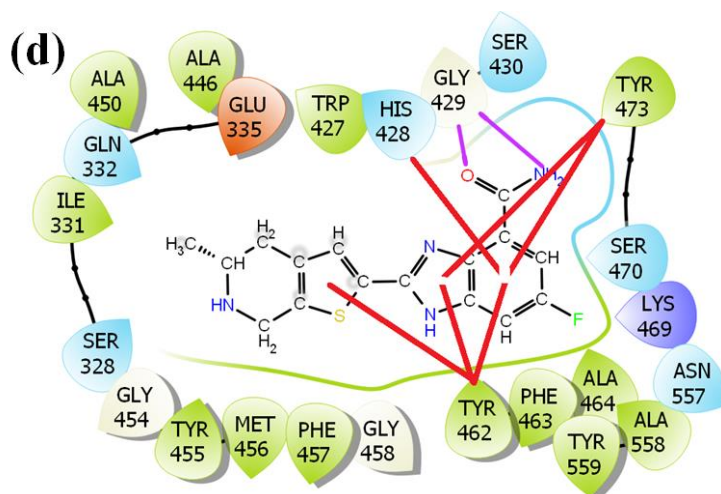
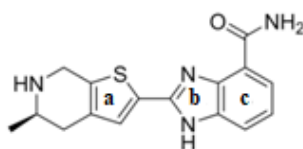


Figure 5.4: 2D-Ligand interaction diagram of FMTTPBC with four PARP-2 enzyme variants. The non-covalent interactions: hydrogen bond interactions (pink) and π - π interactions (red) with the amino acid residues are shown. (a) Gly429Ala, (b) Tyr462Ala, (c) Tyr473Ala and (d) Glu558Ala.

The ligand interactions resided in the catalytic pocket of wild PARP-2 enzyme (429-559 aa) were found to be hydrogen bonding, π -stacking interactions and hydrophobic interactions. The substrate, nicotinamide found to have hydrogen bonding with Gly429 and π - π stacking with Tyr473. Similarly, the complexes of enzyme variant and the ligand are stabilized by hydrogen bonding, π - π interactions and hydrophobic interactions (**Figure 5.4**). It is found that in all the complexes, the tetrahydrothienopyridinyl group showed hydrophobic interactions with Ile331, Ala446, Tyr455, Met456 and Phe457 and the fluoride group with Trp427, Phe463 and Ala464. The ligand rings showed π -stacking possibilities with Tyr455, Tyr462 and Tyr473 residues. In Gly429Ala variant, two hydrogen bond interactions were observed which is similar to the native glycine residue. This suggests that the substitution of glycine with alanine do not significantly alter the binding of ligand at the binding site of Gly429Ala variant. All other three variants, namely Tyr462Ala, Tyr473Ala and Glu558Ala showed two hydrogen bond interactions with Gly429 residue despite of the mutation. In Tyr462Ala and Tyr473Ala variants, the mutated alanine residue showed hydrophobic interaction with fluoride group and carboxamide group of ligand, respectively. In Glu558Ala variant, the negatively charged nature of amino acid residue, Glu, has been changed to hydrophobic nature.

The detailed analysis of ligand orientation and various non-covalent interactions at the binding pocket of each PARP-2 variant are given in **Table 5.1**. It can be seen that the absence of aromatic residues Tyr462 and Tyr473 in Tyr462Ala and Tyr473Ala variants resulted in fewer π -stacking interactions. However, the absence of the residues Gly429 and Glu558 responsible for hydrogen bonding interactions does not affect the formation and stability of protein-ligand complex. In Gly429Ala variant, it is found that -NH_2 group of Ala429 residue forms weak hydrogen bonds between the >C=O of ligand as compared to Gly429 in other three variants. The -NH_2 group of ligand forms similar hydrogen bond at a distance around 2.6 Å with the >C=O of 429th residue in each variant despite the alanine substitution. From Table 1, it is found that the residues Tyr462 and Tyr473 form strong π -stacking interactions with the ligand at a distance range of 4.5-6.0 Å. Other aromatic residues such as Tyr455, His428 and Phe463 form weak π -stacking interactions at a distance range of 6.3-7.5 Å, which are likely to interrupt in biological conditions.

Table 5.1: Various non-covalent interactions observed between PARP-2 variants and ligand, FMTTPBC



Gly429Ala	Tyr462Ala	Tyr473Ala	Glu558Ala
Hydrophobic Interactions			
Ile331, Trp427, Ala429 , Ala446, Tyr455, Met456, Phe457, Tyr462, Phe463, Ala464, Tyr559	Ile331, Trp427, Ala446, Tyr455, Met456, Phe457, Ala462 , Phe463, Ala464	Ile331, Trp427, Ala446, Tyr455, Met456, Phe457, Tyr462, Phe463, Ala464, Ala473	Ile331, Trp427, Ala446, Tyr455, Met456, Phe457, Tyr462, Phe463, Ala464, Ala558
Hydrogen Bond Interactions			
Ala429	Gly429	Gly429	Gly429
$\text{>C=O}_{\text{lig}} \text{ --- HN}_{\text{Ala}}$ 2.33 Å	$\text{>C=O}_{\text{lig}} \text{ --- HN}_{\text{Gly}}$ 2.16 Å	$\text{>C=O}_{\text{lig}} \text{ --- HN}_{\text{Gly}}$ 2.18 Å	$\text{>C=O}_{\text{lig}} \text{ --- HN}_{\text{Gly}}$ 2.18 Å

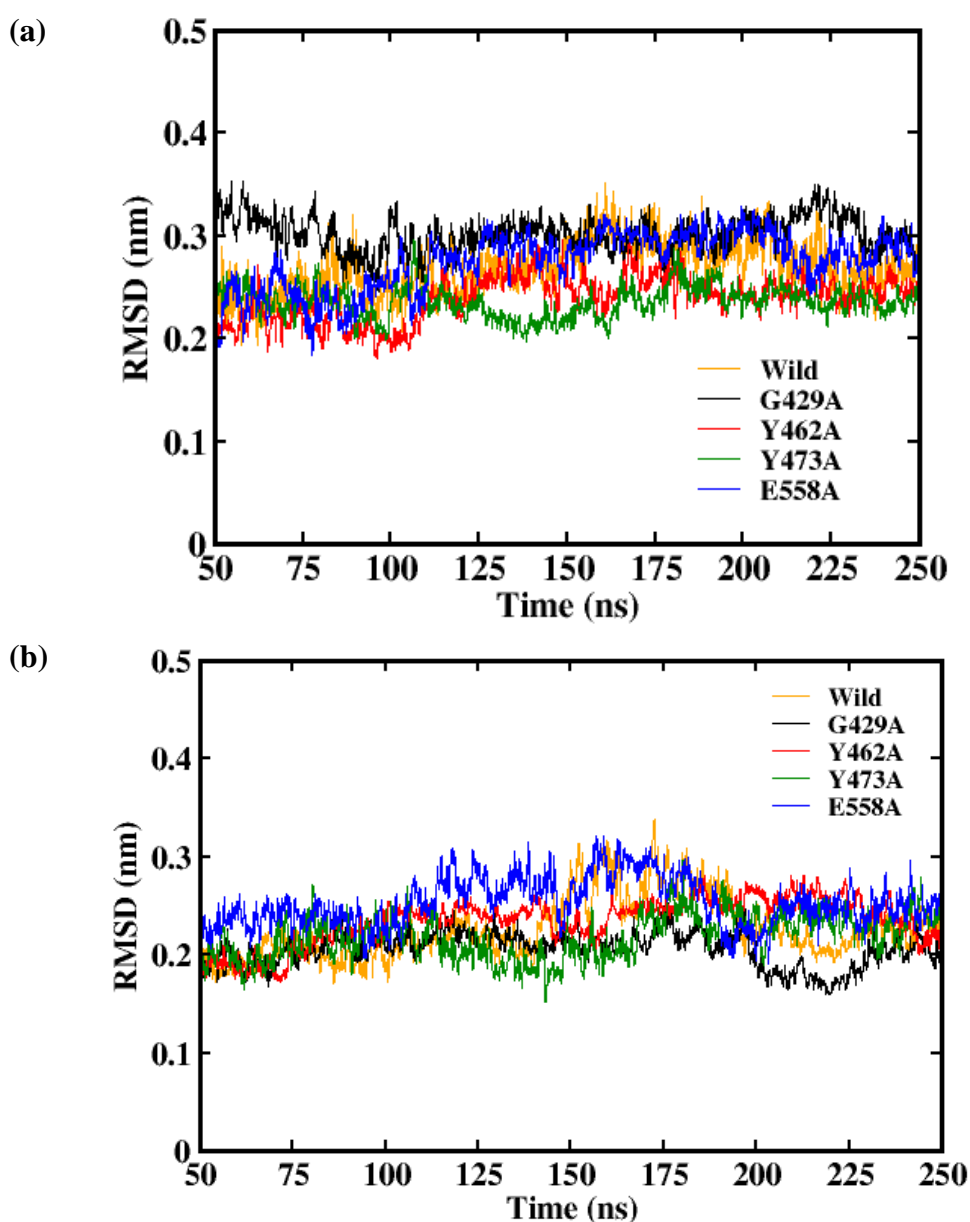
>C=O _{Ala} --- HN _{lig} :	>C=O _{Gly} --- HN _{lig} :	>C=O _{Gly} --- HN _{lig} :	>C=O _{Gly} --- HN _{lig} :
2.67 Å	2.63 Å	2.63 Å	2.58 Å
π-π Stacking Interactions			
a-Tyr455 : 6.34 Å a-Tyr462 : 4.84 Å a-Tyr473 : 6.19 Å b-Tyr462 : 4.61 Å b-Tyr473 : 4.74 Å c-His428 : 6.42 Å c-Tyr462 : 5.73 Å c-Phe463 : 7.39 Å c-Tyr473 : 5.12 Å	a-Tyr455 : 6.44 Å a-Tyr473 : 6.06 Å b-Tyr473 : 4.59 Å c-His428 : 6.39 Å c-Phe463 : 7.46 Å c-Tyr473 : 5.02 Å	a-Tyr455 : 6.33 Å a-Tyr462 : 4.88 Å b-Tyr462 : 4.63 Å c-His428 : 6.43 Å c-Tyr462 : 5.75 Å c-Phe463 : 7.44 Å	a-Tyr455 : 6.43 Å a-Tyr462 : 4.81 Å a-Tyr473 : 5.98 Å b-Tyr462 : 4.59 Å b-Tyr473 : 4.97 Å c-His428 : 6.47 Å c-Tyr462 : 5.73 Å c-Phe463 : 7.41 Å c-Tyr473 : 4.97 Å

***a,b,c** represents the three ligand aromatic rings which can form π -stacking interactions with the aromatic amino acid residues at the catalytic domain of the PARP-2 enzyme.

Next, the docking results were evaluated using MD simulations to get insights about the stability and energy requirements. 250 ns MD simulations were performed to predict the stability of the complex formed between PARP-2 enzyme variants and the ligand FMTNBC. All four simulations for the enzyme complex were found to be converged at this time scale. Further, 250 ns simulations were performed for four PARP-2 variants without ligands to check the stability upon single amino acid substitution. Another set of simulations each for the complex and the variant enzyme were carried out for reproducibility and reliability of the results. All the ligand-bound protein simulations were performed separately by GROMACS 2018 (Abraham et al. 2015) using SPC/E water model (Hess and van der Vegt 2006). The topology was generated using AMBER99SB force field (Hornak et al. 2006). The complex was solvated in a periodic cubic box with distance between the enzymes and the box boundary set to 1.0 nm in all directions. The charge of the system is neutralized by adding one Na⁺ ion. The energy minimization was done by steepest descent algorithm and 10 ns each NVT and NPT equilibration were performed. Finally, a production run of 250 ns was performed by removing restraint to relax the system.

5.4 PROTEIN STRUCTURAL STABILITY

Root Mean Square Deviation (RMSD): The C_{α} RMSD profile for the unbound variants Gly429Ala and Glu558Ala showed higher fluctuations during the course of simulation compared to wild enzyme. This suggests that the unbound variants showed decreased stability compared to the wild enzyme as obtained from CAS results. The variants Tyr462Ala and Tyr473Ala showed fewer fluctuations after 150 ns, which suggest its increased stability (**Figure 5.5 (a)**).



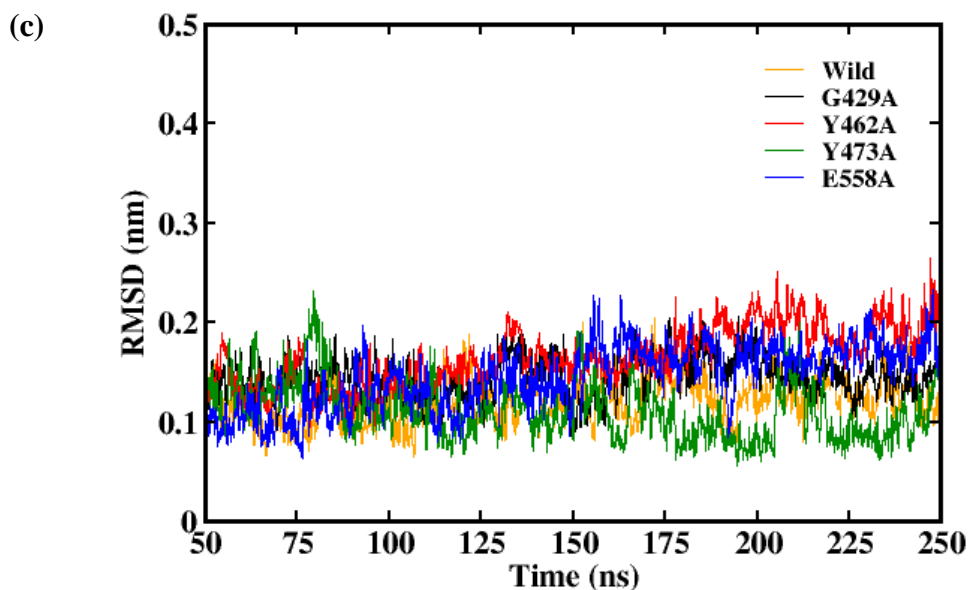


Figure 5.5: RMSD profile of protein backbone C_{α} atoms for last 200 ns simulation trajectory (a) unbound wild and variants (b) bound wild and variants (c) ligand with respect to protein.

The flexible nature of bound enzyme is drastically reduced compared to the unbound enzyme. Higher fluctuations were observed for C_{α} atoms on the backbone of protein variants due to the ligand binding compared to the wild enzyme (**Figure 5.5 (b)**). The significant structural changes are found in Tyr462Ala, Tyr473Ala and Glu558Ala variants after 200 ns trajectory, which may be due to the weak non-covalent interactions of mutant alanine residue with the ligand. The bound Gly429Ala variant is found to fluctuate lesser than the wild enzyme complex, indicating that the hydrogen bond interactions with the ligand are not hampered due to substitution.

Even though, the ligand RMSD converges at a higher value as compared to wild for Tyr462Ala and Gly429Ala variants, the movement of ligand copes well with the protein (**Figure 5.5 (c)**). For Glu558Ala variant, the ligand movement stabilises after 130 ns of the simulation. In case of Tyr473Ala, the protein-ligand interactions are not well-maintained from the beginning of simulation due to the lack of aromatic interactions. This correlates with the docking results that Tyr473 forms multiple aromatic interactions with the three aromatic rings of the ligand.

Root Mean Square Fluctuation (RMSF): The highest value for fluctuations were found in residues Phe313-Thr317, Glu350-Glu355, Pro392-Ser395, and Leu547-

Thr553 which resides away from the catalytic domain of unbound PARP-2 variants as similar to wild enzyme (**Figure 5.6 (a)**). The binding site residues 429-473 aa has fewer fluctuations in Gly429Ala variant compared to the other variants. The ligand bound variants showed a similar observation as in the case of unbound variants. The fluctuations at binding site residues have been reduced in the variants upon ligand binding (**Figure 5.6 (b)**). Among the variants, Tyr473Ala variant showed higher fluctuations at the binding site.

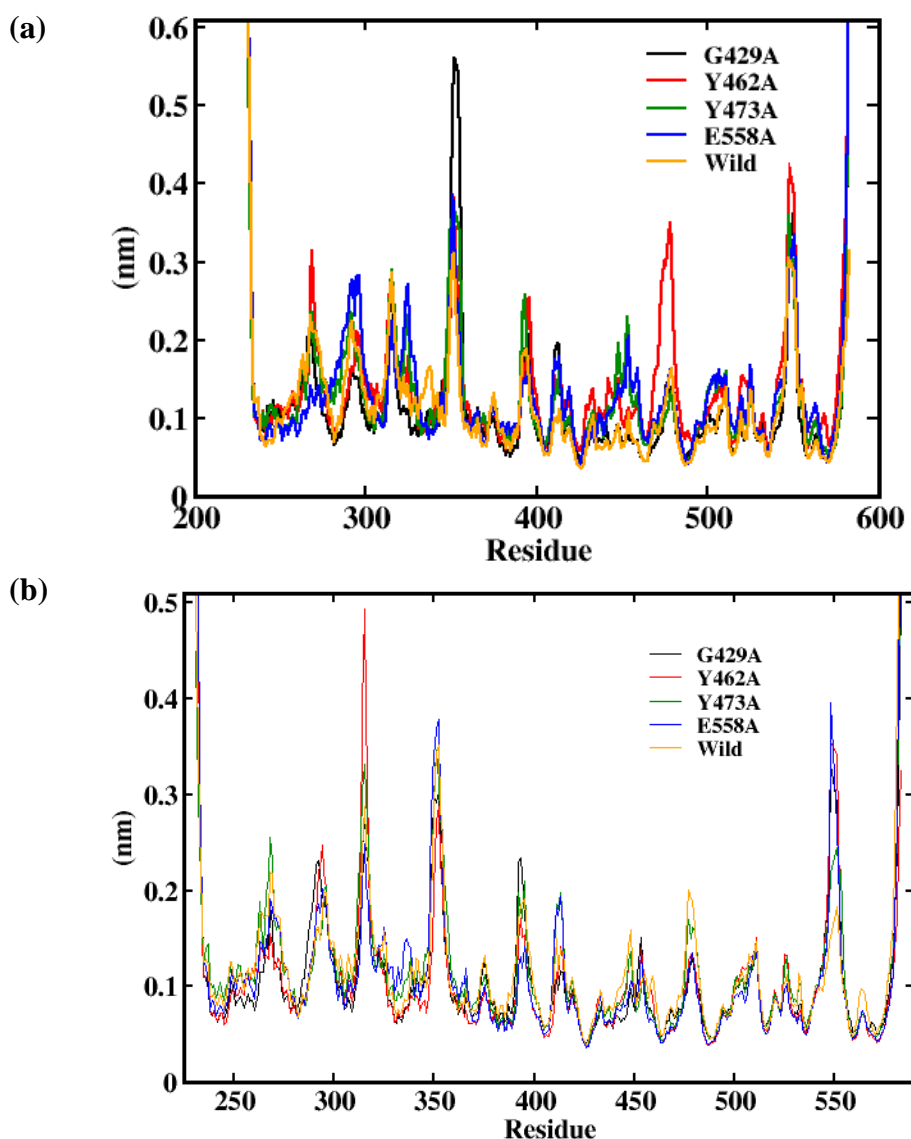


Figure 5.6: Backbone RMSF fluctuations of (a) unbound enzymes and (b) ligand bound enzymes.

Next, it will be interesting to know whether the fluctuations caused changes in the secondary structure of PARP-2 enzyme based on the geometric parameter such as dihedral angle. Dihedral angles (ϕ , ψ) are widely used to characterize the changes in the backbone of protein residues. The deviations in ϕ , ψ angles provide insight into the changes in the secondary structure elements in the protein. **Appendix XVII** shows the free energy landscape (FEL) of dihedral angles (ϕ vs. ψ) for the four ligand bound PARP-2 enzyme variants. The FEL has two well separated regions which belong to α -helical and β -sheet structure. The amino acid residues at position 429, 462, 558 belong to β -sheet region and position 473 belongs to α -helix region. Even though, there is a significant change in the distribution of ϕ angles, a narrow distribution is observed for ψ angles which suggest that there are no considerable changes in the secondary structure of the enzyme upon alanine substitution.

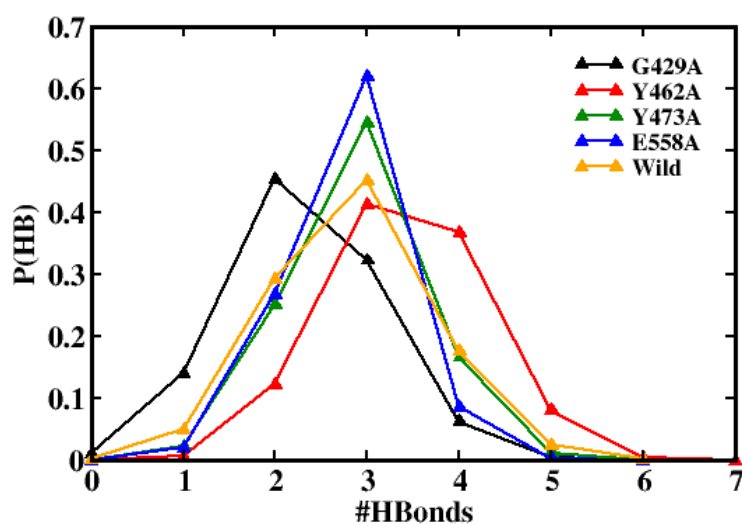
It can be observed that the Gly429Ala, Tyr462Ala and Glu558Ala variants show similar trend in the angle deviation and the corresponding energy minima. These deviations help in the favourable binding of ligand at the cavity of the enzyme through conformational transitions. The alanine substitution at Tyr473 decreased the possibility of transitions as higher energy regions are more in FEL compared to stable regions. It is equally important to understand the geometric changes occurring in the non-covalent interactions during the simulation to reveal the importance of polar and aromatic groups in PARP-2 inhibition.

5.5 IMPORTANCE OF HYDROGEN BONDS

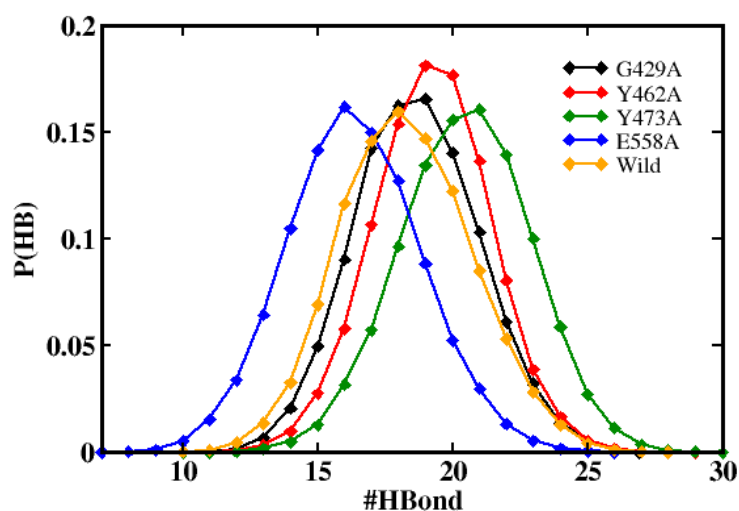
Hydrogen bond interactions play an important role in stabilizing the protein-ligand complex apart from aromatic interactions. The hydrogen bond distribution formed between protein-ligand, ligand-water and cavity-water in wild and variant complexes is shown in **Figure 5.7**. It is found that Gly429Ala, Tyr462Ala, Tyr473Ala and Glu558Ala variants showed maximum occurrence of 6, 7, 5 and 6 hydrogen bonds respectively with the ligand. The average number of hydrogen bonds formed between the ligand and the protein is found to be 3 for all the cases similar to that of wild enzyme except in the Gly429Ala variant which forms 2 hydrogen bonds. The average number of hydrogen bonds formed between ligand and water is found to be 2 which is

lesser in probability compared to wild enzyme. The ligand-water hydrogen bond distribution suggests that the ligand is more solvated in Tyr473Ala variant. The higher solvation of molecule may hamper the protein-ligand interaction in Tyr473Ala variant. The number of cavity-water hydrogen bonds directly implies the solvation of the binding site of the enzyme. It is found that the cavity of Tyr473Ala variant is more solvated whereas the cavity of Glu558Ala variant is less solvated compared to the wild enzyme. Other variants, Gly429Ala and Tyr462Ala showed almost similar trend as the wild enzyme.

(a) Protein-Ligand



(b) Cavity-Water



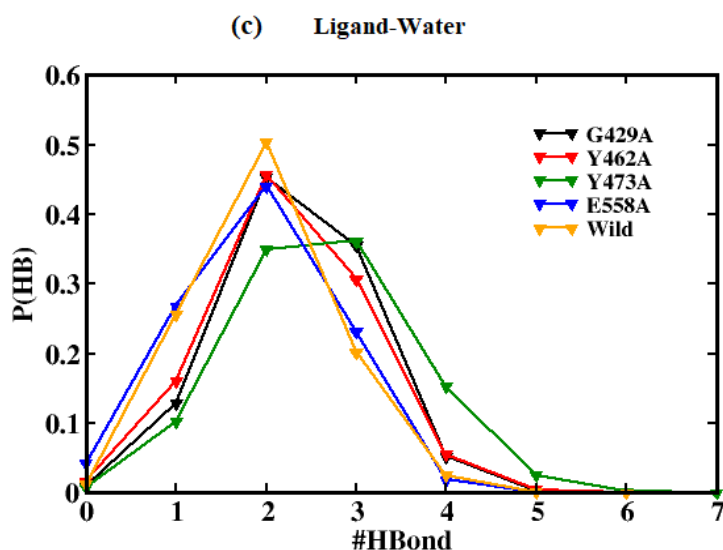


Figure 5.7: Occurrence of hydrogen bond interaction (a) protein-ligand (b) cavity-water and (c) ligand-water.

Further, additional hydrogen bond formation between the residues and the ligand was determined from the MD trajectory (**Table 5.2**). Apart from the Ala429/Gly429 amino acid residue, Ser470 forms stable hydrogen bond with the ligand. Other residues such as Tyr455 and Gln332 form weak hydrogen bonds with the ligand due to its dynamic nature at the binding pocket. The ligand forms water-mediated hydrogen bonds with the residues at the binding site of enzyme variants. Gly429Ala, Tyr462Ala and Tyr473Ala variants showed higher number of water bridges. The highest number of water bridges formed in case of Tyr473Ala variant supports the solvation of both the binding site and the ligand.

Table 5.2: Hydrogen bond occupancy (%) and water bridges between the enzyme variants and the ligand at the catalytic domain of PARP-2

System	Hydrogen Bond Occupancy (%)		Water Bridges
Wild	Lig (H) --- Gly429 (O)	80.0	Glu558 --- water --- Lig (H)
	Lig (H) --- Tyr455 (OH)	17.1	
	Ser470 (H) --- Lig (O)	94.7	
	Gly429 (H) --- Lig (O)	61.3	
	Gly429 (H) --- Lig (N)	12.2	
	Gln332 (H) --- Lig(N)	10.6	

Gly429Ala	Lig (H) --- Ala429 (O)	85	Ser440 (OG) --- water --- Lig (N)
	Ser470 (H) --- Lig (O)	67.5	Glu335 (OE1/2) --- water --- Lig (N)
	Ala429 (H) --- Lig (O)	55.1	Glu558 (OE1) --- water --- Lig (H)
	Ala429 (H) --- Lig (N)	10.5	
Tyr462Ala	Lig (H) --- Gly429 (O)	93.2	Ala462 --- water --- Lig (N)
	Lig (H) --- Tyr455 (OH)	33.2	Tyr455 (HH) --- water --- Lig (N)
	Ser470 (H) --- Lig (O)	94.0	Glu558 (OE2) --- water --- Lig (H)
	Gly429 (H) --- Lig (O)	76.2	
	Gly429 (H) --- Lig (N)	16.1	
	Gln332 (H) --- Lig (N)	24.9	
Tyr473Ala	Lig (H) --- Gly429 (O)	88.8	Glu335 (OE1) --- water --- Lig (H1)
	Lig (H) --- Tyr455 (OH)	4.8	Glu558 (OE1/2) --- water --- Lig (H)
	Lig (H) --- Ser328 (O)	9.0	Tyr455 (HH) --- water --- Lig (H)
	Ser470 (H) --- Lig (O)	95.7	Tyr462 (H) --- water --- Lig (H)
	Gly429 (H) --- Lig (O)	73.9	
	Gly429 (H) --- Lig (N)	12.4	
	Gln332 (H) --- Lig (N)	3.7	
Glu558Ala	Lig (H) --- Gly429 (O)	93.3	Gly437 (H) --- water --- Lig (H)
	Ser470 (H) --- Lig (O)	93.6	
	Gly429 (H) --- Lig (O)	74.4	
	Gly429 (H) --- Lig (N)	6.1	
	Gln332 (H) --- Lig (N)	4.5	

5.6 IMPORTANCE OF π -STACKING INTERACTIONS

In addition to hydrogen bond interactions, π -stacking interactions are mainly found in PARP-2/inhibitor complexes with the tyrosine, histidine and phenylalanine residues at the binding site. The π -stacking interactions are well-characterized using the angle between normal vectors of two aromatic rings from the ligand and the amino acid residue. The geometric criteria for perfect π -stacking interactions constitutes 4.5-5.5 Å for distance and the angle between normal vectors of two aromatic rings should be in range of 0°-90° (Brylinski 2018). An angle of 0°-15° between the vectors indicates

face-to-face π -stacking and a larger angle near to the right angles indicates edge-to-face π -stacking interactions.

The ligand contains three aromatic rings, namely, 6-membered benzene ring, 5-membered imidazole ring and 5-membered thiophene ring. It is found that the residues Tyr462 and Tyr473 form π -stacking interactions with all the three rings in the ligand. The other residues, His428, Phe463 form π -stacking interaction with the benzene ring and Tyr455 forms the interaction with the thiophene ring. The probability distribution of geometric parameters in each variant is plotted in **Appendix XVIII**. The ligand preferred to form small stacking angles at the binding site of PARP-2 variants with the residues His428, Tyr455, Tyr462 and Tyr473 except Phe463.

Among all the π -stacking interactions formed between the amino acid residues and the ligand, Tyr473 and Tyr462 showed maximum stability based on the geometric criteria. Tyr473 formed face-to-face stacking interactions with all the three rings of the ligand at the distance range of 4.5-5.5 Å except in Tyr473Ala variant. Similar observation is found for the residue Tyr462 at a distance of 4.7-6.1 Å except in Tyr462Ala variant. Even though, His428 and Tyr455 form face-to-face stacking interactions, the attraction between the ligands and the residues are weaker compared to Tyr473 based on the distance range of 5.2-6.8 Å. The distance between the benzene ring and His428 is found to be in the range 6.3-6.7 Å which is slightly higher than the distance criteria for π - π interactions. It is found that in all the variants, weak edge-to-face interactions (near right angles) are formed with Phe463 residue at a distance range 7-7.5 Å. Thus, the strength of π -stacking interactions formed between the amino acid residues and the ligand follows the order: Tyr473 > Tyr462 > Tyr455 > His428 > Phe463. Further, the flexibility of the residues was studied by principal component analysis on the backbone C $_{\alpha}$ atoms of the PARP-2 enzyme upon alanine substitution. Next, it will be interesting to quantify the strength of various non-covalent interactions using free energy calculation and per-residue energy contributions.

5.7 PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) was applied on the backbone C_{α} atoms for all the four enzyme variants as well as variant/ligand complexes to determine the conformational changes which assist the ligand binding to the catalytic pocket. It is known that first few principal components of the system describe the overall motion and conformational changes in the protein. Therefore, first 20 eigenvectors of bound and unbound enzyme variants were considered and plotted in **Figure 5.8**. The first nine eigenvectors account for 70%, 72%, 75% and 73% for unbound Gly429Ala, Tyr462Ala, Tyr473Ala and Glu558Ala variants respectively (**Figure 5.8 (a)**). The unbound Tyr473Ala showed a higher flexible behaviour compared to other variants which agrees with the CAS results.

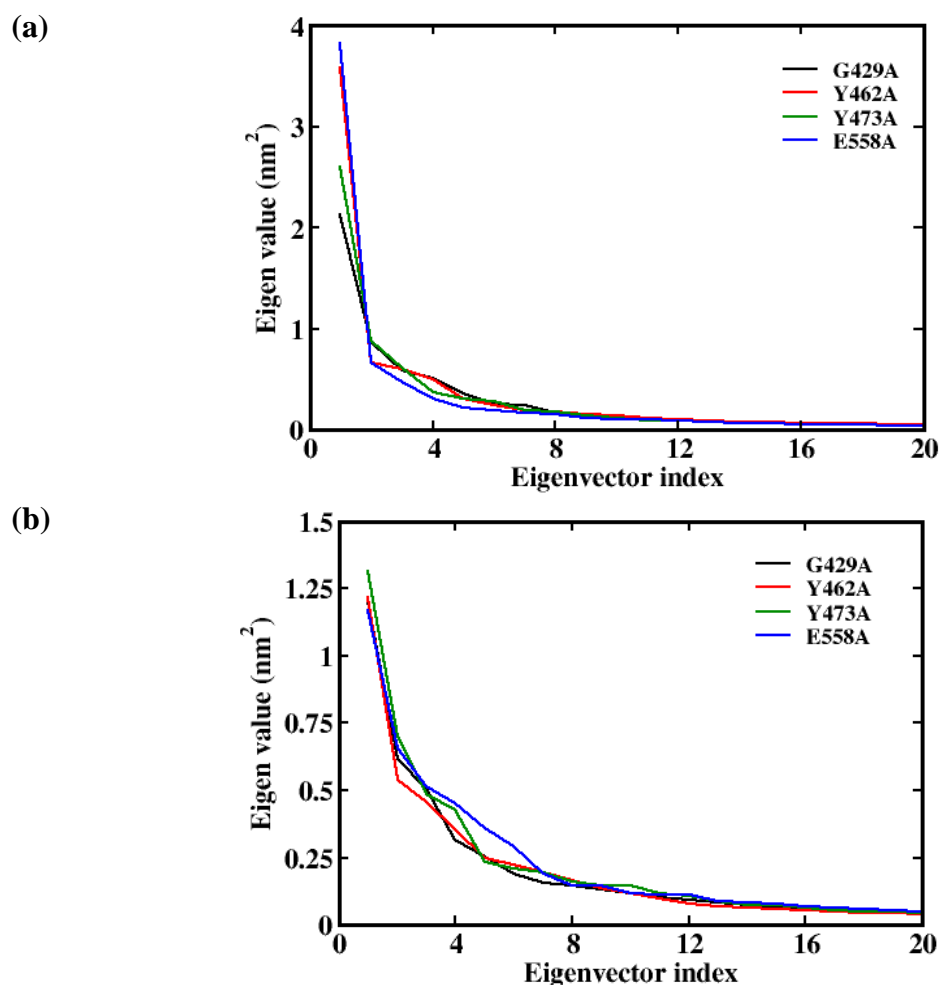


Figure 5.8: Plot of eigenvalue vs. first 20 eigenvector index derived from PCA over 250 ns trajectory for (a) unbound and (b) bound PARP-2 enzyme variants.

For the ligand-bound variants, the first nine eigenvectors contribute for 62%, 63%, 61% and 62% of 250 ns trajectory of Gly429Ala, Tyr462Ala, Tyr473Ala and Glu558Ala variants respectively (**Figure 5.8 (b)**). The decrease in the percentage of correlated motions suggests the protein conformational changes induced by the ligand on binding. It can be found that bound Tyr462Ala variant (63%) showed higher flexible behaviour than bound Tyr473Ala variant (61%). The eigenvalues suggest that favourable correlated motions are observed between the ligand and the adjacent amino acid residues in the catalytic pocket of Tyr462Ala variant. This agrees well with the observed non-covalent interactions between the variants and the ligand.

For better interpretation, the first two principal components (PCs) are plotted in 2D projection plot for both the bound and unbound PARP-2 variants (**Figure 5.9**). It can be seen from PC1-PC2 projection plot that all the variants spread over the phase space than the variant-ligand complexes indicating less correlated motions with the adjacent amino acid residues in variants (**Figure 5.9(a-d)**). This suggests that the correlated motions of each unbound protein variants are highly flexible in nature. The binding of ligand to alanine substituted protein induce different structural changes to the enzyme. The wild-ligand complex spread over the phase space more than the variant complexes (**Figure 5.9 (e)**), indicating that the binding site residues of wild enzyme gain flexibility upon ligand binding. The formation of single dense cluster that spreads around the average structure taken for PCA analysis in Gly429Ala, Tyr462Ala and Glu558Ala variants suggests that the complex undergoes transitions to achieve most stable conformations due to the structural flexibility of the variants. In Tyr473Ala variant, two dense clusters are observed on either sides of the average structure which indicates the presence of two stable independent conformations. However, the ligand bound variants are more compact than the unbound variants and the ligand bound wild enzyme.

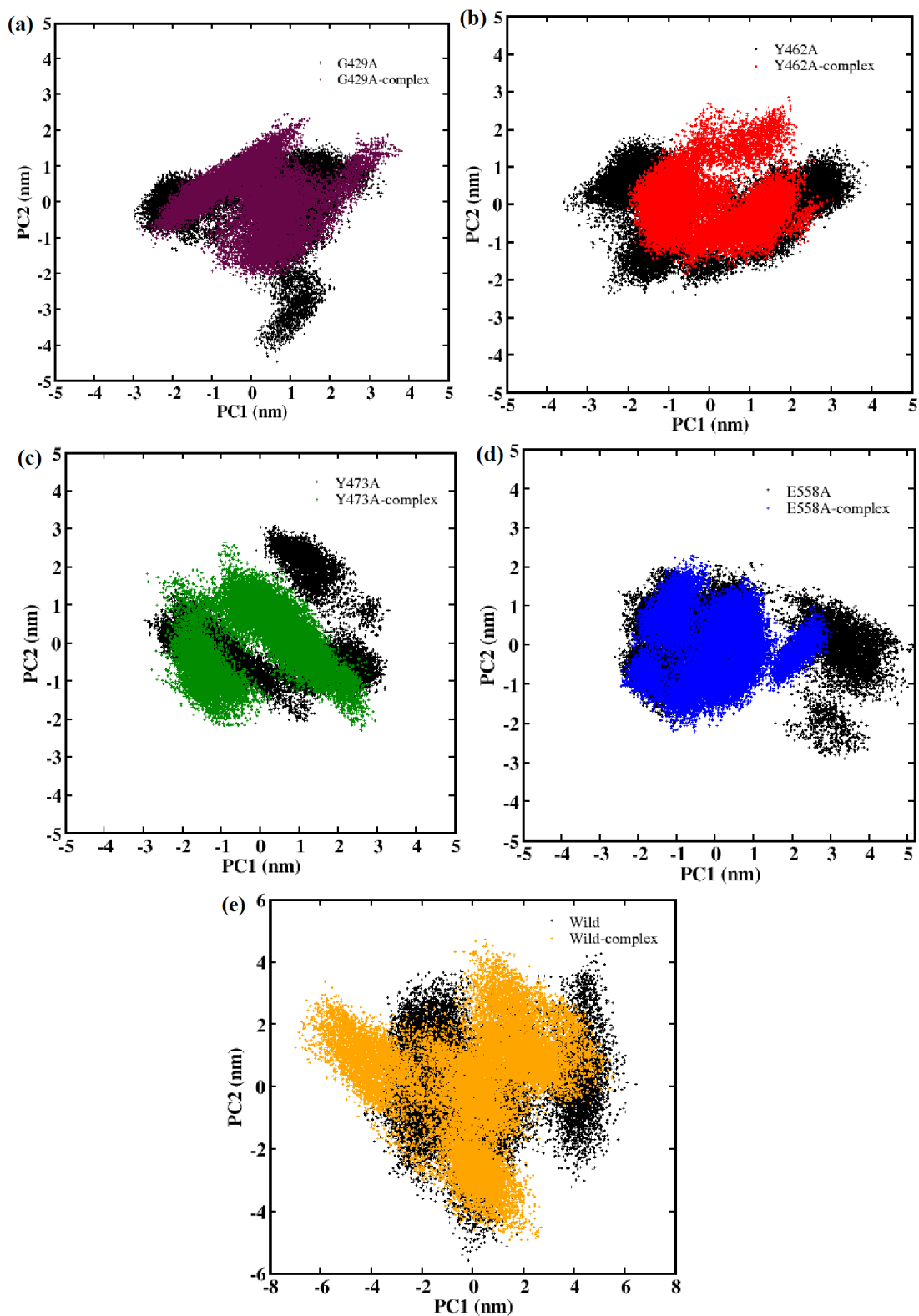


Figure 5.9: First two principal components showing the major motions of protein spread over the phase space. (a) Gly429Ala, (b) Tyr462Ala, (c) Tyr473Ala, (d) Glu558Ala and (e) Wild.

The cosine content of the first three principal components (PCs) for the protein motions is given in **Appendix XIX**. The cosine value (0 to 1) of the principal components indicates whether the simulation trajectory has converged for further sampling of conformational changes in the protein (Khan et al. 2017). The lower cosine value of principal component indicates well-sampled data. The cosine value of PC1 on backbone C- α atoms for the protein-ligand complex of Gly429Ala, Tyr462Ala and Glu558Ala variants are found to be 0.217, 0.416 and 0.326 respectively which suggests that the simulations are reliable to explain the ligand binding at the catalytic pocket. The cosine value of 0.789 for the bound Tyr473Ala variant suggests the randomness of conformations during the simulation. All the unbound variants showed a cosine value around 0.7, indicating that the enzyme achieved a stable conformation after ligand binding except for Tyr473Ala variant. This results correlates well with the solvation of cavity residues observed from the hydrogen bond interactions with water.

Next, the fluctuations in the protein binding site are shown using porcupine plots which show the magnitude and direction of principal components associated with the vectors. The catalytic domain of PARP-2 enzyme contains a β - α -loop- β - α signature motif which constitutes the residues Asp410- Tyr462. In Gly429Ala variant, it is found that mode 2 has larger fluctuations at the loop region and the binding pocket of the protein compared to mode 1. This suggests that the substitution of glycine with alanine do not contribute significantly to the binding of ligand at the binding site of Gly429Ala variant. In mode 3, major fluctuations were observed over the β -sheet region in the variant (**Figure 5.10**). In Tyr462Ala variant, mode 1 showed larger fluctuations at the binding site and loop region than mode 2 and mode 3 (**Figure 5.11**). In Tyr473Ala variant, all the significant modes showed fewer fluctuations (**Figure 5.12**). Glu558Ala variant showed similar behaviour as that of Tyr462Ala variant (**Figure 5.13**). It is observed that depending on the flexibility of catalytic pocket and the ligand, the structural robustness protein-ligand complex can vary from very loose to very tight which have functional implications based on the nature of inhibition (Majewski et al. 2019). Thus from the eigenvector-values, 2D projection plot and the major modes of motions, it can be found that lower structural

robustness or higher flexibility favours the ligand-binding as in Tyr462Ala variant more effectively whereas robustness of catalytic site decreases the binding ability as in Tyr473Ala variant. Also, it is found that the movement of β - α -loop- β - α motif supports the stability of ligand-bound Tyr462Ala variant.

RMSF of the individual principle components can be used as an effective tool to validate the accuracy of calculated PCs by overlapping the overall RMSF of protein backbone with the residue movements of each component from PCA (Antunes et al. 2019). In bound Gly429Ala variant, the residue contribution to mode 1 and mode 2 is from the binding site region which covers 400-500 residues. A large contribution is observed from the residues 310-360 to all the three modes of motions which reside away from the binding site. In the case of bound Tyr462Ala variant, an equal contribution is observed from the binding site residues to all the three modes. A large contribution to mode 1 is observed around the region 310-330, which may result in hydrogen bonding interactions (Gln332 with occupancy of 24.9%). In Tyr473Ala variant, all the contributing fluctuations reside away from the binding site. In Glu558Ala variant, fewer fluctuations are observed from all the residues. Next, the binding free energy calculations were carried out to check the stability of the complexes.

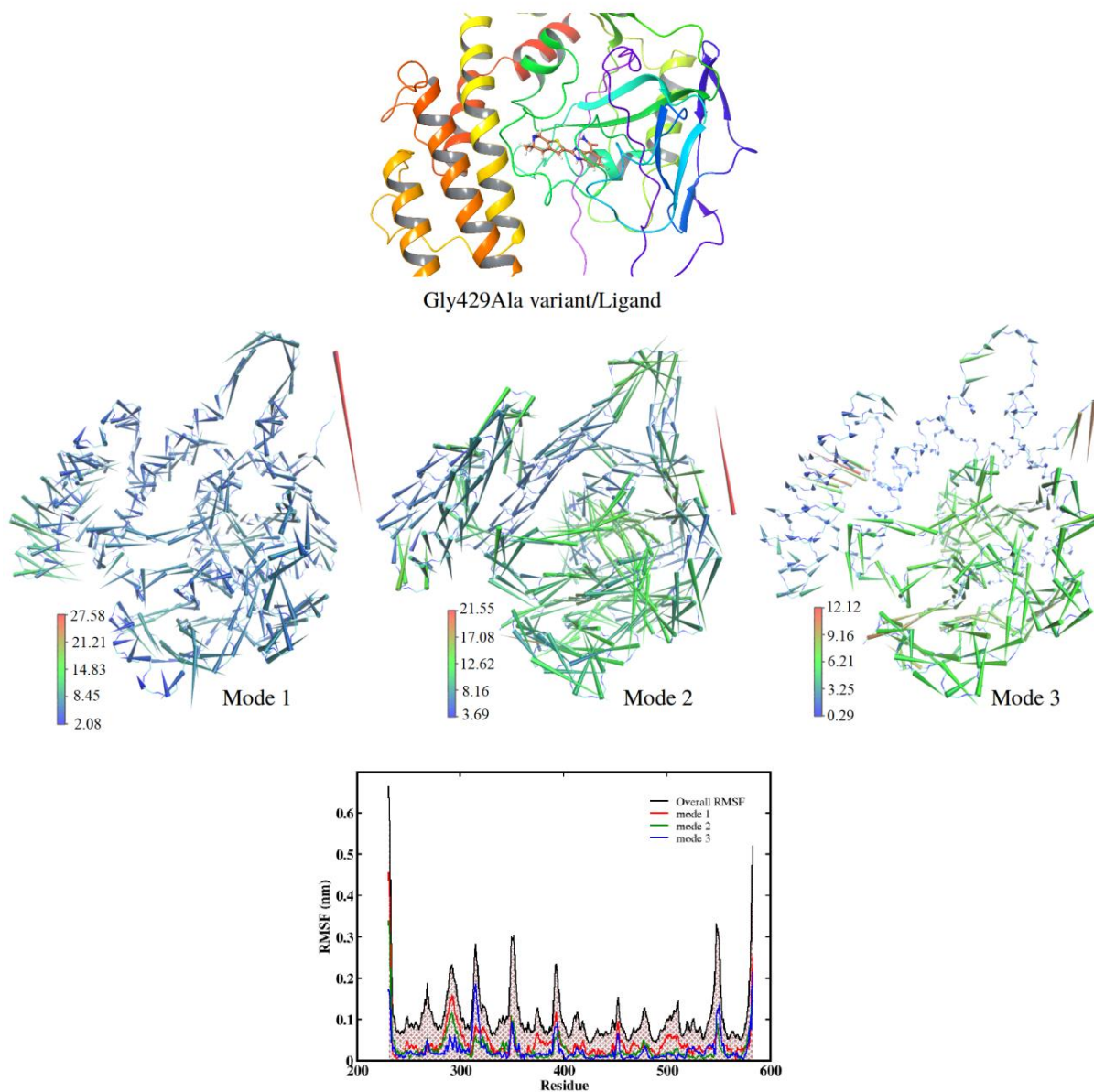


Figure 5.10: Porcupine plots showing prominent motions of enzyme in bound Gly429Ala variant. The protein is represented as backbone with arrow attached to C- α atoms showing the direction and magnitude of prominent motions along mode 1, mode 2 and mode 3. The overall RMSF graph of enzyme in complexes superimposed with the amount of fluctuations of amino acid residues in individual modes (below).

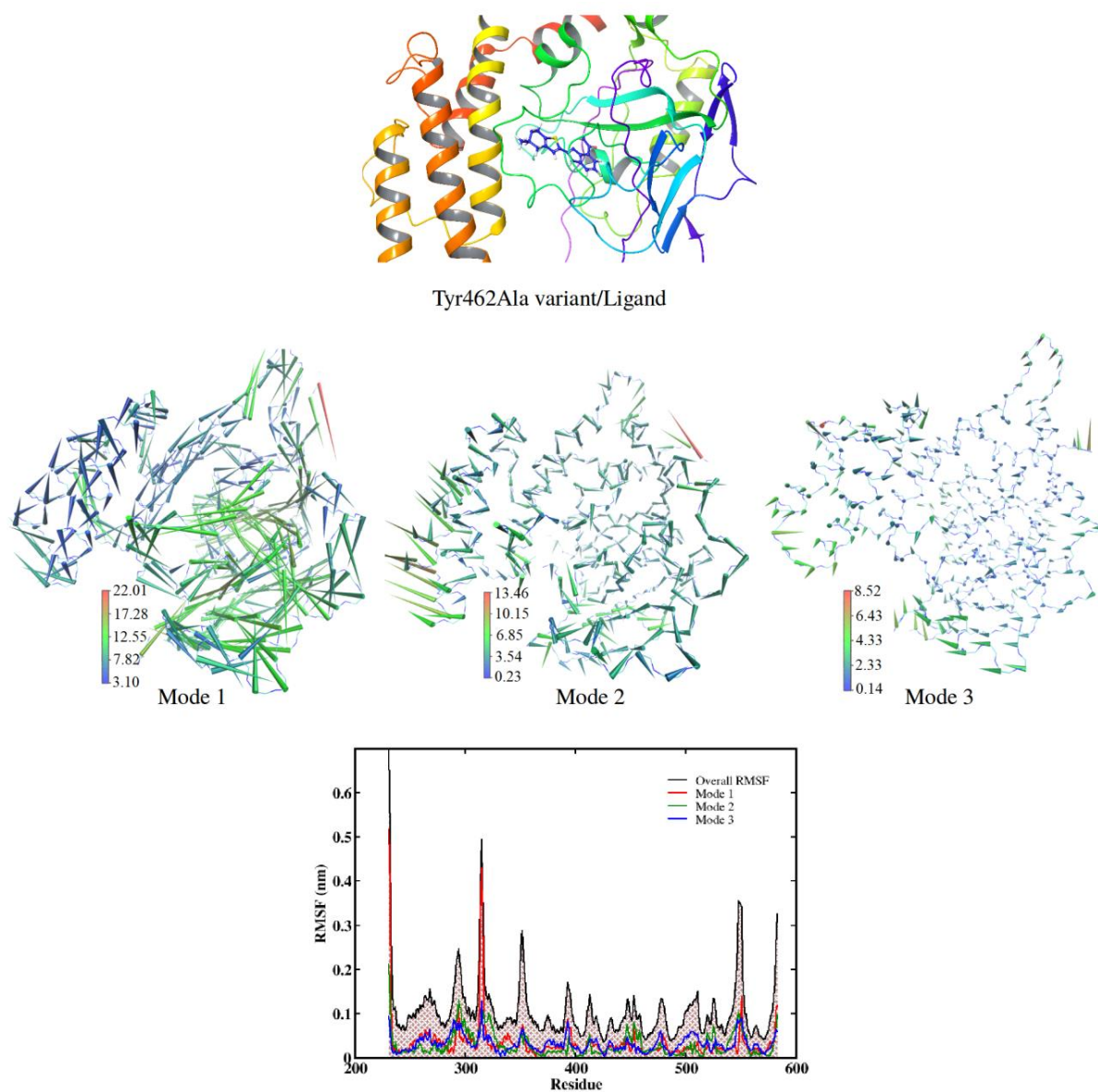


Figure 5.11: Porcupine plots showing prominent motions of enzyme in Tyr462Ala variant complex. The protein is represented as backbone with arrow attached to C- α atoms showing the direction and magnitude of prominent motions along mode 1, mode 2 and mode 3. The overall RMSF graph of enzyme in complexes superimposed with the amount of fluctuations of amino acid residues in individual modes (below).

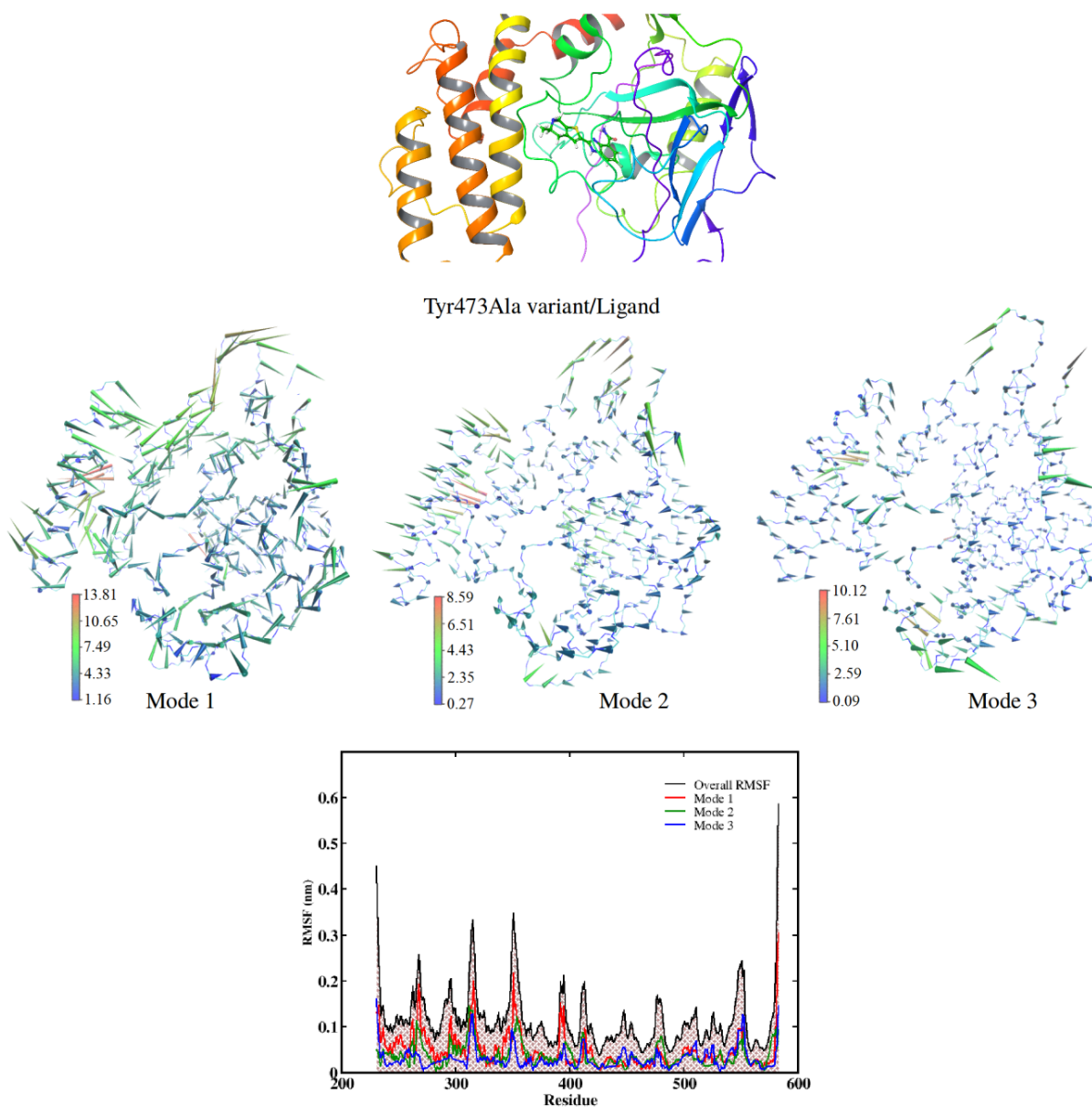


Figure 5.12: Porcupine plots showing prominent motions of enzyme in Tyr473Ala variant complex. The protein is represented as backbone with arrow attached to C- α atoms showing the direction and magnitude of prominent motions along mode 1, mode 2 and mode 3. The overall RMSF graph of enzyme in complexes superimposed with the amount of fluctuations of amino acid residues in individual modes (below).

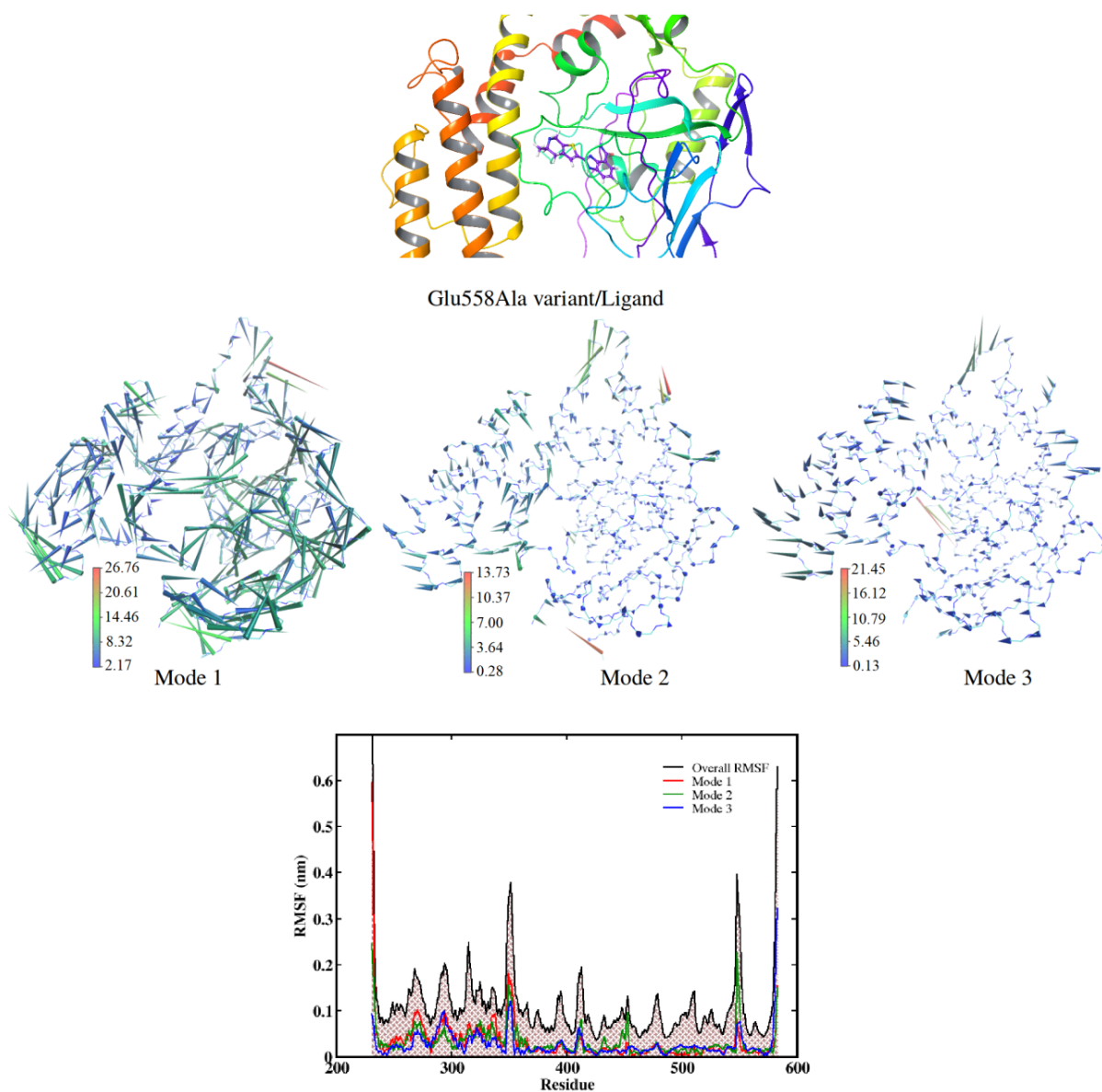


Figure 5.13: Porcupine plots showing prominent motions of enzyme in Glu558Ala variant complex. The protein is represented as backbone with arrow attached to C- α atoms showing the direction and magnitude of prominent motions along mode 1, mode 2 and mode 3. The overall RMSF graph of enzyme in complexes superimposed with the amount of fluctuations of amino acid residues in individual modes (below).

5.8 MM/PBSA ENERGY

The strength of protein-ligand interactions were estimated by MM/PBSA binding energy analysis using default parameters and a low protein dielectric constant, 2. The

binding energy can be decomposed to electrostatic energy, van der Waals energy, SASA energy and polar solvation energy. It is found that the inhibitor can bind to wild more effectively than all the other variants (**Table 5.3**). Among the variants, Gly429Ala, Tyr462Ala and Glu558Ala showed ΔG_{bind} value of -22.59 ± 3.0 , -23.79 ± 2.9 and -21.02 ± 3.6 kcal/mol respectively with almost similar binding pattern with the wild enzyme. A drastic increase in the ΔG_{bind} value (-18.16 ± 3.1 kcal/mol) is observed for Tyr473Ala variant. It can be noted that alanine substitution at Tyr473 greatly affected the contribution of van der Waals energy (-36.04 ± 3.1 kcal/mol) and SASA energy (-3.72 ± 0.2 kcal/mol) to overall binding energy. This may be due to the solvation of ligand which hinders the interaction with the protein. The substitution of alanine at Gly429 (-8.95 ± 2.4 kcal/mol) and Glu558 (-7.60 ± 2.0 kcal/mol) affected the electrostatic contribution to ΔG_{bind} value, due to weak hydrogen bonding interactions. The disruption of water-mediated hydrogen bonds in Glu558Ala variant resulted in the low favourable polar solvation energy. The higher polar solvation energy for other variants is due to the higher percentage of water bridges in these complexes.

Table 5.3: Contribution of energy components to MM/PBSA binding free energy for PARP-2 enzyme variants with FMTTPBC (energy in kcal/mol)

System	van der Waals Energy	Electrostatic Energy	Polar Solvation Energy	SASA Energy	Binding Energy
Wild	-45.67 ± 2.9	-11.72 ± 2.0	36.24 ± 2.9	-4.14 ± 0.2	-25.29 ± 2.9
Gly429Ala	-43.74 ± 2.8	-8.95 ± 2.4	34.41 ± 3.1	-4.14 ± 0.2	-22.59 ± 3.0
Tyr462Ala	-41.52 ± 3.2	-12.50 ± 2.2	34.28 ± 2.7	-4.04 ± 0.2	-23.79 ± 2.9
Tyr473Ala	-36.04 ± 3.1	-12.61 ± 2.1	34.21 ± 3.9	-3.72 ± 0.2	-18.16 ± 3.1
Glu558Ala	-39.68 ± 3.0	-7.60 ± 2.0	30.19 ± 3.7	-3.94 ± 0.2	-21.02 ± 3.6

Overall, the energy contribution is mainly from the van der Waals and the electrostatic interactions which are evident from the energy value for Tyr462Ala variant. Thus both aromatic residues and polar residues are necessary for effective protein-ligand interaction. It can also be noted that the reported mutation, Gly429Ala and Glu558Ala, does not affect the efficacy of drug molecules significantly.

Per-residue Energy Decomposition: Per-residue binding free energy contributions to the variant/ligand complexes were given in **Table 5.4**. The per-residue energy contributions of mutated amino acid residues at the binding site of PARP-2 variants are found to be less as compared to the wild enzyme. In all the variants, a drastic decrease in the SASA and van der Waals energy contribution to overall binding energy is observed from Tyr455 residue. This decrease in the energy contributions can be attributed to the geometrical changes at the binding site occurred due to alanine substitution. In Gly429Ala variant, the alanine found to have similar electrostatic energy contribution as that of glycine. However, a reduction is observed in van der Waals energy from His428, Phe463 and Tyr473 residues. This energy loss is compensated from the lower positive value for polar solvation energy contribution. In Tyr462Ala and Tyr473Ala variants, the loss of aromaticity of the residues greatly affected the hydrophobic contributions to the binding free energy. However, in Tyr462Ala variant, Tyr473 forms stronger aromatic interactions which is evident from higher van der Waals energy contribution and correlates well with the strength of π -stacking interactions obtained in section 3.6. Also, it is found that the ligand is stabilised at the binding site of Tyr462Ala variant by electrostatic interactions with Gly429, Ser470 and Glu558 residues. In Tyr473Ala variant, apart from the decrease in favourable energy contributions such as SASA energy, van der Waals energy and electrostatic energy, there is an increase in the unfavourable polar solvation energy contribution. The higher number of hydrogen bonds with water and water bridges at the binding site correlates well with the higher polar solvation energy in Tyr473Ala variant. The replacement of glutamic acid in Glu558Ala variant negatively affected the electrostatic contribution however it is compensated by a drastic decrease in the unfavourable solvation energy. The aromatic residues tyrosine and histidine contributed to the stable binding energy. The per-residue analysis suggests that the flexible nature of mutated amino acids at the binding site of PARP-2 variants is crucial for well-balanced hydrophobic as well as electrostatic interactions.

