

**IN-SILICO DESIGN OF POTENT THERAPEUTIC
AGENTS AGAINST DIHYDROPTEROATE
SYNTHASE, SARS-COV2, AND PRE-FIBRILLAR
PRION AMYLOIDOSIS**

Thesis

Submitted in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

BRATIN KUMAR DAS

(Reg. No: 177074CY006)



DEPARTMENT OF CHEMISTRY

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA

SURATHKAL, MANGALORE-575025

SEPTEMBER, 2023

**IN-SILICO DESIGN OF POTENT THERAPEUTIC
AGENTS AGAINST DIHYDROPTEROATE
SYNTHASE, SARS-COV2, AND PRE-FIBRILLAR
PRION AMYLOIDOSIS**

Thesis

Submitted in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

BRATIN KUMAR DAS

(Reg. No: 177074CY006)

Under the guidance of

Dr. DEBASHREE CHAKRABORTY



DEPARTMENT OF CHEMISTRY

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA

SURATHKAL, MANGALORE-575025

SEPTEMBER, 2023

DECLARATION

By the Ph.D. Research Scholar

I hereby declare that the Research Thesis entitled “**In-silico Design of Potent Therapeutic Agents Against Dihydropteroate Synthase, SARS-CoV2 Spike Protein and Pre-fibrillar Prion Amyloidosis.**” which is being submitted to the National Institute of Technology Karnataka, Surathkal in partial fulfilment 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.

Bratin Kumar Das .

BRATIN KUMAR DAS

Reg.No. 177074CY006

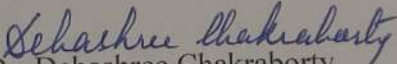
Department of Chemistry

Place: NITK, Surathkal

Date: 14,09,2023

CERTIFICATE

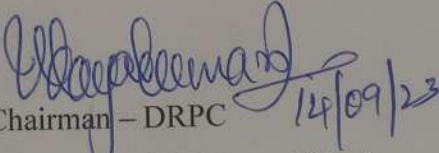
This is to *certify* that the Research Thesis entitled “**In-silico Design of Potent Therapeutic Agents Against Dihydropteroate Synthase, SARS-CoV2 Spike Protein and Pre-fibrillar Prion Amyloidosis**” submitted by **Bratin Kumar Das**, (Register Number: 177074CY006) as the record of the research work carried out by him, is *accepted as the Research Thesis submission* in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy.


Dr. Debashree Chakraborty

Research Guide 14/09/2023

(Name and Signature with Date and Seal)




Chairman – DRPC 14/09/23

(Signature with Date and Seal)

HEAD, DEPARTMENT OF CHEMISTRY
National Institute of Technology Karnataka
Surathkal, Srinivasnagar,
MANGALORE - 575 025, D. K.

ACKNOWLEDGEMENT

I would like to convey my sincere gratitude to my supervisor Prof. Debashree Chakraborty for her excellent guidance throughout my Ph.D. years. Under her supervision I have gained the necessary training to handle simulations of biomolecules at both atomistic and quantum level. I also like to thank her for continuous love, motivation, and support during the ups and downs of my personal life. I am very thankful to be a student under her.

My genuine thanks to the administrative staff for making my time spent at the institute, obstruction-free. I would like to acknowledge NITK Centre for Cyber Physical Systems (CCPS), PARAM Utkarsh CDAC-Bangalore for its excellent computational facilities and I am thankful to DST, SERB (ECR/2016/000707) for financial support. In addition, I would like to extend my special thanks to my lab mates Dilip HN, Pushyaraga PV and Omkar Singh for their insightful comments and suggestions during discussions and their constant support throughout my Ph.D. life. I would like to give special thanks to my friend Dr. Bhupendra Dandekar for his continuous support and motivation during my Ph.D. career. Finally, I would like to thank my family members, my mother Smt. Abha Rani Das, father Tulshi Charan Das, sister Chandramouli and other family members for their consistent encouragement, love, and support in the journey of my life.

Thank you!

Bratin Kumar Das

ABSTRACT

Proteins are central to all biological processes, and their interaction with themselves or other small molecules modulate their function in metabolic, cellular signalling, and immune reactions. Due to their crucial role in biological processes, protein interactions regulate the mechanism of the disease state of organisms and are often targeted by therapeutic agents. However, the major bottleneck in the process of therapeutic design is the lack of structural knowledge regarding the targeted receptors and their dynamic behavior at atomic resolution. The recent advancement in computer technologies and their amalgamation with the chemical and biological processes made a remarkable impact in reducing the cost and complexity of drug discovery processes. The reduction of complexity in computer models helps to understand the underlying interaction and stabilizing forces of drug-receptor complexes. Additionally, the *in-silico* mutagenesis approach allows for testing the efficacy of therapeutic candidates against drug-resistant mutations. The present research illustrates how ligand-based and Molecular dynamics (MD) assisted structure-based techniques can be combined to design potent therapeutics in three different protein targets, such as a bacterial Dihydropteroate Synthase, spike protein of SARS-Cov2, and abnormal peptide aggregation. It is evident from the results that the stability of small organic molecules or peptide epitopes depends on the number of polar functional groups or amino acids involved in the interaction interface. In both cases, the electrostatic energy contribution is comparatively higher than the Van der Waals energy contribution. In the case of peptide aggregation, such interactions can be replaced by small organic molecules to increase the thermodynamic barriers for inhibiting amyloidosis. The insights obtained from the present research work provide a comprehensive strategy to accelerate the highly challenging drug discovery process.

Keywords: Proteins; Therapeutic Target; Computer simulations; MD Simulation; Dihydropteroate Synthase; SARS-Cov2 spike protein; prion amyloidosis

INDEX

CONTENTS		Page No.
CHAPTER 1	<i>INTRODUCTION</i>	1
1.1	IMPORTANCE OF PROTEIN-LIGAND AND PROTEIN-PROTEIN INTERACTION IN BIOLOGICAL PROCESSES	1
1.2	ROLE OF COMPUTER SIMULATIONS IN MODELLING THE BIOMOLECULAR INTERACTIONS	3
1.3	LITERATURE SURVEY AND THE SCOPE OF THE PRESENT WORK	5
	1.3.1 Dihydropteroate Synthase (DHPS)	16
	1.3.2 Spike Protein of SARS-Cov2	17
	1.3.3 Prion Protein and Aggregation Prion Domain	18
1.4	OBJECTIVES OF THE PROPOSED WORK	20
CHAPTER 2	<i>COMPUTATIONAL METHODOLOGY AND ANALYSIS TO ASSESS BIOMOLECULAR INTERACTION AND DYNAMICS</i>	21
2.1.	CHEMOINFORMATICS METHODS FOR LIGAND- BASED DRUG-DISCOVERY	21
	2.1.1 Pharmacophore Mapping	21
	2.1.2 Quantitative structure-activity relationship (QSAR)	22
	2.1.3 Density Functional Theory (DFT)	24
2.2	STRUCTURE-BASED DRUG-DISCOVERY TECHNIQUES	26
	2.2.1 Molecular Docking	26
	2.2.2 Molecular Dynamics Simulation (MD)	29
	2.2.3 Free Energy Calculations	34
	2.2.4 Umbrella Sampling simulations	37

	2.2.5 Free Energy Perturbation	39
2.3	ANALYSIS	40
	2.3.1 Contour Plot analysis	40
	2.3.2 Root Mean Square Deviation (RMSD)	40
	2.3.3 Root Mean Square Fluctuation	41
	2.3.4 Radius of Gyration	41
	2.3.5 Number of Intra-and Intermolecular Hydrogen Bond and it's lifetime	41
	2.3.6 Secondary Structure Analysis	42
	2.3.7 Principal component Analysis	43
	2.3.8 Radial Distribution Function (RDF)	43
	2.3.8 Preferential Binding Coefficient (PBC)	43
	2.3.9 Free Energy Landscape	43
	2.3.10 Mean Square Displacement	44
	2.3.10 ADME toxicity analysis	44
	2.3.12 In-silico Vaccine Design	46
CHAPTER 3	<i>COMPUTATIONAL INSIGHTS INTO FACTOR AFFECTING THE POTENCY OF DIARYL-SULFONE ANALOGS AS ESCHERICHIA COLI DIHYDROPTEROATE SYNTHASE INHIBITORS</i>	51
3.1	BACKGROUND	52
3.2	COMPUTATIONAL DETAILS	53
	3. 2. 1 Ligand Preparation	53
	3.2.2 3D-QSAR modelling	56
	3.2.3 Molecular Docking	57
	3.2.4 DFT Calculation Setup	58
	3.2.5 Relative binding free energy calculation	59
	3.2.6 ADME/Toxicity Prediction Procedure	60

	3.2.7 System Setup for Molecular Dynamics Study	61
3.3	RESULTS AND DISCUSSION	62
	3.3.1 3D-QSAR model construction and Visualization of contour maps	62
	3.3.2 Molecular docking	68
	3.3.3 Frontier molecular orbital analysis	72
	3.3.4 FEP calculations	79
	3.3.5 ADME/Toxicity prediction of top scoring compounds	82
	3.3.6 MD Simulation	84
3.4	CONCLUSION	87
CHAPTER 4	<i>DECIPHERING THE COMPETITIVE INHIBITION OF DIHYDROPTEROATE SYNTHASE BY 8-MARCAPTOGUANINE ANALOGS: ENHANCED POTENCY IN PHENYLSULFONYL FRAGMENTS</i>	89
4.1	BACKGROUND	90
4.2	COMPUTATIONAL DETAILS	89
	4.2.1 Data-set Preparation	91
	4.2.2 Pharmacophore Mapping	93
	4.2.3 Atom-based 3D-QSAR Modelling	97
	4.2.4 Molecular Docking	98
	4.2.5 DFT calculation setup	99
	4.2.6 ADME/Tox Properties Calculation	100
	4.2.7 The MD Simulation Protocol	100
	4.2.8 Definition of Non-bonded Interactions	102
	4.2.8 Free Energy Calculation	102
	4.2.9 Umbrella Sampling	103
4.3	RESULTS AND DISCUSSION	105
	4.3.1 Ligand Based identification of structural determinants for DHPS inhibition	105

	4.3.2 Binding Mode of 8MG-compounds to the DHPS catalytic pocket	111
	4.3.3 DFT Calculation	114
	4.3.4 ADME/Tox profiling	116
	4.3.5 MD Simulation	117
	4.3.6 MM/PBSA binding free energy	127
	4.3.7 Effect of Mutation	130
	4.3.8 Drug unbinding by Umbrella Sampling	132
4.4	CONCLUSION	134
CHAPTER 5	<i>EPITOPE-BASED POTENTIAL VACCINE CANDIDATE FOR HUMORAL AND CELL MEDIATED IMMUNITY TO COMBAT SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS 2 PANDEMIC</i>	137
5.1	BACKGROUND	138
5.2	COMPUTATIONAL DETAILS	139
	5.1 Screening of Potential epitopes	139
	5.2 MD simulation Protocol	141
5.3	RESULTS AND DISCUSSION	141
	5.3.1 Screening of Epitopes	141
	5.3.2 Design of Multimeric Vaccine candidate	146
	5.3.2 Molecular Docking	154
	5.3.3 MD Simulation Study	157
	5.3.4 Free energy calculation by MM/PBSA approach	163
	5.3.5 In-silico Design of Recombinant Plasmid	164
5.4	CONCLUSION	165

CHAPTER 6	<i>SELF ASSEMBLY OF HEXADECAMERIC HUMAN PRION PEPTIDE (127-GYMLGS-132): EFFECT OF TEMPERATURE, PEPTIDE CONCENTRATION, FORCE FEILDS AND SMALL MOLECULE INHIBITOR.</i>	167
6.1	BACKGROUND	168
6.2	COMPUTATIONAL DETAILS	170
	6.2.1 Procedure for System Setup	170
	6.2.2 MD protocol	171
	6.2.3 Umbrella Sampling	174
6.3	RESULTS AND DISCUSSION	175
	6.3.1 Conformational states of monomer and influence of force fields	176
	6.3.2 Aggregate Size Distribution and Effect of NQDA in three different force-fields	178
	6.3.3 Secondary Structure Analysis	187
	6.3.4 Aggregation free energy landscape	190
	6.3.5 Accessing aggregation kinetics by Markov State model	192
	6.3.6 Mechanism of the peptide self-assembly inhibition by NQDA	196
	6.3.7 Role of NQDA on the solvation of peptides	201
	6.3.8 Umbrella Sampling for calculating dimer dissociation energy	204
6.4	CONCLUSION	206
CHAPTER 7	<i>SUMMARY AND CONCLUSIONS</i>	209
APPENDIX		215
REFERENCES		291
LIST OF PUBLICATIONS RELATER TO THE THESIS		319
OTHER PUBLICATIONS		319

CONFERENCE PRESENTATION 320

BIODATA 321

List of Figures	Page No
Figure 1.1	Multidisciplinary approach to model biological interactions 5
Figure 1.2	Crystal structure of dihydropteroate synthase (PDB ID: 1AJ0, resolution: 2.0Å) at their Substrate and Product bound state. 17
Figure 1.3	The three-dimensional structure of SARS-CoV2 spike protein. The three chains of the spike protein are coloured in red, blue, green respectively. 18
Figure 1.4.	(A) The secondary structural arrangement of cellular prion protein and the location of the shortest amyloidogenic motif (shown in blue). The alpha helices and beta-sheets are indicated as H and S respectively. (B)The amino acid sequence and 3D structure of the peptide (127-GYMLGS-132). 19
Figure 2.1	Schematic view of different docking methods. 28
Figure2.2	Flow chart for running MD simulation. 30
Figure2.3	The Thermodynamic cycle for calculating the relative free energy of binding between two states. ΔG_A and ΔG_B are the free energy difference for going unbound state to bound state. ΔG_1 and ΔG_2 are the free energy difference for transforming ligand A to the ligand B. 40
Figure 3.1	Fitness graph between observed activity versus phase-predicted activity for (A) training set (B) test set compounds with the best fit line [$y = 0.88x + 0.71$ ($R^2 = 0.81$)]. (C) Plot of actual value vs. predicted value of external test set $y=0.93x+0.16(R^2=0.74)$. (D) Plot of residual activity vs. predicted value of external test set. 64

Figure 3.2	Visualization of 3D -QSAR contour plot on compound 39 (pEII50= 5.02) Effect of A) presence of the anionic group, B) hydrogen bond donor group, C) hydrophobic group D) electron withdrawing group predicted by generated 3D-QSAR model.	67
Figure 3.3	2D-ligand interaction diagram of A) compound 1, B) compound 3, C) compound 4 and D) compound 50 in the catalytic pocket of 1AJ0.	71
Figure 3.4	The distribution of HOMO orbitals on selected sulfone drug molecules. (A) Compound 1, (B) compound 4, (C) compound 15, (D) compound 21 and (E) compound 50 respectively.	74
Figure 3.5	The location of LUMO orbitals at the pABA binding site without diaryl sulfone compounds. PH2559 and SO ₄ 284 represent the Pterin and Sulfate ion respectively. The location of LUMO orbital at the carbocation of PH2559 indicates the site prone to nucleophilic attack.	75
Figure 3.6	The distribution of HOMO, LUMO orbitals on the atom cluster containing (A) compound 1, (B) compound 4, (C) compound 15, (D) compound 21 and (E) compound 50 and their respective interacting amino acid residues.	77
Figure 3.7	Representative figures of different types of interactions captured by FEP. (A i & ii) Shows gain in favourable salt-bridge interaction due to the presence of phenoxide group (mutating compound 50 to compound 1). (B i&ii) Shows gain in favourable salt-bridge interaction due to the presence of hydroxyl	80

group (mutating compound 50 to compound 4). (C i&ii) Shows gain in favourable salt-bridge and hydrogen bonding interaction due to the presence of phenoxide ion (mutating compound 41 to compound 1). The hydrogen bond, Salt-bridge, π -cation, π -stacking and unfavourable interaction was illustrated by yellow, pink, green, sky and brown dotted line respectively.

- Figure 3.8** (A)The RMSD (\AA) of the simulated positions of 1AJ0 backbone atoms from their initial coordinate in 10 ns MD trajectory. (B) The residue wise RMSF profile of 1AJ0 illustrates the regional changes along the protein chain throughout the trajectories. 85
- Figure 3.9** Histogram represents interaction of compound 4 with different amino acids of 1AJ0 during 10 ns MD trajectory. The pink colour represents the ionic interaction, green colour represents hydrogen bond, violet colour represents hydrophobic interaction and blue colour stands for water mediated hydrogen bonding. 86
- Figure 4.1** The schematic diagram of the methods implemented in the current study. 92
- Figure 4.2** (A) The special distribution of pharmacophoric features of best score hypothesis DDHRR_1. Pharmacophore mapping of the most (B) active and (C) inactive inhibitors. (D) The distance between the central sulphur atom and the hydrophobic groups in active 8-MG compounds. 105

- Figure 4.3** Scatter plot between the experimental and predicted activity of the (A) training set, (B) test set, and (C) external test-set compounds considered in 3D-QSAR studies. (D) Scatter plot between the Residual activity (Predicted Activity-Experimental activity) and predicted activity. The 3D-QSAR contour map fitted on compound 62 ($pK_D= 6.409$) in the context of (E) hydrogen bond donor, (F) Hydrophobic group, and (G) electron-withdrawing group. 106
- Figure 4.4** The docking pose of highly active (A) Compound 62, (B) Compound 61, (C) Compound 51, (D) Compound 55, (E) Compound S1, (F) Compound S3 and inactive (G) Compound 2, (H) Compound 11. The hydrogen bonding, π - π stacking, and π -cation interactions are shown in yellow, sky, and green lines respectively. The interacting amino acids are colored in pink. 113
- Figure 4.5** The time evolution of RMSD profile for DHPS bound with (A) Compound 62, (B) Compound 61, (C) Compound 51, (D) Compound S1, (E) Compound S3, (F) Compound 6, (G) Pteric acid and (H) Apo-DHPS. The black color represents the backbone RMSD and red color represents the ligand RMSD. (Batch 1) 119
- Figure 4.6** Plot showing the RMSF profile of (A) DHPS backbone in the presence of active inhibitors, (B) inactive, Product bound, and *apo*-DHPS. (C) The time evolution of gyration radius of DHPS backbone bound with active, inactive inhibitors, pteric acid, and *apo*-DHPS. (D) The number of average hydrogen bonds between the ligand and DHPS catalytic pocket. 120

Figure 4.7	Plot represents the eigenvalues vs first 30 eigenvector indexes derived from PCA over last 40 ns of MD trajectory for ligand-bound and <i>apo</i> -DHPS structure.	125
Figure 4.8	(A) The secondary structure of the DHPS. Porcupine plots of the first eigenvector for MD simulations of DHPS bound with (B) Compound 62, (C) Compound 61, (D) Compound 51, (E) Compound S1, (F) Compound S3, (G) Compound 6, (H) Pterioic acid and (I) <i>apo</i> -DHPS.	126
Figure 4.9	The per-residue decomposition of MM/PBSA free energy for (A) Compound 62, (B) Compound 61, (C) Compound 51, (D) Compound S1, (E) Compound S3, (F) Compound 6, (G) Pterioic acid.	129
Figure 4.10	The RMSD profile for DHPS bound with (A) compound 51, (B) Compound S3. The RMSF profile for DHPS backbone bound with (C) Compound 51, (D) compound S3. The gyration radius profile for DHPS backbone bound with (E) Compound 51 and (F) Compound S3.	131
Figure 4.11	The Potential of Mean Force (PMF) profile of inhibitor unbinding from the catalytic pocket of DHPS.	133
Figure 5.1	A) The structure of spike glycoprotein in trimeric form. (B) The location of the linear B-cell epitope on the chain B of spike glycoprotein. The yellow, red, pink-, green-, orange- and firebrick-coloured regions are predicted by Emini surface accessibility, Parker Hydrophilicity, Karplus & Schulz flexibility, Kolaskar & Tongaonkar Antigenicity and Chou Fasman method respectively. (C) The location of the conformational B-cell epitopes on the chain B of spike glycoprotein in	144

their open state and closed state. The residues of conformational B-cell epitope are indicating in yellow color. (D) The R1, R2 and R3 are indicating the three different conformational epitope regions in the single chain of spike protein. The location of T-cell peptide epitopes corresponds to (E) MHC-I and (F) MHC-II. The brown, pink, olive, green, red and orange color corresponds to the peptide associated with A*02:01, A*24:02, B*40:01, B*58:01, DRB1*04:01 and DRB1*07:01 respectively.

Figure 5.3 Multiple Sequence alignment of the S-protein of different corona virus species. The highly conserved residues are marked in black and moderately conserved residues are shown in grey colour. The yellow and red highlights are showing the linear B-cell and T-cell epitopes. 146

Figure 5.3 The tertiary structure of chimeric vaccines made of (A) B-cell linear epitopes (B) T-cell linear epitopes. The amino acid sequence of the corresponding vaccines is shown below. The Adjuvant is shown in red; the adjuvant linker is shown in green and the epitope linkers are indicated in blue. 152

Figure 5.4 The non-bonded interaction of the vaccines and conformational epitopes with the immune receptor. (A) The residues of the linear vaccine involved in the formation of hydrogen bond with 7BZ5. The vaccine is shown in blue colour. The interaction of (B) R1 (red), (C) R2 (green) and (D) R3 (brown) with the variable region of 7BZ5. 155

Figure 5.5	The T-cell epitope peptide presented at the peptide presenting groove of (E) HLA-A*02:01, (F) HLA-A*24:01, (G) HLA-B*40:01, (H) HLA-B*58:01, (I) DRB1*04:01, (J) DRB1*07:01. The interacting residues of the receptor is shown in violate colour whereas the residues of vaccine or discontinuous epitope are marked in pink colour.	156
Figure 5.6	(A) The RMSD profile of the chimeric vaccine (Vac-COVID-B) (B) The gyration radius (rGyr) profile of Vac-COVID-B throughout the simulation. (C) The time evolution of RMSD of epitopes located at R1, R2 and R3. (D) The gyration radius of the conformational epitope located at R1, R2, R3. (E) The average number of hydrogen bond between the antibody and vaccine molecules.	158
Figure 5.7	The RMSD profile of the T-cell epitope peptides corresponding to (A) A*-02:01, A*-24:02, B*-40:01 and B*-58:01. (B) DRB1*04:01, DRB1*07:01. The RMSF profile of (C) MHC-I and (D) MHC-II based peptide epitopes. (E) The average number of hydrogen bond formed between the peptide epitope and MHC molecules during the simulation.	161
Figure 5.8	Restriction cloning of pUC19 vector. The red and blue coloured portion indicates the codons for the B-cell, T-cell linear vaccine respectively, inserted in pUC19 vector. The rest of the vector part is represented in black colour.	165
Figure 6.1	The free energy landscape of the prion peptide monomer based on the ϕ/ψ angles in (A) GROMOS-54a7, (B) AMBER99SB-ILDN, and (C) CHARMM-36m force field. The population of different	177

conformational states on the free energy landscapes of prion peptide correspond to (D, E, F) three different force-fields. The different colors have been used for the discretization of microstates. (G, H, I) The transition network (TN) of the prion monomer using ϕ/ψ angles as descriptors.

- Figure 6.2** The time evolution of maximum populated cluster size at different temperature, peptide concentrations for (A, B) GROMOS-54a7 force-field, (C, D) AMBER-99SB-ILDN force-field, (E, F) CHARMM-36m force-field. 180
- Figure 6.3** The dependency of the mean free path of the prion peptides with the temperature and peptide concentration in (A)GROMOS-54a7, (B) AMBER-99SB-ILDN, (C) CHARMM-36m 183
- Figure 6.4** The time evolution of maximum populated cluster size at the absence or presence of NQDA at 1:2 Peptide to NQDA molar ratio in (A) GROMOS-54a7, (B) AMBER-99SB-ILDN, (C) CHARMM-36m. 184
- Figure 6.5** MD snapshots of the prion peptide self-assembly at the prefibrillar stage in GROMOS54a7 force field [20 mM, 300K]. (B) The secondary structure arrangement of the metastable peptide aggregates. (B) The structure of aggregates in the presence of Inhibitor in three different force-fields. The secondary structures such as β -sheet, β -bridge, and coil structures are colored in yellow, olive green, and green color respectively. 186
- Figure 6.6** The secondary structure content of the peptide aggregates at different temperature and peptide concentration in (A, B) GROMOS-54a7, (C, D) 188

	AMBER-99SB-ILDN, (E, F) CHARMM-36m force-field.	
Figure 6.7	The time evolution of secondary structure elements throughout the simulation time scale in absence or presence of NQDA and the final peptide aggregated structure for (A, B) GROMOS 54a7, (C, D) AMBER 99SB ILDN, (E, F) CHARMM 36m force-field.	189
Figure 6.8	Free energy surface involving the number of interchain hydrogen bonds versus solvent accessible surface area in (A) GROMOS-54a7 (B) AMBER-99SB-ILDN, (C) CHARMM-36m. The aggregation energy landscape in presence of NQDA for (D) GROMOS-54a7, (E) AMBER-99SB-ILDN and (F) CHARMM-36m force field.	191
Figure 6.9	The decomposition of metastable states and corresponding transition paths of prion peptide aggregation for (A, B) GROMOS-54a7, (C, D) AMBER-99SB-ILDN and (E, F) CHARMM-36m force-fields.	193
Figure6.10	The residue-wise interaction energy at (A) GROMOS-54a7, (B) AMBER-99SB-ILDN, and (C) CHARMM-36m in the absence (black) or presence (red) of NQDA at 300K.	195
Figure6.11	The residue-wise interaction energy at (A) GROMOS-54a7 [20 mM], (B) AMBER-99SB-ILDN [30 mM], and (C) CHARMM-36m [50 mM] in the absence (black) or presence (red) of NQDA at 300K.	197
Figure 6.12	The time evolution of peptide-peptide hydrogen bond number in the presence or absence of NQDA (A) GROMOMS-54a7 [20 mM], (B) AMBER-99SB-	198

ILDN [30 mM], (C) CHARMM-36m [50 mM] at 300K. The black curve is showing the number of intrapeptide hydrogen bonds in pure water, and the red curve is showing the number of intrapeptide hydrogen bonds in presence of NQDA. (D) The time evolution of peptide-NQDA hydrogen bonds at different force-field at 300K. The 1-dimensional free energy landscape along the intrapeptide hydrogen bond as reaction coordinate for (E) GROMOS-54a7, (F) AMBER 99SB ILDN and (G) CHARMM 36m

Figure 6.13

(A) The intrapeptide Hydrogen bond and π - π stacking interaction between Tyrosine residues within the final oligomeric stage of prion peptides in pure aqueous solution (20 mM, 300K). (B) Replacement of hydrogen bond and Tyrosine-Tyrosine π -stacking interaction by NQDA-molecules. The peptides are illustrated in cartoon format. The β -sheet, β -bridge, coil, and turns are represented in yellow, olive green, and green color respectively. The NQDA molecule is shown in green color.

200

Figure 6.14

The pair correlation function of water molecules from the surface of the largest populated peptide cluster at (A) GROMOS-54a7 [20 mM], (B) AMBER99SB-ILDN [30 mM], (C) CHARMM-36m [50 mM] force-fields at 300K. (D-F) The preferential interaction parameter of peptides in the presence or absence of NQDA for different (D) GROMOS, (E) AMBER and (F) CHARMM force-field. The black curve is showing for pure aqueous solution and the red curve corresponds to the presence of NQDA.

202

List of Tables	Page No.
Table 2.1	The ADME parameters and their permissible range 44
Table 3.1	structural detail, experimental activities, predicted activity of sulfone compounds (1-50) 54
Table 3.2	Regression summary of generated 3D-QSAR model 63
Table 3.3	Calculated pEII ₅₀ for compounds in external test set Cpd-01-to Cpd-10 65
Table 3.4	The docking score of the compounds 1-50. 69
Table 3.5	HOMO- LUMO Energy details of Ligands at their bound state and unbound state 73
Table 3.6	Ligand based ADME and pharmacokinetic properties of top scoring sulfone compounds 82
Table 4.1	Structural details, experimental activity, and predicted activity of 8-Marcapto-guanine compounds (1-62). The superscript “t” indicates the test set compounds in 3D-QSAR modelling. 93
Table 4.2	Summary of performed MD Simulations. MD simulation calculations are run thrice to assess the reproducibility. 101
Table 4.3	The distance and angle between the pharmacophore sites of DDHRR_1 hypothesis. 107
Table 4.4	PLS statistical parameters of the generated 3D-QSAR model 108
Table 4.5	External test set compounds for validating the developed 3D-QSAR model. 111
Table 4.6	Hydrogen bond occupancy between ligand and active site residues of DHPS. UNK indicates the inhibitor and PT1 indicates Pteric acid. 122
Table 4.7	Binding Free energy (kJ/mol) of the 8MG and pteric acid at the DHPS binding pocket by MM/PBSA method (Simulation batch _1) 128
Table 4.8	Binding Free energy (kJ/mol) of the Compound 51 and S3 at the mutated DHPS binding pocket by MM/PBSA method 132

Table 5.1	The highly immunogenic Linear B-cell and T-cell epitope from Spike glycoprotein of SARS-CoVid2	142
Table 5.2	The antigenicity score of selected linear B-cell and T-cell peptides epitopes.	152
Table 5.3	Binding affinities (kJ/mol) of the vaccines towards the immune cell receptors by MM/PBSA method. The ΔG_{elec} , ΔG_{vdw} , ΔG_{sol} , and ΔG_{SASA} are indicating the electrostatic, van der Waals, polar solvation, solvent accessible surface energies respectively	164
Table 6.1	Summary of the Simulated Systems	172
Table 6.2	Average relaxation time or life time of continuous hydrogen bonds between peptide-peptide ($\langle \tau_c^{PP} \rangle$), peptide-water ($\langle \tau_c^{PW} \rangle$) and peptide-NQDA molecules. $\langle \dots \rangle$ indicates time averaging.	201

CHAPTER 1

INTRODUCTION

***Abstract:** In this chapter, a concise introduction to protein-ligand, protein-protein interaction, and its functional role in controlling various disease pathways and biological processes have been discussed. Additionally, a concise literature review, biomolecular system of interest, scope, and objectives of the present research work have been explained.*

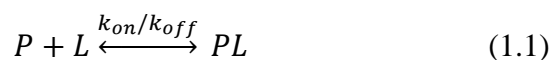
1.1 IMPORTANCE OF PROTEIN-LIGAND AND PROTEIN-PROTEIN INTERACTION IN BIOLOGICAL PROCESSES

The knowledge of protein-ligand interactions (PLIs) and protein-protein interactions (PPIs) is believed to be the core of molecular recognition and essential to all life processes. It is central in various biological processes such as enzyme catalysis, enzyme inhibition, metabolite synthesis, signal transduction, transcriptional regulation, immune reactions (antigen-antibody), etc. The successive evolution of protein function is partially dependent on the highly specific binding site, which tunes the ligand affinities (small molecule or protein subunits) as per the requirement of cells. To regulate the competing biological functions, the cooperativity in ligand binding is critically essential. The cooperative binding often leads to the conformational transition of receptors to high-affinity or low-affinity states, thus regulating the cellular processes. Therefore, understanding the thermodynamic interplays of protein-ligand or protein-protein interaction is necessary to facilitate the discovery and development of the newer generation of drugs or vaccines to combat life-threatening diseases. Moreover, the interaction of ligands can be used to switch protein functions under various physiological conditions (Dunn 2010).

In general, there are three currently existing models for protein-ligand or protein-protein interaction such as “lock-and-key,” “induced fit,” and “conformational selection” (Csermely et al. 2010; Ma et al. 1999; Tobi and Bahar 2005; Tsai et al. 1999). In the first model, both the protein and ligand are considered as rigid, and binding occurs if and only if there is a perfect match between their binding interface, i.e., the

ligand surface and the binding pocket. Thus, a perfectly sized ligand (key) can be inserted into the catalytic cavity (key-hole) of the protein (lock) to accomplish a biologically important reaction or enzyme inhibition. Unlikely, the induced fit model assumes that the protein's binding pocket is flexible and subjected to conformational change during the ligand binding. The above-mentioned model treats protein as a single, stable conformation that changes or undergoes a minor conformational change after ligand binding. However, proteins are dynamic in nature, and a vast collection or ensemble of conformational states/subsets coexist in equilibrium conditions with different populations. The conformational selection model states that ligands selectively bind the appropriate conformational states of the protein and eventually shift the dynamic equilibrium towards that state. This model also introduces the energy landscape theory of protein dynamics on conformational phase space (Henzler-Wildman and Kern 2007).

The protein-protein or protein-ligand complexes in an aqueous solution are believed to be a thermodynamic system that consists of solutes such as protein, small drug molecules, and solvents (water, ion, or various cosolvents). The binding kinetics of ligands with receptor depends on the rate of their association and can be defined as



Where PL indicates, the protein-ligand complex and k_{on} and k_{off} are the rate constant of forward and backward reactions. In such a process, the driving forces for the association and dissociation between protein-ligand are generally derived from the complex interaction and the heat exchanges between all the components of the systems, especially between protein-ligand. The such protein-ligand association is spontaneous when the system reaches equilibrium in constant temperature, and pressure and change of Gibbs free energy (ΔG) of the entire system is negative. The standard free energy (1 atm. pressure and 298K temperature) of protein-ligand binding or ΔG^0 can be defined as

$$\Delta G^{\circ} = -RT \ln K_{on} \quad (1.2)$$

R is the universal gas constant (1.987 cal. K⁻¹. mol⁻¹). The free energy can be expressed in the entropic and enthalpic contribution with the following equation:

$$\Delta G = \Delta H - T\Delta S \quad (1.3)$$

The ΔH and ΔS are the change in entropy and enthalpy of the system for the ligand binding.

The interactions of proteins with small organic molecules (ligand), other proteins, and nucleic acids are mostly governed by non-covalent interactions such as hydrogen bonding, π - π stacking interaction, π -cation interaction, and salt bridge interaction. 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 the protein. This combined force stabilizes interactions in biomolecular collisions in solutions and adhesion between cells.

1.2 ROLE OF COMPUTER SIMULATIONS IN MODELLING THE BIOMOLECULAR INTERACTIONS

It is evident from the previous section that the non-bonded interaction pattern is the key to the stabilization of biomolecular complexes and execution of life processes. This stability further depends on the flexibility of the protein binding pocket and resulting structural rearrangements upon the binding of the moiety (small organic molecules, other proteins, nucleic acids) that occurs in pico-second to microsecond time scale. Therefore, knowledge regarding such molecular events and intricate details of interactions is necessary to understand the disease pathways and design of novel therapeutic strategy against various disease states. There are various experimental techniques are available, starting from X-ray crystallography (XRD), NMR spectroscopy to electron microscopy can be employed to determine structural information of Receptor-Ligand complexes. Additionally, some of the fluorescence microscopic methods can determine the dynamic nature of the biomacromolecules. However, such experimental techniques are time consuming to study the interaction of large ligands libraries (small molecules, peptides etc.) for identifying potent therapeutics.

In this regard computational techniques can accelerate the process of identifying crucial interactions to design effective drug candidates against various bacterial, viral and neurogenerative diseases. The implementation of computers algorithms and

architectures in drug design was first established in 1980 by Merck. Since then, various In-silico methods have been developed to visualize the biomolecular interactions in computer screen and calculate different properties such as free energy, kinetic measures or other macromolecular quantities that can be directly related to experimentally determined observables. The first step towards computer aided molecular modelling is the optimization of protein i.e., the receptor structure to ensure the atomic detailing and the correct protonation of amino acids at binding site. Therefore, the crystallographic structures obtained from XRD or MNR are refined by adding missing amino acid side chain and minimization to remove steric clash. In, the absence of the proper crystal structure of the receptor, homology modelling can be used to design 3D-structure of the protein from other structural homologous of the protein. Further, to optimize the ligand structure, semi-empirical or DFT based quantum techniques are used to ensure correct bond length, angle, torsions, and assignment of the correct partial charges to each atom of the ligand. Various In-silico techniques have been developed such as quantitative structure activity relationship (QSAR), pharmacophore modelling to identify the functional group substitution or spatial arrangement of structural features that ultimately enhance the activity of the drug candidates. Moreover, the quantitative structure property relationship (QSPR) or ADME toxicity algorithms helps to screen hit molecules that are harmful to human body. However, the major bottleneck of the above-mentioned ligand-based techniques is the prediction of non-bonded interaction that governs the stability of ligands at the catalytic pocket of the receptor protein. In this regard, molecular docking-based techniques have been developed which can predict the conformational states for thousands of ligands and their binding affinity for multiple target proteins. Induced fit or ensemble docking algorithms have been developed which can identify the partial conformational drift of enzyme catalytic pocket. Nonetheless, docking algorithms are limited in sampling major conformational interchanges that take place at the receptor binding site during ligand binding and ensuring their dynamic stability. Molecular dynamics simulations are plausible approach to encounter structural flexibility of both ligand and receptor at their bound and unbound states. In recent years the advancement and innovation in computer hardware infrastructure, development of Graphical Processing unit (GPU) and advent of modern MD software allow to generate long trajectories to depict the kinetic profile

or the energy landscape correspond to molecular recognition. Additionally, the MD based enhanced sampling techniques can explain the slow biological processes or rare events such as ligand unbinding kinetics and estimate thermodynamic barriers efficiently. Overall, computer simulations can help to limit expensive chemical synthesis or biological testing and thereby decrease the need for traditional resources in fast-paced process of therapeutic design. The contribution of different areas of science to model biological process is presented in **Figure 1.1**.

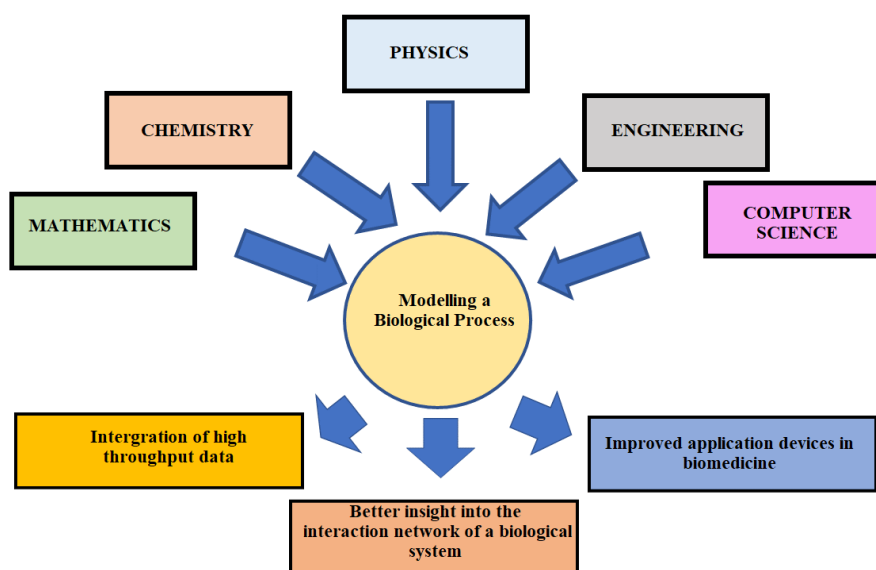


Figure 1.1. Multidisciplinary approach to model biological interactions.

1.3 LITERATURE SURVEY AND THE SCOPE OF THE PRESENT WORK

In the present section, a brief review of the literature has been discussed which focuses on the implication of advanced computational methods toward identifying lead compounds and vaccine candidates against target receptors. The advancement in computer architecture, better algorithms, and the availability of efficient software allow the computation of structural and dynamic properties of several biomolecular assemblies which minimize the huge experimental expense and risk of rejection of therapeutic candidates.

The following literatures are focusing on the implementation of ligand-based strategies to identify crucial atomic arrangement and structural counterparts responsible for biological activity.

- *Leonard et al* developed a QSAR model that can predict the important functional group substitution for three different scaffolds for HIV inhibitory activity. It is found from their results that k-means clustering algorithm on the factor scores of the small molecule descriptor is reliable method for dividing the dataset in to training and prediction set(Leonard and Roy 2006b).
- *Cheng et al.* studied the efficiency of virtual screening of hit compounds based on Pharmacophore model and molecular docking to target eight structurally diverse proteins such as D-alanyl-D-alanine carboxypeptidase, dihydrofolate reductase, thymidine kinase, estrogen receptors, androgen receptors, acetylcholinesterase, HIV-1 protease etc. The results obtained from their study indicate that pharmacophore-based screening method outperform docking based methods to screen potent hit compounds. This case study suggests to pay more attention on the ligand-based techniques in virtual screening in drug discovery process(Chen et al. 2009).
- *Lu et al.* successfully designed pharmacophore and docking guided 3D-QSAR model for formamides derivatives as an antimicrobial agent to block bacterial Enoyl acyl carrier protein reductase. The important functional group substitution at 3D-descriptor space was ensured by comparative molecular field analysis (CoMFA) and comparative molecular indices analysis (CoMSIA). The statistical robustness of the model was validated from the correlation coefficient obtained for training set compounds. It is evident from their results that site-specific incorporation of bulky hydrophobic moiety increases the activity of the ligands(Lu et al. 2012).
- *Martin et al.* worked on the development of methods for dividing training set and test set data for building predictive QSAR model. To

validate a QSAR model in absence of true external test set, statistical external validation is generally used. In such circumstances ration division methods like k-nearest neighbour method (kNN), random forest and hierarchical clustering are better in generating statistically sound QSAR models(Martin et al. 2012a).

- *Vuorinen et al.* developed ligand-based pharmacophore model to screen compounds that can inhibit 17 β -Hydroxysteroid dehydrogenase 2 (17 β -HSD2). This pharmacophore model was employed to screen a database containing 1381 hit compounds and 29 compounds were selected for further in-vitro evaluation. The pharmacophore based virtual screening yields seven compounds that exhibit good biological activity against 17 β -HSD2. Furthermore, two scaffolds, sulfonate and phenylbenzenesulfonamides were obtained which can be used for drug repositioning to inhibit this enzyme.
- *Martinez et al.* constructed 2D-and 3D QSAR models by using a series of experimentally tested inhibitors correspond to dihydrofolate reductase. In both the QSAR model it is found that the biological activity of the inhibitors is mainly influenced by the electronegativity and the size of the functional group substituents(Garro Martinez et al. 2017).
- *Almeida et al.* proposed promising inhibitors of MARK3 protein kinase that expressed in head and neck cancer based on ligand-based screening approaches such as physicochemical properties, overlapping molecular interaction fields and toxicity predictions (Almeida et al. 2014).
- *Peng et al.* aimed to build a large scale, comprehensive and statistically robust 3D-QSAR model that can predict the activity of structurally diverse scaffold targeted for inhibiting Sigma 1 receptor (S1R). Further, the model underwent external test set validation followed by database screening of natural drug compounds. The virtual screening yields two FDA approved drugs (phenyltoloxamine and diphenhydramine) that exhibit low inhibitory constant for S1R receptors (160 nM, 58 nM respectively). This strategy can be used to design potent therapeutic for neuropathic pain(Peng et al. 2019a).

- *Alam et al.* developed a field point-based 3D-QSAR mode to screen 80000 flavonoid compounds and 1480 compounds with activity less than 5 μ M have been screened. Further docking and ADMET toxicity analysis were conducted to select top 25 compounds as a potential inhibitors for tankyrase (Alam and Khan 2019).
- *Pushyara et al.* studied the binding of β -alanine derivatives to the transmembrane allosteric site of the G-protein coupled glucagon receptor (GCGR). Combination of 3D-QSAR, molecular docking, molecular dynamics calculations yield potent compounds that have higher affinity towards GCGR. This study can be a viable initiation point to start drug discovery campaign against type -II diabetes (Venugopal et al. 2020).

The subsequent literatures are focusing on structure-based drug discovery approaches where the non-bonded interactions of ligands or peptides with the receptor binding pocket are the main concern. Further, literature review regarding the application of enhanced sampling approaches on predicting ligand unbinding pathways were also depicted.

- *Osguthorpe et al.* used energy minimization technique to study the binding of trimethoprim at the binding pocket of dihydrofolate reductase. Both all atom force field and united-atom force field were used in this study. Finally, the role of specific residues in ligand binding, interaction energy and entropy loss of ligand were examined. Water molecules, those are not in direct contact with the ligand, were found to have significant interaction energies with the ligand. Thus, the inclusion of at least one shell of waters is be vital for accurate simulations of enzyme complexes (Dauber-Osguthorpe et al. 1988).
- *Swegat et al.* studied protein-ligand binding and unbinding process through MD simulation studies. Phenolic compounds were generally used as bacteriostatic preservatives in pharmaceutical. This article showed how such phenol compounds can be used as a stabilizer for synthetic insulin. It is evident from their study that phenol compounds

mostly interact with His 45, Cys 66, Ser69, Ile610, cys611 and their interaction is local rather than being cooperative (Swegat et al. 2003).

- *Jiao et al.* investigated the binding of charged ligands benzamidine and diazamidine to trypsin employing a polarizable potential energy function and explicit-water molecular dynamics simulations. The binding free energies were computed from the difference between the free energies of decoupling the ligand from water and protein environments. Both the absolute and the relative free energies from the perturbation simulations agree with experimental measurements within 0.5 kcal.mol⁻¹. Comparison of free-energy components sampled from different thermodynamic paths indicates that electrostatics is the main driving force behind benzamidine recognition of trypsin. Overall, their results suggest that the use of a polarizable force field, along with adequate sampling, is necessary for achieving chemical accuracy in molecular simulations of protein–ligand recognition (Jiao et al. 2008).
- *Shivakumar et al.* calculated the solvation free energy of a set of 239 molecules, spanning diverse chemical functional groups commonly found in drugs and drug-like molecules. In this study three different force fields, general AMBER force field (GAFF) with AM1-BCC charges, CHARMM-MSI with CHelpG charges and OPLS_2005 were compared to evaluate their efficiency in calculating the solvation free energy. It was found that OPLS_2005 exhibited high correlation with experimental studies (Shivakumar et al. 2009).
- *Iribarne et al.* studied the repurposing or reposition of phenothiazine derivatives as glutathione reductase (GR) to design antiparasitic lead compounds. Due to the emergence of drug resistance in Trypanothione reductase, the structural homologue GR can be targeted. Molecular docking followed by molecular dynamics were employed to assess the key features of the analogues to inhibit GR. It is evident from their studies that positively charged phenothiazine derivatives have higher affinity towards the catalytic pocket (Iribarne et al. 2009).

- *Jiang et al.* aimed to improve the accuracy in free energy computation by applying free energy perturbation with replica exchange molecular dynamics (FEP/REMD). In this study, FEP/REMD method was extended and combined with an accelerated MD simulations method based on Hamiltonian replica-exchange MD (H-REMD). In this strategy, each system with a given thermodynamic coupling factor λ in the extended ensemble is further coupled with a set of replicas evolving on a biased energy surface. Boosting potentials used to accelerate the interconversion among different rotamer states of the side chains in the neighbourhood of the binding site. These combinations overcome the additional problems arising from the existence of kinetically trapped conformations within the protein receptor. Exchanges are allowed to occur alternatively along the axes corresponding to the thermodynamic coupling parameter λ and the boosting potential. As an case study, the absolute binding free energy of *p*-xylene to the nonpolar cavity of the L99A mutant of the T4 lysozyme was calculated. The tests indicate that the hybrid REMD scheme greatly accelerates the configurational sampling of the amino acid side chains at the binding pocket, thereby improving the convergence of the FEP computations (Jiang and Roux 2010).
- *Sakkiah et al.* studied the interaction of small molecule inhibitors against Sirtuin 2 (SIR2) receptor which plays crucial role in the deacetylation of Lysin residues on histones or different type of proteins. Five different inhibitors such as sirtinol, suramin, mol_6, nf725 and 67 underwent molecular docking followed by molecular dynamics and MM/PBSA calculation. The overall summary of the study indicates that van-der walls components of the binding energy mainly govern the stability of the inhibitors at the binding pocket(Sakkiah et al. 2013).
- *Patel et al.* studied the unbinding of inhibitor from the catalytic pocket of CDK5 by steered molecular dynamics simulation. The unbinding free energy of the inhibitor compounds are found to by good agreement with

experimental IC₅₀ values. This study serves as a plausible paradigm towards employing SMD as structure-based drug discovery approach.

- *Zhou et al.* studied the binding, unbinding kinetics of ligand from three types of G-quadruplexes (basket, propeller, and hybrid) by umbrella sampling simulation studies. It is evident from their study that coulombic interaction is the driving force toward the stability of G-quadruplexes during the unbinding process. The dissociation free energy is found to be in good agreement with experimental results (Zhou et al. 2015a).
- *Balajee et al.* identified inhibitors that have higher affinity towards NS2B and NS3 by virtual screening to designing antiviral therapeutics. Molecular docking and MD simulation studies suggested that inhibitors are mainly stabilized by His51, Ser135, Asp75, Gly133 of the catalytic triad by forming NH—O, OH—N and OH—O hydrogen bonding. The binding affinity of highly potent compound is found to be -74 kcal/mol which is in good agreement with experimental IC₅₀ value (Balajee et al. 2016).
- *Cholko et al.* studied the inhibition of cyclin dependent kinase-Cyclin C enzyme system (CDK8-CylC) by small organic molecules to design anticancer therapeutics. Both Type-I and Type -II inhibitors were used to study the inhibition mechanism and the results obtained from unbiased MD simulation study indicates that Van der Waals energy component is the main contributor for both the types of inhibitors (Cholko et al. 2018).
- 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). Understanding the molecular level mechanism could propose novel insights/hypotheses which serve as a critical step for translational medicine research.

The following literature are based on the computational studies of epitope identification and design of therapeutic enzyme or vaccine candidates.

- *Kumar et al.* identified the epitopes of E5 protein that express in the early phase of human papilloma virus (HPV) infection in cervical cancer. Immunoinformatics approach was adopted to predict the peptide fragments which show affinity towards MHC-II receptors. This study yields a peptide sequence, IPLFLIHTHARFLIT which have highest affinity towards MHC-II correspond to HLA-DRB1*14:01 allele(Kumar et al. 2015).
- *Shi et al.* employed Immunoinformatics assisted screening of B-cell and T-cell epitopes to design vaccine specific to MERS-CoV infection. It is evident from their studies that nucleoprotein of MERS-CoV can be better candidate as an immunogen compared to spike protein in terms of conservancy and eliciting immune response in host body. Total fifteen linear B-cell, ten conformational epitope and ten helper, cytotoxic T-cell epitopes were identified which can be used to design vaccine candidate against MERS-CoV pandemic.(Shi et al. 2015a)
- *Dar et al.* predicted the T-cell epitopes present on the polyprotein of Zika virus to design potential vaccine candidates by bioinformatics tools. The peptide correspond to MHC-I is found to located on NS5 (non-structural protein) and envelop protein whereas MHC-II peptides are mostly located to NS1 and NS2. Further the obtained peptides underwent affinity, promiscuity and antigenicity test for the formulation of vaccine candidates against zika virus infection(Dar et al. 2016).
- *Hundal et al.* developed an open-source algorithm namely pVAC-seq that can identify neoantigens by processing the DNA or RNA sequencing data from the tumour specific mutational repository. The identified epitopes can be used as reference to design anticancer peptide candidates that can have higher affinity towards MHC-I receptors and elicit proliferation of tumour specific CD8+ T-cells by destroying the somatic tumour cells (Hundal et al. 2016).

- *He et al.* aimed to predict antigenic B-cell and T-cell epitopes located on the lectin protein that causes severe allergic reaction in human body. Both the linear B-cell and T-cell epitopes were subjected to molecular docking calculation to screen the epitopes based on their affinity towards respective immune receptors. Finally, four B-cell epitope and two T-cell epitopes were identified and underwent *in vitro* studies to design effective antibody for lectin toxicity(He et al. 2018).
- *Jain et al* used bioinformatic algorithms and molecular docking studies to screen peptides that contain multiple immunogenic epitopes in Ebola virus glycoprotein. Care was taken to select the epitopes that bind wide variety of HLA alleles. Four different peptides such as P1, P2, P5 and P6 was identified which showed remarkable population convergence to be presented at MHC receptor(Jain and Baranwal 2019).
- *Baral et al.* identified immunogenic epitopes located on the glycoprotein of Lassa virus that causes severe haemorrhagic fever in human. The consensus screening of the predicted epitopes yields thirty new B-cell epitopes which lacks experimental evaluation. Molecular Docking and MD simulation studies were employed to assess the stability of epitope-receptor complex. The results obtained from this study indicates that epitopes related to MHC-I allele are mostly stable by higher number of hydrogen bonds. Among the MHC-I epitopes 233-FSRPSPIGY-241 shows higher affinity towards peptide binding groove of MHC molecule(Baral et al. 2020).
- *Nelapati et al.* designed Uricase mutein by replacing the antigenic hotspots located at the wild type bacterial Uricase. The mutation found to increase the affinity of uric acid towards Uricase catalytic pocket. Further, molecular dynamics simulation was employed to ensure the structural stability of muteins after mutation.

The subsequent bibliography aims towards studying the aggregation mechanism of intrinsically disordered peptide and possible therapeutic strategies.

- *Katyal et al.* employed explicit-solvent-all-atom molecular dynamics simulation to study the effect of trehalose molecules in prion peptide

(GNNQQNY) aggregation. The trehalose molecules found to increase the kinetic barrier by trapping small size oligomers (dimer to tetramer) to restrict the evolution of large size oligomer with higher β -sheet content. The faster expulsion of hydration water by trehalose molecule compared to peptides decrease the intra-peptide side-chain and main chain interaction to disassemble toxic oligomers(Katyal and Deep 2017)

- *Zhou et al* studied the effect of Lanosterol on breaking the steric zipper fibril of amyloid β peptide. It is found from their simulation studies that Lanosterol molecules are mainly stabilize by hydrophobic interaction with Phe20 and Phe19. This compound found to have higher activity in suppressing the amyloid toxicity compared to cholesterol due to enhanced hydrophobicity. Further, the inhibitory effect of the same molecule was validated by AFM imaging and ThT fluorescence assay(Zhou et al. 2019a).
- *Pal et al* studied the oligomerization of $\alpha\beta$ 16-22 peptides in aqueous solvent in the presence or absence of ATP molecules with three variants of AMBER force-field. The presence of ATPs found to replace the inter peptide π -stacking interaction by itself to inhibit the formation of antiparallel β -sheet during aggregation process(Pal and Paul 2020).
- *Nie et al.* aimed to identify the mechanism of polyphenol such as gallic acid to stop the interconversion of nontoxic helical state to toxic beta-sheet state. The polar interaction of gallic acid found to be responsible for arresting the interconversion in aqueous medium(Nie et al. 2020).
- *Roy et al* studied the effect of three different compounds such as norepinephrine, benzimidazole and aspirin molecules on the aggregation of human islet amyloid peptide that causes Type II Diabetes in human body. Among those small molecules norepinephrine is found to reduce the β -sheet percentage between the residue stretches Leu12- His18 and Leu27-Gly33. In the same line they also study the effect of amyloid core derived peptides. It is evident from their result that peptide containing D-amino acids have greater efficiency in disrupting the fibril

formation in full length human islet amyloid peptides compared to L-amino acid containing peptide (Roy and Paul 2020, 2021).

- *Kalipillai et al.* studied the charge effect essential for inhibiting the β -hairpin conversion of amyloid β peptide in gold nanoparticle. The results obtained from MD simulation suggest that negatively charged nanoparticles or slabs are better choice compared to positively charged one for shifting the dynamic equilibrium of β peptide from β -sheet to coil or helical state. This study demonstrates excellent therapeutic strategy to cure amyloidosis by nanomaterials (Kalipillai and Mani 2021).
- *Sorout et al.* carried out multiscale molecular dynamics simulation to study the effect of boron nitride nano tubes and sheets with different surface curvature on the conformational transition of amyloid peptide $\alpha\beta$ 1-42. The formation of β -sheet conformation is believed to be the indication of toxicity in aqueous solution. The highest curvature nano tube found to shift the conformational equilibrium to α -helical state whereas the intermediates found to stabilize the peptide in random coil state. This study indicates the role of nanoparticles in decreasing the toxicity of intrinsically disordered peptides by various pathways (Sorout and Chandra 2021).
- *King et al.* applied molecular dynamics simulation to study the effect of five flavonoid compounds such as Morin, Myricetin, Epicatechin, Quercetin and Dihydroquercetin on the aggregation of Islet amyloid polypeptide. It is evident from their study that oligomer formation and emergence of is essentially driven by inter peptide Phe23 interaction. The flavonoid compounds mainly disrupt the π -stacking interaction to arrest the evolution of toxic oligomers.

The above literature survey provides a plausible paradigm towards the usage of state of the art *In-silico* techniques in different areas of translational pharmaceutical research such as enzyme inhibition or vaccine development. From this viewpoint, three different biomolecular system have been considered based on their role in controlling bacterial, viral, and neurodegenerative diseases. In the subsequent sections a brief discussion on the systems of interest and their scope in therapeutic development have been conferred.

1.3.1 Dihydropteroate Synthase (DHPS)

Dihydropteroate synthase (DHPS) is known to be a validated drug target to block folate production in bacterial cells (Domagk 1935). This particular enzyme catalyses the condensation between 7,8-dihydropterine pyrophosphate (DHPP) and *P*-aminobenzoic acid (*p*ABA) to produce 7,8-dihydropteroate, a precursor of tetrahydrofolate (Birmingham Alun and Derrick Jeremy P. 2002; Griffin and Brown 1964). Folate analogs like tetrahydrofolate reported to be a crucial co-factor for synthesizing essential amino acids, nitrogen bases and play an important role in reaction associated with one carbon transfer (Bourne 2014; Rossi et al. 2011). Higher eukaryotic cells consume folate from dietary sources due to the presence of membrane-associated folate transport protein (Matherly and Goldman 2003; Whetstine et al. 2002). On contrary, all bacterial cells and some lower pathogenic eukaryotes lack the folate transport protein hence compelled to synthesize folate *de novo* (Levin et al. 2004). Therefore, the growth of bacterial cells can be prohibited by blocking folate biosynthetic pathway. Few enzymes in the *de novo* pathway, particularly DHPS are not human homologs, therefore makes them a promising target and topic of intense research for developing antimicrobial agents to reduce the mortality of HIV/AIDS and COVID-19 patients (Arooj et al. 2013; Shaw et al. 2014; Suthar et al. 2015; Tari 2012; Zhang et al. 2021).

DHPS is reported to be encoded by *folP* gene (Sánchez-Osuna et al. 2019) and the entire polypeptide chain is folded into the most commonly found TIM barrel-like structure (β/α)₈ in which eight α -helices bundles around the inner cylinder made of eight parallel β -strands (Babaoglu et al. 2004; Baca et al. 2000; Morgan et al. 2011; Yogavel et al. 2018). The N-terminal pole of the α/β barrel is interconnected by rigid loops while the opposite pole consists of extended and flexible loops (Achari et al. 1997a), crucial for ligand recognition and binding (**Figure 1.2**). The active site of DHPS monomer is located at the C-terminal end of the β -barrel which comprises of three conserved sub-sites: the pterin binding site, the *p*ABA binding site, and the anion binding site. The pterin binding pocket is situated in a deep cleft of the C-terminal end of β -barrel, whereas the *p*ABA binding pocket is located at the surface and comprises of two flexible loops (loop1 and loop2) (Achari et al. 1997b; Hampele et al. 1997). Initially,

DHPP binds with DHPS active site and it catalyses the slow release of pyrophosphate moiety from DHPP. The released pyrophosphate moiety plays an important role in stabilizing the conformation of flexible *p*ABA binding loops (Yun et al. 2012a). After *p*ABA binding, the condensation of DHP^+ and *p*ABA occurs through S_N1 reaction to form dihydropteroate. The bi-substrate pocket of DHPS is therefore the target of the substrate or product antagonists to block the folate pathway by producing dead-end conjugates.

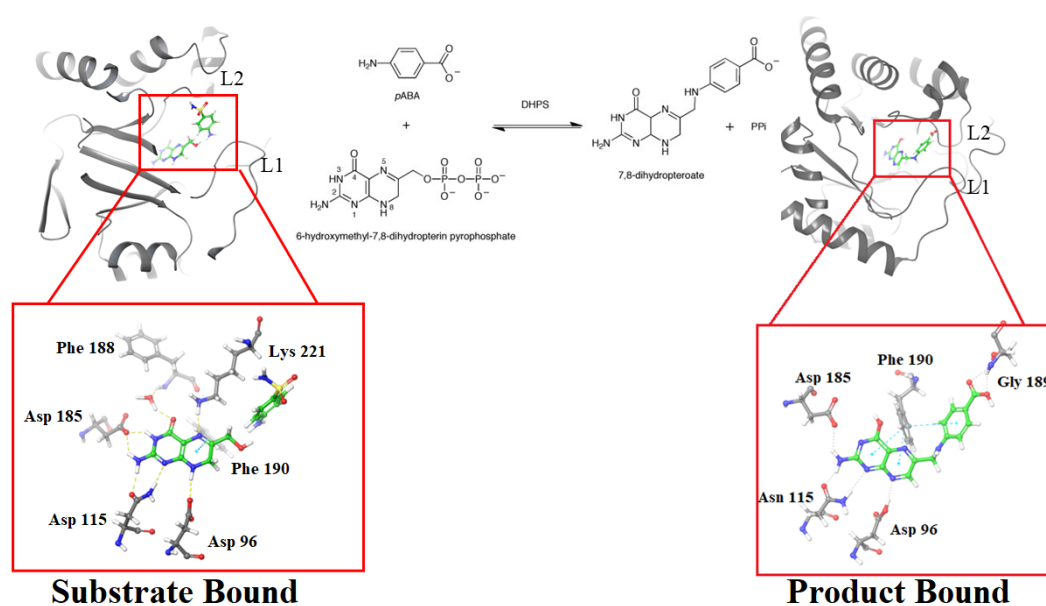


Figure 1.2. Crystal structure of dihydropteroate synthase (PDB ID: 1AJ0, resolution: 2.0Å) at their Substrate and Product bound state.

1.3.2 Spike Protein of SARS-Cov2

The specific interaction of antigens such as viral proteins with the receptors of humoral and cell mediated defence mechanism are crucial to design multimeric vaccines for building adaptive immunity against life-threatening viral diseases. The genome of SARS-CoV-2 is a single stranded positive sense RNA and reported to be largest viral genome till date (Cui et al. 2019; Fehr and Perlman 2015). This RNA genome encodes four poly proteins of SARS-Cov2, like membrane protein (M), nucleocapsid protein (N) and envelop protein (E). The genome entry in the host cell is guided by crown-shaped glycosylated spike protein (S-protein) located at the envelop surface (**Figure 1.3**). The S-protein is a trimeric class I fusion protein which exists in a metastable

prefusion conformation and undergoes a dramatic structural rearrangement to fuse into the host cell membrane(Bosch et al. 2003; Li 2016; Wrapp et al. 2020). The ectodomain of S protein includes the receptor binding S1-sub unit and the trimeric membrane fusion S2 or stalk domain(Shang et al. 2020; Walls et al. 2020; Yan et al. 2020). The S1 subunit receptor-binding domain (RBD) specifically interacts with the host receptor known as Angiotensin-converting enzyme 2 or ACE 2(Yuan et al. 2020). At the moment of binding of S1 domain to a host-cell receptor, the prefusion trimer is destabilized, which results in shedding of the S1 subunit, and transition of the S2 subunit to a stable post fusion conformation (Gui et al. 2017; Walls et al. 2020; Wrapp et al. 2020). The study of crucial activity of the S-protein can therefore provide a breakthrough in vaccine design and development compared to other structural proteins of SARS-CoV-2 for inducing specific immune response against this virus.

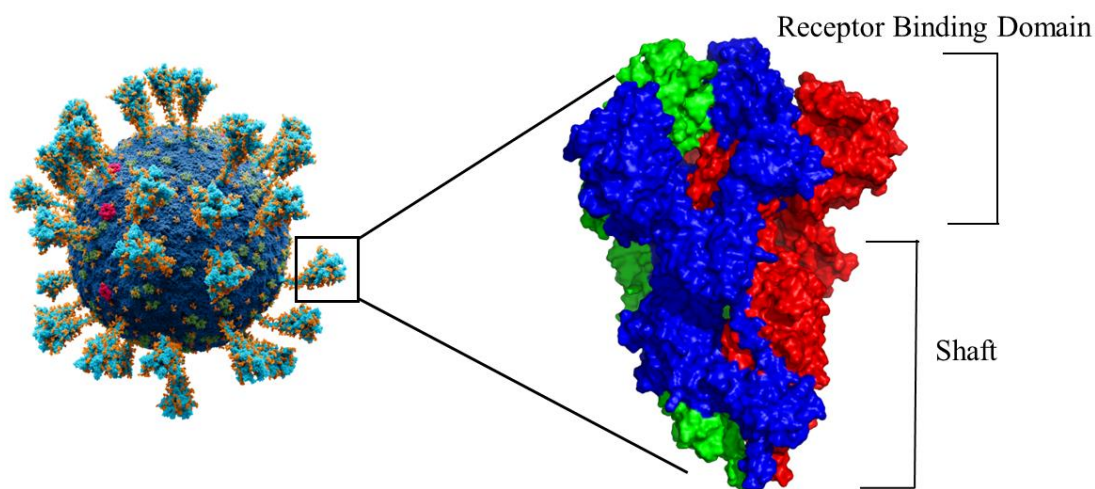


Figure 1.3. The three-dimensional structure of SARS-CoV2 spike protein. The three chains of the spike protein are coloured in red, blue, green respectively. (source)

1.3.3 Prion Protein and Aggregation Prion Domain

Prion is a cell surface protein (**Figure 1.4**) and is dominantly expressed in the peripheral and central nervous systems(Bendheim et al. 1992). The primary functions of this protein include cell adhesion, synapse formation(Santuccion et al. 2005), regulation of circadian rhythm(Tobler et al. 1996), neuroprotection(Guillot-Sestier et al. 2009;

Roucou et al. 2004), controlling ion homeostasis, cell signalling,(Lewis and Hooper 2011), etc. It is known that the transformation of the normal cellular prion isoform (Prp^C) to protease-resistant scrapie isoform (Prp^{Sc}) is the initial point of prion disease. Following the misfolding, an exponential increase in the infected prion protein (PrP^{Sc}) population occurs in an autocatalytic manner and eventually organized mature fibril forms. On the other hand, a large subset of the latest evidence shows that the aggregation of peptides derived from normal prion protein can be another causative agent of prion disease.

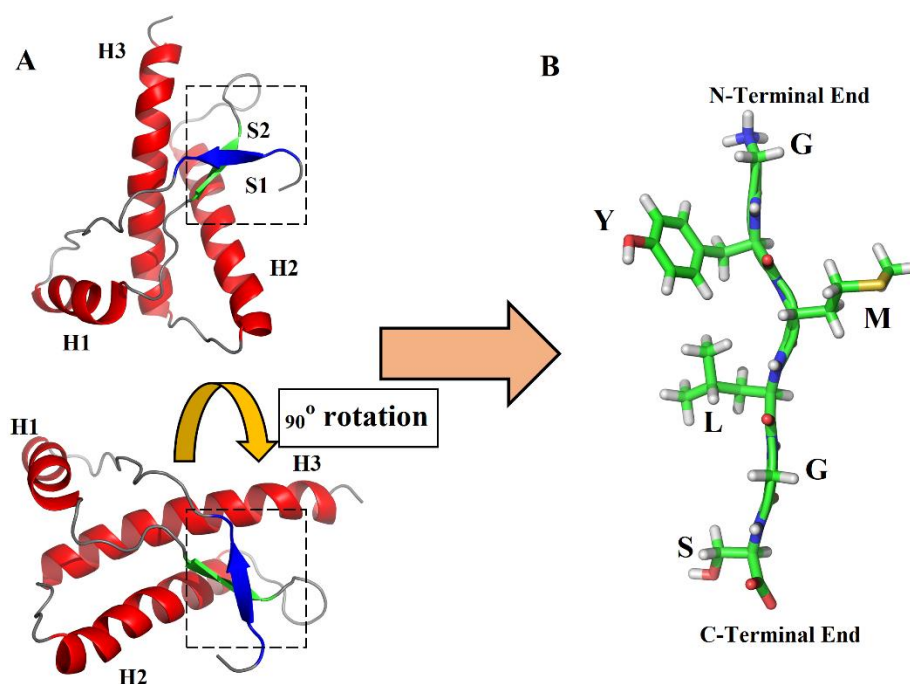


Figure 1.4. (A) The secondary structural arrangement of cellular prion protein and the location of the shortest amyloidogenic motif (shown in blue). The alpha helices and beta-sheets are indicated as H and S respectively. (B) The amino acid sequence and 3D structure of the peptide (127-GYMLGS-132).

Thus, the pre-fibrillar counterparts of such proteinaceous agents are the primary toxic species (Laganowsky et al. 2012; Larson and Lesné 2012) which consist of lower-order oligomers of abnormally folded proteins (Chamachi and Chakrabarty 2016) or assembly of small peptides derived from misfolded proteins (Apostol et al. 2013; Balbirnie et al. 2001; Gallagher-Jones et al. 2018; Nelson et al. 2005a; Nelson and Eisenberg 2006). Recent reports suggest that the appearance of the β -sheet or the zipper-like structures

of pre-fibrillar aggregates are the indication of cytotoxicity(Chatani et al. 2014; Larson and Lesné 2012; Lesné et al. 2006; Walsh et al. 2002) and are considered as the main culprits for nerve cell disruption(Jang et al. 2013). Modelling the aggregation of such lower-order soluble oligomers, their structural and energetic profile remains unclear in the context of prion toxicity. This small size pre-fibrillar aggregates can be targeted by small organic compounds to disintegrate them at the pre-fibrillar stage.

1.4 OBJECTIVES OF THE PROPOSED WORK

Based on the detailed literature review and above-mentioned scopes, the following objective

- Identification of factor affecting the potency of sulfonamide derivatives on antagonist-mediated inhibition of Dihydropteroate synthase and the prediction of possible sulfa-resistant mutations at the active site.
- Determining the resilience of 8-Mercaptoguanine compounds in product antagonism and circumventing the drug resistant mutations at DHPS catalytic pocket.
- Identification of highly potent and conserved antigenic epitopes in the spike protein of SARS-CoV2 and their interaction with immune receptors to design effective vaccine for hard immunity.
- Identification of the metastable lag-phases and thermodynamic barriers of pre-fibrillar prion peptide aggregation and the effect of naphthoquinone dopamine on the aggregation barriers.

CHAPTER 2

COMPUTATIONAL METHODOLOGY AND ANALYSIS TO ASSESS BIOMOLECULAR INTERACTION AND DYNAMICS

Abstract: *This chapter includes a detailed outline of the different in-silico approaches to study the dynamics related to biomolecular associations at atomic resolution. In the case of protein-ligand complexes, cheminformatics-based and structure-based strategies have been implemented to identify crucial interactions for the stability of such associations. Structure-based and free-energy techniques were applied to assess the affinity of the immunogenic protein segments with the immune receptors of the human body. Immunoinformatics approaches were employed to screen the potential epitopes of viral proteins. Further, the mathematical expressions of various analytical techniques were discussed in detail to study the stability of the therapeutically important protein-protein and protein-ligand association.*

2.1. CHEMOINFORMATICS METHODS FOR LIGAND-BASED DRUG- DISCOVERY

2.1.1 Pharmacophore Mapping

The prediction of small-molecule binding to a targeted receptor is believed to be a very crucial part of drug discovery. In the absence of structural knowledge regarding the active site of targeted enzymes, information related to the spatial arrangement of atom groups or functional groups is essential to screen small molecule databases with desirable geometries. This mutual arrangement of atomic features, also known as pharmacophore, is essential for recognizing small molecule hits by enzymes. In general, a pharmacophore is a group of atomic features that are common to a series of biologically active compounds. In order to generate pharmacophoric patterns corresponding to a particular catalytic cavity, a wide range of theoretical and experimental data is routinely used. There are three known aspects in pharmacophore mapping, such as (i) determining the features essential for the biological activity, (ii) molecular conformation identification, and (iii) alignment or superposition of small molecule hits and scoring the pharmacophore models according to their match with

active compounds. The most common pharmacophoric features are (i) hydrogen bond acceptor (A), (ii) hydrogen bond donor (D), (iii) hydrophobic group (H), (iv) negatively ionizable (N), (v) positively ionizable (N) and (vi) aromatic ring (R). The common pharmacophoric features can be identified using a tree-based partitioning strategy to group similar spatial arrangements. The scoring function or the survival score of the pharmacophore hypothesis can be defined as

$$S = W_{site}S_{site} + W_{vec}S_{vec} + W_{vol}S_{vol} + W_{sel}S_{sel} + W_{mrew} \quad (2.1)$$

Where, S and W are the scores and weights respectively correspond to vector, volume, and selectivity. The W_{mrew} is the reward weight for the m number of active compounds that feats well with the pharmacophoric model. After the generation of the highly scored pharmacophoric model, a database of small organic hits can be screened to obtain highly active molecules for enzyme inhibition.

2.1.2 Quantitative structure-activity relationship (QSAR)

Quantitative structure-activity relationship (QSAR) is a statistical method that correlates the binding affinity with the functional group substitution without the knowledge of structural information of the binding pocket. The compounds used for QSAR model generation are generally from similar congeneric series or share a common scaffold, and their biological activity must be estimated by the same enzymatic assay procedure. In this method, electronic, steric, and molecular properties such as lipophilicity and polarizability of a series of molecules are fed to a QSAR model to train it for predicting the binding affinity of untested compounds with similar geometry. Therefore, selecting descriptors that describe the physicochemical and structural properties of the hits is believed to be integral to QSAR modelling. The predictability of the QSAR model depends on the quality of biological activity data, the statistical model, and the chosen descriptors.

In the process of QSAR method development, two theories such as (i) the Wilson method and (ii) Hansch analysis, are considered as the basis of current QSAR models. In both methods, the biological activity is predicted by adding the contributions from different structural features of a ligand. Further, the predicted activity can be correlated with the logarithm of equilibrium constants obtained from enzyme-drug

reactions. In the Wilson method, the presence or the absence of certain structural features is denoted by 1 or 0, and the resulting feature matrix is correlated with the biological activity. The biological activity can be computed as

$$\log (1/C) = \sum_i a_i + \mu \quad (2.2)$$

Where C is the concentration of the compound that causes 50% reduction in the enzyme activity. The a_i indicates the contribution of activity from the i^{th} functional group, and μ stands for the biological activity of the common scaffold. Similarly, according to the Hansch model, biological activity can be expressed in terms of the nonlinear equation and defined as

$$\log (1/C) = a (\log P)^2 + b \log P + c\sigma + \dots + k \quad (2.3)$$

Where P is the n-octanol/water partition coefficient, σ is the Hammett parameter, a, b, c is the regression coefficient, and k is the constant term. It is reported that Hansch analysis is superior compared to Free Wilson analysis due to the less requirement of properties to correlate the experimental activity of the hits. However, the outcome of different QSAR models of the same dataset causes dilemma in employing Hansch analysis. To overcome the limitation and broaden the applicability of both Free Wilson and Hansch analysis, the combination of both theories, known as the mixed method, can be used to correlate biological activity. This can be defined as

$$\log (1/C) = a (\log P)^2 + b \log P + c\sigma + \sum_i a_i + \dots + k \quad (2.4)$$

It can be noted here that the above two methods cannot describe the three-dimensional (3D) properties of ligands essential to derive the atom-based descriptors. In order to correlate the actual conformation of the ligand with the biological activity of the ligands, grid-based or field-based 3D-QSAR models have been developed. The oldest model or the prototype of current 3D-QSAR models is known as DYLOMMS (dynamic lattice-oriented molecular modelling system), which employs principal component analysis (PCA) to retrieve vectors from molecular interaction. This is further modified by advanced statistical methods such as partial least square (PLS) and molecular shape-based descriptors to predict the biological activity. The most important criteria to generate statistically sound 3D-QSAR model is the selection of training set and test set

compounds from the biological dataset. Care needs to be taken to include structurally diverse compounds in training set to cover most of the descriptor space of the overall biological dataset. Therefore, the 75% of the compounds in dataset is generally included in training set to build the 3D-QSAR model. The rest 25% are for testing the efficiency of the QSAR in predicting the biological activity of the compounds from model dataset. Notably, each point in training set needs to be closer to any one point in the training set at descriptor space. There are various algorithms available for dividing the dataset in to training set and test set such as K-means clustering, D-optimal design, Kennard-Stone algorithm, sphere exclusion algorithm etc. After 3D-QSAR model generation the plausible prediction of the model need to be further ensured by external test set compounds that are not been used in model generation.

2.1.3 Density Functional Theory (DFT)

Density functional theory calculations are known to be useful for optimizing the 3-dimensional structure of hit compounds, such as bond length, bond angles, and electronic properties like Mullikan charges, HOMO-LUMO energies, energy gap, vibrational frequency, and dipole moment. The electronic properties of the molecules are believed to play an important role in their pharmacological effects. As an example, the energy of HOMO is directly related to the ionization potential, and the energy of LUMO is related to the electron affinity. The locations correspond to HOMO, and LUMO indicates the possible sites for the electrophilic and nucleophilic attack, respectively, in the molecule. Moreover, the energy gap represents the chemical reactivity of the compounds. The lesser energy gap favours the rapid electron transfer or exchange and thereby makes the molecule highly reactive. Additionally, DFT calculation helps to envision the charge distribution or the potential energy surface of a molecule.