Table 5.4: Per residue binding energy contribution in the PARP-2 variant complexes

	Major Contribution			
	Energy in kcal/mol			
	SASA	van der Waals Energy	Electrostatic Energy	Polar Solvation Energy
wild	Tyr473 ~ -0.34 Tyr462 ~ -0.24 Tyr455 ~ -0.20 Gln332 ~ -0.14 Ser328 ~ -0.11	Tyr473 ~ -4.44 Tyr462 ~ -3.00 His428 ~ -2.84 Tyr455 ~ -2.78 Phe463 ~ -1.14	Glu558 ~ -3.17 Ser470 ~ -2.70 Gly429 ~ -1.69	Glu558 ~ 4.84 Tyr473 ~ 2.14 Ser470 ~ 1.61
Gly429Ala	Tyr473 ~ -0.33 Tyr462 ~ -0.29 Tyr455 ~ -0.13 Gln332 ~ -0.13 Ser328 ~ -0.13	Tyr473 ~ -3.92 Tyr462 ~ -3.11 His428 ~ -2.57 Tyr455 ~ -1.49 Phe463 ~ -1.03	Glu558 ~ -3.05 Ser470 ~ -2.12 Ala429 ~ -1.70 Ser328 ~ -1.43	Glu558 ~ 4.12 Tyr473 ~ 1.73 Ser470 ~ 1.31 Lys469 ~ 1.70
Tyr462Ala	Tyr473 ~ -0.34 Ala462 ~ -0.19 Tyr455 ~ -0.18 Gln332 ~ -0.17 Ser328 ~ -0.06	Tyr473 ~ -4.62 Ala462 ~ -1.39 His428 ~ -2.31 Tyr455 ~ -1.81 Phe463 ~ -1.16	Glu558 ~ -3.00 Ser470 ~ -2.62 Gly429 ~ -1.65 Ser328 ~ -0.81	Glu558 ~ 4.68 Tyr473 ~ 2.18 Ser470 ~ 1.58 Lys469 ~ 1.45
Tyr473Ala	Ala473 ~ -0.10 Tyr462 ~ -0.24 Tyr455 ~ -0.15 Gln332 ~ -0.17 Ser328 ~ -0.10	Ala473 ~ -1.08 Tyr462 ~ -2.92 His428 ~ -2.68 Tyr455 ~ -1.47 Phe463 ~ -0.96	Glu558 ~ -3.22 Ser470 ~ -2.52 Gly429 ~ -1.73 Ser328 ~ -0.99	Glu558 ~ 4.76 Ala473 ~ 0.33 Ser470 ~ 1.67 Lys469 ~ 2.48
Glu558Ala	Tyr473 ~ -0.33 Tyr462 ~ -0.22 Tyr455 ~ -0.08 Gln332 ~ -0.14 Ser328 ~ -0.09	Tyr473 ~ -3.90 Tyr462 ~ -2.29 His428 ~ -2.72 Tyr455 ~ -0.96 Phe463 ~ -0.95	Ala558 ~ -0.52 Ser470 ~ -2.58 Gly429 ~ -1.77 Ser328 ~ -0.83	Ala558 ~ 0.03 Tyr473 ~ 1.96 Ser470 ~ 1.62 Lys469 ~ 2.87

* The mutated residues are given in bold.

5.9 CONCLUSION

PARP inhibition is a promising therapeutic strategy in cancer treatment. However, the clinically trialled drug candidates have no selectivity for PARP enzymes. Here, we aimed to identify the essential amino acid residues required for the inhibition of PARP-2 enzyme using single amino acid substitution. The importance of π - π stacking and hydrogen bond interactions in the inhibition mechanism were determined by mutating the enzyme into four variants: Gly429Ala, Tyr462Ala, Tyr473Ala and Glu558Ala. The $\Delta\Delta G$ value between wild and unbound variants from CAS results showed that Tyr473Ala variant has increased the stability (0.56 kcal/mol); whereas the alanine substitution of Gly429 (-0.6 kcal/mol), Tyr462 (-0.82 kcal/mol) and Glu558 (-0.59 kcal/mol) has decreased the variant stability. Molecular docking of FMTNBC with mutant enzymes suggested that methyl group stabilizes the complex by hydrophobic interactions with Ile331, Tyr455 and Met456. The stability of complex formed by the four variants was confirmed by 250 ns molecular dynamics simulation. The alanine substitution of Tyr473 in unbound protein showed lesser flexibility compared to other three substitutions, Gly429, Tyr462 and Glu558; which correlates with the CAS results. The structural analysis of variant complexes showed that the protein-ligand interactions in Tyr473Ala are not well-maintained. The ligand-water and cavity-water hydrogen bond distribution suggests that the ligand and cavity are more solvated in Tyr473Ala. The strength of aromatic interactions formed between the residues and the ligand follows the order: Tyr473 > Tyr462 > Tyr455 > His428 > Phe463. PCA revealed that fluctuations at the binding sites are essential for effective ligand binding. The bound variant Tyr473Ala showed fewer fluctuations in first three principal components which agree to the least favourable protein-ligand interactions in the complex. Among the variants, it is observed that the binding free energy of Tyr473Ala /FMTNBC complex is least favourable (-18.16 ± 3.1 kcal/mol) and Tyr462Ala /FMTNBC complex is most favourable (-23.79 ± 2.9 kcal/mol). This also suggests that Tyr473 contributes significantly than Tyr462 for efficient ligand binding in wild PARP-2 enzyme. Free energy landscapes of binding site residues as a function of dihedral angles revealed that neither the alanine substitution nor the ligand binding affected the secondary structure of the enzyme. Overall, the study suggests that the

importance of aromatic rings in both the ligand and the binding site of PARP-2 for effective inhibition. Among the aromatic residues, Tyr473 contributes to strong protein-ligand interactions than other residues such as Tyr462, Tyr455, His428 and Phe463. To conclude, π -stacking interactions with Tyr473 are more significant followed by hydrogen bond interactions with Gly429, water-bridges with Glu558 and hydrophobic interactions Ile331, Tyr455, Met456 residues. The outcomes of this study provide insights for the designing of target-specific and isoform-selective PARP-2 inhibitors.

CHAPTER 6

MOLECULAR MECHANISM OF INHIBITION OF COVID-19 MAIN PROTEASE BY β -ADRENOCEPTOR AGONISTS AND ADENOSINE DEAMINASE INHIBITORS USING *IN SILICO* METHODS

Novel coronavirus (COVID-19) which emerged in late 2019 has badly affected the world. No clinically proven drugs are available yet as the targeted therapeutic agents for the treatment of this disease. The viral main protease which helps in replication and transcription inside the host can be an effective drug target. In the present study, we aimed to discover the potential of β -adrenoceptor agonists and adenosine deaminase inhibitors as repurposing drugs against protease inhibitor.

6.1 BACKGROUND

The current global pandemic of COVID-19 has affected more than 213 countries since its emergence in late 2019. This respiratory disease is caused by novel virus strain of family *Coronaviridae*; SARS Cov-2 (nCoV) which is an enveloped, positive-sensed, single stranded RNA betacoronavirus. As of 17 November 2020, 54.8 million people have been infected and 1.3 million people have died of SARS CoV-2. The evidences suggest that nCoV has a zoonotic source similar to SARS (severe acute respiratory syndrome) coronavirus and MERS (Middle East respiratory syndrome) coronavirus (Ahmad et al. 2020). Clinically proven antiviral drugs and vaccines are not available for SARS, MERS and COVID-19 pandemic. Currently, the treatment of COVID-19 uses remdesivir (viral RNA polymerase inhibitor), hydroxychloroquine (anti malarial drug), lopinavir/ritonavir (anti HIV drugs- protease inhibitors) and dexamethasone (corticosteroid medication). Studies show that these medications showed little or no reduction in the mortality rate of hospitalized patients when compared to normal cases. Since inventing a FDA approved drug after all the clinical trials is time taking process, repurposing of already available drugs can be a better solution for the treatment of COVID-19. Therefore, to develop therapeutic strategies for COVID-19, it is necessary to understand the mechanism of action of various viral enzymes.

Coronaviridae have the largest positive stranded RNA genome of 26 to 32 kb among the known RNA viruses (Schoeman and Fielding 2019). The sequence analysis of SARS Cov-2 isolates reveal that the genome encodes for 16 non-structural proteins (Nsp 1-16) which forms replicase/ transcriptase complex (RTC), 4 structural proteins (spike, envelope, membrane, nucleocapsid) and 9 putative accessory factors (Fehr and Perlman 2015; Gordon et al. 2020; Wu et al. 2020). Among these proteins, main protease (M^{pro}) plays a major role in the life cycle of novel coronavirus and is the key enzyme for replication and transcription process. M^{pro} process the precursor polyproteins to form functional proteins inside the virus, mainly the post-translational processing of replicase polyprotein (Wang et al. 2016). The viral main protease which helps in replication and transcription inside the host can be one among the effective drug target. Thus, the objective of the present study is to understand the molecular mechanism of main protease inhibition by potential drug candidates using virtual screening (VS) and molecular dynamics (MD) simulations. Preference was given to the existing drug molecules such as β -adrenoceptor agonists and adenosine deaminase inhibitors which can be used as repurposing drug against M^{pro} inhibitors since β -adrenoceptor agonists are used in the treatment of bronchial asthma (Barisione et al. 2010) whereas adenosine deaminase inhibitors are used in the treatment of cancer and inflammatory disorders (Glazer 1980; Trincavelli 2013). The identification of lead molecules was done by both ligand-based and structure-based virtual screening methods. Also, this work analyzes the detailed molecular mechanism and energy requirements for the ligand to act as effective protease inhibitors. Repurposing of existing drug molecules helps us to reduce the time required to develop new molecules as an effective inhibitor. The insights obtained from this study can be used to design new drug molecules with higher biological activity to treat COVID-19.

6.2 LIGAND SELECTION

The three-dimensional structures of the drug candidates were retrieved from ZINC database based on the Lipinski's rule (Irwin et al. 2012). Approximately, 22621 molecules were obtained from the ZINC database for ligand-based virtual screening. LigPrep module (Schrödinger 2020-1) was used to prepare the ligands. The low energy conformations of the ligand with possible ionization states were generated at

pH value 6.0. A total of 45907 conformations were generated from the dataset after adding hydrogen and removing small fragments. Further, the geometry of resultant drug candidates was refined by LigPrep module using OPLS_2005 force field (Jorgensen et al. 1996; Kaminski et al. 2001).

6.3 PHARMACOPHORE MODEL GENERATION

The crystal structure of COVID-19 main protease with Michael acceptor inhibitor N3 (PDB ID: 6LU7, resolution: 2.16 Å) (Jin et al. 2020) was used as initial structure for pharmacophore modelling. A pharmacophore model is generated using the structural information the bioactive ligand or active site of the target. N3 was developed using computational drug-designing sources and found to inhibit protease effectively with $K_{obs}/[I]$ value of $11300 \pm 880 \text{ M}^{-1}\text{s}^{-1}$ equivalent to dissociation constant (Jin et al. 2020). The protein was prepared at an experimental pH of 6.0 using PROPKA program and Phase module (Schrödinger 2020-1) was used to generate the pharmacophore model based on the protein-ligand interactions using ‘Receptor-ligand complex’ option. Pharmacophore-based screening was done with the prepared ligands from ZINC database. Minimum matches of 4 out of 5 sites were set as the criteria to obtain ligands with desired features. The final hits were ranked according to the Phase fitness score, align score, vector score and volume score (Dixon et al. 2006).

Based on M^{Pro} (6LU7)-imidazole carboxamide complex (receptor-ligand complex), five-featured pharmacophore model, AARRR, consisting of two hydrogen bond acceptor (A3 and A4) and three aromatic rings (R9, R10 and R11) were generated (**Figure 6.1**). This model is expected to identify the drug candidates with desired pharmacophore requirements and thus expected to act as inhibitor by fitting in the active site of protein. The acceptor A3 lies towards Asn142, Gly143 residues and acceptor A4 lies towards Glu166, Leu167 residues. The aromatic rings R10 and R11 were found to be near to aromatic amino acid residues, i.e., Phe140, Hie163, His164 and Pro52, Tyr54 respectively. In covid-19 main protease, the Nε2-protonated tautomeric form (Hie) of histidine amino acid (His) (Li and Hong 2011) is present at the binding site. The ring R9 is found close to Thr25, Thr26 and Cys145. A dataset containing 22621 drug-like molecules were screened using AARRR pharmacophore

model as template to obtain drug candidates with similar pharmacophore. Around 5300 molecules were obtained as top hits for structure-based virtual screening.

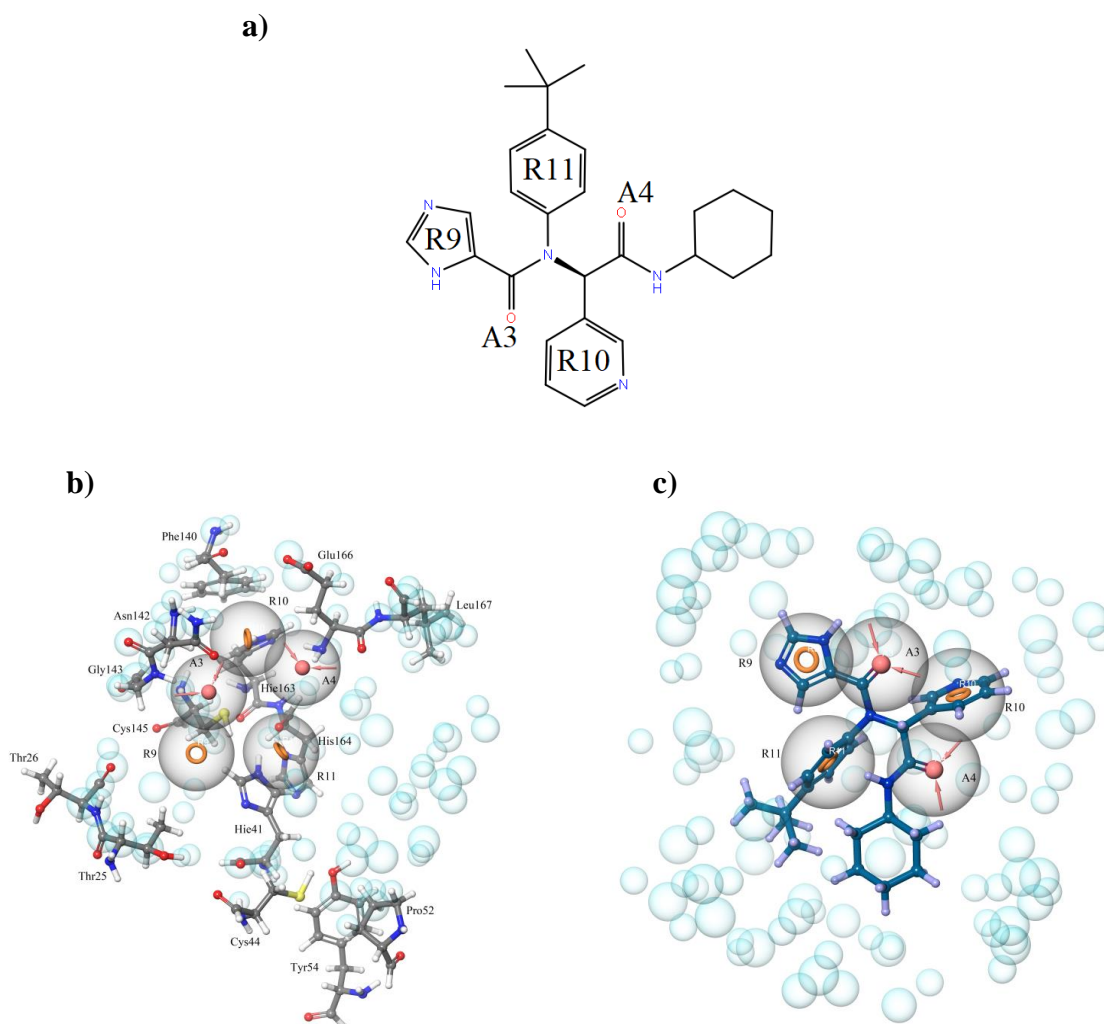


Figure 6.1: (a) Structure of broad spectrum non-covalent inhibitor, derivative of imidazole carboxamide (b) Pharmacophore model (AARRR) generated using Phase module inside the binding pocket of covid-19 main protease (c) Overlay of inhibitor over the generated model. The spheres indicate the excluded volume forbidden for ligands due to enzyme backbone or lipophilic interactions.

6.4 MOLECULAR DOCKING-BASED SCREENING

Ligand-based virtual screening by pharmacophore model identified 5297 hits from ZINC database based on the fitness score. Since the database is larger for molecular docking, the ligands were subjected to high-throughput virtual screening (HTVS). All

molecular docking studies were carried out using Glide module (Schrödinger 2020-1). Docking based screening is done by three steps, HTVS, standard precision (SP) and extra precision (XP). The van der Waals scaling factor for proteins was set to 1.0 cut-off and a grid of 72 Å dimension was generated at protease binding site. A total of 704 molecules obtained from HTVS, were then subjected to docking via SP mode. Out of these compounds, 52 compounds from the database were selected for XP mode of docking. Finally, two top-scored ligands each from both categories were chosen for post dock analysis. For comparison, two other crystal structures of main protease with a broad spectrum non-covalent inhibitor X77 (6W63) and alpha-ketoamide 13b (6Y2G) and Apo M^{Pro} (7KFI) were used for better insight of the ligand binding site of the enzyme **Appendix XX**.

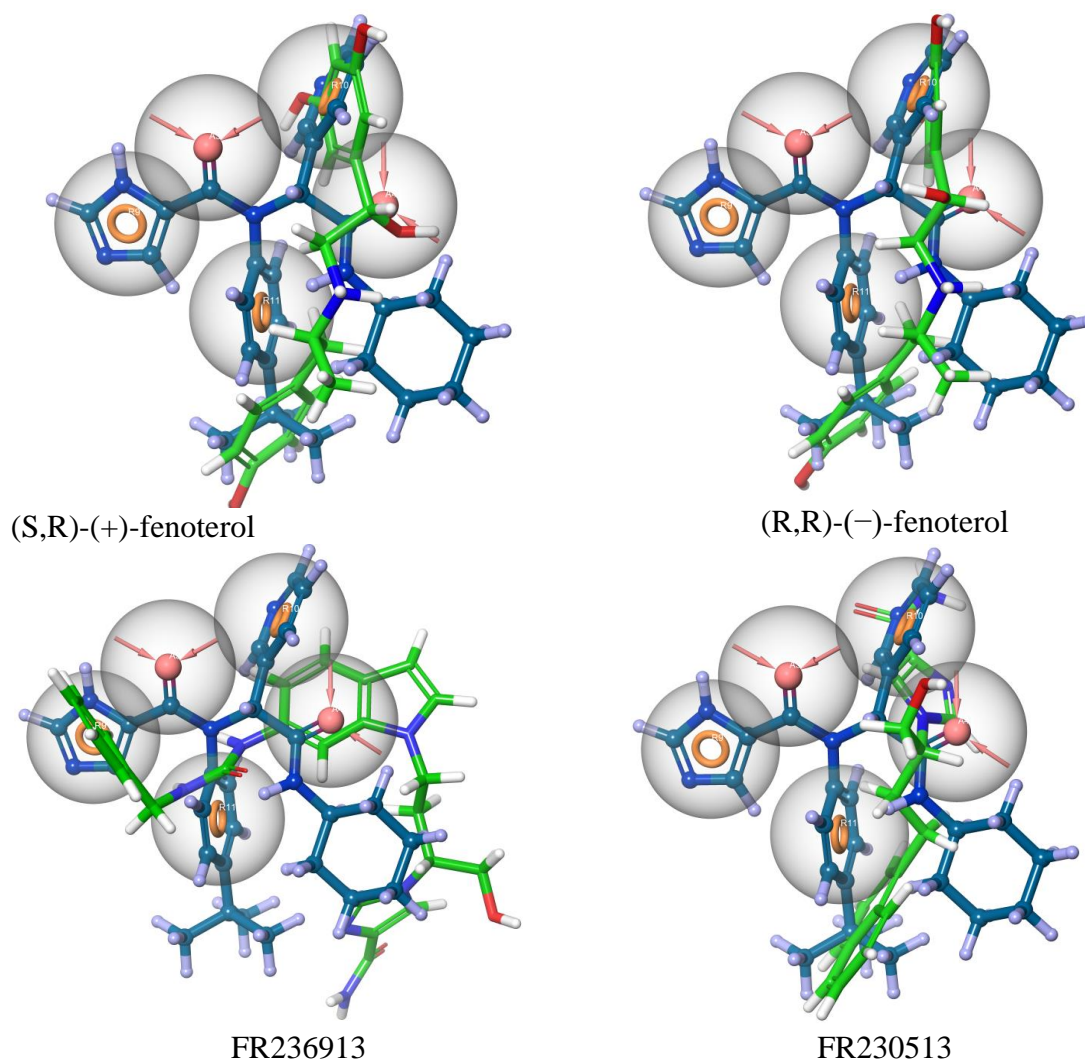


Figure 6.2: Overlay of top-scored molecules over developed pharmacophore model.

The docking score of top-scored hit molecules ranges from -9.491 to -7.526 kcal/mol for adrenoceptor agonists and -9.155 to -8.004 kcal/mol for adenosine deaminase inhibitors (**Appendix XXI**). The four top-scored molecules predicted by XP mode of molecular docking are (S,R)-(+)-fenoterol, (R,R)-(-)-fenoterol, FR236913 and FR230513. Fenoterol is β adrenoceptor agonist which is used as asthma medication (Svedmyr 1985). Compounds, FR236913 and FR230513, are potent adenosine deaminase inhibitors (Terasaka et al. 2004a; b). The overlay of top-scored molecules obtained from XP mode of docking over the developed pharmacophore model is shown in **Figure 6.2**. Both the inhibitor category satisfied at least three chemical features, namely the two aromatic rings and one hydrogen bond acceptor which correlate well with the developed AARRR pharmacophore model. The highest docking scored + (-) Fenoterol molecule possesses two aromatic rings which can hold the ligand at protein catalytic site by stacking interactions. The molecule has two sets of oxygen atoms which can act as hydrogen bond acceptor with Glu166 (H) and Gly143 (H). In case of adenosine deaminase inhibitors (FR236913 and FR230513) two aromatic rings are found in the structure which contributes to the activity but at slightly different sites compared to the developed model. FR236913 molecule has one acceptor carbonyl oxygen which interacts with Gln189 (H); whereas FR230513 has two hydrogen bond acceptor groups. The nitrogen atom present at the indole ring of the inhibitor interacts with Glu166 (H) and the carbonyl oxygen of the inhibitor was found to form random hydrogen bonds with Gly143, Ser144 and Cys145. Thus aromatic rings and hydrogen bond donor-acceptor groups are important chemical features for the inhibitors as obtained by the developed pharmacophore model.

2-D Ligand interaction diagram (**Appendix XXI**) showed the possibilities of hydrogen bonding between acceptor site of ligand and amino acid residues such as Glu166, Gln189, Gly143, Phe140 and Hie163 of covid-19 main protease. Also, there is lesser possibility of hydrogen bond formation with the residues Cys44, Cys145 and π -stacking interaction with the residue Hie41. It is found that the complex is stabilized by hydrophobic interaction as well as polar interaction. The presence of water mediated hydrogen bonds further increase the stability of the complex. The protonated

form of compound, (S,R)-(+)-fenoterol showed the highest docking score (-9.491) and found to have hydrogen bonding interaction with Cys44, Hie163 and Glu166. This compound also showed a water-mediated hydrogen bond with Asn142. The second potent ligand with lowest binding energy (-55.55 kcal/mol) is showed by (R,R)-(-)-fenoterol which is a conformer of fenoterol. From docking studies, it is found that both the enantiomers of fenoterol are found to be potent inhibitors against protease enzyme among adrenoceptor agonists. Similarly, among adenosine deaminase inhibitors, FR236913 and FR230513 are found to be potent ligands against protease. The compound, FR236913 showed hydrogen bonding interaction with Phe140, Gly143, Hie163 and Gln189. Additionally, it showed stacking interaction with Hie41. The compound, FR230513 showed only three hydrogen bonds with residues Phe140, Gly143 and Hie163. Apart from hydrogen bonds, all the complexes are stabilized by hydrophobic interactions and polar interactions.

The top-scored molecules were docked to other crystal structures of main protease (6Y2G and 6W63) and apo form (7KFI) for comparison. The docking score and interactions of native inhibitors of main protease are given in **Appendix XXII**. The interactions of ligands with the amino acid residues at binding sites of 6Y2G, 6W63 and apo form are found to be similar to 6LU7 structure (**Table 6.1**). Mainly, the molecules form hydrogen bond interaction with Gly143, Hie163 and Glu166. The docking score suggests that the screening results of 6LU7 are well correlated with other crystal structures of main protease.

Table 6.1: Docking score (kcal/mol) of top-scored molecules with existing crystal structures of main protease (6LU7, 6W63, 6Y2G) and apo form (7KFI)

Sl No:	Compound	6LU7	6W63	6Y2G	7KFI
Adrenoceptor Agonists					
1	(S,R)-(+)-fenoterol	-9.491	-8.415	-7.289	-7.166
2	(R,R)-(-)-fenoterol	-9.425	-7.716	-7.205	-7.166
3	Arbutamine	-8.195	-7.182	-7.096	-6.150
4	Ractopamine hydrochloride	-7.526	-6.584	-6.774	-5.578

Adenosine deaminase inhibitors					
1	FR236913	-9.163	-6.243	-7.507	-6.124
2	FR230513	-9.049	-6.583	-7.433	-5.559
3	FR221647	-8.221	-5.438	-6.282	-5.371
4	FR233623	-8.004	-6.032	-7.098	-5.086

6.5 ADME/TOXICITY STUDIES

The drug-likeness of a molecule can be predicted by pharmacokinetic (PK) properties. The ADME/Tox profile and molecular descriptors of the top-scored hits were predicted by Qikprop module (Schrödinger Release 2020-1). The molecular weight (MW), number of hydrogen bond donor-acceptor groups and octanol/water partition co-efficient ($P_{o/w}$) of top-scored β -adrenoceptor agonists and adenosine deaminase inhibitors are evaluated based on rule of five to predict their drug-likeness.

The molecules are evaluated by Lipinski's rule of five and percent of human oral absorption (**Table 6.2**). The ADME/Toxicity values are found to be in permissible limit for bioavailable drugs (Section 2.6). All the top-scored hits obeyed Lipinski's rule of five (Lipinski 2004). The percentage of human oral absorption values are found to be in the permissible range (> 80% high and < 25% poor). All the top-scored molecules found to have molecular weight, solvent accessible surface area (SASA), hydrogen bond acceptor-donor groups, octanol/water partition coefficient (QPlogPo/w), polar surface area (PSA) and solubility (QPlogS) in acceptable range. The compound, +/-Fenoterol found to follow the acceptable range of ADMET properties and obeys Lipinski's rule of 5. In the case of FR236913, it is found that dipole moment is 12.06 debye, predicted blood/ brain partition coefficient (QPlogBB) is -2.898, SASA is 817.883 Å², solubility is -5.312, predicted percent of oral absorption is 80.010% and PSA is 138.831 Å² which are towards the upper limit of acceptable range.

Table 6.2: Predicted ADME/Toxicity properties for top-scored molecules from β -adrenoceptor agonist and adenosine deaminase inhibitors

	MW	Dipole Moment	SASA	Acceptor HB	Donor HB	QPlogPo/w	QPlogS
+Fenoterol	303.357	2.14	591.897	5.45	5	1.034	-1.783
-Fenoterol	303.357	1.44	606.354	5.45	5	1.071	-2.007
Arbutamine	317.384	2.56	652.492	5.45	5	1.546	-2.414
Ractopamine	301.385	2.46	622.722	4.7	4	2.524	-2.905
FR236913	446.508	12.06	817.883	8.2	5	2.358	-5.312
FR230513	309.367	4.65	590.194	6.2	3	2.166	-3.629
FR221647	259.307	11.17	526.568	6.2	3	1.131	-2.644
FR233623	323.394	5.65	607.496	6.2	3	2.275	-3.906

	QPP _{Caco}	QPlogBB	QPlogK _{h_{sa}}	Percent Human Oral Absorption	PSA	Rule of 5	Rule of 3
+Fenoterol	24.158	-1.686	-0.318	57.754	100.4	0	1
-Fenoterol	24.027	-1.771	-0.319	57.927	100.7	0	1
Arbutamine	26.698	-1.956	-0.247	61.527	100.8	0	0
Ractopamine	89.370	-1.177	-0.043	77.646	77.7	0	0
FR236913	33.337	-2.898	-0.108	80.010	138.8	0	0
FR230513	210.165	-1.505	-0.062	81.198	91.5	0	0
FR221647	177.212	-1.545	-0.407	73.814	90.2	0	0
FR233623	188.356	-1.592	-0.004	80.984	88.98	0	0

* Permissible ranges of physico-chemical properties are given in Section 2.6

6.6 MOLECULAR DYNAMICS SIMULATION

The stability of complex formed between covid-19 main protease (6LU7) and the top-scored ligands was analyzed by 100 ns simulations by Gromacs 2018.4 (Abraham et al. 2015) using SPC/E water model (Mark and Nilsson 2001). The protein topology was generated using AMBER99SB force field (Hornak et al. 2006). The complex was

solvated in a cubic box with 1 nm edges and neutralized with Na⁺ ions. The energy minimization of system was done by steepest descent algorithm with a maximum of 50000 steps until a convergence tolerance of 10 kJmol⁻¹. A 10 ns NVT and NPT equilibration were carried out at 300 K and 1.0 atm respectively throughout the process (Bussi et al. 2007; Martyna et al. 1992; Parrinello and Rahman 1981). Lastly, a 100 ns production run was performed by removing restraint to relax the protein-ligand system and trajectories were saved at every 10 ps for analysis. Three independent simulations (a total of 12 simulations) were performed for each protein-ligand system to check the reproducibility of the results (**Table 6.3**). A 100 ns simulation for M^{Pro}/N3 (6LU7), M^{Pro}/X77 (6W63) and M^{Pro}/13b (6Y2G) complexes was also performed to analyze the stability of native inhibitor N3 and to compare the efficiency of predicted inhibitors.

Table 6.3: Characteristics of protein (6LU7) -ligand complex obtained from 3 independent molecular dynamics simulations

Properties/ Simulation		S1	S2	S3	SD
RMSD (nm)	MPro/+ -fenoterol	0.180	0.208	0.190	0.014
	MPro/- -fenoterol	0.174	0.188	0.175	0.008
	MPro/FR236913	0.185	0.226	0.194	0.021
	MPro/FR230513	0.198	0.202	0.179	0.012
Rg(nm)	MPro/+ -fenoterol	2.198	2.202	2.184	0.009
	MPro/- -fenoterol	2.216	2.185	2.192	0.016
	MPro/FR236913	2.177	2.195	2.202	0.013
	MPro/FR230513	2.211	2.189	2.222	0.017

Stability of Protein-Ligand Complex: The backbone RMSD profile for main protease complexes are shown in **Figure 6.3**. The RMSD profile suggested that the complexes are stabilized after 25 ns of simulation. The average RMSD value for the complex of Michael acceptor N3 is found to be 0.24 nm. The average RMSD values for the backbone atoms of complex of +-fenoterol, --fenoterol, FR236913 and FR230513 for

3 independent simulations are found to be 0.193 ± 0.014 nm, 0.179 ± 0.008 nm, 0.202 ± 0.021 nm and 0.193 ± 0.012 nm respectively. It is evident from the RMSD profile that the complexes of top-scored hits are stable than the co-crystal inhibitor N3.

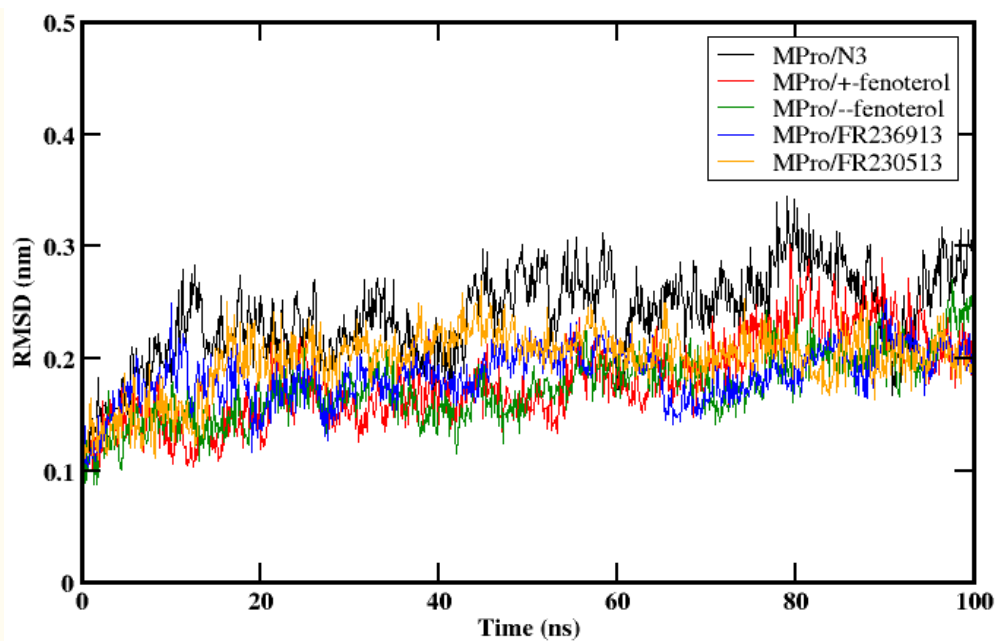


Figure 6.3: RMSD profile of backbone of covid-19 main protease (6LU7) complexed with Michael acceptor inhibitor N3 and top-scored hits over 100 ns trajectory.

RMSF profile describes the behaviour of movements of amino acid residues. Lower value of RMSF indicates rigid structure while higher value indicates loosely bounded structure. The average RMSF value of main protease bound to Michael acceptor N3, +-fenoterol, --fenoterol, FR236913 and FR230513 are found to be 0.116 nm, 0.103 nm, 0.107 nm, 0.109 nm and 0.086 nm respectively (**Figure 6.4 (a)**). The region from Thr25- Arg60 and Thr135- Ile200 which showed interaction with the ligand has lower fluctuations than in the complex of N3, +/-fenoterol and FR230513. The RMSF values of these residues lies in the range of 0.07- 0.09 nm. MPro/+fenoterol complex is found to be more stable with fewer fluctuations. The complex MPro/FR236913 showed an opposite trend with higher fluctuations at this region. The compactness of protease inhibitor complexes were predicted by Rg (radius of gyration) values throughout the 100 ns simulation. The Rg plots are shown in **Figure 6.4 (b)** and the average Rg values for +-fenoterol, --fenoterol, FR236913 and FR230513 complexes are found to be 2.198 ± 0.009 nm, 2.198 ± 0.016 nm, $2.191 \pm$

0.013 nm and 2.207 ± 0.017 nm respectively. The R_g value for N3 is found to be 2.218 nm. From the R_g values, it is clear that all the inhibitor complexes are compact and stable compared to the native inhibitor N3.

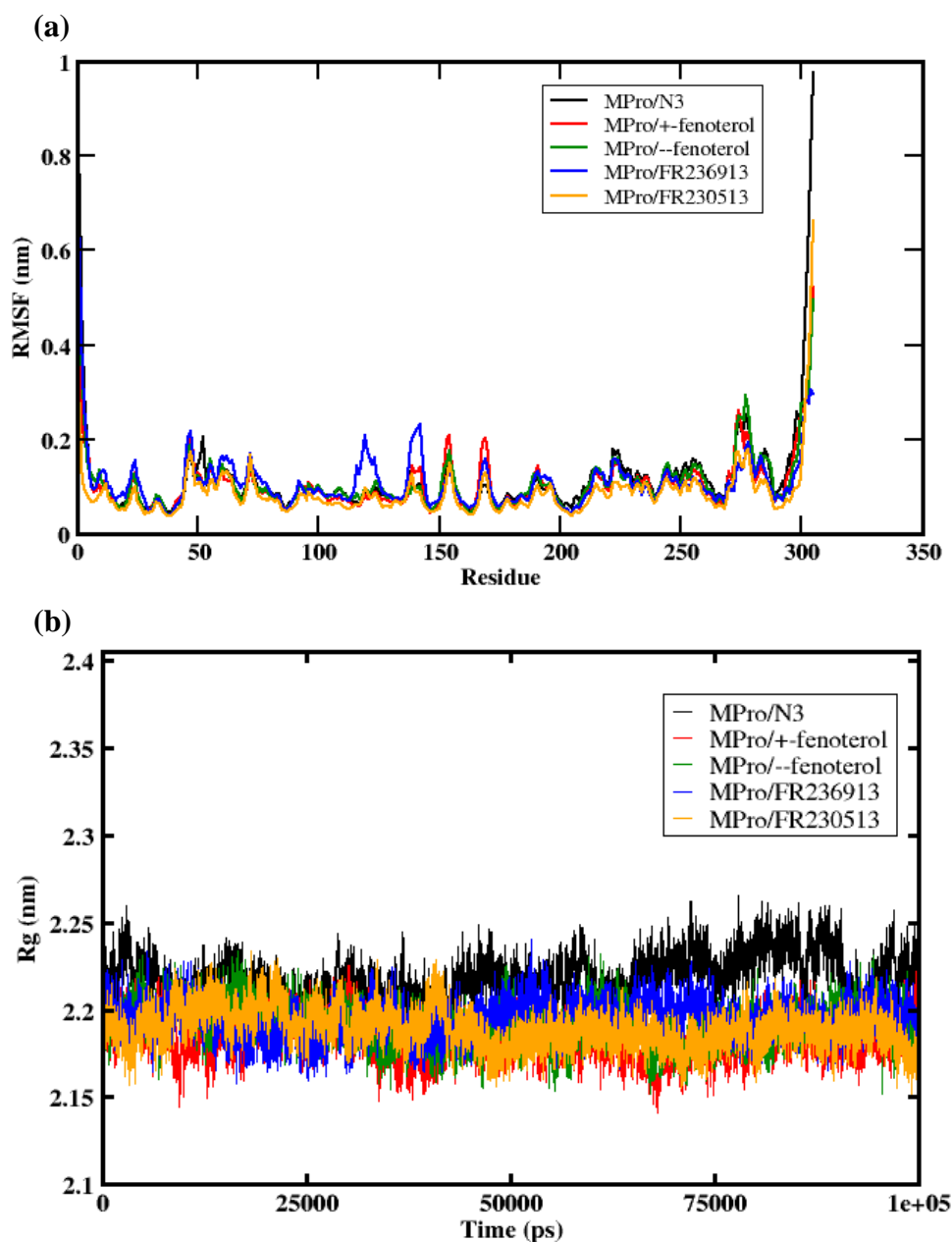
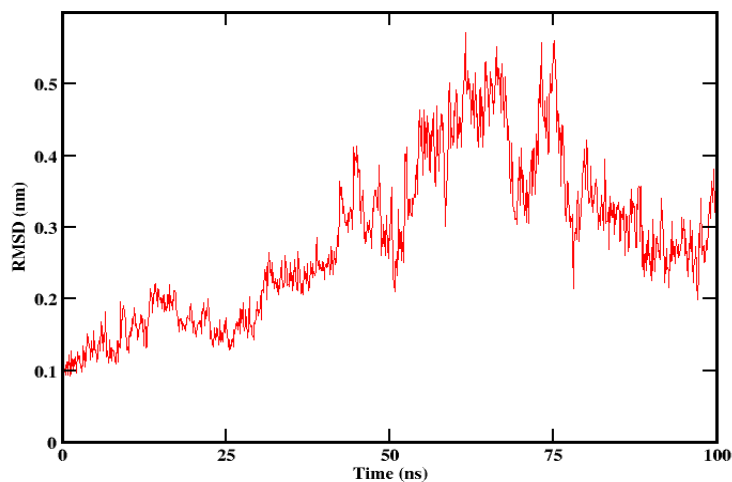


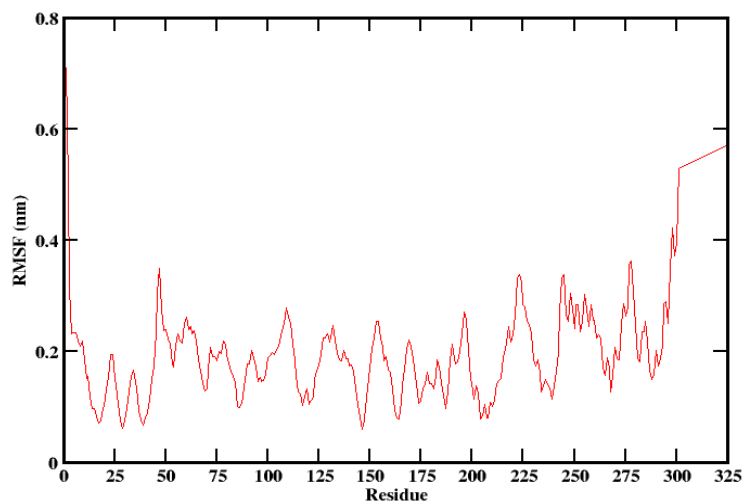
Figure 6.4: (a) Residue-based fluctuations of 6LU7 backbone of complexes (b) R_g (nm) vs. Time (ps) for main protease (6LU7)/ inhibitor complexes.

The main protease bound to inhibitor 13b (6Y2G) found to have higher RMSD value, higher RMSF fluctuations and lower compactness (higher R_g value) compared to above discussed complexes (Mittal et al. 2020) (**Figure 6.5**).

a)



b)



c)

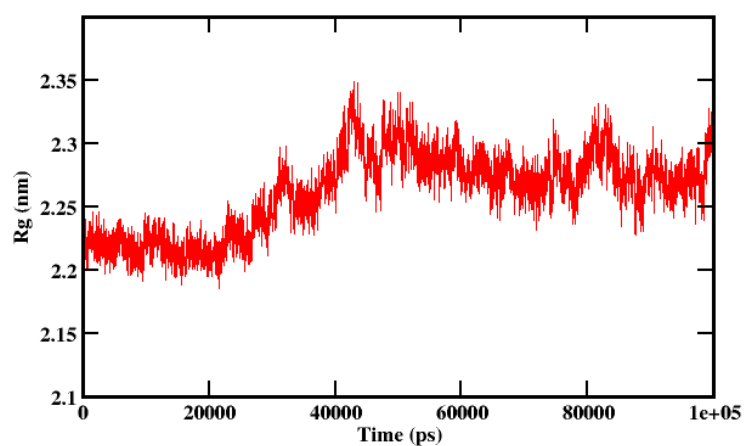


Figure 6.5: a) RMSD profile of backbone of main protease (6Y2G) complexed with alpha-ketoamide 13b, b) Residue-based fluctuations of protein backbone of complex and c) Rg (nm) vs. Time (ps).

Secondary Structure Analysis: Per-residue secondary structure analysis of the protein suggests that there are no considerable changes in the structure of enzyme on inhibitor binding (**Appendix XXIII**). The average occupancy of secondary structure components such as β -sheet, β -bridge, α -helix, bend, coil, 5-helix and 3-helix are shown in **Figure 6.6**. There is an insignificant presence of 5-helix in +/- fenoterol, FR236913 complexes during the course of simulation while 5-helix is absent in FR230513 complex. α -helix which found to be fluctuating at the initial time scale of simulation, stabilizes after 40 ns for all the complexes. β -sheets and β -bridges are found to be stable throughout the course of simulation. The secondary structure analysis suggests that there is no significant structural change at the binding site of protein on ligand binding.

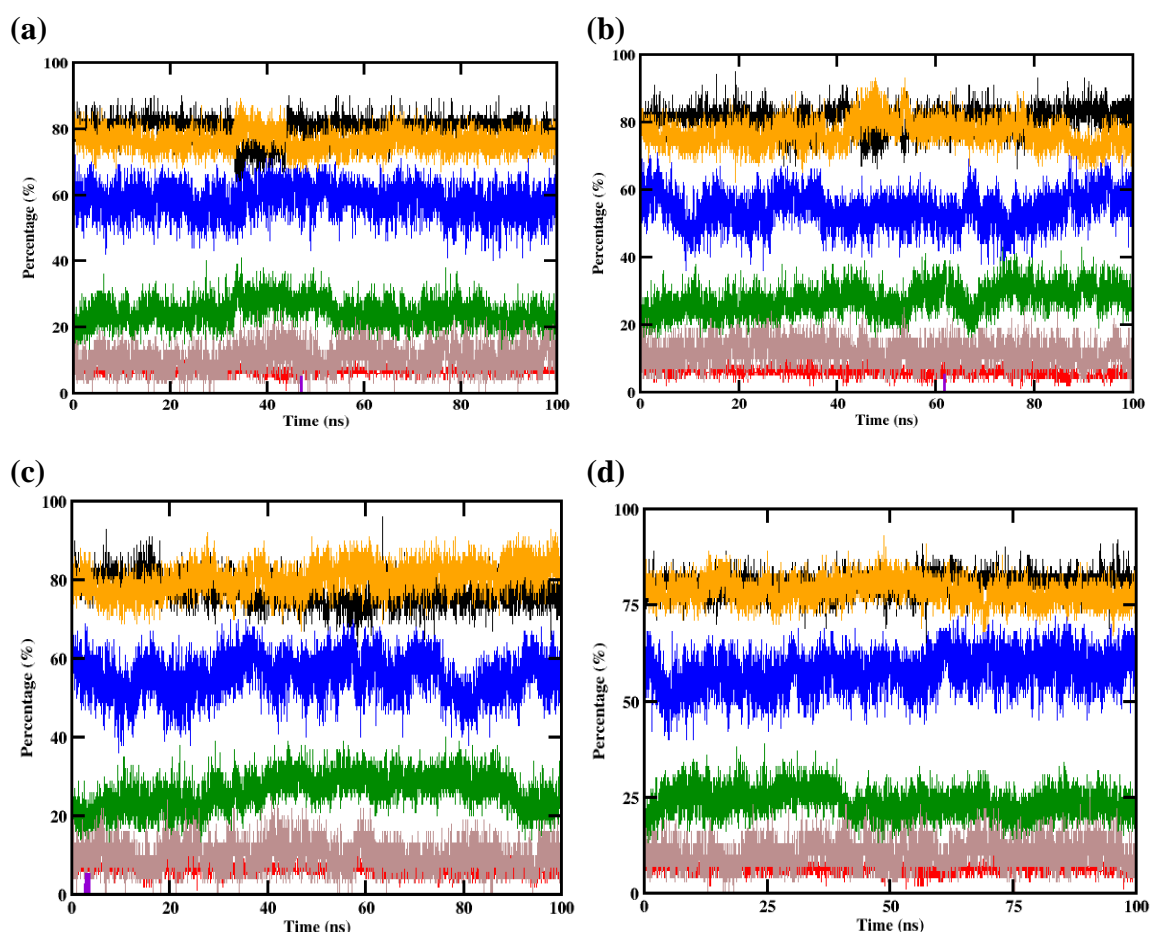


Figure 6.6: Occupancy of secondary structure elements of covid-19 main protease (6LU7) in the inhibitor complex. **A)** MPro/+fenoterol, **B)** MPro/-fenoterol, **C)**

MPro/FR236913 and **D**) MPro/FR230513. (Colour codes- Black: β -sheet, red: β -bridge, green: bend, blue: α -helix, yellow: coil, brown: 3-helix).

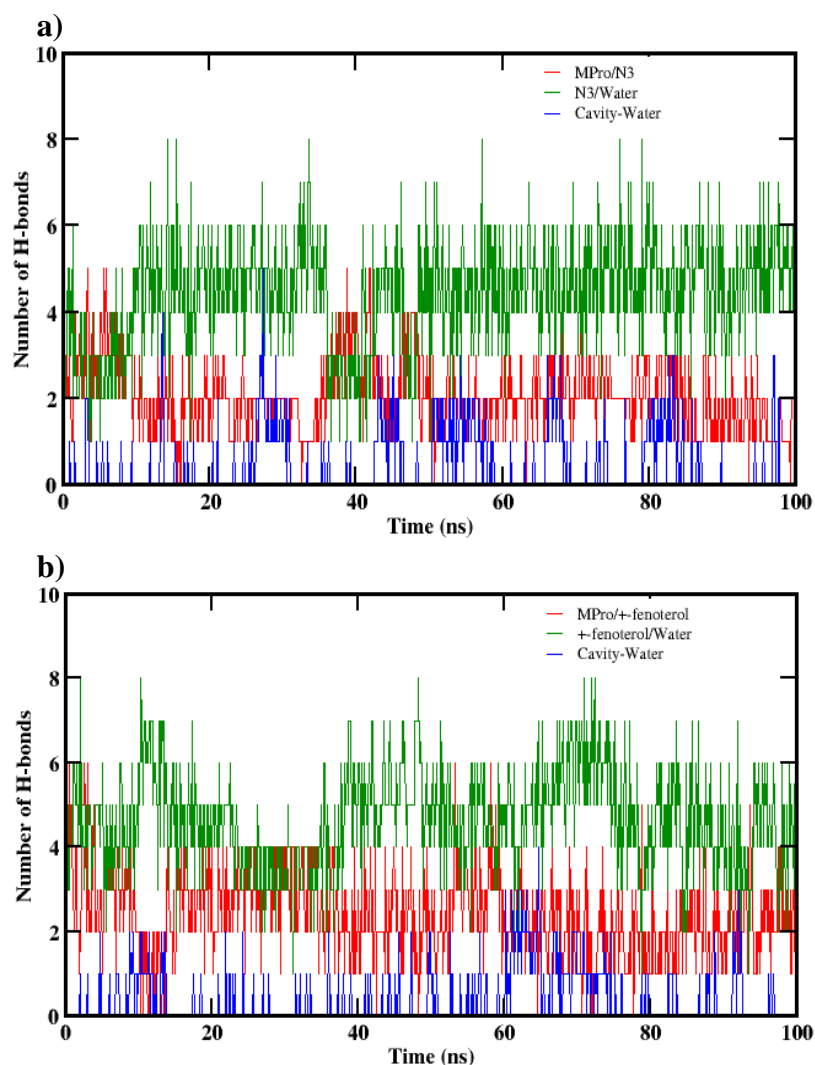
Hydrogen Bonding Interactions: The average number of hydrogen bonds between protein-ligand, ligand-water and binding site residues-water are illustrated in **Figure 6.7**. Apart from the hydrogen bonds between protein and ligand, water-mediated hydrogen bonds provide extra stability to the protein-ligand system. Hydrogen bonding interaction of ligand with Gly143, Hie163 and Glu166 residues are common in complexes (**Table 6.4**). Apart from these interactions, co-crystal inhibitor found to have interactions with His41 with 64.3% occupancy. Compound +-fenoterol forms two extra hydrogen bonds with Phe140 and Asp187 with occupancy of 18% and 22.9% respectively whereas -fenoterol shows three extra hydrogen bonds with Asn142 (13.4%), Asp187 (76.6%) and Cys145 (32.5%). The compound, FR236913 showed least occupancy values for all the hydrogen bonds whereas the compound, FR230513 showed major hydrogen bond occupancy with residues Glu166 (85.4%), Cys145 (59.8%) and Ser144 (41.4%). Even though similar hydrogen bond interactions are observed for all the complexes, difference in the value of hydrogen bond occupancy maybe attributed due to the possibility of acquiring various conformations during the course of simulation. The hydrogen bond occupancy of MPro/X77 (6W63) and MPro/13b (6Y2G) shows that the majority of the hydrogen bond interaction is observed with amino acid residues Glu166 (> 90 %) and Gly143.

Table 6.4: Hydrogen bond occupancy between ligand and residues at the binding site of main protease

System	Hydrogen Bond Occupancy (%)	
	Donor --- Acceptor	
MPro/N3 (6LU7)	N3 (H) --- His41 (N)	65.9
	N3 (H) --- Glu166 (O)	93.8
	N3 (H) --- Thr190 (O)	74.2
	Gln189 (H) --- N3 (O)	49.2
	Glu166 (H) --- N3 (O)	17.8
MPro/+fenoterol	Lig (H) --- Glu166 (O)	27.1
	Lig (H) --- Glu166 (O)	33.8
	Lig (H) --- Hie163 (N)	10.6

	Lig (H) --- Phe140 (O)	18.0
	Lig (H) --- Asp187 (O)	22.9
	Glu166 (H) --- Lig (O)	50.9
	Gly143 (H) --- Lig (O)	16.1
MPro/--fenoterol	Lig (H) --- Asn142 (O)	13.4
	Lig (H) --- Glu166 (O)	12.5
	Lig (H) --- Asp187 (O)	76.6
	Glu166 (H) --- Lig (O)	11.5
	Hie163 (H) --- Lig (O)	19.6
	Cys145 (H) --- Lig (O)	32.5
	Gly143 (H) --- Lig (O)	21.1
MPro/FR236913	Lig (H) --- Phe140 (O)	10.6
	Lig (H) --- Cys44 (O)	12.0
	Lig (H) --- Thr190 (O)	13.1
	Lig (H) --- Thr190 (O)	11.7
	Gln189 (H) --- Lig (O)	14.2
	Glu166 (H) --- Lig (O)	14.2
	Hie163 (H) --- Lig (O)	16.6
MPro/FR230513	Lig (H) --- Ser144 (O)	41.4
	Lig (H) --- Leu141 (O)	19.6
	Lig (H) --- Asn142 (O)	13.9
	Glu166 (H) --- Lig (N)	85.4
	Cys145 (H) --- Lig (O)	59.8
	Ser144 (H) --- Lig (O)	35.0
	Gly143 (H) --- Lig (O)	38.3
	Asn142 (H) --- Lig (O)	25.2
MPro (6W63)/X77	Glu166 (H) --- N3 (O)	96.1
	Hie163 (H) --- N3 (N)	18.5
	Gly143 (H) --- N3 (O)	18.3
	His41 (H) --- N3 (N)	64.3
MPro (6Y2G)/13b	13b (H) --- Asn142 (O)	59.6
	13b (H) --- His164 (O)	14.5
	13b (H) --- Phe140 (O)	43.2
	13b (H) --- His164 (O)	96.7
	13b (H) --- Glu166 (O)	43.3
	Glu166 (H) --- 13b (O)	99.4
	Hie163 (H) --- 13b (O)	79.9
	Cys145 (H) --- 13b (O)	28.8
	Gly143 (H) --- 13b (O)	65.1

Figure 6.7 (a) represents the hydrogen bond distribution of Michael acceptor inhibitor N3 with protein and water. It can be seen from **Figure 6.7 (b&c)** that average number of hydrogen bonds formed between protein and ligand is three. Water plays an important role in stabilising these complexes by forming hydrogen bonds with protein binding site as well as the ligands. From **Figure 6.7 (d)**, it is found that the average number of protein-ligand hydrogen bonds is two and the ligand is solvated more by ligand-water hydrogen bonds. Also, it can be seen in **Figure 6.7 (e)** that the ligand does not form stable hydrogen bonds with protein and the ligand is solvated by water molecules. Both the ligand (FR236913 and FR230513) forms a very few water-mediated hydrogen bonds and the binding site have lesser water molecules. Therefore, hydrogen bond stabilisation is seen maximum in case of +/-Fenoterol complexes as compared to other complexes.



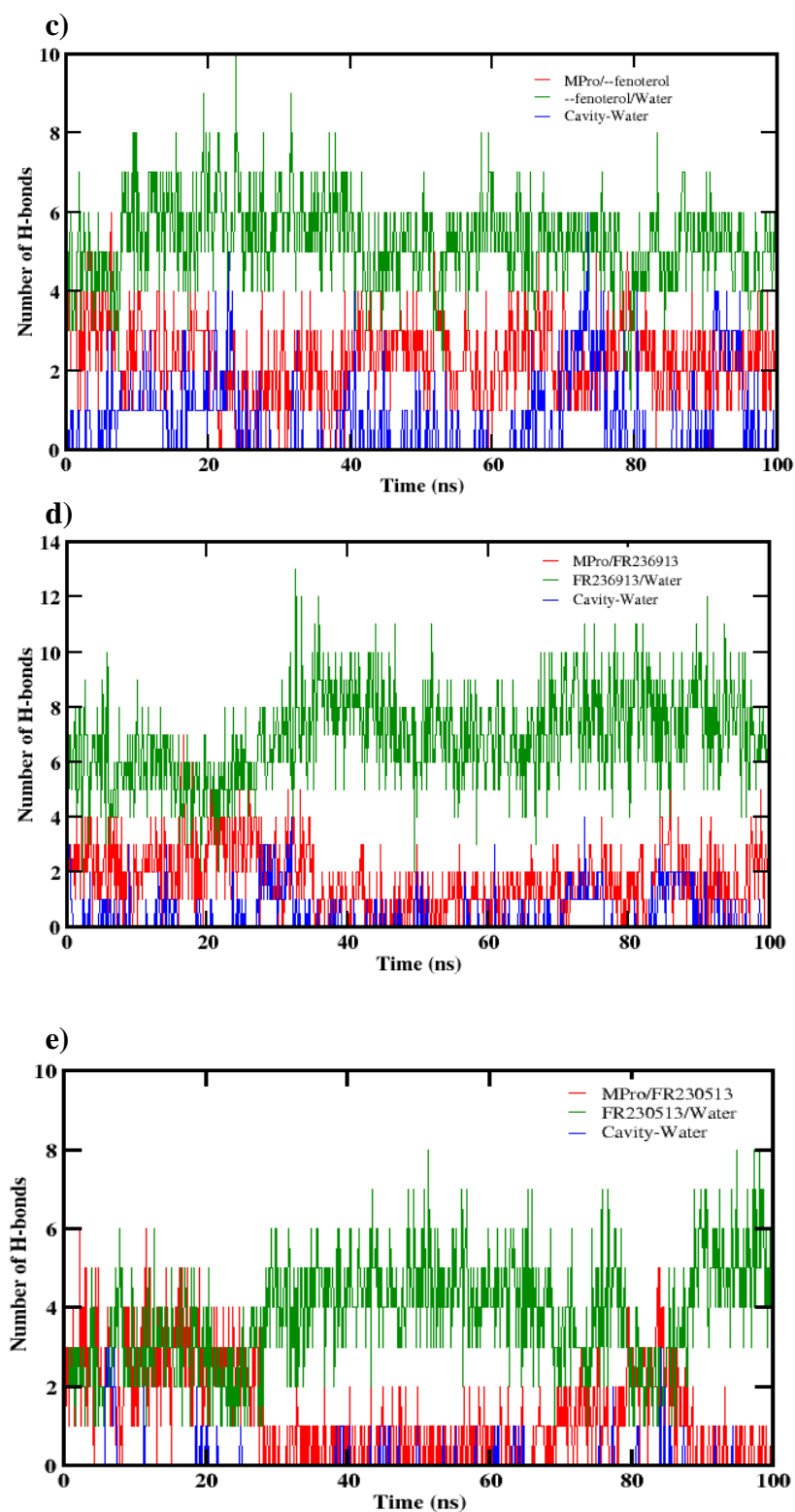


Figure 6.7: Hydrogen bond interactions between protease (6LU7) -inhibitor (red), inhibitor-water (blue) and binding site residues of protease-water (green). **a)** MPro/N3, **b)** MPro/+fenoterol, **c)** MPro/-fenoterol, **d)** MPro/FR236913 and **e)** MPro/FR230513.

6.7 PCA AND FEL ANALYSIS

Principal component analysis (PCA) was applied on backbone atoms for a stable trajectory of 40 ns for all the complexes of protease (6LU7) to analyze the essential dynamics which governs the conformational changes during simulation. The eigenvalue rapidly decreased along the eigenvector index indicating that first three eigenvector contribute significantly to the conformational changes in the protease enzyme during the simulation (**Figure 6.8**). The first ten eigenvectors accounts for 80.71%, 70.97%, 72.54%, 66.05% and 73.43% motions for MPro/N3, MPro/+fenoterol, MPro/--fenoterol, MPro/FR236913 and MPro/FR230513 complexes respectively, indicating that ligand induces conformational changes to protein on binding. PCA analysis suggests that complex of top-scored molecules are more stable than the complex of native inhibitor N3 bound to protease enzyme. PCA analysis also suggests that the compound, FR236913 binds intact to the binding site of enzyme with lowest correlated motions with the residues.

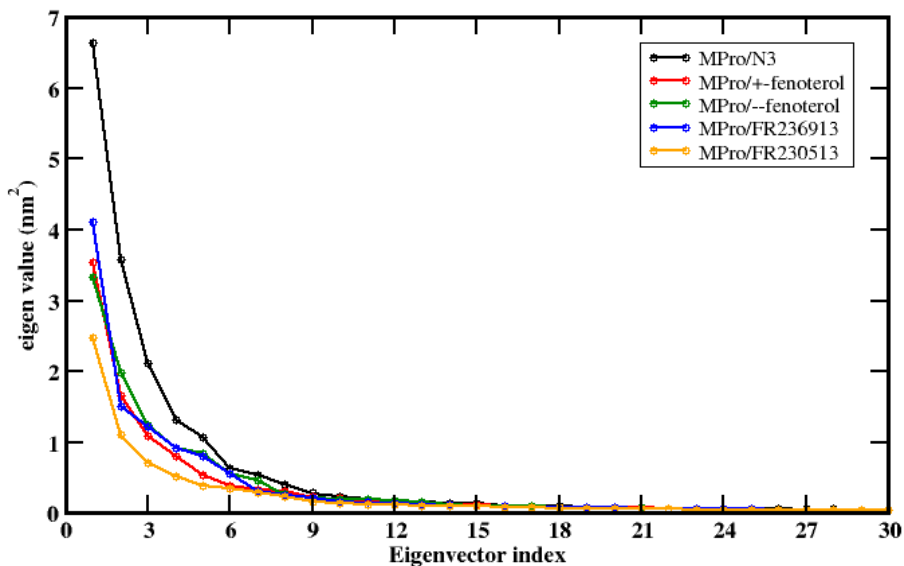
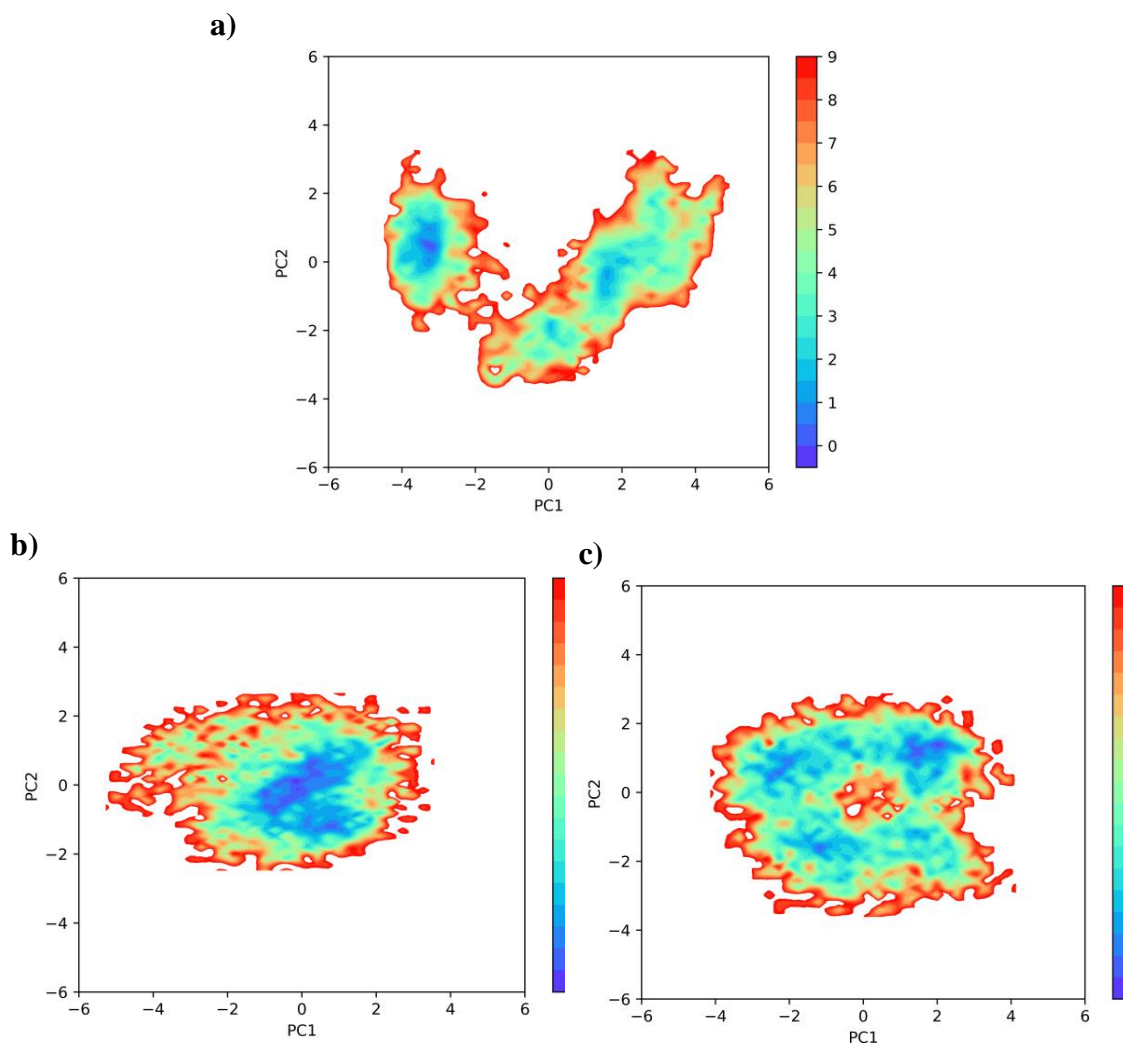


Figure 6.8: Plot of eigenvalue vs. first 30 eigenvector index derived from PCA over a stable trajectory of 40 ns for protein-ligand systems.

Gibbs free energy plots were constructed with PC1 and PC2 as reaction coordinates to relate the obtained structural properties with the thermodynamic information (**Figure 6.9**). The energy minima on plots indicate the stable ligand

conformation at the binding pocket of enzyme over estimated time scale. The FEL of the complexes of +-fenoterol and --fenoterol with the enzyme suggests that the ligand form a stable complex with the protein. The complex can span from one conformation to another easily with the help of interactions such as hydrogen bonds. Thus the flexibility of ligands to form hydrogen bonds with residues as well as water results in stable complex. The complexes of native inhibitor N3 and the other ligands such as FR236913, FR230513 have two well-defined regions separated by energy barriers. This is due to the reduced number of water-mediated hydrogen bonds in the complexes. The hydrogen bond distribution for FR236913 and FR230513 with the enzyme is similar which reflected in the FEL. The solvation of ligands reduces the possibility of change in conformations of the complex, making separate regions for ligand-water and protein-ligand interactions.