In DFT calculation, the central goal is to calculate the electron density ρ , which is related to ground state energy or other molecular properties, and map a wave function to a value. The density functional can be represented as

$$E_0[\rho] = T[\rho] + V[\rho] + EXC[\rho] \quad (2.5)$$

The square brackets in equation (2.6) indicate the functional while $T[\rho]$, $V[\rho]$ correspond to the kinetic and potential energy of the system. The potential energy consists of electron-electron and electron nuclei interaction energy. The third term attributes to exchange-correlation energy and the only unknown. The implication of the Schrödinger equation to replace the complex many-body wave function of a system with electron density simplifies the calculation of the electronic structure of bigger molecules. In this regard, the Hohenberg-Kohn theorem and Kohn-Sham equations are well known. According to Hohenberg-Kohn theorem

Theorem 1: *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 $n(r)$.*

Theorem 2: *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 .*

Therefore, the total energy or the Hamiltonian and the properties of an atomic system can be computed by the electron density. The lowest electron density or ground state density appears at the lowest energy, and other densities correspond to higher energy. In a system of non-interacting particles, the Kohn-Sham equation can be defined as a local effective potential under which the particles move

$$\left(\frac{-\hbar^2}{2m} \nabla^2 + v_{eff}(r) \right) \phi_i(r) = \varepsilon_i \phi_i(r) \quad (2.6)$$

Here, ε_i indicates the energy corresponds to the Kohn-Sham orbital (ϕ_i). Further, the density of the N particle system can be defined as

$$\rho(r) = \sum_i^N |\phi_i(r)|^2 \quad (2.7)$$

The energy of the system in terms of density can be given as

$$E(\rho) = T_s(\rho) + \int dr V_{ext}(r)\rho(r) + E_H(\rho) + E_{xc}(\rho) \quad (2.8)$$

This equation can be compared with eqn (2.6), except the external potential term. The kinetic energy in terms of orbital and Kohn-Sham potential energy term can be expressed as

$$T_s(\rho) = \sum_{i=1}^N \int dr \phi_i^*(r) \left(\frac{-\hbar^2}{2m} \nabla^2 \right) \phi_i(r) \quad (2.9)$$

$$v_{eff}(r) = v_{ext}(r) + e^2 \int \frac{\rho(r')}{|r-r'|} dr' + \frac{\delta E_{xc}(\rho)}{\delta \rho(r)} \quad (2.10)$$

As the exchange-correlation term is unknown, thus the accuracy of the DFT technique depends on the usage of different approximations such as local density approximation (LDA), generalized gradient approximation (GGA) or hybrid functional for exchange correlation functional. In recent days B3LYP hybrid functional which consists of Becke's 3-exchange functional and Lee's four correlational functional is believed to highly efficient in computing the exchange correlation term.

2.2 STRUCTURE-BASED DRUG-DISCOVERY TECHNIQUES

2.2.1 Molecular Docking

Molecular recognition plays a key role in promoting fundamental biomolecular events such as enzyme substrate, drug-protein, and antigen-immune receptor interactions. Detailed understanding of the general principles that govern the nature of the interactions (van der Waals, hydrogen bonding, electrostatic) between the ligands and their protein or nucleic acid targets may provide a framework for designing the desired potency and specificity of potential drug leads for a given therapeutic target. Molecular docking is a method which helps to identify the preferred orientation of one molecule (ligand) to a second (receptor) when bound to each other to form a stable complex. The main objective of docking is to optimize the alignment of ligand (small organic molecules or protein subunit) with respect to target receptor and estimate the binding affinity using scoring function. Therefore, the development of docking algorithm and the empirical scoring function is crucial to improve the accuracy of computer aided drug design or in-silico vaccine design.

The overall binding energy of the system is given by

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + (V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf}) \quad (2.11)$$

where L refers to ligand, P refers to protein (target) and P-L refers to the complex. Further, the ΔS_{conf} indicates the configurational entropy.

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

$$V = W_{vdw} \sum_{i,j} \left(\frac{A_{i,j}}{r_{i,j}^{12}} - \frac{B_{i,j}}{r_{i,j}^6} \right) + W_{hbond} \sum E(t) \left(\frac{C_{i,j}}{r_{i,j}^{12}} - \frac{D_{i,j}}{r_{i,j}^{10}} \right) + E_{ele} \sum \frac{q_i q_j}{e(r_{i,j}) r_{i,j}} + W_{sol} \sum_{i,j} (S_i V_j + S_j V_i) e^{-\frac{r_{i,j}^2}{2\sigma^2}} \quad (2.12)$$

The optimized weighing constant, 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 a directional hydrogen bond depending on 10/12 potential. The third term is Coulomb potential for electrostatics and the final term is desolvation potential.

2.2.1.1 Classification of Molecular Docking and Docking approaches

In terms of flexibility of ligand and receptor, docking can be classified in to following forms

- **Rigid Docking:** In this method both the ligand and receptor are considered as rigid conformation and mainly follows the Emil Fischer “lock and key” model. In this method the search of preferred ligand orientation is limited and only consider the translational, rotational degrees of freedom. In this method the ligand flexibility is accounted using a set of ligand conformers.
- **Flexible Docking:** In the case of induced fit model, it is crucial to consider the flexibilities of both receptor and ligand due to the conformational change for minimum energy complex. However, due to high computational cost the ligand conformation is usually treated as flexible instead of the receptor. In this method the bond angel or dihedrals are considered as rotatable to achieve good fit at the binding pocket of the receptor.

The schematic diagram of different mode of docking is given in **Figure 2.1**.

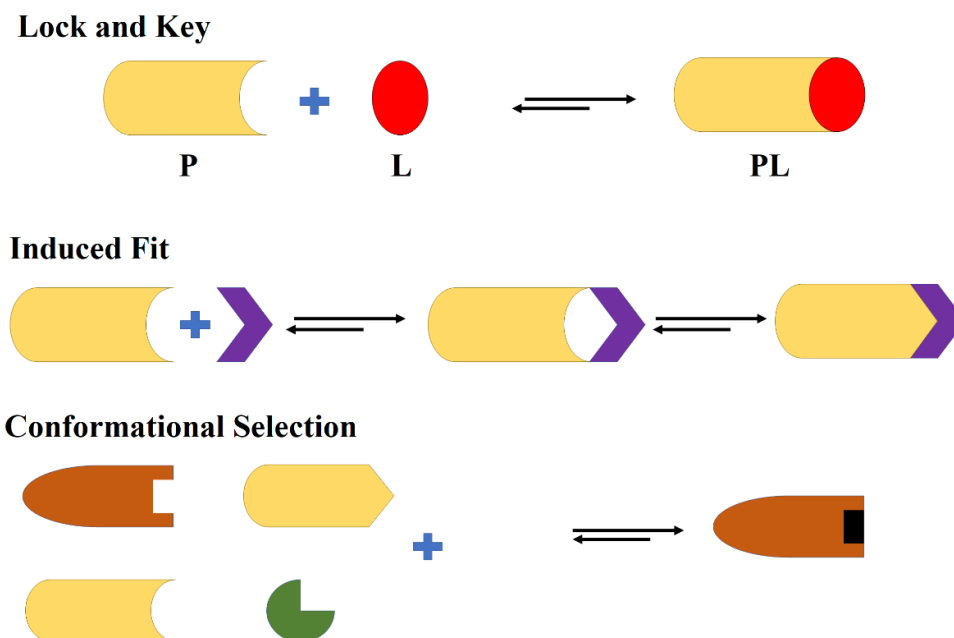


Figure 2.1 Schematic view of different docking methods.

Based on the algorithms regarding the minimum energy state of ligand at receptor binding pocket, docking approaches can be categorized as follows

- Monte Carlo Method: This method involves in making random conformation, rotation, and transition of ligand at the catalytic pocket and the scoring of the ligand conformation using Metropolis criteria. Further, the ligand conformation undergoes a probability function test to validate the ligand orientation.
- Matching or Ligand Fit Approach: The ligands are precisely placed at the active site based on space complimentary approach.
- Fragment Based Approach: In this method ligand fragments are divided into connected fragments that fits well with the binding pocket.
- Distance Geometry: This approach connects the pharmacophore sites for space complementarity.

2.2.1.2 Docking requirements

In order to obtain experimentally relevant Receptor-ligand complex, care need to be taken for ensuring the geometrical integrity of both the receptor or ligand prior docking. In case of receptor the atomistic detailing of binding pocket and the experimental protonation state of the amino acids are necessary to envision the non-bonded interaction between ligand and receptor after docking calculations. The ligand geometry such as bond distance, angle, dihedrals need to be optimized and the protonation state of polar atom need to be assigned correctly. Further, the scoring method need to be implemented to predict the binding affinity.

2.2.2 Molecular Dynamics Simulation (MD)

MD is an extremely powerful *in-silico* technique that involves in solving the classical equation of motion to study the microscopic behavior of atoms or molecules for a certain time period. Therefore, MD simulation techniques are often lauded as “virtual microscopes” and are known to have widespread implications in different areas of science especially in Life-science and material science. The method allows the prediction of the dynamic properties of substances directly from the underlying interactions between the molecules. The fundamental principle of MD simulation lies on the Newton’s second law of motion. From knowledge of the force on each atom, it is possible to determine the acceleration of each atom in the system. Integration of the equations of motion then yields a trajectory that describes the positions, velocities, and accelerations of the particles as they vary with time. The flowchart of the different MD steps is given in **Figure 2.2**. Molecular dynamics simulations for large biomolecules can be time consuming and computationally expensive due it’s complex potential energy surface. From the trajectory, the average values of thermodynamic properties can be determined by integrating the particular property and corresponding probability at each microstate. The average value of any property A can be given by

$$\langle A \rangle = \rho_1 A_1 + \rho_2 A_2 + \rho_3 A_3 + \dots + \rho_n A_n \quad (2.13)$$

For large number of microstates, the average property can be given as

$$\langle A \rangle = \int A(q) \rho(q) dq \quad (2.14)$$

Here, ρ is the probability of the property at each microstate which can be defined in terms of Boltzmann distribution.

$$\rho(q) = \frac{e^{-V(q)/kT}}{\int e^{-V(q)/kT} dq} \quad (2.15)$$

In the above equation the denominator is known as partition function and $V(q)$ is the potential energy correspond to a microstate.

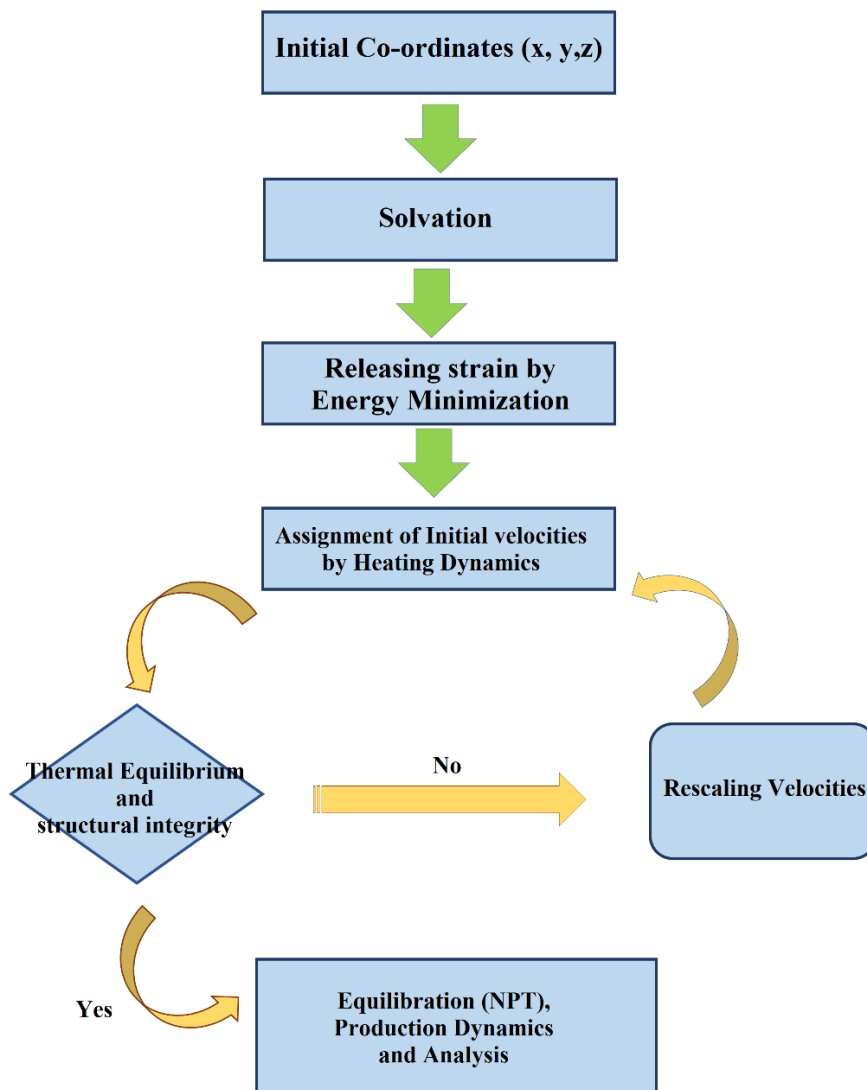


Figure2.2 Flow chart for running MD simulation.

For a system composed of N atoms, the force acting on atom I can be given as

$$F_i = m_i a_i = m_i \frac{dv_i}{dt} = \frac{dp_i}{dt} \quad (2.16)$$

Here, p_i is the linear momentum of the atom i . For a conservative system

$$F = -\nabla_i V = -\frac{dV}{dr_i} \quad (2.17)$$

Combining equation (2.17) and (2.18) we obtain

$$-\frac{dV}{dr_i} = \frac{dp_i}{dt} \quad (2.18)$$

The classical kinetic energy of atom i can be given as

$$T_i = \frac{1}{2} m_i v_i = \frac{1}{2} \frac{m_i^2}{m_i} v^2 = \frac{p_i^2}{2m_i} \quad (2.20)$$

Differentiating the kinetic energy with respect to linear momentum we obtain

$$\frac{dT_i}{dq_i} = \frac{2p_i}{2m_i} = \frac{p_i}{m_i} = \frac{m_i v_i}{m_i} = v_i = \frac{dq_i}{dt} \quad (2.19)$$

The equations (2.19) and (2.21) are known as Hamiltonian equation of motion and need to be solved numerically to visualize the movement of atoms.

2.2.2.1 Integrator Algorithms:

- *Leap-Frog Algorithm:* This algorithm 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.

$$v\left(t + \frac{1}{2} \partial t\right) = v\left(t - \frac{1}{2} \partial t\right) + \frac{\partial t}{m} F(t) \quad (2.20)$$

$$r(t + \partial t) = r(t) + \partial t v\left(t + \frac{1}{2} \partial t\right) \quad (2.21)$$

- *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

$$r(t + \partial t) = r(t) + v(t) \partial t + \frac{1}{2} a(t) \partial t^2 \quad (2.22)$$

$$r(t - \partial t) = r(t) - v(t) \partial t + \frac{1}{2} a(t) \partial t^2 \quad (2.23)$$

Summing (2.7) and (2.8),

$$r(t + \partial t) = 2r(t) - r(t - \partial t) + a(t) \partial t^2 \quad (2.24)$$

- *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.

$$r(t + \partial t) = r(t) + \partial t v(t) + \frac{1}{2} \partial t^2 a(t) \quad (2.25)$$

$$v(t + \partial t) = v(t) + \frac{1}{2} \partial t [a(t) + a(t + \partial t)] \quad (2.26)$$

2.2.2.2 Potential energy and Force Field

The calculation of the potential energy of the system is crucial to estimate the force acting on each atom of the system and helpful in solving the equations of motions. The potential energy of the system can be related by a simple equation of internal coordinates known as force field. In order to simulate the dynamics of biomolecular system various force fields are available such as AMBER (Weiner et al. 1984), CHARMM (Brooks et al. 1983), OPLS (Jorgensen and Tirado-Rives 1988) and GROMOS (van Gunsteren and Berendsen 1987). The potential energy of the system can be defined as

$$V = \sum_{bonds} k_b (b - b_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} \frac{1}{2} V_n [1 + \cos(n\phi - \delta)] + \sum_{impropers} k_\omega (\omega - \omega_0)^2 + \sum_{non-bonded} 4\epsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{min,ij}}{r_{ij}} \right)^6 + \sum_{non-bonded} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right] \quad (2.27)$$

The first four term of the equation is known as the bonded terms and consists of energy related to bond stretching, angle bending and torsional energies. The k is the force constant in the bonded energy term in the potential energy equation. The last two term of equation is related to the non-bonded energy terms (electrostatic and Lennard-Jones energy terms) between two atoms that are separated by more the three consecutive bonds. In case of Lennard-Jones (LJ) energy terms the r_{ij} corresponds to the inter-atomic distance, ϵ_{ij} is the well depth. The radius of the atoms can be determined by the parameter σ or the distance where the LJ potential is zero. In electrostatic energy term ϵ_0 indicates the vacuum permittivity.

2.2.2.3 Periodic Boundary Condition

Periodic boundary conditions are crucial to approximate behavior of a macromolecular system by using the small part known as unit cell and minimize the edge effect in MD simulations. In this method the unit cell is replicated in 3 dimensions with the periodic images to impose no boundary which ensures the connection with the surrounding environment. After imposing the periodic condition, the long range electrostatic interactions are often calculated by Ewald Summation (Wells and Chaffee 2015) and requires charge neutrality of the system. However, to improve the speed and accuracy to compute electrostatic interaction in large-scale MD simulations Particle Mesh Ewald (PME) method have been developed (Essmann et al. 1995a). In most of the MD engines triclinic unit cell is mostly considered which satisfy the following conditions

$$a_x = b_y = c_z \quad (2.28)$$

$$a_x > 0, b_y > 0, c_z > 0 \quad (2.29)$$

$$a_x \leq \frac{1}{2}b_y, b_y \leq \frac{1}{2}c_z, c_z \leq \frac{1}{2}a_x \quad (2.30)$$

2.2.2.4 Temperature and Pressure Control in MD Simulation

The temperature and pressure equilibrium are important criterion in MD simulation to maintain the appropriate phase space density for the system of interest. To capture the statistical characteristics of different microscopic environments different ensembles are available such as microcanonical ensemble (NVE), canonical ensemble (NVT) and isobaric-isothermal ensemble (NPT). In last two ensembles are highly used in biomolecular simulations, and the temperature of the system need to be constant throughout the simulation process. The thermal equilibrium can be achieved by employing various thermostat algorithms, as example: Velocity Rescaling, Berendsen, Nose-Hoover etc. In every case the velocity of the particles can be obtained from Boltzmann distribution which can be given as

$$p(r) \propto \exp(-U(r)/k_B T) \quad (2.31)$$

Further, controlling pressure in the MD simulation is crucial as most of the biological process happens at constant pressure. The pressure of a system containing N particles

can be calculated by *Clausius virial theorem* given as

$$P = \frac{2}{3V}(E_{kin} - \Xi) \quad (2.32)$$

In equation, V is the velocity of the system and E_{kin} correspond to the kinetic energy of the system. The parameter Ξ can be defined as

$$\Xi = \frac{1}{2} \sum_{i < j} r_{ij} \cdot f(r_{ij}) \quad (2.33)$$

Here, $f(r_{ij})$ is the force acting between two particles j and i at distance r_{ij} .

1.2.5 Free Energy Calculations

All the molecular processes such as protein folding, protein-protein/ligand association, are known to be driven by free energy. Therefore, accurate determination of binding free energy of the inhibitors is believed to be crucial in the drug discovery process (Christ et al. 2010; Kumari et al. 2014a; Srinivasan et al. 1998a). A range of computational approaches can be used to estimate binding free energies. These includes thermodynamic integration (TI), free energy perturbation (FEP), linear interaction energy (LIE), molecular mechanics Poission-Boltzmann surface area (MM-PBSA), molecular mechanics Generalized Born surface area (MM-GBSA) etc. (Foloppe and Hubbard 2006; Gohlke et al. 2003; Kollman et al. 2000; Kollman 1993; Meirovitch 2007; Parenti and Rastelli 2012; Ytreberg et al. 2006). Of these FEP and TI are the most rigorous but also the most computationally expensive. As a consequence, these two methods are mostly applicable to small perturbations or structural transitions (Meirovitch 2007). FEP and TI methods are also difficult to employ in cases where large numbers of structurally diverse ligands are being considered as it is necessary to generate a dual/hybrid topology for each protein-ligand pair (Kollman 1993). Endpoint free energy methods estimate the free energy using an ensemble of structures at the initial and final states making these approaches computationally highly efficient (Foloppe and Hubbard 2006; Gohlke et al. 2003; Kollman et al. 2000).

Generally, the binding free energy of the protein with ligand in solvent can be expressed as (Gilson and Honig 1988b; Kollman et al. 2000; Srinivasan et al. 1998b)

$$\Delta G_{binding} = G_{complex} - (G_{protein} + G_{ligand}) \quad (2.34)$$

where, $G_{complex}$ is the total free energy of the protein–ligand complex and $G_{protein}$ and G_{ligand} are total free energies of the isolated protein and ligand in solvent, respectively. Furthermore, the free energy for each individual entity can be defined as (Gohlke et al. 2003; Kollman et al. 2000; Kuhn and Kollman 2000; Lee et al. 2000; Wang and Kollman 2000)

$$G_x = \langle E_{MM} \rangle - TS + \langle G_{solvation} \rangle \quad (2.35)$$

where x is the protein or ligand or protein–ligand complex. $\langle E_{MM} \rangle$ is the average molecular mechanics potential energy in a vacuum. TS refers to the entropic contribution to the free energy in a vacuum where T and S denote the temperature and entropy, respectively. The last term $\langle G_{solvation} \rangle$ is the free energy of solvation.

1.2.5.1 Molecular Mechanics Potential Energy

The vacuum potential energy EMM, includes the energy of both bonded as well as nonbonded interactions, and it is calculated based on the molecular mechanics (MM) force-field parameters (Hornak et al. 2006a; Lindorff-Larsen et al. 2010a; Wang et al. 2004).

$$E_{MM} = E_{bonded} + E_{nonbonded} = E_{bonded} + (E_{vdW} + E_{elec}) \quad (2.36)$$

where E_{bonded} is bonded interactions consisting of bond, angle, dihedral and improper interactions. The nonbonded interactions ($E_{nonbonded}$) include both electrostatic (E_{elec}) and van der Waals (E_{vdW}) interactions and are modelled using a Coulomb and Lennard-Jones (LJ) potential function, respectively.

1.2.5.2 Free Energy of Solvation

The free energy of solvation is the energy required to transfer a solute from vacuum into the solvent. In the MM-PBSA approach, it is calculated using an implicit solvent model. The solvation free energy is expressed as the following two terms (Gilson and Honig 1988b; Honig and Nicholls 1995; Still et al. 1990)

$$G_{solvation} = G_{polar} + G_{nonpolar} \quad (2.37)$$

Where, G_{polar} and $G_{nonpolar}$ are the electrostatic and Van der Waals contributions to the solvation free energy, respectively.

1.2.5.3 Polar Solvation Energy

The electrostatic term, G_{polar} , is estimated by solving the Poisson–Boltzmann (PB) equation,(Ganath et al. 2006; Honig and Nicholls 1995; Srinivasan et al. 1998b) which is given by

$$\nabla \cdot [\varepsilon(r) \cdot \varphi(r)] - \varepsilon(r)k(r)^2 \sin[\varphi(r)] + \frac{4\pi\rho^f(r)}{kT} = 0 \quad (2.38)$$

where $\phi(r)$ is electrostatic potential, $\varepsilon(r)$ is the dielectric constant, and $\rho^f(r)$ is the fixed charge density. The term k^2 is related to the reciprocal of Debye length which is dependent on the ionic strength of the solution.

1.2.5.4 Nonpolar Solvation Energy

The non-electrostatic term of solvation free energy (nonpolar) includes repulsive and attractive forces between solute and solvent that are generated by cavity formation and van der Waals interactions, respectively(Levy et al. 2008; Tan et al. 2007; Wagoner and Baker 2006). It can be expressed as

$$G_{nonpolar} = G_{cavity} + G_{vdW} \quad (2.39)$$

Where, G_{cavity} is work done by the solute to create a cavity in the solvent and depends on the shape and geometry of the solute. G_{vdW} is the attractive van der Waals energy between solvent and solute. These terms can be estimated using a variety of models.

1.2.5.5 SASA-Only Nonpolar Model

The solvent accessible surface area (SASA) model is based on the assumption that the SASA is linearly dependent on the $G_{nonpolar}$ term and can thus be calculated as follows(Gohlke et al. 2003; Kuhn and Kollman 2000; Lee et al. 2000; Srinivasan et al. 1998b; Still et al. 1990):

$$G_{nonpolar} = \gamma A + b \quad (2.40)$$

where γ is a coefficient related to surface tension of the solvent, A is SASA, and b is fitting parameter.

1.2.5.9 Binding Energy Decomposition

Initially the energy components E_{MM} , G_{polar} , and $G_{nonpolar}$ of individual atoms are calculated in the bound as well as the unbound form, and subsequently their contribution to the binding energy ΔR_x^{BE} of residue x is calculated as follows:

$$\Delta R_x^{BE} = \sum_{i=0}^n (A_i^{bound} - A_i^{free}) \quad (2.41)$$

Where, $A_i^{complex}$ and A_i^{free} are the energy of i th atom from x residue in bound and unbound forms, respectively and n is the total number of atoms in the residue. We note that energy contribution summed over all residues is equal to the binding energy, i.e. $\Delta G_{binding} = \sum_{x=0}^m \Delta R_x^{BE}$ where, m is total number of residues in either protein–protein or protein–ligand complexes.

2.2.4 Umbrella Sampling simulations

Umbrella Sampling (US) simulations have long been used to explore the mechanism of ligand-protein unbinding and the determination of binding free energy (Huang et al. 2017b; Miao et al. 2018; Sun et al. 2015; Zhou et al. 2015b). However, the unbinding pathway modelled by such enhanced sampling methods must mimic the natural molecular movement to generate unbiased Potential of mean force (PMF). To steer the unbinding of ligands from the catalytic pocket of protein, a harmonic potential can be added to the system's real or unbiased potential (V_u). The bias potential can be expressed as

$$V_b = V_u + U(\xi) \quad (2.42)$$

ξ is the reaction coordinate (RC) that changes with respect to time during the pulling simulations. The harmonic potential can be defined as follows

$$U_i(\xi) = \frac{1}{2} K_i (\xi_i - \xi_{i0})^2 \quad (2.43)$$

Where ξ_{i0} is the target value of the RC at the window i around which the sampling is essentially confined, and K_i is the spring constant. It is believed that the sampling distribution largely depends on the value of the K_i . Therefore, choosing the value of K_i significantly reduces the number of umbrella sampling windows for generating unbiased, free energy change for a rare event. The force component generated by the bias potential on coordinate x_j can be calculated as

$$\frac{\partial U_i}{\partial x_j} = F_i = -K_i (\xi_i - \xi_{i0}) \frac{\partial \xi_i}{\partial x_j} \quad (2.44)$$

The unbiased, free energy change of the permeation process can be obtained by WHAM (weighted histogram analysis method) algorithm. The free energy can be defined as,

$$G^u(\xi) = -k_B T \ln P^u(\xi) = \frac{1}{\beta} \ln P^u(\xi) \quad (2.45)$$

$P^u(\xi)$ is the unbiased probability distribution along the reaction coordinate, and k_B is the Boltzmann constant. This can be computed using the biased probability ($P^b(\xi)$) distribution from umbrella sampling expressed in terms of partition function as follows

$$P^b(\xi) = \frac{\int \delta[\xi(q) - \xi] e^{-\beta[V_u(q) + U_i(\xi(q))]} dq}{\int e^{-\beta[V_u(q) + U_i(\xi(q))]} dq} = e^{-\beta U_i(\xi(q))} \frac{\int \delta[\xi(q) - \xi] e^{-\beta[V_u(q)]} dq}{\int e^{-\beta[V_u(q) + U_i(\xi(q))]} dq} \quad (2.46)$$

Further, the unbiased probability distribution along the RC can be written as

$$P^u(\xi) = P^b(\xi) e^{\beta U_i(\xi)} \langle e^{-\beta U_i(\xi)} \rangle \quad (2.47)$$

Where the $\langle \dots \rangle$ indicates the ensemble average of the exponential and the unbiased free energy at window i can be written as

$$G_i^u(\xi) = -\frac{1}{\beta} \ln P_i^b(\xi) - U_i(\xi) + X_i \quad (2.48)$$

The calculation of $X_i = -\frac{1}{\beta} \ln \langle e^{-\beta U_i(\xi)} \rangle$ depends on the value of the ensemble average of the exponential at each umbrella window. Further, the ensemble average can be defined as

$$\langle e^{-\beta U_i(\xi)} \rangle = e^{-\beta X_i} = \int P^u(\xi) e^{-\beta U_i(\xi)} d\xi \quad (2.49)$$

The global unbiased probability distribution (P^u) can be calculated as the weighted average of the unbiased probability distribution of each window. Thus,

$$P^u(\xi) = \sum_i^{n_w} w_i(\xi) P_i^u(\xi) \quad (2.50)$$

The weights at each umbrella window can be calculated by minimizing the global probability distribution with two conditions such as $\frac{\partial(P^u(\xi))}{\partial w_i} = 0$ and $\sum_i^{n_w} w_i(\xi) = 1$.

At window i the weight depends on can be expressed as follows

$$w_i(\xi) = \frac{\alpha_i(\xi)}{\sum_j^{n_w} \alpha_j} \quad (2.51)$$

$$\alpha_i(\xi) = n_i e^{-\beta U_i(\xi) + \beta X_i} \quad (2.52)$$

Here, n_i is the number of steps sampled at the window i . Notably, the computation of factor ΔX_i at equation (7) depends on equations (9), (10), and (11), whereas the value of parameter α_i again depends on X_i . Therefore, the potential of mean force calculation is essentially an iterative method and needs to be carried out until the free energy results converge.

2.2.5 Free Energy Perturbation

The above two free energy methods mentioned here fail to capture the free energy change due to the functional group or amino acid mutations. In such cases, overlap sampling and perturbation theory can be combined to compute free energy change from MD trajectories. The free energy perturbation theory or FEP captures the free energy difference between two ligands at the catalytic pocket of protein when ligand A changes to ligand B due to functional group mutation. The equation (Kollman, 1993; Zwanzig, 1954) for calculating relative free energy is given as

$$\Delta\Delta G = \Delta G_A - \Delta G_B = -k_B T \ln \langle \exp(- (U_B - U_A)/k_B T) \rangle_A \quad (2.53)$$

Where, U_A and U_B correspond to the potential energy of A and B states, k_B is the Boltzmann constant, G is the relative free energy difference when ligand A transforms to B and T is absolute temperature. The $\langle \dots \rangle_A$ indicates the ensemble average over a simulation trajectory of state A. The thermodynamic cycle for determining relative free energy calculation is depicted in **Figure 2.3**. In this method, the change of free energy for the transformation of A state to B state is represented as a function of the coupling parameter namely λ . This parameter tunes the level of changes the system go through or the extent of perturbation of Hamiltonian during the transformation of state A to state B. Thus, it is important to plan the number of λ points to describe the change from the initial point ($\lambda=0$) to final point ($\lambda=1$). Care needs to be taken to ensure adequate sampling in each phase space for maximum overlapping at each λ_i point during the transformation and the convergence in free energy. The force-field parameters (χ_j) were scaled from the linear equation of λ_i given as

$$\chi_i = \chi_A \lambda_i + (1 - \lambda_i) \chi_B \quad (2.54)$$

The χ_A and χ_B are the force field parameters corresponding to state A and B.

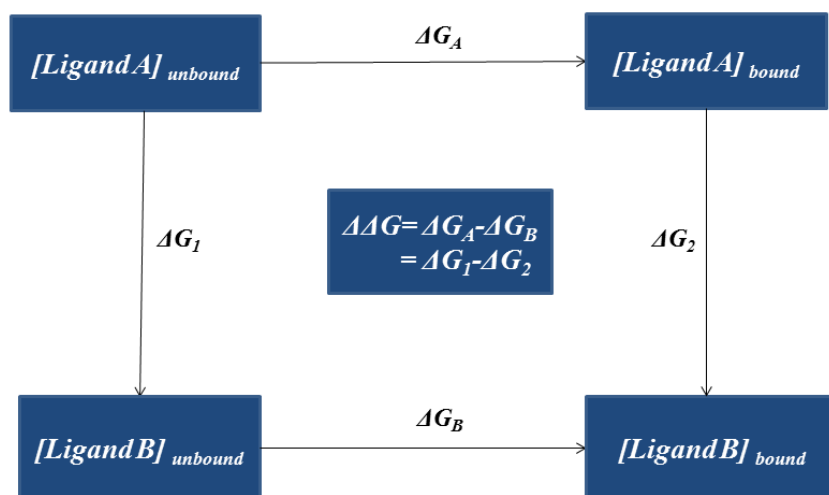


Figure 2.3. The Thermodynamic cycle for calculating the relative free energy of binding between two states. ΔG_A and ΔG_B are the free energy difference for going unbound state to bound state. ΔG_1 and ΔG_2 are the free energy difference for transforming ligand A to the ligand B.

2.3 ANALYSIS

The following analysis was carried out to assess the interaction and dynamics of biomolecules and their corresponding therapeutic agents.

2.3.1 Contour Plot analysis: The contour plots for 3D-QSAR models at 6th PLS was constructed by the Phase tool of Schrodinger software. Further the HOMO, LUMO orbitals were generated by Cubegen utility

2.3.2 Root Mean Square Deviation (RMSD):

The convergence of a biomolecular motion in MD simulation is known to be an indication of equilibrium state and can be measured by the spatial difference between two frames of the trajectory. The RMSD can be defined as

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i^X - r_i^Y)^2} \quad (2.55)$$

In eqn (2.39) N is the number of atoms, I indicate the current atom and r^X , r^Y are the starting and reference states respectively. The *gmx rms* utility of the Gromacs tool have been employed to calculate the RMSD profile correspond to molecule of interest and the target. The reference structures have been aligned before calculating the RMSD.

2.3.3 Root Mean Square Fluctuation:

The measurement of fluctuation is crucial to assess the dynamics of protein for a trajectory of T ns. The root means square fluctuation of residue i for T time period (ns) with respect to reference position (r_i^{ref}) can be defined as

$$RMSF_i = \sqrt{\frac{1}{T} \sum_{t_j=1}^T (r_i(t_j) - r_i^{ref})^2} \quad (2.56)$$

The *gmx rmsf* utility of Gromacs was used to calculate RMSF of proteins residues.

2.3.4 Radius of Gyration:

The measurement of global shape or the compactness of protein is primary measurement for conformational transition during the simulation time scale. The radius of gyration can be defined as

$$R_g = \sqrt{\frac{\sum_i (r_i^2) m_i}{\sum_i m_i}} \quad (2.57)$$

In eqn (2.57) m_i is the mass of atom r_i .

2.3.5 Number of Intra-and Intermolecular Hydrogen Bond and it's lifetime:

The hydrogen bonds between the protein and small molecules are said to be formed if the distance between the donor atoms O, N and acceptor hydrogen H linked to O, N are separated by 0.35 nm, larger than 100° (Stickle et al. 1992). The *gmx hbond* utility of Gromacs was used to calculate the hydrogen bonds formed during the simulation time scale. The occupancy percentage of a specific donor-acceptor pair is crucial to assess their frequency during simulation. The hydrogen bond occupancy can be defined as the ratio between the number of frame where hydrogen bond formed and the total number

of frames in the simulation trajectory. The hydrogen bond can be defined as(Chandra 2000)

$$S_{HB}(t) = \frac{\langle h(0).H(t) \rangle}{\langle h(0)^2 \rangle} \quad (2.58)$$

The hydrogen bond between donor and acceptor is said to be formed if their interatomic distance is < 0.25 nm. Here, $\langle \dots \rangle$ is indicating the average over all the pairs of given type. $H(t)$ is indicating the population parameter which is 1 if a specific hydrogen bond exists from time $t = 0 \rightarrow 1$ or zero otherwise. We calculated the hydrogen bond between protein polar hydrogens (attached to nitrogen and oxygen) and oxygen of water, protein oxygen, and hydrogen of water. We have considered interfacial water molecules up to a cut off value of 0.5 nm for the calculation of protein-water hydrogen bond lifetime.

2.3.6 Secondary Structure Analysis

The change in secondary structure such as alpha helix, extended β -sheet, turns and coils of the protein during simulation was analysed by DSSP programme. The *gmx do_dssp* utility was used to generate the matrix file and *gmx xpm2ps* utility was used to generate the time evolution of secondary structure.

2.3.7 Principal component Analysis

The reduction of high dimensional data obtained from MD trajectory is necessary to extract the major motion or the essential dynamics system that directly related to the function, allosteric regulation and folding of biological macromolecules or peptides. Principle component analysis (PCA) is such computational method to identify the collective motions. The principal component modes of biomolecular dynamics can be calculated by diagonalizing the covariance matrix defined as

$$C_{ij} = \frac{1}{N} \sum_{k=1}^N (x_i^k - \langle x_i \rangle)(x_j^k - \langle x_j \rangle) \quad (2.59)$$

In eqn (2.59) x_i^k and x_j^k are the pair of elements correspond to vector x^k which depict the configuration of the system at time step k . $\langle x_i \rangle$ and $\langle x_j \rangle$ are the time averaged value of the elements for the N structures sampled in the MD simulation. In general, x is the vector that contains all the coordinates of the c_α atom of a protein. The eigen vector correspond to largest eigen value provides the low dimensional subspace that can describe the maximum behaviours of the biomolecules.

2.3.8 Radial Distribution Function (RDF)

Radial distribution function defines the structure of a molecular system as a change in of particle number density with respect to distance from a reference point. This can be defined as

$$g_{\alpha\beta} = \lim_{dr \rightarrow 0} \frac{\rho(r, dr)}{4\pi \left(\frac{N_{\alpha\beta}}{V}\right) r^2 dr} \quad (2.60)$$

Here, r is the distance between the atom pairs of interest. $P(r, dr)$ is number of atom pairs present in the distance spanning from r to $r+ dr$. $N_{\alpha\beta}$ is the number of pairs of two species considered to calculate the pair correlation function. The deviation in the value of $g_{\alpha\beta}$ from the unity indicates the change in order or correlation in the system.

2.3.8 Preferential Binding Coefficient (PBC)

The preferential binding coefficient quantify the extent of interaction between two species in the aqueous solution and can be calculated by the Kirkwood-Buff (KB) theory of preferential interaction. The KB integral ($G_{\alpha\beta}$) can be defined as

$$G_{\alpha\beta} = 4\pi \int_0^{\infty} [\langle g_{\alpha\beta}(r) \rangle - 1] r^2 dr \quad (2.61)$$

Here, α and β are the two species considered for calculating the binding coefficient. Further, the preferential binding coefficient of species α with specie β can be written as

$$v_{\alpha\beta} = \rho (G_{\alpha\beta} - G_{\alpha W}) \quad (2.62)$$

In eqn (2.46) $G_{\alpha W}$ is the KB integral value between species α and water molecules.

2.3.9 Free Energy Landscape:

The free energy surface or landscape technique is widely used to detect the extent of sampling of proteins or peptides in the phases space and number of minima present in a particular biological reaction. The construction of such free energy contour is useful in explaining conformational transition, folding, or unfolding of proteins and the self-assembly process of the peptides. Notably, the selection of reaction coordinate is crucial that can maximally explain the biological process. The free energy is mainly calculated based on the populations correspond to the reaction coordinates of collective variables. The free energy profile corresponds to λ_1 and λ_2 can be defined as

$$\Delta G(\lambda_1, \lambda_2) = -k_B T [\ln P(\lambda_1, \lambda_2) - \ln P_{max}] \quad (2.63)$$

Where, T and k_B are the system temperature and Boltzmann constant respectively. The P indicates the probability density function retrieved from the histogram of the reaction coordinates and P_{max} is the maximum probability density. This maximum probability density value needs to be subtracted to ensure the lowest free energy minima at $\Delta G=0$.

2.3.10 Mean Square Displacement

The extent of deviation for a particle with respect to time is known to the measure of the mobility of the particle. It can be calculated as

$$MSD = \langle |r_i(t_0 + t) - r_i(t_0)|^2 \rangle \quad (2.64)$$

Where $r_i(t_0+t)$ and $r_i(t_0)$ represents the position vectors of i^{th} atom at time t and starting time t_0 . The diffusion co-efficient can be calculated from Einstein's equation

$$D = \lim_{t \rightarrow \infty} MSD / 6t \quad (2.65)$$

2.3.11 ADME toxicity analysis:

It is believed that only improving potency of a particular drug towards its target is not the primary objective in the drug development process, unless taking PK profile and toxicity in to consideration (Ekins et al. 2002). The *in vivo* behaviours of the hit compounds relate the structural properties to the chemical, physical or biological characteristics. Therefore, calculation of specific parameters and the information of their permissible range is crucial to predict the pharmacological performance and ensure the minimal side-effect in human body. The parameters chosen in the current study are listed in **Table 2.1**.

Table 2.1 The ADME parameters and their permissible range

Parameters	Explanation	Permissible Range
MW	Molecular weight of the organic compound	130.0 - 725.0 g
logP _{o/w}	<i>n</i> -octanol water partition coefficient that serve as estimation of lipophilicity and hydrophilicity of the substrate.	-2.0- 6.5

HBD	Number of donar hydrogen bond present in the molecule	0.0- 6.0
HBA	Number of acceptor hydrogen present in the molecule	2.0- 20.0
SASA	Solvent accessible surface area of the hit compounds	300.0- 1000.0 Å ²
PSA	Estimates the surface corresponding to polar atoms especially oxygen and nitrogen.	7.0- 200.0 Å ²
rotor	Number of rotatable bonds in the molecule	0-15
logS	Estimates the water solubility of hit compound.	-6.5- 0.5 mol dm ⁻³
dipole	Estimates the electric polarity of the molecule	1.0- 12.5 debye
logK _{hsa}	Measures the extent of binding blood serum albumin	-1.5- 1.5
logBB	Blood-brain barrier partition coefficient that measures the ability of the hit compound to cross blood brain barrier.	-3.0- 1.2
PCaco- 2	This parameter predicts the human intestinal permeability of a compound.	< 25 poor, > 500 great.
PMDCK	Estimates the small molecule absorption on Madin-Darby Canine Kidney cell line	< 25 poor, > 500 great.
	Human oral absorption percentage	> 80% high and < 25% poor

The combination of first four parameters is known as Lipinski's rule of five and the combination of PCaco, logS and number of primary metabolites is known as Jorgensen rule of five. This thumb rules are the primary assessment of drug likeness.

2.3.12 In-silico Vaccine Design

2.3.12.1 Multiple Sequence Alignment and domain analysis

The evolutionary diversity of various classes of corona-virus S-protein was primarily determined by Multiple sequence analysis (MSA). We used MEGA7(Kumar et al. 2001, 2016) software for MSA and the Phylogenic tree was constructed by ClustalW tool(Larkin et al. 2007), employing neighbour joining algorithm with default parameters and 1,000 bootstrap replicas. In this paper, S-protein amino acid sequences of all coronavirus species (α -coronavirus, γ -coronavirus, β -coronavirus such as SARS-CoVid, MERS-CoVid and SARS-CoVid2, δ -coronavirus) were considered which cause severe lungs diseases in human or animal. Additionally, S-protein sequences of SARS-CoVid2 retrieved from most affected countries in the globe were included in MSA and Phylogenetic tree analysis. All the sequences that are included in this study were retrieved from NCBI database (<https://www.ncbi.nlm.nih.gov/>).

Further, the major domains of spike glycoprotein of SARS-CoVid-2 were identified by Pfam database (<https://pfam.xfam.org/>).

2.3.12.2 B-cell epitope prediction

In order to design anti-SARS-CoVid2-neutralizing antibodies, the antigenic sites of S-protein were determined using IEDB (Immune-Epitope-Database and Analysis-Resource) tool (<http://tools.iedb.org/main/>). The linear B-cell epitopes were predicted by set-of physicochemical parameters such as the exposed surface propensity by Emini et. al(Emini et al. 1985b), hydrophilicity by Parker et. al(Parker et al. 1986b), flexibility by Karplus et. al(Karplus and Schulz 1985b), antigenic propensity by Kolaskar et. al(Kolaskar and Tongaonkar 1990a) and β -turn propensity by Chou et. al(Chou and Fasman 1977) etc. In addition, the linear epitopes were also predicted by Bepipred Linear epitope prediction method(Larsen et al. 2006). The surface accessible probability (S_n) of an amino-acid located at the n^{th} position of any amino acid sequence can be defined as

$$S_n = [\prod_{i=1}^6 \delta_{n+4-i}] * (0.37)^{-6} \quad (2.66)$$

Where δ_n stands for the fractional surface probability of an amino acid located at position n. The probability value greater than 1.0 indicates the tendency of a residue to

be located at the surface of protein. The hydrophilicity profile of the amino-acid sequence SARS-CoVid2 s-protein was predicted by correlating the hydrophilicity HPLC (High-performance liquid chromatography) parameters of twenty synthetic peptides(Parker et al. 1986b). The antigenicity of a peptide is believed to be depending on the flexibility of the side chain of the amino-acid residues(Westhof et al. 1984). The B-factors (temperature factor) of the residues present in S-protein is known to be necessary for measuring the segmental flexibility. The normalized B-value of a C_α atom of residue was calculated by the following equation

$$B_{norm} = (B + D_P) / (\langle B \rangle_P + D_P) \quad (2.67)$$

Where, $\langle B_P \rangle$ is the average B-value of all C_α atoms of protein P and D_P is the tuning factor of the protein. The values of B_{norm} more than 1.0 indicates the antigenic region of a protein. Further, the antigenic propensity of the S-protein residues was calculated from the relation mentioned below

$$A_P = f_{Ag} / f_s \quad (2.68)$$

Where, f_{Ag} stands for the frequency of occurrence of each amino-acids at the known antigenic regions and f_s indicates the frequency of occurrence of the residues at the surface of query protein(Kolaskar and Tongaonkar 1990a). Further, the continuous residue stretches with higher β -turn, β -sheet propensities were also determined to identify the most antigenic regions of S-protein. VaxiJen 2.0(Doytchinova and Flower 2007) was employed to evaluate the antigenicity of the predicted epitopes. The antigenic score of the predicted epitopes with an antigenicity score of 0.4 believed to initiate adequate defensive immune reaction. B-cell discontinuous or conformational epitope that are scattered along the protein sequence were identified by DiscoTope2.0(Larsen et al. 2006) with a discotope score threshold of -3.7 was used to predict discontinuous epitopes. The Pymol was used to examine the positions of selected linear and discontinuous epitopes on the 3D structure of SARS-CoV-2 spike-protein.

2.3.12.3 T-cell epitope prediction

The development of memory B-cell in human body depends on the recognition of cytotoxic T-lymphocyte (CTL) epitopes presented at the surface of T-cell receptors. The memory B-cell is believed to fight against the re-infection (secondary immune

response) to mitigate the effects of disease-causing pathogen(Akkaya et al. 2020). IEDB tool was employed to predict the peptides that are presented by the major types of T-cell receptors (TCRs) such as Major Histocompatibility complex Class I (MHC Class I) and Class II. The prediction of both HLA (human leukocyte antigen) class-I and class-II T-cell epitopes are mainly based on artificial neural network (ANN) implemented in NetMHCpan tool(Andreatta and Nielsen 2016; Nielsen and Andreatta 2016). In the present case we considered A*02:01, A*24:01, B*40:01, B*58:01 alleles for MHC-I based and DRB1*04:01, DRB1*07:01 for MHC-II based epitopes due to their importance in case of SARS-CoVid2(Grifoni et al. 2020). In order to filter the T-cell epitope peptides with lower affinity, the Percentile rank threshold was set to 1% for MHC-I and 10% for MHC-II. Further, the antigenicity of each T-cell epitopes was calculated by VaxiJen (Doytchinova and Flower 2007) (v2.0).

2.3.12.4 Multi-epitope Vaccine design

The continuous B-cell epitopes from S-protein predicted by different physicochemical methods mentioned above were linked together by GPGPG linkers in sequential manner. In order to induce regulatory immune response, the Cholera Toxin B (CTB) adjuvant was fused by EAAAK linker at the N-terminal end of the vaccine construct. The similar procedure was followed for MHC-I and MHC-II based linear T-cell epitopes. The tertiary structure of the two linear vaccine construct was generated by trRosetta(Yang et al. 2020) web server (<https://yanglab.nankai.edu.cn/trRosetta/>). Furthermore, the 3D-models of the vaccine was validated by ProSA-webserver (Wiederstein and Sippl 2007) (<https://prosa.services.came.sbg.ac.at/prosa.php>) and Ramachandran plot implemented in Schrodinger tool.

2.3.12.5 In-silico Cloning of Vaccine Construct

The reverse translation along with the codon optimization of the vaccine constructs were carried out with Codon Adaptation tool(Grote et al. 2005) (<http://www.jcat.de/>) in order to generate the cDNA sequence of the vaccines. The K-12 strain of *E.coli* was used to express the vaccines. The expression level of the vaccines was evaluated by codon adaptation index (CAI) and overall GC content. Finally, the cDNA sequence of

the vaccines was inserted into pUC19 (2686 bp) employing SnapGene tool (<http://www.snapgene.com>).

The size of the vaccine construct consists of b-cell linear epitope (Vac-COVID-B) is found to be 399 base pair long whereas; the vaccine construct made of T-cell linear peptide epitopes (Vac-COVID-T) is 333 base pair (bp). The CAL value > 0.8 and GC composition between 30-70% of a cDNA segment is believed to be the benchmark for good expression of the gene in the host system(Sharp and Li 1987). The CAI value of both vaccines are found to be 1.0 and the GC content values for Vac-COVID-B, Vac-COVID-T are 60.65%, 56.75% respectively. This indicates the efficient expression of the adapted codons in *E.coli* K12 strain. Finally, we inserted the cDNA sequences of the vaccines computationally at the multiple cloning site of pUC19 vector. The size of the recombinant plasmid is 3401 bp and can be named as pUC-COVID. This study implies an efficient cloning strategy for the chimeric vaccines.

CHAPTER 3

COMPUTATIONAL INSIGHTS INTO FACTOR AFFECTING THE POTENCY OF DIARYL-SULFONE ANALOGS AS ESCHERICHIA COLI DIHYDROPTEROATE SYNTHASE INHIBITORS

Abstract: Dihydropteroate synthase (DHPS) is an alluring target for designing novel drug candidates to prevent infections caused by pathogenic *Escherichia coli* strains. Diaryl Sulfone (SO) compounds are found to inhibit DHPS competitively with respect to the substrate pABA (p-aminobenzoate). The extra aromatic ring of diaryl sulfone compounds found to stabilize them in highly flexible pABA binding loops. In this present study, a statistically significant 3D-QSAR model was developed using a data set of diaryl sulfone compounds. The favourable and unfavourable contributions of substitutions in sulfone compounds were illustrated by contour plot obtained from the developed 3D-QSAR model. Molecular docking calculations were performed to investigate the putative binding mode of diaryl sulfone compounds at the catalytic pocket. DFT calculations were carried out using SCF approach, B3LYP-6-31G (d) basis set to compute the HOMO, LUMO energies and their respective location at pABA binding pocket. Further, the developed model was validated by FEP (Free Energy Perturbation) calculations. The calculated relative free energy of binding between the highly potent and less potent sulfone compound was found to be -3.78 kcal/mol which is comparable to the experimental value of -5.85kcal/mol. A 10 ns molecular dynamics simulation of inhibitor and DHPS confirmed its stability at pABA catalytic site. Outcomes of the present work provide deeper insight in designing novel drug candidates for pathogenic *Escherichia coli* strains.

Keywords: Dihydropteroate Synthase; 3D-QSAR; Molecular Docking; Density Functional Theory; Free Energy Perturbation; Molecular Dynamics Simulation; Diaryl Sulfone.

3.1 BACKGROUND

Targeting DHPS (Qi et al. 2011; Zhao et al. 2016a) and its mutational studies (Yun et al. 2012a) have become an interesting topic for the development of antibacterial drugs candidates to prevent antibiotic-resistant pathogenic bacterial infections. Therefore, a deeper insight regarding diaryl sulfone binding between two flexible loops of EcDHPS will be helpful for synthesizing novel type of sulfa drugs with improved affinity to EcDHPS catalytic pocket. Computer simulation of biomolecules proved to be an important tool for better understanding of protein-ligand interactions and their stabilization patterns (Srivastava and Tiwari 2017) at the molecular level. Recently, DFT studies of protein-ligand complexes are reported to be effective for understanding them at their electronic level (Tao et al. 2009). The HOMO, LUMO orbital energies and their location at the protein-ligand complex are found to be important for predicting the type of non-bonded interactions which are crucial to achieve favourable recognition of ligands by protein molecules ultimately resulting to their stability in catalytic pocket (Correa-Basurto et al. 2012). DFT calculations also facilitate the mechanistic investigation of product formation in the catalytic pocket of enzymes (Malkhasian and Howlin 2016). Another important technique to measure the stability of the protein ligand complex is FEP (Free energy perturbation). The FEP/REST (Free energy perturbation/ replica exchange with solute tempering) are reported (Lenselink et al. 2016; Wang et al. 2015) to provide a rigorous algorithm to compute the difference in binding affinity due to presence or absence of specific functional groups in inhibitors. Several FEP studies were carried out to explain the effect of in-silico mutagenesis and SAR (structure-activity relationship) of different antagonist bound to enzymes (Chen et al. 2013; Goldfeld et al. 2015; Keränen et al. 2014). FEP calculations are also reported to be efficient in reducing the computational cost compared to absolute free energy calculations (Bash et al. 1987; Jorgensen and Ravimohan 1998). Therefore, the utility of these tools is found to be crucial to determine the factors affecting the stabilization of diaryl sulfone derivatives at DHPS catalytic pocket. The relative stability of the protein-ligand complex calculated by FEP can be compared with the experimental data.

In this present chapter, we selected a dataset of sulfone compounds to develop a model which clearly describes the effect of substitutions, affecting the potency of

sulfone compounds. Initially, 3D-QSAR model was developed to understand the structure-activity relationship (SAR) of diaryl sulfone molecules. The effect of the spatial arrangement of structural features on DHPS inhibition was explained by the contour plot obtained from 3D-QSAR model. Next molecular docking of all the sulfone molecules was carried out at the binding pocket of *pABA*. The result of molecular docking was correlated with the experimental activity value (pEI_{50}). Molecular docking helped in predicting the probable binding pose of sulfone molecule with respect to *pABA*. Further, we employ DFT (Density Functional Theorem) calculations to clarify the ligand-protein stability and interactions at electronic level. FEP/REST calculations were performed to compute the change in binding free energy due to functional group mutation. Lastly, 10 ns molecular dynamics was performed to confirm the stability of one of the top scoring compounds. The developed model provides insightful information regarding the mode of binding and important substitution required to stabilize diaryl sulfone compounds at EcDHPS catalytic pocket. This study will help in developing a better generation of Sulfa drugs or pterin-sulfa compounds with improved affinity and therapeutic activity to prevent the infection caused by pathogenic *E.coli*.

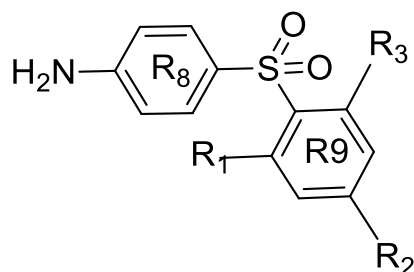
3.2 COMPUTATIONAL DETAILS

3.2.1 Ligand Preparation

A biological data set consisting of fifty diaryl sulfone(SO) derivatives(**Table 3.1**) was chosen from literature(De Benedetti et al. 1987, 1989; Hevener et al. 2010a; Lopez de Compadre et al. 1987) were used in the present study. The selected molecules from the biological dataset shared the same assay procedure (Richey and Brown 1969) with variation in substitution and potency profiles. The biological activity of the dataset was represented by the EI_{50} (Enzyme Inhibition Index₅₀) values and reported to have inhibitory activity spanned from 67.3 μ M to 0.21 μ M. In spite of having higher potency, compound 1 was not considered for 3D-QSAR calculations due to the absence of molecules having potency in the same activity range (Golbraikh et al. 2003). EI_{50} values were imported in Maestro (**Schrödinger Release 2017-2**: Maestro, Schrödinger, LLC, New York, NY, 2017) project table panel and converted to pEI_{50} . The 3D structures of the ligand were constructed using the builder panel in Maestro. The

geometry of the ligands was optimized after generating the structure using the Ligprep module (v4.9, schrodinger2017-2). Partial atomic charges were assigned and the possible ionization states were generated at pH of 8.2 ± 0.1 . The OPLS_2005 (Kaminski et al. 2001) force field was used for energy minimization of each ligand until it reached a root mean square deviation (RMSD) cut off of 0.01 \AA . Then the resulting structures were used for the modelling studies.

Table 3.1. structural detail, experimental activities, predicted activity of sulfone compounds (1-50).



Compounds	R1	R2	R3	Exp Activity (pEII ₅₀ /EXP)	Pred Activity (pEII ₅₀ /PRED)	**RA
1	ONa	ONa	H	8.302	-	-
2	CH ₃	O ⁻	H	6.63	6.55	-0.08
3	CH ₃	O ⁻	CH ₃	6.68	6.60	0.08
4	Cl	OH	H	6.58	6.59	0.01
5 ^t	Cl	O ⁻	H	6.58	6.57	-0.01
6	CH ₃	OH	H	6.46	6.46	0.00
7	CH ₃	CH ₃	O ⁻	6.3	6.02	-0.28
8 ^t	ONa	NH ₂	H	6.29	6.31	0.02
9 ^t	CH ₃	OH	CH ₃	6.29	6.29	0.00
10	OH	OH	H	6.14	6.17	0.03
11	ONa	OH	OH	5.83	5.80	-0.03
12 ^t	OH	O ⁻	OH	5.83	5.60	-0.23
13	H	ONa	H	5.82	5.83	0.01

14 ^t	NH ₂	NH ₂	H	5.79	5.82	0.03
15 ^t	CH ₃	CH ₃	CH ₃	5.77	5.85	0.08
16	CH ₃	OCH ₃	H	5.75	5.77	0.02
17 ^t	CH ₃	CH ₃	OH	5.75	5.72	-0.03
18 ^t	CH ₃	OCH ₃	CH ₃	5.75	6.03	0.28
19	CH ₃	CH ₃	H	5.72	5.68	-0.04
20	OH	OH	OH	5.71	5.76	0.05
21 ^t	Cl	Cl	H	5.65	5.86	0.21
22	NH ₂	H	H	5.64	5.62	-0.02
23	H	OH	H	5.61	5.64	-0.03
24	H	NH ₂	H	5.57	5.57	0.00
25	H	N(C ₂ H ₅) ₂	H	5.57	5.53	-0.04
26	CH ₃	CH ₃	OCH ₃	5.52	5.53	0.01
27	OCH ₃	OCH ₃	H	5.49	5.58	0.09
28	Cl	Cl	Cl	5.48	5.81	0.33
29	H	N(CH ₃) ₂	H	5.44	5.37	-0.07
30 ^t	H	NHOH	H	5.34	5.44	-0.10
31	H	NHC ₂ H ₅	H	5.34	5.33	-0.01
32	H	COOH	H	5.29	5.29	0.00
33	H	COONa	H	5.29	5.34	0.05
34 ^t	NH ₂	NHCHO	H	5.23	4.91	-0.32
35	H	NHCOCH ₃	H	5.23	5.26	0.03
36 [*]		<i>m</i> -NH ₂		5.22	5.24	0.02
37	H	CH ₃	H	5.21	5.18	-0.03
38	H	OCH ₃	H	5.18	5.12	-0.06
39 ^t	H	H	H	5.02	5.19	0.17
40	H	Br	H	4.96	4.97	0.01
41	OCH ₃	OCH ₃	OCH ₃	4.89	4.87	-0.02
42	H	F	H	4.74	4.73	-0.01
43 ^t	H	Cl	H	4.68	4.76	0.08
44	H	CONH ₂	H	4.64	4.73	0.09

45	H	COOCH ₃	H	4.46	4.40	-0.06
46	NO ₂	NO ₂	H	4.44	4.46	0.02
47 ^t	H	COCH ₃	H	4.36	4.37	0.18
48	H	NO ₂	H	4.34	4.35	0.01
49	H	CON(C ₂ H ₅) ₂	H	4.30	4.46	0.16
50	H	CN	H	4.17	4.07	-0.1

1) ^tDefines that the compound is in the test-set list for 3D-QSAR analysis.

2) *Defines the functional group is in the *meta* position of the R9 pharmacophoric ring only

3) **Residual activity (RA) defines the difference between phase predicted activity and experimental activity of diaryl sulfone compounds.

3.2.2 3D-QSAR modelling

Phase (v 4.9) (Dixon et al. 2006) was used to generate 3D-QSAR models for the diaryl sulfone-based inhibitors of DHPS. The prepared structures of SO derivatives were selected in workspace navigator panel in Maestro (**Schrödinger Release 2017-2**: Maestro, Schrödinger, LLC, New York, NY, 2017) interface and imported to the atom-based 3D-QSAR model development panel with their respective biological activity value (pIIE₅₀). Before building the 3D-QSAR model, all prepared ligands were aligned using common scaffold based alignment, a type of flexible ligand alignment tool in Phase and Largest Common Bemis-Murcao scaffold (Bemis and Murcko 1996) was selected as a parameter. The atom-based 3D-QSAR model considers all the atoms into account; whereas pharmacophore based QSAR model does not consider the ligand feature beyond pharmacophore model. Consequently, atom based 3D-QSAR proved to be more useful for predicting the true structure-activity relationship (Shah et al. 2010; Verma et al. 2010) of diaryl sulfone compound at the *p*ABA binding pocket. The dataset was randomly divided, 75% structures (to maintain the standard 3:1 ratio) were kept as training set and rest molecules were automatically considered as test-set by using “Automated Random

Selection” option present in the Phase module. Care was taken in order to include the most active and inactive molecules in training set (Dixon et al. 2006; Golbraikh et al. 2003). The maximum number of Partial Least Square (PLS) factor was N/5 (Where N in the number of training set molecules). Use of higher PLS factor could cause over fitting of data, hence optimal six PLS factor was used (Polański et al. 2002). The 3D-QSAR model was generated by keeping a grid spacing of 1Å. The equation for calculating the predicted activity can be given as

$$PEII_{50/PRED} = y + a_1HBD_1 + \dots + a_nHBD_n + b_1A_1 + \dots + b_nA_n + c_1S_1 + \dots + c_nS_n + \dots + d_1EW_1 + \dots + d_nEW_n \quad (3.1)$$

Where, HDB, A, S and EW stand for the descriptors related to hydrogen bond donor, anionic, Steric or hydrophobic group and the electron withdrawing group. Further, $a_{1..n}$, $b_{1..n}$, $c_{1..n}$ and $d_{1..n}$ are the empirical constants derived from PLS method. Further, contour plot analysis was performed to interpret and recognize the important pharmacophoric requirements at spatial sites of the structures by the cubic 3D grid. The biological activity of the training set molecules was evaluated by the generated 3D-QSAR model in order to estimate the quality of the 3D-QSAR model. Lastly the stability and predictivity of the developed 3D-QSAR model was examined with an external test set of ten diaryl sulfone compounds (Lopez de Compadre et al. 1987).

3.2.3 Molecular Docking

The co-crystal structure of DHPS with sulfonamide (PDB ID: 1AJ0, resolution 2.0Å) was obtained from Protein Data Bank (Achari et al. 1997b) and the docking of the fifty drug molecules were carried out using an automated docking program namely Autodock (v4.2.6) which considers the protein to be rigid during docking. Prior to docking, DHPS structure was prepared (Sastry et al. 2013) using Autodock graphical user interface (Morris et al. 2009). The co-crystallized ligand, water molecules were removed and Gasteiger charges (Gasteiger and Marsili 1980), polar hydrogen were added to the protein structure. The energy minimization of protein was performed till RMSD of 0.30 Å conjugate gradient steps using OPLS-2005 force-field (Sastry et al. 2013). The minimized ligands were taken and the bond associated with the sulphone groups were set as rotatable bonds (Lopez de Compadre et al. 1987). A 3D grid was created at the binding site of the protein having a size of 30Å × 24Å × 22Å with a default

spacing of 0.375Å, using an Auto-Grid algorithm to quantify the binding affinity of ligand at the catalytic site of DHPS. The grid includes all eight active site residues namely Thr62, Arg63, Phe190, Ser219, Arg220, Lys221, His257 and SO₄284 (Achari et al. 1997b). The molecular docking simulations were performed using Lamarckian Genetic Algorithm (MORRIS et al. n.d.). The initial population and the number of energy evaluations were fixed to 150 and 2.5×10^6 respectively for all molecular docking simulations. The co-crystallized sulphanilamide was removed and re-docked into the same position and RMSD between them was calculated to assess the reliability of docking results (Tripuraneni and Azam 2016). Several clusters were obtained for the docking orientation within RMS (Root Mean Square) deviation of 0.5Å and the lowest energy cluster obtained for each sulfone compounds were used for further analysis.

3.2.4 DFT Calculation Setup

Molecular orbital calculations were performed to explore the binding site of DHPS (i.e., the *p*ABA binding pocket) to analyse the protein-ligand interaction at electronic level. It was evident from the frontier molecular orbital theory that HOMO of the ligands interacts with LUMO positioned in the binding pocket of the protein (Correa-Basurto et al. 2012). In order to determine the location of HOMO, LUMO orbitals in ligands single point energy calculations were performed by Gaussian09 software package (“Gaussian 09 Citation | Gaussian.com” n.d.), using SCF(Self Consistent Field) approach(Tomasi et al. 2005). B3LYP(Lee et al. 1988a) functional and 6-31G(d) (Otsuka Takao et al. 2015) basis set was used to compute single point energy of sulfone compounds. The recognition ability of the protein was explained by evaluating the HOMO and LUMO energies of the ligands (Correa-Basurto et al. 2012). The HOMO, LUMO calculations of the binding site residues was also performed to explain the ligand-binding mechanism at the catalytic pocket of DHPS. Since the calculation of the whole protein molecule with ligand is computationally expensive so we took only the interacting amino acids and ligands for this study. The generated polypeptide cluster was capped with N-actyl group at the N-terminal end and N-methyl amide group at C-terminal end (Duan et al. 2007). The interacting amino acid residues at *p*ABA binding pocket (Achari et al. 1997b)were considered for single point energy calculation at B3LYP/6-31G(d) level. The resulting amino acid cluster was found to have 176 atoms.

The cubegen utility in Gaussian09 software package was used to generate the HOMO, LUMO orbitals. The docked poses of highly potent and less potent compounds along with the interacting amino acids were chosen for calculating HOMO, LUMO energies using abovementioned basis set and functional. The location of HOMO, LUMO orbitals on the ligand-protein complex was visualized by Gabedit interface (v2.4.7).

3.2.5 Relative binding free energy calculation

FEP/REST (Free Energy Perturbation/ Replica exchange with solute tempering) calculations were performed to compute the binding free energy difference between the less potent and highly potent sulfone compounds (these compounds share the same congeneric scaffold) using Desmond MD (v4.8) (Bowers et al. 2006) suite distributed by Schrödinger. The incorporation of REST method in FEP calculations enhances sampling in phase space with an efficient hopping protocol. The efficient sampling of relevant conformations of ligands helps to measure relative binding affinity accurately within an easily accessible simulation time period (Wang et al. 2012). In FEP/REST calculation a small region of interest especially the localized region surrounding the binding pocket including the ligand is heated up by keeping rest of the system cold (Liu et al. 2005). FEP/REST captures the free energy difference of two similar systems through an alchemical transformation pathway. This pathway involves an array of discrete steps starting from initial lambda window ($\lambda = 0$) to final lambda window ($\lambda = 1$) (Wang et al. 2011). The free energy difference was calculated applying Bennett Acceptance ratio (BAR) method (Bennett 1976) and the error for each free energy calculation was estimated by bootstrapping (Paliwal and Shirts 2011; Pohorille et al. 2010). Desmond (v 4.8) programmes are reported to provide a good single node, parallel performance compare to other MD packages (Bowers et al. 2006). The OPLS-2005 force field was employed to generate the essential force field parameters required for performing the FEP simulations due to its accuracy in predicting the free energy of solvation of drug-like molecules (Shivakumar et al. 2010). The pose view file of minimized 1AJ0 and sulfone compounds was imported in the FEP panel of Desmond (v4.8), the core RMSD difference between two compounds was 0.004 Å. The complexes of protein and ligand molecules were solvated in SPC (Simple point-charge) water model (Ferguson David M. 1995) in an orthorhombic box with 5 Å buffer size

and minimized with the limited memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) method (“Practical Methods of Optimization, 2nd Edition,” n.d.). Initially a 100 ps MD simulation was carried out under NVT ensemble applying a force constant of 50 kcal/mol/Å² for minimizing the protein and ligand heavy atom. 1 fs time step was maintained for aforementioned 100 ps simulation. The solute heavy atoms were restrained to their initial position and 10 K temperature was maintained during the simulation. Next, two MD simulations of 12 ps, 36 ps were run under NVT and NPT ensemble respectively maintaining the same force constant to equilibrate the system with 1 fs time step. The system was relaxed with 240 ps MD simulation under NPT ensemble without applying any restrains on heavy atoms keeping time step as 1 fs. Ultimately, 5 ns production simulations were continued under NPT ensemble for both solvent and protein complex using Nose- Hoover thermostats(Martyna et al. 1992), Martyna-Tobias-Klein barostat(Martyna et al. 1994) at effective temperature of 300K temperature and 1 atm pressure. For production run 1fs time step was maintained throughout the simulation time. FEP/REST simulations were carried out with 12 lambda windows and 1.2 ps interval was kept to exchange replicas between two neighbouring lambda windows (Fukunishi et al. 2002). The acceptance ratio of replica exchange was maintained at 0.3. The convergence analysis of free energy during the course of simulation was also performed to check whether the simulation time was sufficient for free energy calculation. The relative binding energy at various thermodynamic states was calculated from the trajectories using the *fepmapper_sid2pdf.py* python script for further analysis (Bowers et al. 2006).

3.2.6 ADME/Toxicity Prediction Procedure

The drug-like nature of diaryl sulfone compounds were predicted by evaluating their pharmacokinetic (PK) profiles. The ADME/Tox properties of the top ten dock scored SO compounds were calculated using QikProp (v5.7)(Jorgensen and Duffy 2000) (Schrodinger-2018/3). The structures were already prepared with Ligprep module and imported to QikProp panel for ADME calculation. The molecular weight (MW), octanol/water partition co-efficient (logP_{o/w}), dipole moment, H-bond donor and H-bond acceptor etc. properties were calculated for top scored diaryl sulfone compounds

and evaluated in accordance with “Lipinski’s rule of 5” (Lipinski et al. 1997) to access their drug-likeness.

3.2.7 System Setup for Molecular Dynamics Study

The dynamic behaviour of diaryl sulfone bound EcDHPS (PDB ID: 1AJ0, 2.00 Å) was observed through all atom molecular dynamics simulation study. The docking predicted binding mode of highly active sulfone compounds were evaluated by observing their dynamic nature in EcDHPS active site. Desmond (v5.4) software package (Bowers et al. 2006; Shivakumar et al. 2010) was employed to perform the molecular dynamics of 1AJ0 and compound 4 in its bound state. OPLS_2005 all atom force field (Kaminski et al. 2001; Shivakumar et al. 2009, 2010; William L. Jorgensen et al. 1996) was applied to generate the necessary topology and parameter files required for molecular dynamics simulation. The entire system was inserted in the centre of an orthorhombic periodic box with 10Å buffer region from the protein surface to fill with water. The periodic box volume of 1AJ0/compound 4 complex was 313617 Å³. The whole system was solvated with SPC (simple point-charge) water molecules (Mark and Nilsson 2001a) and neutralized by adding thirteen Cl⁻ ions. The solvated protein structure in the periodic box contained 29076 atoms including 4381 atoms of 1AJ0 and compound 4. Next, the system was minimized applying steepest descent algorithm and gradient threshold was kept at 25 kcal/mol/Å. The maximum number of iterations during minimization was kept 2000 steps until a convergence threshold of 1.0 kcal/mol/Å was attained. Subsequently, NVT (constant number of atoms N, volume V and temperature T) equilibration was performed at 310K for 5 ns using Nose-Hoover thermostat (Martyna et al. 1992)(thermostat relaxation time = 200 ps)with a time step of 2 fs. Next to NVT, NPT (constant number of atoms N, pressure P and temperature T) equilibration was performed using Nose-Hoover thermostat and Martyna-Tobias-Klein barostat (Martyna et al. 1994)(P=1 atm, barostat relaxation time= 200 ps and thermostat relaxation time= 200 ps) at 310K with a time step of 2 fs for 5 ns. The solute (protein and ligand) heavy atoms were restrained during the equilibrations, applying force constant of 100 Kcal/mol/Å². Lastly, the equilibrated system was taken for production run. The restrains on solute heavy atoms were removed and 10 ns production MD was carried out in NPT ensemble using Nose-Hoover thermostat and Martyna-Tobias-Klein

barostat with a time step of 2 fs. multiple time step RESPA integration algorithm was implemented for all simulation steps with a time step of 2fs for bonded, 2fs for ‘near’ non-bonded and 6fs for ‘far’ non-bonded interactions. Simulation data were retrieved at each 10 ps and the visual inspection of three-dimensional structures, trajectories were done using Maestro graphical interface.

3.3. RESULTS AND DISCUSSION

3.3.1 3D-QSAR model construction and Visualization of contour maps

The atom-based 3D-QSAR analysis generated a statistically significant model along with good predicted activities of test set compounds when PLS factor was 6. The Biological activity predicted by the generated 3D-QSAR model is documented in **Table 3.1**. The PLS regression summary for generated 3D-QSAR is represented in **Table 3.2**. The low value of standard deviation (SD) (0.1171) and root-mean-square deviation obtained from the present study define the significance and reliability of the model. Moreover, high regression coefficient (0.9737) for training set, in addition with the stability ranges from 0.862 to 0.58 on maximum scale of 1, F value (Variance ratio) of 166.5 with smaller P value (<0.005) and Pearson-r of 0.9183 reflected the relevance and confidence of the model respectively (**Table 3.2**). The stability for predicting unknown compounds in the test set for the generated 3D-QSAR model was indicated by the low value of RMSE (Root Mean Square Error), P (Significance level of variance ratio) and SD (**Table 3.2**). The 3D-QSAR was validated by its closeness in predicting the activity of test set ligands (**Table 3.1**). **Figure 3.1** illustrates the scatter plots which describe that the experimental and predicted activity of sulfone derivatives showed moderate difference and good linear correlation between experimental and Phase predicted biological activity values.

The developed 3D-QSAR model was further validated by predicting the experimental activity of compounds not included in model development (external test set). The predicted pEII₅₀ values of the compounds in external test set are documented in **Table 3.3**. A scatter plot of experimental vs. predicted pEII₅₀ values of external test set is illustrated in **Figure 3.1-C**. The plot of predicted activity vs. the residual activity is also shown in **Figure 3.1-D**, which was used to identify the outlier of the developed 3D-QSAR model.

Table 3.2. Regression summary of generated 3D-QSAR model

PLS	SD	R ²	F	P	Stability	RMSE	Q ²	Pearson-r
1.	0.3744	0.6811	68.3	1.92e-09	0.862	0.39	0.6446	0.8572
2.	0.2992	0.8027	67.1	1.19e-11	0.828	0.36	0.6936	0.8709
3.	0.2160	0.9005	90.5	3.94e-15	0.702	0.31	0.7720	0.8925
4.	0.1719	0.9391	111.8	3.48e-17	0.647	0.34	0.7312	0.8713
5.	0.1418	0.9600	134.2	1.16e-18	0.632	0.29	0.8067	0.9092
6.	0.1171	0.9737	166.5	5.02e-20	0.58	0.30	0.7948	0.9183

It was found that, the QSAR model was able to predict the experimental activity with a correlation coefficient (R²) of 0.74 and cross-validation coefficient (Q²) of 0.61. The R² value and Q² above 0.5 indicate the generated 3D-QSAR model to be good and able to endorse the experimental inhibitory activity (pEII₅₀) of compounds included in external test set (Golbraikh and Tropsha 2000a). Moreover, **Figure 3.1-D** illustrates no outlier in this study. Hence, the model can be considered as stable. In order to interpret the effect of the spatial arrangement of structural features on biological activity, contour plot analysis was performed. The structural features were the presence of H-bond donor group, hydrophobic group, electron withdrawing group and an anionic group. contribution was shown by red cubes. Comparisons of most significant favourable and unfavourable interactions are represented in **Figure 3.2**. Their individual positive contribution was represented by blue cubes and negative Since the substitutions were mainly situated at R9 ring, the contour plot appeared at 2', 4' and 6' position of this region. The list of functional groups presents at 2', 4', 6' site of the R9 ring of all 50 compounds is given in **Table 3.1** to correlate the contour plot with the substitution. The generated 3D-QSAR model was implemented on compound 39 (pEII₅₀= 5.02) for better visual inspection of the model. The effects of different substituents are discussed as follows.