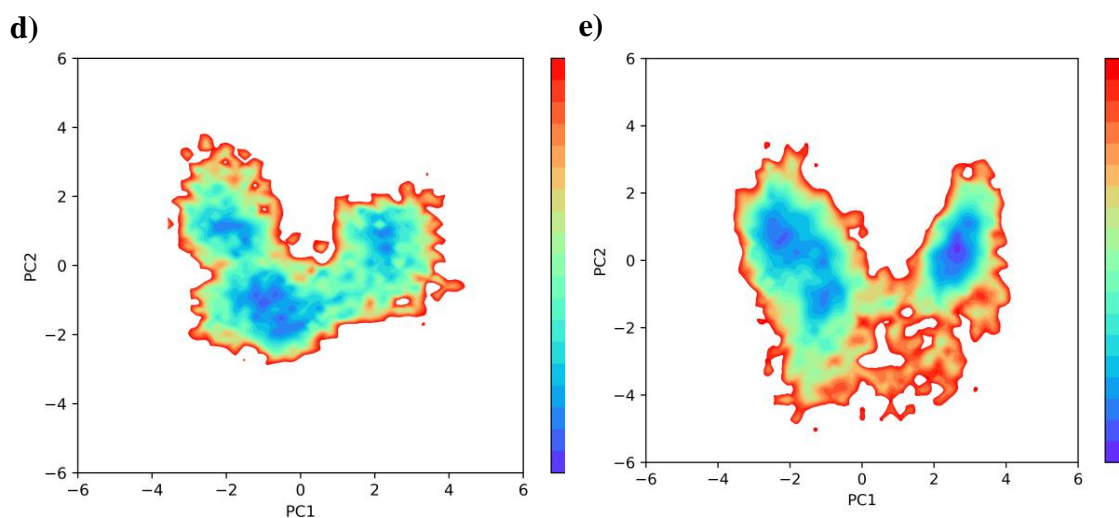


Figure 6.9: FEL from a stable 40 ns trajectory using the reaction coordinates as the projection of backbone atoms of protease (6LU7) onto the first two principal components. **a)** MPro/N3, **b)** MPro/+fenoterol, **c)** MPro/-fenoterol, **d)** MPro/FR236913 and **e)** MPro/FR230513.

6.8 MM/PBSA FREE ENERGY

To estimate the strength of protein-ligand interactions, binding energy analysis was carried out by MM/PBSA method for the last 75 ns of MD trajectory. The binding free energy of MPro/+fenoterol and MPro/-fenoterol are -52.27 and -47.36 kcal/mol respectively which indicates that the complexes are stable when compared to MPro/N3 (-33.31 kcal/mol), MPro/X77 (-24.25 kcal/mol) and MPro/13b (-27.92 kcal/mol). MPro/FR230513 and MPro/FR236913 has showed a stable binding free energy of -20.83 kcal/mol and -19.79 kcal/mol respectively (**Table 6.5**). The most important contribution to the overall binding free energy is from van der Waals energy and electrostatic energy. Electrostatic contribution arises from the hydrogen bonding interactions directly from the ligands or due to the presence of water-mediated hydrogen bonds. For +/-fenoterol the electrostatic contribution is found to be higher due to the presence of higher number of hydrogen bonds between protein and ligand. For all complexes, the contribution from van der Waals energy is quite significant which indicates the importance of hydrophobic interaction. SASA energy has a positive effect on overall energy. Lesser contribution of electrostatic energy in FR236913 and FR230513 complexes result in decrease in the binding energy

compared to fenoterol complexes. It can be seen that the average number of hydrogen bonds formed between MPro/ligand and cavity/water are reduced in case of MPro/FR236913 and MPro/FR230513 complexes. The high positive value of polar solvation energy and lower contribution of electrostatic energy disfavours the ligand binding at protease enzyme which is evident from the free energy of FR236913 and FR230513 complexes. The positive value for polar energy is due to the solvation of ligand by the water molecules, thereby reducing its interaction with the protein. MM/PBSA binding free energy values of complexes of N3, FR236913 and FR230513 suggests that FR236913 and FR230513 have similar effect on protease enzyme as its native inhibitor N3. Similarly, FEL and free energy analysis suggests that +/- fenoterol can be an effective inhibitor against protease enzyme. It is found that the adrenoceptor inhibitors shows more stable binding energy that the adenosine deaminase inhibitors. MM/PBSA energy components explain the importance of aromatic rings, hydrophobic core and hydrogen bond donor-acceptor groups in ligands as well as at the binding site. The binding free energy obtained from other two independent simulation shows good correlation.

Table 6.5: Contribution of energy components to MM/PBSA binding free energy for covid-19 main protease with potential hits (energy in kcal/mol)

System	van der Waals Energy ($\times 0.24$)	Electrostatic Energy ($\times 0.24$)	Polar Solvation Energy ($\times 0.24$)	SASA ($\times 0.24$)	Binding Energy ($\times 0.24$)
Co-crystal complexes					
MPro (6LU7)/N3	-254.015 \pm 17.26	-71.180 \pm 13.29	209.180 \pm 19.74	-23.049 \pm 1.16	-138.788 \pm 22.00
MPro (6W63)/X77	-206.926 \pm 13.15	-48.228 \pm 14.91	174.388 \pm 16.84	-20.276 \pm 1.00	-101.042 \pm 15.47
MPro (6Y2G)/13b	-234.467 \pm 22.70	-102.426 \pm 14.08	244.155 \pm 21.88	-23.763 \pm 2.00	-116.501 \pm 17.85
Simulation- 1					
MPro (6LU7)/ +/-fenoterol	-143.451 \pm 14.53	-262.777 \pm 25.02	203.943 \pm 25.09	-15.501 \pm 1.07	-217.785 \pm 14.99

MPro (6LU7)/ --fenoterol	-149.823 ± 12.09	-233.150 ± 30.14	201.554 ± 28.10	-15.933 ± 0.92	-197.351 ± 14.80
MPro (6LU7)/ FR236913	-186.225 ± 19.09	-42.974 ± 15.75	165.437 ± 27.34	-18.687 ± 1.50	-82.449 ± 15.91
MPro (6LU7)/ FR230513	-173.976 ± 9.75	-39.467 ± 9.31	143.109 ± 9.45	-16.449 ± 0.77	-86.782 ± 10.54

Simulation- 2					
MPro/+ fenoterol	-122.490 ± 22.35	-232.139 ± 22.66	165.665 ± 25.71	-13.681 ± 1.85	-202.645 ± 20.03
MPro/- fenoterol	-145.650 ± 13.89	-261.856 ± 28.15	207.421 ± 23.36	-16.317 ± 0.83	-216.402 ± 16.56
MPro/ FR236913	-203.832 ± 16.73	-65.136 ± 13.55	199.929 ± 21.05	-20.530 ± 1.39	-89.570 ± 15.468
MPro/ FR230513	-170.790 ± 10.39	-39.113 ± 9.66	143.968 ± 9.36	-16.299 ± 0.79	-82.234 ± 10.31

Simulation- 3					
MPro/+ fenoterol	-151.936 ± 13.45	-243.471 ± 25.94	207.032 ± 29.54	-15.933 ± 0.96	-204.208 ± 18.36
MPro/- fenoterol	-141.754 ± 13.66	-242.629 ± 29.30	190.267 ± 34.70	-15.394 ± 1.61	-209.510 ± 13.57
MPro/ FR236913	-177.256 ± 29.24	-52.982 ± 15.29	167.877 ± 29.82	-18.365 ± 1.93	-80.726 ± 24.45
MPro/ FR230513	-171.901 ± 13.31	-17.144 ± 12.87	121.829 ± 14.55	-16.223 ± 1.04	-83.439 ± 12.89

6.9 CONCLUSION

The present work describes the combined approach of ligand-based and structure-based virtual screening to obtain potential drug candidates as covid-19 main protease inhibitor. The drug-like molecules from ZINC database were screened and ranked based on docking score, fitness and Lipinski's rule of five. The molecular docking suggested 8 potential hits (4 each from adrenoceptor agonists and deaminase inhibitors) which showed lower binding energy with the protease. Out of 8, four molecules such as +/-fenoterol, FR236913 and FR230513 were selected as top-scored hits. Docking suggested that hydrogen bond interactions with residues Gly143,

Glu166 and Gln189 are crucial for ligand binding. The generated pharmacophore model contains three aromatic rings which indicate the possibility of stacking interactions with the residues. ADME/Toxicity prediction suggested that FR236913 has lower drug-likeness compared to other hit molecules. The stability of complex formed by four top-scored hits from the docking and co-crystal inhibitor N3 were analyzed by 100 ns MD simulation. The complexes of +/-fenoterol, FR236913 and FR230513 were found to be stable than the native inhibitor N3, from RMSD, RMSF and Rg values. The complexes of adrenoceptor inhibitors (+/-fenoterol) were stabilized by water-mediated hydrogen bonds with protein and ligand. PCA and FEL analysis suggests that flexibility of binding site residues helps the ligand to interact effectively with the enzyme. MM/PBSA calculations suggested that van der Waals energy and electrostatic energy are the crucial for the stability of complexes. The top-scored molecules predicted from molecular docking found to be an efficient inhibitor against main protease enzyme. The outcomes from this *in silico* study can be used to design and synthesize main protease inhibitors against novel coronavirus.

CHAPTER 7

UNDERSTANDING THE ROLE OF WATER ON TEMPERATURE-DEPENDENT STRUCTURAL MODIFICATIONS OF SARS COV-2 MAIN PROTEASE BINDING SITES

Thermally stable and labile proteases are found in microorganisms. Protease mediates the cleavage of polyproteins in the virus replication and transcription process. In this work, we study the effect of temperature and water on the structural modifications of SARS CoV-2 main protease, both, monomer and dimer.

7.1 BACKGROUND

Proteins are sensitive and behave differently in lower and higher temperature conditions. The temperature effect can either alter or break the non-covalent interactions thereby causing changes in the three-dimensional folding of the protein chain. Both, thermally stable, as well as labile proteases, are found in microorganisms, plants and animals. The human coronaviruses including SARS-CoV-2 showed a higher average mean lifetime in 10-60 °C temperature range and decontamination above 70 °C (Yap et al. 2020). Studies showed that the differences in an external environment like temperature disables the outermost structural proteins, spike protein, in the coronavirus to bind ACE2 (Angiotensin Converting Enzyme 2) (Rath and Kumar 2020). It is reported that the SARS-CoV-2 RBD (Receptor Binding Domain) affinity decreases at a higher temperature above 40 °C from both experimental and simulation studies (He et al. 2020; Rath and Kumar 2020; Zhou et al. 2021). As for the main protease enzyme, crystallographic studies have been carried out in cryogenic conditions, which bias enzyme conformations. Experimental evidence suggests that the protease enzyme exists in a monomer-dimer equilibrium, in which the dimer form is active whereas the monomer form is inactive (Darke et al. 1996; Khayat et al. 2004; Zhong et al. 2008). This equilibrium can be affected by temperature and small molecule inhibitors (Silvestrini et al. 2021). Theoretically, Chen et. al proposed that only a protomer is active in SARS 3C-like protease dimer

using MD simulations (Chen et al. 2006). Since SARS CoV-2 belongs to similar family and genus as SARS CoV, it is essential to understand the mechanism of protease action to develop efficient small molecule inhibitors. Many researches on drug discovery focussed more on SARS CoV-2 main protease as an excellent target (Achutha et al. 2020; Huff et al. 2022; Huynh et al. 2020; Maffucci and Contini 2020; Venugopal and Chakraborty 2021). So it will be interesting to study the dynamic behaviour of both monomer and dimer forms as well as the temperature/ water induced changes in the system. Recently, Kordzadeh et. al showed the temperature and pH-dependent conformational fluctuations on whole SARS CoV-2 main protease using molecular dynamics simulations (Kordzadeh and Saadatabadi 2021). This study ignores the structural and dynamic changes of protein and water structure at various binding sites of protease. Even though high-resolution crystal structure of SARS-CoV-2 M^{Pro} was recorded (Ebrahim et al. 2021), the molecular level mechanism and the solvent effect on M^{Pro} folding, stability and biological activity have not been reported for the wide temperature range.

In this work, we study the temperature effect on the structural modifications of SARS CoV-2 main protease, both, monomer and dimer using atomistic molecular dynamic simulations at four different temperatures. This allows us to understand the impact of temperature on the active dimer and inactive monomer main protease. Specifically, the role of amino acid residues in substrate binding pocket, catalytic dyad, dimerization site and allosteric sites in monomer and dimerized protease were determined using various geometric criteria of non-bonded interactions. The dynamic nature of water molecules at the interface of ligand binding sites is crucial for maintaining the activity of the protein. The interfacial water molecules may favour or disfavour the energetics of binding sites which affect the binding of chemical moieties (Cui et al. 2018; Morningstar-Kywi et al. 2022; Singh and Chakraborty 2021). Generally, this fact is overlooked while studying the degradation process of protein. Here, the effect of water molecules was studied using radial distribution functions, orientational tetrahedral order parameter, free energy $\Delta G(r)$ and hydrogen bond dynamics between key residues at the above-mentioned sites and water molecules. Dihedral analysis at these sites determines the transition between the secondary

structure elements in the protein (Dayalan et al. 2006). Principal Component Analysis (PCA) characterizes the essential dynamics which govern the amino acid residue fluctuations at the pocket due to temperature change. Further, the structural evolution and the stability of the whole monomer and dimer during simulation are interpreted using network analysis based on the root mean square deviation for the C α atom in the protein. The global free energy of folding was determined based on the changes in the percentage of α -helices and β -sheets for each temperature. Additionally, the results of simulations were compared using two water models: SPC/E model and modified TIP3P model.

7.2 MOLECULAR DYNAMICS SIMULATION PROTOCOL

The model of SARS COV-2 main protease (33.83 kDa) was retrieved from Protein Data Bank (PDB ID: 6Y2E, resolution: 1.75 Å) (Zhang et al. 2020). The enzyme model is devoid of small molecules and consists of 306 amino acid residues. The protein structure was solvated in a cubic box with distance between box boundary and protein structure as 1.0 nm in all directions using SPC/E water model (Hess and van der Vegt 2006) and the system is neutralized with 4 Na⁺ ions. The energy minimization of the system was done by the steepest descent algorithm with a maximum of 50000 steps until a convergence tolerance of 1000 kJmol⁻¹nm⁻¹. All the molecular dynamics simulations were performed by GROMACS 2018.4 (Abraham et al. 2015) using CHARMM27 force field for proteins (MacKerell et al. 1998; Mackerell et al. 2004). A 10 ns NVT equilibration was performed using velocity-rescale temperature coupling method (Bussi et al. 2007) with a time step of 2 fs. Next, a 10 ns NPT equilibration was carried out with a time step of 2 fs at 1 atm using Nose-Hoover thermostat (Martyna et al. 1992) and Parrinello-Rahman pressure coupling scheme (Parrinello and Rahman 1981) respectively. The long-range electrostatic interactions were calculated by particle mesh Ewald method (Essmann et al. 1995) with a cut-off 1.2 nm and Fourier spacing of 0.16 nm. The short-range van der Waals cut-off was fixed to 1.2 nm. LINCS constraints (Hess et al. 1998) were used to restrain the bond involving hydrogen atoms. The structural and dynamic behaviour of the protein were analyzed for a wide temperature range, from 278 K to 383 K. Finally, a production run of 200 ns was performed until the convergence of

RMSD at lower temperature. The trajectories were saved at every 10 ps and GROMACS tools were used for analysis. The simulations were replicated twice (a total of 8 monomer simulations) to check the reproducibility of the results and error bars are given wherever necessary. The model retrieved from PDB is the monomer form of protease. However, the global stoichiometry of the protease is homodimer (Sun et al. 2022; Zhang et al. 2020). Therefore, for the dimer simulations, protein-protein docking was carried out using HDOCK server (Yan et al. 2020). The structure retrieved from PDB (ID: 6Y2E) is considered as the target and the average structure obtained from the MD simulation of monomer at each temperature is considered as the substrate for protein-protein docking. The docking energy score obtained between the ligand and the target protease is -555.86 kcal/mol. The geometry of dimer obtained from docking is chosen as the initial configuration for the dimer simulations. We have done a total of 8 dimer simulations. Further, a set of 200 ns simulations (4 simulations each for monomer and dimer) were performed using modified TIP3P water model (m-TIP3P) to determine the effect of water model on the temperature dependent behaviour of main protease. Additional, five 200 ns simulations each for 288 K, 298 K, 308 K, 318 K, 338 K and 358 K were performed for monomer protease to determine the entropy and enthalpy contributions to the free energy. A total of 6 μ s simulations were performed.

7.3 TEMPERATURE EFFECT ON M^{Pro} ENZYME

The temperature-dependent stability of monomer and dimer were studied based on structural parameters and solvent effect. 200 ns molecular dynamics simulations were performed using SPC/E and m-TIP3P water model at 4 different temperatures, 278 K (cold denaturation), 310 K (physiological), 348 K (higher temperature reported for enzyme activity (Lawyer et al. 1993)) and 383 K (extreme thermophilic conditions (Vieille and Zeikus 2001)), to obtain the insights into the conformational evolution and stability of SARS CoV-2 main protease enzyme. The dimer is composed of two monomer proteases, which are denoted as chain A and chain B. Both the chains are showing similar results (**Appendix XXIV**). Therefore, we selected one chain for comparing the structure and dynamics of the main protease dimer form with the monomer. For the dimer, the monomer-monomer interaction energy is calculated

between the two chains of the dimer. At 310 K, the interaction energy is found to be stable than other temperatures.

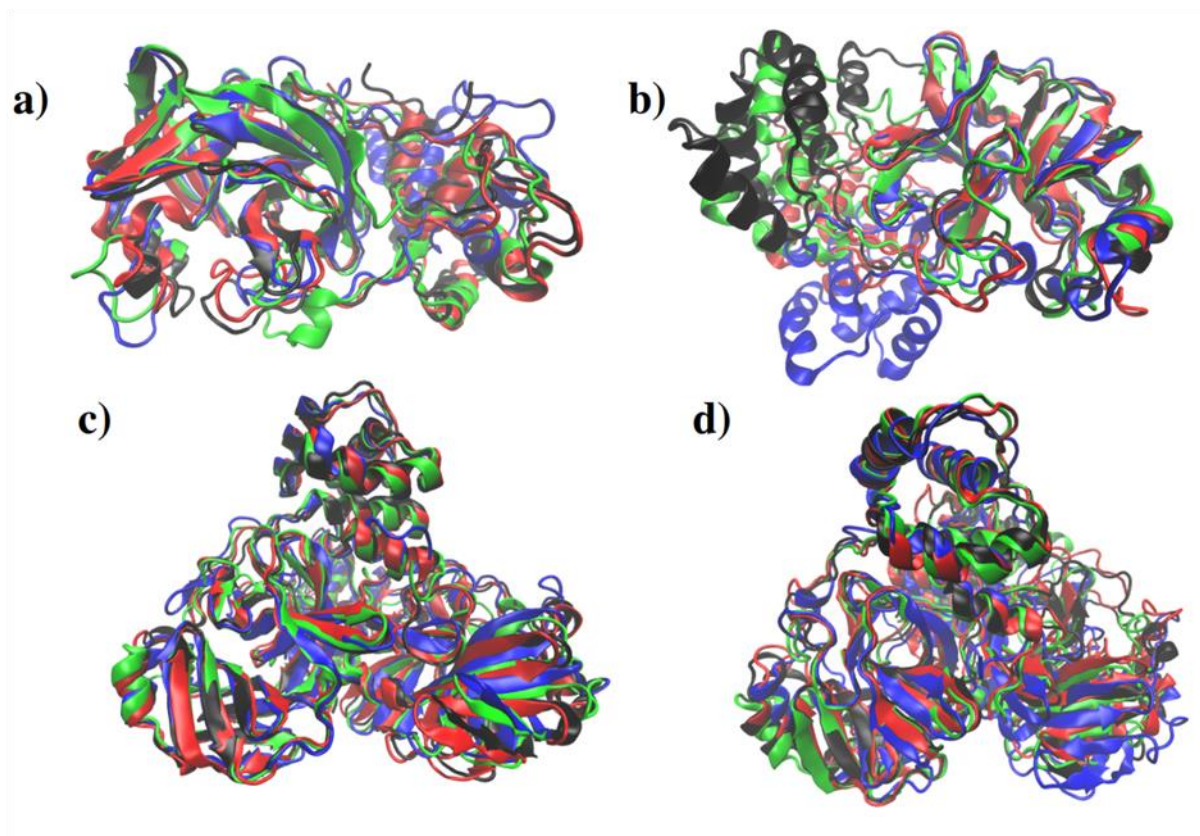


Figure 7.1: Global structures of SARS CoV-2 M^{Pro} monomer (**a, b**) and dimer (**c, d**) at 278 K (black), 310 K (red), 348 K (green) and 383 K (blue) in SPC/E (**a, c**) and m-TIP3P (**b, d**) water model.

The global structure of SARS CoV-2 main protease protomer and dimer is found to be similar despite the increase in temperature and difference in the water model as shown in **Figure 7.1**. Per residue secondary structure analysis of protease using do_dssp tool (Kabsch and Sander 1983) at different temperatures shows no considerable changes in the average occupancy of total secondary structure components in the protein (**Appendix XXV**). The percentage of secondary structure element obtained by do_dssp tool can be correlated with the calculated percentage from X-Ray diffraction technique (α -helix: 0.31%, β -sheet: 0.27%, unstructured region: 0.42%). However, it can be observed that β -bridges and α -helices are formed with the increase in temperature whereas native α -helices are destroyed in the structure. The secondary

structure element plots reveal the degradation of the enzyme with the rise in temperature which confirms the change of the M^{Pro} conformational space (**Appendix XXVI**). It is found that the monomer and the dimer-chain A conformations at 310 K (human body temperature) are more stable than other temperatures with ΔG_{fold} values of -3.20 ± 0.25 and -3.43 ± 0.13 kJ/mol (**Figure 7.2**).

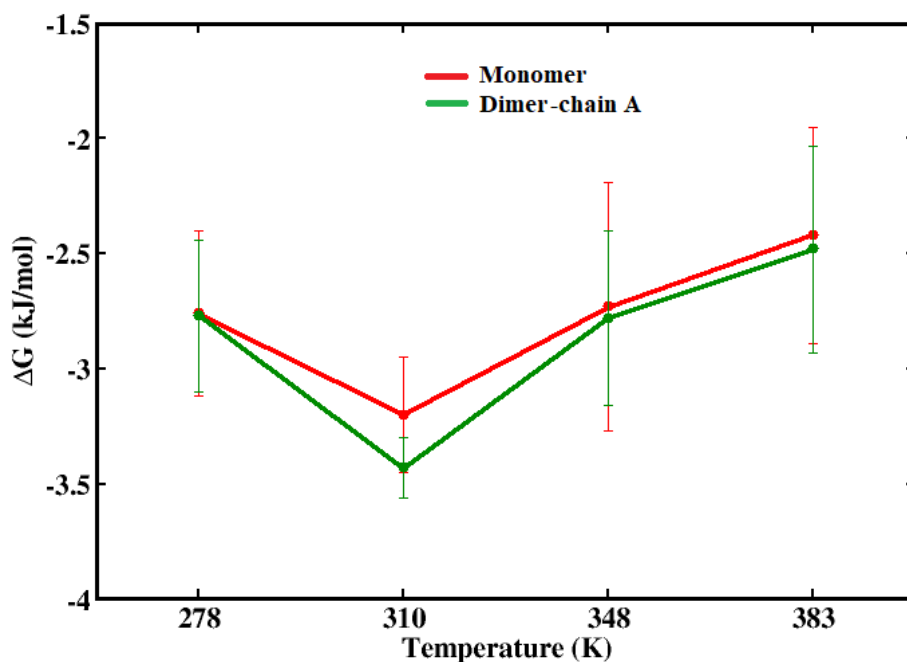
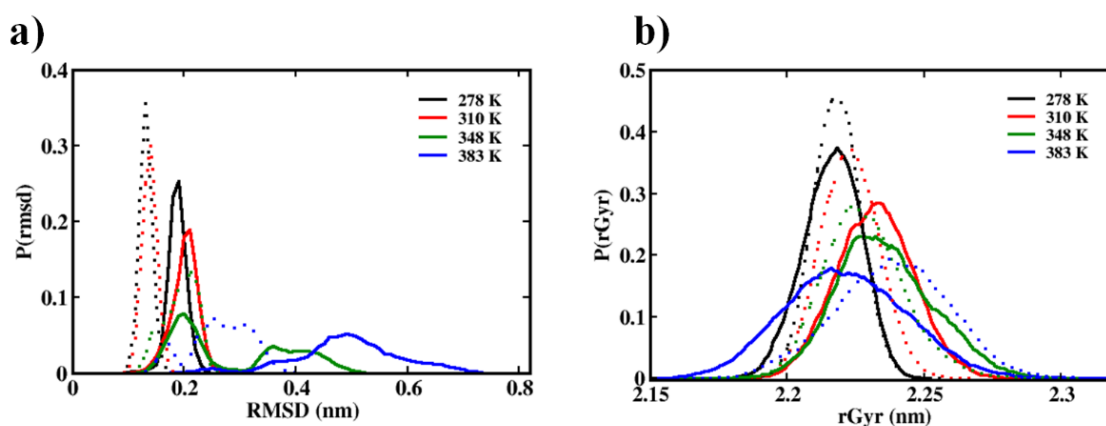


Figure 7.2: Global ΔG from folding/unfolding of protease vs. Temperature, calculated by eqn. 2.45.

The probability distribution of C_{α} -atom RMSD for both monomer and chain A of the dimer (dimer-chain A) in SPC/E water model is shown in **Figure 7.3(a)**. It is found that at 278 K and 310 K, the protein has a single probable structure for both dimer-chain A and monomer. At 348 K, the conformation of protease gradually evolves during simulation due to thermal fluctuations. The system of monomer has well equilibrated within 25 ns of the simulation time with a value of 0.187 ± 0.02 and 0.200 ± 0.02 nm at 278 K and 310 K respectively. At temperatures 348 K and 383 K, marginally higher RMSD values are observed. The dimer-chain A has achieved the rigidity after dimerization is evident from the fewer RMSD fluctuations with a value of 0.131 ± 0.01 and 0.137 ± 0.01 nm at 278 K and 310 K respectively. It can be observed that the rigidity is lost with the rise in temperature as seen in 348 K and 383

K, which is evident from the tail region. The compactness of the protease monomer decreases with rising in temperature due to the thermal fluctuations at the α -helical region of the protein (Lobanov et al. 2008). At 383 K, due to random motion of the protein molecules, frequent contacts are made and broken at different stages which are not stable resulting to broader distribution. Similar trend is observed for dimer-chain A (**Figure 7.3(b)**). However it is observed that the monomer is more compact than the dimer at 383 K. Further, this observation is supported by the correlation of inter-residue distance (**Appendix XXVII**). The correlation map shows the contact formation (orange region) and contact rupture (blue region) between amino acid residues. At 348 K, the non-bonded interactions were broken compared to the lower temperature. At 383 K, contact formation is more than the contact rupture (Mercadante et al. 2018). It is evident from the RMSD and the Rg values that the dimer-chain A has fewer fluctuations than the single-chain protease at lower temperatures. At 348 K and 383 K, both the systems behave differently compared to lower temperature. SASA provides the surface area of protein exposed to the solvent molecule. A change in SASA value indicates a change in the tertiary conformations of the protein (**Figure 7.3(c)**). The main protease has the lowest SASA value at 278 K and 310 K indicates that the residues are less exposed to the environment. This also suggests that the protein undergo considerable thermal degradations at 348 K and 383 K in comparison to lower temperature. The SASA value for monomer is higher at 348 K suggesting higher solvation around the protein due to the degradation of protein. This anomaly is observed because 348 K is the melting temperature for the main protease. The dimer-Chain A showed a similar SASA value as the monomer at 278 K, 310 K and 383 K which suggests that the interaction of two monomers does not affect the solvent accessibility of monomer units. At 348 K, the single peak SASA value for dimer-chain A is observed which is quite different from the monomeric form. From secondary structure analysis, it is found that at 348 K, monomer undergoes thermal fluctuations leading to the degradation of α -helix and β -sheet at domain I and domain II respectively (**Appendix XXVI**). However in case of dimer, the interaction of two protease chains protects the secondary structure at domain II and domain III. Therefore, at 348 K, the solvent exposure is less in dimer as compared to monomer unit. Among the dimer, there is an increase in the SASA value with rise in

temperature. Hydrogen bond interactions between protein-protein and protein-water are considered an important factor for the stabilization and activity of protease enzyme. **Figure 7.3(d)** shows the average number of hydrogen bonds formed between protein/protein and protein/water. It is found that there is no significant decrease in number of intra-protein hydrogen bonds when the temperature rises from 288 K to 310 K while there is a significant reduction in the average number of hydrogen bond when the temperature rises to 348 K and 383 K. Both monomer and dimer-chain A show a decrease in the number of hydrogen bonds with temperature change. Similarly, protein-water hydrogen bonding showed a gradual decrease with the temperature rise. The number of water molecules decreases with the temperature inside the binding cavity which suggests the expansion of protein structure in both monomer and dimer-chain A. However, at higher temperatures, a tail region is observed as the water molecules can easily move in and out due to the cavity expansion. The number of water molecules in the substrate-binding site is shown in **Figure 7.3(e)**. The isotropic temperature factor (B-factor) was calculated by eqn. 2.30. Apart from the two terminal ends of the protein, the highest fluctuations are observed for amino acid residues in the range Arg40- Glu55 and Asn180- Ile200 at higher temperatures (**Figure 7.3(f)**). It is observed that from 348 K- 383 K, the amino acid residues from Lys100- Phe150 show higher fluctuations compared to the lower temperature. It is believed that residues in the regions Arg40- Glu55 and Lys100- Phe150 contribute to the catalytic activity of the protease (Ferreira et al. 2021). At higher temperatures, the fluctuations are found to be higher in the above-mentioned regions for single chain and dimerized main protease.



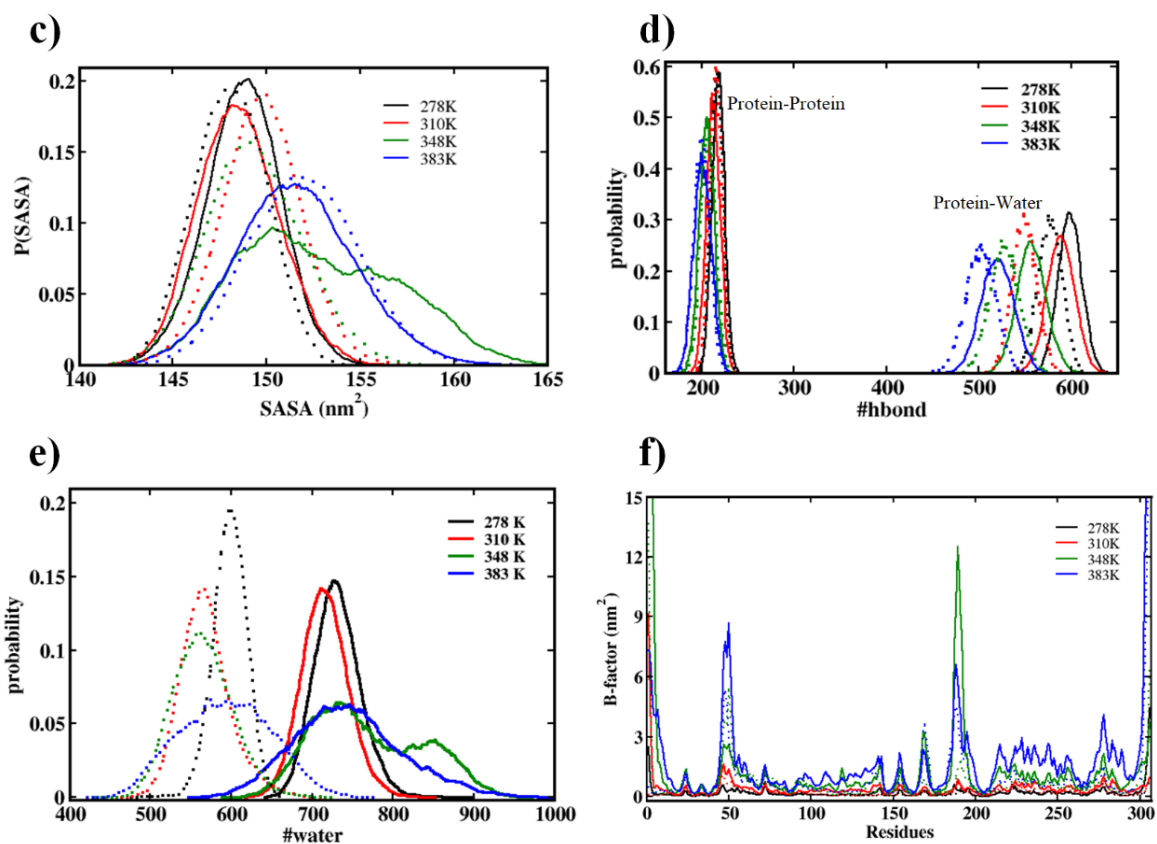


Figure 7.3: Probability distribution of (a) C_α-RMSD (b) radius of gyration of protein (c) SASA (d) Number of M^{Pro}-M^{Pro}/ M^{Pro}-water hydrogen bonds of protein (e) number of water molecules in the substrate binding site during 200 ns simulation and (f) Calculated B-factor of C_α atoms with amino acid residues for M^{Pro}. The solid and dotted lines represent monomer and dimer-Chain A respectively in SPC/E water model.

7.4 TEMPERATURE EFFECT ON VARIOUS BINDING SITES

Even though the structural analyses such as RMSD, Rg, RMSF and SASA give information about the global changes in the protein structure, it may ignore the temperature-induced conformational changes at the ligand-binding sites and dimerization site. The structural analyses were performed selectively for important amino acid residues as given in **Table 7.1**. The catalytic dyad and the substrate-binding site are located between the cleft of domain I and domain II while the dimerization site and allosteric site is regulated mostly by domain III (Mengist et al.

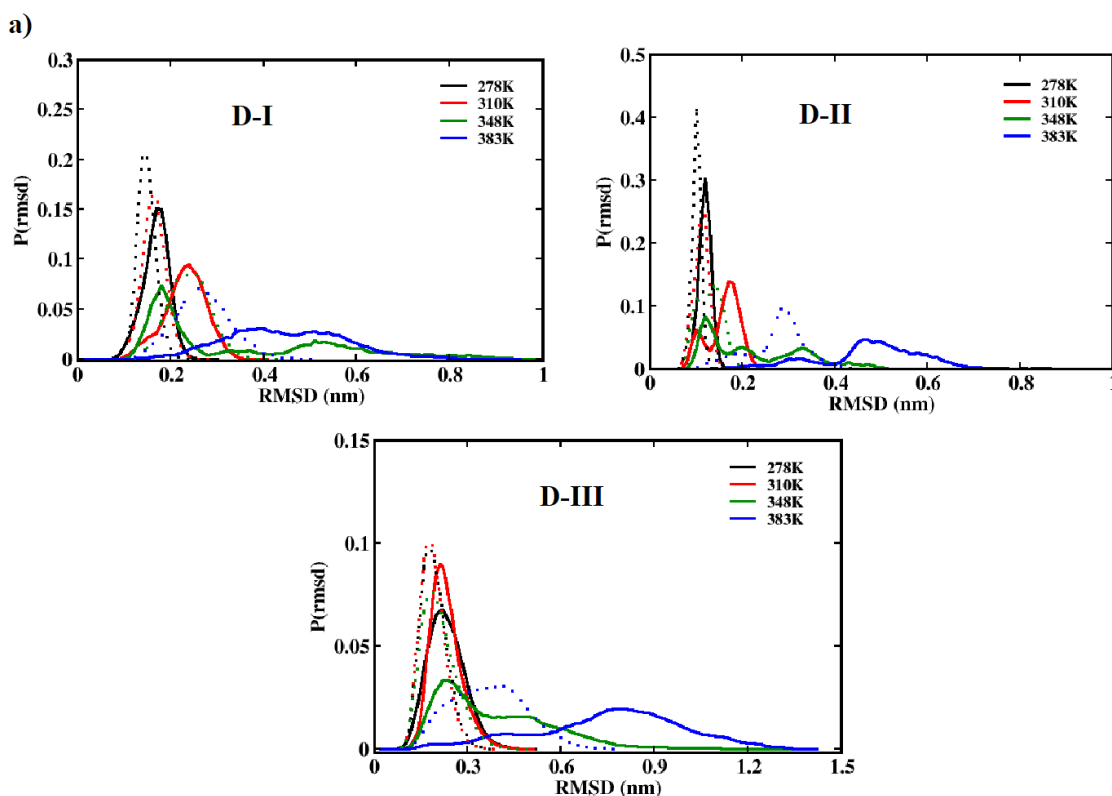
2021). Domain I consists of 10-99 residues which can be regarded as ($\alpha + \beta$) regions, domain II consists of 100-182 residues which are β -sheet regions and domain III consists of 198-303 residues which are α -helix regions.

Table 7.1: Important amino acid residues taken for analysis and their biological function

Residues	Function/ Role in M ^{Pro}
His41 , Cys145	Catalytic dyad (Huang et al. 2004)
His41, Met49, Gly143, Ser144, His163, His164, Met165, Glu166 , Leu167, Asp187, Arg188, Gln189, Thr190, Ala191, Gln192	Substrate binding site (Mengist et al. 2021)
Arg4, Ser10, Gly11, Glu14, Asn28, Cys117 , Ser139, Phe140, Ser147, Glu290, Arg298	Dimerization site (Goyal and Goyal 2020)
Met6, Phe8, Gln127, Ser139, Asp295, Arg298 , Gln299	Allosteric site (El-Baba et al. 2020)

These structural properties for the amino acid residues present at substrate-binding site, dimerization site and allosteric site were determined for different temperatures rather than the whole protein for a better understanding of conformational changes occurring at the binding pockets. The basic structural analyses of binding site amino acid residues are shown in **Figure 7.4**. The C _{α} RMSD profile of amino acid residues present at various binding sites with respect to the domain (I, II, III) movements in monomer suggests different dynamics of protease binding sites at lower and higher temperatures. The higher RMSD values at 348 K and 383 K indicate the flexible nature of all the three domains in the protein as well as the binding site residues. Domain I and domain III motions are more flexible compared to domain II due to the α -helix regions (**Appendix XXVIII**). The highly flexible nature of domain III at higher temperature may have implications on the dimerization of protease which is crucial for its biological function. Similar behaviour is observed in dimer-chain A for

lower temperatures. Unlike the monomer, the single chain in the dimer showed a stable movement at 348 K. At 383 K, the chain became unstable as expected. The R_g values indicate that with temperature rise the compactness of binding site residues decreases. In case of monomer form, the compactness of binding site residue decreases with rise in temperature except 310 K. The R_g value at 310 K is due to the stable global structure of protease as evident from ΔG_{fold} value. In dimer form, the average value of R_g at each temperature is almost similar, while the value of tail region is increasing which suggests the loss of rigidity with rise in temperature. This is because most of the residues lie in the α -helical region which is more flexible than the β -sheet element. An increasing trend is observed for the solvent-exposed surface area of the protease single chain and dimer. Exception at 348 K for SASA value is due to the presence of more coiled structure.



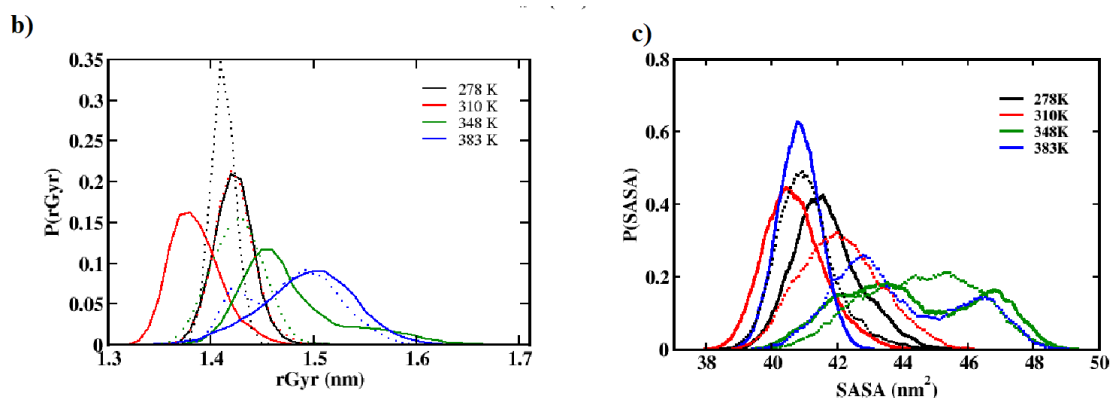


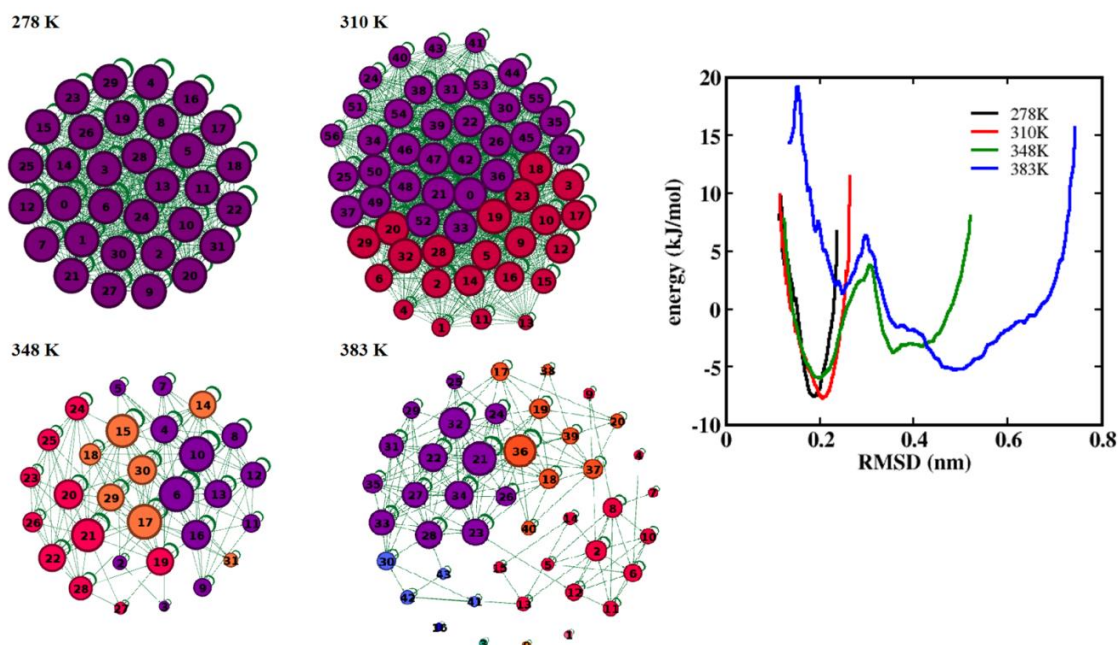
Figure 7.4: Probability distribution of (a) C_{α} -RMSD of binding site residues with respect to D-I, D-II, D-III of protein (b) radius of gyration of binding site residues (c) SASA of binding site residues. The binding site residues are considered as per **Table 7.1**. The solid and dotted lines represent monomer and dimer-Chain A respectively in SPC/E water model.

7.5 CONFORMATIONAL DYNAMICS OF THE PROTEIN

Further, we evaluated the change in the local structures in the conformational space due to the thermal fluctuations by network analysis using RMSD as the criterion. Around 60-80 prominent structures which cover the major conformations of protein were extracted from 200 ns trajectory for each temperature using gromos clustering method (Daura et al. 2004). Next, to find out the similarity/connections between two conformations, we calculated the RMSD between these conformations. Two conformations were considered to be connected if the RMSD value between them is less than 2.0 \AA (Carugo and Pongor 2008). Since we compared the change in the conformation for all the temperatures, we preferred RMSD criteria 2 \AA where we get a single prominent conformation for 278 K. Conformations obtained from other temperatures are compared with respect to 278 K. The network distribution for monomer and dimer-A is given in **Figure 7.5** with communities represented in different colours. It can be seen that only a single community is present at 278 K and are densely connected to each other. At 310 K, two communities are mainly found where the violet community corresponds to evolved conformations and is densely connected. At higher temperatures, many communities with fewer connections are

observed which explains the conformational evolution due to thermal fluctuations. This trend is further confirmed by RMSD based PMF (potential of mean force) calculation (eqn. 2.28, **Figure 7.5**). It can be seen that at 278 K and 310 K, the monomer has similar free energy minima around -7.5- -7.7 kJ/mol with different RMSD range at 0.187 and 0.205 nm respectively. For dimer-chain A, a single minimum in the range of -8.4- -9.0 kJ/mol with RMSD 0.13 and 0.14 nm is observed at 278 K and 310 K respectively. With the increase in the temperature, the free energy curve becomes shallow having multiple global minima. This explains the presence of structures with different energy values. These results correlate well with the network distribution results. The above results suggest that main protease possesses unique structural and dynamic properties at lower and higher temperatures. The energy levels for the monomer are found to be much shallower than the dimer at higher temperatures. The unique structural and dynamic properties observed at different temperatures are further analyzed by PCA analysis and deviation in dihedrals angles considering the various binding sites.

a)



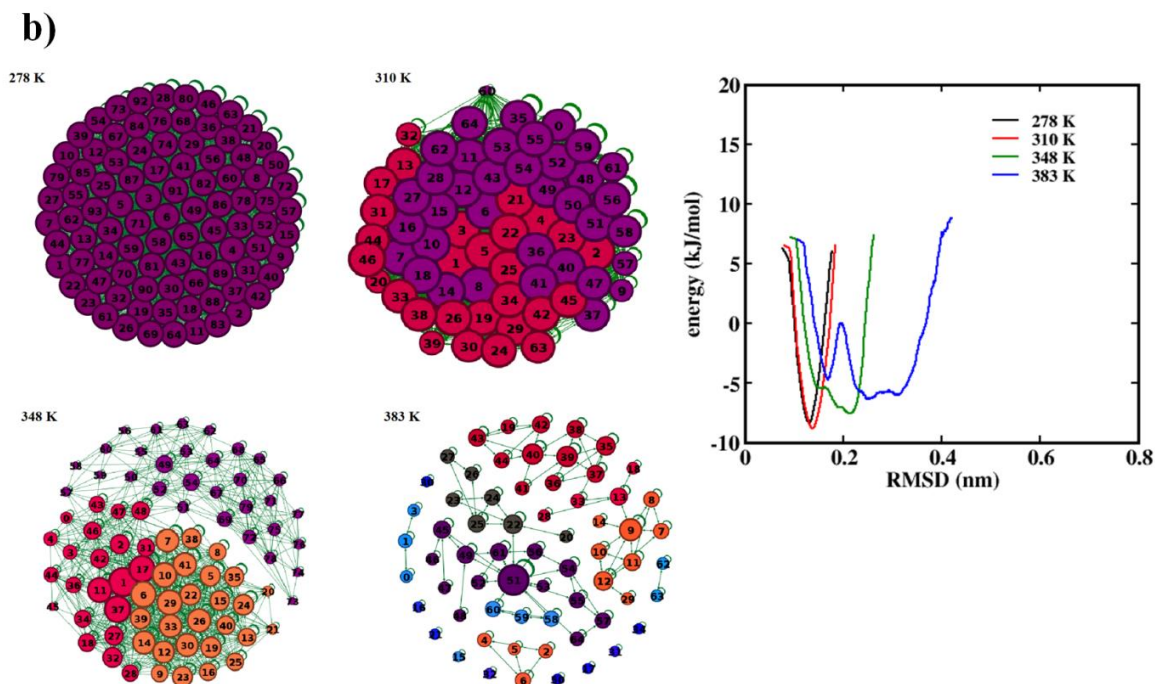


Figure 7.5: Network distributions of most prominent conformations of main protease from 200 ns trajectory at different temperatures in SPC/E water model. PMF profile for SARS CoV-2 main protease as a function of RMSD (Right) (a) monomer (b) dimer-Chain A.

7.6 PRINCIPAL COMPONENT ANALYSIS

PCA was performed for backbone C_{α} atoms of key amino acid residues in both the monomer and dimer-Chain A to characterize the essential dynamics which govern the conformational changes throughout 200 ns simulation (**Figure 7.6**). The first eigenvectors correspond to all the possible motions induced by the temperature. The eigenvalue plot indicates the highly flexible nature of the binding site residues. For better representation, the first two eigenvectors are plotted in 2D projection plot for both monomer and dimer enzyme binding pocket. In the case of the monomer, the scatter plot spread over the phase space more than the dimer. Also, it is observed that at 278 K and 310 K, the conformational changes are not very different but at 348 K and 383 K, the plots deviated to a great extent indicating the temperature-induced flexibility in the protein. The binding site of dimer-Chain A is less flexible due to the structural stability achieved during the dimerization. The PCA analysis shows the

flexibility of amino acid residues at the binding pocket of the monomer, agreeing with the structural analysis. Even though the dimer-Chain A is less flexible compared to the monomer, it is found that the flexibility of the binding site increased with temperature rise. The structural flexibility observed at different temperatures is characterized by deviation in dihedral angles considering the various binding sites.

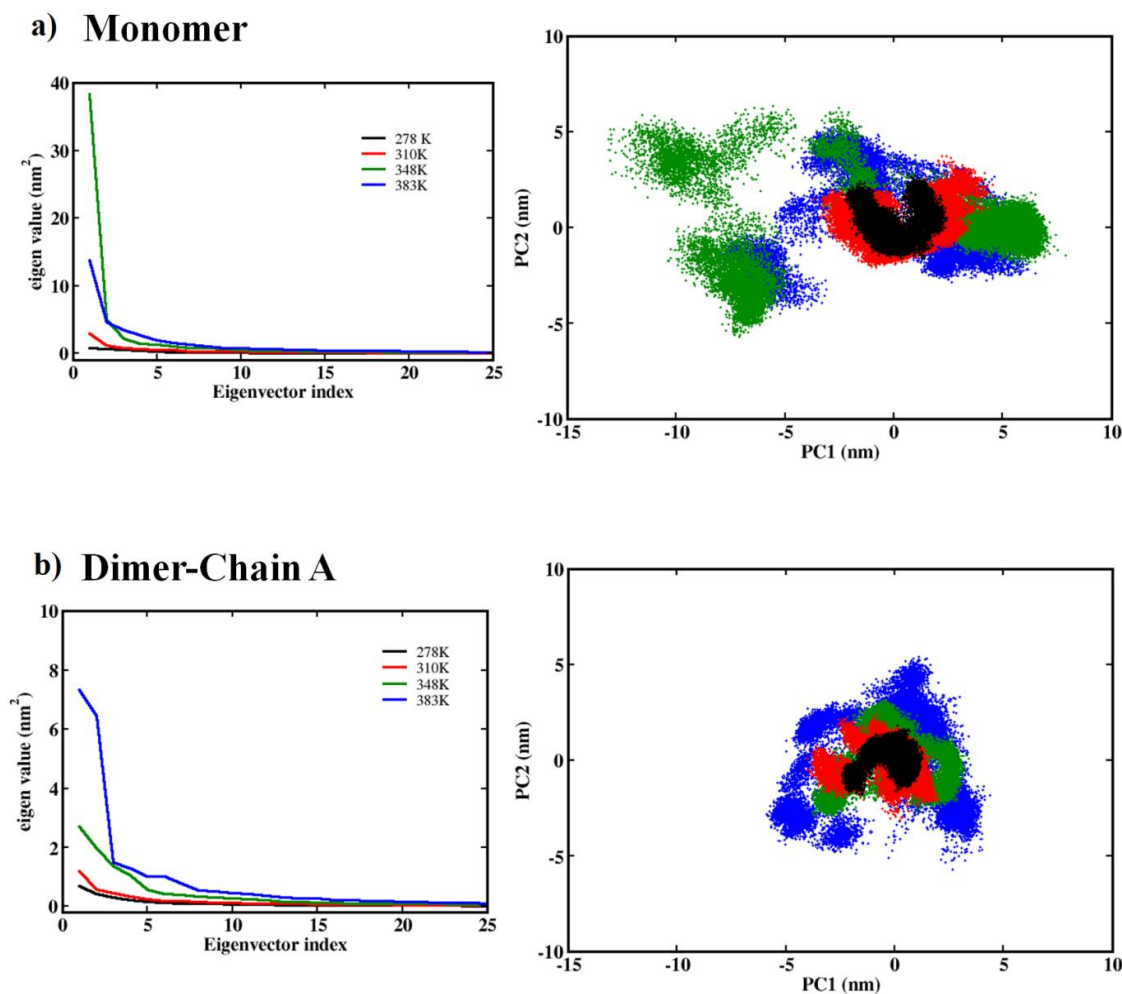


Figure 7.6: Plot of eigenvalue vs first 25 eigenvector index derived from PCA over 200 ns MD trajectory for (a) M^{Pro} monomer and (b) M^{Pro} dimer-Chain A. 2-D projection of first two principal motions (PC1 and PC2) of main protease (Right).

7.7 CHANGES IN DIHEDRAL ANGLES AT VARIOUS BINDING SITES

The changes in the backbone of protein residues can be characterized with the help of dihedral angles (ϕ , ψ). The deviations in ϕ , ψ angles provide insight into the changes

in the secondary structure elements due to the temperature-dependent fluctuations at these binding sites. The time evolution of dihedral angle deviations for the key residues at four different binding sites is presented in **Appendix XXIX**. Each dot in the scatter plots shows the combination of ϕ , ψ angles of residues and the colour indicates the time scale at which the current combination is present. It can be seen that broader distribution of ϕ , ψ angles are observed at 348 K and 383 K except for amino acid residues at the catalytic dyad. At lower temperatures, the catalytic dyad is found to be intact with fewer fluctuations providing better stability and biological activity to the protein. Even though deviations are observed for 348 K and 383 K, the residues showed a tendency to revert to the initial conformation during simulation. At 278 K, the residues at the substrate-binding site, dimerization site and allosteric site tend to remain in the initial structure during the simulation. With temperature rise, it can be seen that there are more deviations in ϕ , ψ angles which might be due to the changes in the local and non-local interactions. This includes the making and breaking of non-covalent interactions such as hydrogen bonds, salt bridges, van der Waals interactions and aromatic interactions. This also suggests the local structural changes at the secondary structure elements of protein binding sites which might get neglected in global conformational analysis. The results for dimer-chain A show a similar trend as monomer (**Appendix XXX**).

It became more important to see the water role at the binding site, as there are many deviations at the protein binding sites with rise in temperature. The structure and dynamics properties of water molecules were calculated to understand the role of water in the previously mentioned binding sites as given in (**Table 7.1**). The amino acid residues, His41, Cys117, Glu166 and Arg298 were selected from the catalytic dyad, dimerization site, substrate-binding site and allosteric site respectively to analyze the role of water. These residues are selected based on their interaction with the organic moieties/inhibitors at four different binding sites.

7.8 RADIAL DISTRIBUTION FUNCTION

The radial distribution functions (RDF) between the pair of atoms are used to analyse the structural properties of the system. For this, we have plotted the RDF between

oxygen atoms and hydrogen atom of water molecules $g_{ow-ow}(r)$, $g_{Hw-Ow}(r)$ (**Figure 7.7**) and $g_{C\alpha-Ow}$ pair correlation functions of His41, Cys117, Glu166 and Arg298 amino acids (**Figure 7.8**) at four different temperatures. It can be seen that there are three distinct peaks in **Figure 7.7(a)**. The first peak is the highest and sharp, which is found at the separation of about 0.345 nm from centred atom's position. The second and third peaks are relatively shorter and wider, which are located approximately at positions 0.57 nm and 0.68 nm, respectively. It can be seen that the peak height of the RDFs decreases with the increase in temperature and higher solvation shell peaks are not well defined. Therefore, it can be concluded that there is a change in the hydration shell of the water molecules with the rise in temperature which becomes less compact.

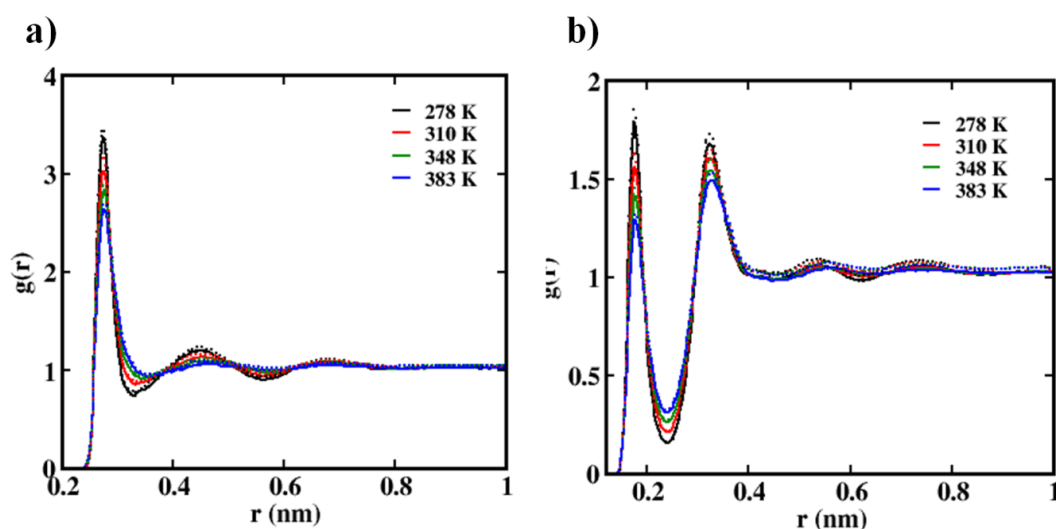


Figure 7.7: RDF of (a) Oxygen–Oxygen and (b) Hydrogen–Oxygen of water molecules at different temperatures.

Further, to study the detailed overview of the arrangement of water molecules around C_{α} atom of the amino acids mentioned at the binding sites, we plotted the C_{α} -Ow RDF around His41, Cys117, Glu166 and Arg298. Therefore, we calculated the C_{α} -Ow RDF of four selected amino acid residues at different temperatures in **Figure 7.8**. Here also, it is found that the solvation of the water molecules becomes broader with the increase in the temperature. In the case of the monomer, the water molecules are more altered around C_{α} atom of His41 at lower temperatures and showing the sharp peak but at higher temperatures the peak is broad. It is observed that Cys117 have fewer neighbouring water molecules while Glu166 showed a narrow and sharp

peak (**Figure 7.8 (c)**) which suggests that the water molecules present here are more ordered around the binding site at lower temperature. In the monomer, the C_{α} -Ow RDF of Arg298 indicates the low number of water molecules is present in the first solvation shell, thereby less interaction with the water. In the case of the dimer, the C_{α} -Ow RDF of all selected amino acids has a narrow and sharp peak which indicates the homogeneity of the water molecule compared to the monomer. Further, it would be interesting to check the tetrahedral arrangement of water molecules around the C_{α} atom of key amino acid residues in the solvation shell by orientational order parameter which is discussed in a later section.

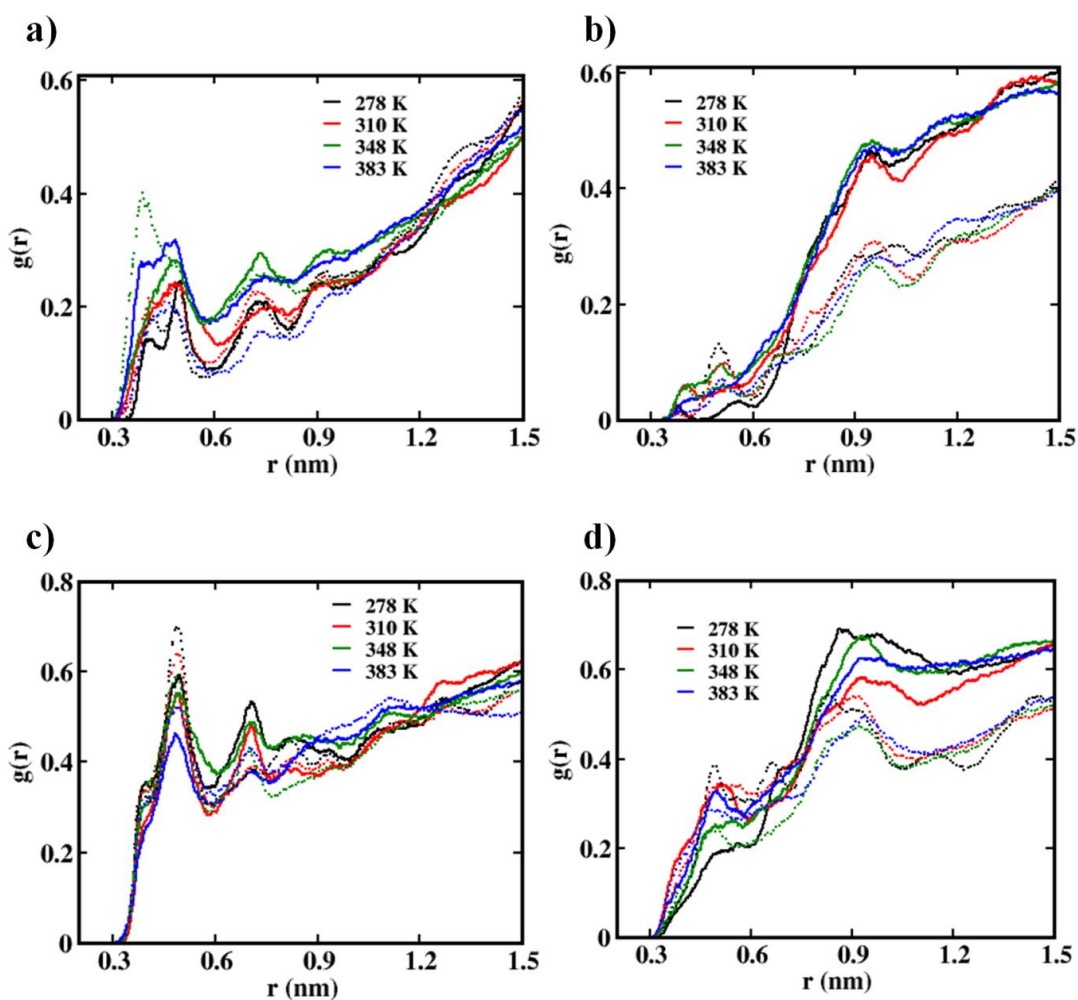
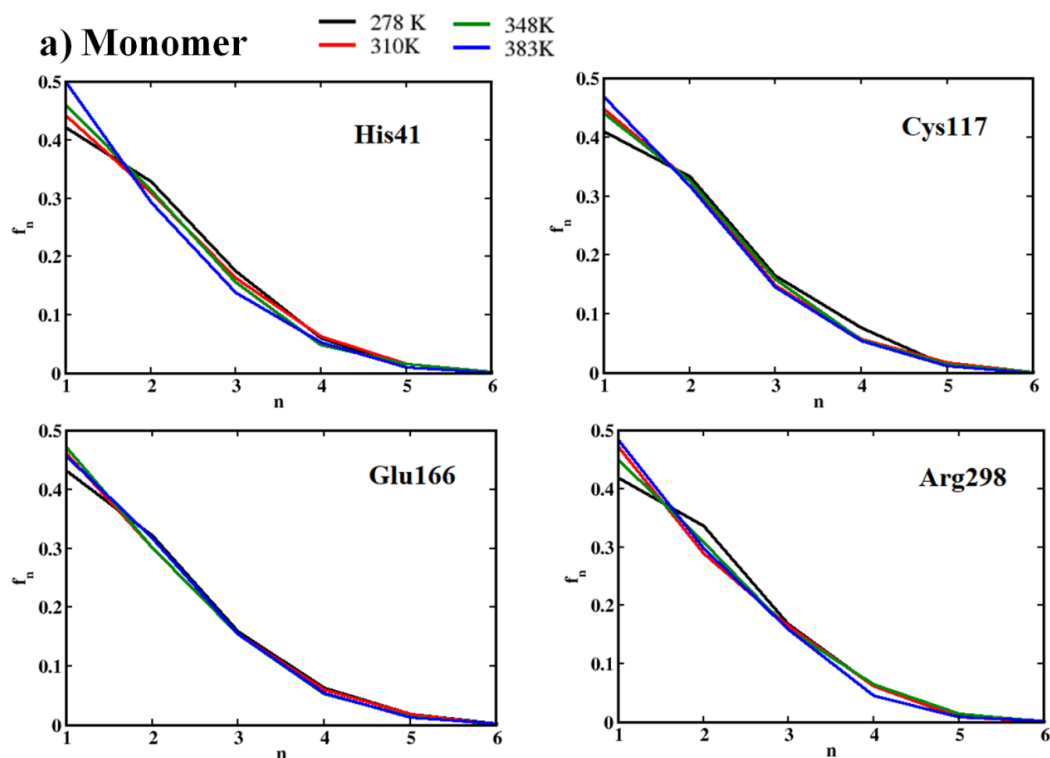


Figure 7.8: Radial distribution function of C_{α} -Ow of key amino acid residues at various binding site of protease at different temperatures in SPC/E water model. Solid line and dash line represent monomer and dimer-Chain A respectively. **(a)** His41 **(b)** Cys117 **(c)** Glu166 **(d)** Arg298.

7.9 NUMBER OF HYDROGEN-BONDED WATER MOLECULES

Further, the distorted structure of water molecules near the interface of His41, Cys117, Glu166 and Arg298 residues was confirmed by calculating the fraction (f_n) of oxygen atoms of water molecules that contains n number of water-water hydrogen bonds. The hydrogen bond distance criteria between oxygen-oxygen are considered as 3.25 Å, to incorporate the flexible hydrogen bonds (Singh and Chakraborty 2021). A cut-off value of 6.0 Å was selected to define the interfacial water molecules present around C_α of selected amino acid residues of protease based on the first solvation shell in the RDF. The distance criteria will select the neighbouring water molecules within this cut-off value obtained from RDF.

In all the cases, it has been observed that the probability of lower coordinated (mono-coordinated) water molecules increases at higher temperatures (Figure 7.9). At the lower temperature, the fraction of higher coordinated water molecules is higher as compared to 348 K and 383 K. At the higher temperature the probability of higher coordinated water molecules is less. Further, we have calculated the tetrahedral order parameter to see the orientation of the water molecules around the binding site residues.