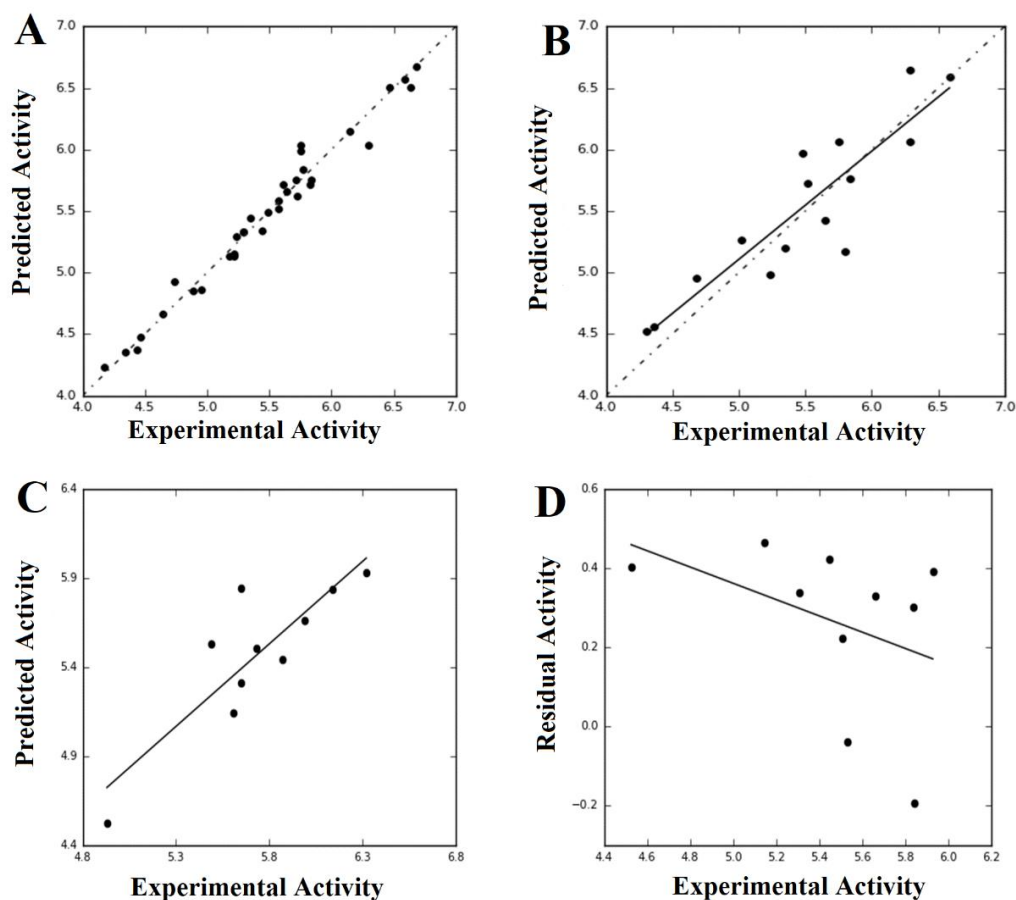


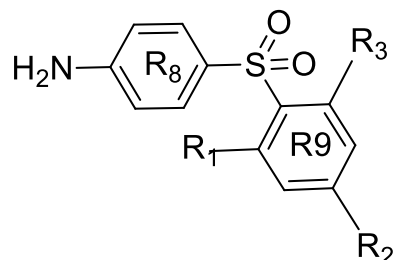
Figure 3.1. Fitness graph between observed activity versus phase-predicted activity for (A) training set (B) test set compounds with the best fit line [$y = 0.88x + 0.71$ ($R^2 = 0.81$)]. (C) Plot of actual value vs. predicted value of external test set $y=0.93x+0.16$ ($R^2=0.74$). (D) Plot of residual activity vs. predicted value of external test set.

3.3.1.1 Presence of Anionic Group

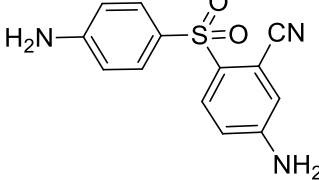
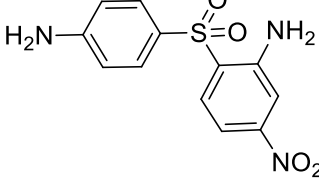
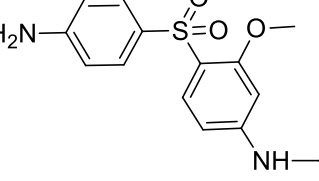
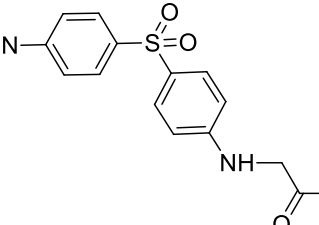
It can be identified from **Table 3.1** that highly active compound 2 ($pEI_{50} = 6.63$), compound 3 ($pEI_{50} = 6.68$), compound 4 ($pEI_{50} = 6.58$) and moderately active compound 11 ($pEI_{50} = 5.83$) have phenoxide group in 2', 4' and 6' position of R9 pharmacophoric ring, which is anionic or electron donating in nature. The blue cube region at 2', 4' and 6' position (**Figure 3.2-A**) is representing the favourable and unfavourable contribution of anionic groups which completely agrees with the

aforementioned compounds. Although compound 1 ($pEI_{50} = 8.302$) was not considered in the 3D-QSAR calculation, it is having phenoxide ions at 2' and 4' position.

Table 3.3. Calculated pEI_{50} for compounds in external test set **Cpd-01**-to **Cpd-10**



compounds	Structures	Experimental activity	Predicted Activity	RA
Cpd-01		6.320	5.928	0.392
Cpd-02		6.140	5.837	0.303
Cpd-03		5.990	5.661	0.329
Cpd-04		5.870	5.446	0.424
Cpd-05		5.610	5.145	0.465
Cpd-06		5.730	5.570	0.223

Cpd-07		5.650	5.834	-0.193
Cpd-08		4.930	4.526	-0.404
Cpd-09		5.490	5.530	-0.040
Cpd-10		5.650	5.310	-0.340

Presence of electron donating groups at R9 ring increases the electron density in R9 ring and SO₂ unit of SO compounds hence increase the potency of ligands (Lopez de Compadre et al., 1987). A remarkable decrease in biological activity was found for compound 50 (pEII₅₀= 4.17), compound 46 (pEII₅₀= 4.44), compound 44 (pEII₅₀= 4.64) due to lack of anionic or electron donating group in the blue cube region (**Figure 3.2-A**) of R9 ring. However, a too bulky group at 6' position of the R9 ring may give some unfavourable anionic contribution in ligands.

3.3.1.2 Presence of Hydrogen Bond Donor Group

For hydrogen bond donor aspect, the blue cubes at 2' and 4' indicate the preference of hydrogen bond donor (**Figure 3.2-B**) groups at those positions. Small groups like –NH₂, –OH at 2' and 4' position of R9 ring generally contributes to favourable hydrogen bonding interactions. It is found from **Table 3.1** that most of the highly active compounds like compound 4, compound 6 (pEII₅₀= 6.46), compound 8 (pEII₅₀= 6.29), compound 9 (pEII₅₀= 6.29) and compound 10 (pEII₅₀= 6.14) have –OH group in 2' position and 4' position, hence increment in activity was found for those compounds.

Whereas the presence of bulky substituents like $-\text{CONH}_2$ groups and absence of hydrogen bond donor groups in those positions result in a decrease in biological activity of compound 44 ($\text{pEII}_{50}=4.64$).

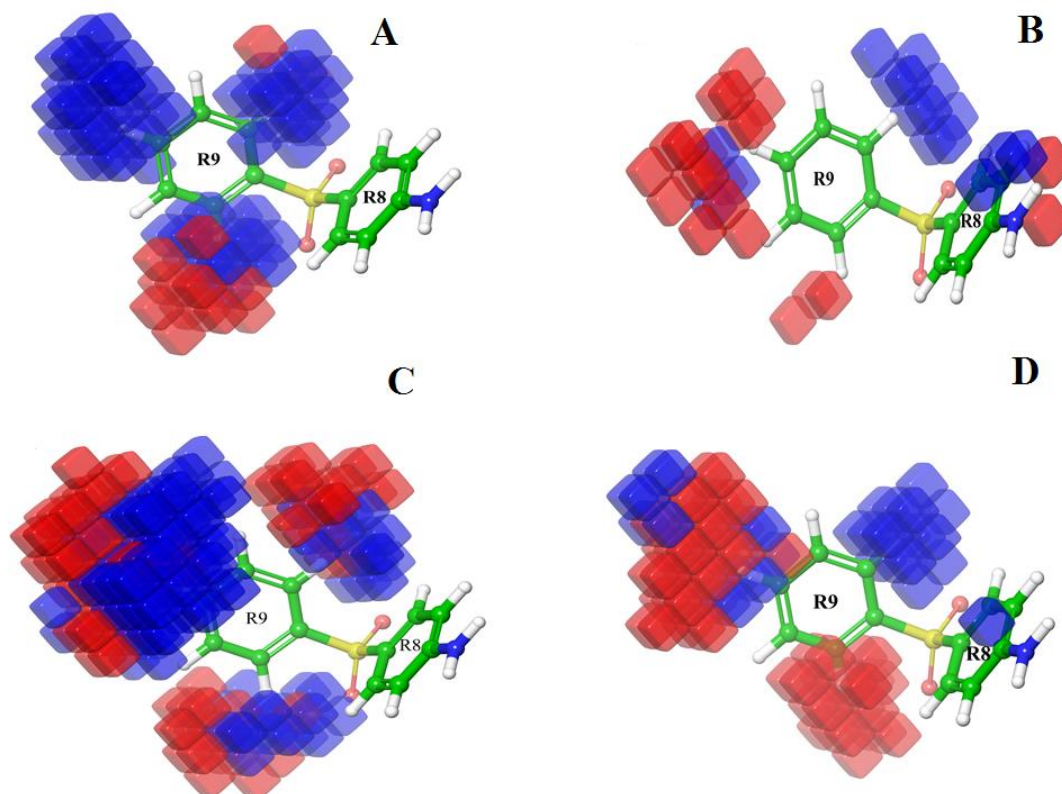


Figure 3.2. Visualization of 3D-QSAR contour plot on compound 39($\text{pEII}_{50}= 5.02$) Effect of A) presence of the anionic group, B) hydrogen bond donor group, C) hydrophobic group D) electron withdrawing group predicted by generated 3D-QSAR model.

3.3.1.3 Presence of Hydrophobic group

The other important component that impacts on biological activity is a hydrophobic character, as displayed in **Figure 3.2-C** which represents the presence of blue cubes at 2', 4', 6' positions of the R9 pharmacophoric site due to the presence of small hydrophobic groups for favourable hydrophobic interaction with 1AJ0 catalytic pocket. This assumption is supported by highly active compound 3, compound 7($\text{pEII}_{50}= 6.30$), compound 9 and moderately active compound 15 ($\text{pEII}_{50}=5.77$), compound 17($\text{pEII}_{50}= 5.75$) and compound 19($\text{pEII}_{50}=5.72$) because $-\text{CH}_3$ is substituted at 2' and 4' positions

(**Table 3.1**) in aforementioned ligands. The presence of hydrophobic groups can induce stiffness in the covalent bond which is attached with sulphonyl group, hence decreases the entropy of ligands, and increase the activity (Lopez de Compadre et al., 1987). Compound 50, compound 48 (pEII₅₀=4.34), compound 44 lack hydrophobic groups in 2' and 4' which is responsible for the low 1AJ0 inhibitory activity.

3.3.1.4 Presence of Electron-Withdrawing Group

The existence of red cubes at 2' and 4' position of the R9 pharmacophoric feature indicates the unfavourable position of electron withdrawing group (**Figure 3.2-D**). The presence of electron withdrawing group in R9 pharmacophoric ring reduces the electron density in this ring and sulphonyl unit, hence decreases the inhibitory activity of sulfone compounds. Compound 50 (pEII₅₀=4.17), compound 46 (pEII₅₀=4.44), compound 48 (pEII₅₀=4.34), compound 47 (pEII₅₀=4.36) and compound 49 (pEII₅₀=4.3) have electron withdrawing group at 4' and 2' position resulting decrease in biological activity (**Table 1**).

3.3.2 Molecular docking

Molecular docking suggested that the interactions were mainly influenced by hydrogen bonding and π - π stacking interactions due to the presence of two aromatic rings in sulfone compounds. The critical distances for donor-acceptor and donor-donor were set to 0.35 nm and 0.245 nm respectively to identify the hydrogen bond between ligand and protein complexes from molecular docking studies. Further, the cutoff for \angle DDA was set to 30°. Further, the π - π stacking interaction between the aromatic rings of ligand and the aromatic amino acid sidechains at the catalytic pocket of protein is said to be formed if the inter-centroid distance is less than 0.8 nm and the angle between the normal of one or both rings and the centroid-centroid vector must be between 0° to 60° or 120° to 180°. In case of T-shaped π - π stacking interaction, the above-mentioned angle must be ~90°. Additionally, at least one atom of each ring needs to be within 0.45 nm to form π - π stacking. In case of π -cation interaction, the distance and angle between the π -system and the cation centre must be 0.6 nm and 60° $\leq\theta\leq$ 90° respectively. It is found from **Appendix I** and **Table 3.4** that as we move from compounds substituted with electron donating groups to electron withdrawing groups, the inhibitory potency of

sulfone drug molecules decreases. The substitution at ortho-position is attributed to the stability of complex as it helped the molecule to fit into the binding site more rigidly (Lopez de Compadre et al., 1987). The interactions were mainly resided in the region of Thr62 to His 257 due to the presence of catalytic site in this portion. It was found that Thr 62, Arg220 and Ser222 involved in hydrogen bonding network with most of the sulfone (SO) derivatives (Compound 1-50) and Phe190, His257 exhibited π - π stacking interaction with the R8, R9 aromatic rings present in the SO molecules.

Table 3.4. The docking score of the compounds 1-50.

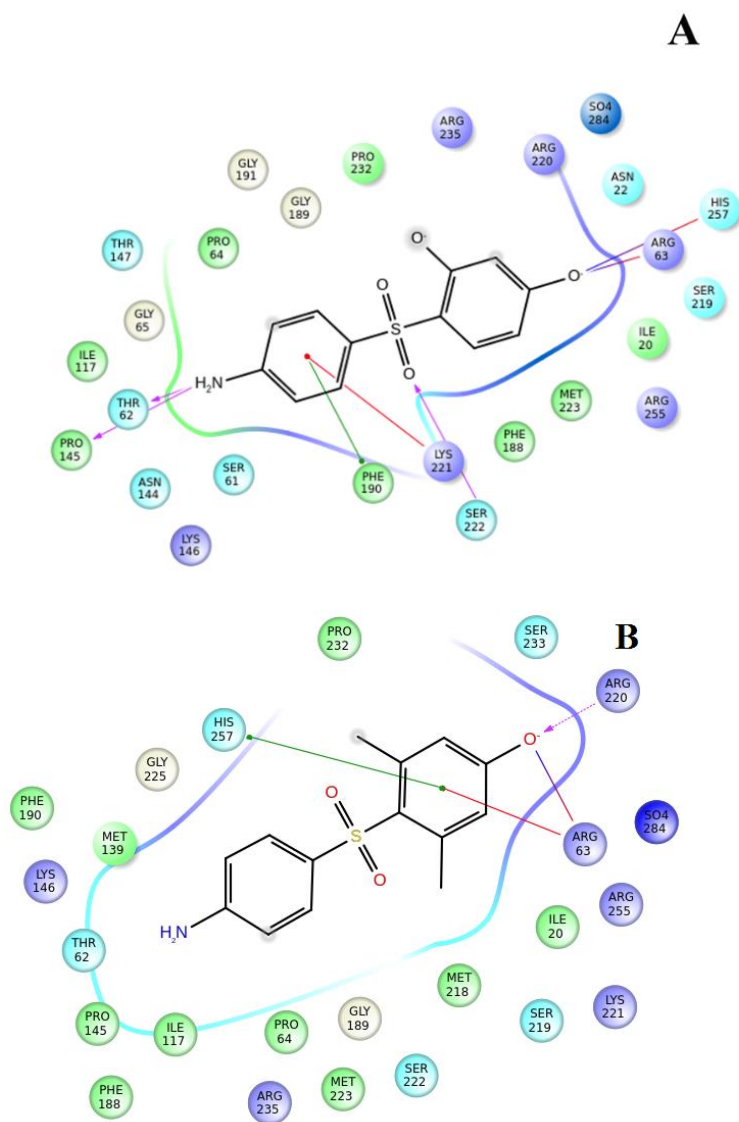
Compound No	Docking Score (kcal/mol)	Compound No.	Docking Score (kcal/mol)
1	-6.09	26	-5.22
2	-5.68	27	-5.15
3	-5.88	28	-5.39
4	-5.7	29	-4.83
5	-5.66	30	-5.16
6	-5.86	31	-4.77
7	-5.01	32	-4.81
8	-4.97	33	-5.31
9	-5.89	34	-5.47
10	-5.22	35	-4.63
11	-5.42	36	-5.05
12	-4.74	37	-4.97
13	-5.14	38	-5.28
14	-4.43	39	-4.88
15	-5.67	40	-5.01
16	-5.89	41	-4.91
17	-5.06	42	-4.71
18	-5.76	43	-4.76
19	-5.50	44	-5.41
20	-4.74	45	-4.62
21	-5.66	46	-4.06
22	-5.13	47	-4.05
23	-5.09	48	-5.72
24	-4.49	49	-4.22
25	-4.80	50	-4.03

The binding energies of sulfone derivatives were found to span from -6.09 Kcal/mol to -4.03 Kcal/mole at the *p*ABA binding site of DHPS and showed good correlation with experimental activity value (**Table 4.3**). The ligand interaction diagram of docked compounds was illustrated in **Figure 3.3**. The inhibitory activity of highest potent drug molecule (Compound 1) can be attributed due to the presence of two electron donating

groups -ONa (which easily dissociates as O^- and Na^+) at *ortho* and *para* position in the ring R9 which enhance the electron density in common moiety (4-NH₂-C₆H₄SO₂) (ring R8) (De Benedetti et al., 1989) supporting for stronger π - electron interactions. The anionic functional group of the ligand was found to form salt bridges with Arg63 and His257 residues, resulting in extra stabilization of compound 1 at the *p*ABA binding pocket (**Figure 3.3-A**). Similar observations were also implied by the generated 3D-QSAR model that the presence of negative ionic groups at R1, R2, and R3 are desirable for biological activity. The protein-ligand complex of compound 1 was found to be stabilized by three hydrogen bonding interaction, π - π interaction, π -cation interaction, and salt bridge formation (**Figure 3.3-A**). The carbonyl groups of Thr62 and Pro145 residues were contributing hydrogen bonding interactions with amine hydrogen of compound 1 at a distance of 1.98Å (Thr62-C=O --- H-N) and 1.92Å (Pro145- C=O --- H-N) respectively. Sulphonyl oxygen of compound 1 showed hydrogen bonding with the amide hydrogen of Ser222 residue (S=O ---H-N- Ser222). Phe190 and Lys221 exhibited π - π stacking and π - cation interactions respectively with the ring R8 of the most potent drug. In addition to these interactions, extra stabilization was found between the phenoxide ions in ligand and the amino acid residues Arg63 and His257 due to salt bridge formation. The binding energy of highly potent compound 3 and compound 4 also showed good agreement in accordance with the 3D-QSAR results. The methyl groups present at 2', 6' position in the R9 ring of compound 3 was found to involve in hydrophobic interactions with Pro232, Arg220 and Lys221 residues respectively and presented in **Appendix II**. Both the compounds 3 and 4 exhibited one hydrogen bonding interaction in the complex (O^- --- H-N- Arg220 and H-O --- H-N- Arg220 respectively) and also showed π - π stacking interaction with His257 (**Figure 3.3-B & C**). Arg63 and Lys221 were found to form π -cation interaction with the ring R9 and the ring R8 respectively. The lowest potent drug molecule (compound 50) showed hydrogen bonding interactions with the amino acid residues Thr62 (Thr62C=O--HN), Ser222 (Ser222NH---O=S) and π -cation interaction with Lys221. The formation of the salt bridge was absent in the compound 50 due to the absence of anionic group (**Figure 3.3 D**). In addition, compound 50 was found to make steric clashes with the side chain of Lys 221 (**Appendix III-D**). Fewer non-bonded interactions along with

unfavourable steric clashes destabilize the **compound 50** in flexible *p*ABA binding site and cause lower potency of this compound.

The reliability of docking results was assessed by re-docking the co-crystallized ligand (*p*ABA) of protein DHPS (PDB ID: 1AJ0) into the binding site of the protein and calculating RMSD values between re-docked pose and the X-ray crystal structure. It is considered that RMSD values up to 2Å are reliable (Kramer et al., 1999). The docking pose and the interactions obtained after re-docking matches with the known crystal structure conformation with RMSD value of 1.21Å (**Appendix IV**).



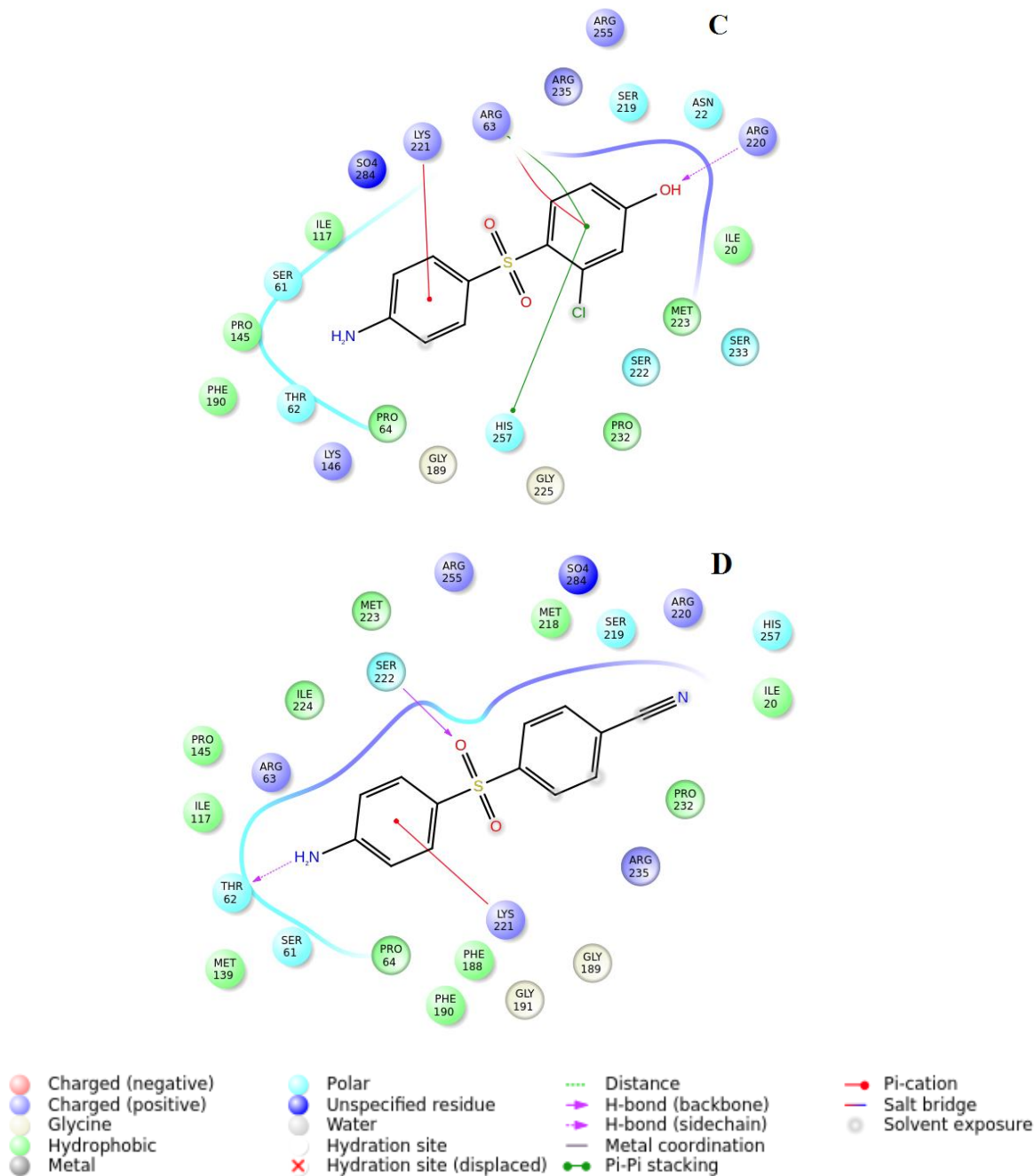


Figure 3.3. 2D-ligand interaction diagram of A) compound 1, B) compound 3, C) compound 4 and D) compound 50 in the catalytic pocket of 1AJ0.

3.3.3 Frontier molecular orbital analysis

The HOMO, LUMO energies of highly active compound 1, compound 4, moderately active compound 15 and less active compound 21, compound 50 were calculated and documented **Table 3.5**. The abovementioned compounds were chosen to observe how

functional group substitutions in R9 ring of diaryl sulfone compounds contribute for better stability of these derivatives in *p*ABA binding pocket of EcDHPS. Compound 1 was chosen to observe the effect of two phenoxide groups (O⁻) (anionic in nature) present at 2' and 4' position of R9 ring. Compound 4 was chosen to see the effect of –OH group (hydrogen bond donor group) located at 4' position of R9 ring. Moderately active compound 15 was found to be important to investigate the effect of –CH₃ substitution at 2', 4' and 6' position of the same ring. We also studied the effect of electron withdrawing group in compound 21, compound 50 having chloro (-Cl) group at 2', 4' position and cyano group (-CN) at 4' position of R9 ring respectively. The variation in conformational energy due to substitutions in R9 was assessed from the differences of HOMO, LUMO orbital energies of chosen sulfone compounds and listed in **Table 3.5**.

Table 3.5. HOMO- LUMO Energy details of Ligands at their bound state and unbound state

HOMO- LUMO Energy Gap Details								
Sl. No	Compound				Complex			
	E _{compound} (eV)	*E _{HOMO} (eV)	*E _{LUMO} (eV)	*E _{gap} (eV)	E _{complex} (eV)	*E _{HOMO} (eV)	*E _{LUMO} (eV)	*E _{gap} (eV)
1	-33133.41	-5.977	-1.240	4.737	-146150.07	-0.774	-0.456	0.318
4	-43591.84	-5.877	-1.388	4.489	-135677.67	-0.508	-0.167	0.341
15	-32248.50	-5.856	-1.318	4.538	-90284.429	-0.444	-0.095	0.349
21	-54051.36	-6.095	-1.892	4.203	-126509.33	-0.212	0.083	0.295
50	-31550.25	-6.236	-2.220	4.016	-86158.636	-1.499	-1.142	0.357

*E_{HOMO} and E_{LUMO} defines the energy of HOMO, LUMO orbitals and E_{gap} = E_{LUMO} - E_{HOMO}.

From **Table 3.5** it is evident that HOMO, LUMO energy differences, E_{gap} (eV) of highly potent and less potent compounds are in same range. This shows that substitution of various functional groups at R9 ring does not play much role in improving

conformational stability and recognition ability. However, the location of HOMO, LUMO orbitals and the total energy will be different when these ligands are bound to the protein. Therefore, it will be interesting to study the position of HOMO, LUMO and their energies in interacting amino acids and ligand complex. We have presented the HOMO, LUMO energy details of the complexes in **Table 3.5** which will be discussed later. The distribution of HOMO orbitals on the sulfone molecules are illustrated in **Figure 3.5**. Red and blue colours represent the positive and negative lobes of HOMO orbitals located at the sulfone compounds. It is clear from **Figure 3.4-A, B** that HOMO orbitals cover the entire R8 and R9 position for highly potent compound 1 and compound 4. Moderately active compound 15 found to have some portion of HOMO orbitals distributed over R9 ring (**Figure 3.4 C**).

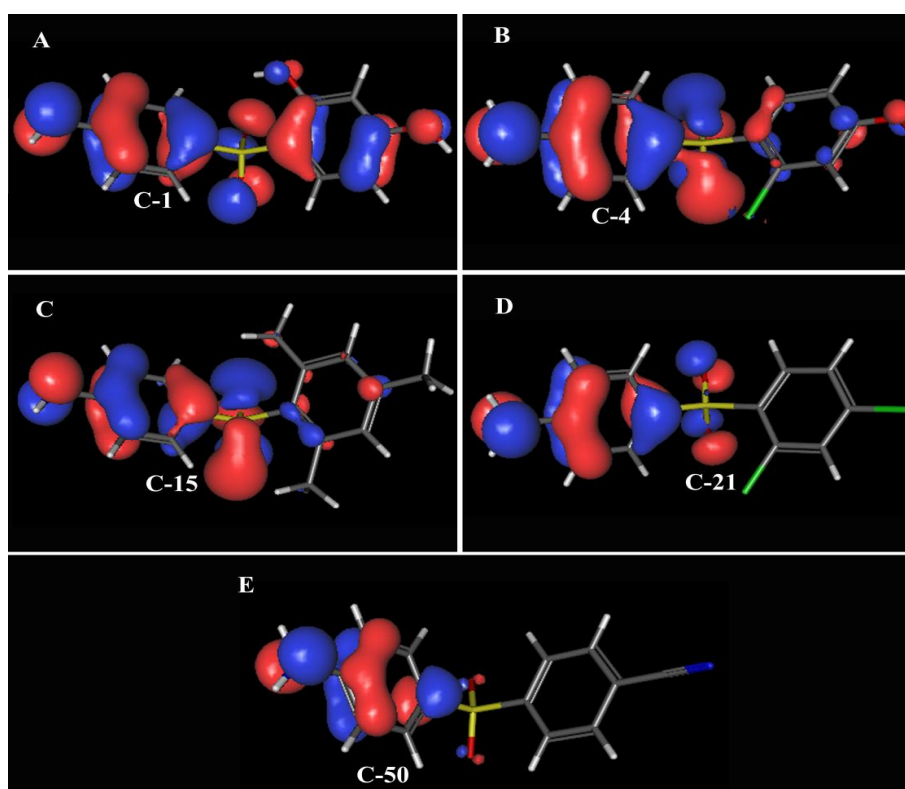


Figure 3.4. The distribution of HOMO orbitals on selected sulfone drug molecules. (A) Compound 1, (B) compound 4, (C) compound 15, (D) compound 21 and (E) compound 50 respectively.

The accumulation of HOMO orbitals on R9 ring indicates the influence of electron donating group in increasing the electron density over this aromatic ring. In comparison

to highly active and moderately active compounds, HOMO orbitals were absent in R9 position for less active compound 21, compound 50 (**Figure 3.4 D** and **3.5 E**). HOMO orbitals were mainly located at R8 and NH₂ portion of compound 21 and 50. This is due to the presence of electron withdrawing group at R9 ring which decreases the electron density over this ring. **Figure 3.5** illustrates the distribution of LUMO orbitals over the interacting amino acid residues located at *p*ABA binding pocket of EcDHPS. The pink and green colours represent the positive and negative lobes of LUMO orbitals.

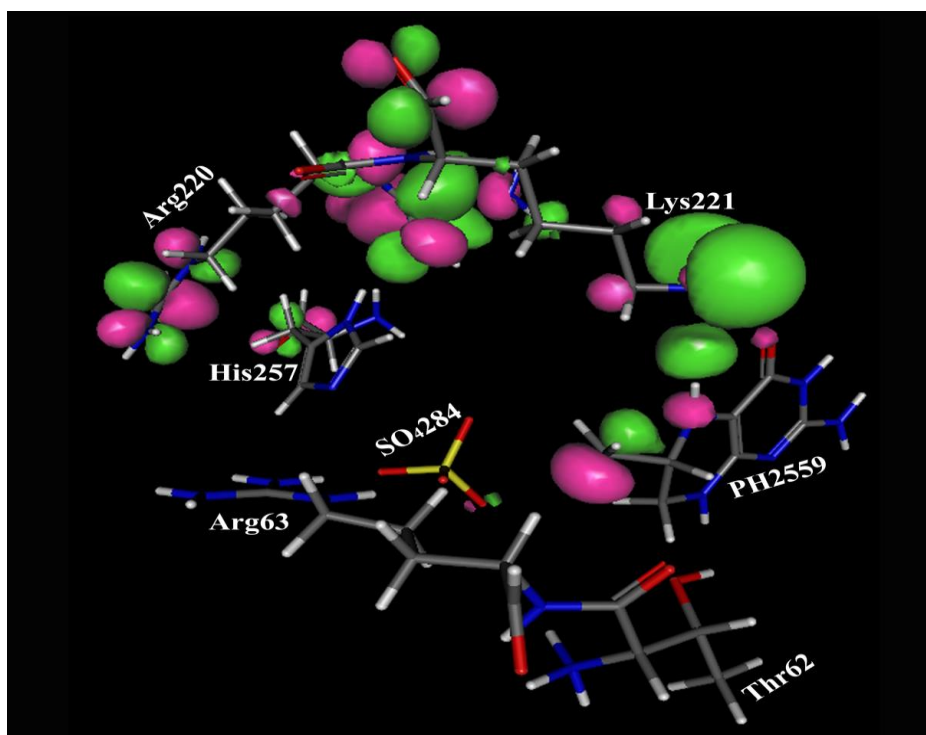


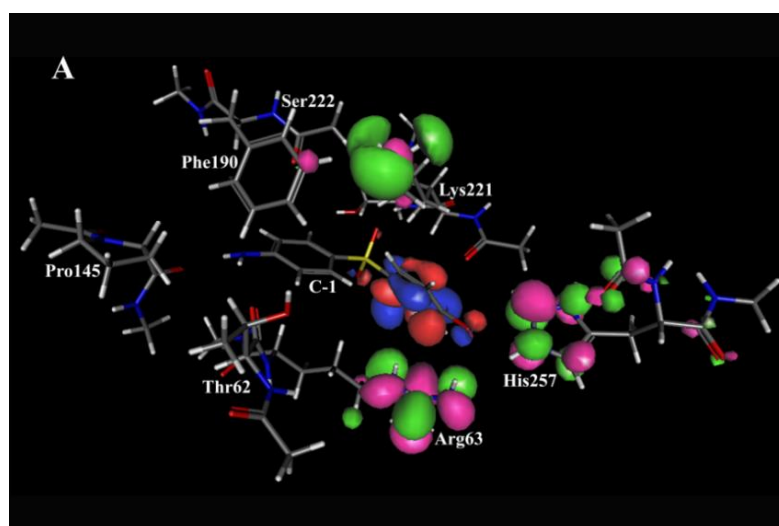
Figure 3.5. The location of LUMO orbitals at the *p*ABA binding site without diaryl sulfone compounds. PH2559 and SO₄284 represent the Pterin and Sulphate ion respectively. The location of LUMO orbital at the carbocation of PH2559 indicates the site prone to nucleophilic attack.

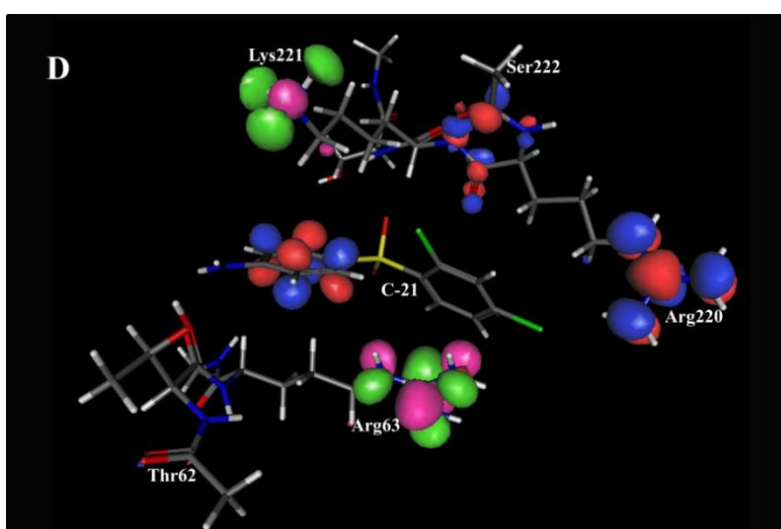
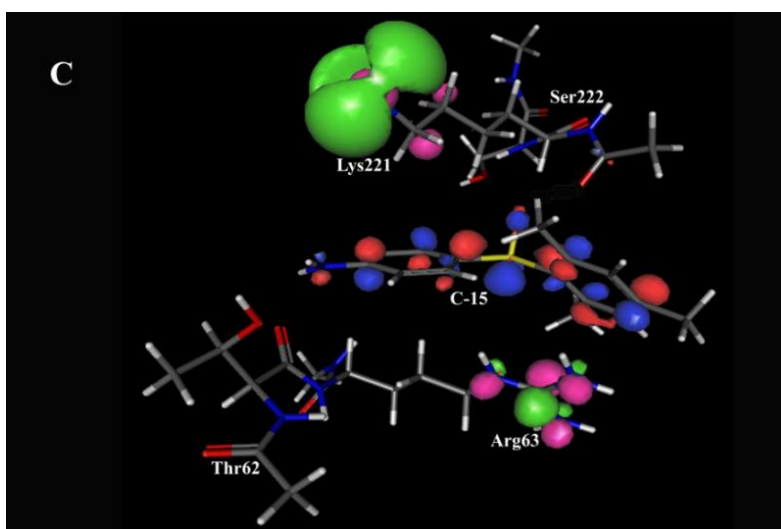
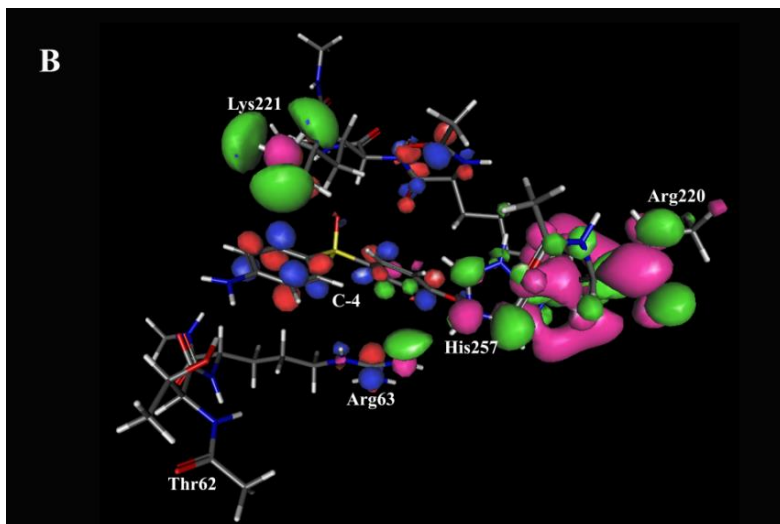
According to the frontier molecular orbital theory, we know that HOMO of the ligands interacts with LUMO located at the catalytic pocket of protein. Therefore, it would be interesting to see the position of the LUMO orbitals present in the catalytic pocket. We have shown the orientation of the interacting amino acids over the sulfone compounds in **Appendix III**. **Figure 3.5** shows that LUMO orbitals are mainly located on the carbocation of pterin moiety, Arg220, Lys 221 and His257. The presence of LUMO

orbitals at the carbocation of pterin moiety (**Figure 3.5**) and HOMO at NH₂ portion of SO (**Figure 3.5**) compounds indicates the formation of a pterin-sulfa complex by the nucleophilic attack. The HOMO positions at the R9 ring and the LUMO located at Arg220 and His257 indicates that the substitutions at R9 ring interacts with those amino acid residues and stabilizes the ligands at the pABA pocket. The closer proximity of HOMO of the ligand and the LUMO of the interacting amino acid in compound 1, compound 4 and compound 10 strongly suggests the reason of higher binding affinity of the protein-ligand complex due to the presence of electron donating, anionic groups in R9 ring. Absence of HOMO orbitals at the R9 ring of less active compound 21 and compound 50 gives less stability to these ligands at the pABA binding pocket. This can be correlated with our docking studies. Since, it is computationally expensive to do the DFT calculation for the whole protein molecule, so we chose the interacting amino acids and the ligand to see the HOMO, LUMO orbitals. We calculated the total energy of each atom clusters consist of each aforementioned sulfone compounds and interacting amino acid residues obtained in docking studies using self-consistent field approach. The HOMO, LUMO energy values are listed in **Table 3.5** and the HOMO, LUMO energy gap are also calculated. The lower energy gap for each atom cluster indicates the easy movement of electrons between HOMO, LUMO orbitals. **Figure 3.7** shows the distribution of HOMO, LUMO orbitals at each complex. The total energy of the cluster containing the atoms of interacting amino acid residues with highly potent compound and the least potent compound was found to be -33133.41 and -31550.25 eV respectively which indicates the greater stability of compound 1 than compound 50 at pABA binding pocket. **Figure 3.6 A** illustrates the location of HOMO orbitals in compound 1 and the LUMO region on pABA binding site. The presence of LUMO orbitals at Arg63, Lys221 and His257 indicated their interaction with the HOMO orbitals located at R9 ring and phenoxide groups of compound 1. These interactions found to stabilize the compound 1 between the two flexible loops of pABA binding pocket. Particularly, the presence of LUMO orbitals on Arg63 indicated its role in stabilizing sulfone compound after its binding at the catalytic pocket because LUMO orbitals did not appear in the amino acid cluster of the catalytic pocket of the protein alone (**Figure 3.6**). **Figure 3.6 B** also shows that location of HOMO orbitals on R8 and R9 ring of compound 4. The location of LUMO orbitals in this complex was found to

be similar as that of the amino acid complex containing compound 1. HOMO located on both R8 and R9 ring indicates the π -interactions of compound 4 with LUMO located at Lys221, His 257 and Arg 63 of *p*ABA pocket. **Figure 3.6 C** illustrates the position of HOMO, LUMO orbitals on the complex consist of compound 15 and interacting amino acid residues. The location of LUMO orbitals over Arg63 and Lys 221 indicates π -interaction of these amino acids with the HOMO positioned in R8 and R9 ring of compound 15. This interaction stabilizes compound 15 at *p*ABA binding pocket and play an important role in improving the potency of this compound.