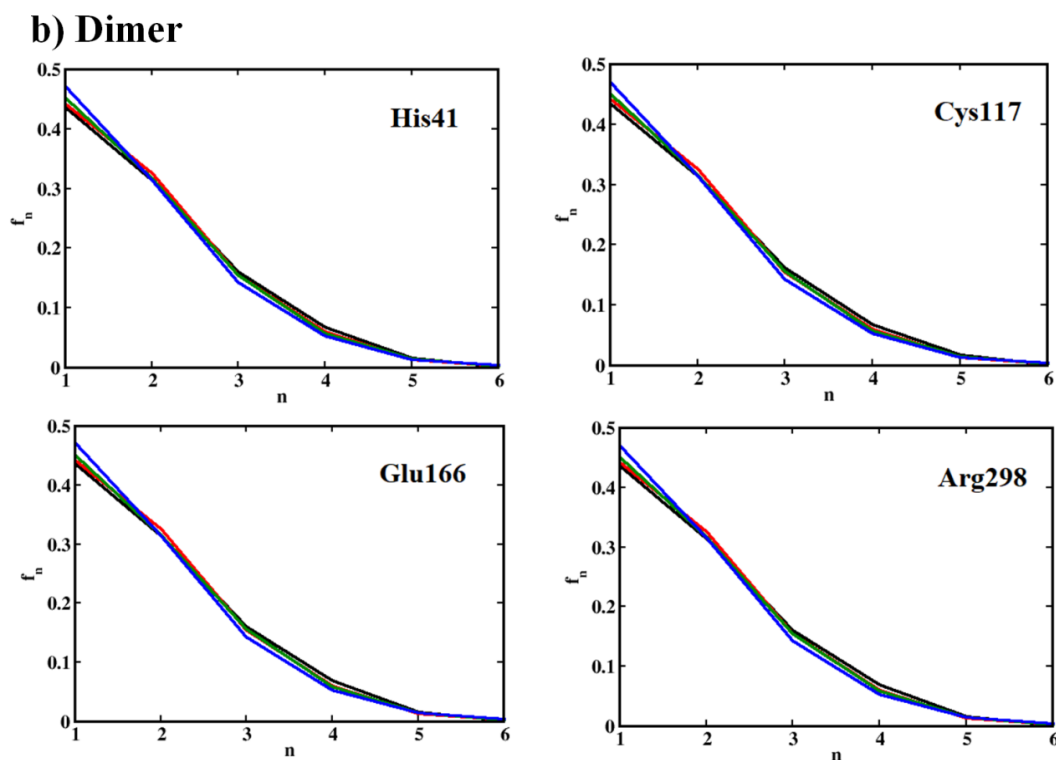


Figure 7.9: Fraction of water molecules having n number of hydrogen bonds within a distance of 6.0 \AA from C_α of amino acid residues of (a) M^{Pro} monomer and (b) dimer-Chain A for SPC/E water model at different temperatures.

7.10 ORIENTATIONAL TETRAHEDRAL ORDER PARAMETER

The structure of water molecules near the protein surface can be characterized based on the tetrahedral order parameter (Section 2.7.4). The probability distribution of S_g is plotted for the water molecules which are present at a cut-off of 6.0 \AA from C_α atoms of His41, Cys117, Glu166 and Arg298 residues selected from various ligand binding sites (**Figure 7.10**). It can be seen that the peaks corresponding to the monomer are broader compared to the dimer, which suggests more heterogeneity of the water molecules in the monomer compared to dimer-chain A. The sharp and narrow distribution of S_g for dimer-Chain A indicates less heterogeneity of the water molecules. This homogeneous distribution of water molecules in dimer systems are observed irrespective of the temperature.

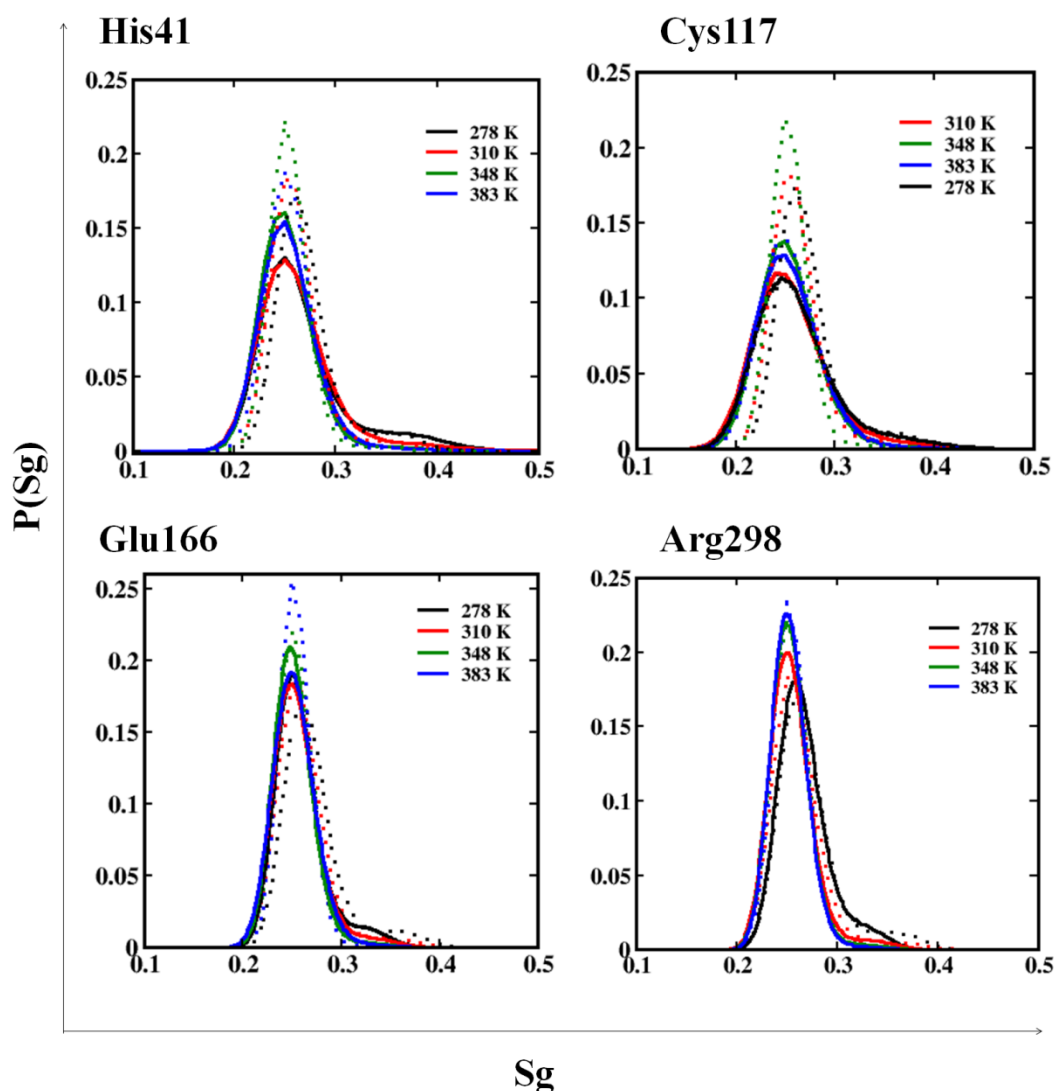


Figure 7.10: Orientational tetrahedral order parameter of water molecules within 6.0 Å from C_{α} of key amino acid residues at various binding site of monomer (solid line) and dimer-Chain A (dash line) in SPC/E water model at different temperatures.

In the case of monomer, a tail region towards the higher S_g value is observed indicating that the water molecules have a more distorted tetrahedral arrangement. Hence, the water structure in the first solvation shell around the binding site residues of the monomer has a more disrupted and distorted tetrahedral structure with a lower density as compared to dimer-Chain A. Further, with the increment of the temperature this tail region decreases both for the dimer and the monomer which suggests that the water molecules are randomly and homogeneously distributed at the interface of the

protein at higher temperature as compared to the lower temperature. This can be explained based on the interstitial water molecules present in the cavity formed from the binding site residues. The interstitial water molecules can lead to non-tetrahedral arrangement of water molecules called high density water (Biswas et al. 2018). With the rise in temperature, the interstitial water molecules moves out of the cavity due to thermal expansion, which may leads to more tetrahedral like water at higher temperatures. Next, we focused on the thermodynamic properties in terms of free energy, entropy and enthalpy of the water molecule with the binding site amino acid residues.

7.11 FREE ENERGY vs. TEMPERATURE

The hydrogen-bonding network between the water molecules is perturbed due to the interaction of water molecules with these amino acids residues with the rise in temperatures which in turn changes the free energy. The thermodynamic properties were determined by Finite-Difference Approximation (Section 2.7.6) (Pettitt and Rossky 1986; Smith and Haymet 1993). Additional 5 MD simulations were performed to calculate the entropy-enthalpy contributions as shown in **Table 7.2**. The hydration shell formed by the water molecules around the C_{α} of binding site amino acids residues at 288 K, 308 K and 348 K is shown in **Appendix XXXI**. The RDFs for C_{α} -O_w pairs for additional temperatures (298 K, 318 K, 338 K and 358 K) are given in **Appendix XXXII**.

Table 7.2: T and ΔT selected for entropy and enthalpy contributions

T	ΔT	T- ΔT	T+ ΔT
288 K	10 K	278 K	298 K
308 K	10 K	298 K	318 K
348 K	10 K	338 K	358 K

The entropy and enthalpy contributions to the free energy for each binding site residue are given in **Appendix XXXIII**. The free energy of C_{α} -O_w pairs calculated as a function of pair distance for monomer residues at different temperatures are presented in **Figure 7.11**. With the rise in temperature, the free energy of the binding

pocket increases and become more positive as compared to the lower temperature. This can be explained based on the number of hydrogen bonded water molecules present inside the cavity formed by the binding sites. With the rise in temperature, there is presence of more fractions of broken hydrogen bonds in water molecules (**Figure 7.9**) and the water-water hydrogen bond strength also decreases. Therefore, the protein-solvent interaction becomes weaker resulting in a positive free energy.

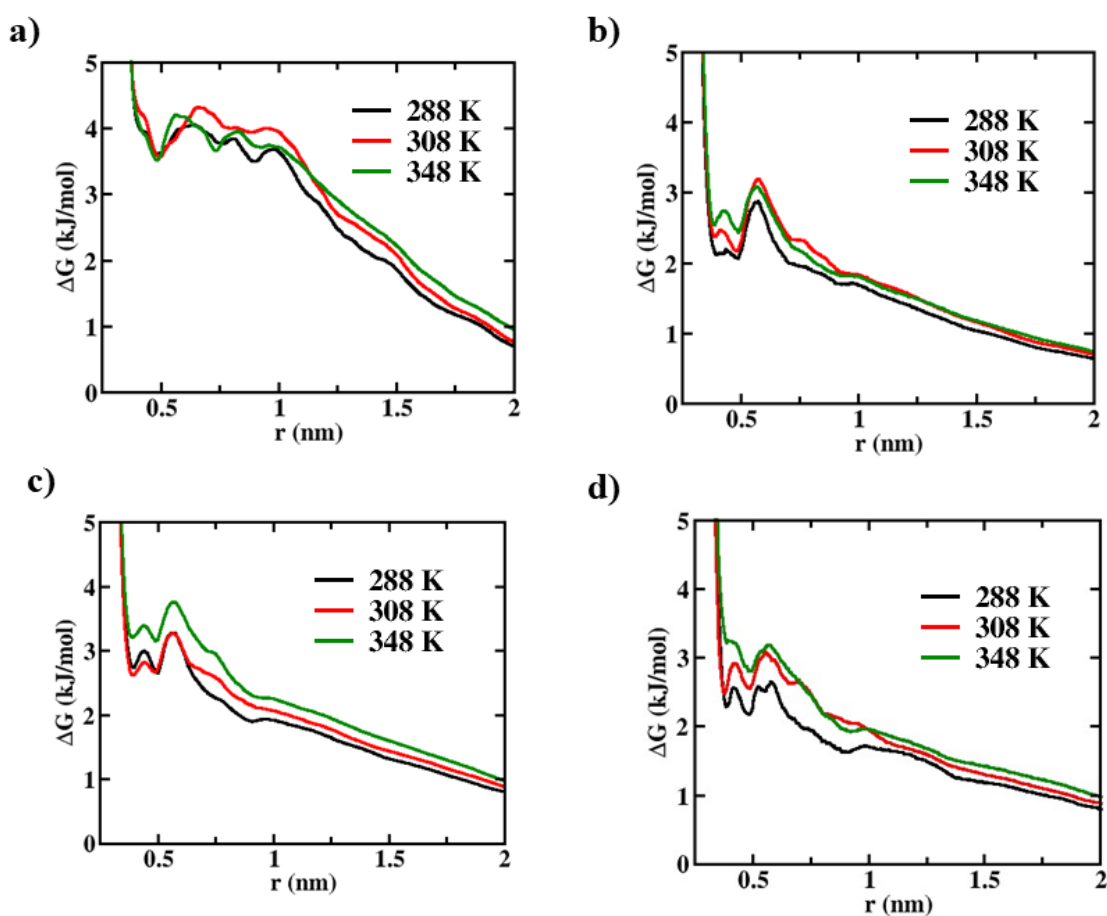


Figure 7.11: Free energy function of C_{α} -Ow of amino acid residues at various binding site of protease at different temperatures for monomer protease in SPC/E water model. (a) Catalytic dyad (b) Substrate binding site (c) Dimerization site (d) Allosteric site.

It is found that the free energy of the allosteric site and substrate binding site is lower at 288 K and higher at 348K. The $\Delta\Delta G$ value for the water interaction is 0.3 and 0.55 kJ/mol for the substrate-binding site and allosteric site, respectively, between 288

K and 348 K. The low $\Delta\Delta G$ value (0.3 kJ/mol) indicates that the water can be easily replaced at the substrate-binding site by small organic molecules. The higher value of $\Delta\Delta G$ (0.55 kJ/mol) for the allosteric site indicates the more stability of the amino acid residues due to the water molecules. It may be concluded that the energy expense of binding the small molecules at substrate binding is less compared to the allosteric site. Even though the dimerization site residues show an increase in free energy at higher temperatures, ΔG values at 288 K and 308 K show a similar trend. At 348 K higher value for free energy can be attributed due to the thermal fluctuations of protein and the free energy differences between the 288 K and 348 K are found to be $\Delta\Delta G = 0.51$ kJ/mol. In the catalytic dyad, the free energy is almost similar at all temperatures. It is observed that there is the presence of stable conformation of the protein at lower temperatures which are unstable at the higher temperature. It can be found that the interaction of free energy is governed by the arrangement of water molecules and the fraction of hydrogen-bonded water molecules around the C_α atom of key amino acid residues in the solvation shell.

7.12 HYDROGEN BOND DYNAMICS

The continuous hydrogen bond autocorrelation functions were calculated as described in Section 2.7.5. The hydrogen bond lifetime of water molecule around His41 and Glu166 residues is comparatively higher as compared to Cys117 and Arg298 residues for M^{Pro} monomer at 278 K and 310 K. The lifetime is even more at cold denaturation temperature, 278 K than 310 K. The His41 and Glu166 in catalytic dyad and substrate binding site are buried between the cleft of domain I and domain II of the protease. This forms a cavity that holds the water molecules firmly leading to strong hydrogen bond interaction between water molecules. At the high temperatures 348 K and 383 K the hydrogen bond lifetime decreases around the residues of the binding site. The higher hydrogen bond lifetime of water molecules indicates that the binding sites in the protease are stable. This is supported by the observation that the number of water molecules decreases with a rise in temperature (**Figure 7.3(e)**) due to the expansion of binding sites. The hydrogen bond lifetime in ps of water molecules around the amino acids has been shown in **Table 7.3**. In the case of the dimer, a similar trend for water-water hydrogen bond lifetime is observed with a lower value due to its protein

structural rigidity. The two monomers interact in such a way that the binding sites are buried inside the homodimer. The number of water molecules residing in the binding sites of dimer is lesser than the monomer. Therefore, the water-water hydrogen bond lifetime is lower in case of dimer as compared to the monomer, even though the trend is similar. Thus, it can be concluded that the hydrogen-bonded water molecules stabilize the binding sites of the protein.

Table 7.3: Hydrogen bond lifetime (ps) of water molecules around the key amino acid residues at the various sites of main protease in SPC/E water model

Temperature	Monomer				Dimer			
	His41	Cys117	Glu166	Arg298	His41	Cys117	Glu166	Arg298
278 K	4.67	2.21	3.07	2.33	0.92	0.78	0.81	0.98
310 K	1.29	1.24	1.83	1.22	0.60	0.56	0.59	0.54
348 K	0.85	0.72	0.74	0.76	0.40	0.39	0.42	0.39
383 K	0.57	0.51	0.52	0.60	0.32	0.30	0.31	0.30

7.13 STRUCTURE AND DYNAMICS OF M^{Pro} IN mTIP3P WATER MODEL

The temperature-dependent stability of monomer and dimer were studied based on structural parameters and solvent effect using mTIP3P water model. It is observed that the global conformation of the main protease monomer and dimer remains similar despite of the rise in temperature (**Figure 7.1**). The local structural evolutions are found due to certain changes in the secondary structure elements in the protease. Unlike SPC/E model, structural analyses like C_α-RMSD, rGyr and SASA suggest that the protease monomer undergoes thermal fluctuations starting from 310 K (**Figure 7.12**). The dimer-Chain A is found to be more rigid with fewer conformational evolutions.

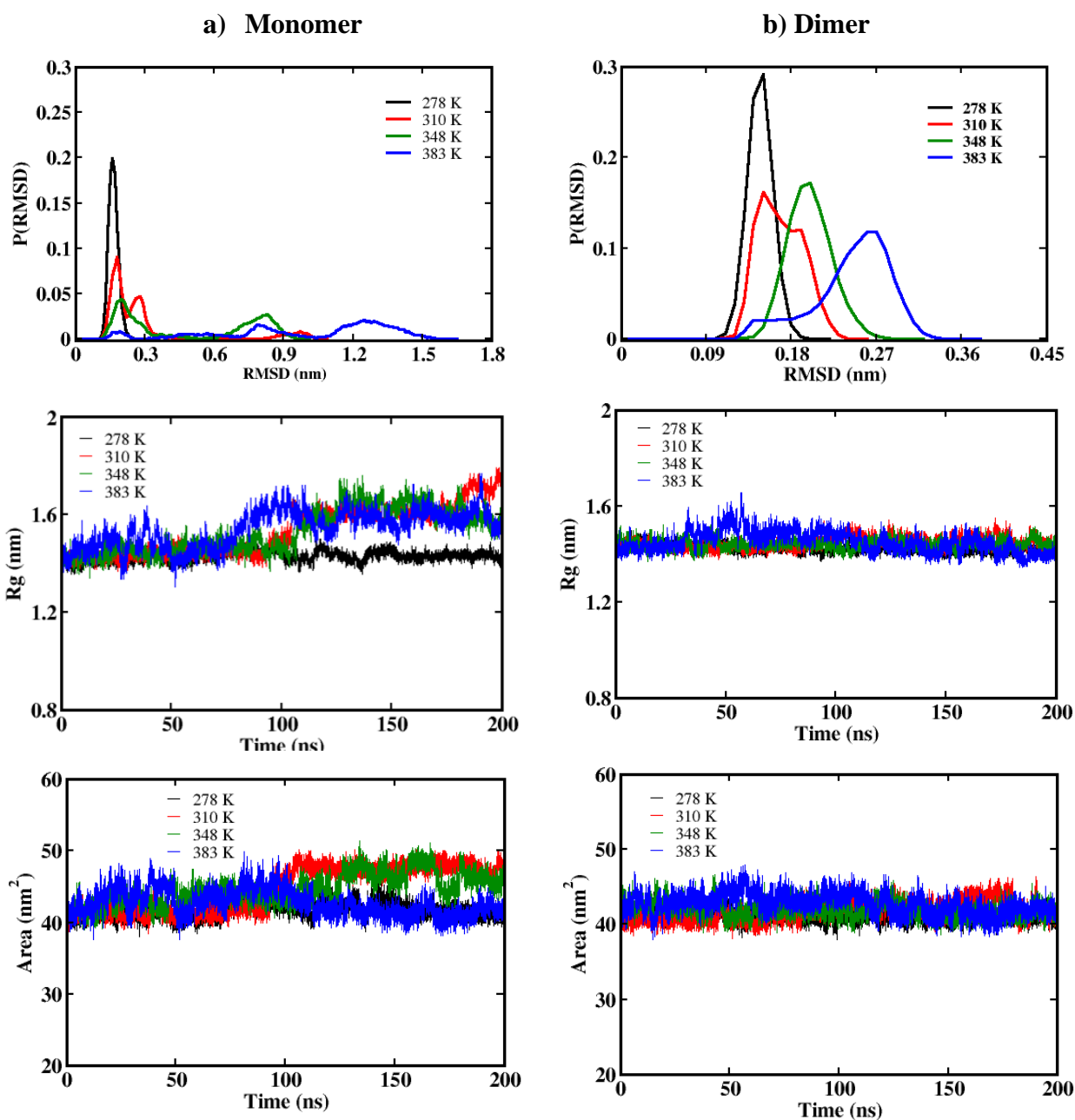


Figure 7.12: Probability distribution of C_α -RMSD; time evolution of radius of gyration and SASA in mTIP3P water model. (a) monomer (b) dimer-Chain A.

Like SPC/E water model, the solvent exposure is higher in the monomer than the dimer with the rise in temperature as expected. In the case of monomer and dimer, the water molecules present in the first solvation shell of Arg298 is less as compared to other binding site residues in the mTIP3P water model which shows a similar trend to SPC/E water model (**Figure 7.13**). Like SPC/E water model, it has been observed

that the probability of the lower coordinated water molecules is found at higher temperatures around all residues in both monomer and dimer (**Appendix XXXIV**). The probability distribution of order parameter around the binding site residues shows that the TIP3P water molecules are more heterogeneous in monomer compared to dimer chain-A which shows a similar trend to SPC/E water model (**Appendix XXXV**).

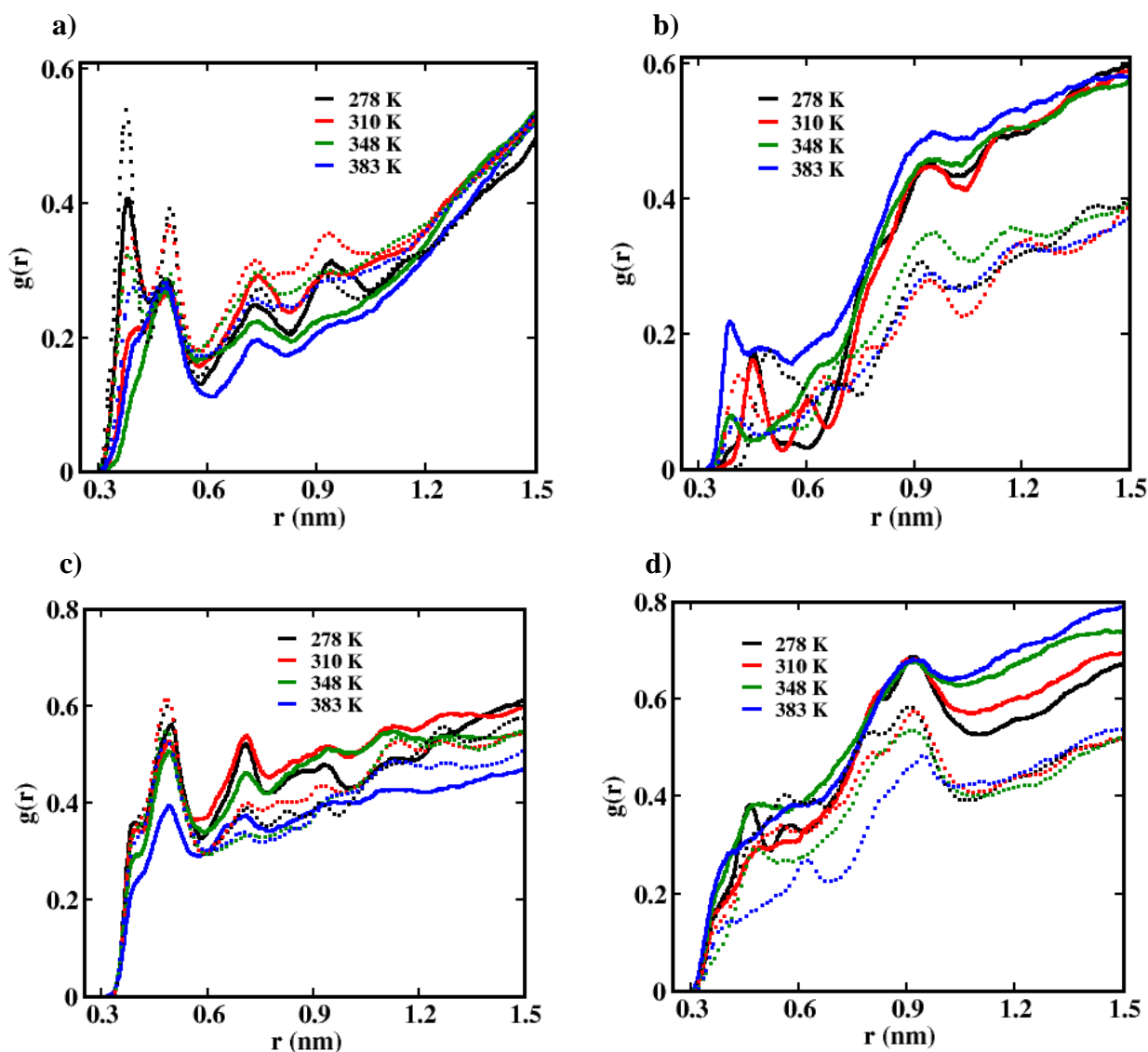


Figure 7.13: Radial distribution function of C_{α} -Ow of key amino acid residues at various binding site of protease at different temperatures in mTIP3P water model. Solid line and dash line represent monomer and dimer-Chain A respectively. (a) His41 (b) Cys117 (c) Glu166 (d) Arg298.

7.14 CONCLUSION

Temperature dependence on the conformational evolution of SARS CoV-2 M^{Pro} using classical molecular dynamics simulations reveals that the global structure of the enzyme is intact with minimum changes while significant changes are observed at the various binding sites of the protease. The various binding sites of the main protease can be divided into catalytic dyad, substrate-binding site, dimerization site and allosteric site. The basic structural changes evaluated by RMSD, rGyr and RMSF describe the thermal fluctuations in the protein with the rise in the temperature. Even though the dimerization of protease helps in the structural stability and function, the dimer turns to be unstable with rise in the temperature. We studied two different water models, namely the CHARMM-SPC/E and the CHARMM-TIP3P water model system. Network analysis shows the conformational evolution of the protease at higher temperatures. The structural analyses performed selectively for important amino acid residues at the various binding site showed the flexible nature of domain III at higher temperatures may affect the dimerization of the main protease which is crucial for its biological function. The dihedral analysis and PCA analysis around the binding sites concluded the flexible behaviour of both monomer and dimer-Chain A with an increase in temperature. The water structural properties were analysed in terms of radial distribution functions, tetrahedral order parameter (S_g) values and fraction of the number of hydrogen bonds around various binding sites of the main protease. The water structure in the solvation shell is found to be more heterogeneous at lower temperatures around the binding site amino acid residues. The higher hydrogen bond lifetime of the water molecules shows that the water molecules have a considerable effect on the stability of the binding sites. It is found that the $\Delta\Delta G$ value of interaction between water and the binding site amino acids is around 0.5 kJ/mol for both dimerization and allosteric binding site and 0.3 kJ/mol for substrate binding site. This explains the exposure of amino acid residues to the water molecules. The results of the two different water models showed a similar trend; however, the SPC/E water model showed a more pronounced effect compared to the TIP3P water model.

CHAPTER 8

SUMMARY AND CONCLUSIONS

This chapter mainly includes a summary of the whole research work and the major conclusions obtained from the research. This also includes a brief scope regarding the future work which can effectively guide drug discovery process.

8.1 SUMMARY

This thesis has addressed mainly about the interaction, the mechanism and the energy criteria of small organic molecules with the protein targets to obtain valuable insights which can be used in drug discovery process. A detailed summary regarding the whole research work is pointed below:

- ♣ Both ligand-based and structure-based approaches were employed to understand the factors affecting the inhibition of various protein systems such as G-protein coupled receptors, poly ADP-ribose polymerase enzyme and SARS CoV-2 main protease.
- ♣ Pharmacophore modelling and 3D-QSAR techniques were employed on small organic molecules to analyse the chemical features favourable for the molecule to act as an efficient drug molecule.
- ♣ Molecular docking and molecular dynamics simulations were employed to understand the interaction and binding affinity of small organic molecules at the binding site of respective targets. The energy requirements for stable protein-ligand complex formation were evaluated by free energy calculations.
- ♣ The electronic level properties of the small organic molecules, the interacting amino acid residues and the protein-ligand complexes were investigated using global reactivity descriptors. The drug-target non-covalent interactions were characterized by HOMO-LUMO visualization.
- ♣ The level of drug exposure to the tissues, the bioavailability and the biological activity of the small organic molecules were determined based on the pharmaceutically important parameters such as blood-brain barrier value, cell permeability, human serum albumin binding and Lipinski's rule.

- ♣ The key amino acid residues responsible for poly ADP-ribose polymerase-2 by benzimidazole derivatives were determined by single amino acid substitution at the catalytic domain of the enzyme by alanine residue.
- ♣ The temperature dependency or inactivation of SARS CoV-2 main protease (M^{Pro}) monomer and dimer forms is evaluated in presence of two different water models: SPC/E and modified TIP3P models.
- ♣ The structural evolution/degradation of protease with respect to rise in temperature is determined by network analysis and dihedral analysis.
- ♣ The water structure around the various binding sites of main protease is evaluated by radial distribution function (RDF), fraction of hydrogen-bonded water molecules, orientational tetrahedral order parameter and thermodynamic properties.
- ♣ The water-water hydrogen bond lifetime calculated for the various binding sites of main protease revealed the contribution of water molecules in stabilizing the binding sites of the protein.

8.2 CONCLUSIONS

The major conclusions drawn from the whole research work is as follows:

- ♣ A statistically robust and highly predictive 3D-QSAR model was developed using 58 β -alanine based GCGR antagonists with significant variation in structure and potency profile. The position of hydrophobic/non-polar substituents increases the inhibitory activity while the electron withdrawing groups found to have unfavourable contribution to the potency of the inhibitor.
- ♣ The hydrogen bonding network between the polar region of β -alanine derivatives with Arg346, Ser350, Leu399, Asn404 and Lys405 amino acid residues at TM7 region and the proper alignment of hydrophobic region towards TM5 facilitates the GCGR allosteric inhibition.
- ♣ The importance of hydrophobic and hydrogen bond interactions for effective allosteric inhibition of GCGR for diabetic therapy is further supported by HOMO-LUMO analysis, simulation analysis and MM/GBSA free energy analysis.

- ♣ The inhibition study of PARP-2 enzyme with a set of 32 benzimidazole derivatives revealed that the ligand forms hydrogen bonds with Gly429 and π - π stacking interaction with Tyr473 at the binding pocket. Also, the non-polar groups in the ligand such as methyl group stabilize the complex by hydrophobic interactions with Ile331, Tyr455 and Met456 residues.
- ♣ For effective PARP-2 enzyme inhibition in cancer treatment, there should be strong π -stacking interactions coupled with hydrogen bond interactions and water bridges between the benzimidazole derivatives and the amino acid residues at the binding site.
- ♣ PARP-2 enzyme showed higher binding affinity towards benzimidazole derivatives compared to PARP-1 enzyme in terms of per residue energy contribution and aromatic ring stacking interactions. This guides us to design and develop isoform selective PARP enzymes.
- ♣ Single amino acid mutations at the binding site of PARP-2 revealed that aromatic residues are crucial for stable inhibitor complex. The presence of aromatic rings in the ligand favours the effective inhibition of the enzyme by forming π - π stacking interactions with tyrosine residues at the binding pocket.
- ♣ For effective PARP-2 inhibition, π -stacking interactions with Tyr473 are more significant followed by hydrogen bond interactions with Gly429, water-bridges with Glu558 and hydrophobic interactions Ile331, Tyr455, Met456 residues.
- ♣ Four new inhibitors (P1, P2, P3 and P4) were proposed with higher activity out of which the three molecules appeared to be promising candidates for PARP-2 inhibition, which is supported by MD simulation, HOMO-LUMO energy gap, free energy analysis and ADME/Toxicity properties.
- ♣ Ligand-based and structure-based virtual screening methods revealed that β -adrenoceptor agonists and adenosine deaminase inhibitors which are used in asthma and cancer/inflammatory disorders respectively can be effectively used as repurposing drugs against SARS CoV-2 main protease.
- ♣ Three major hydrogen bonds with Gly143, Glu166 and Gln189 residues at the protease binding pocket is crucial for better enzyme inhibition.

- ♣ In all the above cases, *in silico* approaches identified that water-mediated hydrogen bonds/ water-bridges stabilizes the ligand at the binding pocket of protein targets.
- ♣ In all the cases, it is found that van der Waals energy, SASA energy and electrostatic energy contributes for thermodynamically stable binding of ligand with the protein target.
- ♣ Further, it is found that the temperature does not affect the global conformation of SARS CoV-2 main protease whereas a local conformational space evolved due to thermal fluctuations. The maximum conformational changes are observed at 348 K which is near the melting temperature.
- ♣ Dimerization and allosteric sites were found to be more affected and flexible with rise in temperature. The free energy, $\Delta\Delta G_{\text{water-interaction}}$ indicates that the energy expense of ligand binding at substrate binding site is less than allosteric site. The SPC/E water model showed more pronounced effect compared to mTIP3P water model.

The *in silico* methods reveal that the small organic molecules/ ligands interacts at the binding site of respective targets with different affinity. It is found that the binding affinity of the ligands towards the respective targets can be altered by incorporating the favourable functional groups such as hydrophobic groups, hydrogen bond acceptor-donor groups or polar groups. The proper interpretation of the non-covalent interactions responsible for the protein-ligand interactions help to alter the binding affinity and drug efficacy. Overall, the work suggests that optimized non-covalent interactions such as hydrogen bonding interactions, hydrophobic interactions and aromatic interactions contribute to the effective ligand binding. The future work can be extended to study the protein-ligand interactions by altering the external factors such as pH, temperature and co-solvents. This may helps us to understand the specific and non-specific forces which can stabilize the complex more effectively. The knowledge of molecular-level mechanism and energy criteria of enzyme inhibition helps in the designing and synthesizing new bioavailable drug candidates more effectively.

APPENDIX

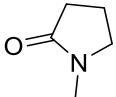
Appendix I: Detailed structure, experimental activity (Exp pK_i), predicted activity (Pred pK_i) and residual activity of glucagon receptor antagonists

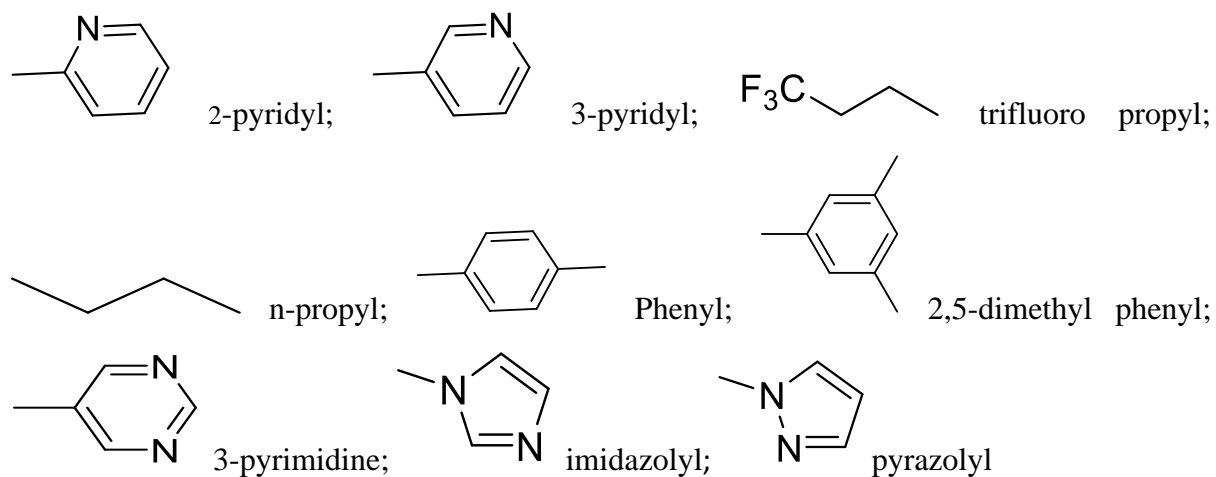
Cpd	X	H1	R2	R3	H2	Exp pK _i	Pred pK _i	Residual Activity
Class I: Pyrazole ether series of inhibitors								
1	O	n-propyl	pyrazolyl	Phenyl	2-Me	4.833	4.846	0.013
2	O	n-propyl	pyrazolyl	Phenyl	3-OMe	5.666	5.662	-0.004
3	O	n-propyl	pyrazolyl	Phenyl	4-OMe	5.703	5.689	-0.014
4 ^t	O	n-propyl	pyrazolyl	Phenyl	4-CF ₃	5.839	5.854	0.015
5	O	n-propyl	pyrazolyl	Phenyl	4-CF ₃	5.924	5.951	0.027
6 ^t	O	n-propyl	pyrazolyl	Phenyl	3-Me	5.955	5.993	0.038
7	O	n-propyl	pyrazolyl	Phenyl	3-Cl	5.963	5.944	-0.019
8	O	n-propyl	pyrazolyl	Phenyl	4-Me	6.014	6.011	-0.003
9 ^t	O	n-propyl	pyrazolyl	Phenyl	4-Cl	6.222	6.025	-0.197
10	O	n-propyl	pyrazolyl	Phenyl	3-Et	6.255	6.249	-0.006
11	O	n-propyl	pyrazolyl	2-pyridyl	4-Et	6.670	6.681	0.011
12	O	n-propyl	pyrazolyl	3-pyridyl	4-CF ₃	6.900	6.878	-0.022
Class II: R-substituted pyrazole ethers and aminopyrazoles series								
13	O	n-propyl (+)	pyrazolyl	Phenyl	4-CF ₃	6.367	6.375	0.008
14 ^t	O	n-propyl (-)	pyrazolyl	Phenyl	4-CF ₃	7.041	6.878	-0.163
15	NH	n-propyl (-)	pyrazolyl	Phenyl	4-CF ₃	7.174	7.148	-0.026
16	NH	trifluoro propyl (-)	pyrazolyl	Phenyl	4-CF ₃	7.319	7.293	-0.026
17	NH	n-propyl (+)	pyrazolyl	Phenyl	4-CF ₃	7.721	7.711	-0.010
18	NH	Cyclopentyl (-)	pyrazolyl	Phenyl	4-CF ₃	7.770	7.765	-0.005
19	NH	trifluoro propyl (+)	pyrazolyl	Phenyl	4-CF ₃	7.921	7.907	-0.014
20 ^t	NH	Cyclopentyl (+)	pyrazolyl	Phenyl	4-CF ₃	8.046	7.374	-0.672
Class III: Ether and amino derivatives with different hydrophobic core								
21	O	n-propyl	Pyrimidine	Pyrazolyl	4-CF ₃	5.959	5.948	-0.011
22	O	n-propyl	Phenyl	Pyrazolyl	4-CF ₃	6.051	5.999	-0.052

23	NH	n-propyl	Phenyl	Imidazolyl	4-CF ₃	6.169	6.193	0.024
24	NH	n-propyl	2-pyridyl	Pyrazolyl	4-CF ₃	6.174	6.150	-0.024
25	O	n-propyl	3-pyridyl	Pyrazolyl	4-CF ₃	6.638	6.645	0.007
26 ^t	O	n-propyl	Phenyl	Imidazolyl	4-CF ₃	6.658	6.689	0.031
27	NH	n-propyl	3-pyridyl	Pyrazolyl	4-CF ₃	6.670	6.772	0.102
28	O	n-propyl	3-MePh	Pyrazolyl	4-CF ₃	6.699	6.707	0.008
29	NH	n-propyl	Phenyl	Pyrazolyl	4-CF ₃	6.796	6.751	-0.045
30	NH	n-propyl	3-pyrimidine	Pyrazolyl	4-CF ₃	7.000	7.027	0.027
31	O	n-propyl	3,5-diMePh (-S)	Pyrazolyl	4-CF ₃	7.194	7.192	-0.002
32	O	n-propyl	3,5-diMePh (+R)	Pyrazolyl	4-CF ₃	7.854	7.835	-0.019
33	NH	n-propyl	3,5-diMePh	Pyrazolyl	4-CF ₃	7.921	7.963	0.042
Class IV: Ethers substituted with different alkyl side chains								
34	O	Ethyl	Phenyl	Pyrazolyl	4-CF ₃	5.854	5.889	0.035
35	O	cyclo- Propyl	Phenyl	Pyrazolyl	4-CF ₃	5.886	5.858	-0.028
36 ^t	O	iso- Propyl	Phenyl	Pyrazolyl	4-CF ₃	6.143	6.134	-0.009
37 ^t	O	cyclo- Butyl	Phenyl	Pyrazolyl	4-CF ₃	6.155	6.048	-0.107
38	O	iso- Butyl	Phenyl	Pyrazolyl	4-CF ₃	6.260	6.266	0.006
39	O	cyclo- Pentyl	Phenyl	Pyrazolyl	4-CF ₃	6.377	6.366	-0.011
40	O	tert- Butyl	Phenyl	Pyrazolyl	4-CF ₃	6.769	6.754	-0.015
Class V: Ethers with substitution on heterocyclic ring								
41	O	n-propyl	Phenyl	Pyrazolyl	3-C(O) NMe ₂	5.187	5.249	0.062
42	O	n-propyl	Phenyl	Pyrazolyl	3-OMe	5.357	5.366	0.009
43	O	n-propyl	Phenyl	Pyrazolyl	H	5.387	5.486	0.099
44 ^t	O	n-propyl	Phenyl	Pyrazolyl	3-CN	5.387	5.565	0.178
45	O	n-propyl	Phenyl	Pyrazolyl	3-Me	5.721	5.732	0.011
46 ^t	O	n-propyl	Phenyl	Pyrazolyl	3-F	5.769	5.957	0.188
47	O	n-propyl	Phenyl	Pyrazolyl	3-Cl	5.959	5.904	-0.055

48	O	n-propyl	Phenyl	Pyrazolyl	3-CF ₃ , 4-Me	5.959	5.957	-0.002
49	O	n-propyl	Phenyl	Pyrazolyl	-(CH ₂) ₄	6.161	6.146	-0.015
50 ^t	O	n-propyl	Phenyl	Pyrazolyl	-CH=CH CH=CH	6.538	6.194	-0.344

Class VI: N- and C-linked 5-membered cyclic compounds

51		Phenyl	Phenyl	4-CF ₃	6.121	6.104	-0.017
52	Cyclo-butyl	3-pyridyl	Pyrazolyl	4-CF ₃	6.208	6.254	0.046
53	N-methyl pyrrolidine	Phenyl	Phenyl	4-CF ₃	6.258	6.298	0.040
54 ^t	Cyclobutyl	2,5-diMePh _{trans}	Pyrazolyl	4-CF ₃	6.496	6.920	0.424
55 ^t	Cyclobutyl	Phenyl	Pyrazolyl	4-CF ₃	6.521	6.834	0.313
56 ^t	Cyclobutyl	2,5-diMePh _{cis}	Pyrazolyl	4-CF ₃	7.066	6.920	-0.146
57	Cyclobutyl	Phenyl	Phenyl	4-CF ₃	7.658	7.678	0.020
58	Cyclobutyl	3-pyrimidine	Phenyl	4-CF ₃	7.854	7.885	0.031



^t defines the test set compounds considered for 3D- QSAR analysis

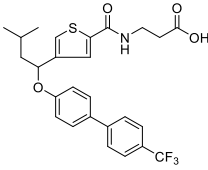
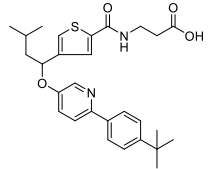
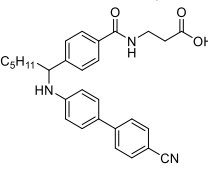
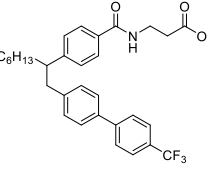
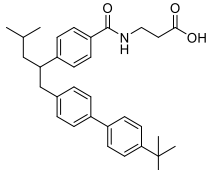
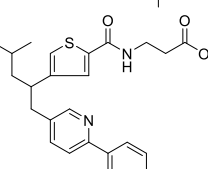
+/- indicates the enantiomers of the compounds

+R/ -S indicates the absolute configuration of chiral centres of the compounds

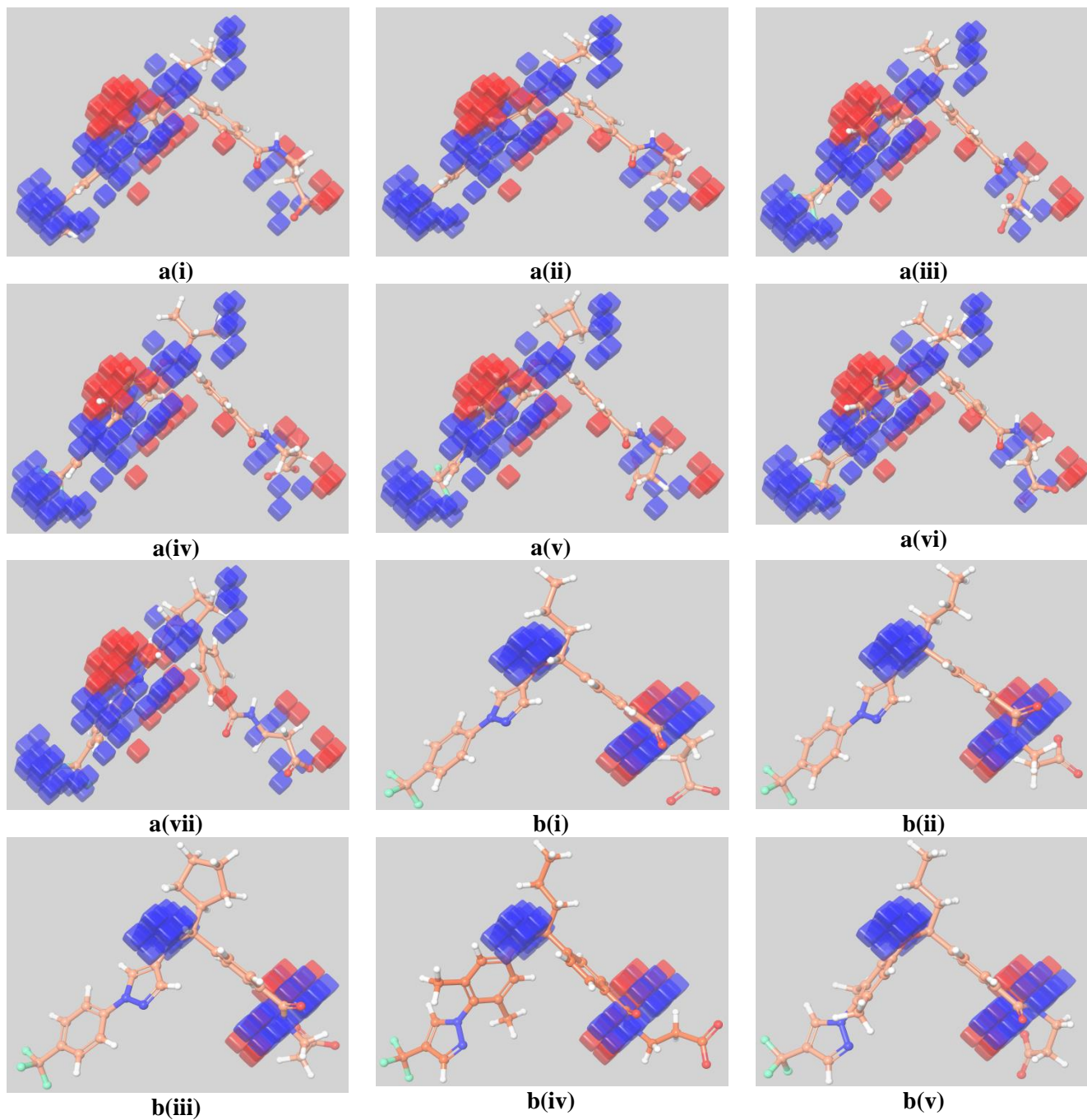
Residual activity is defined as the difference between predicted activity and experimental activity

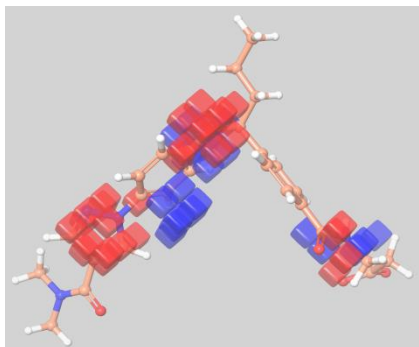
Appendix II: Predicted biological activity of external test set of 15 compounds

Compound	Structure	Experimental pK _i	Predicted pK _i	Residual Activity
1		7.097	7.000	0.097
2		7.585	7.041	0.544
3		7.292	7.146	0.146
4		7.959	7.321	0.638
5		7.155	7.099	0.056
6		7.137	7.195	-0.058
7		7.585	7.147	0.438
8		6.595	6.409	0.186
9		6.577	6.386	0.191

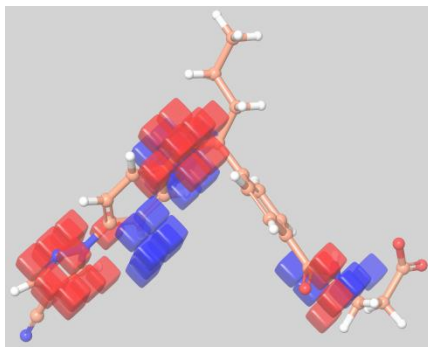
10		6.409	6.390	0.019
11		6.529	6.389	0.140
12		6.762	6.762	0.000
13		6.635	6.881	-0.246
14		6.575	6.613	-0.038
15		6.096	6.344	-0.248

Appendix III: QSAR contour plots visualized in terms of (a) (i)-(vii) favourable and unfavourable hydrophobic interactions on Compound 5, 12, 35, 36, 37, 40, 58 (b) (i)-(v) favourable and unfavourable hydrogen bond donor groups on Compound 13, 17, 20, 25, 50 and (c) (i)-(v) favourable and unfavourable electron withdrawing groups on Compound 41, 44, 46, 47, 50.

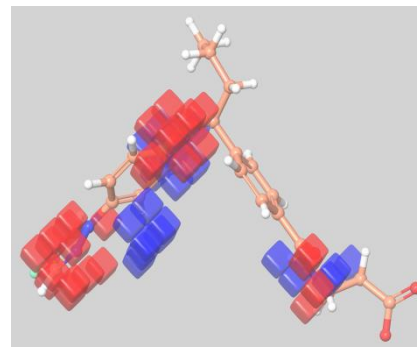




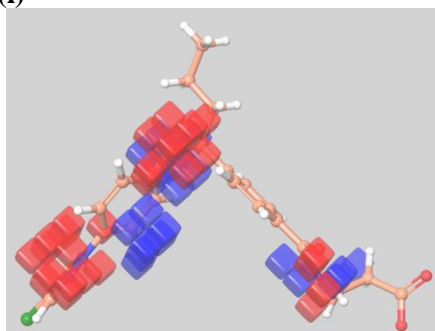
c(i)



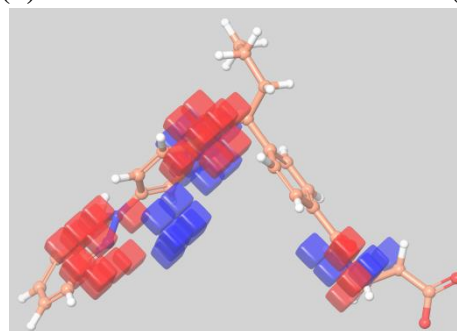
c(ii)



c(iii)



c(iv)



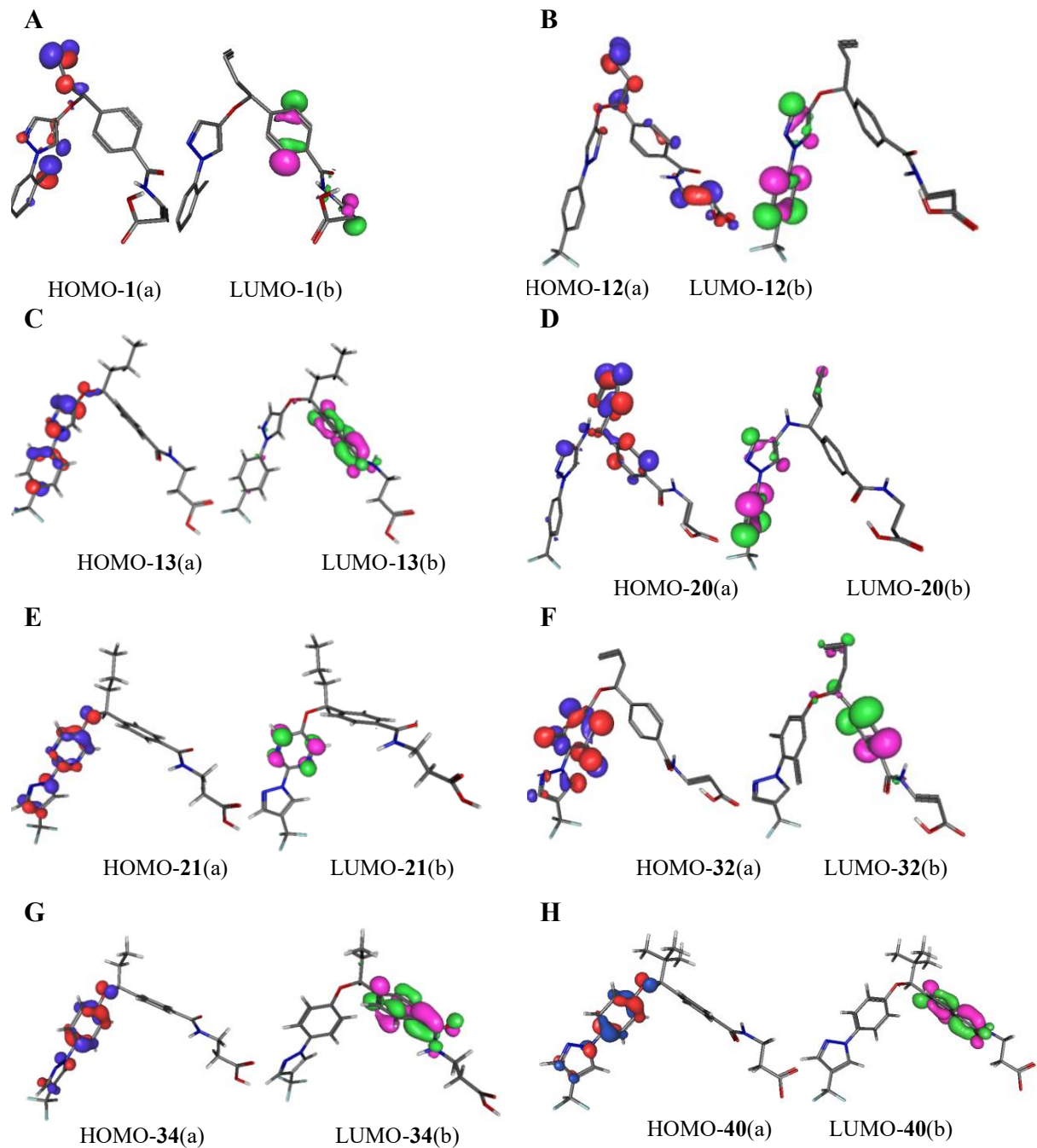
c(v)

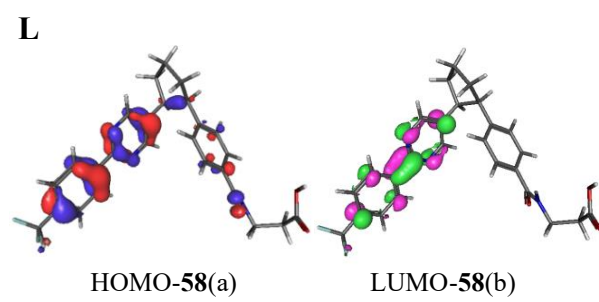
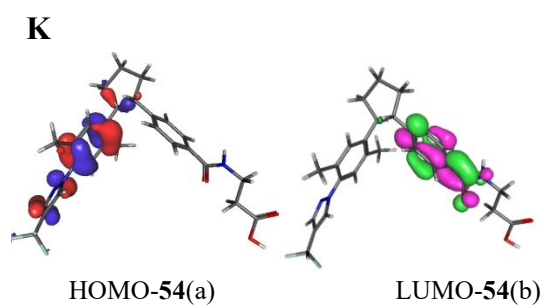
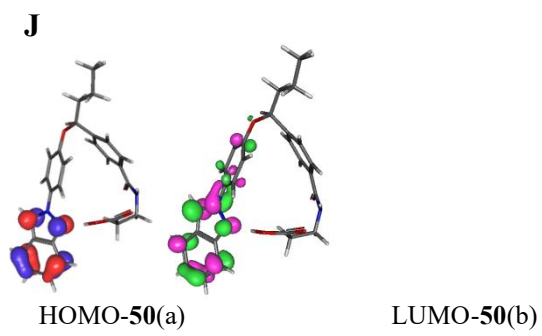
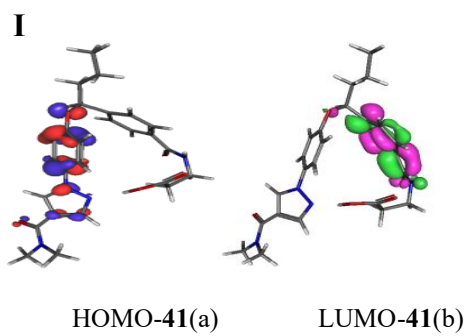
Appendix IV: Energy terms contributing to docking energy of protein- ligand complex (kcal/mol)

Cpd	Intermolecular Energy	Internal Energy	Torsional Free Energy	Unbound System's Energy	Binding Free energy
1	-8.61	-1.54	3.28	-1.54	-5.33
2	-9.99	-1.35	3.58	-1.35	-6.41
3	-9.90	-0.95	3.58	-0.95	-6.32
4	-9.78	-1.66	3.58	-1.66	-6.20
5	-10.40	-1.10	3.58	-1.10	-6.82
6	-9.96	-2.08	3.28	-2.08	-6.68
7	-10.17	-1.91	3.28	-1.91	-6.89
8	-10.19	-1.77	3.28	-1.77	-6.91
9	-10.81	-1.11	3.28	-1.11	-7.53
10	-10.77	-1.58	3.58	-1.58	-7.19
11	-11.14	-1.21	3.58	-1.21	-7.56
12	-11.37	-1.75	3.58	-1.75	-7.79
13	-10.92	-1.54	3.58	-1.54	-7.34
14	-11.18	-1.11	3.58	-1.11	-7.60
15	-11.30	-1.91	3.58	-1.91	-7.72
16	-11.17	-1.66	3.88	-1.66	-7.29
17	-11.33	-2.0	3.58	-2.0	-7.75
18	-11.17	-1.90	3.28	-1.90	-7.89
19	-11.59	-1.75	3.88	-1.75	-7.71
20	-11.54	-1.92	3.28	-1.92	-8.25
21	-10.38	-1.83	3.58	-1.83	-6.80
22	-10.21	-1.53	3.28	-1.53	-6.93
23	-10.75	-1.07	3.58	-1.07	-7.17
24	-10.77	-0.86	3.58	-0.86	-7.19
25	-11.07	-1.82	3.28	-1.82	-7.78
26	-10.86	-1.40	3.28	-1.40	-7.58
27	-9.85	-0.77	3.58	-0.77	-6.28

28	-11.29	-0.79	3.58	-0.79	-7.71
29	-11.11	-1.20	3.58	-1.20	-7.53
30	-10.87	-1.90	3.58	-1.90	-7.29
31	-11.01	-1.53	3.58	-1.53	-7.43
32	-11.28	-2.13	3.58	-2.13	-7.70
33	-10.78	-1.58	3.58	-1.58	-7.20
34	-9.79	-1.13	3.28	-1.13	-6.51
35	-9.69	-1.21	2.98	-1.21	-6.70
36	-10.50	-1.11	3.28	-1.11	-7.22
37	-10.60	-0.82	3.28	-0.82	-7.32
38	-10.57	-1.24	3.28	-1.24	-7.28
39	-10.42	-1.55	2.98	-1.55	-7.43
40	-9.46	-1.41	3.28	-1.41	-6.18
41	-9.97	-1.71	3.58	-1.71	-6.39
42	-9.69	-1.93	3.28	-1.93	-6.41
43	-9.61	-2.17	2.98	-2.17	-6.62
44	-9.57	-1.65	2.98	-1.65	-6.59
45	-10.00	-1.93	2.98	-1.93	-7.01
46	-10.37	-1.54	2.98	-1.54	-7.39
47	-10.25	-1.39	2.98	-1.39	-7.26
48	-10.49	-1.52	3.28	-1.52	-7.21
49	-10.96	-1.30	3.28	-1.30	-7.68
50	-10.89	-1.78	3.28	-1.78	-7.61
51	-9.81	-0.52	2.68	-0.52	-7.13
52	-10.28	-1.68	2.68	-1.68	-7.60
53	-9.37	-0.00	2.68	-0.00	-6.68
54	-8.59	-1.63	2.68	-1.63	-5.91
55	-10.19	-1.68	2.68	-1.68	-7.51
56	-10.44	-1.67	2.68	-1.67	-7.76
57	-10.42	-0.92	2.68	-0.92	-7.73
58	-10.35	-1.05	2.68	-1.05	-7.67

Appendix V: The HOMO-LUMO orbital positions mapped on selected ligands at B3LYP/ 6-31G(d,p) level of calculation





Appendix VI: Contribution of energy components to MM/ GBSA binding free energy for ligands at the binding site of 5XEZ (kcal/mol)

Compound	ΔG Bind	ΔG Bind Cov	ΔG Bind HBond	ΔG Bind Sol/SA	ΔG Bind Sol/GB	ΔG Bind Vdw
1	-41.097	0.846	-0.958	-11.771	7.410	-36.623
2	-37.197	-0.156	-0.333	-13.373	8.795	-32.129
3	-43.780	4.134	-0.878	-13.013	7.248	-41.277
4	-35.480	4.623	-0.545	-12.897	7.156	-33.812
5	-45.107	2.126	-1.438	-15.158	8.510	-39.147
6	-41.743	4.137	-0.892	-10.490	7.950	-42.444
7	-45.608	2.814	-0.981	-13.233	7.199	-41.408
8	-39.881	3.341	-0.426	-14.590	7.250	-35.450
9	-50.771	0.578	-0.299	-16.205	7.368	-42.214
10	-41.533	1.389	-0.794	-13.480	7.090	-35.748
11	-44.611	1.182	-0.298	-15.778	7.642	-37.360
12	-46.730	1.796	-1.326	-15.892	8.157	-39.467
13	-38.177	8.364	-1.620	-15.738	8.715	-37.898
14	-42.071	12.740	-1.627	-15.844	9.084	-46.424
15	-48.514	3.532	-0.887	-15.913	7.918	-43.164
16	-52.935	11.007	-1.979	-11.444	-0.997	-49.522
17	-39.598	9.282	-1.626	-15.609	8.494	-40.140
18	-56.129	-2.648	-1.237	-15.913	9.469	-45.800
19	-44.680	8.724	-1.364	-16.386	10.116	-45.772
20	-63.475	-2.589	-0.677	-16.853	7.162	-50.516
21	-48.766	9.464	-0.779	-15.791	7.809	-49.470
22	-43.378	5.197	-0.322	-15.305	8.237	-41.186
23	-53.360	-1.786	-1.542	-15.617	9.365	-43.786
24	-37.081	12.562	-1.325	-16.009	9.368	-41.677
25	-50.776	1.363	-1.590	-14.596	9.076	-45.028
26	-46.196	6.016	-1.583	-14.732	9.017	-44.913
27	-34.675	13.347	-0.483	-14.397	7.948	-41.092

28	-47.731	-1.766	-1.733	-12.626	10.089	-41.694
29	-42.022	6.305	-1.163	-16.204	9.160	-40.121
30	-44.087	6.883	-1.511	-16.645	9.639	-42.454
31	-41.551	3.979	-1.472	-15.675	10.384	-38.767
32	-51.755	-1.806	-1.572	-16.124	10.086	-42.339
33	-41.636	4.362	-1.428	-15.519	9.720	-38.771
34	-41.651	-0.816	-0.924	-11.804	8.437	-36.543
35	-46.149	4.334	-0.024	-15.152	7.594	-42.902
36	-46.661	1.629	-0.383	-15.231	7.419	-40.096
37	-49.351	3.908	-1.664	-17.703	10.274	-44.166
38	-43.783	4.135	-0.878	-13.013	7.248	-41.274
39	-50.097	6.395	-0.171	-18.232	9.490	-48.178
40	-54.334	6.284	-0.759	-16.500	7.224	-49.986
41	-49.127	5.198	-1.733	-18.637	10.419	-44.374
42	-45.214	6.824	-0.735	-15.462	7.673	-43.513
43	-44.840	7.106	-0.099	-16.142	9.195	-44.900
44	-38.036	3.461	-1.192	-14.379	7.499	-33.425
45	-42.553	4.148	-0.488	-14.833	8.006	-39.387
46	-44.538	5.218	-0.310	-15.166	8.046	-42.326
47	-43.810	7.493	-0.081	-16.858	9.446	-43.809
48	-48.296	5.241	-0.832	-15.961	7.881	-44.625
49	-44.632	7.344	-1.249	-17.022	8.501	-42.206
50	-64.178	-1.295	-0.674	-16.783	7.528	-52.954
51	-38.942	5.748	-0.453	-15.302	7.884	-36.817
52	-45.669	4.166	-1.497	-14.456	9.424	-43.583
53	-41.620	9.008	-0.002	-20.487	3.708	-33.847
54	-39.325	7.372	-1.807	-15.828	11.939	-41.001
55	-57.093	-4.270	-1.399	-15.099	8.529	-44.853
56	-45.857	8.311	-0.787	-15.182	7.730	-45.929
57	-43.519	0.143	-0.846	-13.946	7.318	-36.188
58	-45.765	3.811	-1.420	-14.659	9.369	-42.866

Appendix VII: Characteristics of protein-ligand complex obtained from 5 independent molecular dynamics simulations