The positions of HOMO, LUMO orbitals on the complex containing compound 21 and compound 50 with interacting amino acid are depicted in **figure 3.6 D** and **3.6 E** respectively. In both the complexes the HOMO orbitals were absent at R9 ring in comparison to highly active and moderately active compounds. In these cases, HOMO orbitals are located in R8 and NH₂ position sulfone compounds which interact with LUMO lies on Lys221 and Phe190. The absence of HOMO orbitals at R9 ring can be a strong reason for destabilizing the compounds between two flexible loops of *p*ABA binding pocket.





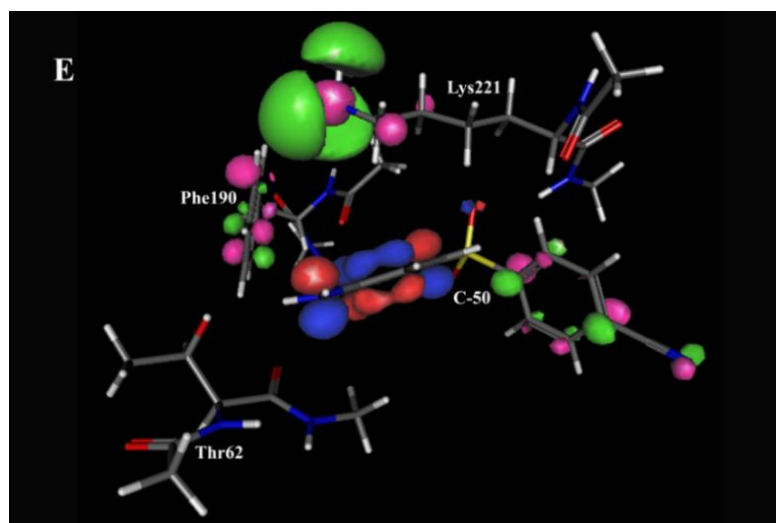


Figure 3.6. The distribution of HOMO, LUMO orbitals on the atom cluster containing (A) compound 1, (B) compound 4, (C) compound 15, (D) compound 21 and (E) compound 50 and their respective interacting amino acid residues.

3.3.4 FEP calculations

The relative binding free energy between the lowest potent (Compound 50, $pEI_{50}=4.17$, compound 41, $pEI_{50}=4.89$) and highly potent compounds in the data list (compound 1, $pEI_{50}=8.302$ and compound 4, $pEI_{50}=6.59$) were calculated using the FEP/REST method (Wang et al., 2015, 2012). The experimental relative binding free energy of Ligand A and B was calculated using the following equation (Uciechowska et al., 2012)

$$\Delta G_A = -RT \ln(IC^A_{50}) \quad (3.2)$$

$$\Delta G_B = -RT \ln(IC^B_{50}) \quad (3.3)$$

$$\Delta \Delta G_{exp} = \Delta G_A - \Delta G_B \quad (3.4)$$

Where IC_{50}^A and IC_{50}^B are the biological activities (Inhibitory Concentration fifty) of ligand A and B respectively. We took compound 50 as our ligand A and compound 1 as ligand B. Compound 50 found to be involved in steric clashes with Ile20 and His257 (First ligand, **Figure 3.7A-i**), but when the cyano group was mutated into phenoxide ($-O^-$) group at 2' and 4' position of R9 ring, gain in favourable interaction was achieved (**Second-ligand, Figure 3.7A-ii**) by forming salt bridges with Arg 63 and His 257

residues at the pABA binding site of DHPS. The relative free energy gain was calculated based on equation (2).

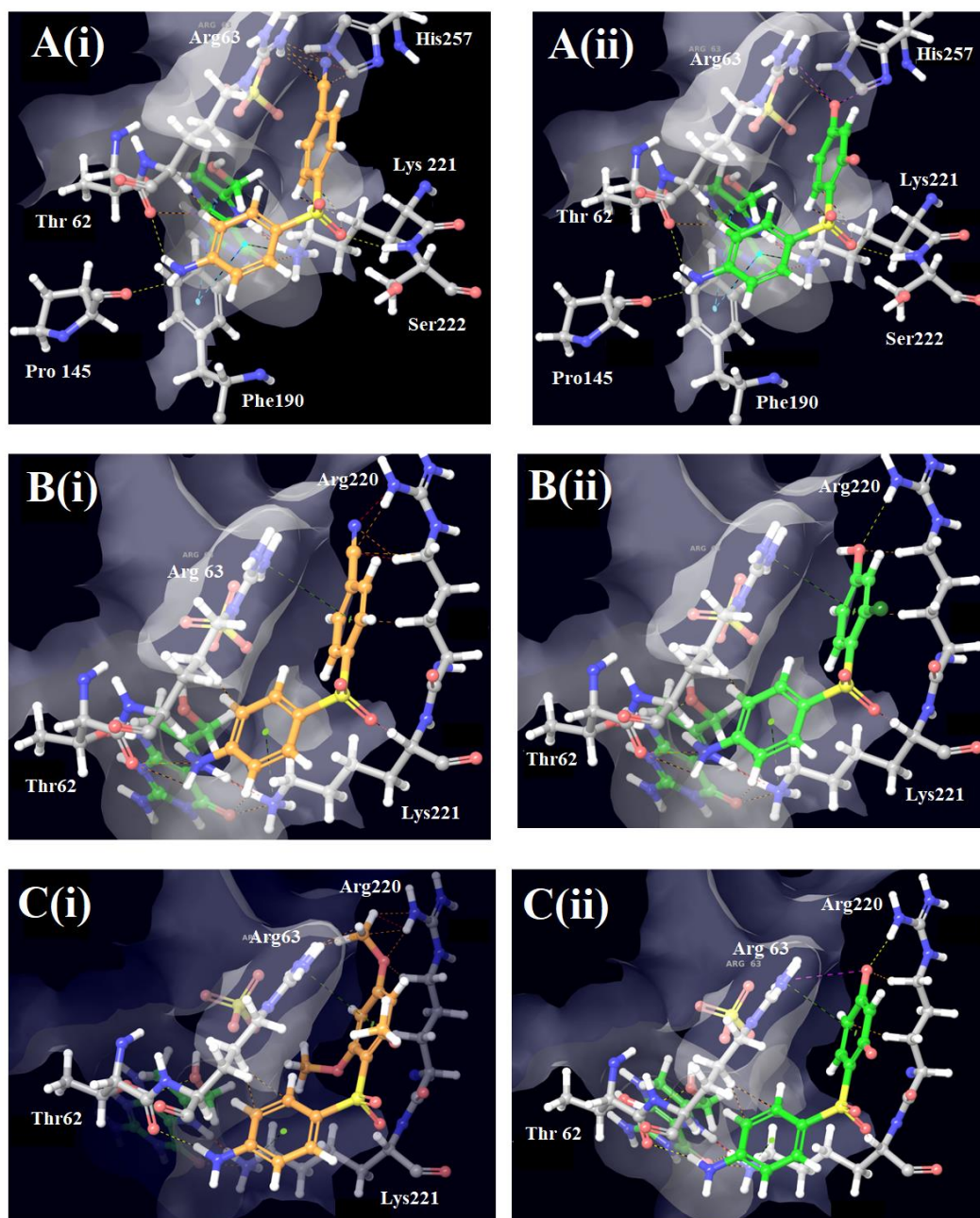


Figure 3.7. Representative figures of different types of interactions captured by FEP. (A I & ii) Shows gain in favourable salt-bridge interaction due to the presence of phenoxide group (mutating compound 50 to compound 1). (B i&ii) Shows gain in favourable salt-bridge interaction due to the presence of hydroxyl group (mutating

compound 50 to compound 4). (C i&ii) Shows gain in favourable salt-bridge and hydrogen bonding interaction due to the presence of phenoxide ion (mutating compound 41 to compound 1). The hydrogen bond, Salt-bridge, π -cation, π -stacking and unfavourable interaction was illustrated by yellow, pink, green, sky and brown dotted line respectively.

Mutating the cyano group ($-\text{C}\equiv\text{N}$) to phenoxide ($-\text{O}^-$) group, the experimental binding free energy ($\Delta\Delta\text{G}_{\text{EXP}}$) was found to be increased by -5.85 Kcal/mol due to gain in favourable salt bridge interactions. From FEP simulation, the computed relative free energy ($\Delta\Delta\text{G}_{\text{CAL}}$) difference between compound 50 and compound 1 was found to be -3.78 kcal/ mol. Similarly, the increase in experimental binding free energy ($\Delta\Delta\text{G}_{\text{EXP}}$) of -3.42 Kcal/mol was observed when cyano group (Compound 50, **Figure 3.7B-i**) at 4' position of R9 ring, was mutated into a hydroxyl group (Compound 4, **Figure 3.7 B-ii**) at the same position of sulfone compounds. The cyano group found to be involved in steric clashes with Ile20 and His257. When the cyano group was mutated to hydroxyl ($-\text{OH}$) group, the resulted free energy gains due to the formation of a hydrogen bond with Arg220 was found to be -2.58 kcal/ mol. To see the unfavourable steric effect, we calculated the relative free energy difference when compound 41 is mutated to compound 1, illustrated in **Figure 3.7 C-i&ii**. The methoxy group at 4' and 6' position of compound 41 was found to involve steric clashes with the side chain of Arg63, Arg220, and His257. When the methoxy groups at 2' and 4' position is mutated into phenoxide ion ($-\text{O}^-$) in compound 1, the gain in favourable hydrogen bonding and salt bridge interaction with Arg63 and arg220 results in an increase in experimental binding free energy by -4.83 kcal/mol. This effect is captured by our FEP calculation by computing binding free energy difference ($\Delta\Delta\text{G}_{\text{CAL}}$) between these two compounds of -1.94 kcal/ mol.

The timeline representation of free energy difference was found to reach a plateau within the simulation time (**Appendix V**) which indicates the convergence of free energy calculations. The relative free energy profile during the simulation is shown in **Appendix V** and the exchange density of FEP replicas over window is presented in **Appendix VI**.

3.3.5 ADME/Toxicity prediction of top scoring compounds

It is believed that only improving potency of a particular drug towards its target is not the primary objective in the drug development process, unless taking PK profile and toxicity in to consideration (Ekins et al. 2002). The various important pharmacokinetic properties of compound 1 to compound 10 and the permissible range are documented in **Table 3.6**.

Table 3.6 Ligand based ADME and pharmacokinetic properties of top scoring sulfone compounds, substrate (pABA) and

Compound No.	Molecular Weight	Donor HB	Acceptor HB	QPlog P _{o/w}	Dipole	QPlogS
1	265.283	2.5	5.500	0.754	8.606	-2.407
2	277.337	2.5	5.750	1.099	8.766	-2.718
3	263.311	2.5	5.750	0.938	8.150	-2.709
4	283.729	2.5	5.750	0.966	8.380	-2.713
5	283.729	2.5	5.750	0.938	8.380	-2.709
6	263.311	2.5	5.750	1.843	8.150	-3.131
7	277.337	1.5	4.750	0.754	9.078	-2.407
8	265.283	2.5	5.500	1.099	8.606	-2.718
9	277.337	2.5	5.750	0.754	8.776	-2.407
10	265.283	2.5	5.500	0.523	8.606	-2.157
pABA	137.138	3.5	5.500	-0.492	5.204	-1.056
Sulfanilamide	172.201	2.5	3.000	0.793	8.034	-0.961
Mafinide	186.228	4.0	5.500	-1.120	7.473	-0.024
Sulfadiazine	250.275	2.5	7.500	0.036	8.325	-2.858
Sulfacetamide	214.239	2.5	6.000	-0.153	9.417	-1.857
Sulfisoxazole	267.302	2.5	7.000	1.004	13.034	-2.923
Dapson	248.299	3.0	6.000	0.968	10.335	-2.754
Sulfamethoxazole	253.275	2.5	7.000	0.468	12.761	-2.413
Sulfadoxine	310.327	2.5	8.500	0.894	10.539	-2.891
Compound No.	rotor	PSA	QPlogKhasa	QPlogBB	Present Human Oral Absorption	
1	5	102.06	-0.454	-1.562	66.72	
2	4	77.015	-0.308	-1.111	76.306	
3	4	81.911	-0.384	-1.193	73.291	
4	4	82.410	-0.414	-1.114	73.421	
5	4	82.410	-0.414	-1.114	73.421	
6	4	81.911	-0.384	-1.193	73.291	
7	4	76.651	-0.111	-0.984	84.069	
8	5	102.06	-0.454	-1.562	66.729	
9	4	77.015	-0.308	-1.881	76.306	
10	5	102.06	-0.454	-1.881	66.729	

pABA	2	75.811	-0.832	-0.844	62.363
Sulfanilamide	3	91.720	-0.779	-1.219	59.595
Mafinide	4	92.565	-0.764	-0.937	46.890
Sulfadiazine	4	101.088	-0.687	-1.282	68.728
Sulfacetamide	3	106.307	-0.634	-1.323	63.791
Sulfisoxazole	4	102.237	-0.506	-1.362	73.921
Dapson	4	88.261	-0.491	-1.378	71.165
Sulfamethoxazole	4	103.363	-0.605	-1.383	70.057
Sulfadoxine	6	109.567	-0.568	-1.301	77.540

The top scored compounds were thoroughly evaluated with the basic parameters of “Lipinski’s rule of 5” and other pharmacokinetic parameters. Generally, violation of Lipinski’s rule more the two is considered to be forbidden for orally active compounds. All the pharmacokinetic properties of the top scored compounds documented in **Table 3.6** were found to be under permissible limit.

The values of polar surface area and rotatable bond count of top scored compounds known to have great impact on their oral bioavailability. The number of rotatable bonds and the polar surface area in top ranked sulfone compounds are found to be in the range of 4-5 and 77-102 Å² respectively. These values found to be under recommended ranges thus expected to have good bioavailability. It can be observed from **Table 3.6** that the molecular weight of the top scored ligands are in the range of 263.31 to 283.72 which is acceptable for drug-like compounds (Lipinski et al. 1997).

Moreover, the number of hydrogen bond donor groups and acceptor group found to be below the threshold limit (H-bond-donor should be less than or equal to 5, similarly the threshold value for number of H-bond acceptor is 10). This indicates good adsorption of diaryl sulfone in system circulation. Additionally, the present human oral absorption value found to be within 66.72-84.062%, indicating moderate to high adsorption of the top scoring compound. It is believed that a chemical compound will be drug like if their water/octanol partition co-efficient is less than 5. It is obvious from table 4 that the water/octanol partition co-efficient much below the threshold level. The values of solvent accessible surface area (SASA) and polar solvent area of all top scoring sulfone derivatives are in found to be in accepted range (Duffy and Jorgensen 2000). It is found from **Table 3.6** that all the top scored inhibitors did not violet the “Lipinski’s rule of 5” (Lipinski et al. 1997) and “Jorgensen’s rule of 3”. Further, the human serum albumin binding affinity values and blood brain barrier coefficient values are found to be under

permissible range. The predicted ADME properties of ten highly active drug compounds indicate acceptable pharmacokinetic and less toxicity profiles for Phase 1 clinical trials. Further, the ADME toxicity of properties the sulfone compounds are compared with the established drugs against DHPS available in market. It is evident from **Table 3.6** that the top scored sulfone compounds have similar toxicity profile with the available drugs in the market. Notably, the highly active sulfone compounds have higher solubility and percent oral absorption compared to Mafinide, sulfadiazine, and sulfacetamide.

3.3.6 MD Simulation

The dynamic behaviour of ligand at the binding pocket of enzyme is important to evaluate the stability of that particular ligand inside the active site. A 10 ns molecular dynamics simulation of 1AJ0/compound 4 provides further insights into molecular interaction of compound 4 in motion. The Root Mean Square Deviation (RMSD) of the enzyme backbone with respect to its initial position increased up to 1.64 Å for first 4 nanoseconds and stabilized around 1.5 Å for rest of the trajectory. Moreover, the ligand movement cope well with the RMSD of binding site (**Figure 3.8A**). The average RMSD of enzyme (**Figure 3.8A**) backbone and heavy atoms are found to be 1.44 Å and 1.86 Å respectively. This reflects minute structural change of 1AJ0 during simulation from the crystal structure of its own. The residue wise Root Mean Square Fluctuation (RMSF) of 1AJ0 was illustrated in **figure 3.8 B**. The detailed inspection of RMSF helped in characterizing the regional changes in the protein chain throughout the course of simulation. The maximum value of C α , backbone, heavy atom (3.01, 3.07 and 3.33 Å respectively) RMSF was found for residue Glu277 which resides in a flexible region of Ser274-Glu 282(C-terminal end). Additionally, Gly31 found to have high C α RMSF of 2.50 Å at the N-terminal flexible region from Val23 to Ser36. Both highly fluctuating loops are found to reside away from binding pocket. Interestingly, the RMSF value of protein backbone residue at the catalytic site was found to be in the range of 0.51 Å-1.17 Å indicating the stability of catalytic region under motion.

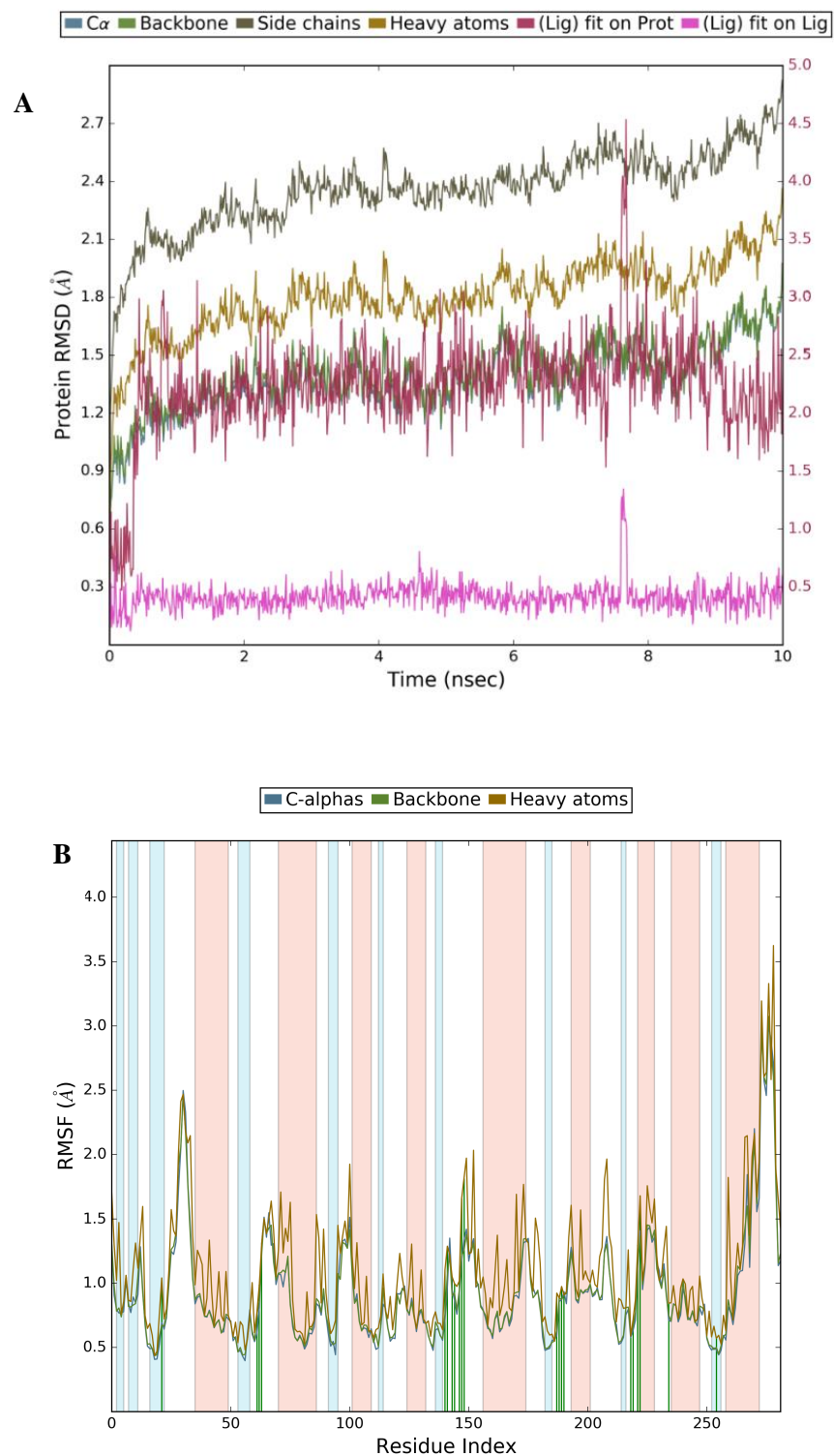


Figure 3. 8. (A)The RMSD (Å) of the simulated positions of 1AJ0 backbone atoms from their initial coordinate in 10 ns MD trajectory. (B) The residue wise RMSF profile

of 1AJ0 illustrates the regional changes along the protein chain throughout the trajectories.

The protein-ligand contacts during simulation are illustrated in **Figure 3.9** and **Appendix VII**. Compound 4 was found to exhibit ionic interaction in the stable region (**Figure 3.9**) of Thr62-Pro64 and Arg255-His257. Hydrogen bonding found to be formed majorly with Pro145, Ser222 and hydrophobic interaction was dominated by Phe190 throughout the simulation time. The region between Met223-Arg235 found to exhibit water mediated hydrogen bond with **compound 4**. The region from Gly65-Gly143 and Gly191-Ser219 no interaction was found due to higher fluctuation (**Figure 3.8 B** and **Appendix VII**).

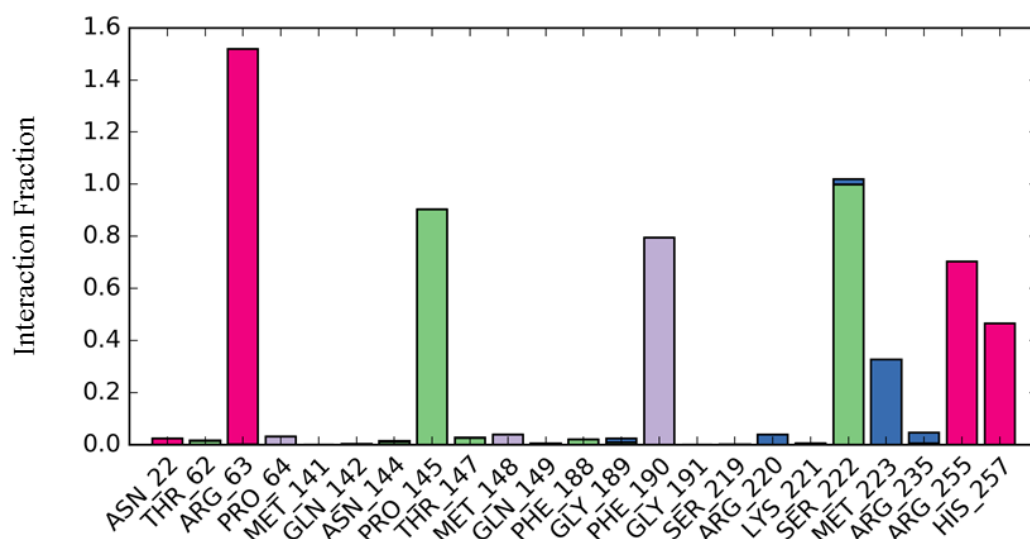


Figure 3.9. Histogram represents interaction of **compound 4** with different amino acids of 1AJ0 during 10 ns MD trajectory. The pink colour represents the ionic interaction, green colour represents hydrogen bond, violet colour represents hydrophobic interaction and blue colour stands for water mediated hydrogen bonding.

During the course of simulation interaction observed with Arg63, Pro145, Phe190, Ser222, Met223, Arg255 and His 257 located to the active site are expected to be responsible for the inhibitory activity of compound 4. A low mean RMSD of 0.375 Å of compound 4 indicates less conformational modification during simulation. The low

solvent accessible surface area (SASA) (7.04-73.90 Å²), high polar solvent area (PSA) (184-187 Å²) and molecular surface area (MolSA) (233-238 Å²) of compound 4 further supports its stabilization during 10 ns molecular dynamics simulation (**Appendix VIII**).

3.4 CONCLUSION

In this present chapter, 3D-QSAR modelling, molecular docking, DFT calculations, FEP calculations and MD simulation were performed to explore the effect of different substitution on sulfone derivatives to stabilize them in *p*ABA binding loop of EcDHPS catalytic pocket. A highly predictive 3D-QSAR model was developed which helped in understanding how the structure of sulfone compounds affects the biological activity. Contour plot mapping was performed to observe the spatial arrangement of favourable and unfavourable functional groups and their contribution to inhibit the EcDHPS enzyme. Presence of electron donating group and hydrogen bond donor at 2' and 4' position of the R9 ring were found to increase the potency of sulfone compounds. Molecular docking was done to find the possible binding pose of sulfone compounds at the *p*ABA binding pocket. The docking study suggested that Thr62, Arg63, Pro145, Phe190, Lys 221, Ser222 and His257 amino acids are the key binding residues in the active site of EcDHPS. The highest potent compound 1 contains two dissociated phenoxide ion which forms the salt bridge with Arg63 and His257. Possibly the mutation in Arg63, the His257 region can disrupt the sulfone (SO) binding at DHPS catalytic pocket due to their active role in stabilizing the R9 ring. DFT studies showed the stabilization pattern and reaction mechanism of sulfone compound at electronic level. The HOMO, LUMO interaction pattern between sulfone compounds and *p*ABA binding site are also supported by docking studies. The presence of electron withdrawing group at 4' position directly affects the electron density over the R9 ring and an SO₂ moiety of the ligand which resulted compound 50 to be least potent. Presence of hydrophobic groups found to have both favourable and unfavourable contribution depending on the size of the group and site of substitution. Further, our conclusion was validated by FEP calculations by capturing relative binding free energies between less potent and highly potent compounds. The predicted ADME of top scored drug candidates are found to be in acceptable ranges. Lastly a 10 ns MD

simulation of highly active compound 4 indicates the fixation of this inhibitor in pABA catalytic pocket. This combinatorial computational study contributes a set of useful structural guidelines focusing on pABA binding pocket, which will greatly help in designing sulfone-based and novel pterin-sulfa compounds for the EcDHPS-selective inhibitors to prevent infection caused by pathogenic and antibiotic-resistant *E. coli* strains.

CHAPTER 4

DECIPHERING THE COMPETITIVE INHIBITION OF DIHYDROPTEROATE SYNTHASE BY 8-MARCAPTOGUANINE ANALOGS: ENHANCED POTENCY IN PHENYLSULFONYL FRAGMENTS

Abstract: *The emergence of sulfa-drug resistance and reduced efficacy of pterin-based analogs towards Dihydropteroate synthase (DHPS) inhibition dictate a pressing need of developing novel antimicrobial agents for immune-compromised patients. Recently, a series of 8-Mercaptoguanin (8-MG) derivatives synthesized for 6-Hydroxymethyl-7,8-dihydropterin pyrophosphokinase (experimental $K_D \sim 100 \cdot 0.36$) showed remarkable homology with the pteric-acid and serve as a template for product antagonism in DHPS. The present work integrates ligand-based drug discovery techniques with structure-based docking, enhanced MD simulation, and MM/PBSA techniques to demonstrate the essential features of 8-MG analogs which make it a potent inhibitor for DHPS. The delicate balance in hydrophilic, hydrophobic substitutions on the 8-MG core is the crucial signature for DHPS inhibition. It is found that the dynamic interactions of active compounds are mainly dominated by consistent hydrogen bonding network with Asp 96, Asn 115, Asp 185, Ser 222, Arg 255 and π - π stacking, π -cation interactions with Phe 190, Lys 221. Further, two new 8-MG compounds containing N-phenylacetamide (compound S1, $\Delta G_{bind-eff} = -62.03$ kJ/mol) and phenylsulfonyl (compound S3, $\Delta G_{bind-eff} = -71.29$ kJ/mol) fragments were found to be the most potent inhibitor of DHPS, which stabilize the flexible pABA binding loop, thereby increasing their binding affinity. MM/PBSA calculation shows electrostatic energy contribution to be the principal component in stabilizing the inhibitors in the binding pocket. This fact is further confirmed by the higher energy barrier obtained in umbrella sampling for this class of inhibitors.*

Keywords: *Dihydropteroate synthase; 8-mercaptoguanine; Phenylsulfonyl fragment; Molecular dynamics simulation; Umbrella Sampling*

4.1 BACKGROUND

The point mutations at the flexible *p*A_{BA} binding loops of DHPS (Yun et al. 2012b) in various pathogenic microorganisms, hinders the utilization of sulfonamide mediated antimicrobial therapy (Griffith et al. 2018; Ho and Juurlink 2011; Sköld 2000). Additionally, the use of sulfa-drugs against bacterial infections have been limited due to rigorous immunological reactions (i.e. sulfa allergy) and toxicity that may cause breathing problem, loss of appetite, vomiting, nausea, fever, etc (Mondal et al. 2015). As a consequence, the attention shifted towards targeting the structurally rigid pterin binding pocket to bypass sulfa drug resistance and its side effects. Although, there are many reports regarding the synthesis, structural and computational studies of pterin analogues on successful inhibition of DHPS (Azzam et al. 2020; Babaoglu et al. 2004; Hevener et al. 2010b), the major bottleneck of these inhibitors are lower solubility and higher selectivity towards the pterin binding pocket (Zhao et al. 2012).

Therefore, a big thrust of the drug-discovery community has been redesigning DHPS inhibitors, which are capable of circumventing sulfa-drug resistance as well as have higher solubility. Accordingly, structure-based drug-discovery schemes (Dennis et al. 2014, 2016; Hammoudeh et al. 2013; Zhao et al. 2016b) have been applied in designing DHPS inhibitors, which not only bind the structurally rigid pterin binding pocket but also occupy the triosephosphate and flexible *p*A_{BA} binding pocket. Recent *in silico* efforts by Chakraborty and co-workers (Das et al. 2019) have also presented a free energy basis of FDA-approved sulfa drugs followed by the interpretation of crucial mutations towards sulfa drug resistance. The results obtained from this study also provided the idea to use product antagonists for retaining the inhibitory activity in face of some adverse point mutations at flexible loop regions. Additionally, a series of 8-Mercaptoguanine derivatives against 6-Hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK), reported by Dennis and Co-workers offer lower-order K_D (~100-0.36 μ M) value towards the enzymes of folate biosynthetic pathway (Dennis et al. 2016, 2018) and shown excellent homology with the natural product of DHPS. However, the key interaction of the compounds with the catalytic pocket and associated conformational dynamics of the receptor remains elusive and poses a fundamental challenge in the development of DHPS inhibitors.

In this context, the present chapter addresses a pertinent query about the key factors that stabilize the 8-MG analogs at the simultaneous Pterin, pyrophosphate, pABA binding pockets of DHPS and increases the binding affinity. Here, we applied a robust *in-silico* approach to discover the potential hits for DHPS inhibition which have the ability to circumvent the sulfa drug resistance (**Figure 4.1**) as well as show higher affinity towards DHPS catalytic pocket. Specifically, we combined ligand-based drug discovery techniques such as Pharmacophore based virtual screening (Pal et al. 2019; Venugopal and Chakraborty 2021), 3D-QSAR(Ajmani et al. 2006; Peng et al. 2019b), Density functional theory (DFT)(Andrade-Ochoa et al. 2018) to determine crucial functional-group substitution responsible for the potency of 8-MG compounds as well as screen large drug library. Next, structure-based molecular docking(Shahzad et al. 2020), all-atom MD simulation(Das and Chakraborty 2020; Koneru et al. 2019), MM/PBSA(Genheden and Ryde 2015; Kumari et al. 2014a) studies were applied on top screened lead compounds to understand the conformational dynamics and free energetics of potent 8-MG analogs at the DHPS catalytic pocket. Furthermore, the enhanced sampling-based free energy simulation technique, namely umbrella sampling simulations(Kästner 2011a; Sun et al. 2015; Zhou et al. 2015b) were employed to cross-validate the affinity or resilience of DHPS inhibitors against sulfa-resistant mutations and dissociation from the catalytic cavity. Results obtained from this study reveal insights into the potency and future development of 8-MG derivatives as an antibacterial agent.

4.2 COMPUTATIONAL DETAILS

4.2.1 Data-set Preparation

In order to develop a common pharmacophore and 3D-QSAR model, the primary step was ligand preparation. A set of sixty-two 8-mercaptoguanine compounds (**Table 4.1**) having an inhibitory effect against dihydropteroate synthase (DHPS) were retrieved from literature(Dennis et al. 2018)and used in the present study. All compounds used for the modeling study were reported to share the same assay procedure(Seabrook and Newman 2013). The biological activity data (K_D) of those compounds were found to span over four orders of magnitude and ranges from 420 to 0.39 μ M. The biological activities of all the compounds were converted into pK_D values ($pK_D=6-\log_{10}(K_D)$) for

ease in modeling studies. The three-dimensional structures of the compounds were constructed in Maestro (Schrödinger Release 2017-2: Maestro, Schrödinger, LLC, New York, NY, 2017) builder panel.

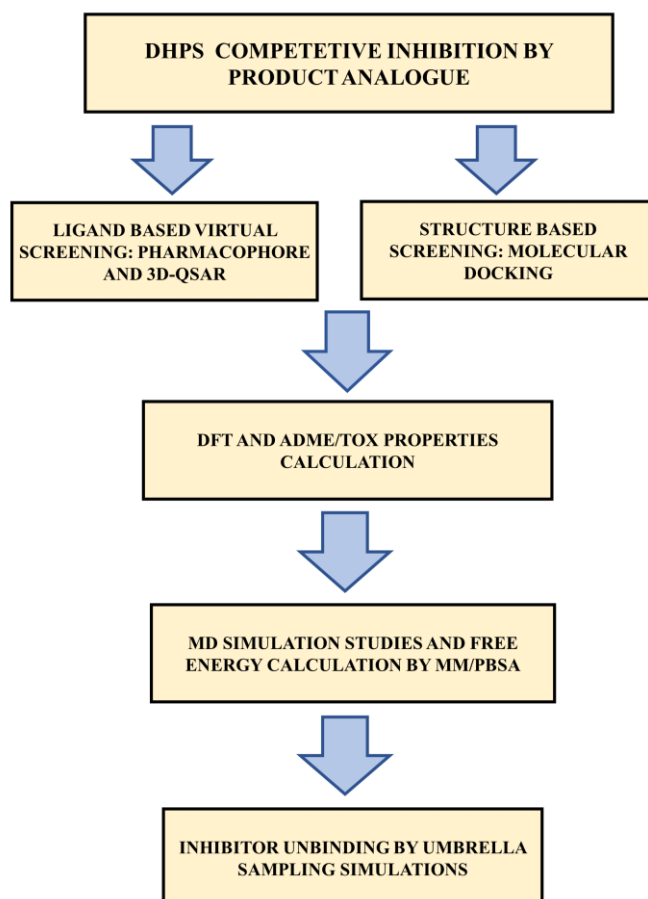


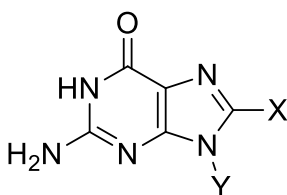
Figure 4.1. The schematic diagram of the methods implemented in the current study.

Next, all the 3D structures of the inhibitors were imported to the Ligprep module (Schrödinger Release 2017-2: LigPrep, Schrödinger, LLC, New York, NY, 2017) in order to optimize the geometry and generate possible ionization state present at the pH of 6.9. OPLS_2005 force field (Kaminski et al. 2001; William L. Jorgensen et al. 1996) was employed to minimize the energy of each compound using a root mean square deviation (RMSD) cut-off of 0.01 Å. The resulting structures were used for further modelling studies.

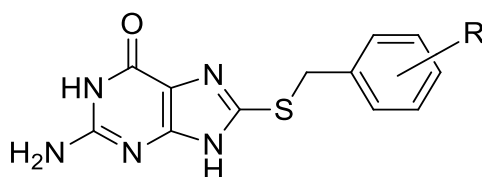
4.2.2 Pharmacophore Mapping

Pharmacophore is a part of molecular structure that is necessary for recognizing the ligand by a biological macromolecule. Phase(Dixon et al. 2006) was employed to find out the common pharmacophoric features of all chosen inhibitors of DHPS. Phase provides a set of five pharmacophoric features, hydrogen bond donor (D), hydrogen bond acceptor (A), hydrophobic group (H), negatively ionizable (N), positively ionizable (P), and an aromatic ring (R). Inhibitors having K_D value greater than 5.3 was considered as active and the threshold of K_D for inactive molecule was 4.2.

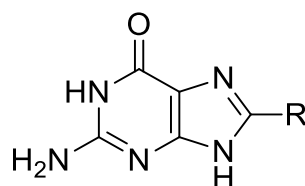
Table 4.1. Structural details, experimental activity, and predicted activity of 8-Mercapto-guanine compounds (1-62). The superscript “t” indicates the test set compounds in 3D-QSAR modelling.



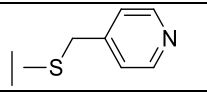
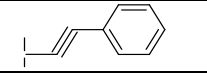
Compound	X	Y	$pK_{D(\text{exp})}$	$(pK_{D(\text{pred})})$	RA ($pK_{D(\text{res})}$)
1	SH	Ethyl	3.377	3.376	-0.001
2	SH	Benzyl	3.456	3.303	-0.152
3	SCH ₃	H	3.481	3.616	0.135
4 ^t	OH	H	3.690	4.039	0.349
5	N-morpholino	H	3.796	3.870	0.074
6	SH	CH ₂ Benzyl	3.796	3.790	-0.006
7	OH	Methyl	3.996	4.049	0.053
8 ^t	SH	H	4.118	3.798	-0.32
9	Methyl	H	4.149	4.119	-0.03
10	SH	Methyl	4.155	4.140	-0.014
11	H	Methyl	4.237	4.282	0.046
12	Br	H	4.377	4.228	-0.149

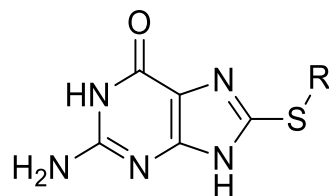


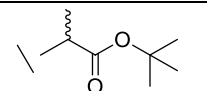
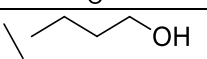
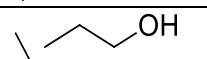
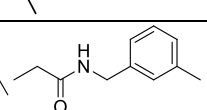
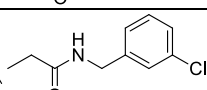
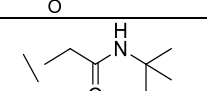
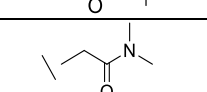
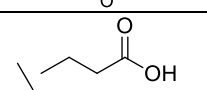
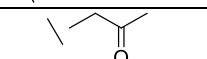
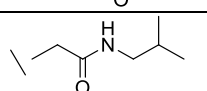
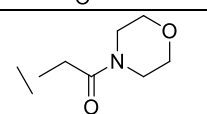
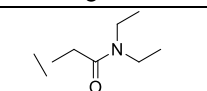
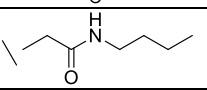
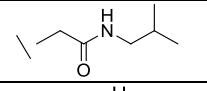
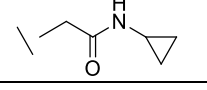
Compound	R	(pK _{D(exp)})	(pK _{D(pred)})	RA (pK _{D(res)})
13	2-nitro	4.481	4.492	0.011
14	2-fluoro, 3-methyl	4.509	4.624	0.116
15 ^t	3,5- dimethyl	4.854	4.526	-0.328
16	2-bromo	4.854	4.855	0.001
17 ^t	2,3- dimethyl	4.854	4.745	-0.108
18 ^t	2,2- difluoro	4.854	4.988	0.134
19	2-methyl, 4-fluoro	4.880	5.035	0.155
20	2-cyano	4.896	4.863	-0.032
21	2,4- difluoro	4.921	4.985	0.064
22	2,5- dimethyl	4.939	5.016	0.077
23	3,4- difluoro	4.959	5.008	0.049
24	3-cyano	4.959	4.99	0.036
25	4-fluoro	4.959	4.995	0.036
26 ^t	3-methyl	4.959	4.750	-0.208
27	2-methyl	5.022	4.936	-0.085
28	2-fluoro	5.071	4.935	-0.135
29	3-methyl, 4-fluoro	5.086	5.109	0.023
30	3-methoxy	5.125	5.166	0.041
31 ^t	2-chloro	5.155	4.897	-0.257

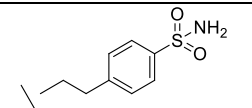
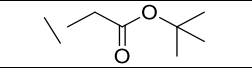
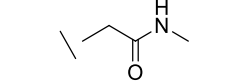
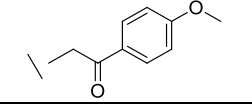
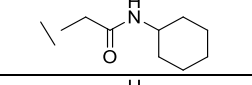
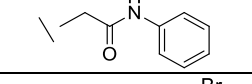
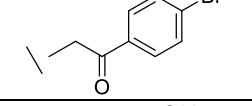
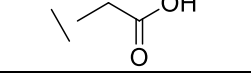


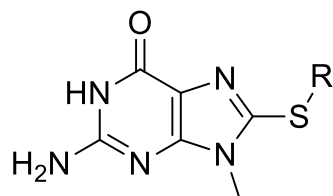
Compound	R	(pK _{D(exp)})	(pK _{D(pred)})	RA(pK _{D(res)})
32		4.796	4.795	-0.001
33		4.820	4.663	-0.161

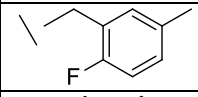
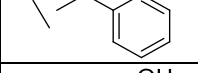
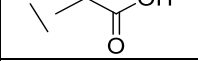
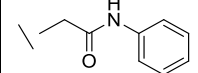
34 ^t		4.824	4.938	0.114
35		5.009	5.021	-0.003



Compound	R	(pK _D (exp))	(pK _D (pred))	RA(pK _D (res))
36		4.658	4.784	0.128
37 ^t		4.699	4.511	-0.147
38.		4.740	4.395	-0.344
39.		4.917	4.867	-0.050
40.		5.004	5.023	0.018
41.		5.018	4.935	-0.082
42.		5.086	5.109	0.023
43		5.102	5.136	0.033
44 ^t		5.108	4.855	-0.252
45.		5.125	5.169	0.044
46.		5.149	5.176	0.027
47.		5.161	5.169	0.008
48 ^t		5.220	5.110	-0.11
49 ^t		5.310	5.169	-0.140
50.		5.319	5.386	0.068

51.		5.337	5.231	-0.105
52.		5.344	5.274	-0.06
53 ^t		5.347	5.442	0.095
54 ^t		5.409	5.252	-0.15
55.		5.538	5.536	0.002
56 ^t		5.688	5.909	0.22
57.		5.690	5.638	-0.05
58.		6.08	6.16	0.08



Compound	R	Experimental Activity (pK _D (exp))	Predicted Activity (pK _D (pred))	Residual Activity (pK _D (res))
59		4.523	4.546	0.023
60.		4.721	4.934	0.213
61.		6.377	6.339	-0.037
62.		6.409	6.479	0.07

The rest of the compounds were considered moderately active. This yields ten active molecules and ten inactive molecules which were used in pharmacophore model generation and corresponding scoring of that hypothesis. During hypothesis settings “number of pharmacophoric features” was kept at five and a maximum of thirty

hypotheses was generated. All the hypotheses were scored under the Phase Hypo Scoring method using default parameters for vector, site, volume, and energy terms. The best-scored hypothesis was chosen to understand the molecular fragment responsible for crucial non-bonding interactions.

To assess the predictive power of the model the best-scored hypothesis was tested against a set of decoy compounds (Kirchmair et al. 2008). Decoys are organic molecules that are assumed to be inactive against a pharmacological target. Due to the unavailability of decoys for DHPS, Decoy Finder (Cereto-Massagué et al. 2012) was employed to generate target-specific decoy compounds for assessing the performance of generated hypothesis. The inputs for Decoy Finder include a set of active compounds for a biological macromolecule, known as “queries” and a set of molecules from which the decoys will be selected. Next, the highly scored hypothesis was tested against 252 decoy compounds obtained from Decoy-Finder. A database containing 81,549 druglike molecules was generated by Phase-Database builder utility based on the hits collected from ZINC, PubChem and Schrodinger database. Finally, the validated hypothesis was subjected to database screening in order to screen new 8MG-compound with varied substitution.

4.2.3 Atom-based 3D-QSAR Modelling

The quantitative structure-activity relationship (QSAR) methods have been applied widely in computer-aided drug-design approaches (Hoekman 1996; Puzyn et al. 2010). It is believed that the classification of the training set and test set is the key to build a predictive 3D-QSAR model (Andrada et al. 2017; Leonard and Roy 2006a). Many classification techniques, such as: (1) Random selection, (2) Selection based on biological activity, (3) Kennard-stone algorithm, and (4) k-means clustering is reported in the literature. However, the first two methods are reported to fail in building QSAR models with optimal predictability. To the best of our knowledge, the non-hierarchical k-means clustering with partial least square (PLS) technique is highly reliable to build a statistically significant 3D-QSAR model (Andrada et al. 2017; Martinez et al. 2016). The k-means clustering technique was implemented on the factor scores of the original variable matrix corresponds to each 8MG-compound considered in this study. The variable matrix consists of response variables (pK_i), predictor variable (molar

refractivity (MR), ADME properties, reactivity descriptors), and topological descriptor variables predicted by PaDEL software (Yap 2011). The details of the descriptors used in this study are given in Supplementary Section. Factor analysis on this variable matrix indicates that three factors can explain the maximum variance of the dataset. Therefore, k-means clustering on the three-factor scores of all sixty-two 8-MG derivatives was carried out to group molecules with maximum similarities. The ratio between the entire training set and test set compounds was 3:1 which is known to be standard for 3D-QSAR modeling (Golbraikh and Tropsha 2000b; Umamatheswari et al. 2010). Atom-based ligand alignment tool of phase was employed to align all selected inhibitors prior to 3D-QSAR calculation (Bemis and Murcko 1996). The highest number of PLS in the 3D-QSAR model is generally kept at $N/5$ which is 6 in the present case (N is the number of molecules in the training set). It is known that a higher number of PLS may cause statistical overfitting of the data (Polański et al. 2002; Tropsha et al. 2011). Therefore, we use six PLS factors as an incremental increase in the statistical significance to improve the predictability. The contour map obtained from the generated 3D-QSAR model was analyzed to interpret the effect of three-dimensional arrangements of structural features on DHPS inhibition. The quality of the 3D-QSAR model was evaluated by predicting the activity of external test set compounds which were not considered during model development (Andrada et al. 2017; Martin et al. 2012b).

4.2.4 Molecular Docking

4.2.4.1 Protein Preparation

The co-crystal structure of DHPS with 8-MG analog (PDB ID: 5U10) was retrieved from Protein Data Bank (<http://www.rcsb.org>) and further processed with protein preparation wizard (Sastry et al. 2013) of Schrödinger software. The polar hydrogens were added to the protein structure at the pH of 6.9 to mimic the experimental ionization state. The hydrogen bond orientations were subsequently optimized and minimized. The conserved water molecule in the catalytic pocket was preserved. Lastly, the protein heavy-atoms were minimized by the AMBER99SB force field with an RMSD cut-off value of 0.30 Å by Gromacs-2016.5. In the course of minimization, the protein-heavy atoms were constrained, and hydrogen torsion parameters were turned off, to permit free rotation of hydrogen atoms. The receptor was minimized in vacuum.

4.2.4.2 Receptor Grid Generation

The minimized 3D structure of 5U10 was imported to the Maestro GUI to prepare grid box for docking studies using Glide/ receptor-grid generation tool (Friesner et al. 2004; Halgren et al. 2004). The crystal structure of the ligand was selected to identify the centroid of the active site of 5U10. A 3D-grid box was defined within a 20 Å radius around the co-crystal ligand structure keeping van der walls scaling factor 1.0 Å and partial charge cut-off of 0.25. Prior to dock all the 8-MG derivatives, the co-crystal ligand structure of 5U10 was removed and re-docked to assess the accuracy of the docking calculation.

4.2.4.3 XP Docking

After achieving proper RMSD value between the co-crystal and re-docked ligand the grid box was selected for docking all sixty-two 8-MG derivatives at the catalytic pocket of 5U10. The van-der-walls scaling factor was kept at 0.80 during docking calculation with a partial charge cut-off value of 0.15. During the docking calculations, the receptor atoms were kept rigid while the rotatable bonds of the inhibitors were allowed to move flexibly inside the binding pocket of 5U10. The docking of all the inhibitors was performed in extra precision mode(Friesner et al. 2006). After docking of each inhibitor, the strength of protein-ligand interactions was determined using the empirical scoring function implemented in Glide.

4.2.5 DFT calculation setup

The electronic properties of drug candidates are known to play a crucial role in eliciting pharmacological responses in the human body(Matysiak 2007). The most active, least active and top three screened compounds (collected from database screening) identified by molecular docking were subjected to DFT calculation to study the quantum chemical properties of the molecules such as MO, density, molecular electrostatic potential (MEP), and dipole moments, etc. Primarily the geometry of chosen ligands was optimized in Gaussian09("Gaussian 09 Citation | Gaussian.com" n.d.) program using the combination of gradient correction of the exchange functional by Becke (B3LYP) (Becke 1993) and the correlation function of Lee, Yang, and Parr (Lee et al. 1988b) with both 6-31G**(d,p), 6-311G**(d,p) basis sets. All the quantum chemical

calculations were carried out in the aqueous phase by employing the polarizable conductor-continuum model (CPCM) (Takano and Houk 2005) and in vacuum. In order to explain the chemical reactivity of ligand, the frontier orbitals such as highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) were calculated. The location of HOMO indicates the electron-donating portion of the ligand, whereas LUMO conveys the tendency to accept the electron from the interacting amino acids of the protein. The electronic excitation energy was computed from the band-gap energy (HOMO-LUMO energy gap) to assess the reactivity, stability of the drug candidates (Zhan et al. 2003; Zheng et al. 2013). Lastly, the location of HOMO-LUMO over the catalytic pocket residues was identified to explain the binding mechanism at the electronic level. The *Cubegen* utility of Gaussion09 was employed to generate the cube files for visualizing the HOMO-LUMO orbitals. The MOs were visualized in Gauss-View (v-6.0) GUI for further analysis.

4.2.6 ADME/Tox Properties Calculation

In the drug discovery process, the ADME properties of the hit compounds are believed to play a crucial role to ensure safe drug administration. The QikProp (Jorgensen et al. 2006) module of Schrödinger is employed to compute drug-likeness of highly active 8-MG derivatives for competitive inhibition of DHPS. The octanol-water partition coefficient ($\log P_{o/w}$), H-bond donor, H-bond acceptor, molecular weight (MW), dipole moment, blood-brain barrier and membrane permeability, etc properties of the hit compounds are calculated to evaluate their drug-like properties.

4.2.7 The MD Simulation Protocol

All the MD simulations were performed in GROMACS-2016.5 (Abraham et al. 2015a) using AMBER99SB force field (Hornak et al. 2006b; Nguyen et al. 2014). The details of the atomistic-MD simulations of protein-ligand systems are given in **Table 4.2**. The topology of the selected inhibitors was prepared using ACPYPE (Sousa da Silva and Vranken 2012a) software package. All the four protein-ligand systems were solvated separately in cubic boxes with TIP3P water molecules (Mark and Nilsson 2001b). A 1 nm buffer space was maintained between the surface of the protein and the edge of boxes for sufficient solvation. The approximate volume of each box was set to be

421.875 nm³. The overall charge of the system was neutralized by adding six Na⁺ ions and 0.15 M NaCl concentration was maintained. In order to remove bad contacts during solvation, energy minimization was performed using the steepest descent algorithm using 50,000 steps until a gradient threshold of 10 kJ/mol was reached. The bonds connected to hydrogen atoms in the protein are constrained by LINCS(Hess et al. 1997a) algorithm, and SETTLE algorithm (Miyamoto and Kollman 1992a) was used to constrain the geometry of the water molecules. We have used leap-frog integrator algorithm to integrate the equation of motion with a time step of 2fs. All the systems were initially heated up to 300K for 1ns and further equilibrated for 4 ns. The temperature of the systems was maintained by V-rescale (Bussi et al. 2007a) coupling algorithm ($\tau_t=0.1$ ps).

Table 4.2. Summary of performed MD Simulations. MD simulation calculations are run thrice to assess the reproducibility.

Protein	Ligand	Number of Water molecule	Protein+ Ligand Atoms	Box Length (nm)	Simulation time (ns) and sampling interval (ps, subscript)
Dihydropteroate Synthase (5U10)	Compound 62	14564	4086	7.86	100 ₁₀
	Compound 61	14584	4075	7.86	100 ₁₀
	Compound 51	14573	4087	7.86	100 ₁₀
	Compound S1	14581	4089	7.86	100 ₁₀
	Compound S3	14568	4109	7.86	100 ₁₀
	Compound 6	14574	4065	7.86	100 ₁₀
	Pterioic Acid	14563	4089	7.86	100 ₁₀
	Apo	14594	4054	7.86	100 ₁₀
Mutated Dihydropteroate Synthase	Compound 51	14567	4079	7.86	100 ₁₀
	Compound S3	14543	4096	7.86	100 ₁₀

After temperature equilibration, the system condition is changed into an isothermal-isobaric ensemble from the canonical ensemble for 1ns and further equilibrated up to 9 ns. During NPT equilibration the box volume was allowed to fluctuate isotropically to attain a steady-state and the system pressure was maintained by Parrinello-Rahman (Parrinello and Rahman 1981a) coupling algorithm ($\tau_p= 2$ ps). After attaining appropriate density of the system, we removed the restrains to allow the free or unbiased

movement of protein-ligand complex for 100 ns. The cut-off radius for calculating short range electrostatic and van der Waals interaction was set to 1.2 nm. The long-range electrostatic interactions were calculated by the Particle Mesh Ewald method (Darden et al. 1993; Essmann et al. 1995b). Further, the trajectories were saved at every 10 ps for our analysis. The mutation in the DHPS catalytic pocket was introduced employing Pymol software and the simulation protocol was same with the wild type protein-ligand simulations. To check the reproducibility, we further repeated the MD simulation of each system twice.

4.2.8 Definition of Non-bonded Interaction

The ligand molecules are said to be hydrogen bonded if the donor (D)-acceptor (A), D-D distance and the <DDA angle satisfy some specific criteria. The critical distances for donor-acceptor and donor-donor were set to 0.35 nm and 0.245 nm respectively to identify the hydrogen bond between ligand and protein complexes from molecular docking studies. Further, the cutoff for <DDA was set to 30° (Chandra 2000).

The π - π stacking interaction between the aromatic rings of ligand and the aromatic amino acid sidechains at the catalytic pocket of protein is said to be formed if the inter-centroid distance is less than 0.8 nm and the angle between the normal of one or both rings and the centroid-centroid vector must be between 0° to 60° or 120° to 180°. In case of T-shaped π - π stacking interaction, the above-mentioned angle must be ~90°. Additionally, at least one atom of each ring needs to be within 0.45 nm to form π - π stacking (P. Dimitrijević et al. 2012).

The geometric criteria for π -cation interaction are defined in terms of the distances and the angle between the π -system and the cation centre. The distance must be lesser than 0.6 nm and the angle (θ) must be $60^\circ \leq \theta \leq 90^\circ$ (Liang and Li 2018).

4.2.9 Free Energy Calculation

The endpoint free energy method MM/PBSA (Kumari et al. 2014a) was employed to compute the binding affinity of the compounds selected for MD simulation. It is evident from the literature that the thermodynamic condition for enzyme inhibition is isothermal -isobaric i.e. NPT. Hence, the last 40 ns production MD trajectory of protein-

ligand complexes were considered for MM/PBSA calculation. In general, the binding free energy of the ligand (L) at the catalytic pocket of receptor molecule (R) to form a complex (RL) can be defined as

$$\Delta G_{bind} = \Delta G_{RL} - (\Delta G_L + \Delta G_P) \quad (4.1)$$

Further, the eqn. 4.1 can be decomposed as

$$\Delta G_{bind} = \Delta H - T\Delta S \approx \Delta E_{MM} + \Delta G_{sol} - T\Delta S \quad (3.2)$$

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{ele} + \Delta E_{vdw} \quad (3.3)$$

$$\Delta G_{sol} = \Delta G_{GB/PB} + \Delta G_{np} \quad (3.4)$$

$$\Delta G_{np} = \gamma \cdot SASA + b \quad (3.5)$$

Where, ΔE_{MM} is the changes in gas phase molecular mechanics energy, which includes internal energies (ΔE_{int}) arising from the bond, atom, dihedral terms, electrostatic interaction (ΔE_{ele}), and van der Waals (ΔE_{vdw}) interaction terms. The solvation free energy term (ΔG_{sol}) consists of polar and nonpolar solvation energies between the solute and continuum solvent. The polar solvation energy term is generally computed by the Poisson-Boltzmann equation for obtaining electrostatic potential of solute in the solution while the nonpolar solvation is obtained from the linear equation of the solvent-accessible surface area of the solute (Gilson and Honig 1988a; Wang et al. 2019). The change in conformational entropy i.e. $-T\Delta S$ can be calculated by normal mode analysis (Srinivasan et al. 1998a; Wright et al. 2014). However, due to the high computational cost, the change in conformational entropy of similar ligands is often neglected.

4.2.10 Umbrella Sampling

Umbrella Sampling (US) simulations have long been used to explore the mechanism of ligand-protein unbinding and the determination of binding free energy (Huang et al. 2017b; Miao et al. 2018; Sun et al. 2015; Zhou et al. 2015b). However, the unbinding pathway modelled by such enhanced sampling methods must mimic the natural molecular movement to generate unbiased Potential of mean force (PMF). Therefore, identification of correct pulling direction is important to generate initial conformations for the US (You et al. 2019). In the present case, we employed steered molecular

dynamics, an enhanced sampling method to pull the highly active 8MG derivative, and compared the unbinding free energetics with one inactive compound. In order to define RC, we selected the vector connecting the C β of Leu 215 and the central Sulphur atom of 8-MG compounds. The reason behind selecting Leu 215 as a reference point of dissociation is due to its smallest root mean square fluctuation (RMSF) value. The rate of ligand movement was set to 10 nm/ns and the ligand was pulled up to 4 nm from the reference point. The elastic constant k_i was set to 1000 kJ/mol.nm². The starting conformation for steered MD (SMD) simulation is taken from the last frame of our 100 ns conventional MD simulation performed above. The SMD simulations were carried out at 300K, 1 bar using Nose-Hoover temperature and Parrinello-Rahaman pressure coupling algorithm(Lemkul and Bevan 2010a). Further, conformations are taken at every 0.2 nm distance as a starting conformation for the US, and 10 ns US was performed at each window. The unbiased PMF was constructed from each US window by WHAM(Kumar et al. 1992a) and MBAR protocol(Shirts and Chodera 2008).

4.3. RESULTS AND DISCUSSION

4.3.1 Ligand Based identification of structural determinants for DHPS inhibition

We have used pharmacophore modelling and 3D-QSAR studies to primarily identify the crucial structural counterparts of 8-MG based molecules required for DHPS inhibition. A Total of thirty pharmacophore hypotheses are found to be predicted by PHASE and documented in **Appendix IX**. The DDHRR_1 (**Figure 4.2-A**) is chosen as the best pharmacophore hypothesis due to its highest Phase Hypo score of 1.34 and survival score of 5.81 which indicate good reliability and predictability of the model.

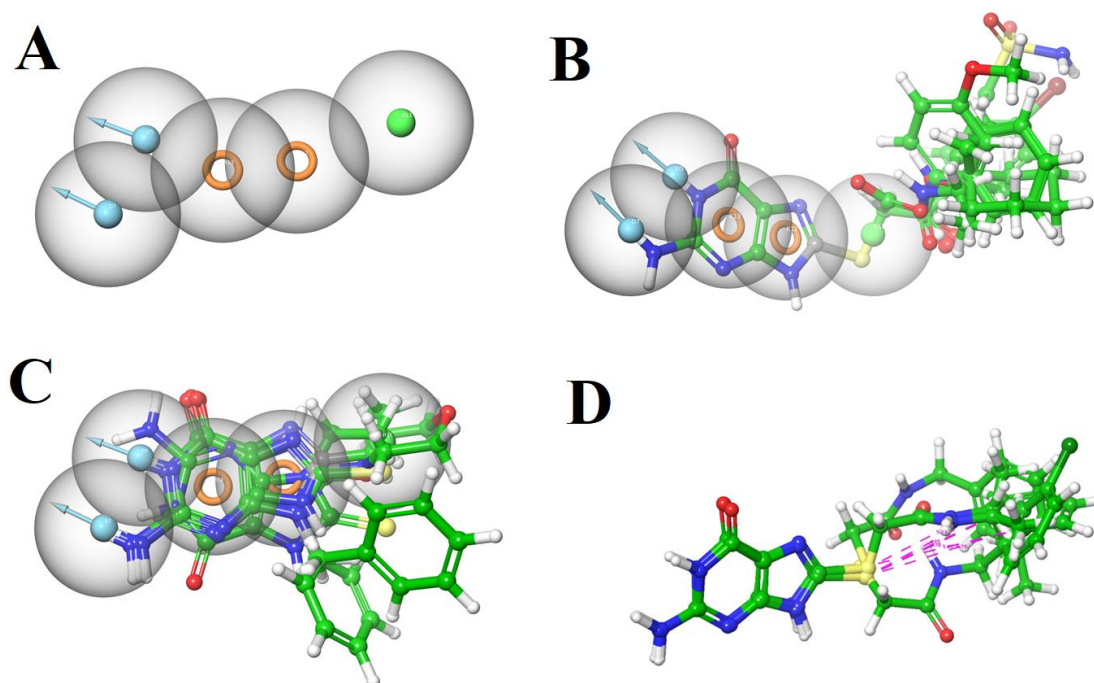


Figure 4.2. (A) The special distribution of pharmacophoric features of best score hypothesis DDHRR_1. Pharmacophore mapping of the most (B) active and (C) inactive inhibitors. (D) The distance between the central sulphur atom and the hydrophobic groups in active 8-MG compounds.

This pharmacophore hypothesis consists of five pharmacophoric features such as two hydrogen bond donors (D), one hydrophobic centre (H), and two aromatic ring (R) that are found to be present in all the active compounds. The inactive compounds are found to lack the hydrophobic features that are important for the

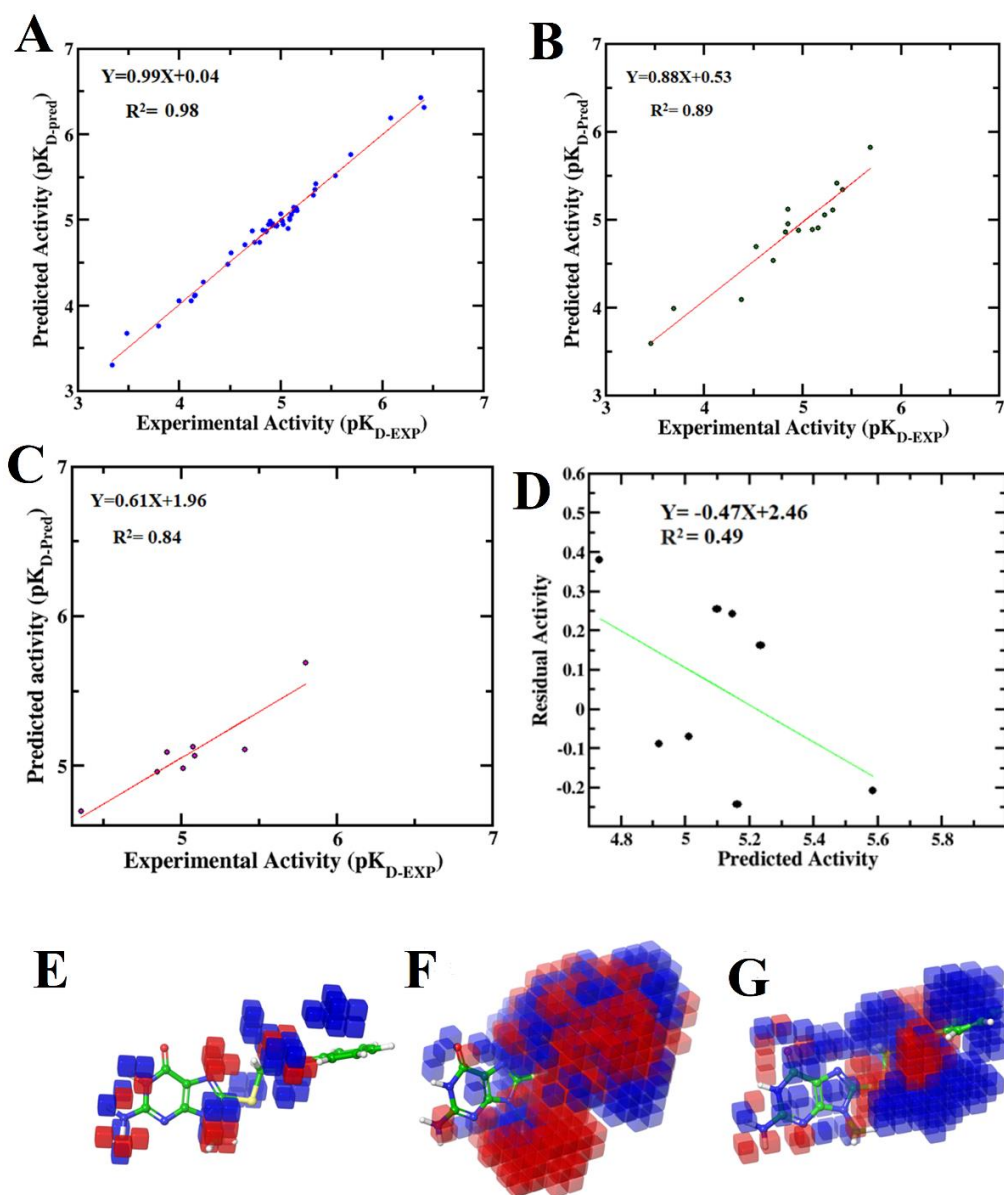


Figure 4.3. Scatter plot between the experimental and predicted activity of the (A) training set, (B) test set, and (C) external test-set compounds considered in 3D-QSAR studies. (D) Scatter plot between the Residual activity (Predicted Activity-Experimental activity) and predicted activity. The 3D-QSAR contour map fitted on compound 62 ($pK_D=6.409$) in the context of (E) hydrogen bond donor, (F) Hydrophobic group, and (G) electron-withdrawing group.

competitive inhibition of DHPS. The distance and angle between the pharmacophoric sites are believed to be a crucial attribute in competitive inhibitions (**Table-4.3**). The active and inactive molecules are aligned on the pharmacophore hypothesis and shown in **Figure 4.2 B, C**. It is evident from **Figure 4.2-B, C** that the reason behind the difference in the active and inactive compounds is the intrinsic distances and angles of pharmacophoric sites. Further to assess the ability of the DDHRR_1 model to discriminate the bioactive compounds for DHPS inhibition, we tested the model against 252 decoy dataset compounds obtained from the decoy finder server.

Table 4.3. The distance and angle between the pharmacophore sites of DDHRR_1 hypothesis.

Site 1	Site 2	Distance (Å)	Site 1	Site 2	Site 3	Angle (°)
D5	D7	3.48	D7	D5	H11	55.25
D7	H11	7.08	D5	D7	H11	94.49
D5	H11	8.13	D7	H11	D5	25.26
D5	R13	3.45	D7	R13	D5	71.32
D7	R13	2.30	R13	D7	D5	69.73
D5	R12	5.15	D7	D5	R13	69.91
D7	R12	4.23	R13	D5	R12	16.1
R13	R12	2.07	R13	R12	D5	27.54
R12	H11	3.07	D5	R13	R12	136.35
R13	H11	5.11	D5	R12	H11	162.45
			R12	D5	H11	6.53
			R12	H11	D5	11.01
			R13	R12	H11	167.37
			R12	H11	R13	5.08
			R12	R13	H11	7.54
			D7	R13	H11	142.80
			R13	D7	H11	25.86
			R13	H11	D7	11.32
			D5	R13	H11	142.79
			R13	D5	H11	22.38
			R13	H11	D5	14.86

The pharmacophoric model is found to retrieve 100% hit compounds from the decoy dataset. The contribution of the active molecule in enrichment, the robust initial enhancement (RIE) value was calculated for all the pharmacophore models. The RIE for the DDHRR_1 model is found to be 18.14 indicating the superiority of the model ranking over the random distribution. Lastly, taking the DDHRR_1 model as a 3D-

structural query, we performed screening to retrieve bioactive compounds as DHPS inhibitors from the database (containing 81,549 hits) made by Phase. Database screening yields 4128 hits that can act as DHPS product analogues.

Next, the 3D-QSAR model is analyzed to unveil the effect of functional group substitution on the biological activity of 8-MG derivatives for DHPS inhibition. It is believed that atom-based 3D-QSAR modelling is more efficient in predicting structure-activity relationships compared to pharmacophore-based 3D-QSAR models. To reduce the number of descriptors, the original variables correspond to sixty-two 8MG compounds are converted to latent variables or factor scores. A non-hierarchical method such as k-means clustering (Everitt n.d.) was employed on the factor scores of all the compounds and five clusters are formed. The deviation of the compounds to their corresponding cluster is depicted in **Appendix X**. The test set compounds are selected based on proximity to the centroids in each cluster and the remaining compounds are considered as the training set. The biological activity of the 8-MG compounds predicted by our model is shown in **Table 4.1**. In addition, the scatter plot of both training set and test set compound illustrated in **Figure 4.3-D, E** indicates an excellent linear correlation and moderate difference for the biological activity of 8-MG compounds predicted by our 3D-QSAR model. The PLS regression summary is depicted in **Table 4.4**.

Table 4.4. PLS statistical parameters of the generated 3D-QSAR model

Factors	SD	R ²	Stability	F	P	RMSE	Q ²	Pearson-R
1	0.4337	0.5441	0.958	51.3	7.45e ⁻⁰⁹	0.25	0.7531	0.8890
2	0.3749	0.6727	0.859	43.2	6.51e ⁻¹¹	0.25	0.7369	0.8658
3	0.3125	0.7780	0.624	47.9	1.84e ⁻¹³	0.24	0.7626	0.8817
4	0.2410	0.8712	0.272	67.7	2.89e ⁻¹⁷	0.23	0.7801	0.8870
5	0.1806	0.9295	0.127	102.8	2.23e ⁻²¹	0.21	0.8016	0.9081
6	0.1349	0.9617	0.0637	158.9	2.38e ⁻²⁵	0.21	0.8185	0.9083

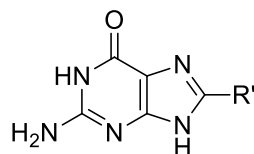
SD: Standard deviation of the regression, R²: Correlation Coefficient, F: Variance ratio, P: significance level of variance ratio, RMSE: Root mean square error, Q²: cross-validation coefficient, Pearson-R: Value for the correlation between the predicted and observed activity for the test set.

The lower values of RMSE (0.21) and standard deviation (0.1349) at the sixth PLS imply that the data used for the model is reliable for 3D-QSAR analysis. The predictive power of our 3D-QSAR model is assessed by the cross-validation coefficient (Q^2) which is found to be 0.8185, indicating excellent predictive power of the model. In addition, the relevance of the model is reflected by the regression coefficient of 0.9617 for the training set compounds. The large value of F (158) and the range of stability from 0.958 to 0.063 on the maximum scale of 1 supports the statistical significance of this regression model. Furthermore, the degree of confidence of the 3D-QSAR model is supported by the lower value of P ($2.38e^{-25}$) (Bhole et al. 2021). The predictive power of the 3D-QSAR model is further validated by the external test set (Dennis et al. 2014) (**Table 4.5**). The scatter plot of experimental activity vs. predicted activity and predicted activity vs. residual activity are depicted in **Figure 4.3-C**. The regression coefficient (R^2) of 0.79 indicates the good predictability of the activity of the compounds that are not considered in model development. It is believed that a 3D-QSAR model with an R^2 value above 0.5 is a good model to predict the activity of unknown compounds (Golbraikh et al. 2003; Golbraikh and Tropsha 2000c). It is evident from **Figure 4.3-D** that there are no outlier compounds predicted by our 3D-QSAR model which further indicates the stability of the model.

Furthermore, contour plot analysis is performed for 8MG derivatives to inspect the relevance of spatial arrangement of structural determinants to explain their activity. In **Figure 4.3** the favorable zone (blue cubes) and the unfavorable zone (red cubes) for the biological activity of the compounds considered here are shown by applying the 3D-QSAR model at 6th PLS on most active compound 62 in the training set. The blue cubes located at the $-\text{CONH}$ group and $-\text{COOH}$ group linked to the H11 pharmacophoric site of compound 62 ($\text{pK}_D = 6.409$) and compound 61 ($\text{pK}_D = 6.377$) respectively indicate the preference of hydrogen bond donor group in this place (**Figure 4.3-E**). This is further supported by the presence of $-\text{CONH}$ group in blue cube region for highly active such as compound 50 ($\text{pK}_D = 5.319$), compound 53 ($\text{pK}_D = 5.347$), compound 55 ($\text{pK}_D = 5.538$) and compound 56 ($\text{pK}_D = 5.688$). Moreover, the absence of such hydrogen bond donor in compound 59 ($\text{pK}_D = 4.523$), 60 ($\text{pK}_D = 4.721$) and 32-34 (pK_D values 4.795-4.824) decrease their activity group. The red cubes located near the R12 pharmacophoric region indicate the unfavorable region for the hydrogen

bond donating group. The fact is supported by the lesser activity of compound 55 ($pK_D = 4.721$) and compound 56 ($pK_D = 4.721$) compare to the highest active compound 61. For hydrophobic interaction attributes, the blue cubes located at the benzene ring of compound 62 and compound 57 ($pK_D = 5.690$) indicate a favorable region for hydrophobic interaction with DHPS (**Figure 4.3-F**). This is further supported by the highly active compound 53 ($pK_D = 5.347$) and 54 ($pK_D = 5.409$). The location of the hydrophobic group attached with the H11 pharmacophoric site is found to be important in enhancing the activity of the compound. The placement of hydrophobic moieties greater than 4.7 Å from the Sulphur atom is found to increase the activity of 8-MG compounds (**Figure 4.3-D**). The red cubes located within the 4.7 Å of the Sulphur atom indicate the unfavorable region of hydrophobic groups. The activity of compound 5 ($pK_D = 3.796$) and compound 9 ($pK_D = 4.149$) is decreased due to the presence of hydrophobic groups at this position. Furthermore, the red cubes located near R12 pharmacophoric site indicate the presence of bulky hydrophobic group which decreases the activity of 8-MG based compounds. This assumption is supported by the low activity of compound 1 ($pK_D = 3.337$), compound 2 ($pK_D = 3.456$) and compound 6 ($pK_D = 3.796$). Further, we analyzed the effect of the spatial arrangement of electron-withdrawing groups on the activity of 8-MG compounds (**Figure 4.3-G**). The blue cubes located at the benzene ring indicate the preference of electron-withdrawing groups in this region. The presence of Cl at 3rd position, Br in 4th position and -SO₂-group in 4th position of benzene ring for compound 40 ($pK_D = 5.004$), compound 51 ($pK_D = 5.337$), compound 57 ($pK_D = 5.609$) respectively increases the activity. The red cubes located near the H11 pharmacophoric site indicate the unfavorable region for the electron-withdrawing group. It is evident from the 3D-QSAR contour map that the presence of -CONH groups or ketone groups as a linker between the benzene ring and H11 pharmacophoric site enhances the activity of 8MG-based compounds. In addition, the presence of the electron-withdrawing group increases the activity of the compounds for DHPS inhibition.