Properties/ Simulation		S1	S2	S3	S4	S5	SD
RMSD (Å)	C-alpha	5.748	7.164	5.748	6.467	6.887	0.65
	Backbone	5.745	7.146	5.745	6.455	6.875	0.64
	Side chain	6.657	8.008	6.657	7.251	7.480	0.58
	Heavy atoms	6.179	7.600	6.179	6.831	7.135	0.62
	Lig wrt prot	5.164	5.777	5.164	3.816	3.791	0.89
	Lig wrt lig	1.254	2.585	1.254	1.328	1.436	0.57
RMSF range (Å)	c-alpha	2.549- 4.602	2.405- 5.787	2.549- 4.602	2.317- 4.316	2.112- 4.586	-
	Backbone	2.548- 4.564	2.441- 5.886	2.548- 4.564	2.336- 4.532	2.123- 4.554	-
	Side chain	2.668- 6.041	2.563- 7.370	2.668- 6.041	2.372- 5.044	2.188- 6.748	-
	Heavy atoms	2.600- 5.415	2.501- 6.172	2.600- 5.415	2.375- 4.786	2.192- 5.864	-
Ligand	Rg (Å)	5.264	6.323	5.264	5.489	5.722	0.44
	MolSA (Å ²)	439.852- 462.211	446.761- 465.869	439.852- 462.211	438.782- 463.614	442.776- 463.970	-
	SASA (Å ²)	277.069- 449.461	241.283- 426.983	277.069- 449.461	252.101- 457.269	262.310- 478.872	-
	PSA (Å ²)	140.240- 180.367	136.685- 184.685	140.240- 180.367	137.452- 184.016	145.036- 181.619	-

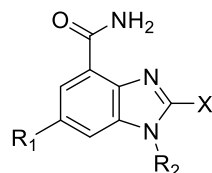
Appendix VIII: Predicted ADME/Toxicity properties of top scored glucagon receptor antagonists

Cpd	pK _i	MW	Dipole Moment	SASA	Donor HB	Acceptor HB	QPlogP _{o/w}	QPlogS
20	8.046	500.520	7.1049	738.556	2	5.50	3.753	-6.130
19	7.921	528.453	4.8958	737.326	2	5.50	3.984	-6.466
33	7.921	502.535	10.8124	758.536	2	5.25	6.532	-6.998
32	7.854	503.520	5.6430	860.222	1	5.50	7.073	-9.043
58	7.854	483.489	8.6880	790.552	1	5.50	6.200	-8.309
18	7.770	500.520	7.1340	842.873	2	5.50	4.064	-8.081
17	7.721	474.482	4.6355	813.770	2	5.50	3.699	-7.402
57	7.658	481.514	4.3137	795.874	1	3.50	7.403	-9.199
16	7.319	528.453	2.0435	843.910	2	5.50	4.351	-8.458
31	7.194	503.520	5.6430	764.573	1	5.25	6.895	-7.308
15	7.174	474.482	2.6968	719.087	2	5.50	3.387	-5.627
54	7.066	499.532	6.3752	738.836	1	4.50	6.738	-7.948
14	7.041	475.467	4.8841	709.198	1	5.25	6.127	-6.174
30	7.000	476.457	7.7491	730.145	2	7.50	4.867	-5.842

Rotor	PSA	QPlogK _{hs} ^a	QPlogBE	QPPC _{aco}	QPPMDCK	% Human Oral Absorption	Rule of 5	Rule of 3
8	109.414	0.896	-0.459	44.202	77.051	56.802	1	1
9	109.932	0.762	-0.297	40.552	316.267	58.276	1	1
9	100.932	0.940	-0.670	279.128	733.275	71.192	2	1
9	100.321	1.253	-1.344	113.887	278.392	72.906	2	1
5	102.461	1.037	-1.101	96.84	220.959	80.481	1	1
8	109.305	1.056	-1.128	18.268	40.118	58.744	1	2
9	108.773	0.856	-1.108	20.002	44.528	66.171	0	2
5	81.913	1.533	-0.849	155.56	366.819	91.937	1	1
9	114.402	0.971	-1.055	14.361	139.742	55.528	1	2
9	96.638	1.042	-0.570	358.57	966.906	77.716	2	1
9	109.956	0.689	-0.519	44.819	78.341	66.342	0	0
4	91.013	1.310	-0.532	186.076	473.353	88.705	1	1
9	108.081	0.785	-0.775	236.887	414.478	84.233	1	1
9	124.622	0.293	-1.077	119.864	295.746	79.219	0	1

* Permissible ranges of physico-chemical properties are given in Section 2.6

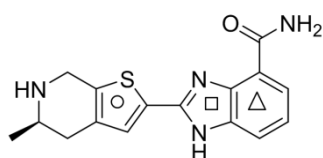
Appendix IX: Detailed structure, experimental activity and docking energy of benzimidazole carboxamide inhibitors against PARP-1/2 enzyme



Sl No:	R ₁	R ₂	X	PARP-2		PARP-1	
				IC ₅₀ (μM)	Binding energy (kcal/mol)	IC ₅₀ (μM)	Binding energy (kcal/mol)
C-1	F	H		14	-7.940	16	-8.870
C-2	F	H		30	-7.930	8	-8.850
C-3	F	Me		42	-7.460	18	-9.010
C-4	F	-		49	-7.270	25	-8.640
C-5	H	H		53	-7.310	14	-9.910
C-6	F	H		64	-7.200	42	-8.390
C-7	F	H		70	-7.830	15	-9.690
C-8	F	H		73	-7.680	27	-8.110
C-9	F	H		83	-7.630	58	-8.250
C-10	F	H		86	-7.190	25	-9.340
C-11	F	H		87	-7.840	40	-8.650
C-12	F	-		99	-7.910	25	-9.320
C-13	F	H		109	-7.660	22	-8.370
C-14	H	H		113	-7.82	58	-8.540

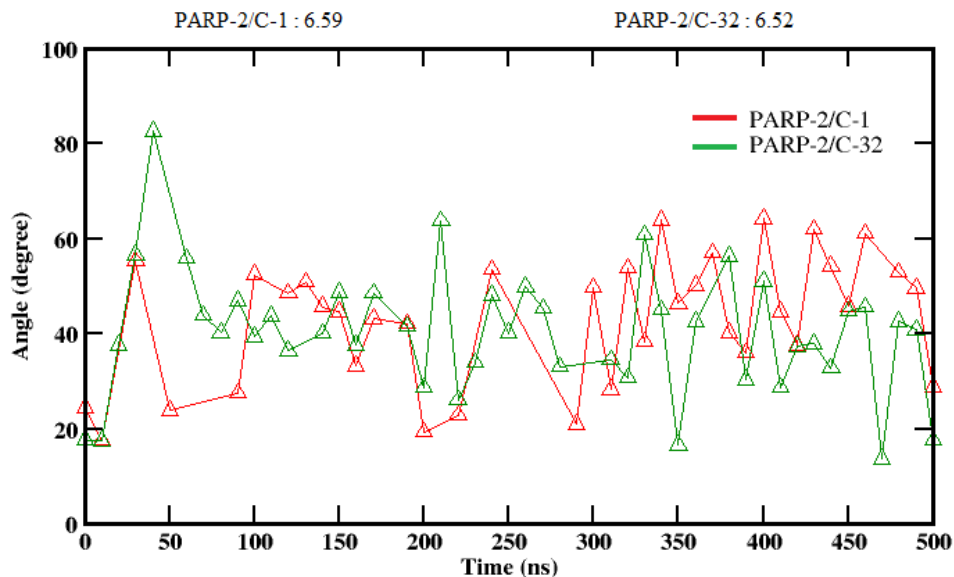
C-15	F	Me		117	-7.070	20	-9.810
C-16	F	H		124	-7.400	17	-8.490
C-17	H	H		131	-7.530	123	-8.170
C-18	H	H		140	-7.550	90	-8.360
C-19	F	Me		144	-7.500	244	-8.380
C-20	F	H		145	-7.790	25	-9.120
C-21	F	H		149	-7.360	37	-8.210
C-22	F	H		196	-7.430	27	-8.630
C-23	F	H		219	-7.600	36	-9.130
C-24	F	H		239	-7.680	14	-9.260
C-25	H	H		247	-7.500	55	-8.820
C-26	Me	H		281	-7.570	345	-8.200
C-27	F	H		324	-7.820	86	-9.000
C-28	F	Me		328	-7.550	153	-8.800
C-29	F	H		348	-7.750	130	-8.840
C-30	F	H		356	-7.440	92	-8.770
C-31	F	Et		359	-7.120	250	-8.940
C-32	Cl	H		382	-6.090	51	-8.890

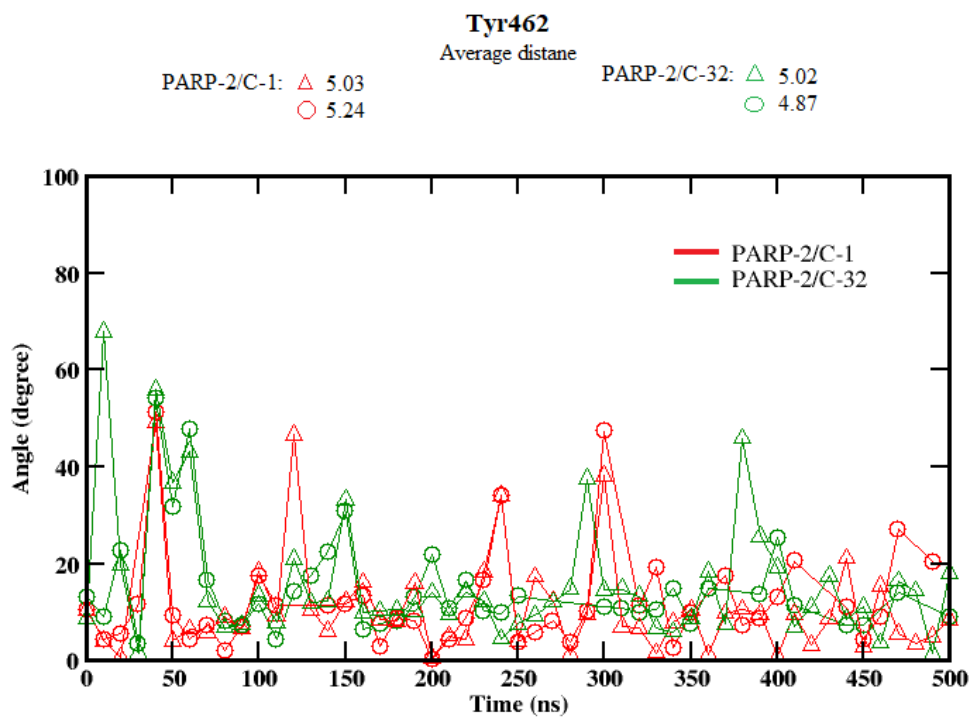
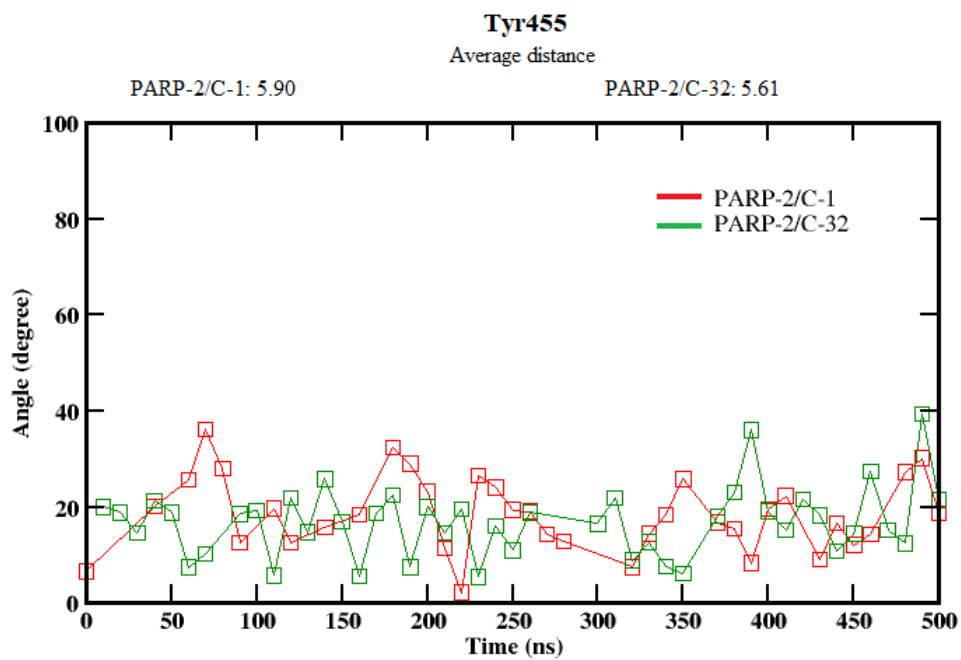
Appendix X: The angle between the normal vectors of two rings in stacking interaction formed between ligand and aromatic amino acid residues of PARP-2 enzyme at the catalytic pocket over 500ns simulation. The geometrical shapes (circle, square and triangle) represent the interaction of amino acid residue with different rings present in the ligand. His428 and Phe463 were found to have only one interaction with the benzene ring (triangle) of benzimidazole core whereas Tyr455 found to interact with the imidazole (square) ring. Tyr462 showed interaction with benzene ring and thiophene ring (circle) whereas Tyr473 showed π -stacking interaction with all the three aromatic rings present in the ligand. Average distance is given in Å



His428

Average distance



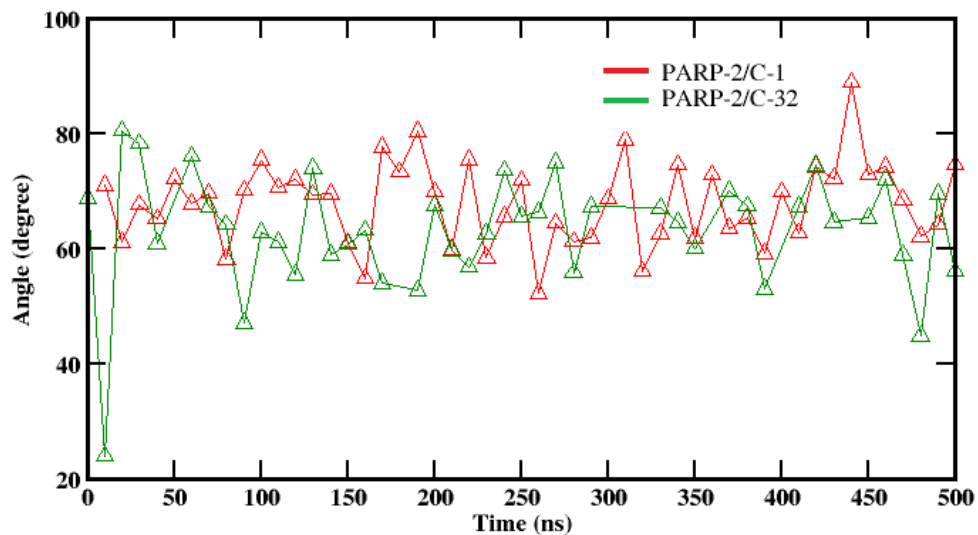


Phe463

Average distance

PARP-2/C-1 : 7.34

PARP-2/C-32 : 7.54

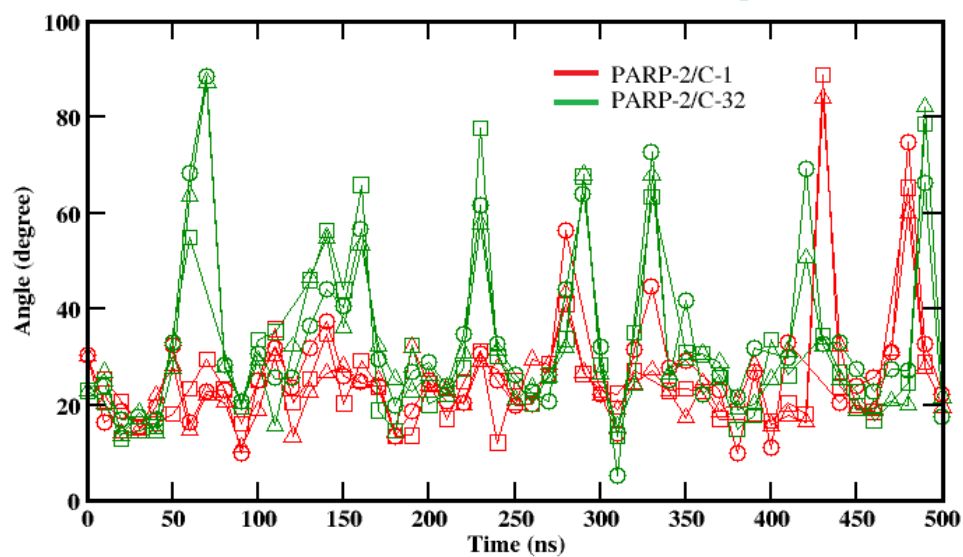


Tyr473

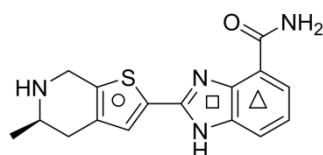
Average distance

PARP-2/C-1: \triangle 4.43
 \square 4.15
 \circ 4.64

PARP-2/C-32: \triangle 4.30
 \square 5.00
 \circ 5.00



Appendix XI: The angle between the normal vectors of two rings in stacking interaction formed between ligand and aromatic amino acid residues of PARP-1 enzyme at the catalytic pocket over 500ns simulation. The geometrical shapes (circle, square and triangle) represent the interaction of amino acid residue with different rings present in the ligand. His862, Tyr896 and Phe897 were found to have only one interaction with the benzene ring (triangle) of benzimidazole core whereas Tyr889 found to interact with the thiophene ring (circle). Tyr907 showed π -stacking interaction with benzene ring and imidazole ring (square) ring present in the ligand. Average distance is given in Å

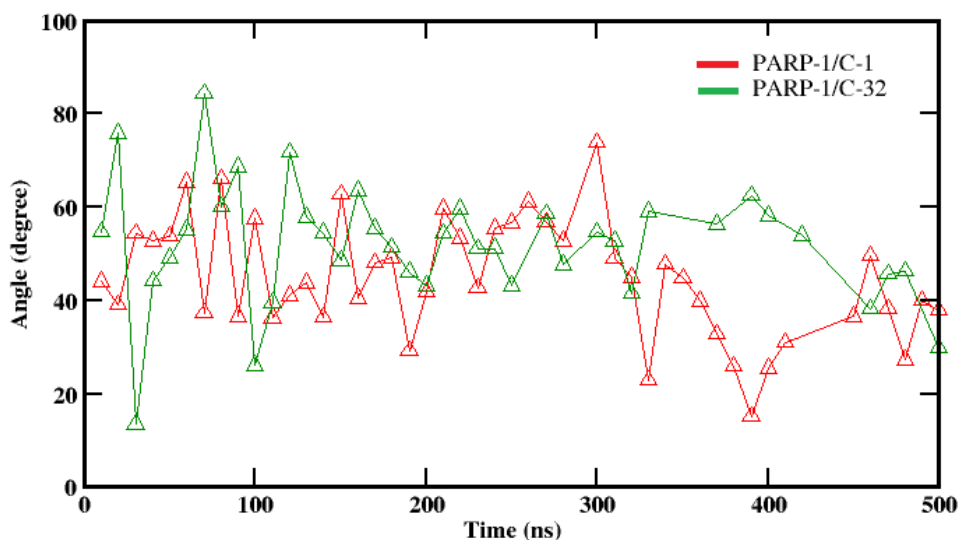


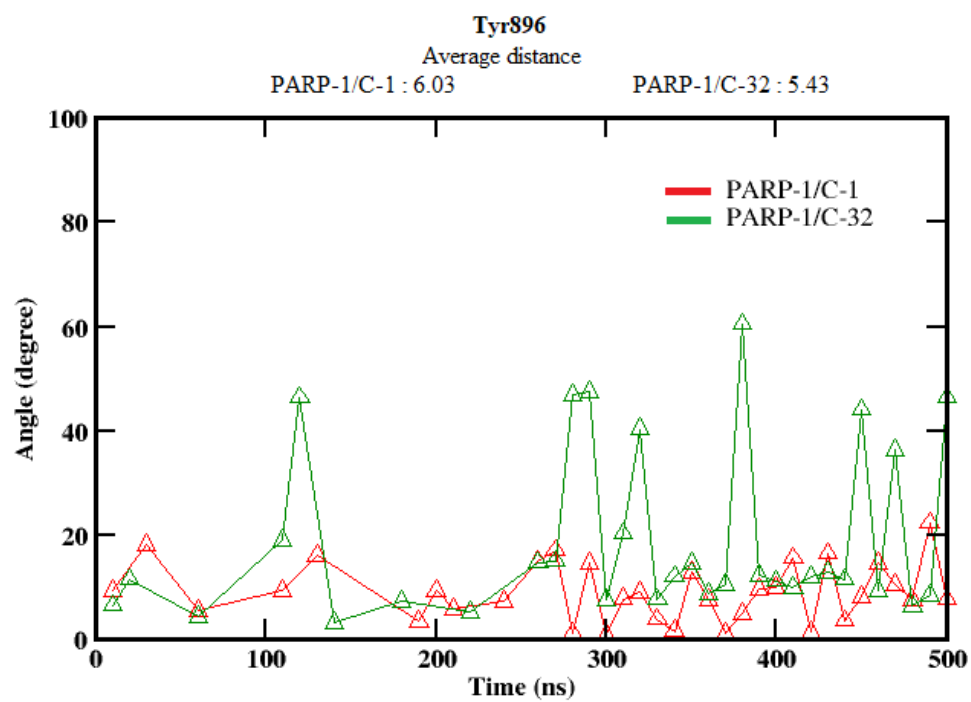
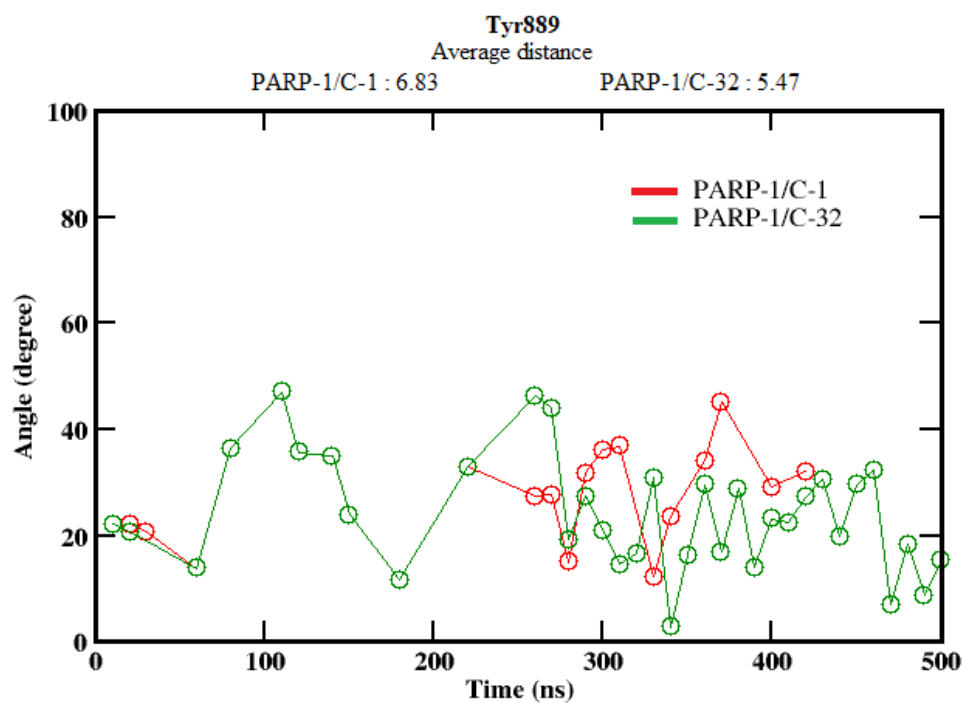
His862

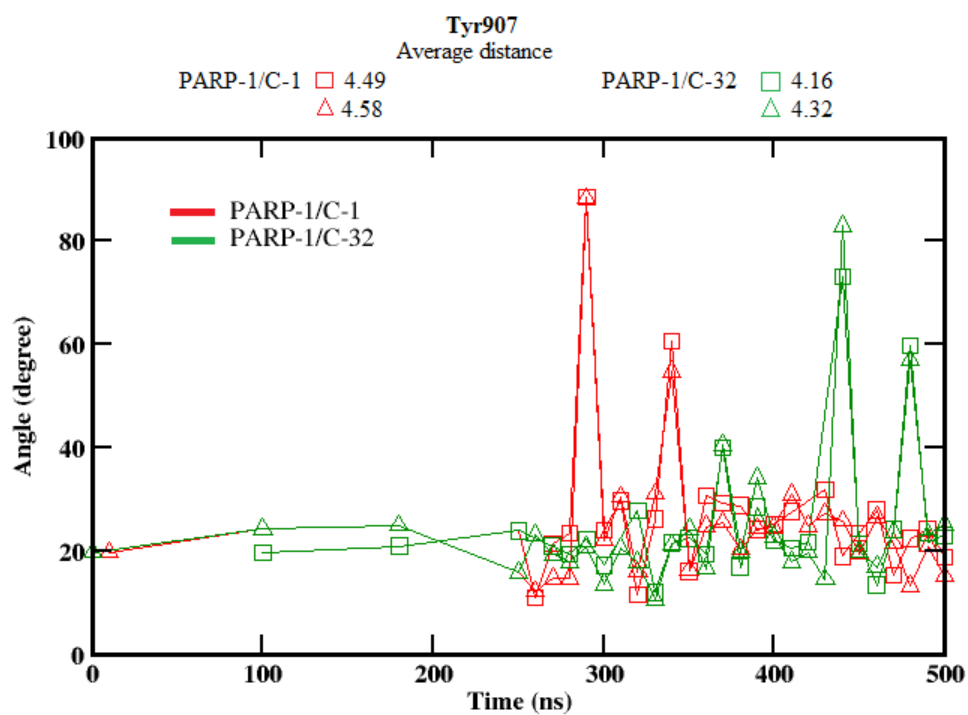
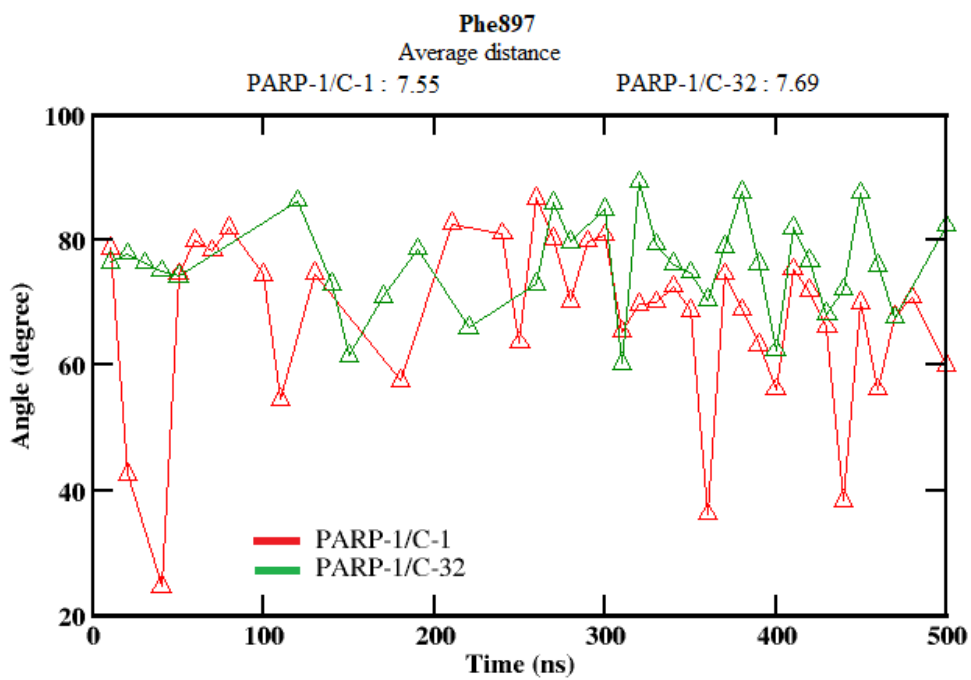
Average distance

PARP-1/C-1 : 4.72

PARP-1/C-32 : 4.78

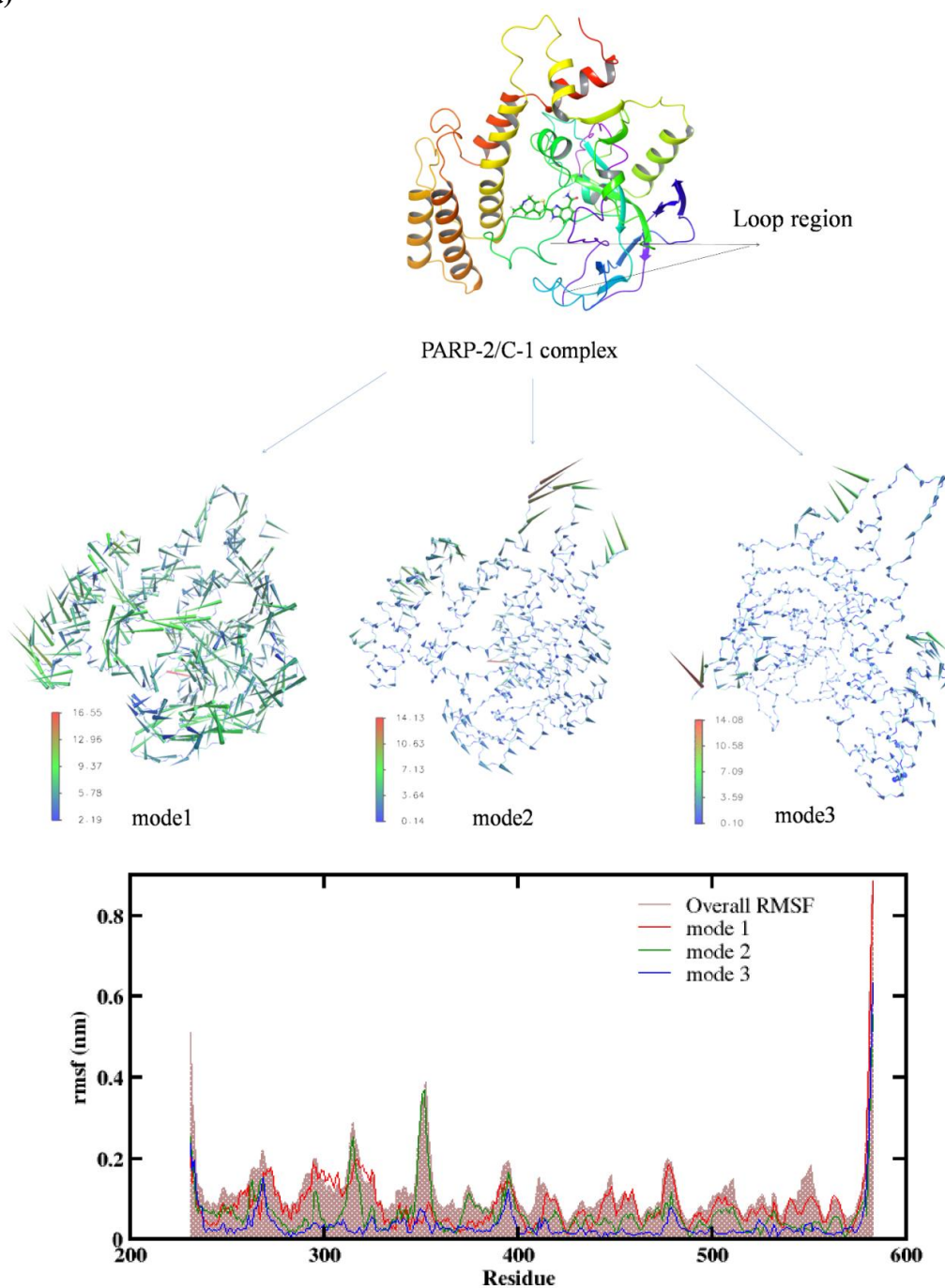




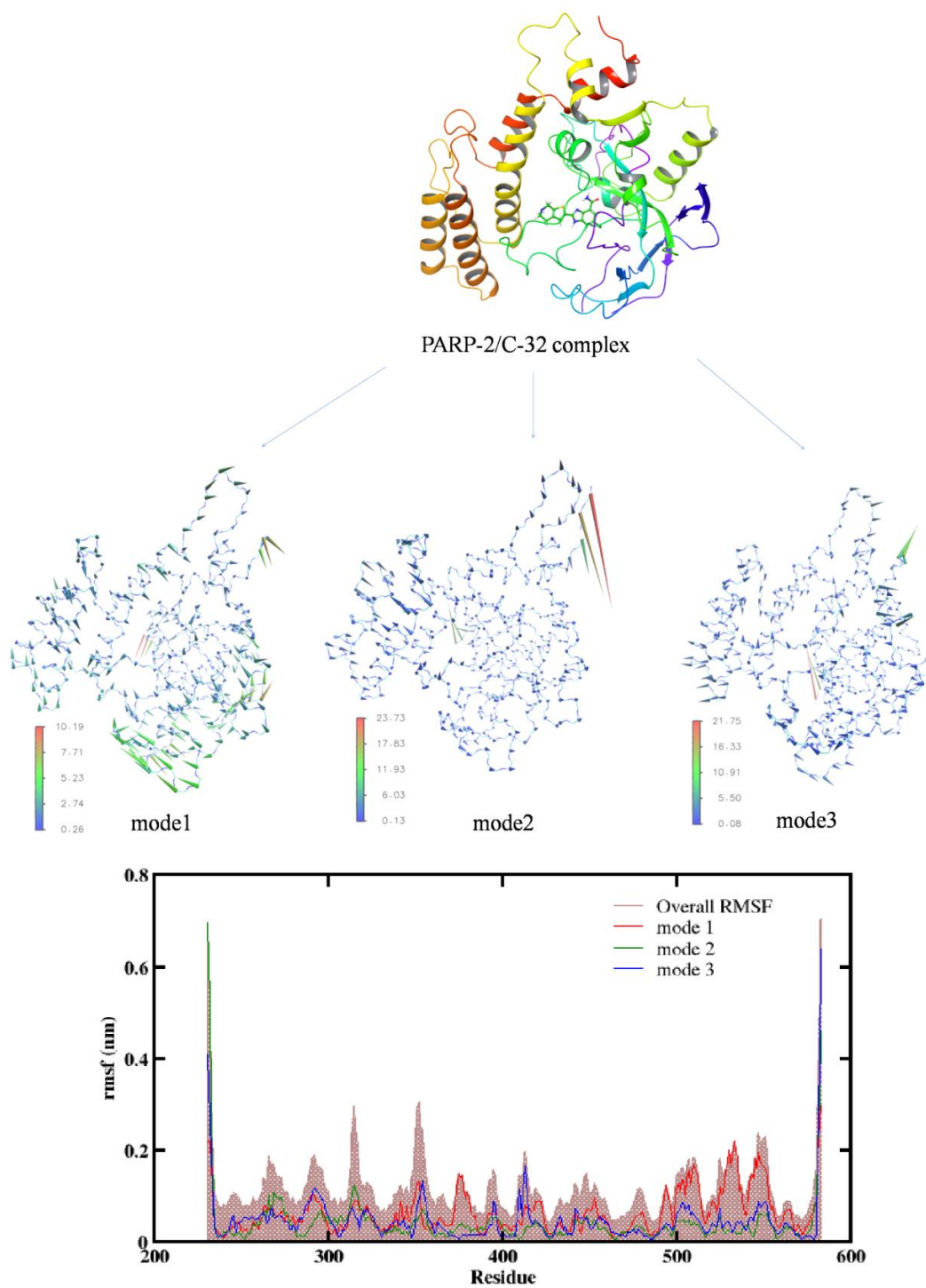


Appendix XII: Porcupine plots showing prominent motions of enzyme in complexes. The protein is represented as backbone with arrow attached to C- α atoms showing the direction and magnitude of prominent motions along mode 1, mode 2 and mode 3. The overall RMSF graph of enzyme in complexes superimposed with the amount of fluctuations of amino acid residues in individual modes (below) (a) PARP-2/C-1 complex (b) PARP-2/C-32 complex and (c) PARP-2 enzyme

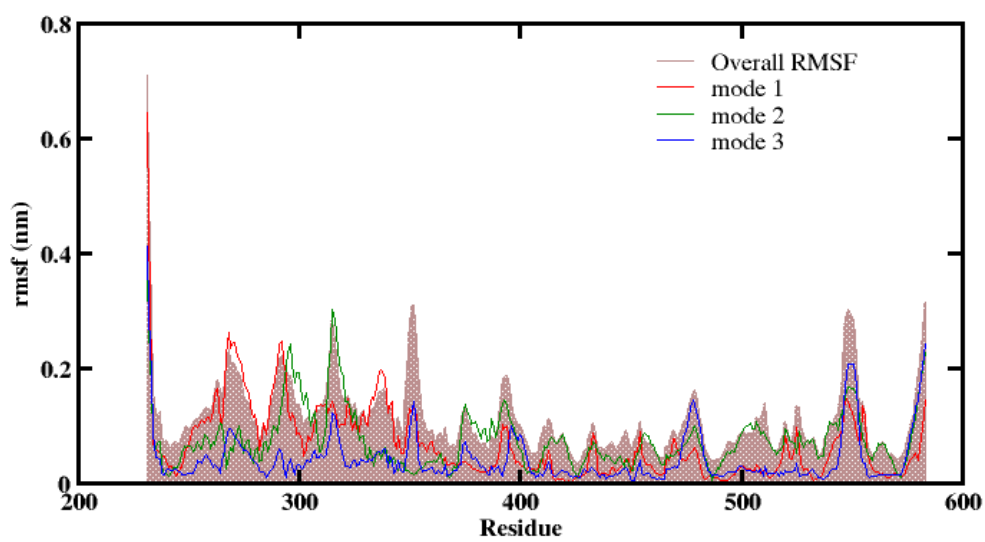
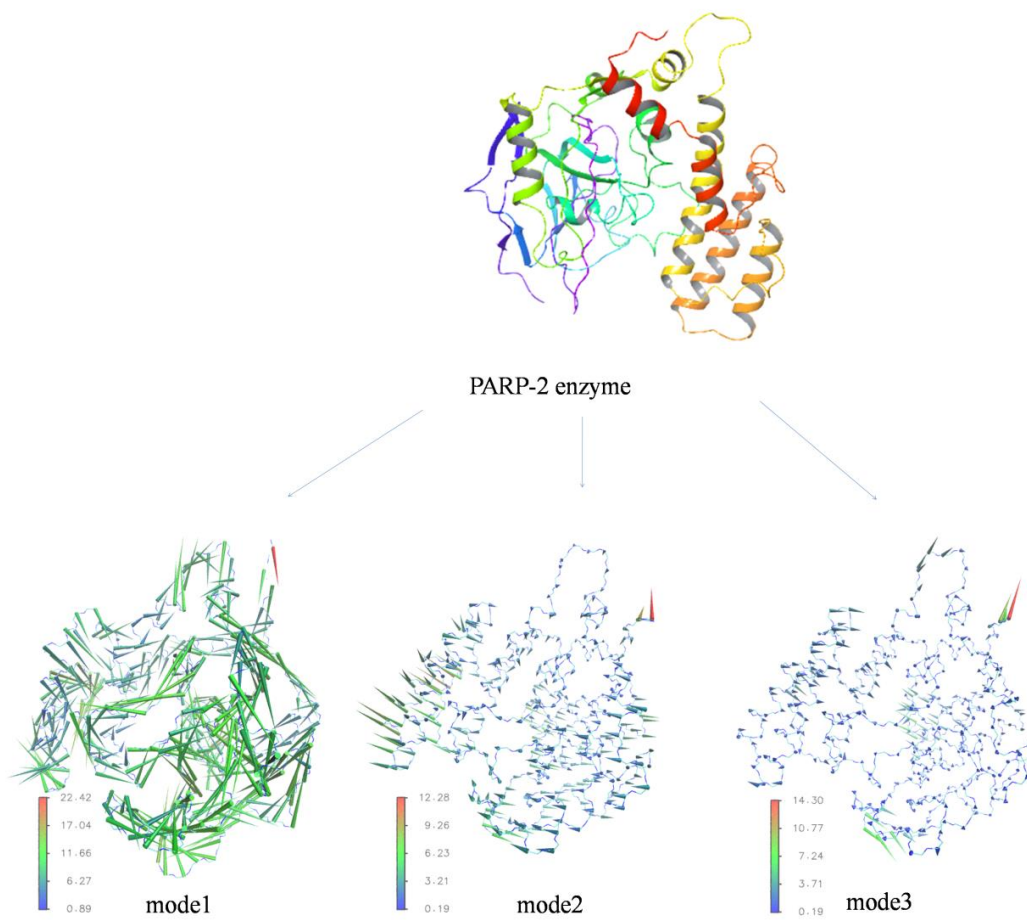
(a)



(b)



(c)



Appendix XIII: Hydrogen bond occupancy between the predicted drug molecules and the amino acid residues at binding pocket of enzyme

System	Hydrogen Bond Occupancy	
PARP-2/P-1	Lig (H) --- Tyr462 (O)	88.8%
	Lys469 (H) --- Lig (O)	51.0%
	Met456 (H) --- Lig (O)	49.8%
	Gly429 (H) --- Lig (N)	77.3%
	Gln332 (H) --- Lig (O)	48.2%
PARP-2/P-2	Gly429 (H) --- Lig (N)	15.5%
PARP-2/P-3	Lig (H) --- Tyr473 (OH)	10.8%
	Met456 (H) --- Lig (O)	21.9%
	Asn434 (H) --- Lig (O)	21.5%
	Ser430 (H) --- Lig (O)	24.7%
	Gly429 (H) --- Lig (N)	60.5%
	His428 (H) --- Lig (N)	29.9%
	Gln332 (H) --- Lig (N)	27.9%
PARP-2/P-4	Lig (H) --- Glu558 (O)	27.1%
	Lig (H) --- Gly429 (O)	12.0%
	Lig (H) --- His428 (N)	31.5%
	Lys469 (H) --- Lig (O)	68.5%
	Gln324 (H) --- Lig (N)	12.0%

Appendix XIV: Contribution of energy components to MM/PBSA binding free energy for PARP enzymes with predicted drug molecules (kcal/mol)

System	van der Waals Energy	Electrostatic Energy	Polar Solvation Energy	SASA	Binding Energy
PARP-2/P-1	-57.43 ± 3.4	-21.0 ± 3.9	59.4 ± 3.8	-5.4 ± 0.2	-24.4 ± 3.5
PARP-2/P-2	-62.76 ± 3.1	-5.9 ± 2.9	47.5 ± 5.4	-5.7 ± 0.2	-26.6 ± 4.4
PARP-2/P-3	-57.14 ± 3.8	-13.1 ± 2.9	47.9 ± 5.7	-5.6 ± 0.3	-28.0 ± 3.2
PARP-2/P-4	-60.80 ± 4.2	-14.3 ± 3.5	50.2 ± 4.5	-5.8 ± 0.3	-30.7 ± 4.0

Appendix XV: Per residue binding energy contribution in the predicted complexes

Energy Term	Major Contribution			
	Energy in kcal/mol			
	PARP-2/P-1	PARP-2/P-2	PARP-2/P-3	PARP-2/P-4
SASA	Gln332 ~ -0.23	Ser328 ~ -0.21	Ser328 ~ -0.11	Ser328 ~ -0.15
	His428 ~ -0.14	Gln332 ~ -0.21	Gln332 ~ -0.28	Gln332 ~ -0.16
	Tyr455 ~ -0.18	His428 ~ -0.14	His428 ~ -0.13	His428 ~ -0.13
	Tyr462 ~ -0.40	Gly454 ~ -0.21	Tyr455 ~ -0.19	Gly454 ~ -0.15
	Tyr473 ~ -0.40	Tyr455 ~ -0.10	Tyr462 ~ -0.16	Tyr455 ~ -0.12
		Tyr462 ~ -0.17	Tyr473 ~ -0.23	Tyr462 ~ -0.27
		Tyr473 ~ -0.34	Tyr473 ~ -0.33	
van der Waals Energy	His428 ~ -2.60	His428 ~ -2.57	His428 ~ -3.02	His428 ~ -2.26
	Tyr455 ~ -2.54	Tyr455 ~ -1.34	Tyr455 ~ -2.65	Tyr455 ~ -1.78
	Met456 ~ -1.46	Tyr462 ~ -1.62	Met456 ~ -0.78	Tyr462 ~ -3.47
	Tyr462 ~ -5.45	Phe463 ~ -1.58	Tyr462 ~ -1.81	Phe463 ~ -0.76
	Tyr473 ~ -3.81	Tyr473 ~ -3.81	Tyr473 ~ -3.46	Tyr473 ~ -3.74
Electrostatic Energy	Ser328 ~ -1.22	Ser328 ~ -2.52	Ser328 ~ -1.42	Ser328 ~ -1.72
	Gln332 ~ -2.45	Gln332 ~ -1.40	Gln332 ~ -2.56	Gln332 ~ -1.29
	Trp427 ~ -1.73	Trp427 ~ -0.36	Trp427 ~ -0.61	Trp427 ~ -0.56
	Gly429 ~ -1.53	Gly429 ~ -1.57	Gly429 ~ -1.63	Gly429 ~ -1.04
	Arg431 ~ -1.37	Arg431 ~ -1.64	Ser430 ~ -1.72	Lys469 ~ -4.77
	Lys469 ~ -4.50		Lys469 ~ -1.47	
Polar Solvation Energy	Ser328 ~ 1.37	Ser328 ~ 1.53	Ser328 ~ 1.28	Ser328 ~ 1.73
	Gln332 ~ 2.35	Gln332 ~ 0.79	Gln332 ~ 2.16	Gln332 ~ 1.12
	Gly429 ~ 2.14	Gly429 ~ 2.05	Gly429 ~ 2.22	Gly429 ~ 1.61
	Tyr462 ~ 3.19	Tyr462 ~ 0.80	Tyr462 ~ 1.25	Tyr462 ~ 1.36
	Lys469 ~ 6.28	Lys469 ~ 0.31	Lys469 ~ 2.61	Lys469 ~ 6.35
	Tyr473 ~ 2.10	Tyr473 ~ 1.48	Tyr473 ~ 1.10	Tyr473 ~ 2.00
	Glu558 ~ 1.77	Glu558 ~ 2.23	Glu558 ~ 1.30	Glu558 ~ 0.44

Appendix XVI: Predicted free energy change ($\Delta\Delta G = \Delta G_{\text{wild}} - \Delta G_{\text{variant}}$) upon alanine substitution

Mutation	$\Delta\Delta G_{\text{Pred}}$ (kcal/mol)	Outcome	Mutation	$\Delta\Delta G_{\text{Pred}}$ (kcal/mol)	Outcome
E231A	0.0	Increased Stability	Y260A	-1.76	Decreased Stability
S232A	0.95	Increased Stability	N261A	0.14	Increased Stability
Q233A	0.06	Increased Stability	T262A	-0.58	Decreased Stability
L234A	-2.77	Decreased Stability	K263A	0.18	Increased Stability
D235A	0.15	Increased Stability	K264A	0.18	Increased Stability
L236A	0.15	Increased Stability	P266A	-0.36	Decreased Stability
V238A	-0.88	Decreased Stability	L267A	-3.25	Decreased Stability
Q239A	0.08	Increased Stability	G268A	1.57	Increased Stability
E240A	-0.02	Decreased Stability	K269A	0.8	Increased Stability
L241A	-3.26	Decreased Stability	L270A	-2.75	Decreased Stability
I242A	-3.17	Decreased Stability	T271A	-0.24	Decreased Stability
K243A	0.8	Increased Stability	V272A	0.84	Increased Stability
L244A	0.15	Increased Stability	Q274A	-0.3	Decreased Stability
I245A	-3.2	Decreased Stability	I275A	0.5	Increased Stability
C246A	-0.03	Decreased Stability	K276A	0.8	Increased Stability
N247A	0.31	Increased Stability	G278A	0.48	Increased Stability
V248A	0.84	Increased Stability	Y279A	-1.19	Decreased Stability
Q249A	0.42	Increased Stability	Q280A	0.39	Increased Stability
M251A	-1.04	Decreased Stability	S281A	2.15	Increased Stability
E252A	-0.02	Decreased Stability	L282A	-3.54	Decreased Stability
E253A	-0.02	Decreased Stability	K283A	0.8	Increased Stability
M254A	-0.58	Decreased Stability	K284A	0.72	Increased Stability
M255A	-1.04	Decreased Stability	I285A	-3.16	Decreased Stability
M256A	0.53	Increased Stability	E286A	-0.07	Decreased Stability
E257A	-0.02	Decreased Stability	D287A	1.24	Increased Stability
M258A	-1.16	Decreased Stability	C288A	0.06	Increased Stability
K259A	0.39	Increased Stability	I289A	-2.31	Decreased Stability

G292A	-2.6	Decreased Stability	K325A	0.8	Increased Stability
Q293A	0.19	Increased Stability	E326A	-0.3	Decreased Stability
H294A	-0.22	Decreased Stability	L327A	-3.23	Decreased Stability
G295A	0.02	Increased Stability	S328A	1.58	Increased Stability
R296A	0.48	Increased Stability	E329A	-0.02	Decreased Stability
L298A	-3.28	Decreased Stability	K330A	-1.39	Decreased Stability
M299A	0.73	Increased Stability	I331A	-2.29	Decreased Stability
E300A	-0.02	Decreased Stability	Q332A	0.42	Increased Stability
C302A	-0.7	Decreased Stability	L333A	-3.24	Decreased Stability
N303A	1.61	Increased Stability	L334A	-3.27	Decreased Stability
E304A	-0.1	Decreased Stability	E335A	-0.02	Decreased Stability
F305A	-2.37	Decreased Stability	L337A	-3.27	Decreased Stability
Y306A	-1.26	Decreased Stability	G338A	1.57	Increased Stability
T307A	2.06	Increased Stability	D339A	1.24	Increased Stability
R308A	-0.09	Decreased Stability	I340A	-2.72	Decreased Stability
I309A	-4.19	Decreased Stability	E341A	-0.02	Decreased Stability
P310A	-0.36	Decreased Stability	I342A	0.5	Increased Stability
H311A	-1.83	Decreased Stability	I344A	0.5	Increased Stability
D312A	-0.19	Decreased Stability	K345A	0.8	Increased Stability
F313A	-3.07	Decreased Stability	L346A	0.15	Increased Stability
G314A	-2.15	Decreased Stability	V347A	-0.81	Decreased Stability
L315A	-0.27	Decreased Stability	K348A	0.68	Increased Stability
R316A	0.28	Increased Stability	T349A	0.81	Increased Stability
T317A	0.81	Increased Stability	E350A	0.64	Increased Stability
P318A	-0.36	Decreased Stability	L351A	-0.27	Decreased Stability
P319A	-0.36	Decreased Stability	Q352A	0.26	Increased Stability
L320A	0.09	Increased Stability	S353A	0.14	Increased Stability
I321A	-3.82	Decreased Stability	P354A	0.67	Increased Stability
R322A	0.11	Increased Stability	E355A	0.52	Increased Stability
T323A	-0.85	Decreased Stability	H356A	-0.03	Decreased Stability
Q324A	0.42	Increased Stability	P357A	2.47	Increased Stability

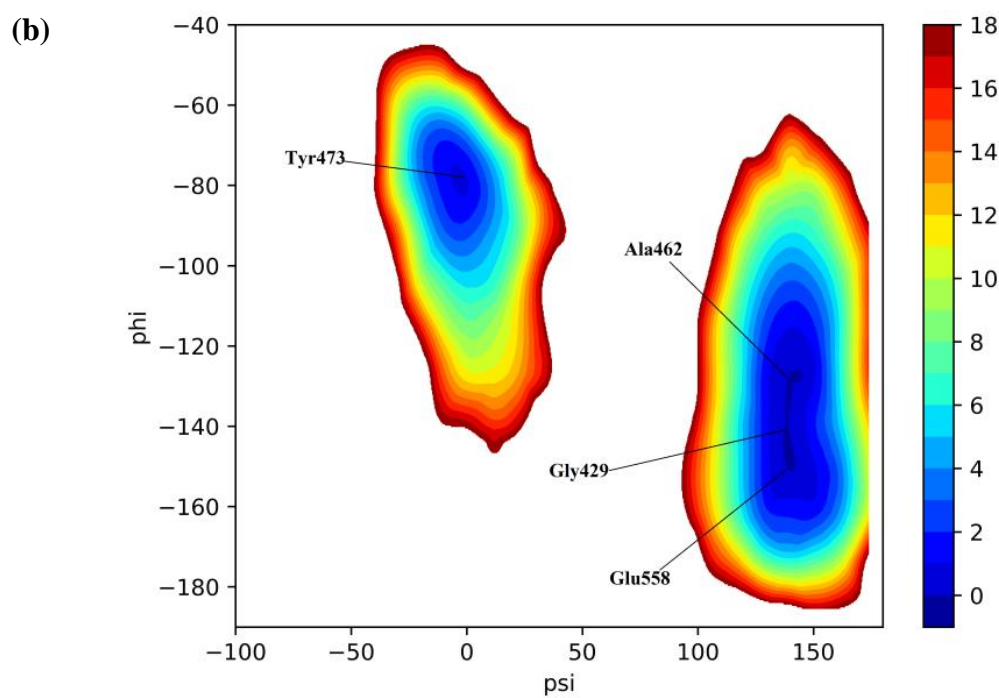
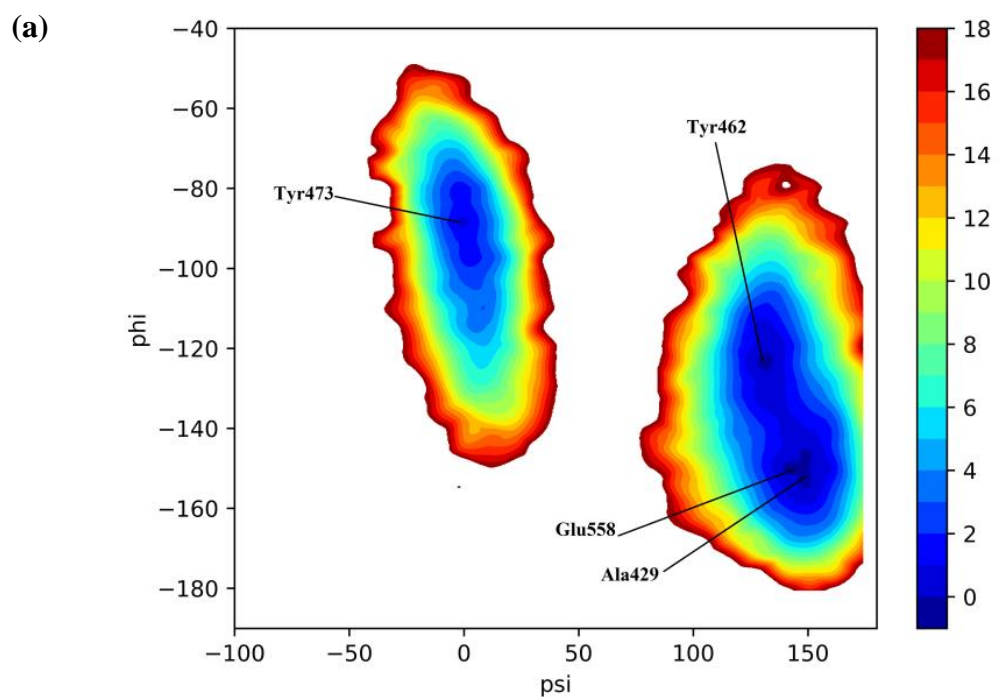
L358A	-2.63	Decreased Stability	H390A	-0.97	Decreased Stability
D359A	-0.17	Decreased Stability	P392A	0.67	Increased Stability
Q360A	-1.46	Decreased Stability	T393A	-0.07	Decreased Stability
H361A	0.61	Increased Stability	H394A	0.12	Increased Stability
Y362A	0.12	Increased Stability	S395A	0.4	Increased Stability
R363A	0.41	Increased Stability	D396A	0.02	Increased Stability
N364A	1.61	Increased Stability	Y397A	-2.36	Decreased Stability
L365A	-3.25	Decreased Stability	T398A	-0.48	Decreased Stability
H366A	-0.02	Decreased Stability	M399A	-1.75	Decreased Stability
C367A	-1.46	Decreased Stability	T400A	-0.48	Decreased Stability
L369A	-3.97	Decreased Stability	L401A	-4.0	Decreased Stability
R370A	-0.59	Decreased Stability	L402A	-0.97	Decreased Stability
P371A	1.7	Increased Stability	D403A	0.08	Increased Stability
L372A	-3.34	Decreased Stability	L404A	-3.93	Decreased Stability
D373A	0.7	Increased Stability	F405A	-3.7	Decreased Stability
H374A	-0.5	Decreased Stability	E406A	-0.59	Decreased Stability
E375A	0.06	Increased Stability	V407A	-2.85	Decreased Stability
S376A	0.3	Increased Stability	E408A	-0.4	Decreased Stability
Y377A	0.56	Increased Stability	K409A	-1.67	Decreased Stability
E378A	-0.02	Decreased Stability	D410A	0.91	Increased Stability
F379A	-2.03	Decreased Stability	G411A	-3.13	Decreased Stability
K380A	0.8	Increased Stability	E412A	-0.3	Decreased Stability
V381A	-0.46	Decreased Stability	K413A	0.8	Increased Stability
I382A	-3.17	Decreased Stability	E414A	-0.02	Decreased Stability
S383A	1.49	Increased Stability	F416A	-2.46	Decreased Stability
Q384A	0.42	Increased Stability	R417A	0.3	Increased Stability
Y385A	-1.28	Decreased Stability	E418A	0.06	Increased Stability
L386A	-3.24	Decreased Stability	D419A	-0.21	Decreased Stability
Q387A	0.17	Increased Stability	L420A	-3.47	Decreased Stability
S388A	1.02	Increased Stability	H421A	0.31	Increased Stability
T389A	0.41	Increased Stability	R423A	-0.51	Decreased Stability

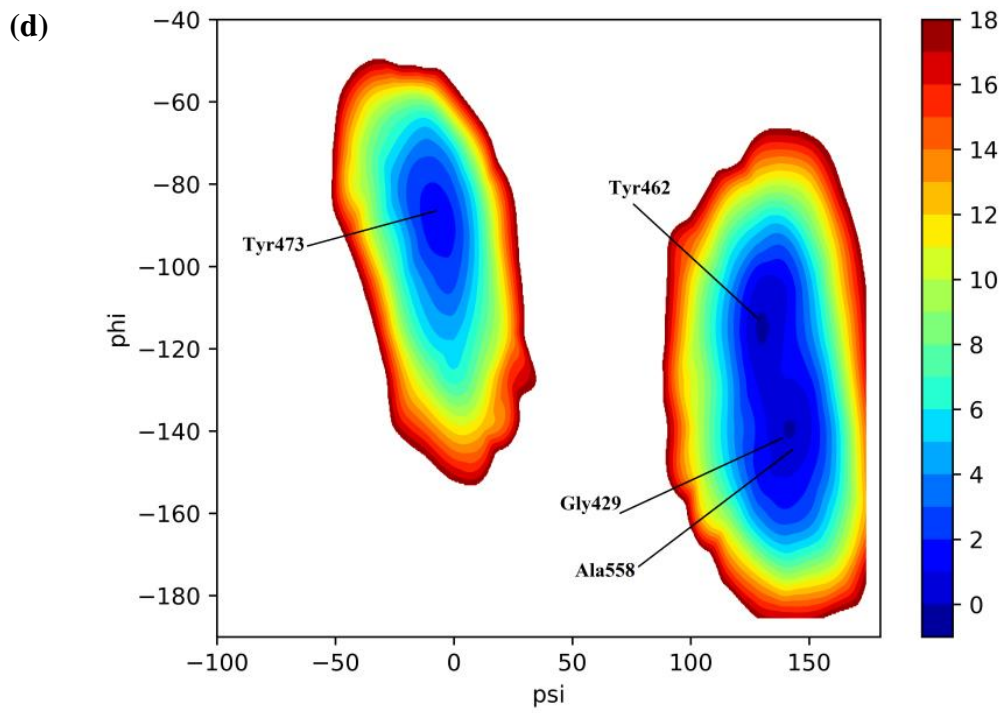
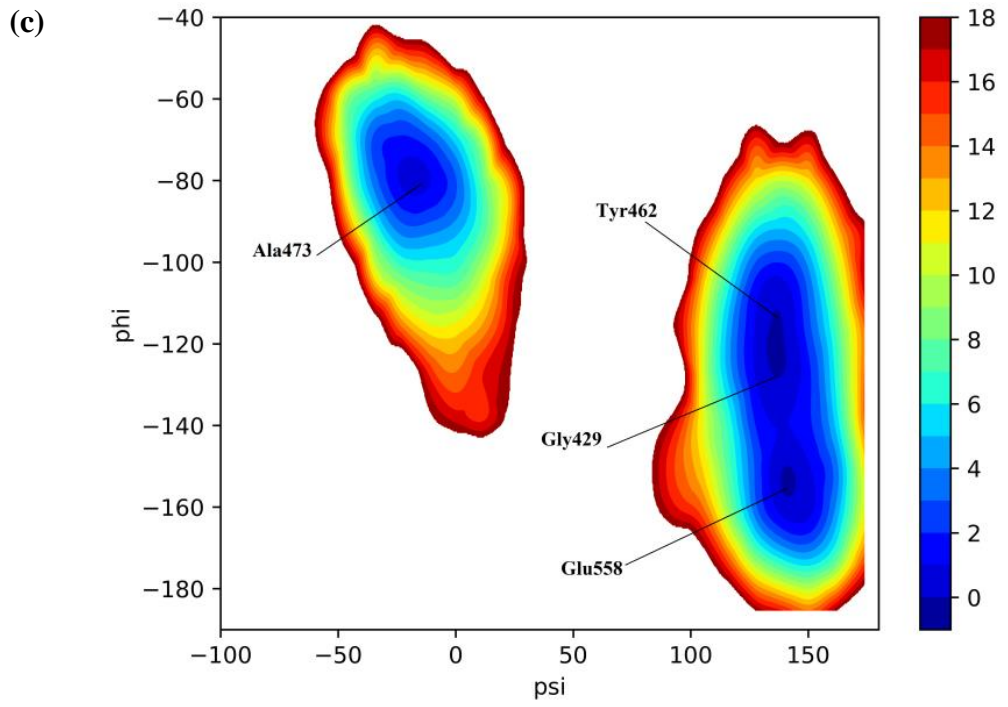
M424A	-1.43	Decreased Stability	F457A	-3.08	Decreased Stability
L425A	-3.96	Decreased Stability	G458A	-1.86	Decreased Stability
L426A	-3.83	Decreased Stability	K459A	-0.93	Decreased Stability
W427A	-2.6	Decreased Stability	G460A	-0.7	Decreased Stability
H428A	-0.51	Decreased Stability	I461A	-4.57	Decreased Stability
G429A	-0.6	Decreased Stability	Y462A	-0.82	Decreased Stability
S430A	0.08	Increased Stability	F463A	-3.82	Decreased Stability
R431A	0.47	Increased Stability	D465A	0.4	Increased Stability
M432A	0.73	Increased Stability	M466A	-1.54	Decreased Stability
S433A	1.71	Increased Stability	S467A	1.65	Increased Stability
N434A	0.08	Increased Stability	S468A	1.66	Increased Stability
W435A	-2.11	Decreased Stability	K469A	0.31	Increased Stability
V436A	-0.88	Decreased Stability	S470A	0.79	Increased Stability
G437A	0.63	Increased Stability	Y473A	0.56	Increased Stability
I438A	-2.74	Decreased Stability	C474A	-0.85	Decreased Stability
L439A	-3.22	Decreased Stability	F475A	0.24	Increased Stability
S440A	1.26	Increased Stability	S477A	-0.57	Decreased Stability
H441A	1.14	Increased Stability	R478A	0.05	Increased Stability
G442A	-2.15	Decreased Stability	L479A	-0.27	Decreased Stability
L443A	-3.43	Decreased Stability	K480A	0.99	Increased Stability
R444A	-0.37	Decreased Stability	N481A	0.03	Increased Stability
I445A	0.15	Increased Stability	T482A	-0.5	Decreased Stability
P447A	-0.36	Decreased Stability	G483A	-1.25	Decreased Stability
P448A	0.67	Increased Stability	L484A	-3.89	Decreased Stability
E449A	0.06	Increased Stability	L485A	-3.92	Decreased Stability
P451A	-1.29	Decreased Stability	L486A	-3.91	Decreased Stability
I452A	0.5	Increased Stability	L487A	-3.86	Decreased Stability
T453A	1.26	Increased Stability	S488A	1.42	Increased Stability
G454A	1.57	Increased Stability	E489A	-1.33	Decreased Stability
Y455A	0.52	Increased Stability	V490A	-2.71	Decreased Stability
M456A	0.14	Increased Stability	L492A	-3.34	Decreased Stability

G493A	-1.32	Decreased Stability	F528A	0.78	Increased Stability
Q494A	0.43	Increased Stability	V529A	-1.72	Decreased Stability
C495A	-1.13	Decreased Stability	T530A	-0.5	Decreased Stability
N496A	-0.15	Decreased Stability	L531A	-3.78	Decreased Stability
E497A	-0.4	Decreased Stability	N532A	-0.35	Decreased Stability
L498A	-0.97	Decreased Stability	G533A	-3.13	Decreased Stability
L499A	-0.27	Decreased Stability	S534A	1.28	Increased Stability
E500A	0.22	Increased Stability	T535A	-0.5	Decreased Stability
N502A	0.07	Increased Stability	V536A	-2.85	Decreased Stability
P503A	0.67	Increased Stability	P537A	1.96	Increased Stability
K504A	0.18	Increased Stability	L538A	-2.71	Decreased Stability
E506A	0.06	Increased Stability	G539A	-0.59	Decreased Stability
G507A	1.57	Increased Stability	P540A	1.37	Increased Stability
L508A	0.15	Increased Stability	S542A	0.25	Increased Stability
L509A	0.28	Increased Stability	D543A	0.08	Increased Stability
Q510A	0.55	Increased Stability	T544A	-0.09	Decreased Stability
G511A	-3.13	Decreased Stability	G545A	-2.6	Decreased Stability
K512A	0.68	Increased Stability	I546A	-0.65	Decreased Stability
H513A	-0.66	Decreased Stability	L547A	-0.52	Decreased Stability
S514A	1.23	Increased Stability	N548A	1.34	Increased Stability
T515A	0.14	Increased Stability	G551A	-0.09	Decreased Stability
K516A	-0.31	Decreased Stability	Y552A	0.68	Increased Stability
G517A	-0.6	Decreased Stability	T553A	-0.12	Decreased Stability
L518A	-0.21	Decreased Stability	L554A	0.09	Increased Stability
G519A	0.1	Increased Stability	N555A	0.03	Increased Stability
K520A	0.25	Increased Stability	Y556A	-1.95	Decreased Stability
M521A	-0.34	Decreased Stability	N557A	1.12	Increased Stability
P523A	-0.27	Decreased Stability	E558A	-0.59	Decreased Stability
S524A	0.14	Increased Stability	Y559A	-2.51	Decreased Stability
S525A	1.58	Increased Stability	I560A	-4.41	Decreased Stability
H527A	1.14	Increased Stability	V561A	-2.85	Decreased Stability

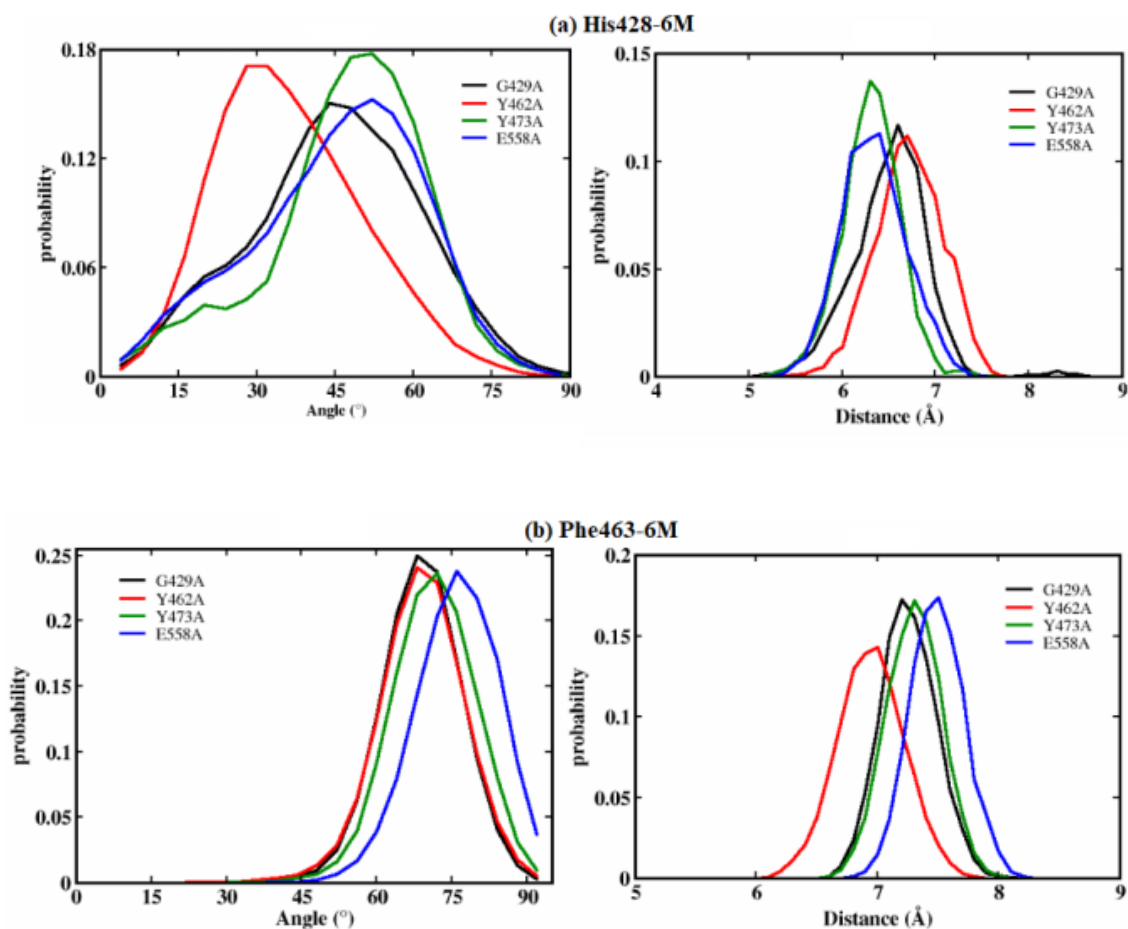
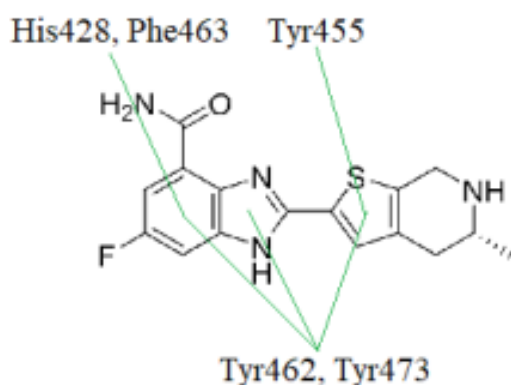
Y562A	-1.63	Decreased Stability	L573A	-3.84	Decreased Stability
N563A	0.14	Increased Stability	K574A	-1.67	Decreased Stability
P564A	2.47	Increased Stability	V575A	-2.71	Decreased Stability
N565A	1.49	Increased Stability	Q576A	-0.2	Decreased Stability
Q566A	-1.16	Decreased Stability	F577A	-3.8	Decreased Stability
V567A	-2.85	Decreased Stability	N578A	0.56	Increased Stability
R568A	-1.98	Decreased Stability	F579A	-0.6	Decreased Stability
M569A	-1.75	Decreased Stability	L580A	-0.52	Decreased Stability
R570A	-0.59	Decreased Stability	Q581A	0.38	Increased Stability
Y571A	-2.35	Decreased Stability	L582A	-0.52	Decreased Stability
L572A	-3.96	Decreased Stability	W583A	0.0	Increased Stability

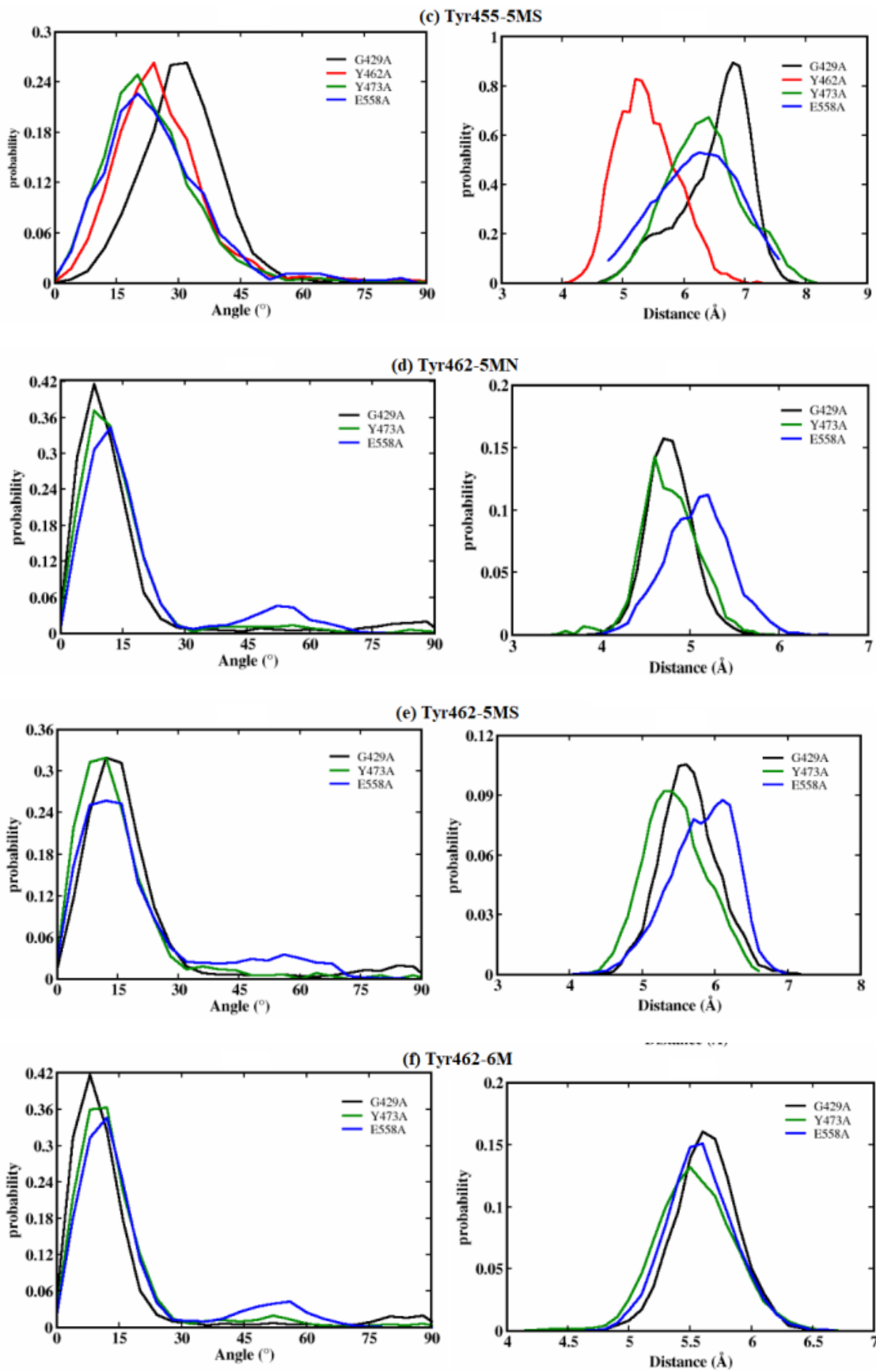
Appendix XVII: Free energy landscapes of binding site residues as a function of dihedral angles (ψ, ϕ in degrees). **(a)** Gly429Ala, **(b)** Tyr462Ala, **(c)** Tyr473Ala and **(d)** Glu558Ala

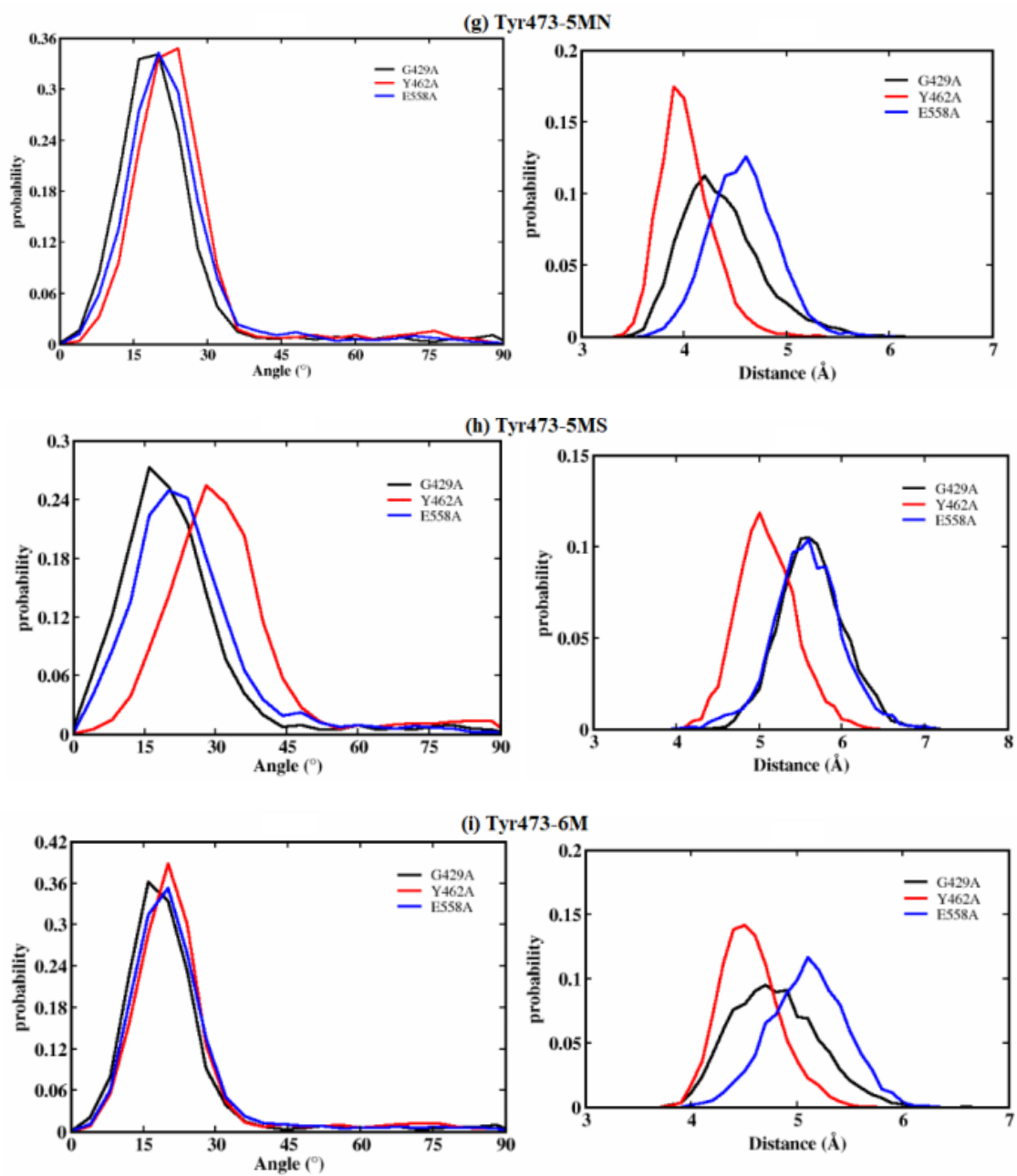




Appendix XVIII: Probability distribution of the normal vector angles and the distance between the two rings forming π -stacking interactions. 6M: 6-membered benzene ring in the ligand, 5MN: 5-membered imidazole ring in the ligand and 5MS: 5-membered thiophene ring in the ligand





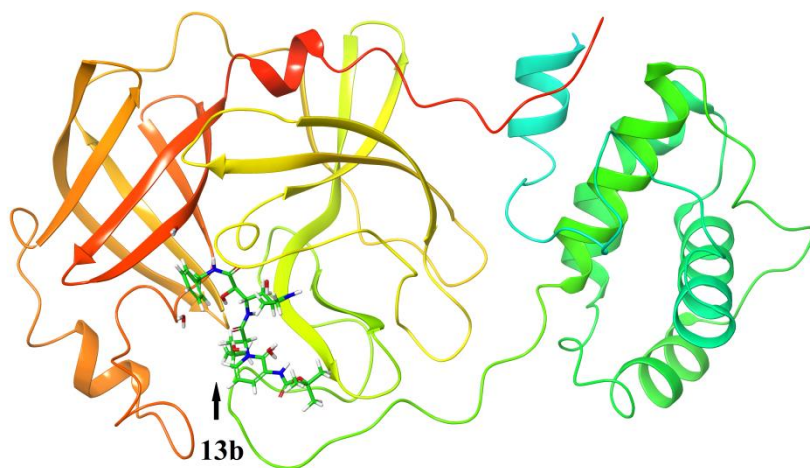


Appendix XIX: Cosine content on the principal components of backbone C- α atoms

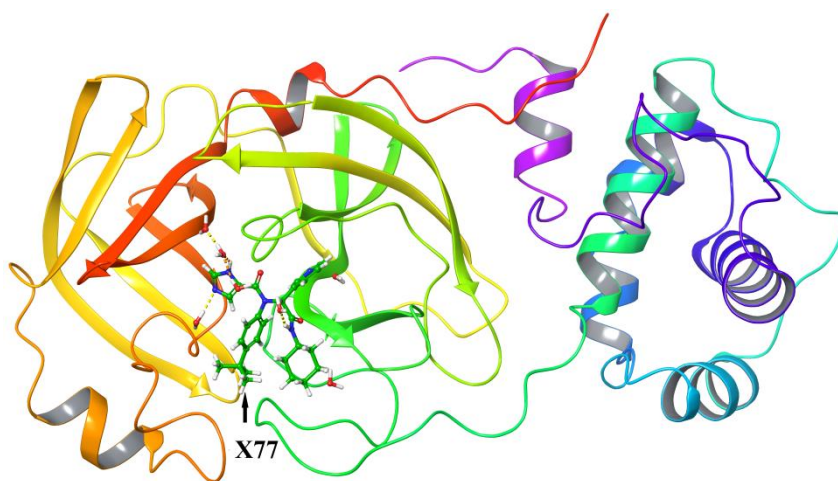
Systems	PC1	PC2	PC3
PARP-2 wild	0.712	0.005	0.017
PARP-2 wild/ligand	0.360	0.031	0.016
Gly429Ala variant	0.791	0.075	0.042
Gly429Ala variant/ligand	0.217	0.039	0.509
Tyr462Ala variant	0.783	0.045	0.059
Tyr462Ala variant/ligand	0.416	0.389	0.509
Tyr473Ala variant	0.755	0.112	0.037
Tyr473Ala variant/ligand	0.789	0.016	0.055
Glu558Ala variant	0.705	0.017	0.147
Glu558Ala variant/ligand	0.326	0.309	0.035

Appendix XX: Crystal structure of SARS CoV-2 main protease (M^{pro}) complex with **a)** alpha-ketoamide 13b (6Y2G) and **b)** broad spectrum non-covalent inhibitor X77 (6W63). **c)** Crystal structure of SARS CoV-2 main protease apo form (7KFI)

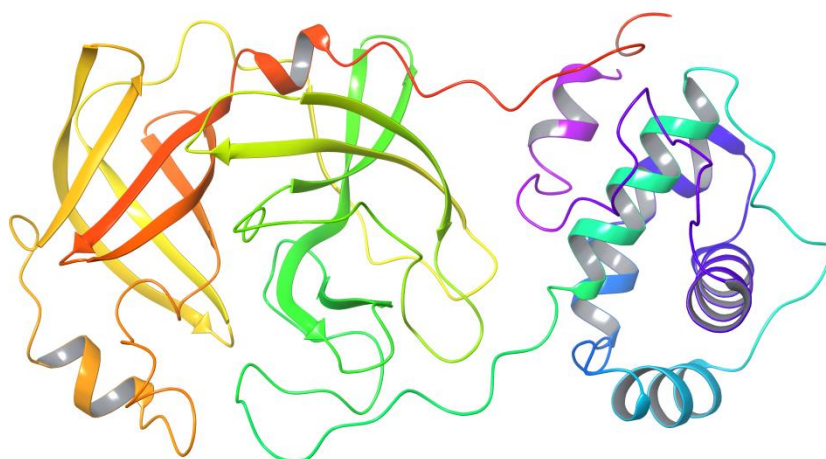
a)



b)



c)

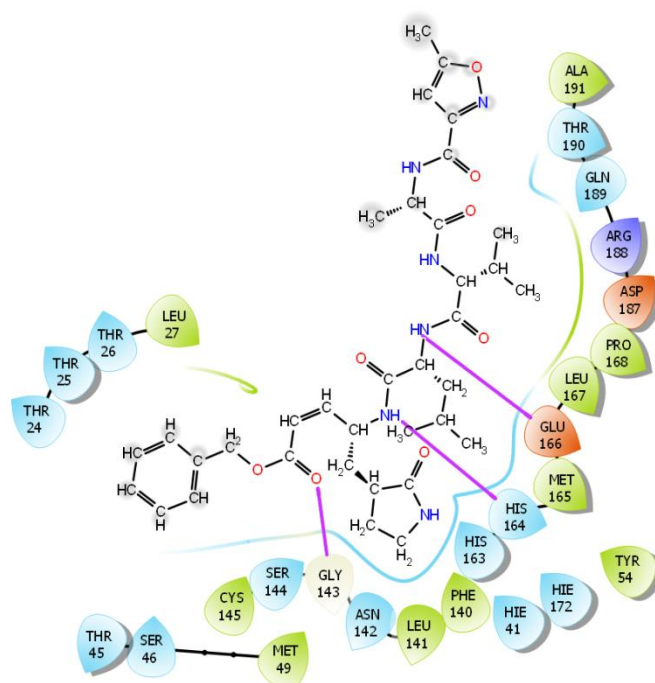


Appendix XXI: Molecular structure, common name, docking score and 2D ligand interaction diagrams of top-scored hit molecules with main protease (6LU7)

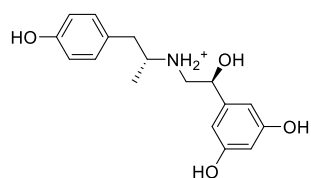
Structure and Name
Docking Score (kcal/mol)

2D Ligand Interaction Diagram

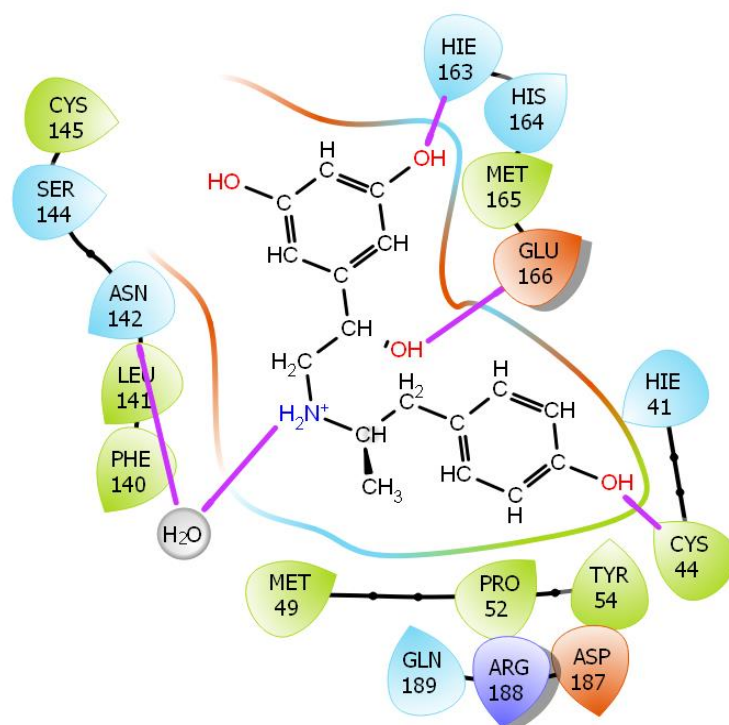
Inhibitor, N3
DS = -7.128

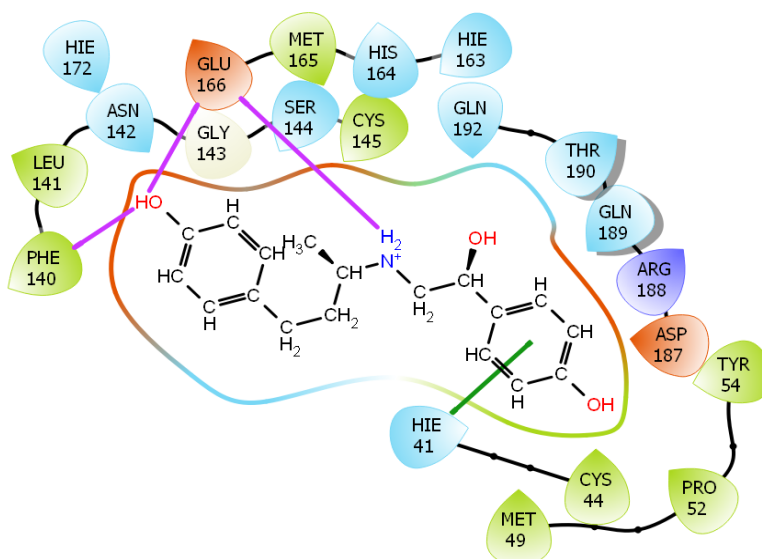
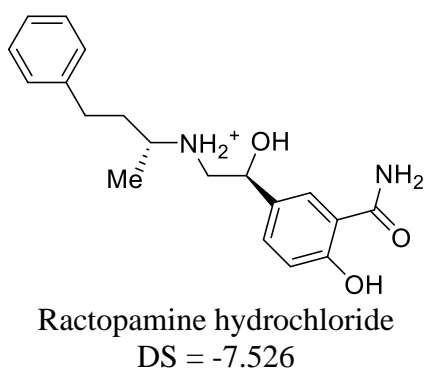
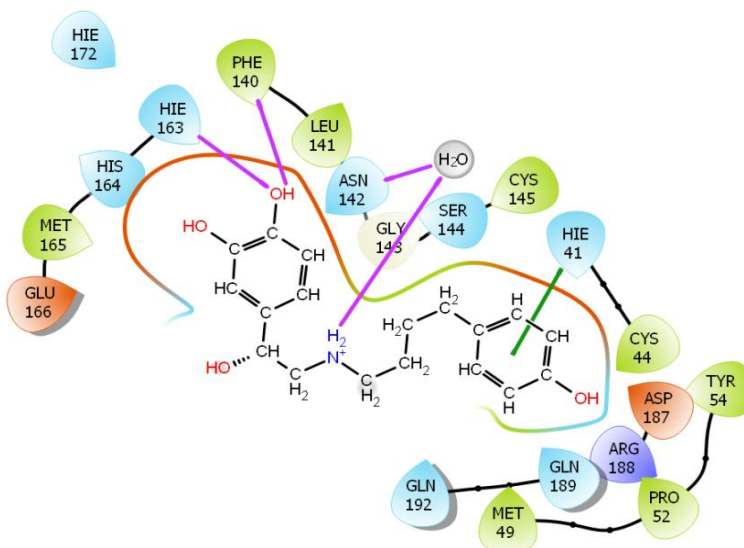
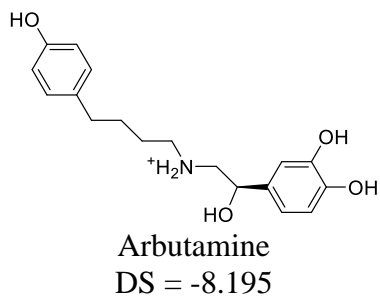
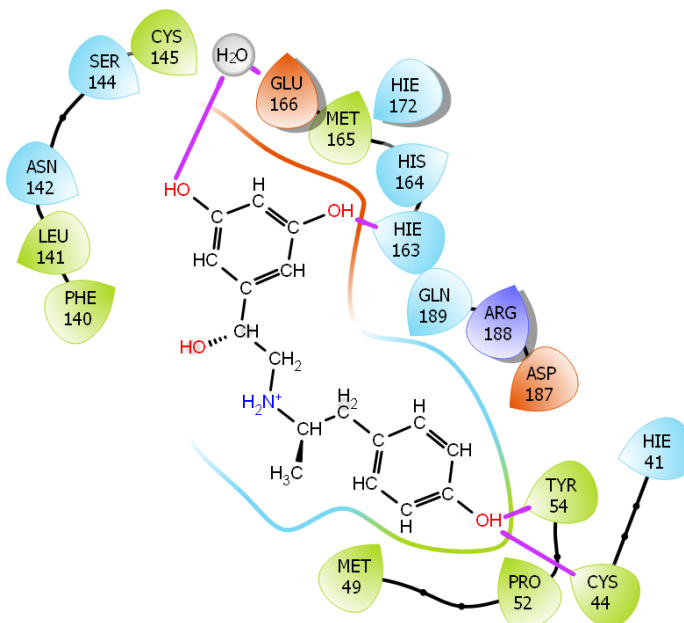
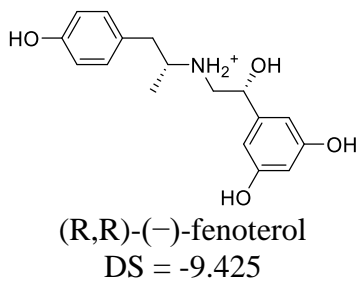


Adrenoceptor agonists

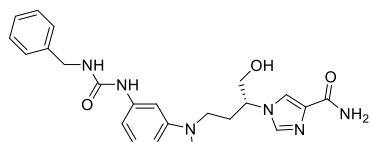


(S,R)-(+)-fenoterol
DS = -9.491

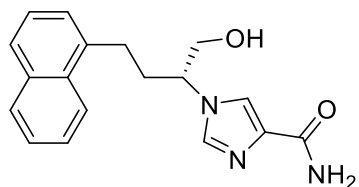
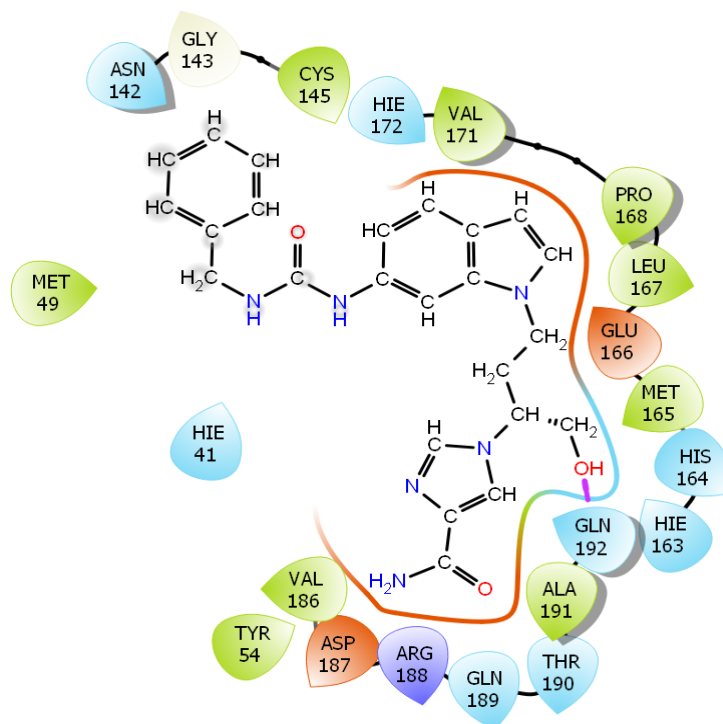




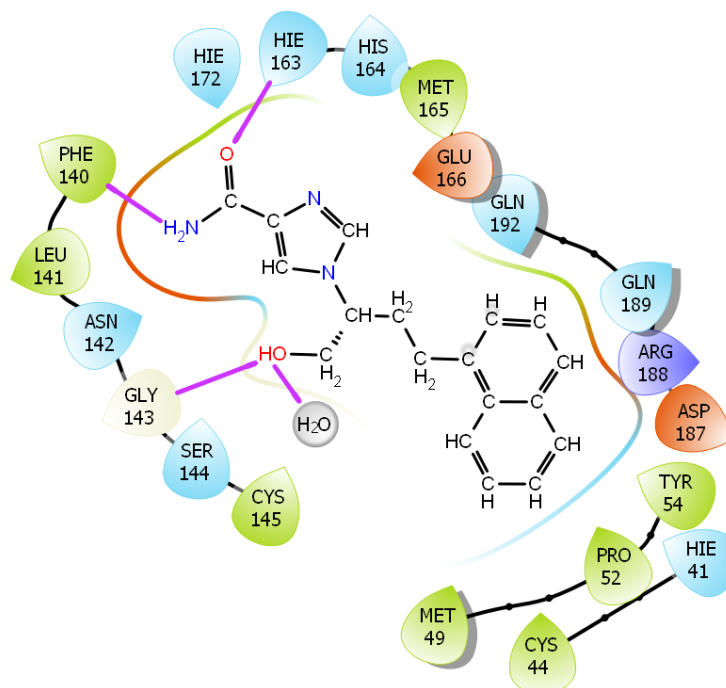
Adenosine deaminase inhibitors

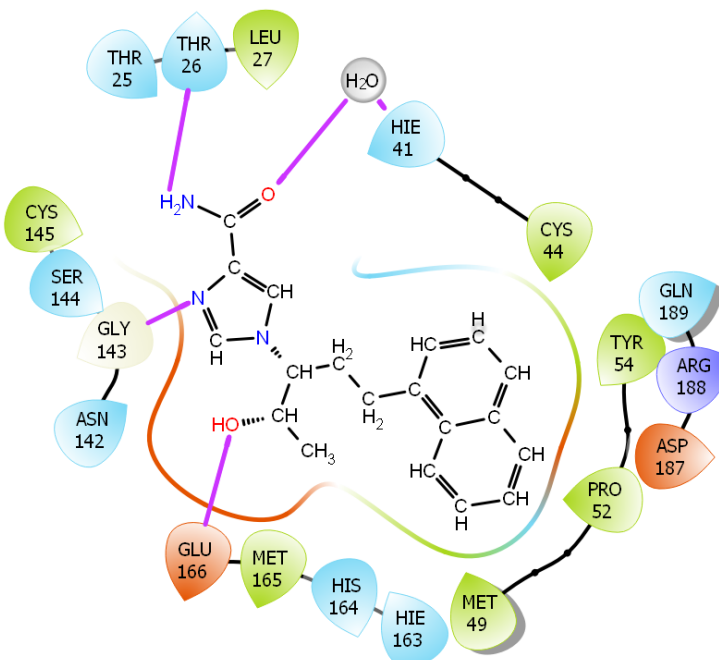
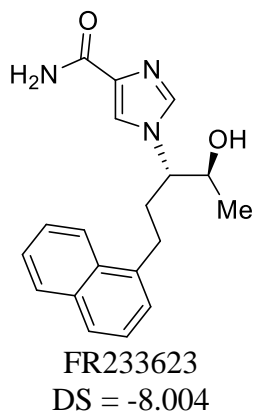
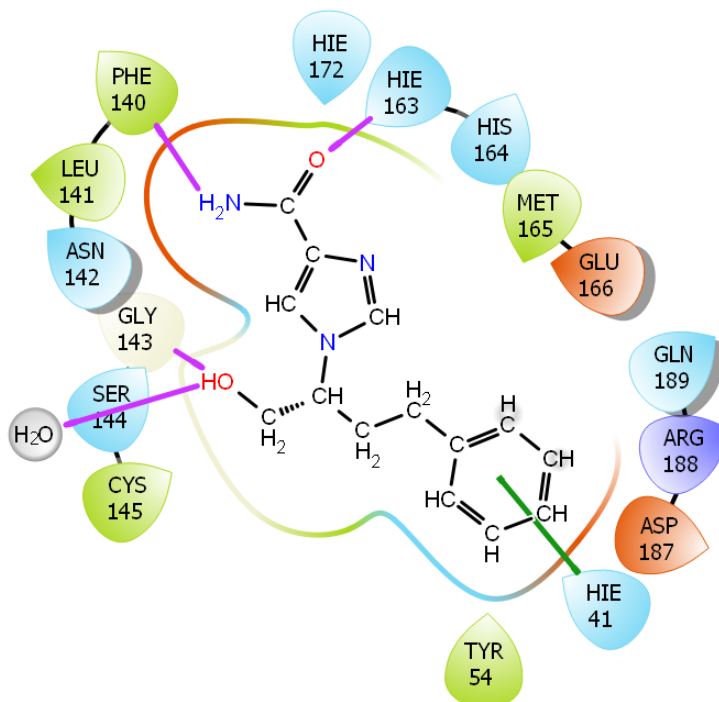
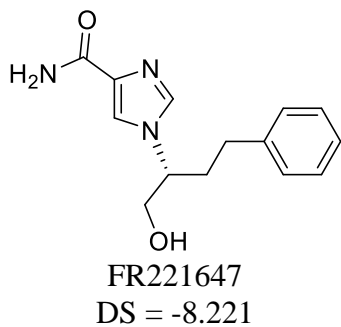


FR236913
DS = -9.163



FR230513
DS = -9.049





***LID Legend**

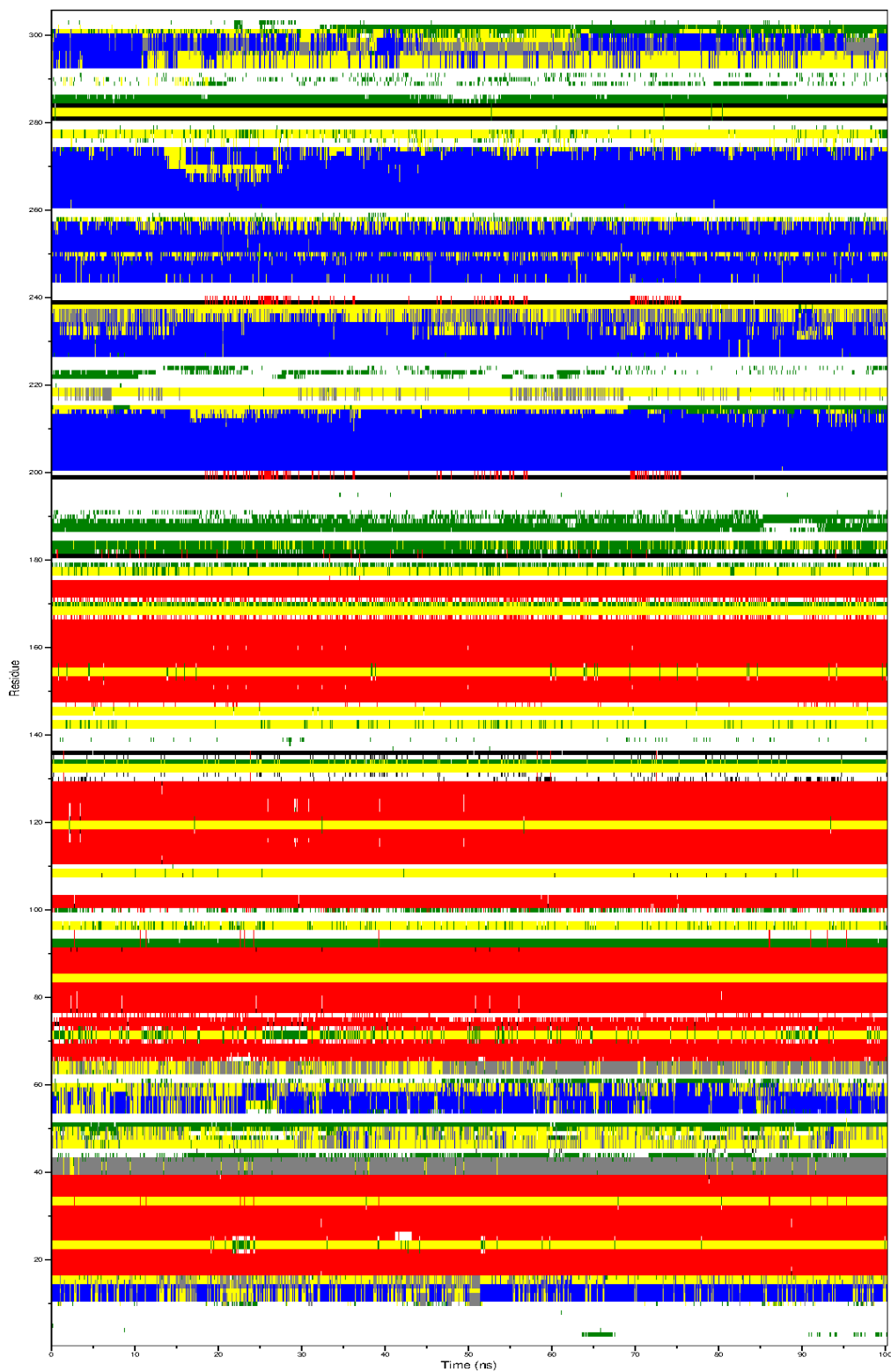
- | | | | |
|--------------------|----------------------------|--------------------|------------------|
| Charged (negative) | Polar | Distance | Pi-cation |
| Charged (positive) | Unspecified residue | H-bond | Salt bridge |
| Glycine | Water | Halogen bond | Solvent exposure |
| Hydrophobic | Hydration site | Metal coordination | |
| Metal | Hydration site (displaced) | Pi-Pi stacking | |

Appendix XXII: Docking score and interaction of existing co-crystal ligands

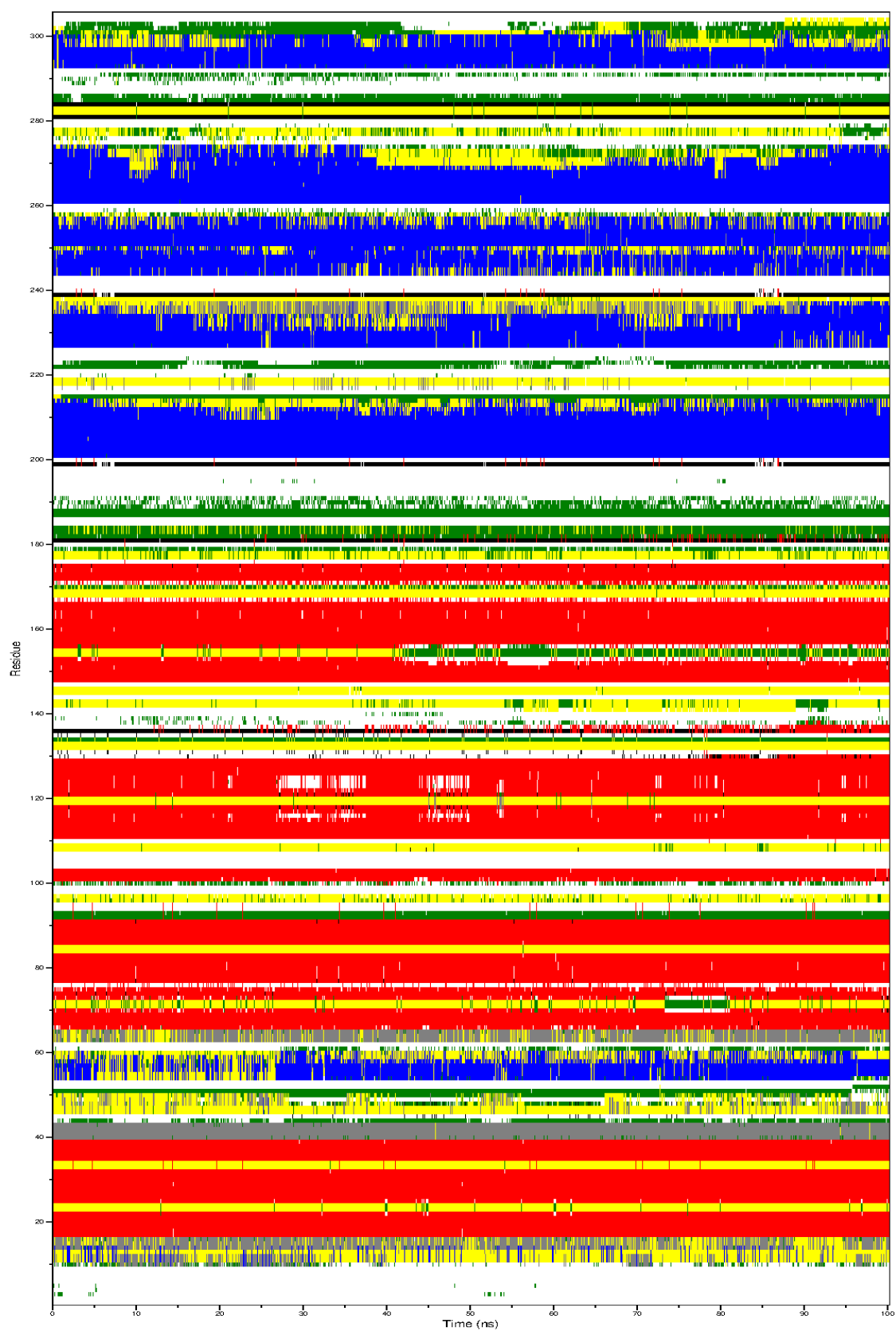
Protein Structure	Compound	Docking Score (kcal/mol)	Interactions
6W63	X77	-7.940	<p>Hie41, Phe103, Gly143, His164, Glu166</p>
6Y2G	13b	-7.574	<p>Gly143, Cys145, Hie163, His164, Glu166</p>

Appendix XXIII: Per-residue secondary structure analysis of main protease (6LU7) in the complex. **a)** MPro/+fenoterol, **b)** MPro/-fenoterol, **c)** MPro/FR236913 and **d)** MPro/FR230513

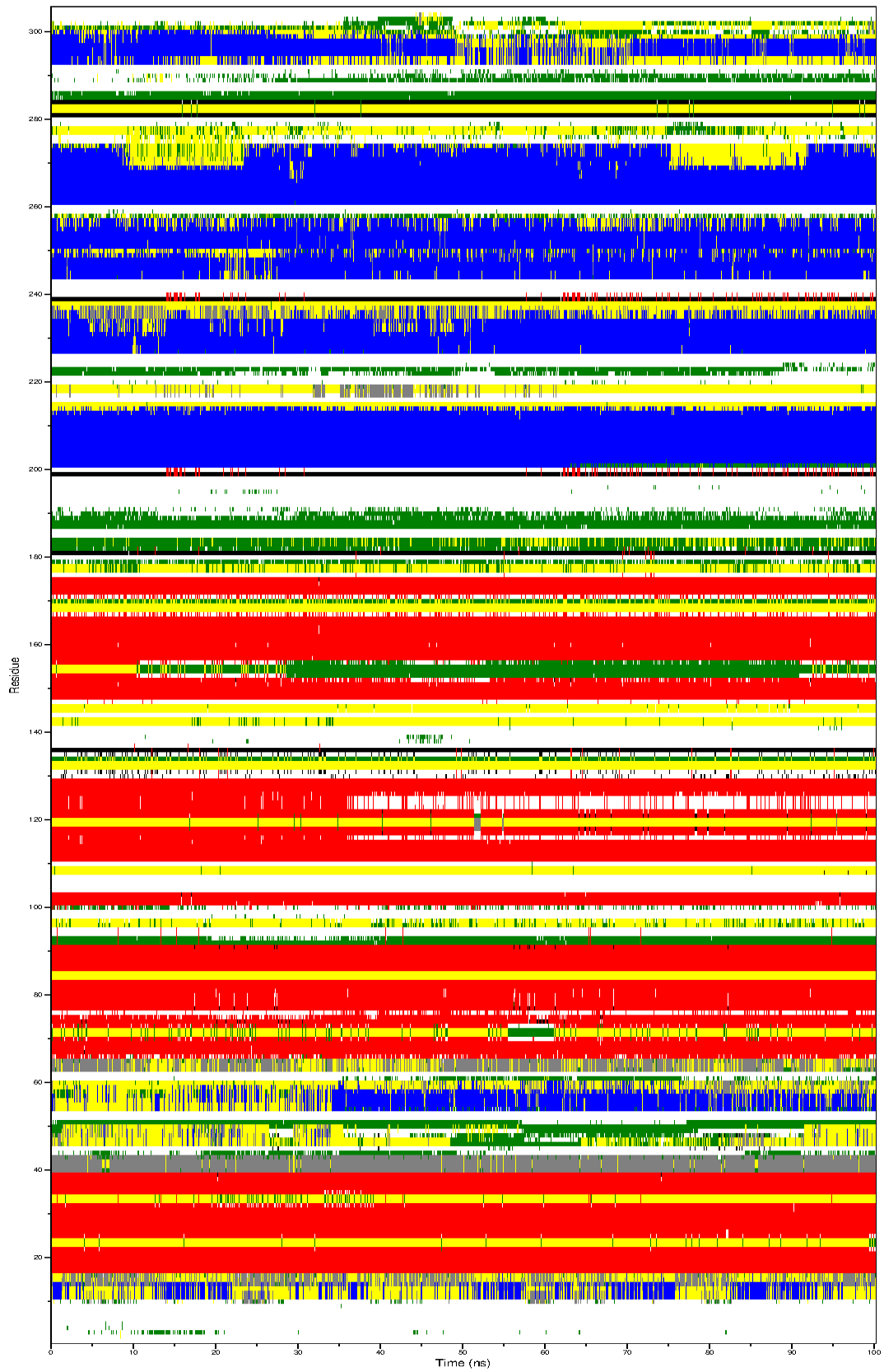
a)



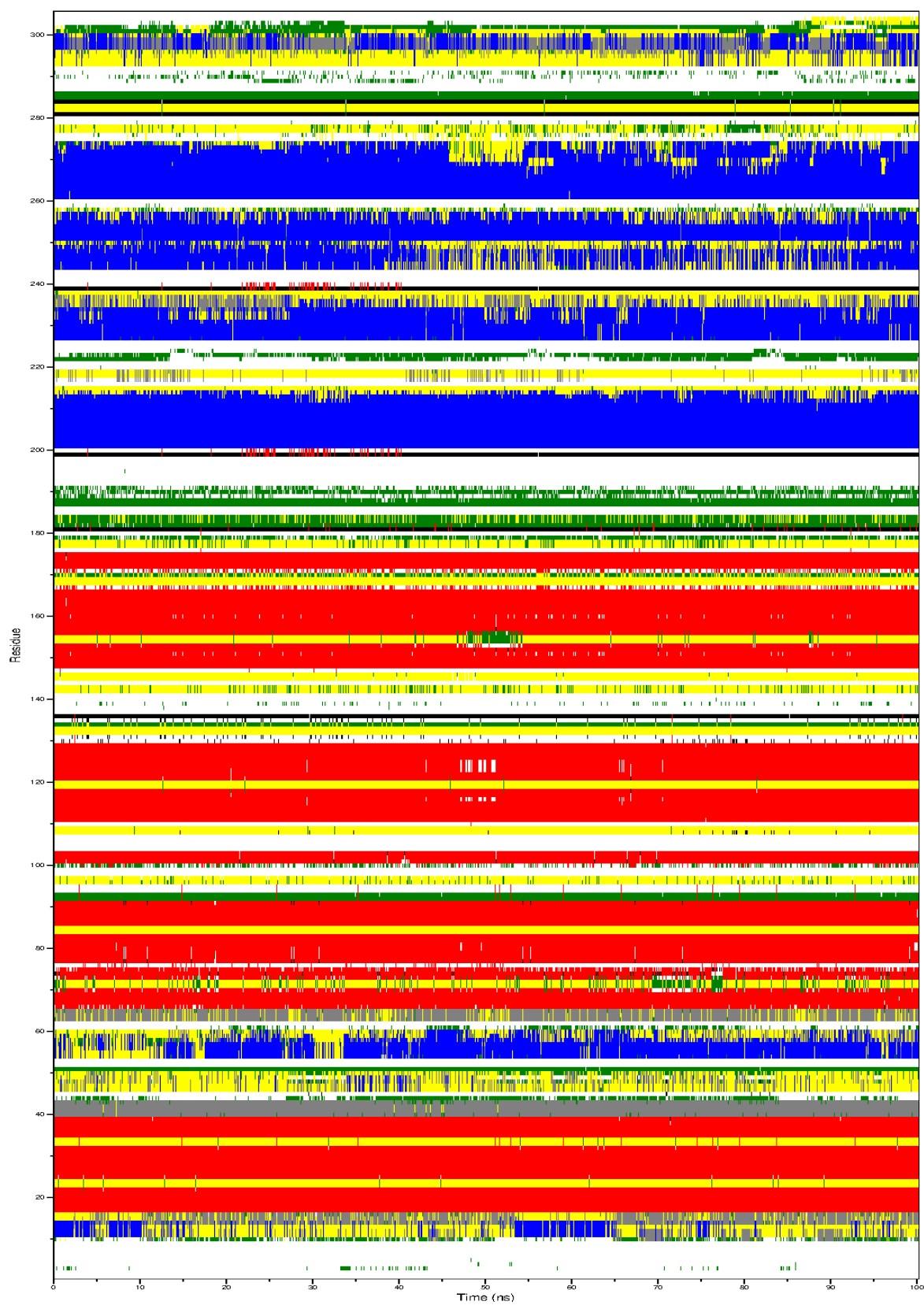
b)



c)

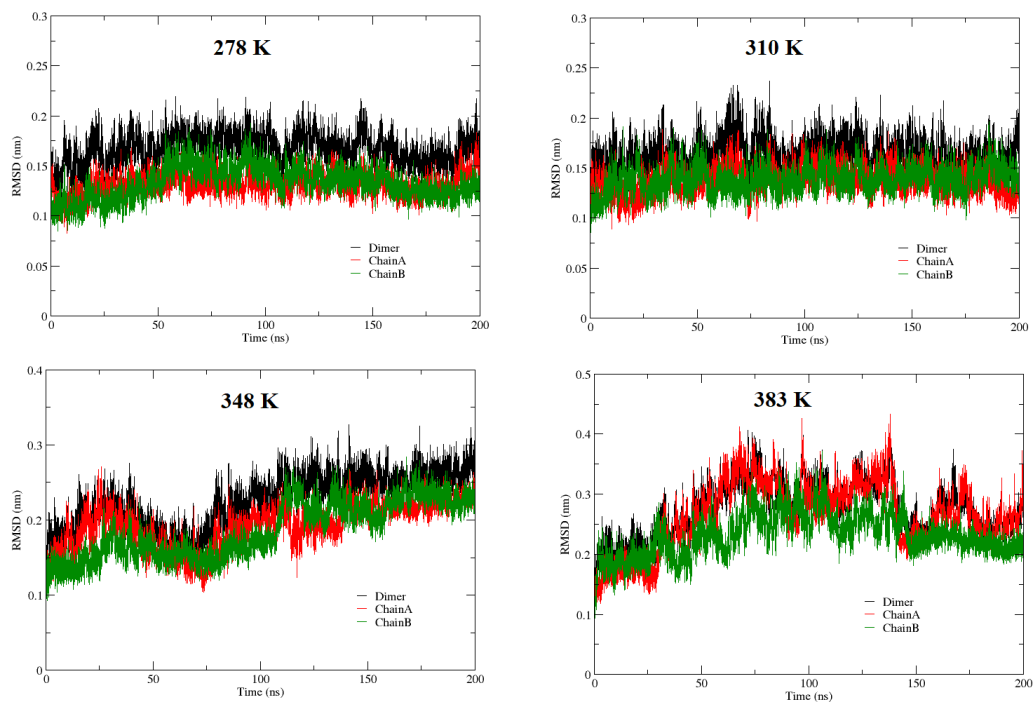


d)

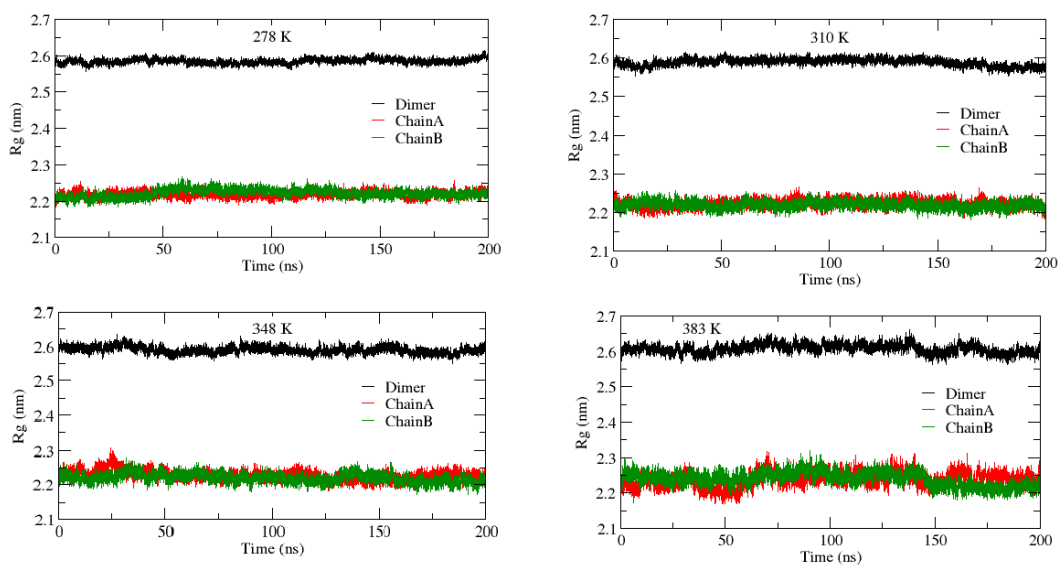


Appendix XXIV: a) RMSD profile of the dimer, Chain A and Chain B at different temperatures in SPC/E water. b) Radius of gyration (rGyr) of dimer, Chain A and Chain B at different temperatures as a function of time in SPC/E water. c) Dimer Interaction energy (kJ/mol) vs. Temperature (K)

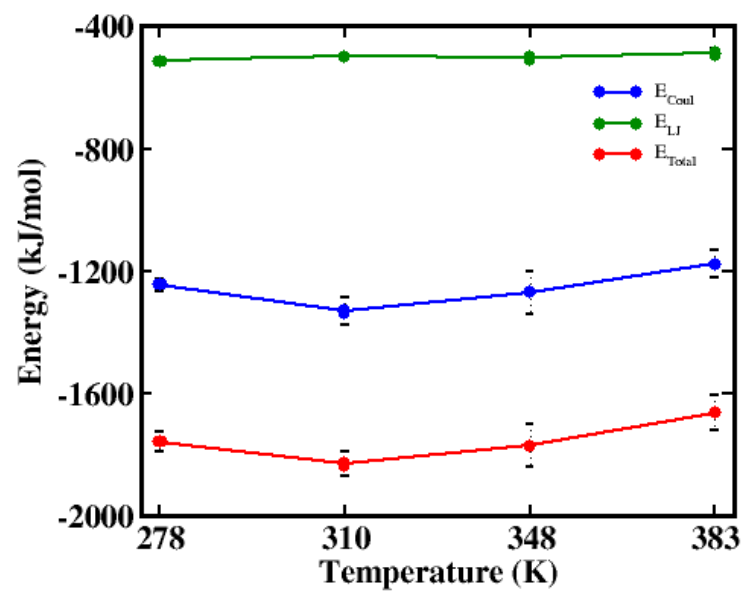
a)



b)



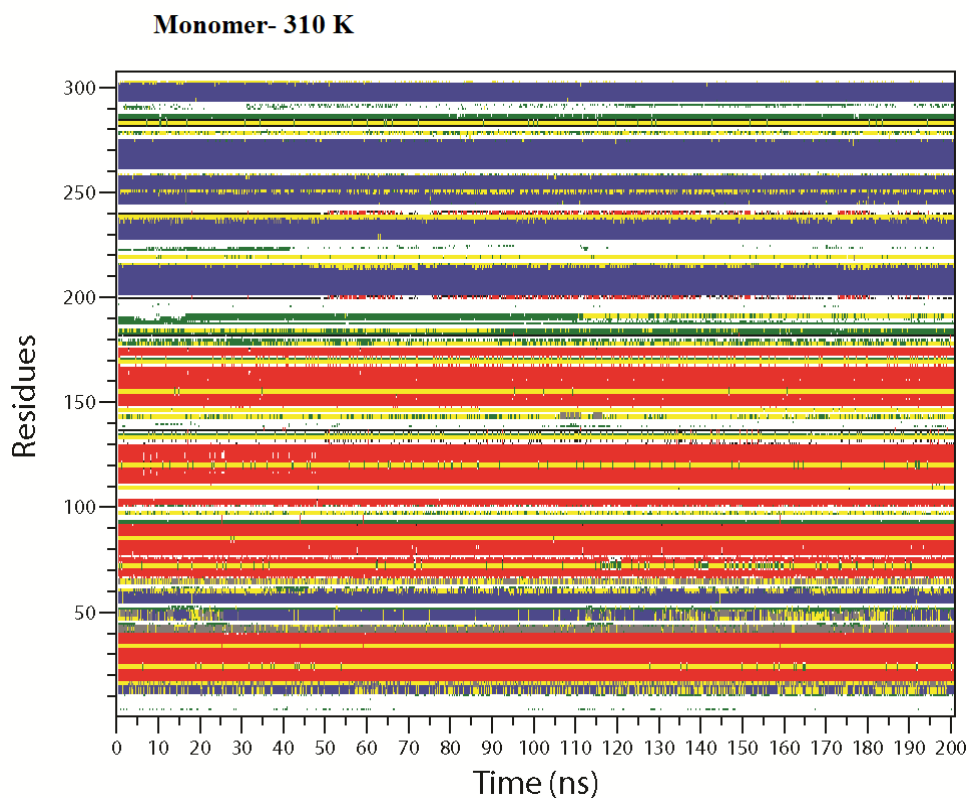
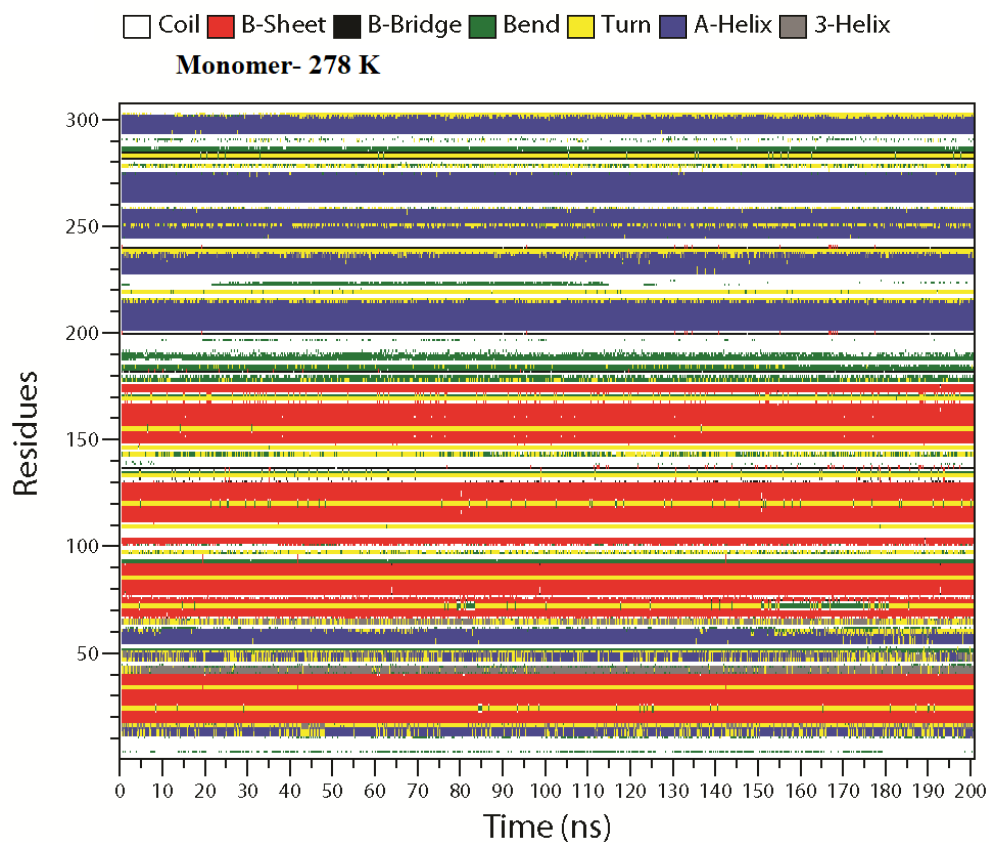
c)



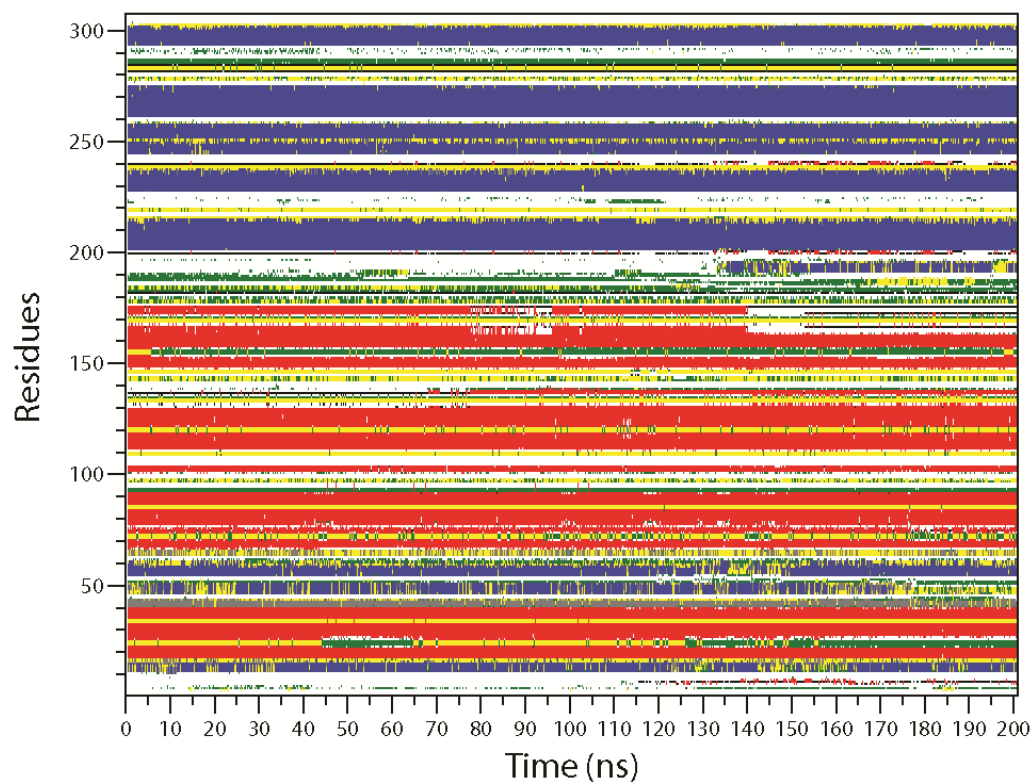
Appendix XXV: Average occupancy of secondary structure components obtained for 2 sets of simulations each for monomer and dimer- chains using SPC/E water model

Temperature	Elements	Percentage (%)		
		Monomer	Dimer chain-A	Dimer chain-B
278 K	Coil	0.26, 0.26	0.27, 0.27	0.27, 0.26
	β -sheet	0.26, 0.27	0.26, 0.26	0.26, 0.27
	β -bridge	0.02, 0.02	0.02, 0.02	0.02, 0.02
	Bend	0.07, 0.07	0.06, 0.07	0.06, 0.05
	Turn	0.13, 0.13	0.13, 0.13	0.13, 0.14
	α -helix	0.23, 0.23	0.23, 0.24	0.24, 0.24
	3-helix	0.02, 0.02	0.02, 0.02	0.02, 0.02
310 K	Coil	0.26, 0.27	0.27, 0.27	0.27, 0.28
	β -sheet	0.27, 0.25	0.26, 0.26	0.26, 0.25
	β -bridge	0.02, 0.02	0.02, 0.02	0.02, 0.02
	Bend	0.06, 0.08	0.07, 0.07	0.06, 0.07
	Turn	0.14, 0.13	0.13, 0.13	0.13, 0.13
	α -helix	0.23, 0.23	0.23, 0.24	0.23, 0.23
	3-helix	0.02, 0.02	0.02, 0.02	0.02, 0.02
348 K	Coil	0.27, 0.27	0.27, 0.27	0.27, 0.27
	β -sheet	0.25, 0.26	0.25, 0.25	0.25, 0.26
	β -bridge	0.02, 0.02	0.02, 0.02	0.02, 0.02
	Bend	0.08, 0.08	0.09, 0.08	0.08, 0.08
	Turn	0.13, 0.12	0.13, 0.13	0.13, 0.13
	α -helix	0.23, 0.23	0.23, 0.22	0.23, 0.23
	3-helix	0.02, 0.02	0.02, 0.02	0.02, 0.02
383 K	Coil	0.24, 0.26	0.26, 0.27	0.28, 0.27
	β -sheet	0.26, 0.25	0.26, 0.25	0.25, 0.25
	β -bridge	0.02, 0.02	0.02, 0.02	0.02, 0.02
	Bend	0.10, 0.10	0.08, 0.08	0.09, 0.09
	Turn	0.13, 0.13	0.14, 0.13	0.13, 0.13
	α -helix	0.22, 0.22	0.22, 0.23	0.23, 0.23
	3-helix	0.02, 0.03	0.02, 0.02	0.01, 0.02

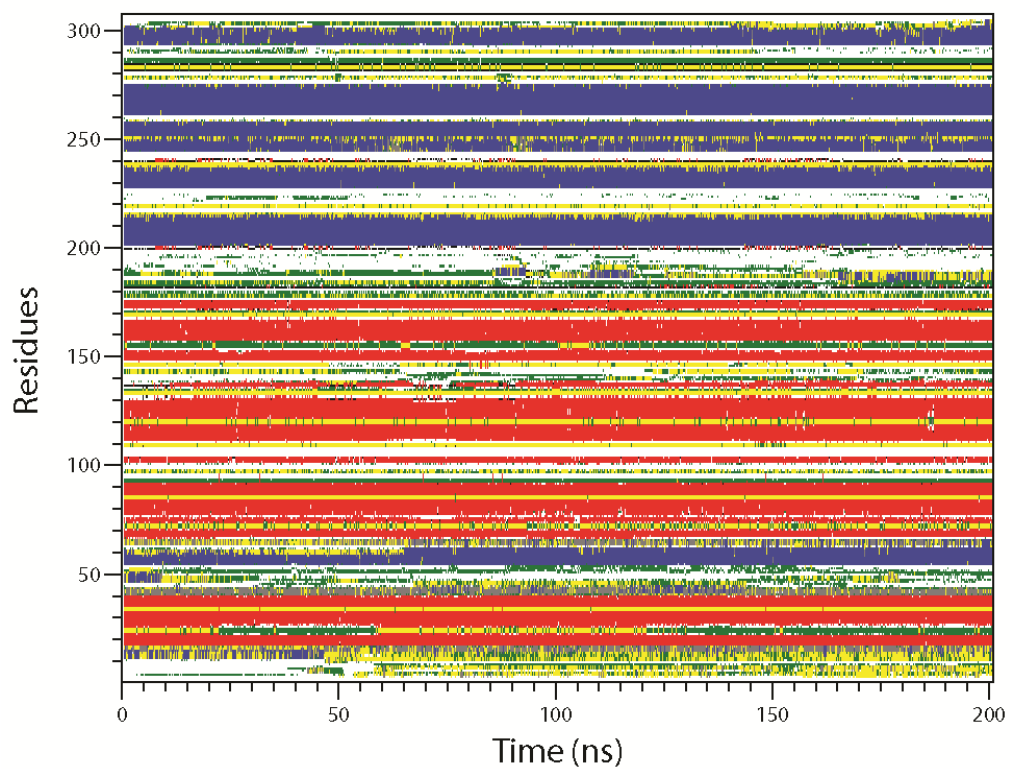
Appendix XXVI: Secondary structure elements for the protease calculated by DSSP for the monomer and dimer (chain A) in SPC/E water model



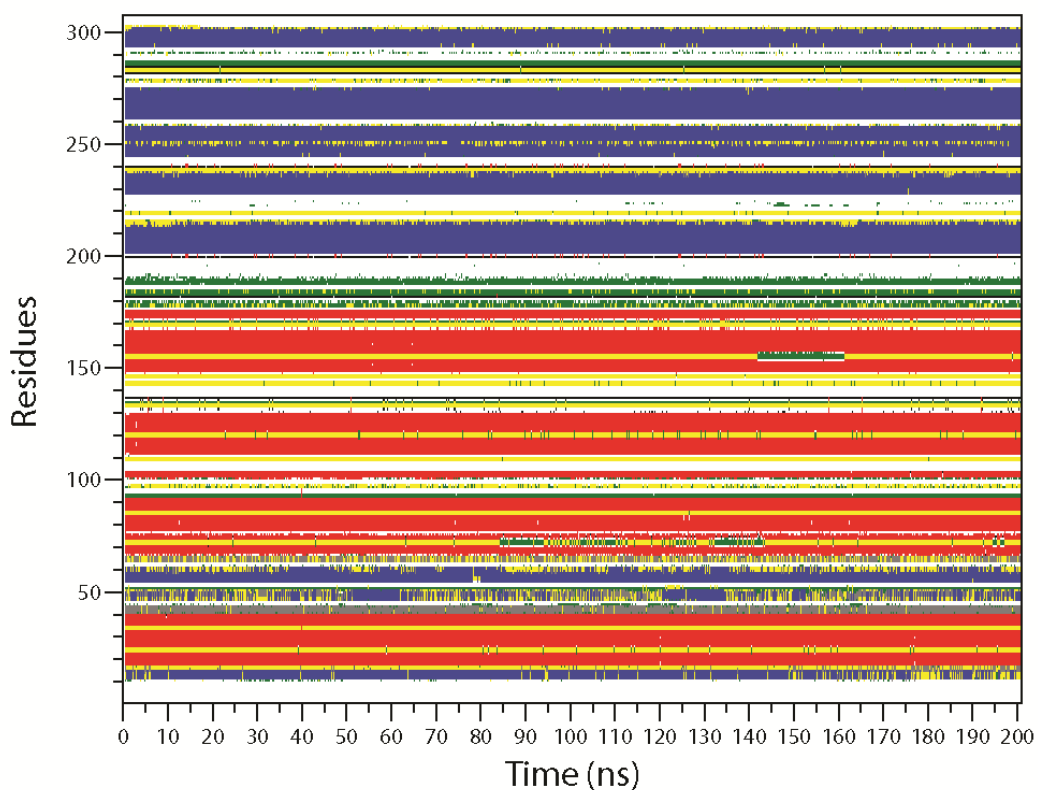
Monomer- 348 K



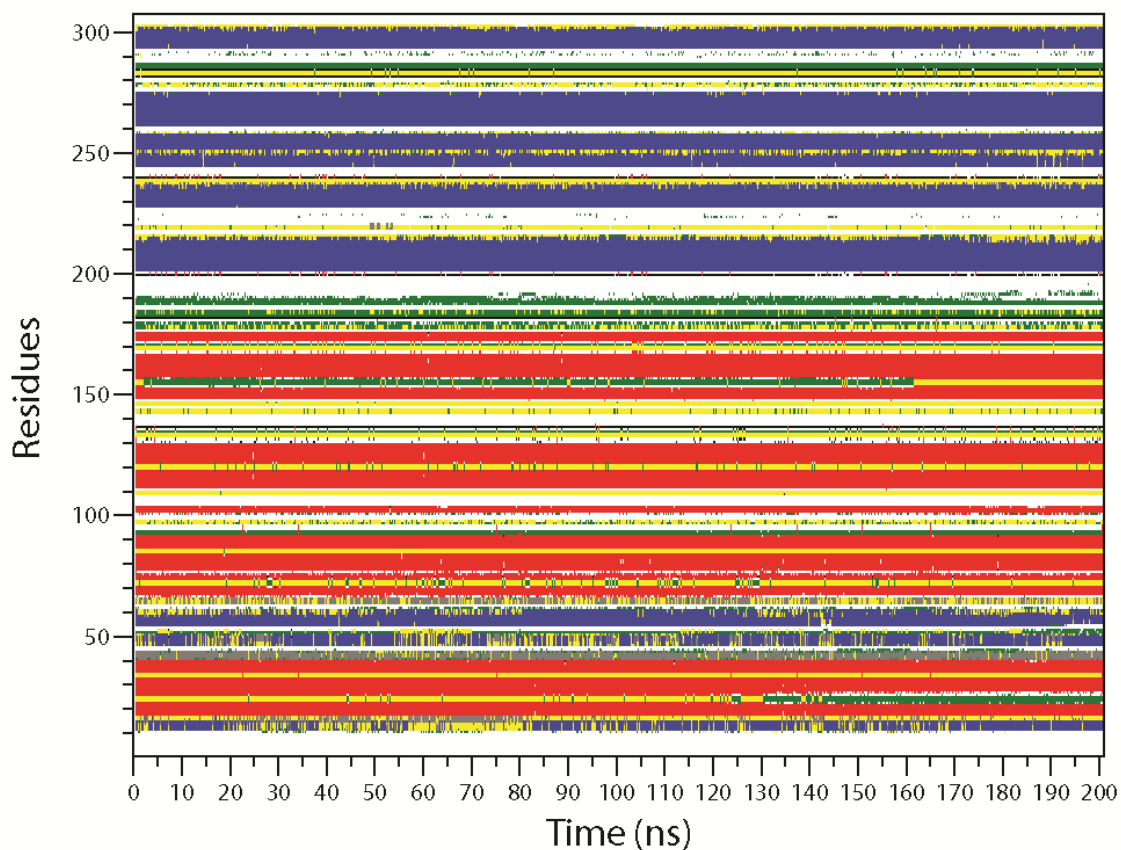
Monomer- 383 K



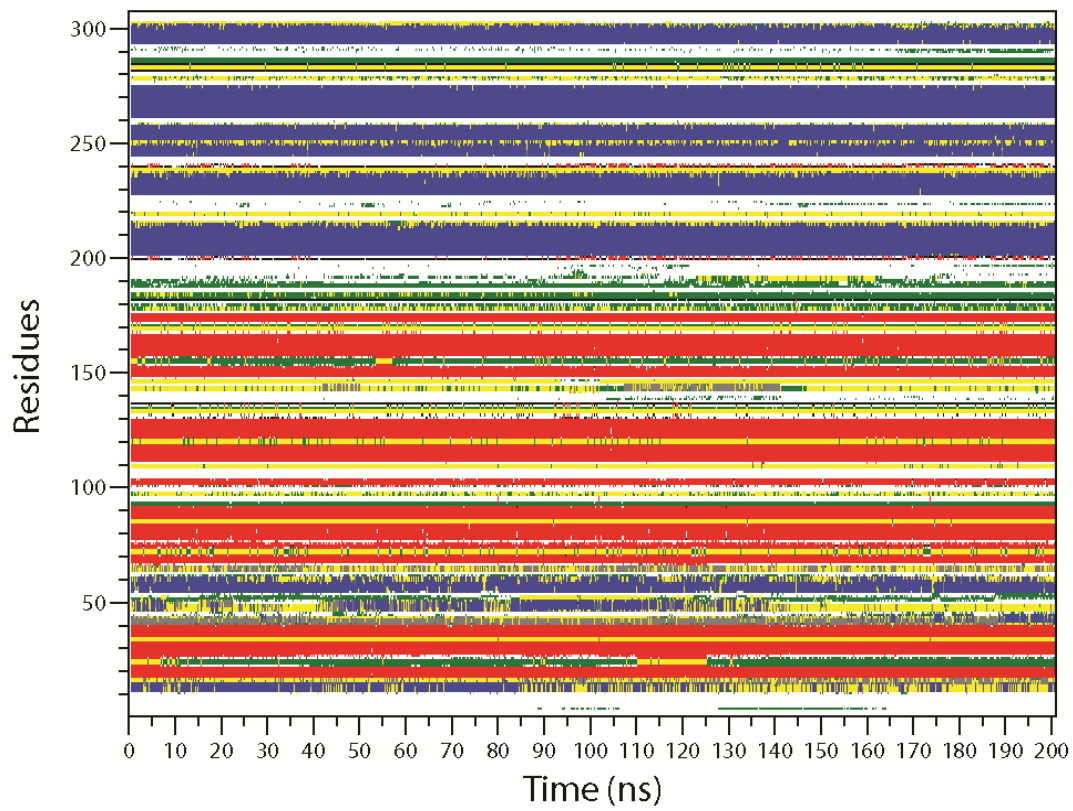
Dimer (Chain A)- 278 K



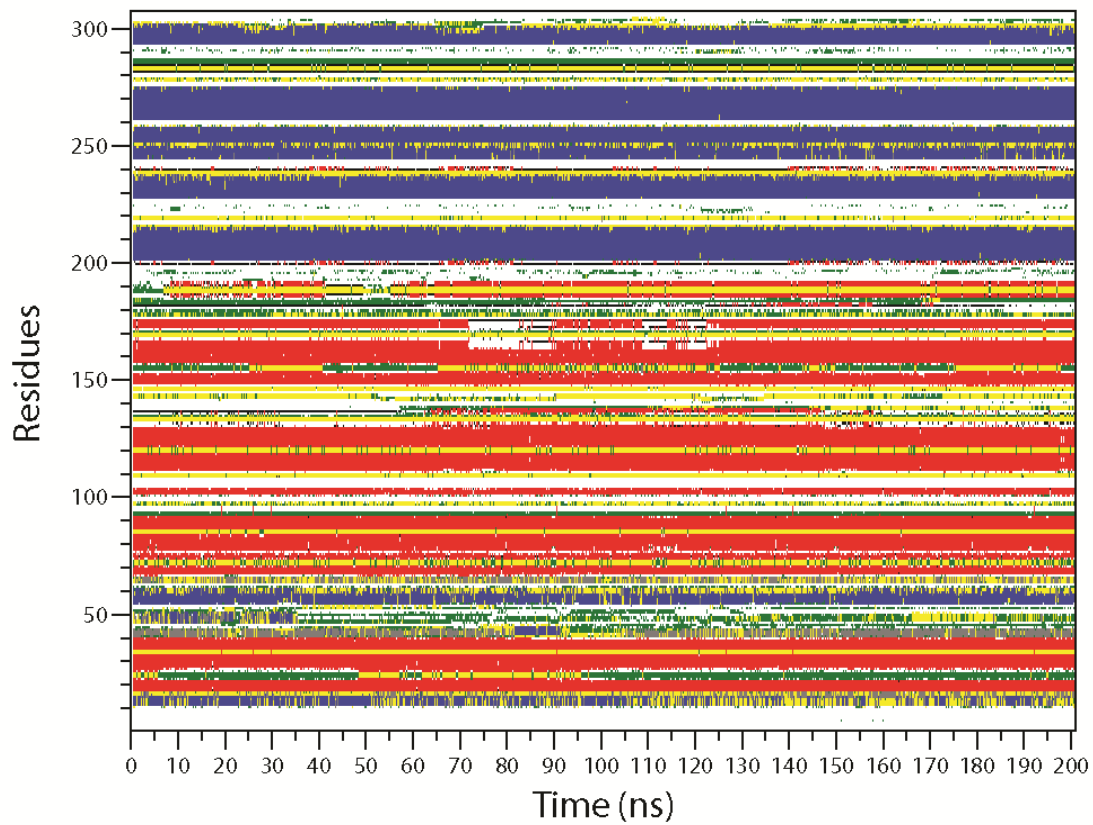
Dimer (Chain A)- 310 K



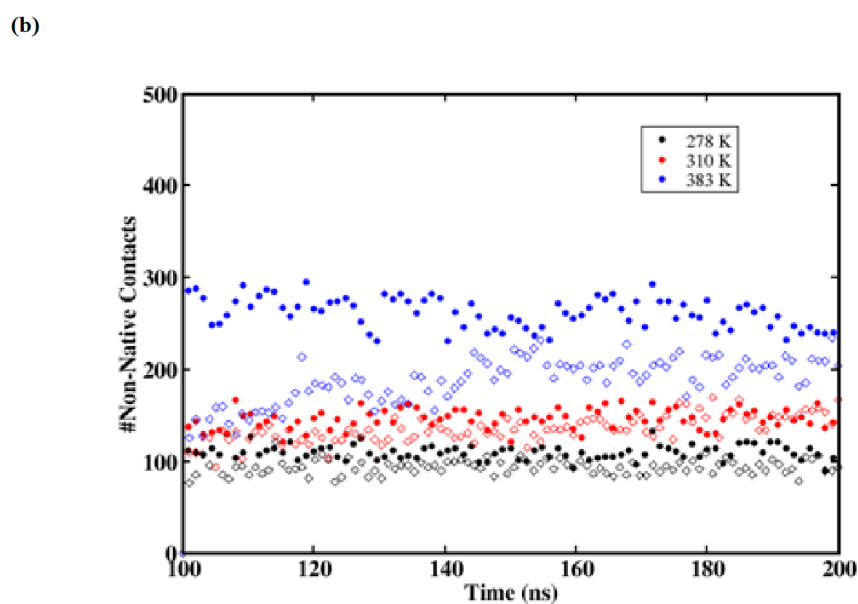
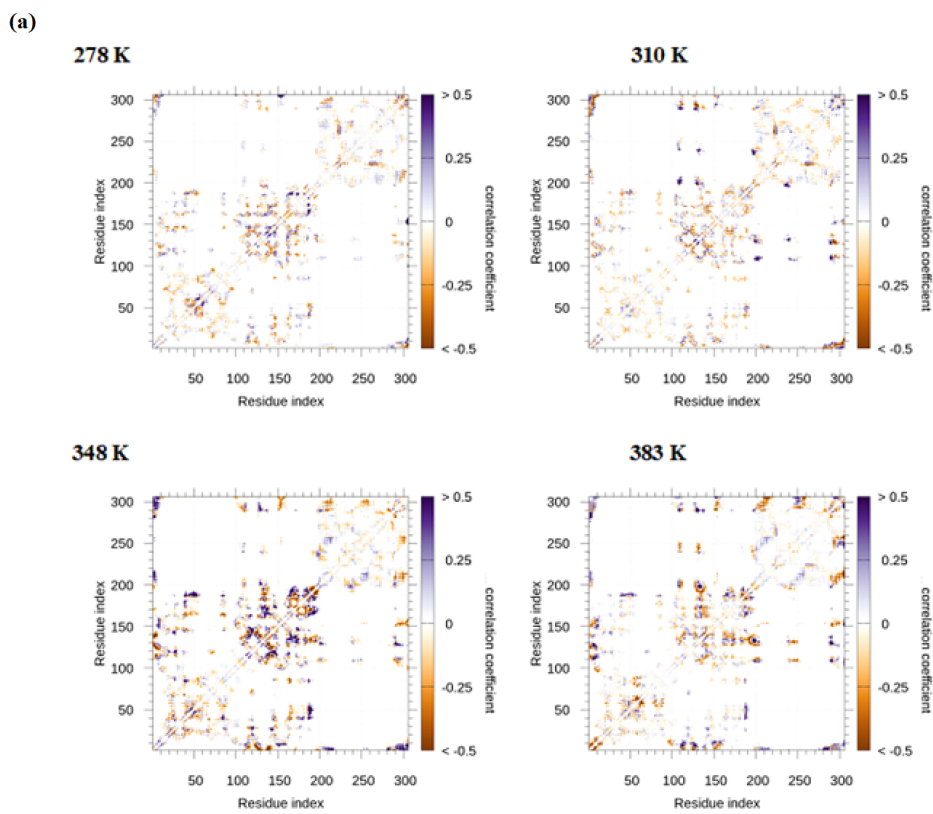
Dimer (Chain A)- 348 K



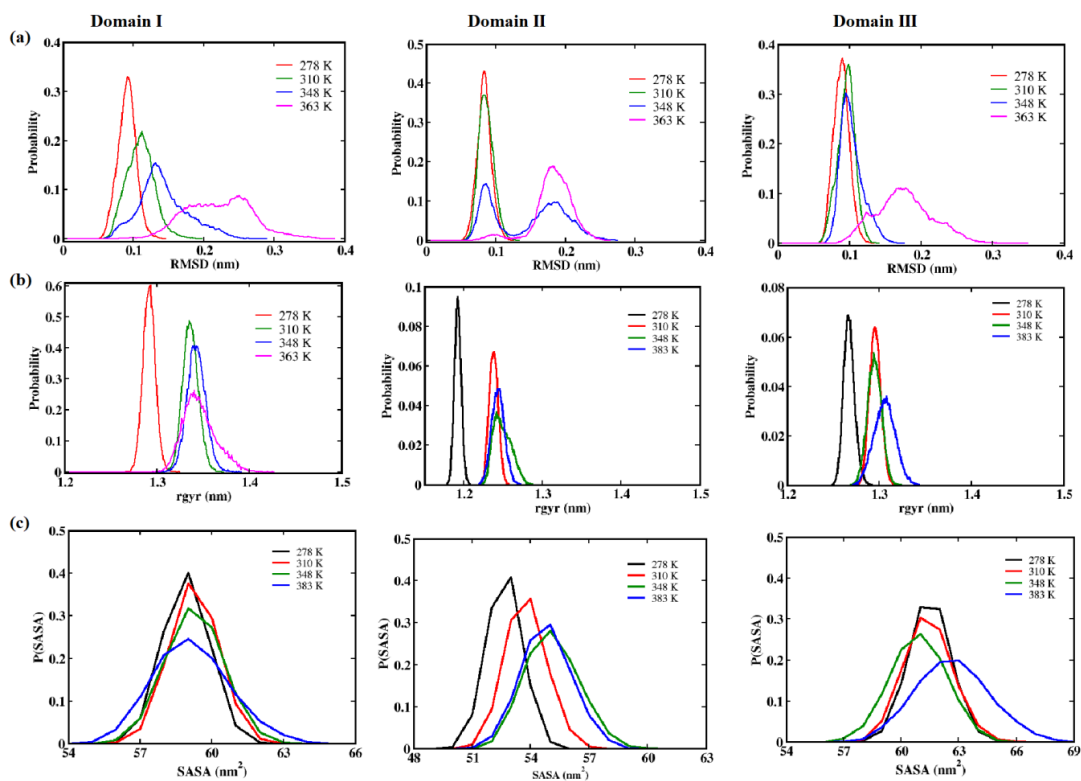
Dimer (Chain A)- 383 K



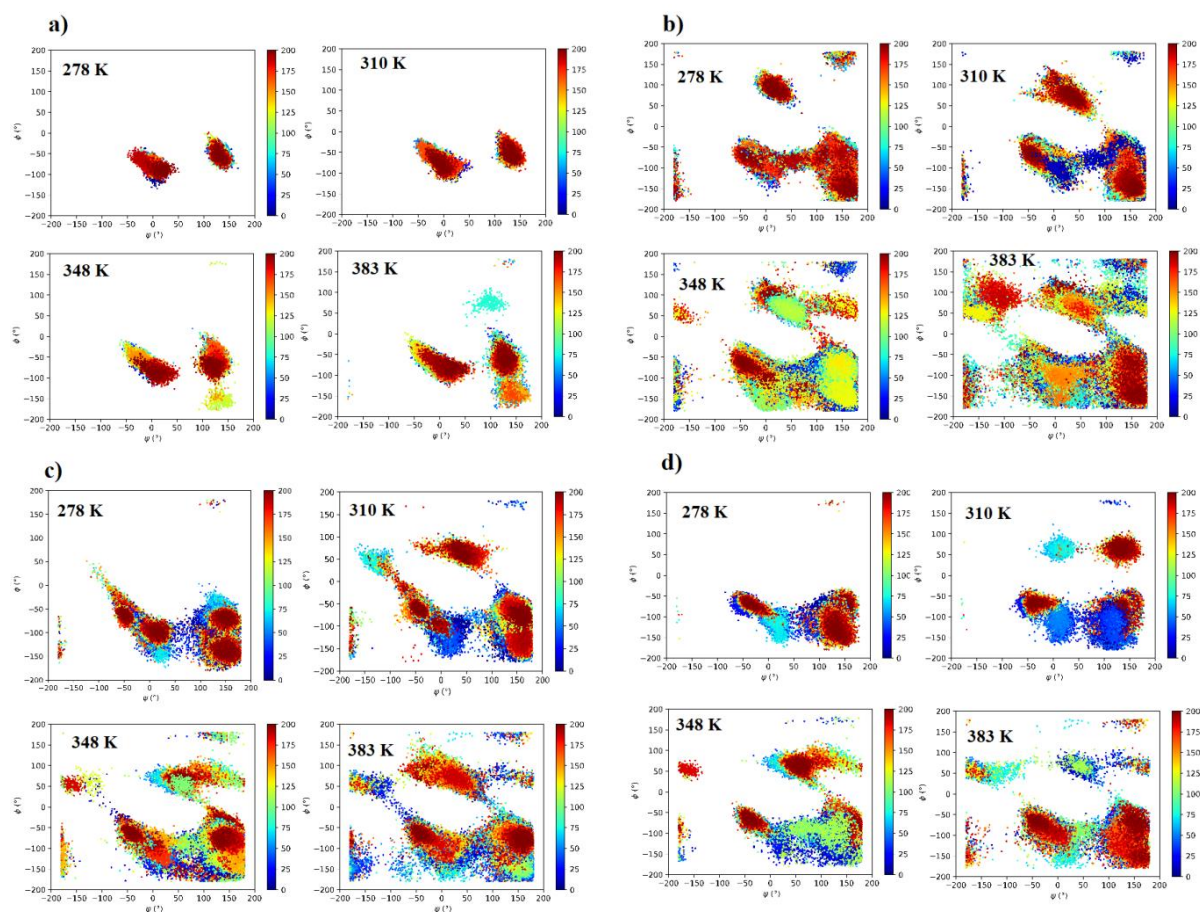
Appendix XXVII: (a) Correlation of inter-residue distance at different temperature for monomer. The blue colour shows the residue contact ruptures and the orange colour shows the residue contact formation (b) Change in the number of non-native contacts in monomer (circle) and dimer-chain A (diamond) as a function of time for last 100 ns MD trajectory



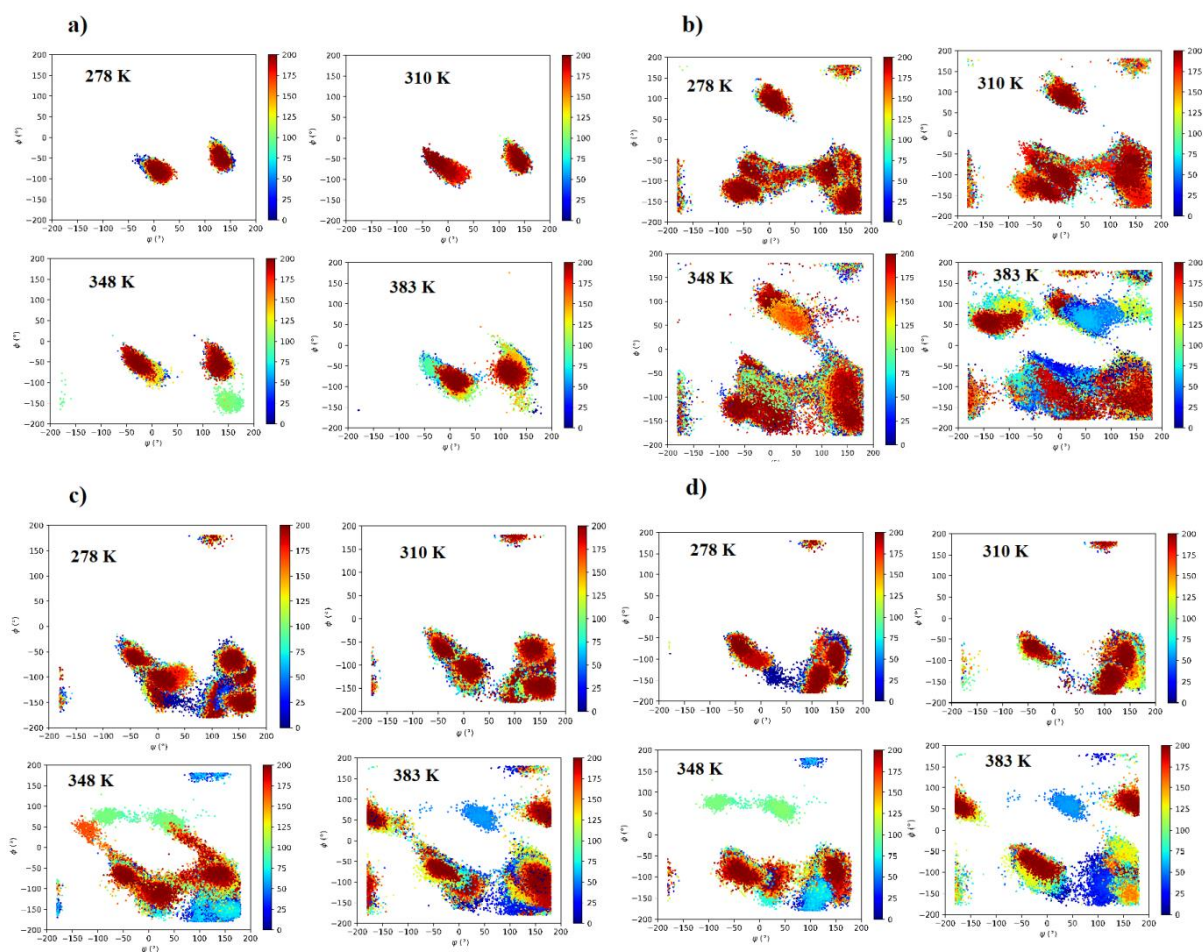
Appendix XXVIII: Probability distribution of (a) C_{α} -RMSD (b) rGyr and (c) SASA for domain I, II, III for monomer protease



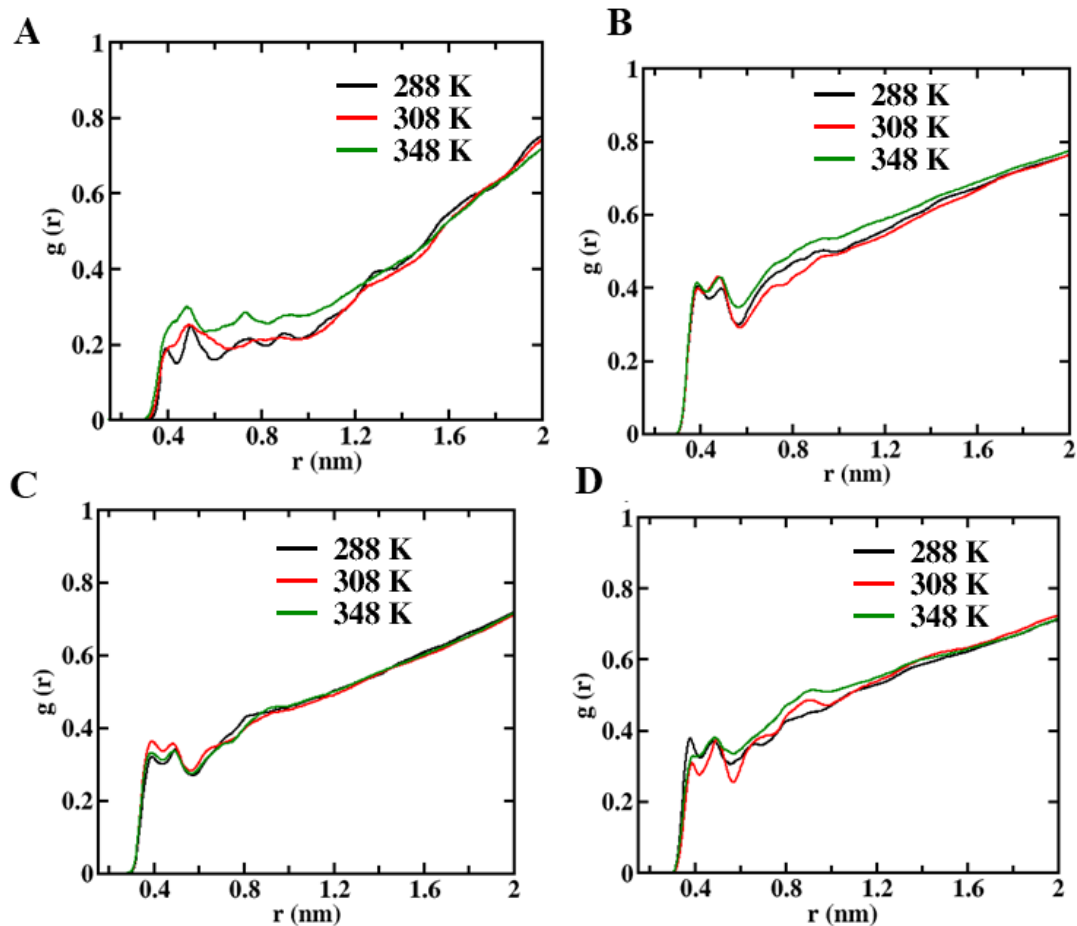
Appendix XXIX: Dihedral angle (ϕ , ψ) transitions of amino acid residues at various sites of monomer during the course of simulations **(a)** Catalytic dyad **(b)** Substrate binding site **(c)** Dimerization site **(d)** Allosteric site. The colour scale represents the simulation time in ns



Appendix XXX: Dihedral angle (ϕ , ψ) transitions of amino acid residues at various sites of dimer-chain A during the course of simulations in SPC/E water model **(a)** Catalytic dyad **(b)** Substrate binding site **(c)** Dimerization site **(d)** Allosteric site. The colour scale represents the simulation time in ns

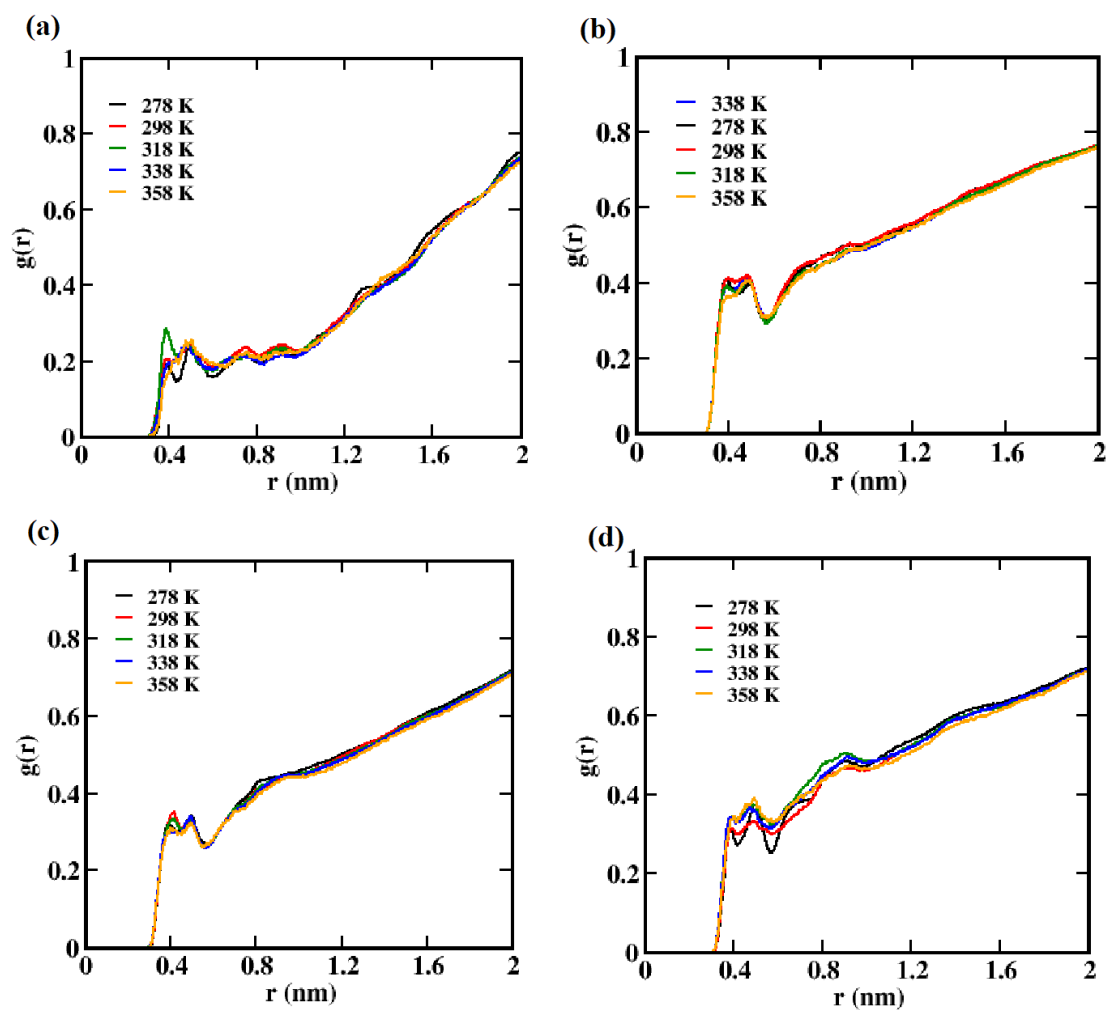


Appendix XXXI: Radial distribution function of C_{α} -Ow of amino acid residues at various binding site of protease at different temperatures in SPC/E water model. (a) Catalytic dyad (b) Substrate binding site (c) Dimerization site (d) Allosteric site

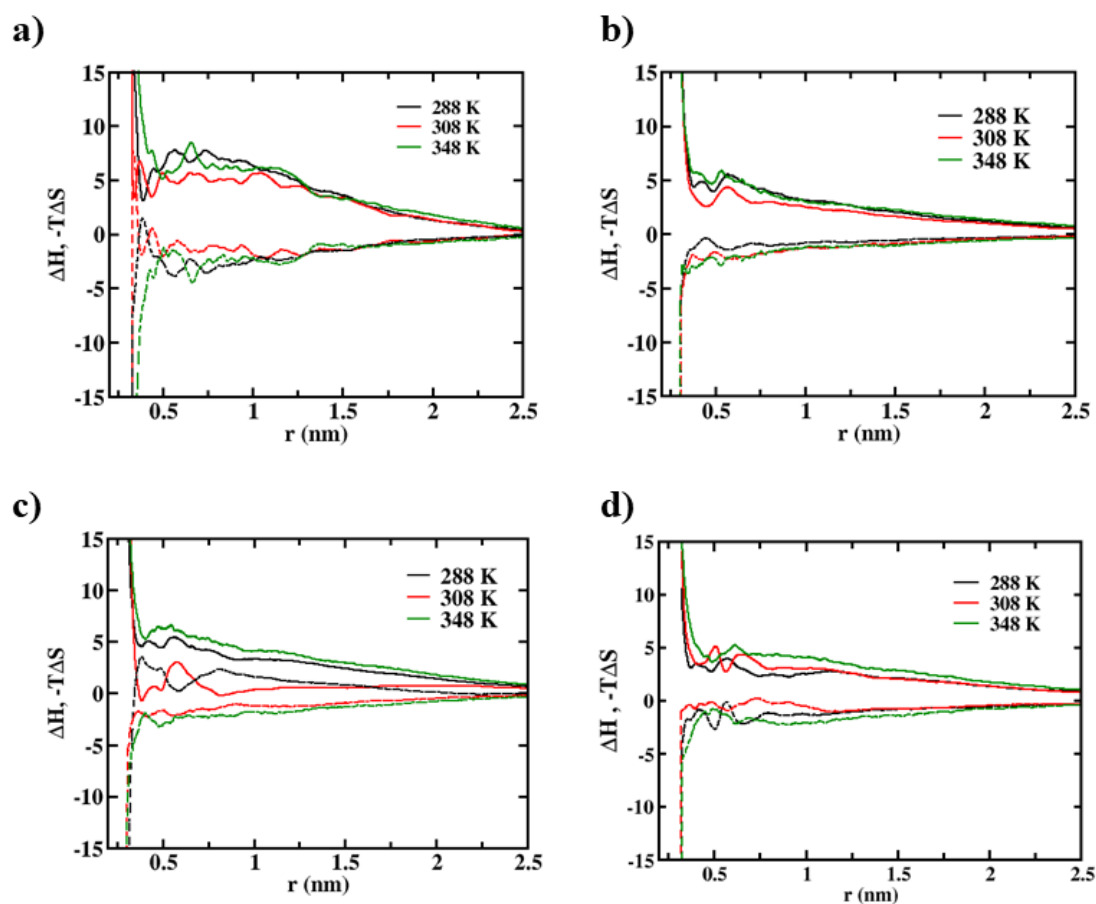


Appendix XXXII: Radial distribution function of C_{α} -Ow of amino acid residues at various binding site of protease at 5 additional temperatures in SPC/E water model.

(a) Catalytic dyad **(b)** Substrate binding site **(c)** Dimerization site **(d)** Allosteric site

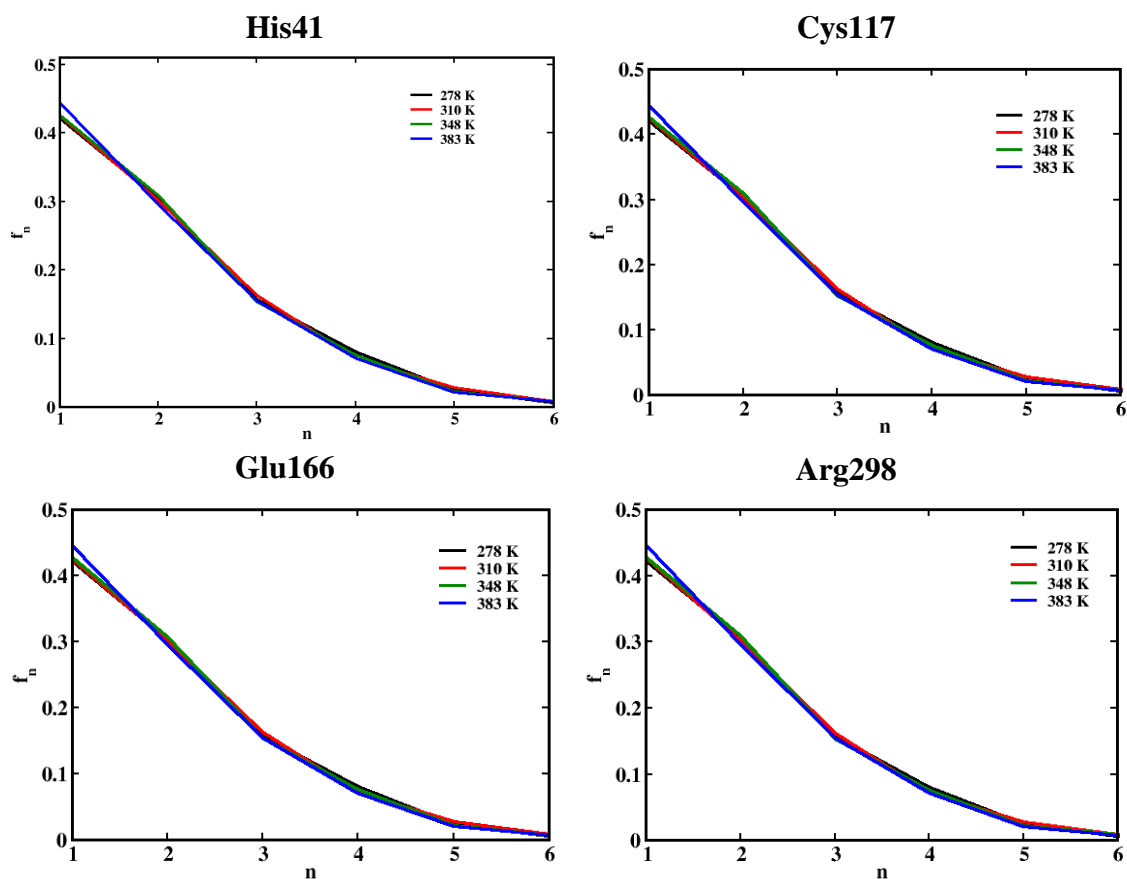


Appendix XXXIII: Enthalpy (solid lines) and entropy (dashed lines) contribution to free energy at different temperatures for monomer protease in SPC/E water model. **(a)** Catalytic dyad **(b)** Substrate binding site **(c)** Dimerization site **(d)** Allosteric site

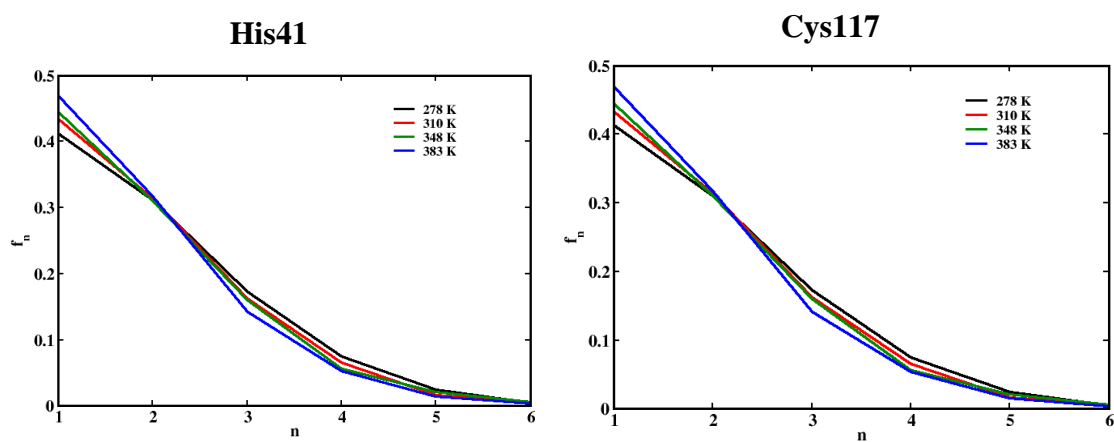


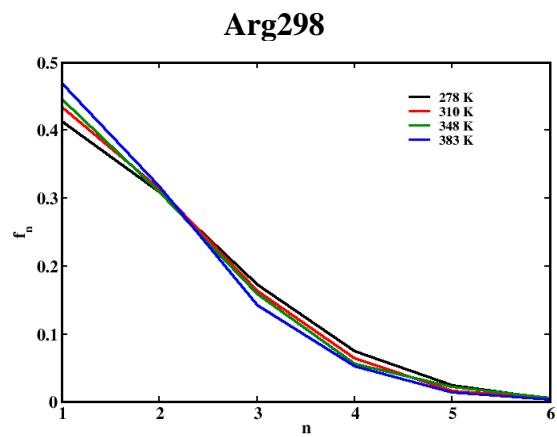
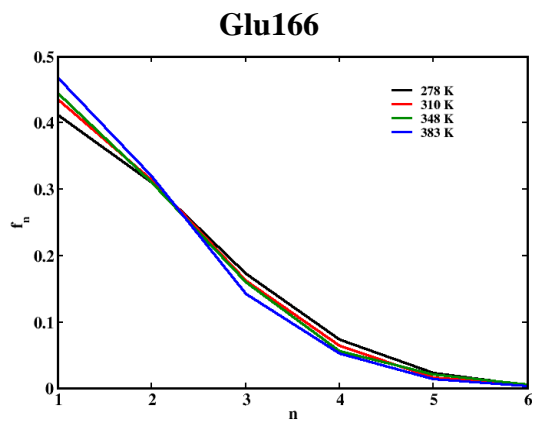
Appendix XXXIV: Fraction of water molecules having n number of hydrogen bonds within a distance of 6.0 \AA from C_α of amino acid residues of (a) M^{Pro} monomer and (b) dimer-Chain A for mTIP3P water model at different temperatures

(a) Monomer



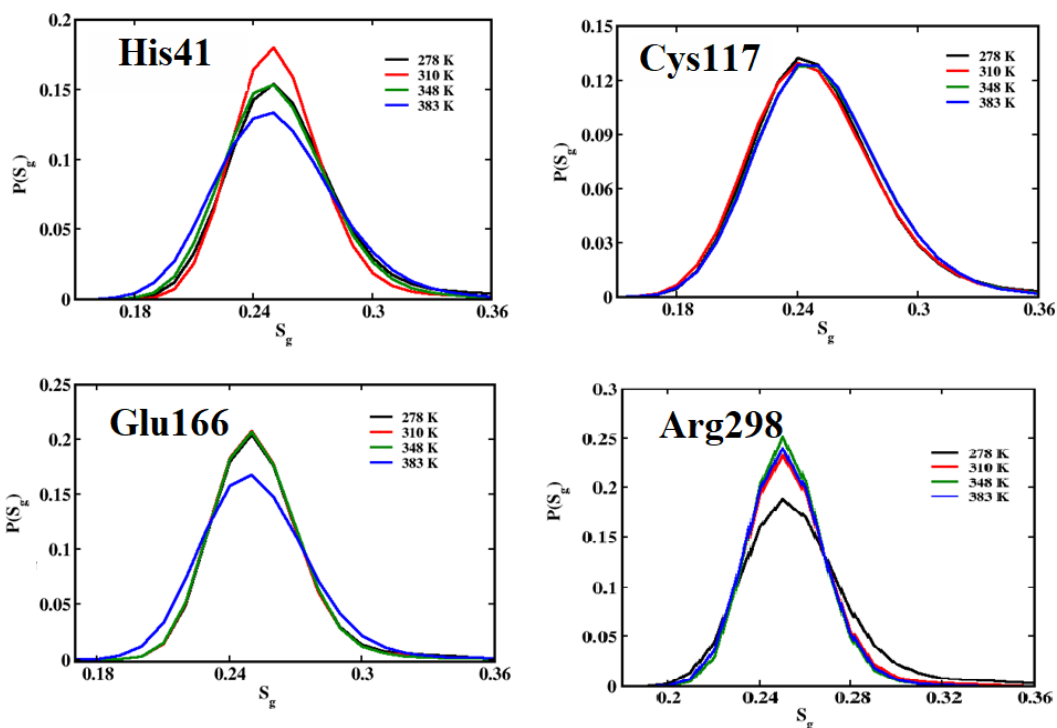
(b) Dimer



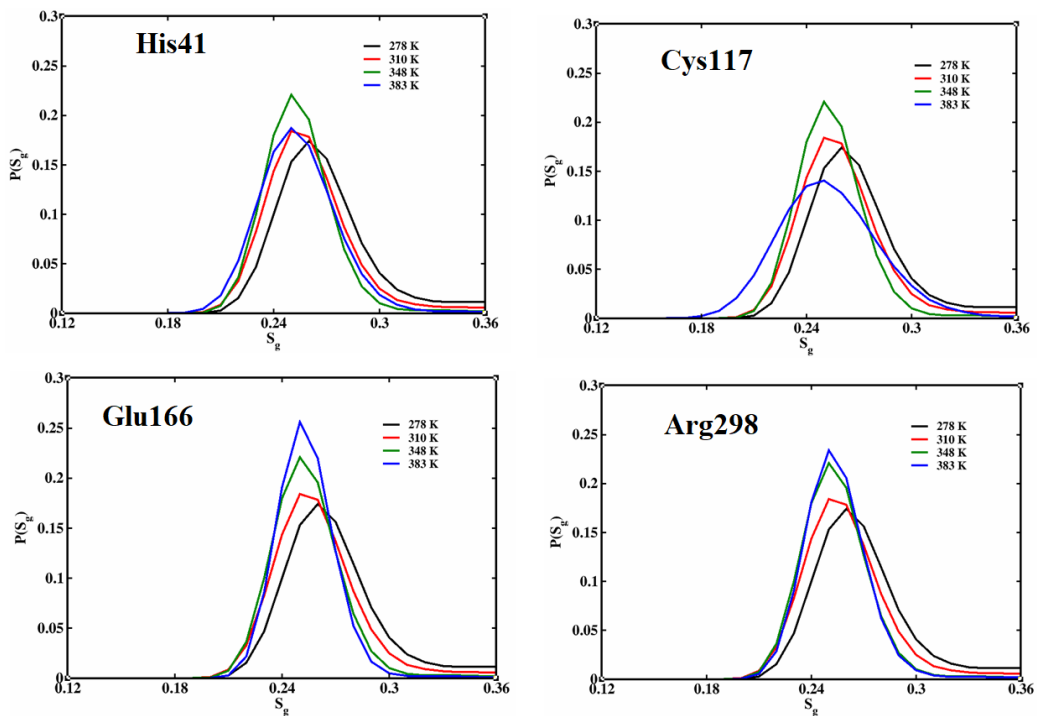


Appendix XXXV: Orientational tetrahedral order parameter of water molecules within 6.0 Å from C_α of key amino acid residues at various binding site of (a) monomer and (b) dimer-chain A in mTIP3P water model at different temperatures

a) Monomer



b) Dimer



REFERENCES

- Abdulrahman, H. L., Uzairu, A., and Uba, S. (2021). "QSAR, Ligand Based Design and Pharmacokinetic Studies of Parviflorons Derivatives as Anti-Breast Cancer Drug Compounds Against MCF-7 Cell Line." *Chemistry Africa*, 4(1), 175–187.
- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015). "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers." *SoftwareX*, 1–2, 19–25.
- Abramyan, T. M., Snyder, J. A., Thyparambil, A. A., Stuart, S. J., and Latour, R. A. (2016). "Cluster analysis of molecular simulation trajectories for systems where both conformation and orientation of the sampled states are important." *J. Comput. Chem.*, 37(21), 1973–1982.
- Achutha, A. S., Pushpa, V. L., and Suchitra, S. (2020). "Theoretical Insights into the Anti-SARS-CoV-2 Activity of Chloroquine and Its Analogs and In Silico Screening of Main Protease Inhibitors." *J. Proteome Res.*, 19(11), 4706–4717.
- Adhikari, N., Banerjee, S., Baidya, S. K., Ghosh, B., and Jha, T. (2022). "Ligand-based quantitative structural assessments of SARS-CoV-2 3CLpro inhibitors: An analysis in light of structure-based multi-molecular modeling evidences." *J. Mol. Struct.*, 1251, 132041.
- Ahmad, T., Khan, M., Haroon, Musa, T. H., Nasir, S., Hui, J., Bonilla-Aldana, D. K., and Rodriguez-Morales, A. J. (2020). "COVID-19: Zoonotic aspects." *Travel Med. Infect. Dis.*, 101607.
- Ajmani, S., Jadhav, K., and Kulkarni, S. A. (2009). "Group-Based QSAR (G-QSAR): Mitigating Interpretation Challenges in QSAR." *QSAR Comb. Sci.*, 28(1), 36–51.
- Aktaş, A., Tüzün, B., Aslan, R., Sayin, K., and Ataseven, H. (2021). "New anti-viral drugs for the treatment of COVID-19 instead of favipiravir." *J. Biomo. Struct. Dyn.*, 39(18), 7263–7273.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). “Protein Function.” *Molecular Biology of the Cell*. 4th edition.

Alexpandi, R., De Mesquita, J. F., Pandian, S. K., and Ravi, A. V. (2020). “Quinolines-Based SARS-CoV-2 3CLpro and RdRp Inhibitors and Spike-RBD-ACE2 Inhibitor for Drug-Repurposing Against COVID-19: An in silico Analysis.” *Front. Microbiol.*, 11, 1796.

Almeida, J., Volpini, J., Poiani, J., and Silva, C. (2014). “Ligand-Based Drug Design of Novel MARK-3 Inhibitors in Cancer.” *Curr. Bioact. Compd.*, 10(2), 112–123.

Alves, C. N., Martí, S., Castillo, R., Andrés, J., Moliner, V., Tuñón, I., and Silla, E. (2007). “Calculation of binding energy using BLYP/MM for the HIV-1 integrase complexed with the S-1360 and two analogues.” *Bioorg. Med. Chem.*, 15(11), 3818–3824.

Amé, J.-C., Rolli, V., Schreiber, V., Niedergang, C., Apiou, F., Decker, P., Muller, S., Höger, T., Murcia, J. M., and Murcia, G. de. (1999). “PARP-2, A Novel Mammalian DNA Damage-dependent Poly(ADP-ribose) Polymerase.” *J. Biol. Chem.*, 274(25), 17860–17868.

Amé, J.-C., Spenlehauer, C., and Murcia, G. de. (2004). “The PARP superfamily: Review articles.” *Bioessays*, 26(8), 882–893.

Anandakrishnan, R., Izadi, S., and Onufriev, A. V. (2019). “Why Computed Protein Folding Landscapes Are Sensitive to the Water Model.” *J. Chem. Theory Comput.*, 15(1), 625–636.

Anighoro, A. (2020). “Underappreciated Chemical Interactions in Protein–Ligand Complexes.” *Quantum Mechanics in Drug Discovery*, Methods in Molecular Biology, A. Heifetz, ed., New York, NY: Springer US, 75–86.

Antunes, D., Jorge, N. A. N., Garcia de Souza Costa, M., Passetti, F., and Caffarena, E. R. (2019). “Unraveling RNA dynamical behavior of TPP riboswitches: a comparison between *Escherichia coli* and *Arabidopsis thaliana*.” *Sci. Rep.*, 9(1), 4197.

- Aoyagi-Scharber, M., Gardberg, A. S., Yip, B. K., Wang, B., Shen, Y., and Fitzpatrick, P. A. (2014). “Structural basis for the inhibition of poly(ADP-ribose) polymerases 1 and 2 by BMN 673, a potent inhibitor derived from dihydropyridophthalazinone.” *Acta Crystallogr. F Struct. Biol. Commun.*, 70(9), 1143–1149.
- Ashworth, A. (2008). “A Synthetic Lethal Therapeutic Approach: Poly(ADP) Ribose Polymerase Inhibitors for the Treatment of Cancers Deficient in DNA Double-Strand Break Repair.” *J. Clin. Oncol.*, 26(22), 3785–3790.
- Awad, I. E., Abu-Saleh, A. A.-A. A., Sharma, S., Yadav, A., and Poirier, R. A. (2022). “High-throughput virtual screening of drug databanks for potential inhibitors of SARS-CoV-2 spike glycoprotein.” *J. Biomol. Struct. Dyn.*, 40(5), 2099–2112.
- Baby, K., Maity, S., Mehta, C. H., Suresh, A., Nayak, U. Y., and Nayak, Y. (2020). “Targeting SARS-CoV-2 Main Protease: A Computational Drug Repurposing Study.” *Arch. Med. Res.*, S0188440920309577.
- Bai, P., Canto, C., Brunyánszki, A., Huber, A., Szántó, M., Cen, Y., Yamamoto, H., Houten, S. M., Kiss, B., Oudart, H., Gergely, P., Menissier-de Murcia, J., Schreiber, V., Sauve, A. A., and Auwerx, J. (2011). “PARP-2 Regulates SIRT1 Expression and Whole-Body Energy Expenditure.” *Cell Metab.*, 13(4), 450–460.
- Balajee, R., Srinivasadesikan, V., Sakthivadivel, M., and Gunasekaran, P. (2016). “*In Silico* Screening, Alanine Mutation, and DFT Approaches for Identification of NS2B/NS3 Protease Inhibitors.” *Biochem. Res. Int.*, 2016, 1–13.
- Barisione, G., Baroffio, M., Crimi, E., and Brusasco, V. (2010). “Beta-Adrenergic Agonists.” *Pharmaceuticals*, 3(4), 1016–1044.
- Bastian, M., Heymann, S., and Jacomy, M. (2009). “Gephi: An Open Source Software for Exploring and Manipulating Networks.”
- Becke, A. D. (1988). “Density-functional exchange-energy approximation with correct asymptotic behavior.” *Phys. Rev. A*, 38(6), 3098–3100.

- Becke, A. D. (1993). "Density-functional thermochemistry. III. The role of exact exchange." *J. Chem. Phys.*, 98(7), 5648–5652.
- Beeman, D. (1976). "Some multistep methods for use in molecular dynamics calculations." *J. Comp. Phys.*, 20(2), 130–139.
- Bharathi, A. C., Yadav, P. K., and Syed Ibrahim, B. (2016). "Sequence diversity and ligand-induced structural rearrangements of viper hyaluronidase." *Mol. Biosyst.*, 12(4), 1128–1138.
- Biswas, S., Chakraborty, D., and Mallik, B. S. (2018). "Interstitial Voids and Resultant Density of Liquid Water: A First-Principles Molecular Dynamics Study." *ACS Omega*, 3(2), 2010–2017.
- Blinder, S. M. (1965). "Basic Concepts of Self-Consistent-Field Theory." *Am. J. Phys.*, 33(6), 431–443.
- Bolelli, K., Ertan-Bolelli, T., Unsalan, O., and Altunayar-Unsalan, C. (2020). "Fenoterol and dobutamine as COVID-19 main protease inhibitors: A virtual screening study." *J. Mol. Struct.*, 129449.
- Brendler, T., Al-Harrasi, A., Bauer, R., Gafner, S., Hardy, M. L., Heinrich, M., Hosseinzadeh, H., Izzo, A. A., Michaelis, M., Nassiri-Asl, M., Panossian, A., Wasser, S. P., and Williamson, E. M. (2021). "Botanical drugs and supplements affecting the immune response in the time of COVID -19: Implications for research and clinical practice." *Phytother. Res.*, 35(6), 3013–3031.
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983). "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations." *J. Comput. Chem.*, 4(2), 187–217.
- Brylinski, M. (2018). "Aromatic interactions at the ligand-protein interface: Implications for the development of docking scoring functions." *Chem. Biol. Drug. Des.*, 91(2), 380–390.

- Bussi, G., Donadio, D., and Parrinello, M. (2007). “Canonical sampling through velocity rescaling.” *J. Chem. Phys.*, 126(1), 014101.
- Caldararu, O., Kumar, R., Oksanen, E., Logan, D. T., and Ryde, U. (2019). “Are crystallographic *B* -factors suitable for calculating protein conformational entropy?” *Phys. Chem. Chem. Phys.*, 21(33), 18149–18160.
- Carugo, O., and Pongor, S. (2008). “A normalized root-mean-square distance for comparing protein three-dimensional structures.” *Protein Sci.*, 10(7), 1470–1473.
- Chau, P.-L., and Hardwick, A. J. (1998). “A new order parameter for tetrahedral configurations.” *Mol. Phys.*, 93(3), 511–518.
- Chen, H., Wei, P., Huang, C., Tan, L., Liu, Y., and Lai, L. (2006). “Only One Protomer Is Active in the Dimer of SARS 3C-like Proteinase.” *J. Biol. Chem.*, 281(20), 13894–13898.
- Chen, X., Huan, X., Liu, Q., Wang, Y., He, Q., Tan, C., Chen, Y., Ding, J., Xu, Y., Miao, Z., and Yang, C. (2018). “Design and synthesis of 2-(4,5,6,7-tetrahydrothienopyridin-2-yl)-benzoimidazole carboxamides as novel orally efficacious Poly(ADP-ribose)polymerase (PARP) inhibitors.” *Eur. J. Med. Chem.*, 145, 389–403.
- Chew, W. C., Liu, A. Y., Salazar-Lazaro, C., Na, D.-Y., and Sha, W. E. I. (2019). “Hamilton Equations, Commutator, and Energy Conservation.” *Quantum Rep.*, 1(2), 295–303.
- Chirico, N., and Gramatica, P. (2012). “Real External Predictivity of QSAR Models. Part 2. New Intercomparable Thresholds for Different Validation Criteria and the Need for Scatter Plot Inspection.” *J. Chem. Inf. Model.*, 52(8), 2044–2058.
- Chojnacka, K., Witek-Krowiak, A., Skrzypczak, D., Mikula, K., and Młynarz, P. (2020). “Phytochemicals containing biologically active polyphenols as an effective agent against Covid-19-inducing coronavirus.” *J. Funct. Foods.*, 73, 104146.

Cholko, T., Chen, W., Tang, Z., and Chang, C. A. (2018). “A molecular dynamics investigation of CDK8/CycC and ligand binding: conformational flexibility and implication in drug discovery.” *J. Comput. Aided. Mol. Des.*, 32(6), 671–685.

Choudhary, S., Malik, Y. S., and Tomar, S. (2020). “Identification of SARS-CoV-2 Cell Entry Inhibitors by Drug Repurposing Using in silico Structure-Based Virtual Screening Approach.” *Front. Immunol.*, 11, 1664.

Ciccia, A., and Elledge, S. J. (2010). “The DNA Damage Response: Making It Safe to Play with Knives.” *Mol. Cell*, 40(2), 179–204.

Cohen, A. J., Mori-Sánchez, P., and Yang, W. (2008). “Insights into current limitations of density functional theory.” *Science*, 321(5890), 792–794.

Congreve, M., Oswald, C., and Marshall, F. H. (2017). “Applying Structure-Based Drug Design Approaches to Allosteric Modulators of GPCRs.” *Trends Pharmacol. Sci.*, 38(9), 837–847.

Correa-Basurto, J., Bello, M., Rosales-Hernández, M. C., Hernández-Rodríguez, M., Nicolás-Vázquez, I., Rojo-Domínguez, A., Trujillo-Ferrara, J. G., Miranda, R., and Flores-Sandoval, C. A. (2014). “QSAR, docking, dynamic simulation and quantum mechanics studies to explore the recognition properties of cholinesterase binding sites.” *Chem. Biol. Interact.*, 209, 1–13.

Correa-Basurto, J., Ramos-Morales, F. R., Matus, M. H., Rosales-Hernández, M. C., Mancilla-Percino, T., Trujillo-Ferrara, J., and Ilizaliturri-Flores, I. (2012a). “Docking and DFT Studies to explore the Topoisomerase II ATP Pocket employing 3-Substituted 2,6-Piperazindiones for drug design.” *Mol. Simul.*, 38(13), 1072–1084.

Correa-Basurto, J., Ramos-Morales, F. R., Matus, M. H., Rosales-Hernández, M. C., Mancilla-Percino, T., Trujillo-Ferrara, J., and Ilizaliturri-Flores, I. (2012b). “Docking and DFT Studies to explore the Topoisomerase II ATP Pocket employing 3-Substituted 2,6-Piperazindiones for drug design.” *Mol. Simul.*, 38(13), 1072–1084.

- Costantino, G., Macchiarulo, A., Camaioni, E., and Pellicciari, R. (2001). "Modeling of Poly(ADP-ribose)polymerase (PARP) Inhibitors. Docking of Ligands and Quantitative Structure–Activity Relationship Analysis." *J. Med. Chem.*, 44(23), 3786–3794.
- Cui, D., Zhang, B. W., Matubayasi, N., and Levy, R. M. (2018). "The Role of Interfacial Water in Protein–Ligand Binding: Insights from the Indirect Solvent Mediated Potential of Mean Force." *J. Chem. Theory Comput.*, 14(2), 512–526.
- Cusinato, J., Cau, Y., Calvani, A. M., and Mori, M. (2021). "Repurposing drugs for the management of COVID-19." *Expert Opin. Ther. Pat.*, 31(4), 295–307.
- Daddam, J. R., Sreenivasulu, B., Peddanna, K., and Umamahesh, K. (2020). "Designing, docking and molecular dynamics simulation studies of novel cloperastine analogues as anti-allergic agents: homology modeling and active site prediction for the human histamine H1 receptor." *RSC Adv.*, 10(8), 4745–4754.
- Dalvi, V. H., and Rosky, P. J. (2010). "Molecular origins of fluorocarbon hydrophobicity." *Proc. Natl. Acad. Sci. USA*, 107(31), 13603–13607.
- Darke, P. L., Cole, J. L., Waxman, L., Hall, D. L., Sardana, M. K., and Kuo, L. C. (1996). "Active Human Cytomegalovirus Protease Is a Dimer." *J. Biol. Chem.*, 271(13), 7445–7449.
- Daura, X., Gademann, K., Jaun, B., Gunsteren, W. F. van, and Mark, A. E. (2004). "Peptide Folding: When Simulation Meets Experiment." *Angew. Chem. Int. Ed.*, 38(1-2), 236-240.
- Dawoud, N. T. A., El-Fakharany, E. M., Abdallah, A. E., El-Gendi, H., and Lotfy, D. R. (2022). "Synthesis, and docking studies of novel heterocycles incorporating the indazolylthiazole moiety as antimicrobial and anticancer agents." *Sci. Rep.*, 12(1), 3424.
- Dayalan, S., Gooneratne, N. D., Bevinakoppa, S., and Schroder, H. (2006). "Dihedral angle and secondary structure database of short amino acid fragments." *Bioinformatics*, 1(1), 78–80.

DeFrees, K., Kemp, M. T., ElHilali-Pollard, X., Zhang, X., Mohamed, A., Chen, Y., and Renslo, A. R. (2019). “An empirical study of amide–heteroarene π -stacking interactions using reversible inhibitors of a bacterial serine hydrolase.” *Org. Chem. Front.*, 6(11), 1749–1756.

Delaney, C. A., Wang, L. Z., Kyle, S., White, A. W., Calvert, A. H., Curtin, N. J., Durkacz, B. W., Hostomsky, Z., and Newell, D. R. (2000). “Potentiation of temozolomide and topotecan growth inhibition and cytotoxicity by novel poly(adenosine diphosphoribose) polymerase inhibitors in a panel of human tumor cell lines.” *Clin. Cancer Res.*, 6(7), 2860–2867.

Desiraju, G. R., and Steiner, T. (2001). *The weak hydrogen bond: in structural chemistry and biology*. International Union of Crystallography monographs on crystallography, Oxford: Oxford University Press.

Devi, T. S., Rajitha, G., Swathi, K., Sudheer Kumar, K., and Umamaheswari, A. (2020). “Synthesis, Molecular Docking Studies and Biological Evaluation of N-Acylarylhydrazones as Anti-Inflammatory Agents.” *GeNeDis 2018*, Advances in Experimental Medicine and Biology, P. Vlamos, ed., Cham: Springer International Publishing, 137–148.

Dingerdissen, H. M., Torcivia-Rodriguez, J., Hu, Y., Chang, T.-C., Mazumder, R., and Kahsay, R. (2018). “BioMuta and BioXpress: mutation and expression knowledgebases for cancer biomarker discovery.” *Nucleic Acids Res.*, 46(D1), D1128–D1136.

Ditchfield, R., Hehre, W. J., and Pople, J. A. (1971). “Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules.” *J. Chem. Phys.*, 54(2), 724–728.

Dixon, S. L., Smondirev, A. M., Knoll, E. H., Rao, S. N., Shaw, D. E., and Friesner, R. A. (2006). “PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results.” *J. Comp. Aided Mol. Des.*, 20(10–11), 647–671.

Dolatkhah, Z., Javanshir, S., Sadr, A. S., Hosseini, J., and Sardari, S. (2017). “Synthesis, Molecular Docking, Molecular Dynamics Studies, and Biological Evaluation of 4 *H* - Chromone-1,2,3,4-tetrahydropyrimidine-5-carboxylate Derivatives as Potential Antileukemic Agents.” *J. Chem. Inf. Model.*, 57(6), 1246–1257.

Donawho, C. K., Luo, Y., Luo, Y., Penning, T. D., Bauch, J. L., Bouska, J. J., Bontcheva-Diaz, V. D., Cox, B. F., DeWeese, T. L., Dillehay, L. E., Ferguson, D. C., Ghoreishi-Haack, N. S., Grimm, D. R., Guan, R., Han, E. K., Holley-Shanks, R. R., Hristov, B., Idler, K. B., Jarvis, K., Johnson, E. F., Kleinberg, L. R., Klinghofer, V., Lasko, L. M., Liu, X., Marsh, K. C., McGonigal, T. P., Meulbroek, J. A., Olson, A. M., Palma, J. P., Rodriguez, L. E., Shi, Y., Stavropoulos, J. A., Tsurutani, A. C., Zhu, G.-D., Rosenberg, S. H., Giranda, V. L., and Frost, D. J. (2007). “ABT-888, an Orally Active Poly(ADP-Ribose) Polymerase Inhibitor that Potentiates DNA-Damaging Agents in Preclinical Tumor Models.” *Clin. Cancer Res.*, 13(9), 2728–2737.

Duffy, J. L., Kirk, B. A., Konteatis, Z., Campbell, E. L., Liang, R., Brady, E. J., Candelore, M. R., Ding, V. D. H., Jiang, G., Liu, F., Qureshi, S. A., Saperstein, R., Szalkowski, D., Tong, S., Tota, L. M., Xie, D., Yang, X., Zafian, P., Zheng, S., Chapman, K. T., Zhang, B. B., and Tata, J. R. (2005). “Discovery and investigation of a novel class of thiophene-derived antagonists of the human glucagon receptor.” *Bioorg. Med. Chem. Lett.*, 15(5), 1401–1405.

Ebrahim, A., Riley, B. T., Kumaran, D., Andi, B., Fuchs, M. R., McSweeney, S., and Keedy, D. A. (2022). “The temperature-dependent conformational ensemble of SARS-CoV-2 main protease (M^{PRO}).” *IUCrJ.*, 9(5), 682-694.

El-Baba, T. J., Lutomski, C. A., Kantsadi, A. L., Malla, T. R., John, T., Mikhailov, V., Bolla, J. R., Schofield, C. J., Zitzmann, N., Vakonakis, I., and Robinson, C. V. (2020). “Allosteric Inhibition of the SARS-CoV-2 Main Protease: Insights from Mass Spectrometry Based Assays**.” *Angew. Chem. Int. Ed.*, 59(52), 23544–23548.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). “A smooth particle mesh Ewald method.” *J. Chem. Phys.*, 103(19), 8577–8593.

Farres, J., Martin-Caballero, J., Martinez, C., Lozano, J. J., Llacuna, L., Ampurdanes, C., Ruiz-Herguido, C., Dantzer, F., Schreiber, V., Villunger, A., Bigas, A., and Yelamos, J. (2013). "Parp-2 is required to maintain hematopoiesis following sublethal - irradiation in mice." *Blood*, 122(1), 44–54.

Fehr, A. R., and Perlman, S. (2015). "Coronaviruses: An Overview of Their Replication and Pathogenesis." *Coronaviruses*, Methods in Molecular Biology, H. J. Maier, E. Bickerton, and P. Britton, eds., New York, NY: Springer New York, 1–23.

Feig, M., Onufriev, A., Lee, M. S., Im, W., Case, D. A., and Brooks, C. L. (2004). "Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures." *J. Comput. Chem.*, 25(2), 265–284.

Ferraro, M., D'Annessa, I., Moroni, E., Morra, G., Paladino, A., Rinaldi, S., Compostella, F., and Colombo, G. (2019). "Allosteric Modulators of HSP90 and HSP70: Dynamics Meets Function through Structure-Based Drug Design." *J. Med. Chem.*, 62(1), 60–87.

Ferreira, J. C., Fadl, S., Villanueva, A. J., and Rabeh, W. M. (2021). "Catalytic Dyad Residues His41 and Cys145 Impact the Catalytic Activity and Overall Conformational Fold of the Main SARS-CoV-2 Protease 3-Chymotrypsin-Like Protease." *Front. Chem.*, 9, 692168.

Filipski, K. J., Bian, J., Ebner, D. C., Lee, E. C. Y., Li, J.-C., Sammons, M. F., Wright, S. W., Stevens, B. D., Didiuk, M. T., Tu, M., Perreault, C., Brown, J., Atkinson, K., Tan, B., Salatto, C. T., Litchfield, J., Pfefferkorn, J. A., and Guzman-Perez, A. (2012). "A novel series of glucagon receptor antagonists with reduced molecular weight and lipophilicity." *Bioorg. Med. Chem. Lett.*, 22(1), 415–420.

Foloppe, N., and Hubbard, R. (2006). "Towards predictive ligand design with free-energy based computational methods?" *Curr. Med. Chem.*, 13(29), 3583–3608.

Ford, M. G., Weimar, T., Köhli, T., and Woods, R. J. (2003). “Molecular dynamics simulations of galectin-1-oligosaccharide complexes reveal the molecular basis for ligand diversity: Galectin-1 Ligand Diversity.” *Proteins*, 53(2), 229–240.

Frauenfelder, H., Sligar, S., and Wolynes, P. (1991). “The energy landscapes and motions of proteins.” *Science*, 254(5038), 1598–1603.

Frisch, M., Trucks, G., Schlegel, H., Scuseria, G., Robb, M., Cheeseman, J., Scalmani, G., Barone, V., Mennucci, B., Petersson, G., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H., Izmaylov, A., Bloino, J., Zheng, G., Sonnenberg, J., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J., Peralta, J., Ogliaro, F., Bearpark, M., Heyd, J., Brothers, E., Kudin, K., Staroverov, V., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J., Iyengar, S., Tomasi, J., Cossi, M., Rega, N., Millam, J., Klene, M., Knox, J., Cross, J., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R., Yazyev, O., Austin, A., Cammi, R., Pomelli, C., Ochterski, J., Martin, R., Morokuma, K., Zakrzewski, V., Voth, G., Salvador, P., Dannenberg, J., Dapprich, S., Daniels, A., Farkas, Foresman, J., Ortiz, J., Cioslowski, J., and Fox, D. (2009). “Gaussian 09, Revision B.01.” *Gaussian 09, Revision B.01*, Gaussian, Inc., Wallingford CT.

Gaba, M., Punam, G., and Sarbjot, S. (2010). “An overview on Molecular Docking.” *Int. J. Drug Dev. Res.*, 2(2), 219-231.

Gajdoš, P., Jeżowicz, T., Uher, V., and Dohnálek, P. (2016). “A parallel Fruchterman–Reingold algorithm optimized for fast visualization of large graphs and swarms of data.” *Swarm Evol. Comput.*, 26, 56–63.

Gasteiger, J., and Marsili, M. (1980). “Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges.” *Tetrahedron*, 36(22), 3219–3228.

Genheden, S., and Ryde, U. (2015). “The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities.” *Expert Opin. Drug Dis.*, 10(5), 449–461.

Ghanbarzadeh, S., Ghasemi, S., Shayanfar, A., and Ebrahimi-Najafabadi, H. (2015). "2D-QSAR study of some 2,5-diaminobenzophenone farnesyltransferase inhibitors by different chemometric methods." *EXCLI J.*; 14:Doc484; ISSN 1611-2156.

Glazer, Robert I. (1980). "Adenosine deaminase inhibitors: Their role in chemotherapy and immunosuppression." *Cancer Chemother. Pharmacol.*, 4(4).

Golbraikh, A., Shen, M., Xiao, Z., Xiao, Y.-D., Lee, K.-H., and Tropsha, A. (2003). "Rational selection of training and test sets for the development of validated QSAR models." *J. Comput. Aided Mol. Des.*, 17(2-4), 241-253.

Golbraikh, A., and Tropsha, A. (2002). "Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection." *J. Comput. Aided Mol. Des.*, 16(5-6), 357-369.

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., O'Meara, M. J., Rezelj, V. V., Guo, J. Z., Swaney, D. L., Tummino, T. A., Hüttenhain, R., Kaake, R. M., Richards, A. L., Tutuncuoglu, B., Foussard, H., Batra, J., Haas, K., Modak, M., Kim, M., Haas, P., Polacco, B. J., Braberg, H., Fabius, J. M., Eckhardt, M., Soucheray, M., Bennett, M. J., Cakir, M., McGregor, M. J., Li, Q., Meyer, B., Roesch, F., Vallet, T., Mac Kain, A., Miorin, L., Moreno, E., Naing, Z. Z. C., Zhou, Y., Peng, S., Shi, Y., Zhang, Z., Shen, W., Kirby, I. T., Melnyk, J. E., Chorba, J. S., Lou, K., Dai, S. A., Barrio-Hernandez, I., Memon, D., Hernandez-Armenta, C., Lyu, J., Mathy, C. J. P., Perica, T., Pilla, K. B., Ganesan, S. J., Saltzberg, D. J., Rakesh, R., Liu, X., Rosenthal, S. B., Calviello, L., Venkataramanan, S., Liboy-Lugo, J., Lin, Y., Huang, X.-P., Liu, Y., Wankowicz, S. A., Bohn, M., Safari, M., Ugur, F. S., Koh, C., Savar, N. S., Tran, Q. D., Shengjuler, D., Fletcher, S. J., O'Neal, M. C., Cai, Y., Chang, J. C. J., Broadhurst, D. J., Klippsten, S., Sharp, P. P., Wenzell, N. A., Kuzuoglu-Ozturk, D., Wang, H.-Y., Trenker, R., Young, J. M., Cavero, D. A., Hiatt, J., Roth, T. L., Rathore, U., Subramanian, A., Noack, J., Hubert, M., Stroud, R. M., Frankel, A. D., Rosenberg, O. S., Verba, K. A., Agard, D. A., Ott, M., Emerman, M., Jura, N., Zastrow, M. von, Verdin, E., Ashworth, A., Schwartz, O., Enfert, C. d', Mukherjee, S., Jacobson, M., Malik, H. S., Fujimori, D. G., Ideker, T., Craik, C. S., Floor, S. N., Fraser, J. S., Gross,

J. D., Sali, A., Roth, B. L., Ruggero, D., Taunton, J., Kortemme, T., Beltrao, P., Vignuzzi, M., García-Sastre, A., Shokat, K. M., Shoichet, B. K., and Krogan, N. J. (2020). “A SARS-CoV-2 protein interaction map reveals targets for drug repurposing.” *Nature*, 583(7816), 459–468.

Goyal, B., and Goyal, D. (2020). “Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy.” *ACS Comb. Sci.*, 22(6), 297–305.

Gui, B., Gui, F., Takai, T., Feng, C., Bai, X., Fazli, L., Dong, X., Liu, S., Zhang, X., Zhang, W., Kibel, A. S., and Jia, L. (2019). “Selective targeting of PARP-2 inhibits androgen receptor signaling and prostate cancer growth through disruption of FOXA1 function.” *Proc. Natl. Acad. Sci. USA*, 116(29), 14573–14582.

Güner, O. F. (Ed.). (2000). *Pharmacophore perception, development, and use in drug design*. IUL biotechnology series, LaJolla, CA: International University Line.

Gunsteren, W. F. van, and Berendsen, H. J. C. (1987). “Groningen Molecular Simulation (GROMOS) Library Manual.” Biomos, Groningen, The Netherlands.

Gutierrez-Villagomez, J. M., Campos-García, T., Molina Torres, J., López, M. G., and Vázquez-Martínez, J. (2020). “Alkamides and Piperamides as Potential Antivirals Against the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2).” *J. Phys. Chem. Lett.*, acs.jpcclett.0c01685.

Guzman-Perez, A., Pfefferkorn, J. A., Lee, E. C. Y., Stevens, B. D., Aspnes, G. E., Bian, J., Didiuk, M. T., Filipinski, K. J., Moore, D., Perreault, C., Sammons, M. F., Tu, M., Brown, J., Atkinson, K., Litchfield, J., Tan, B., Samas, B., Zavadoski, W. J., Salatto, C. T., and Treadway, J. (2013). “The design and synthesis of a potent glucagon receptor antagonist with favorable physicochemical and pharmacokinetic properties as a candidate for the treatment of type 2 diabetes mellitus.” *Bioorg. Med. Chem. Lett.*, 23(10), 3051–3058.

Hadaji, E. G., Bourass, M., Ouammou, A., and Bouachrine, M. (2017). “3D-QSAR models to predict anti-cancer activity on a series of protein P38 MAP kinase inhibitors.” *J. Taibah Univ. Sci.*, 11(3), 392–407.

Hayward, S., Kitao, A., and Gō, N. (2008). “Harmonic and anharmonic aspects in the dynamics of BPTI: A normal mode analysis and principal component analysis.” *Protein Sci.*, 3(6), 936–943.

He, J., Tao, H., Yan, Y., Huang, S.-Y., and Xiao, Y. (2020). “Molecular Mechanism of Evolution and Human Infection with SARS-CoV-2.” *Viruses*, 12(4), 428.

Hess, B., Bekker, H., and Berendsen, H. J. C. (1998). “LINCS: A linear constraint solver for molecular simulations.” *J. Comput. Chem.*, 18(12), 10.

Hess, B., and Vegt, N. F. A. van der. (2006). “Hydration Thermodynamic Properties of Amino Acid Analogues: A Systematic Comparison of Biomolecular Force Fields and Water Models.” *J. Phys. Chem. B*, 110(35), 17616–17626.

Hockney, R. W., and Eastwood, J. W. (1988). *Computer simulation using particles*. Bristol [England] ; Philadelphia: A. Hilger.

Hoffmann, M., and Rychlewski, J. (2002). “Density Functional Theory (DFT) and Drug Design.” *Reviews of Modern Quantum Chemistry*, WORLD SCIENTIFIC, 1767–1803.

Homeyer, N., and Gohlke, H. (2012). “Free Energy Calculations by the Molecular Mechanics Poisson–Boltzmann Surface Area Method.” *Mol. Inform.*, 31(2), 114–122.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). “Comparison of multiple Amber force fields and development of improved protein backbone parameters.” *Proteins*, 65(3), 712–725.

Hospital, A., Goñi, J. R., Orozco, M., and Gelpí, J. L. (2015). “Molecular dynamics simulations: advances and applications.” *Adv. Appl. Bioinform. Chem.*, 8, 37–47.

Hottiger, M. O. (2015). “Nuclear ADP-Ribosylation and Its Role in Chromatin Plasticity, Cell Differentiation, and Epigenetics.” *Annu. Rev. Biochem.*, 84(1), 227–263.

- Hou, T., Wang, J., Li, Y., and Wang, W. (2011). "Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations." *J. Chem. Inf. Model.*, 51(1), 69–82.
- Huang, C., Wei, P., Fan, K., Liu, Y., and Lai, L. (2004). "3C-like Proteinase from SARS Coronavirus Catalyzes Substrate Hydrolysis by a General Base Mechanism." *Biochemistry*, 43(15), 4568–4574.
- Huff, S., Kummetha, I. R., Tiwari, S. K., Huante, M. B., Clark, A. E., Wang, S., Bray, W., Smith, D., Carlin, A. F., Endsley, M., and Rana, T. M. (2022). "Discovery and Mechanism of SARS-CoV-2 Main Protease Inhibitors." *J. Med. Chem.*, 65(4), 2866–2879.
- Huynh, T., Wang, H., and Luan, B. (2020). "In Silico Exploration of the Molecular Mechanism of Clinically Oriented Drugs for Possibly Inhibiting SARS-CoV-2's Main Protease." *J. Phys. Chem. Lett.*, 11(11), 4413–4420.
- Ioakimidis, L., Thoukydidis, L., Mirza, A., Naeem, S., and Reynisson, J. (2008). "Benchmarking the Reliability of QikProp. Correlation between Experimental and Predicted Values." *QSAR Comb. Sci.*, 27(4), 445–456.
- Iribarne, F., Paulino, M., Aguilera, S., and Tapia, O. (2009). "Assaying phenothiazine derivatives as trypanothione reductase and glutathione reductase inhibitors by theoretical docking and Molecular Dynamics studies." *J. Mol. Graph. Model.*, 28(4), 371–381.
- Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., and Coleman, R. G. (2012). "ZINC: A Free Tool to Discover Chemistry for Biology." *J. Chem. Inf. Model.*, 52(7), 1757–1768.
- Itoh, Y., Nakashima, Y., Tsukamoto, S., Kurohara, T., Suzuki, M., Sakae, Y., Oda, M., Okamoto, Y., and Suzuki, T. (2019). "N⁺-C-H···O Hydrogen bonds in protein-ligand complexes." *Sci Rep*, 9(1), 767.

Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J. F., Honig, B., Shaw, D. E., and Friesner, R. A. (2004). “A hierarchical approach to all-atom protein loop prediction.” *Proteins*, 55(2), 351–367.

Jang, W. D., Jeon, S., Kim, S., and Lee, S. Y. (2021). “Drugs repurposed for COVID-19 by virtual screening of 6,218 drugs and cell-based assay.” *Proc. Natl. Acad. Sci. USA*, 118(30), e2024302118.

Jazayeri, A., Doré, A. S., Lamb, D., Krishnamurthy, H., Southall, S. M., Baig, A. H., Bortolato, A., Koglin, M., Robertson, N. J., Errey, J. C., Andrews, S. P., Teobald, I., Brown, A. J. H., Cooke, R. M., Weir, M., and Marshall, F. H. (2016). “Extra-helical binding site of a glucagon receptor antagonist.” *Nature*, 533(7602), 274–277.

Jhoti, H., and Leach, A. R. (Eds.). (2007). *Structure-based drug discovery*. Dordrecht: Springer.

Jiang, G., and Zhang, B. B. (2003). “Glucagon and regulation of glucose metabolism.” *Am. J. Physiol.-Endocrinol. Metab.*, 284(4), E671–E678.

Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Liu, X., Yang, X., Bai, F., Liu, H., Liu, X., Guddat, L. W., Xu, W., Xiao, G., Qin, C., Shi, Z., Jiang, H., Rao, Z., and Yang, H. (2020). “Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors.” *Nature*, 582(7811), 289–293.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). “Comparison of simple potential functions for simulating liquid water.” *J. Chem. Phys.*, 79(2), 926–935.

Jorgensen, W. L., and Duffy, E. M. (2002). “Prediction of drug solubility from structure.” *Adv. Drug Deliv. Rev.*, 54(3), 355–366.

Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996). “Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids.” *J. Am. Chem. Soc.*, 118(45), 11225–11236.

Jorgensen, W. L., and Tirado-Rives, J. (1988). “The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin.” *J. Am. Chem. Soc.*, 110(6), 1657–1666.

Kabsch, W., and Sander, C. (1983). “Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features.” *Biopolymers*, 22(12), 2577–2637.

Kaiser, J. (2011). “Combining Targeted Drugs to Stop Resistant Tumors.” *Science*, 331(6024), 1542–1545.

Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2001). “Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides.” *J. Phys. Chem. B*, 105(28), 6474–6487.

Karelson, M., Lobanov, V. S., and Katritzky, A. R. (1996). “Quantum-Chemical Descriptors in QSAR/QSPR Studies.” *Chem. Rev.*, 96(3), 1027–1044.

Karplus, M. (2014). “Development of Multiscale Models for Complex Chemical Systems: From H₂ to Biomolecules (Nobel Lecture).” *Angew. Chem. Int. Ed.*, 53(38), 9992–10005.

Kaserer, T., Beck, K., Akram, M., Odermatt, A., and Schuster, D. (2015). “Pharmacophore Models and Pharmacophore-Based Virtual Screening: Concepts and Applications Exemplified on Hydroxysteroid Dehydrogenases.” *Molecules*, 20(12), 22799–22832.

Kaushik, A. C., Kumar, S., Wei, D. Q., and Sahi, S. (2018). “Structure Based Virtual Screening Studies to Identify Novel Potential Compounds for GPR142 and Their Relative Dynamic Analysis for Study of Type 2 Diabetes.” *Front. Chem.*, 6, 23.

Kavitha, R., Karunagaran, S., Chandrabose, S. S., Lee, K. W., and Meganathan, C. (2015). “Pharmacophore modeling, virtual screening, molecular docking studies and

density functional theory approaches to identify novel ketohexokinase (KHK) inhibitors.” *Biosystems*, 138, 39–52.

Ke, Wang, Zhang, Zhong, Wang, Zeng, and Ba. (2019). “The Role of PARPs in Inflammation—and Metabolic—Related Diseases: Molecular Mechanisms and Beyond.” *Cells*, 8(9), 1047.

Khan, S., Farooq, U., and Kurnikova, M. (2017). “Protein stability and dynamics influenced by ligands in extremophilic complexes – a molecular dynamics investigation.” *Mol. BioSyst.*, 13(9), 1874–1887.

Khayat, R., Batra, R., Beberitz, G. A., Olson, M. W., and Tong, L. (2004). “Characterization of the Monomer–Dimer Equilibrium of Human Cytomegalovirus Protease by Kinetic Methods.” *Biochemistry*, 43(2), 316–322.

Klebe, G. (2000). “Recent developments in structure-based drug design.” *J. Mol. Med.*, 78(5), 269–281.

Kohn, W., and Sham, L. J. (1965). “Self-Consistent Equations Including Exchange and Correlation Effects.” *Phys. Rev.*, 140(4A), A1133–A1138.

Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham, T. E. (2000). “Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models.” *Acc. Chem. Res.*, 33(12), 889–897.

Kollman, Peter. (1993). “Free energy calculations: Applications to chemical and biochemical phenomena.” *Chem. Rev.*, 93(7), 2395–2417.

Kordzadeh, A., and Saadatabadi, A. R. (2021). “Effects of the Temperature and the pH on the Main Protease of SARS-CoV-2: A Molecular Dynamics Simulation Study.” *Biointerface Res. Appl. Chem.*, 12(6), 7239–7248.

- Kore, P. P., Mutha, M. M., Antre, R. V., Oswal, R. J., and Kshirsagar, S. S. (2012). "Computer-Aided Drug Design: An Innovative Tool for Modeling." *Open J. Med. Chem.*, 02(04), 139–148.
- Kramer, B., Rarey, M., and Lengauer, T. (1999). "Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking." *Proteins: Struct. Funct. Genet.*, 37(2), 228–241.
- Kraus, W. L. (2015). "PARPs and ADP-Ribosylation: 50 Years ... and Counting." *Mol. Cell*, 58(6), 902–910.
- Kumar, D., Kumari, K., Vishvakarma, V. K., Jayaraj, A., Kumar, D., Ramappa, V. K., Patel, R., Kumar, V., Dass, S. K., Chandra, R., and Singh, P. (2020a). "Promising inhibitors of main protease of novel corona virus to prevent the spread of COVID-19 using docking and molecular dynamics simulation." *J. Biomol. Struct. Dyn.*, 1–15.
- Kumar, N., Hendriks, B. S., Janes, K. A., Graaf, D. de, and Lauffenburger, D. A. (2006). "Applying computational modeling to drug discovery and development." *Drug Discov. Today*, 11(17–18), 806–811.
- Kumar, S., Bhardwaj, V. K., Singh, R., Das, P., and Purohit, R. (2022). "Identification of acridinedione scaffolds as potential inhibitor of DENV-2 C protein: An in silico strategy to combat dengue." *J. Cell. Biochem.*, 123(5), 935–946.
- Kumar, Y., Singh, H., and Patel, C. N. (2020b). "In silico prediction of potential inhibitors for the Main protease of SARS-CoV-2 using molecular docking and dynamics simulation based drug-repurposing." *J. Infect. Public Health*, S1876034120305268.
- Kurukulasuriya, R., Sorensen, B. K., Link, J. T., Patel, J. R., Jae, H.-S., Winn, M. X., Rohde, J. R., Grihalde, N. D., Lin, C. W., Ogiela, C. A., Adler, A. L., and Collins, C. A. (2004). "Biaryl amide glucagon receptor antagonists." *Bioorg. Med. Chem. Lett.*, 14(9), 2047–2050.

Lane, J. R., Chubukov, P., Liu, W., Canals, M., Cherezov, V., Abagyan, R., Stevens, R. C., and Katritch, V. (2013). “Structure-Based Ligand Discovery Targeting Orthosteric and Allosteric Pockets of Dopamine Receptors.” *Mol. Pharmacol.*, 84(6), 794–807.

Lawyer, F. C., Stoffel, S., Saiki, R. K., Chang, S. Y., Landre, P. A., Abramson, R. D., and Gelfand, D. H. (1993). “High-level expression, purification, and enzymatic characterization of full-length *Thermus aquaticus* DNA polymerase and a truncated form deficient in 5’ to 3’ exonuclease activity.” *Genome Res.*, 2(4), 275–287.

Lee, C., Yang, W., and Parr, R. G. (1988). “Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.” *Phys. Rev. B*, 37(2), 785–789.

Lee, E. C. Y., Tu, M., Stevens, B. D., Bian, J., Aspnes, G., Perreault, C., Sammons, M. F., Wright, S. W., Litchfield, J., Kalgutkar, A. S., Sharma, R., Didiuk, M. T., Ebner, D. C., Filipski, K. J., Brown, J., Atkinson, K., Pfefferkorn, J. A., and Guzman-Perez, A. (2014). “Identification of a novel conformationally constrained glucagon receptor antagonist.” *Bioorg. Med. Chem. Lett.*, 24(3), 839–844.

Lengauer, T., and Rarey, M. (1996). “Computational methods for biomolecular docking.” *Curr. Opin. Struct. Biol.*, 6(3), 402–406.

Lewis, R. A. (2011). “Chapter 4. The Development of Molecular Modelling Programs: The Use and Limitations of Physical Models.” *RSC Drug Discovery*, D. J. Livingstone and A. M. Davis, eds., Cambridge: Royal Society of Chemistry, 88–107.

Li, F., Li, X., Liu, X., Zhang, L., You, L., Zhao, J., and Wu, H. (2011a). “Noncovalent interactions between hydroxylated polycyclic aromatic hydrocarbon and DNA: Molecular docking and QSAR study.” *Environ. Toxicol. Pharmacol.*, 32(3), 373–381.

Li, J., Abel, R., Zhu, K., Cao, Y., Zhao, S., and Friesner, R. A. (2011b). “The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling.” *Proteins*, 79(10), 2794–2812.

- Li, S., and Hong, M. (2011). "Protonation, Tautomerization, and Rotameric Structure of Histidine: A Comprehensive Study by Magic-Angle-Spinning Solid-State NMR." *J. Am. Chem. Soc.*, 133(5), 1534–1544.
- Lien, E. J., Guo, Z.-R., Li, R.-L., and Su, C.-T. (1982). "Use of Dipole Moment as a Parameter in Drug–Receptor Interaction and Quantitative Structure–Activity Relationship Studies." *J. Pharm. Sci.*, 71(6), 641–655.
- Lionta, E., Spyrou, G., Vassilatis, D., and Cournia, Z. (2014). "Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances." *Curr. Top. Med. Chem.*, 14(16), 1923–1938.
- Lipinski, C. A. (2004). "Lead- and drug-like compounds: the rule-of-five revolution." *Drug Discov. Today Technol.*, 1(4), 337–341.
- Lobanov, M. Yu., Bogatyreva, N. S., and Galzitskaya, O. V. (2008). "Radius of gyration as an indicator of protein structure compactness." *Mol. Biol.*, 42(4), 623–628.
- Luzar, A., and Chandler, D. (1996). "Hydrogen-bond kinetics in liquid water." *Nature*, 379(6560), 55–57.
- MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998). "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins." *J. Phys. Chem. B*, 102(18), 3586–3616.
- Mackerell, A. D., Feig, M., and Brooks, C. L. (2004). "Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations." *J. Comput. Chem.*, 25(11), 1400–1415.

- Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013). "Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments." *J. Comp. Aided Mol. Des.*, 27(3), 221–234.
- Maffucci, I., and Contini, A. (2020). "In Silico Drug Repurposing for SARS-CoV-2 Main Proteinase and Spike Proteins." *J. Proteome Res.*, 19(11), 4637–4648.
- Majewski, M., Ruiz-Carmona, S., and Barril, X. (2019). "An investigation of structural stability in protein-ligand complexes reveals the balance between order and disorder." *Commun. Chem.*, 2(1), 110.
- Malkhasian, A. Y. S., and Howlin, B. J. (2016). "Docking and DFT studies on ligand binding to Quercetin 2,3-dioxygenase." *J. Biomol. Struct. Dyn.*, 34(11), 2453–2461.
- Mark, P., and Nilsson, L. (2001). "Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K." *J. Phys. Chem. A*, 105(43), 9954–9960.
- Martić, M., Tatić, I., Marković, S., Kujundžić, N., and Koštrun, S. (2004). "Synthesis, biological activity and molecular modeling studies of novel COX-1 inhibitors." *Eur. J. Med. Chem.*, 39(2), 141–151.
- Martyna, G. J., Klein, M. L., and Tuckerman, M. (1992). "Nosé–Hoover chains: The canonical ensemble via continuous dynamics." *J. Chem. Phys.*, 97(4), 2635–2643.
- Martyna, G. J., Tobias, D. J., and Klein, M. L. (1994). "Constant pressure molecular dynamics algorithms." *J. Chem. Phys.*, 101(5), 4177–4189.
- Mendes-Pereira, A. M., Martin, S. A., Brough, R., McCarthy, A., Taylor, J. R., Kim, J., Waldman, T., Lord, C. J., and Ashworth, A. (2009). "Synthetic lethal targeting of *PTEN* mutant cells with PARP inhibitors." *EMBO Mol. Med.*, 1(6-7), 315–322.
- Meng, E. C., Shoichet, B. K., and Kuntz, I. D. (1992). "Automated docking with grid-based energy evaluation." *J. Comput. Chem.*, 13(4), 505–524.

Meng, X.-Y., Zhang, H.-X., Mezei, M., and Cui, M. (2011). “Molecular docking: a powerful approach for structure-based drug discovery.” *Curr. Comput. Aided Drug Des.*, 7(2), 146–157.

Mengist, H. M., Dilnessa, T., and Jin, T. (2021). “Structural Basis of Potential Inhibitors Targeting SARS-CoV-2 Main Protease.” *Front. Chem.*, 9, 622898.

Mercadante, D., Gräter, F., and Daday, C. (2018). “CONAN: A Tool to Decode Dynamical Information from Molecular Interaction Maps.” *Biophys. J.*, 114(6), 1267–1273.

Mittal, L., Kumari, A., Srivastava, M., Singh, M., and Asthana, S. (2020). “Identification of potential molecules against COVID-19 main protease through structure-guided virtual screening approach.” *J. Biomol. Struct. Dyn.*, 1–19.

Morningstar-Kywi, N., Wang, K., Asbell, T. R., Wang, Z., Giles, J. B., Lai, J., Brill, D., Sutch, B. T., and Haworth, I. S. (2022). “Prediction of Water Distributions and Displacement at Protein–Ligand Interfaces.” *J. Chem. Inf. Model.*, acs.jcim.1c01266.

Moroni, F., Formentini, L., Gerace, E., Camaioni, E., Pellegrini-Giampietro, D. E., Chiarugi, A., and Pellicciari, R. (2009). “Selective PARP-2 inhibitors increase apoptosis in hippocampal slices but protect cortical cells in models of post-ischaemic brain damage.” *Br. J. Pharmacol.*, 157(5), 854–862.

Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998). “Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function.” *J. Comput. Chem.*, 19(14), 1639–1662.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009). “AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility.” *J. Comput. Chem.*, 30(16), 2785–2791.

Muratov, E. N., Amaro, R., Andrade, C. H., Brown, N., Ekins, S., Fourches, D., Isayev, O., Kozakov, D., Medina-Franco, J. L., Merz, K. M., Oprea, T. I., Poroikov, V., Schneider, G., Todd, M. H., Varnek, A., Winkler, D. A., Zakharov, A. V., Cherkasov,

A., and Tropsha, A. (2021). “A critical overview of computational approaches employed for COVID-19 drug discovery.” *Chem. Soc. Rev.*, 50(16), 9121–9151.

Narkhede, R. R., Pise, A. V., Cheke, R. S., and Shinde, S. D. (2020). “Recognition of Natural Products as Potential Inhibitors of COVID-19 Main Protease (Mpro): In-Silico Evidences.” *Nat. Prod. Bioprospect.*, 10, 297-306.

Oliver, A. W. (2004). “Crystal structure of the catalytic fragment of murine poly(ADP-ribose) polymerase-2.” *Nucleic Acids Res.*, 32(2), 456–464.

Pang, X., Zhou, L., Zhang, M., Zhang, L., Xu, L., Xie, F., Xie, F., Yu, L., and Zhang, X. (2012). “Two Rules on the Protein-Ligand Interaction.” *Open Conf. Proc. J.*, 3(1), 70–80.

Papeo, G., Posterl, H., Borghi, D., Busel, A. A., Caprera, F., Casale, E., Ciomei, M., Cirila, A., Corti, E., D’Anello, M., Fasolini, M., Forte, B., Galvani, A., Isacchi, A., Khvat, A., Krasavin, M. Y., Lupi, R., Orsini, P., Perego, R., Pesenti, E., Pezzetta, D., Rainoldi, S., Riccardi-Sirtori, F., Scolaro, A., Sola, F., Zuccotto, F., Felder, E. R., Donati, D., and Montagnoli, A. (2015). “Discovery of 2-[1-(4,4-Difluorocyclohexyl)piperidin-4-yl]-6-fluoro-3-oxo-2,3-dihydro-1 *H* -isoindole-4-carboxamide (NMS-P118): A Potent, Orally Available, and Highly Selective PARP-1 Inhibitor for Cancer Therapy.” *J. Med. Chem.*, 58(17), 6875–6898.

Parr, R. G., and Yang, W. (1994). *Density-functional theory of atoms and molecules*. International series of monographs on chemistry, New York, NY: Oxford Univ. Press [u.a.].

Parrinello, M., and Rahman, A. (1981). “Polymorphic transitions in single crystals: A new molecular dynamics method.” *J. Appl. Phys.*, 52(12), 7182–7190.

Patil, R., Das, S., Stanley, A., Yadav, L., Sudhakar, A., and Varma, A. K. (2010). “Optimized Hydrophobic Interactions and Hydrogen Bonding at the Target-Ligand Interface Leads the Pathways of Drug-Designing.” *PLoS ONE*, (S. Hannenhalli, ed.), 5(8), e12029.

- Paul, S., and Chandra, A. (2005). "Hydrogen Bond Properties and Dynamics of Liquid–Vapor Interfaces of Aqueous Methanol Solutions." *J. Chem. Theory Comput.*, 1(6), 1221–1231.
- Perina, D., Mikoč, A., Ahel, J., Četković, H., Žaja, R., and Ahel, I. (2014). "Distribution of protein poly(ADP-ribosyl)ation systems across all domains of life." *DNA Repair*, 23, 4–16.
- Pettitt, B. M., and Rossky, P. J. (1986). "Alkali halides in water: Ion–solvent correlations and ion–ion potentials of mean force at infinite dilution." *J. Chem. Phys.*, 84(10), 5836–5844.
- Petukhov, M., Rychkov, G., Firsov, L., and Serrano, L. (2004). "H-bonding in protein hydration revisited." *Protein Sci.*, 13(8), 2120–2129.
- Piao, L., Chen, Z., Li, Q., Liu, R., Song, W., Kong, R., and Chang, S. (2019). "Molecular Dynamics Simulations of Wild Type and Mutants of SAPAP in Complexed with Shank3." *Int. J. Mol. Sci.*, 20(1), 224.
- Pirhadi, S., Shiri, F., and Ghasemi, J. B. (2015). "Multivariate statistical analysis methods in QSAR." *RSC Adv.*, 5(127), 104635–104665.
- Polański, J., Gieleciak, R., and Bąk, A. (2002). "The Comparative Molecular Surface Analysis (COMSA) – A Nongrid 3D QSAR Method by a Coupled Neural Network and PLS System: Predicting pK_a Values of Benzoic and Alkanoic Acids." *J. Chem. Inf. Comput. Sci.*, 42(2), 184–191.
- Rácz, A., Bajusz, D., and Héberger, K. (2015). "Consistency of QSAR models: Correct split of training and test sets, ranking of models and performance parameters." *SAR QSAR Environ. Res.*, 26(7–9), 683–700.
- Rath, S. L., and Kumar, K. (2020). "Investigation of the Effect of Temperature on the Structure of SARS-CoV-2 Spike Protein by Molecular Dynamics Simulations." *Front. Mol. Biosci.*, 7, 583523.

Reddy, K. B., Rao, G. M., Kumar, B. V., Choudhury, C., and Rajan, K. S. (2018). “Design and Synthesis of Benzimidazole-4-carboxamides as Potent Poly(ADP-Ribose) Polymerase-1 (PARP-1) Inhibitors.” *Curr. Bioact. Compd.*, 14(2), 100–111.

Reddy, K. K., and Singh, S. K. (2014). “Combined ligand and structure-based approaches on HIV-1 integrase strand transfer inhibitors.” *Chem. Biol. Interact.*, 218, 71–81.

Ren, X., Zeng, R., Wang, C., Zhang, M., Liang, C., Tang, Z., and Ren, J. (2017). “Structural insight into inhibition of REV7 protein interaction revealed by docking, molecular dynamics and MM/PBSA studies.” *RSC Adv.*, 7(44), 27780–27786.

Riccio, A. A., Cingolani, G., and Pascal, J. M. (2016). “PARP-2 domain requirements for DNA damage-dependent activation and localization to sites of DNA damage.” *Nucleic Acids Res.*, 44(4), 1691–1702.

Riva, L., Yuan, S., Yin, X., Martin-Sancho, L., Matsunaga, N., Pache, L., Burgstaller-Muehlbacher, S., De Jesus, P. D., Teriete, P., Hull, M. V., Chang, M. W., Chan, J. F.-W., Cao, J., Poon, V. K.-M., Herbert, K. M., Cheng, K., Nguyen, T.-T. H., Rubanov, A., Pu, Y., Nguyen, C., Choi, A., Rathnasinghe, R., Schotsaert, M., Miorin, L., Dejoze, M., Zwaka, T. P., Sit, K.-Y., Martinez-Sobrido, L., Liu, W.-C., White, K. M., Chapman, M. E., Lendy, E. K., Glynne, R. J., Albrecht, R., Ruppin, E., Mesecar, A. D., Johnson, J. R., Benner, C., Sun, R., Schultz, P. G., Su, A. I., García-Sastre, A., Chatterjee, A. K., Yuen, K.-Y., and Chanda, S. K. (2020). “Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing.” *Nature*, 586(7827), 113–119.

Robinson, P. K. (2015). “Enzymes: principles and biotechnological applications.” *Essays Biochem.*, 59, 1–41.

Roy, K., Kar, S., and Das, R. N. (2015a). *A Primer on QSAR/QSPR Modeling Fundamental Concepts*. Cham: Springer International Publishing ;

Roy, K., Kar, S., and Das, R. N. (2015b). *Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment*. Amsterdam: Elsevier Academic Press.

Ruf, A., Menissier de Murcia, J., Murcia, G. de, and Schulz, G. E. (1996). "Structure of the catalytic fragment of poly(AD-ribose) polymerase from chicken." *Proc. Natl. Acad. Sci. USA*, 93(15), 7481–7485.

Sakkiah, S., Arooj, M., Kumar, M. R., Eom, S. H., and Lee, K. W. (2013). "Identification of Inhibitor Binding Site in Human Sirtuin 2 Using Molecular Docking and Dynamics Simulations." *PLoS ONE*, (C. Verma, ed.), 8(1), e51429.

Schoeman, D., and Fielding, B. C. (2019). "Coronavirus envelope protein: current knowledge." *Viol. J.*, 16(1), 69.

Schreiber, V., Dantzer, F., Ame, J.-C., and Murcia, G. de. (2006). "Poly(ADP-ribose): novel functions for an old molecule." *Nat. Rev. Mol. Cell Biol.*, 7(7), 517–528.

Schrödinger, LLC. (2015). "The PyMOL Molecular Graphics System, Version 1.8."

Sha, W., Wu, X., and Keong, K. G. (2011). "Molecular dynamics (MD) simulation of the diamond pyramid structure in electroless copper deposits." *Electroless Copper and Nickel–Phosphorus Plating*, Elsevier, 82–103.

Shah, B., Modi, P., and Sagar, S. R. (2020). "In silico studies on therapeutic agents for COVID-19: Drug repurposing approach." *Life Sci.*, 252, 117652.

Shen, D.-M., Lin, S., and Parmee, E. R. (2011). "A survey of small molecule glucagon receptor antagonists from recent patents (2006 – 2010)." *Expert Opin. Ther. Pat.*, 21(8), 1211–1240.

Shieh, W. M., Amé, J.-C., Wilson, M. V., Wang, Z.-Q., Koh, D. W., Jacobson, M. K., and Jacobson, E. L. (1998). "Poly(ADP-ribose) Polymerase Null Mouse Cells Synthesize ADP-ribose Polymers." *J. Biol. Chem.*, 273(46), 30069–30072.

Shoichet, B. K. (2004). "Virtual screening of chemical libraries." *Nature*, 432(7019), 862–865.

Shoichet, B. K., Kuntz, I. D., and Bodian, D. L. (1992). "Molecular docking using shape descriptors." *J. Comput. Chem.*, 13(3), 380–397.

Silvestrini, L., Belhaj, N., Comez, L., Gerelli, Y., Lauria, A., Libera, V., Mariani, P., Marzullo, P., Ortore, M. G., Palumbo Piccionello, A., Petrillo, C., Savini, L., Paciaroni, A., and Spinozzi, F. (2021). “The dimer-monomer equilibrium of SARS-CoV-2 main protease is affected by small molecule inhibitors.” *Sci. Rep.*, 11(1), 9283.

Singh, O., and Chakraborty, D. (2021). “Preferential binding affinity of ions and their effect on structure and dynamics of water near antimicrobial peptide.” *J. Mol. Liq.*, 344, 117789.

Singh, R., Bhardwaj, V. K., Das, P., and Purohit, R. (2022). “Identification of 11 β -HSD1 inhibitors through enhanced sampling methods.” *Chem. Commun.*, 58(32), 5005–5008.

Sliwoski, G., Kothiwale, S., Meiler, J., and Lowe, E. W. (2013). “Computational Methods in Drug Discovery.” *Pharmacol. Rev.*, 66(1), 334–395.

Smith, D. E., and Haymet, A. D. J. (1993). “Free energy, entropy, and internal energy of hydrophobic interactions: Computer simulations.” *J. Chem. Phys.*, 98(8), 6445–6454.

Spassov, D. S., Atanasova, M., and Doytchinova, I. (2023). “A role of salt bridges in mediating drug potency: A lesson from the N-myristoyltransferase inhibitors.” *Front. Mol. Biosci.*, 9, 1066029.

Stahl, M., Guba, W., and Kansy, M. (2006). “Integrating molecular design resources within modern drug discovery research: the Roche experience.” *Drug Discov. Today*, 11(7–8), 326–333.

Stewart, J. J. P. (1989). “Optimization of parameters for semiempirical methods I. Method.” *J. Comput. Chem.*, 10(2), 209–220.

Sun, Z., Wang, L., Li, X., Fan, C., Xu, J., Shi, Z., Qiao, H., Lan, Z., Zhang, X., Li, L., Zhou, X., and Geng, Y. (2022). “An extended conformation of SARS-CoV-2 main protease reveals allosteric targets.” *Proc. Natl. Acad. Sci. U.S.A.*, 119(15), e2120913119.

- Svedmyr, N. (1985). “Fenoterol: A Beta 2 -adrenergic Agonist for Use in Asthma; Pharmacology, Pharmacokinetics, Clinical Efficacy and Adverse Effects.” *Pharmacotherapy*, 5(3), 109–126.
- Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R. (1982). “A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters.” *J. Chem. Phys.*, 76(1), 637–649.
- Szabo, A., and Ostlund, N. (2018). *Modern quantum chemistry: introduction to advanced electronic structure theory / Attila Szabo, Neil S. Ostlund.*
- Tan, K. P., Singh, K., Hazra, A., and Madhusudhan, M. S. (2021). “Peptide bond planarity constrains hydrogen bond geometry and influences secondary structure conformations.” *Curr. Res. Struct. Biol.*, 3, 1–8.
- Tanford, C. (1968). “Protein Denaturation.” *Adv. Protein Chem.*, Elsevier, 121–282.
- Teague, S. J. (2003). “Implications of protein flexibility for drug discovery.” *Nat. Rev. Drug Discov.*, 2(7), 527–541.
- Terasaka, T., Kinoshita, T., Kuno, M., and Nakanishi, I. (2004a). “A Highly Potent Non-Nucleoside Adenosine Deaminase Inhibitor: Efficient Drug Discovery by Intentional Lead Hybridization.” *J. Am. Chem. Soc.*, 126(1), 34–35.
- Terasaka, T., Kinoshita, T., Kuno, M., Seki, N., Tanaka, K., and Nakanishi, I. (2004b). “Structure-Based Design, Synthesis, and Structure–Activity Relationship Studies of Novel Non-nucleoside Adenosine Deaminase Inhibitors.” *J. Med. Chem.*, 47(15), 3730–3743.
- Tian, X., Liu, Y., Zhu, J., Yu, Z., Han, J., Wang, Y., and Han, W. (2018). “Probing inhibition mechanisms of adenosine deaminase by using molecular dynamics simulations.” *PLoS ONE*, (C. M. Soares, ed.), 13(11), e0207234.

Topham, C. M., Srinivasan, N., and Blundell, T. L. (1997). "Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitution and propensity tables." *Protein Eng. Des. Sel.*, 10(1), 7–21.

Tozzini, V. (2010). "Multiscale Modeling of Proteins." *Acc. Chem. Res.*, 43(2), 220–230.

Trincavelli, M. L. (2013). "Unveiling the binding mode of adenosine deaminase inhibitors to the active site of the enzyme: implication for rational drug design: Presented by Maria P. Abbracchio." *Purinergic Signal.*, 9(1), 1–3.

Tropsha, A. (2010). "Best Practices for QSAR Model Development, Validation, and Exploitation." *Mol. Inform.*, 29(6–7), 476–488.

Tuckerman, M., Berne, B. J., and Martyna, G. J. (1992). "Reversible multiple time scale molecular dynamics." *J. Chem. Phys.*, 97(3), 1990–2001.

Tuckerman, M. E., Berne, B. J., and Rossi, A. (1991). "Molecular dynamics algorithm for multiple time scales: Systems with disparate masses." *J. Chem. Phys.*, 94(2), 1465–1469.

Tuckerman, M. E., and Martyna, G. J. (2000). "Understanding Modern Molecular Dynamics: Techniques and Applications." *J. Phys. Chem. B*, 104(2), 159–178.

Venugopal, P. P., and Chakraborty, D. (2021). "Molecular mechanism of inhibition of COVID-19 main protease by β -adrenoceptor agonists and adenosine deaminase inhibitors using *in silico* methods." *J. Biomol. Struct. Dyn.*, 1–16.

Venugopal, P. P., M, S., and Chakraborty, D. (2021). "Theoretical insights into molecular mechanism and energy criteria of PARP-2 enzyme inhibition by benzimidazole analogues." *Proteins*, 89(8), 988–1004.

Verlet, L. (1967). "Computer 'Experiments' on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules." *Phys. Rev.*, 159(1), 98–103.

Verma, J., Khedkar, V., and Coutinho, E. (2010). “3D-QSAR in Drug Design - A Review.” *Curr. Top. Med. Chem.*, 10(1), 95–115.

Vieille, C., and Zeikus, G. J. (2001). “Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability.” *Microbiol. Mol. Biol. Rev.*, 65(1), 1–43.

Vrbanac, J., and Slauter, R. (2017). “ADME in Drug Discovery.” *A Comprehensive Guide to Toxicology in Nonclinical Drug Development*, Elsevier, 39–67.

Wang, F., Chen, C., Tan, W., Yang, K., and Yang, H. (2016). “Structure of Main Protease from Human Coronavirus NL63: Insights for Wide Spectrum Anti-Coronavirus Drug Design.” *Sci. Rep.*, 6(1), 22677.

Wang, R., Cong, Y., Li, M., Bao, J., Qi, Y., and Zhang, J. Z. H. (2020). “Molecular Mechanism of Selective Binding of NMS-P118 to PARP-1 and PARP-2: A Computational Perspective.” *Front. Mol. Biosci.*, 7, 50.

Waterbeemd, H. van de (Ed.). (1995). *Chemometric Methods in Molecular Design: van de Waterbeemd/Chemometric. Methods and Principles in Medicinal Chemistry*, Weinheim, Germany: Wiley-VCH Verlag GmbH.

Wei, H., and Yu, X. (2016). “Functions of PARylation in DNA Damage Repair Pathways.” *Genom. Proteom. Bioinform.*, 14(3), 131–139.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., and Weiner, P. (1984). “A new force field for molecular mechanical simulation of nucleic acids and proteins.” *J. Am. Chem. Soc.*, 106(3), 765–784.

Worth, C. L., Preissner, R., and Blundell, T. L. (2011). “SDM--a server for predicting effects of mutations on protein stability and malfunction.” *Nucleic Acids Res.*, 39(suppl), W215–W222.

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., and Zhang, Y.-Z. (2020). “A new coronavirus associated with human respiratory disease in China.” *Nature*, 579(7798), 265–269.

Xu, Z., Yang, Z., Liu, Y., Lu, Y., Chen, K., and Zhu, W. (2014). “Halogen Bond: Its Role beyond Drug–Target Binding Affinity for Drug Discovery and Development.” *J. Chem. Inf. Model.*, 54(1), 69–78.

Xue, X., Yu, H., Yang, H., Xue, F., Wu, Z., Shen, W., Li, J., Zhou, Z., Ding, Y., Zhao, Q., Zhang, X. C., Liao, M., Bartlam, M., and Rao, Z. (2008). “Structures of Two Coronavirus Main Proteases: Implications for Substrate Binding and Antiviral Drug Design.” *J. Virol.*, 82(5), 2515–2527.

Yan, S., Appleby, T., Larson, G., Wu, J. Z., Hamatake, R. K., Hong, Z., and Yao, N. (2007). “Thiazolone-acylsulfonamides as novel HCV NS5B polymerase allosteric inhibitors: Convergence of structure-based drug design and X-ray crystallographic study.” *Bioorg. Med. Chem. Lett.*, 17(7), 1991–1995.

Yan, Y., Tao, H., He, J., and Huang, S.-Y. (2020). “The HDOCK server for integrated protein–protein docking.” *Nat. Protoc.*, 15(5), 1829–1852.

Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R., and Rao, Z. (2003). “The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor.” *Proc. Natl. Acad. Sci. USA*, 100(23), 13190–13195.

Yang, S.-Y. (2010). “Pharmacophore modeling and applications in drug discovery: challenges and recent advances.” *Drug Discov. Today*, 15(11–12), 444–450.

Yap, T. F., Liu, Z., Shveda, R. A., and Preston, D. J. (2020). “A predictive model of the temperature-dependent inactivation of coronaviruses.” *Appl. Phys. Lett.*, 117(6), 060601.

Yelamos, J., Farres, J., Llacuna, L., Ampurdanes, C., and Martin-Caballero, J. (2011). “PARP-1 and PARP-2: New players in tumour development.” *Am. J. Cancer Res.*, 1(3), 328–346.

Yerragunta, V., Patil, P., Srujana, S., Devi, R., Gayathri, R., Srujana, and Divya, A. (2014). "Benzimidazole Derivatives and Its Biological Importance: A Review." *PharmaTutor Magazine*, 2.

Yousefi, H., Mashouri, L., Okpechi, S. C., Alahari, N., and Alahari, S. K. (2021). "Repurposing existing drugs for the treatment of COVID-19/SARS-CoV-2 infection: A review describing drug mechanisms of action." *Biochem. Pharmacol.*, 183, 114296.

Yousefinejad, S., and Hemmateenejad, B. (2015). "Chemometrics tools in QSAR/QSPR studies: A historical perspective." *Chemom. Intell. Lab. Syst.*, 149, 177–204.

Zhang, H., Qiao, A., Yang, D., Yang, L., Dai, A., Graaf, C. de, Reedtz-Runge, S., Dharmarajan, V., Zhang, H., Han, G. W., Grant, T. D., Sierra, R. G., Weierstall, U., Nelson, G., Liu, W., Wu, Y., Ma, L., Cai, X., Lin, G., Wu, X., Geng, Z., Dong, Y., Song, G., Griffin, P. R., Lau, J., Cherezov, V., Yang, H., Hanson, M. A., Stevens, R. C., Zhao, Q., Jiang, H., Wang, M.-W., and Wu, B. (2017). "Structure of the full-length glucagon class B G-protein-coupled receptor." *Nature*, 546(7657), 259–264.

Zhang, L., Hao, G.-F., Tan, Y., Xi, Z., Huang, M.-Z., and Yang, G.-F. (2009). "Bioactive conformation analysis of cyclic imides as protoporphyrinogen oxidase inhibitor by combining DFT calculations, QSAR and molecular dynamic simulations." *Bioorg. Med. Chem.*, 17(14), 4935–4942.

Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., and Hilgenfeld, R. (2020). "Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors." *Science*, 368(6489), 409–412.

Zhong, N., Zhang, S., Zou, P., Chen, J., Kang, X., Li, Z., Liang, C., Jin, C., and Xia, B. (2008). "Without Its N-Finger, the Main Protease of Severe Acute Respiratory Syndrome Coronavirus Can Form a Novel Dimer through Its C-Terminal Domain." *J. Virol.*, 82(9), 4227–4234.

Zhou, J., Ji, M., Zhu, Z., Cao, R., Chen, X., and Xu, B. (2017). "Discovery of 2-substituted 1 H -benzo[d]imidazole-4-carboxamide derivatives as novel poly(ADP-

ribose)polymerase-1 inhibitors with in vivo anti-tumor activity.” *Eur. J. Med. Chem.*, 132, 26–41.

Zhou, Y., Wang, F., Tang, J., Nussinov, R., and Cheng, F. (2020). “Artificial intelligence in COVID-19 drug repurposing.” *Lancet Digit. Health*, S2589750020301928.

Zhou, Z., Yang, Z., Ou, J., Zhang, H., Zhang, Q., Dong, M., and Zhang, G. (2021). “Temperature dependence of the SARS-CoV-2 affinity to human ACE2 determines COVID-19 progression and clinical outcome.” *Comput. Struct. Biotechnol. J.*, 19, 161–167.

LIST OF PUBLICATIONS

1. **Venugopal, P. P.**, Das, B. K., Soorya, E., and Chakraborty, D. (2019). “Effect of hydrophobic and hydrogen bonding interactions on the potency of β -alanine analogs of G-protein coupled glucagon receptor inhibitors.” *Proteins*, 1- 18.
2. **Venugopal, P. P.**, M, S., and Chakraborty, D. (2021). “Theoretical insights into molecular mechanism and energy criteria of PARP-2 enzyme inhibition by benzimidazole analogues.” *Proteins*, 89(8), 988–1004.
3. **Venugopal, P. P.**, and Chakraborty, D. (2021). “Molecular mechanism of inhibition of COVID-19 main protease by β -adrenoceptor agonists and adenosine deaminase inhibitors using *in silico* methods.” *Journal of Biomolecular Structure and Dynamics*, 1–16.
4. **Venugopal, P. P.**, Singh, O., and Chakraborty, D. (2022). “Understanding the role of Water on Temperature-dependent Structural Modifications of SARS CoV-2 Main Protease Binding Sites.” *Journal of Molecular Liquids*, 363, 119867.
5. **Venugopal, P. P.** and Chakraborty, D. “Single amino acid substitution at catalytic domain of poly (ADP-ribose) polymerase 2 identifies residues essential for enzyme inhibition”. (In preparation)

Other Publications

1. Das, B. K., **Venugopal, P. P.**, and Chakraborty, D. (2019). “Computational insights into factor affecting the potency of diaryl sulfone analogs as Escherichia coli dihydropteroate synthase inhibitors.” *Computational Biology and Chemistry*, 78, 37–52.
2. Vranda Shenoy, K., **Venugopal, P. P.**, Reena Kumari, P. D., and Chakraborty, D. (2021). “Effective inhibition of mild steel corrosion by 6-bromo-(2,4-dimethoxyphenyl)methylidene]imidazo [1,2-a]pyridine-2-carbohydrazide in 0.5 M HCl: Insights from experimental and computational study.” *Journal of Molecular Structure*, 1232, 130074.

3. Kumari, D., **Venugopal, P. P.**, P. D., R. K., and Chakraborty, D. (2021). “Exploring the potential role of quercetin in corrosion inhibition of aluminium alloy 6063 in hydrochloric acid solution by experimental and theoretical studies.” *Journal of Adhesion Science and Technology*, 1–26.
4. Singh, O., **Venugopal, P. P.**, Mathur, A., and Chakraborty, D. (2021). “Temperature-Dependent Conformational Evolution of SARS CoV-2 RNA Genome Using Network Analysis.” *J. Phys. Chem. B*, 125(38), 10672–10681.
5. Singh, O., **Venugopal, P. P.**, Mathur, A., and Chakraborty, D. (2022). “Exploring the multiple conformational states of RNA genome through interhelical dynamics and network analysis.” *Journal of Molecular Graphics and Modelling*, 116, 108264.
6. Shankara, S. D., Isloor, A. M., Kudva, A. K., Raghu, S. V., Jayaswamy, P. K., **Venugopal, P. P.**, Shetty, P., and Chakraborty, D. (2022). “2,5-Bis(2,2,2-trifluoroethoxy)phenyl-tethered 1,3,4-Oxadiazoles Derivatives: Synthesis, In Silico Studies, and Biological Assessment as Potential Candidates for Anti-Cancer and Anti-Diabetic Agent.” *Molecules*, 27(24), 8694.
7. Saha, E., Chhetri, A., **Venugopal, P. P.**, Chakraborty, D., and Mitra, J. (2023). “A chemically robust amine-grafted Zn(II)-based smart supramolecular gel as a regenerative platform for trace discrimination of nitro-antibiotics and assorted environmental toxins.” *Journal of Materials Chemistry C*, 11(9), 3252–3261.
8. Singh, O., **Venugopal, P. P.**, and Chakraborty, D. “Sensitivity of the RNA Conformations in SPC/E and mTIP3P Water Models: Effect of the Hydration Shells.” *Journal of Physical Chemistry B* (Under Review).

LIST OF CONFERENCES

1. **Pushyaraga P V**, Bratin Kumar Das, Debashree Chakraborty. (2018). “Molecular Docking and DFT studies on Diaryl Sulfone derivatives as *Escherichia coli* DHPS Inhibitors”. Paper presented in “Sustainable Innovations in Green Chemistry and New Technological Developments (ICSIG-2018)” December 11- 12, Ernakulam
2. **Pushyaraga P V**, Debashree Chakraborty. (2019). “3D-QSAR, Molecular Docking, Molecular Dynamics Simulation and DFT Analysis of Cyclic Guanidine Derivatives as Glucagon Receptor Antagonists”. Paper presented in “K.V. Thomas Endowment Seminar & Second International Symposium on New Trends in Applied Chemistry (NTAC-2019)” January 14- 15, Kochi
3. **Pushyaraga P V**, Debashree Chakraborty. (2019). “Comparative Study of Stability of Drugs with Various Scaffolds as Glucagon Receptor Antagonists using Molecular Docking and MD Simulation”. Poster presented in “Accelerating Biology 2019: Towards Thinking Machines” February 5- 7, Pune
4. **Pushyaraga P V**, Debashree Chakraborty. (2020). “In *Silico* Screening and Density Functional Theory Approaches for Identification of Effective Benzimidazole-based PARP-2 Inhibitors”. Poster presented in “International Conference on Drug Discovery” February 29- March 3, BITS-Hyderabad.

WORKSHOPS ATTENDED

- Five days Faculty Development Programme on “**Computational Techniques in Chemistry & Materials Science – How to use Contemporary Software Tools** ” on 14th–17th & 19th September, 2020 organized by Department of Applied Science and Department of Material Science & Technology, School of Natural and Applied Sciences, MAKAUT, West Bengal, India in association with Islamic University of Science and Technology, Jammu and Kashmir, India.
- Two days workshop on "**Advanced Molecular Modelling and Simulation: SARS CoV-2 Case Study**" organized by Quantumzyme LLP, Bangalore, India, 29th and 30th January 2021.
- Virtual Workshop on “Molecular Dynamics Simulations for Beginners using LAMMPS” on 13th-17th September 2021.
- Hands-on training program on “Advanced Molecular Docking” organized by SIAS Research Forum, 16th-22nd October 2021.

BIODATA

Name: Pushyaraga P V

PhD Registration Number: 177048CY005

e-mail id: pushya1970@gmail.com

EDUCATIONAL QUALIFICATIONS

Examination	Institution	Board	Duration	Marks
PhD (CHEMISTRY)	National Institute of Technology, Surathkal	NITK	2017- 2022	8.67/ 10 (Course work)
MSc (CHEMISTRY)	National Institute of Technology, Surathkal	NITK	2015- 2017	8.83/ 10
BSc (CHEMISTRY)	BCM College for Women, Kottayam, Kerala	MG University, Kottayam, Kerala	2012-2015	92.75%
HSC	Lisieux English School Nadel Vaikom Kottayam Kerala	CBSE	2011- 2012	87.8%
SSC	Lisieux English School Nadel Vaikom Kottayam Kerala	CBSE	2009- 2010	9.2/ 10

ACHIEVEMENTS

- **GATE 2019** qualified
- **Best Poster Award by ACS Applied Biomaterials**, in an international symposium “3D -QSAR, Molecular Docking, Molecular Dynamics Simulation and DFT Analysis of Cyclic Guanidine Derivatives as Glucagon

Receptor Antagonists” by **Pushyaraga PV and D. Chakraborty**, selected by *ACS Applied Biomaterials* Division at 2nd International Symposium on New Trends in Applied Chemistry 2019 held on 14th-15th Jan 2019.

- **Invited lecture** on “An Introduction to Molecular Dynamics” as a part of Seven Days Online Faculty Development Programme on Online Chemistry Teaching organized by the Department of Chemistry, MES Ponnani College in association with SIAS Research Forum during 10th-16th July 2021.
- **Summer Research Fellowship by IASc- INSA- NASI** under the guidance of Dr. Amit Prakash Sharma, International Centre for Genetic Engineering and Biotechnology, New Delhi.

Place: Surathkal

Date: 12-04-2023