Table 4.5. External test set compounds for validating the developed 3D-QSAR model.



Compound No.	Structure	Experimental Activity(pK _{D-EXP})	Predicted Activity (pK _{D-PRED})	Residual Activity
1		5.408	5.163	-0.245
2		5.086	5.014	-0.072
3		5.795	5.586	-0.209
4		5.075	5.236	0.161
5		4.847	5.1	0.235
6		4.357	4.735	0.378
7		5.011	4.92	-0.091
8		4.910	5.15	0.24

4.3.2 Binding Mode of 8MG-compounds to the DHPS catalytic pocket

Molecular docking studies of the 8MG-compounds retrieved from both literature and database screening were carried out at the DHPS catalytic pocket for visual inspection of crucial amino acid interactions of the inhibitors responsible for DHPS inhibition. We validated our docking results by measuring the RMSD between the crystal and redocked structure of the ligand (PDB ID: 5U10, 2.04 Å) which is found to be 0.04 nm (**Appendix XI**). RMSD value lesser than 0.2 nm is considered to be best for predicting the accurate binding pose of unknown inhibitors(Kramer et al. 1999). The docking scores

and Glide predicted binding free energy (Glide Emodel) of all the compounds are depicted in **Appendix XII**. The best binding conformations of highly active and inactive 8MG derivatives are shown in **Figure 4.3**. It is evident from our docking results that hydrogen bonding, π -cation, and π -stacking interactions are key interactions that anchor the ligand inside the catalytic pocket. All the interactions are found to be located between the amino acid stretch of Pro64 to Met139 and Asp185 to Arg 225 indicating the pronouncing binding pocket of DHPS. The highly active compounds found to be involved in hydrogen bond interaction with Asn96, Asn115, Gly189, Asn115, Asp185, Lyn 221 that reside inside the catalytic pocket of DHPS. The R12 and R13 pharmacophoric features of 8-MG rings are found to form π - π stacking interaction with Thr62 and Arg225 residues which play an important role in stabilizing the inhibitors in DHPS active site (**Appendix XIII**). We also generated the map of hydrophobic and hydrophilic fields for highly active inhibitors and depicted in **Appendix XIV**. It is evident from **Appendix XIV** that the DHPS pocket mostly consists of hydrophilic amino acids. However, small hydrophobic portions are also present in the active site. The benzene rings of highly active inhibitors are found to reside near the hydrophobic part of the surface which is found to be correlated with the 3D-QSAR contour plot of the hydrophobic feature. The guanine ring of the 8MG derivatives is found to be buried inside the hydrophilic part located at the pterin binding site of the pocket. The docked pose of highly active compounds inside the DHPS pocket is shown in **Figure 4.4A-F**. The imidazole ring of compound 62 is found to be involved with π - π stacking interaction with Phe 190. There is also an equal probability of forming π -cation interaction with the guanidinium of Arg255(Gallivan and Dougherty 1999). Moreover, the nitrogen atoms substituted in the 8MG ring are found to accepted hydrogen bonds from Asn115 (8MG-N---H-NH-Asn115, 2.51 Å) and donate hydrogen bond to Asn115(Asn115-C=O---H-8MG, 1.68Å), Asp185(C-O⁻¹---H-8MG, 1.79 Å), Lyn221(8MG-NHCO----HN-Lyn221, 2.48 Å). One water-mediated hydrogen bond is found to anchor the C=O of 8-MG ring to Gly217, Gly187, and Asp185. Another highly active compound 51 is found to be involved in hydrogen bonding interaction with Ser222 (NH-----O=S-8MG, 2.63Å), Gly189(C=O-----HN-8MG, 2.00Å).

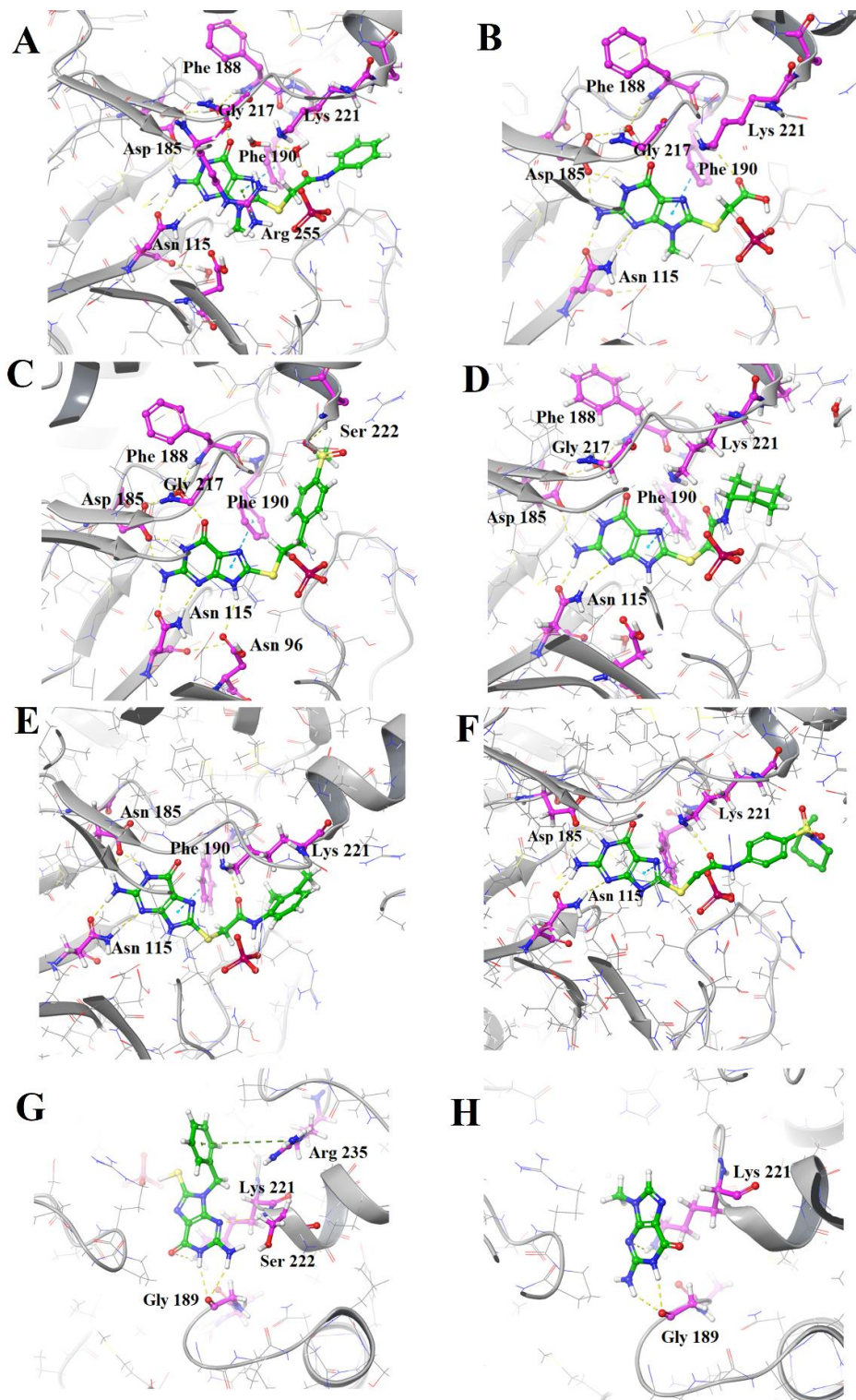


Figure 4.4. The docking pose of highly active (A) Compound 62, (B) Compound 61, (C) Compound 51, (D) Compound 55, (E) Compound S1, (F) Compound S3 and inactive (G) Compound 2, (H) Compound 11. The hydrogen bonding, π - π stacking, and

π -cation interactions are shown in yellow, sky, and green lines respectively. The interacting amino acids are coloured in pink.

The imidazole ring of compound 51 is found to involved hydrogen bond interaction with Asp96 (C=O-----HN-IME-8MG, 2.78Å) (**Figure 4.4-C**). We also perform XP docking of 4128 compounds (obtained from database screening) with respect to DDHRR_1 hypothesis and found five active compounds which can inhibit DHPS. The structures of the compounds are provided in **Appendix XV**. Compound S1 (2-((2-amino-6-oxo-6,9-dihydro-1H-purin-8-yl) thio)-*N*-(2,3-dimethylphenyl) acetamide) and compound S3 (2-((2-amino-6-oxo-6,9-dihydro-1H-purin-8-yl) thio)-*N*-(4-((2-methylpiperidin-1-yl) sulfonyl) phenyl) acetamide) is found to obtain highest docking score (**Appendix XVI**). The 8-MG rings of these compounds are found to have similar interactions as found in the case of highly docked compounds discussed above (**Figure 4.4-E, F**). In contrast, the inactive compound 2 and 11 are found to bind at the surface of the DHPS catalytic pocket and their guanine ring is found to interact with Gly 189 and Lys 221 (**Figure 4.4-G, H**). Such non-specific interactions decrease the inhibitory effect of the inactive compounds and destabilize them inside the DHPS catalytic pocket. Therefore, it is evident from docking studies that the marcaptoguanine ring of the inhibitors is primarily stabilizing the base of the ligand, situated deep inside the pterin-binding pocket and the conjugation from the Sulphur atom restricts the shallow region of pyrophosphate and *p*ABA binding pocket. Such binding pose of the inhibitors is found to cover all the binding pockets of the DHPS catalytic cavity and circumvent the sulfa-resistance mutations often located at the *p*ABA binding pocket.

4.3.3 DFT Calculation

In order to rationalize and interpret the chemical reactivity, kinetic stability of 8MG compounds DFT study has been employed. The energy corresponds to HOMO, LUMO of highly active (compound 62, 61, 51, 56, 58, S1, S3), inactive (compound 6, 2, 11) inhibitors, and product are given in **Appendix XVII**. The HOMO, LUMO energy is found to be varied significantly in vacuum and aqueous environment which are calculated in two different basis-set such as 6-31G**(d, p), 6-311G**(d, p). It is observed from **Appendix XVII** that HOMO, LUMO energies of 8MG compounds are

lowest in the aqueous solution of 6-311G***(d, p)*, thus considered for further analysis. The HOMO energy ranges from -5.971 eV to -6.672 whereas the LUMO energy ranges from -0.903 eV to -3.125. It can be seen that there is not much difference in the HOMO energy but LUMO energy differs for each molecule. Hence, LUMO energy can be considered as the principal signficator for the biological activity of the 8-MG inhibitors. Further, the band gap (energy difference between HOMO and LUMO) is calculated to assess the kinetic stability of the inhibitors(Bharathi et al. 2016). It is clear from **Appendix XVII** that inactive compounds have lower kinetic stability compare to active compounds due to a higher band-gap ($E_g > 5$). Among the highly active inhibitors, compound 51 is found to have the lowest band-gap indicating the highest stability among the inhibitors. It can be further noted that the band-gap for pterioic acid is much lower compared to all the active inhibitors. Thus, 8-MG derivatives are chemically more inert compare to the product (pteroic acid) at the DHPS catalytic pocket which is helpful for better competitive inhibition of DHPS. The HOMO and LUMO regions are known to be directly proportional to the electron-rich and deficient regions of the molecule. Therefore, these regions indicate the reactive regions of inhibitor molecules for stable interaction with catalytic pocket residues. The occupancy of the HOMO, LUMO orbitals on the 8MG derivatives are depicted in **Appendix XVIII- XX**. It can be found that the location of LUMO significantly differs in the case of active compounds. The electron reaches HOMO located at the guanine ring (R12 and R13 pharmacophoric site), which can provide strong hydrogen bonding and cation- π interaction. While the LUMO electron density majorly delocalized at the substitutions associated with the central Sulphur atom (**Appendix XVIII**). The non-overlapping position of HOMO-LUMO in 8-MG derivatives may provide better stability and interaction with the DHPS catalytic pocket. Similar observation is obtained from compound S1 and S3(**Appendix XIX**). Unlike the active compounds, the location of HOMO-LUMO is found to be completely opposite in the case of pterioic acid (**Appendix XX**). This can be the reason behind the higher chemical reactivity of this compound compare to 8MG analogs which help in maintaining kinetic equilibrium between substrate and product at DHPS active site. Furthermore, the HOMO- LUMO location is found to be placed in the guanine ring of compound 2, 6 and 11 which may

hinder the stability and nonbonded interaction at the protein's active site, thereby decreasing the activity of these compounds (**Figure A9**).

The global reactivity descriptors such as dipole moment, chemical potential (μ), electronegativity (χ), hardness (η), softness (s) and electrophilicity index (ω) were calculated to explain the stabilization and interaction of 8-MG derivatives at DHPS catalytic pocket. It can be seen from **Appendix XXI** that compound 51, S3 is found to have a higher dipole moment due to higher number of polar atoms, thereby increasing their activity. Among highly active inhibitors, the chemical hardness of compound 51 is much lower, indicating higher biological activity of the compound (Venugopal et al. n.d.). The hardness of pteric acid is the lowest compare to all the inhibitors considered for DFT studies which further validates that the chemical reactivity of the inhibitors is comparatively lesser than the actual product of DHPS. It can be noted from **Appendix XXI** that the electronegativity and electrophilicity index of the active inhibitor are higher compare to inactive inhibitors. Therefore, the active inhibitors have well-defined electron-rich and deficient regions in their molecular geometry to stabilized itself at the active site of DHPS.

4.3.4 ADME/Tox profiling

In order to avoid poor pharmacokinetics, oral bioavailability, and side effects, the determination of ADME/Tox properties of the drug candidates is believed to be crucial in the drug discovery process (Llorach-Pares et al. 2018; Nolte et al. 1998). The highly active 8MG compounds ($pK_D > 5.0$) with high docking scores are considered to be evaluated by Lipinski's rule of five, the percentage of human oral absorption. The ADME/Tox data of the 8-MG compounds are shown in **Appendix XXII**. The molecular weights of all the 8MG derivatives are found to be ranging from 250–350 Da and other factors of Lipinski's rule ($\log P$, MW, HBA, HBD) are also found to be in the permissible range. Moreover, the percent oral absorption values of the compounds are above 25%. This signifies good oral absorption of the hits in the human body. Further, reduced values of QPlogS of the 8-MG compound validate the higher solubility and drug-likeness of the compounds. It can be observed that some of the active drug candidates have the probability to block HGRT K^+ channel and poor Caco-cell line permeability. However, the most active compounds (Compound 51, Compound 62,

Compound 61 and Compound S1, Compound S3) have considerable values for the above two pharmacokinetic factors. Further, the blood-brain barrier permeability values of most of the active compounds exhibited positive results which signifies that 8-MG compounds have the least possibility to affect the central nervous system. The QPlogK_{hsa} values of active compounds indicate that the drug candidates have lesser interaction with human serum albumin. Lastly, the polar surface area (PSA) values of the active compounds indicate good permeability across the cell monolayer.

4.3.5 MD Simulation

4.3.5.1 Dynamic stability of protein-ligand complex

In order to obtain further insights into the dynamic stability and conformational change of the protein-inhibitor complex, 100 ns long MD simulation of the highly active, inactive inhibitor bound DHPS and *apo*-DHPS was performed. We also carried out MD simulation of DHPS-pterolic acid (product) complex for the same time scale to compare its dynamic behavior with inhibitors bound state. The docked conformations of the compounds are taken as the initial state for MD simulations. The root-mean-square deviation (RMSD) of the protein backbone and inhibitor compounds are shown in **Figure 4.5** to determine the systematic deviation of protein-inhibitor complexes. The upper range of backbone RMSD for DHPS in all simulations is found to be < 0.3 nm which indicates structural stability of the protein during motion (Carugo and Pongor 2001). The RMSD of highly active compound 62 is stabilized at an average value of 0.13 nm throughout the course of simulation (**Figure 4.5-A**) while compound 61 maintained the average RMSD value of 0.1 nm up to ~74 ns and then converged at a higher value of 0.17 nm (**Figure 4.5-B**) for the rest of the simulation. The RMSD of compound 51 is found to have fluctuated till ~55 ns for exploring stable conformation and then reached steady-state with a lower RMSD value of 0.08 nm (**Figure 4.5-C**). Highly active compounds obtained from our pharmacophore-based virtual screening are found to reach a maximum RMSD value of ~ 0.2 nm then stabilized at the end of the simulation (**Figure 4.5-D & E**). It can be seen from **Figure 4.5** that compound 62 and 61 have the most stable conformation compares to other active compounds. In contrast, the RMSD of the highly inactive compound 6 showed several fluctuations throughout the simulation which indicates conformational instability of the inhibitor at

the catalytic pocket of DHPS (**Figure 4.5-F**). Next, we calculated the RMSD of the pteronic acid-DHPS complex to compare its stability with highly active inhibitors. The RMSD of pteronic acid is found to be stabilized at 0.1 nm which indicates that the conformational dynamics of the product is similar to the active 8-MG derivatives. It can be noted from **Figure 4.5-H** that the backbone RMSD profile of the *apo*-DHPS is higher compared to inhibitor bound or product bound state. This indicates higher conformational stability of DHPS in their ligand-bound state. To assess the convergence in MD simulation, we calculated RMS average correlation (Galindo-Murillo et al. 2015) and given in **Appendix XXIII**. It is evident from **Appendix XXIII** that the correlation profile for compound 51, compound 61, and compound S3 bound DHPS backbone approach quickly to the final structure compared to other inhibitor bound states. This further indicates the convergence of the RMS deviation for the above-mentioned 8-MG bound DHPS backbone. Further, the RMSF profiles of DHPS backbone residues in both ligand-bound and *apo*-state are displayed in **Figure 4.6(A-B)** to depict the highly mobile or flexible segments of the protein molecule.

The average RMSF profile of the DHPS backbone is found to be almost similar in all highly active inhibitor-bound states. However, the residue stretches from 62 to 77 is found to have a comparatively higher RMSF value (0.4 nm) with respect to the rest of the DHPS backbone, indicating the higher flexibility of this region during motion (**Figure 4.6-A**). Apart from this most of the backbone regions of the DHPS showed an RMSF value less than 0.1 nm due to the compact tertiary structural arrangement. The degree of fluctuation at the catalytic pocket residues (Asp 96 - Met 139 & Asp 185 – His257) is found to be < 0.2 which helps the formation of the stable protein-inhibitor complex. In the case of inactive compounds, the residue stretch 62-77 of DHPS is found to fluctuate more (~ 0.58 nm) compared to the active bound state (below 0.42 nm). Therefore, the flexibility of regions 62-77 is essentially restricted by active inhibitor binding for a stable protein-ligand complex. Further, it can be noted that the RMSF value of the residue stretches 222-239 is also stabilized by the active 8MG compounds. The C-terminal end of DHPS is found to fluctuate significantly in both active-bound and inactive bound states due to the solvent-exposed unstructured loops.

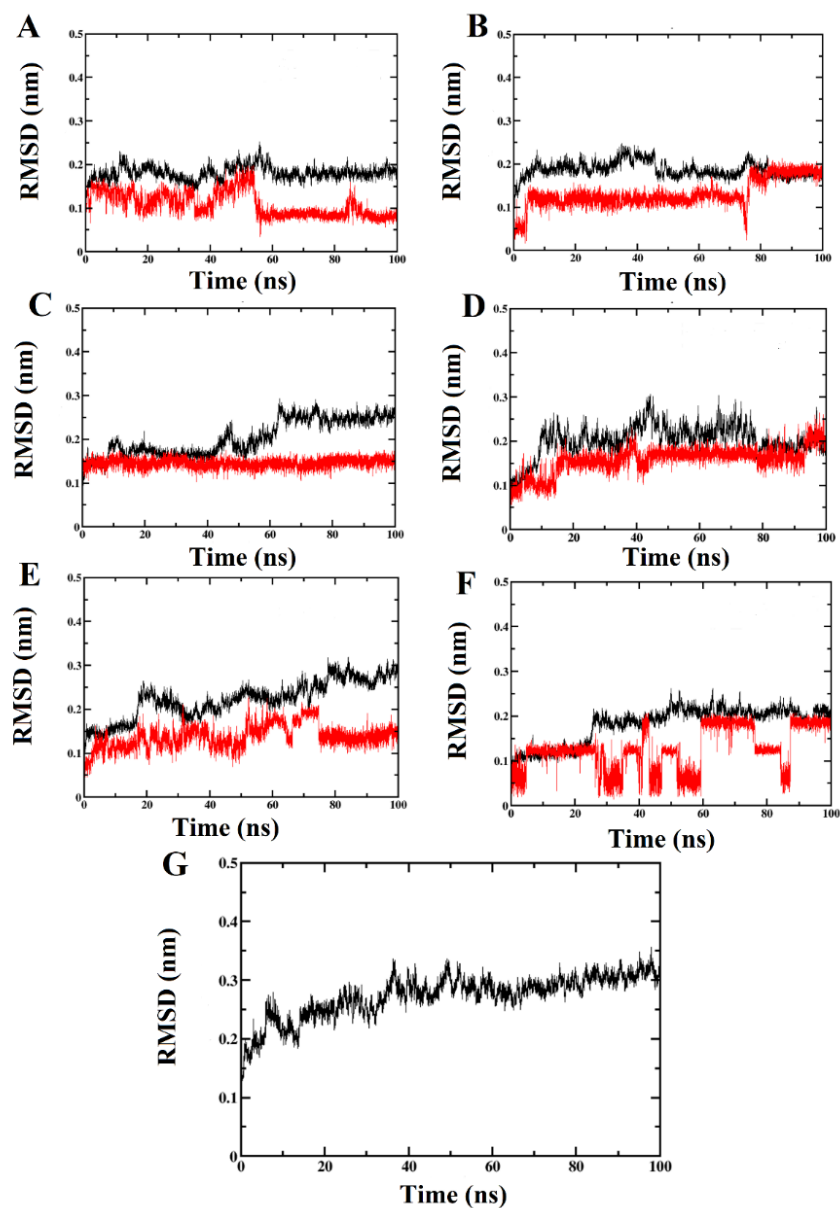


Figure 4.5. The time evolution of RMSD profile for DHPS bound with (A) Compound 62, (B) Compound 61, (C) Compound 51, (D) Compound S1, (E) Compound S3, (F) Compound 6, (G) Pteric acid and (H) Apo-DHPS. The black color represents the backbone RMSD and red color represents the ligand RMSD. (Batch 1).

The compactness of ligand-bound and apo-state of DHPS-8MG complex was evaluated by calculating the gyration radius (rGyr) throughout the simulation time (**Figure 4.6-C**). The rGyr value of the DHPS backbone in all the simulations ranges from 1.72nm to 1.79 nm with an average rGyr value of 1.75 nm. This confirms the fact that inhibitor binding does not cause the unfolding of the DHPS enzyme over the simulation period.

The dynamic stability of the protein-ligand complexes corresponds to simulation_batch 2 and simulation_batch 3 are given in **Appendix XXIII** and **Appendix XXIV**. The trend of the results is found to be almost similar compare to simulation batch 1.

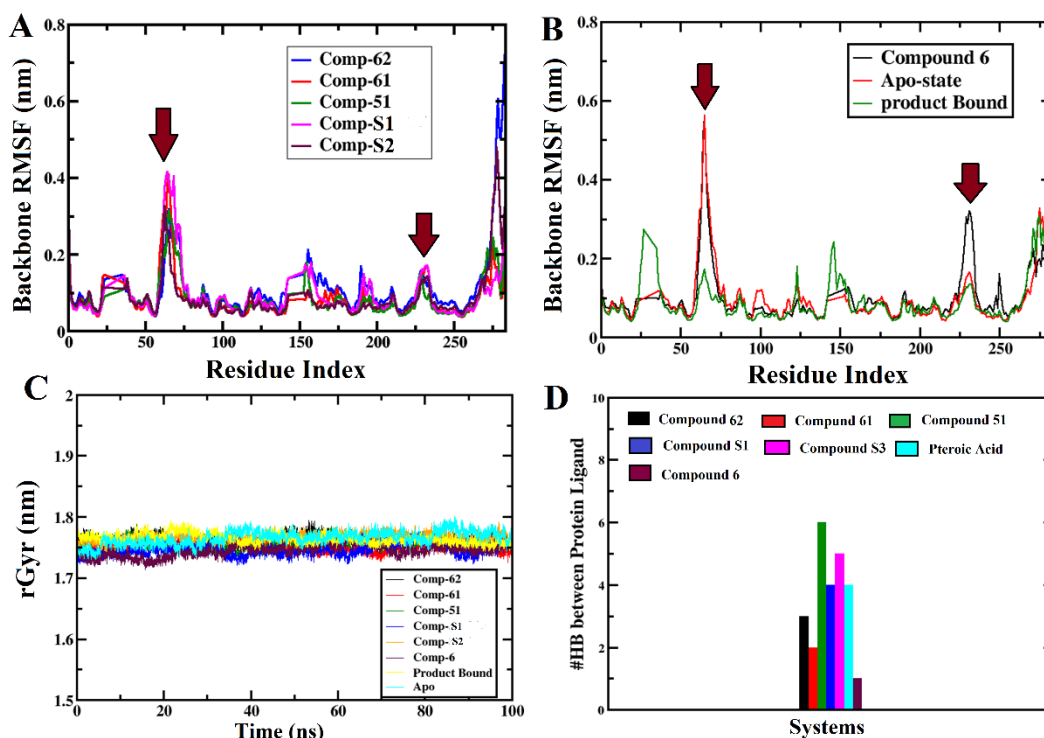


Figure 4.6. Plot showing the RMSF profile of (A) DHPS backbone in the presence of active inhibitors, (B) inactive, Product bound, and *apo*-DHPS. (C) The time evolution of gyration radius of DHPS backbone bound with active, inactive inhibitors, pterico acid, and *apo*-DHPS. (D) The number of average hydrogen bonds between the ligand and DHPS catalytic pocket.

4.3.5.2 non-bonded interactions

The stability of the docking predicted interaction profile of the inhibitors is assessed by MD simulation studies. The average number of hydrogen-bonding interactions involved in stabilizing the inhibitors inside the DHPS catalytic pocket was determined and depicted in **Figure 4.6-D**. We also determined the percentage of hydrogen bond occupancy (HBO) for the donor and acceptor pairs (**Table 4.6**). It is found that compound 51 formed the highest hydrogen bonds with DHPS catalytic pocket compare to other 8-MG derivatives indicating the substitution of sulphonamide group enhances

the electrostatic interaction to stabilize the outer part of the compound at the flexible loop of DHPS catalytic pocket. Similarly, compound S3 is found to form higher hydrogen bond compare to the S1 which further indicates the importance of the sulphonamide group. Compound 61 is found to have less hydrogen bond interaction compare to other active compounds. The residues such as Asn 115, Asp 185, Lys 221, Ser 222, and Arg 255 are found to be common which contribute hydrogen bonds to stabilize the active compounds. Additionally, Leu 184 (HBO: 98.0%), Leu 215 (HBO: 64.8%), Gly 215 (HBO: 47.8%) for compound 62, Asp 96 (HBO: 33.1%), Gly 189 (HBO: 25.7%) for compound 51 and Thr 62 (HBO: 381 %) for compound S3 stabilizes them at the catalytic pocket of DHPS. Therefore, it is evident from our MD study that most of the docking predicted interactions were present throughout the trajectory. The water-mediated hydrogen bonds between the active site residues and ligand are also crucial for anchoring the ligand in the catalytic pocket. The visual inspection of our MD trajectory suggested that Glu 60, Asp 116, Asp 165, Asp 185, Lys 201, and Leu 215 are involved in water-mediated hydrogen bonds with the active inhibitor molecules.

Apart from hydrogen bonding, the ligand or drug molecules are believed to be stabilized at the enzyme catalytic pocket by a delicate balance of weak interactions such as π - π stacking (Chen et al. 2018; Thakuria et al. 2019) and cation- π (Ma and Dougherty 1997; Scrutton and Raine 1996) interaction. In the present scenario, the contribution of such non-covalent interactions to stabilize the highly active inhibitors are assessed from our MD trajectories. The distance criteria for π - π stacking interaction are reported to be 3-8 Å, while the angle between the normal vector of two aromatic rings should be 0°-120°(Thakuria et al. 2019). The Phe 190 is found to be involved in π -stacking interaction with the guanine ring of the highly active inhibitors buried deep inside the DHPS catalytic pocket. This is assessed by the time evolution of the distance and the angle between the aromatic ring of Phe 190 and the marcaptoguanine ring of active inhibitors (**Appendix XXVI** and **XXVII**). In all the cases the average distance is found to be ranging from 0.58 nm-0.75 nm and the angle between the normal vectors is 90°-120°. In the case of compound 62, π stacking interaction is absent between the marcaptoguanine ring and Phe 190. However, Phe 188 is found to make such interaction with the substitute CONH-C₆H₆ ring and stabilizes the solvent-exposed part of the

ligand. Further, the marcaptoguanine ring of the active inhibitors is displayed cation- π interaction with Lys 221 and Arg 255(Kumar et al. 2018). The time evolution of the distance between the marcaptoguanine ring and the positively charged atom of both the residues is found to be below 0.7 nm (**Appendix XXVIII, XXIX**)indicating pronounced cation- π interaction to add stability for protein-inhibitor complex (Xiu et al. 2009).

Table 4.6. Hydrogen bond occupancy between ligand and active site residues of DHPS. UNK indicates the inhibitor and PT1 indicates Pterioic acid.

Compound 62				
Pair ID	Donor-	acceptor	Atom Number	Occupancy (%)
1	UNK283 (H1)	Leu184 (O)	4073-2529	98.0
2	Arg255 (H11)	UNK283 (O2)	3604-4064	71.3
3	Gly217 (H)	UNK283 (O1)	3038-4061	47.8
4	Leu215 (H)	UNK283 (N5)	3003-4059	64.8
Compound 61				
1	UNK283 (H7)	Asp56 (OD1)	4072-689	21.3
2	Lys221 (HZ1)	UNK283 (O1)	3113-4059	46.7
3	Asn115 (D21)	UNK283 (N2)	1591-4052	98.8
Compound51				
1	UNK283 (H11)	Gly189 (O)	4085-2589	25.7
2	UNK283 (N11)	Asp96 (OD2)	4087-1297	15.6
3	UNK283 (N11)	Asp96 (OD1)	4087-1296	17.5
4	UNK283 (H1)	Asp185(OD2)	4075-2539	17.9
5	UNK283 (H1)	Asp185 (OD1)	4075-2538	36.1
6	UNK283 (H1)	Asn115 (OD1)	4075-1589	35.8
7	UNK283 (H)	Asp185 (OD2)	4074-2539	36.0

8	UNK283 (H)	Asp185 (OD1)	4074-2538	60.7
9	Arg255 (H11)	UNK283 (N8)	3604-4057	14.0
10	Ser222 (H)	UNK283 (N22)	3126-4071	25.8
11	Ser222 (H)	UNK283 (O24)	3119-4073	33.0
12	Ser222 (H)	UNK283 (O23)	3119-4072	35.9
13	Lys221 (HZ1)	UNK283 (O7)	3113-4056	68.0
14	Asn115 (D21)	UNK283 (N1)	1591-4050	91.4
Compound 6				
1	Ser239 (HG)	UNK283 (O1)	3373-4053	31.3
2	Arg220 (H21)	UNK283 (N3)	3092-4055	20.0
3	Arg220 (HE)	UNK283 (N3)	3086-4055	43.8
4	Arg220 (H)	UNK283 (O1)	3073-4053	42.5
Compound S1				
1	UNK283 (H1)	Asp185 (OD2)	4074-2539	98.3
2	UNK283 (H1)	Asn115 (OD1)	4074-1589	24.3
3	UNK283 (H15)	Asp185 (OD2)	4088-2539	72.6
4	UNK283 (H15)	Asp185 (OD1)	4088-2538	29.0
5	Arg255 (H11)	UNK283 (N2)	3604-4053	39.1
6	Lys221 (HZ1)	UNK283 (O2)	3113-4065	22.6
7	Lys221 (HZ1)	UNK283 (O1)	3113-4059	16.9
8	Lys221 (HZ1)	UNK283 (N3)	3113-4054	67.0
9	Asn115 (D21)	UNK283 (N2)	1591-4053	38.6
Compound-S3				
1	UNK283 (H6)	Thr62 (OG1)	4087-759	38.1
2	UNK283 (H2)	Asp185 (OD2)	4083-2539	98.1
3	UNK283 (H2)	Asn115 (OD1)	4083-1589	96.3

4	UNK283 (H23)	Asp185 (OD2)	4184-2539	90.7
5	Lys221 (HZ1)	UNK283 (N1)	3113-4052	88.3
6	Asn115 (D21)	UNK283 (N5)	1591-4061	16.1
7	Asn115 (D21)	UNK283 (N3)	1591-4055	70.1
Pterioic acid				
1	PT1 283 (HO1)	Asp185 (OD2)	4247-2709	74.8
2	PT1 283 (HO1)	Asp185 (OD1)	4247-2708	23.2
3	PT1 283 (H1)	Asn115 (OD1)	4243-1629	91.8
4	Ser222 (H)	PT1 283 (O23)	3289-4242	20.1
5	Lys221 (HZ1)	PT1 283 (O1)	3283-4231	13.8
6	Lys221 (HZ1)	PT1 283 (N6)	3283-4223	84.6
7	Asn115 (D21)	PT1 283 (N9)	1631-4225	92.4

4.3.5.3 Principal Component Analysis (PCA)

In the present scenario, the last 40 ns of the trajectories were used for analyzing the principal motions of the protein in an inhibitor-bound state. The eigenvectors are calculated and given in **Figure 4.7** to comprehensively analyze the effect of inhibitor binding on protein motion. It is found that the total motion of inhibitor bound and apo DHPS monomer structure are dispersed over 789 eigenvectors.

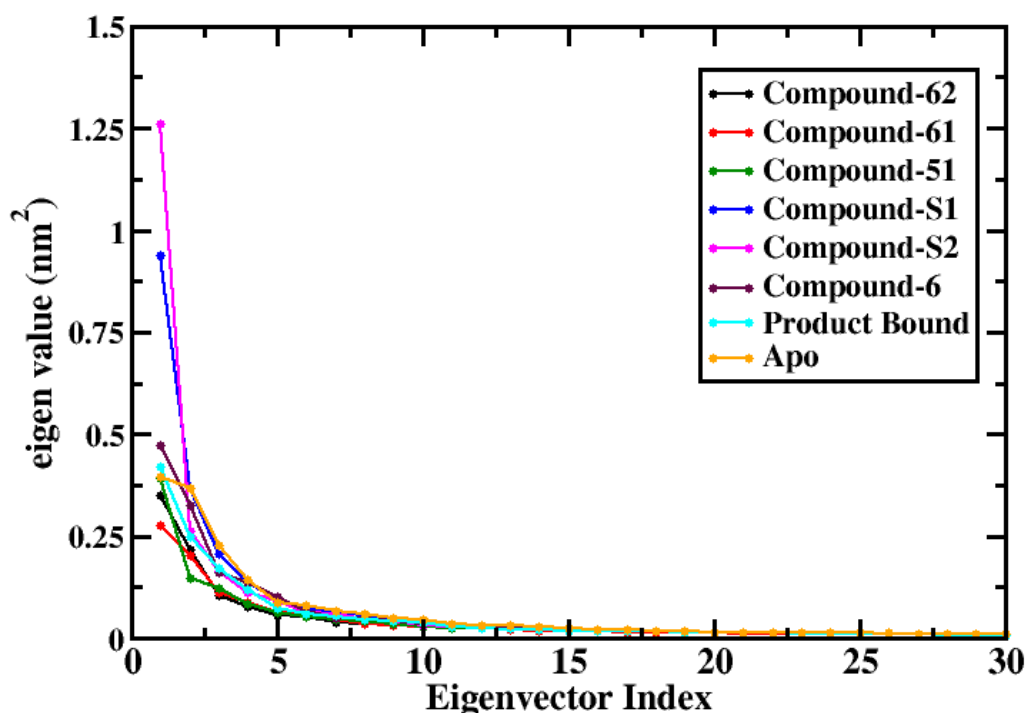


Figure 4.7 Plot represents the eigenvalues vs first 30 eigenvector indexes derived from PCA over last 40 ns of MD trajectory for ligand-bound and *apo*-DHPS structure.

The rapid decrease of the eigenvectors towards the Eigen index indicates that the initial 5 eigenvectors are mostly contributing to the collective motions of DHPS-8MG complexes during MD. It is found that first eigenvector corresponds to 42.65%, 43.79%, 45.78%, 61.05%, 64.94%, 49.2%, 44.51% and 48.3% motions for DHPS bound with compound 62, 61, 51, S1, S3, 6, Pteric acid and Apo-DHPS respectively (**Figure 4.7**). Thus, it can be inferred that the inhibitor-bound DHPS systems are stable compare to the DHPS bound with highly active compounds obtained from database screening. Additionally, the lower collective motions of the Apo-DHPS backbone indicate that the amalgamation of product analogs increases the principal motion crucial for minute conformational transition. The first principal movement (PC1) of the ligand-bound and apo-DHPS are shown in **Figure 4.8-C, D, E, F, G, H, I, J**. The length of the cone represents motion magnitude and the direction is shown by the pointing of the arrow. It is found that essential motions of apo and product-bound DHPS are only associated with flexible loop regions (**Figure 4.8-B**) that are crucial for ligand recognition.

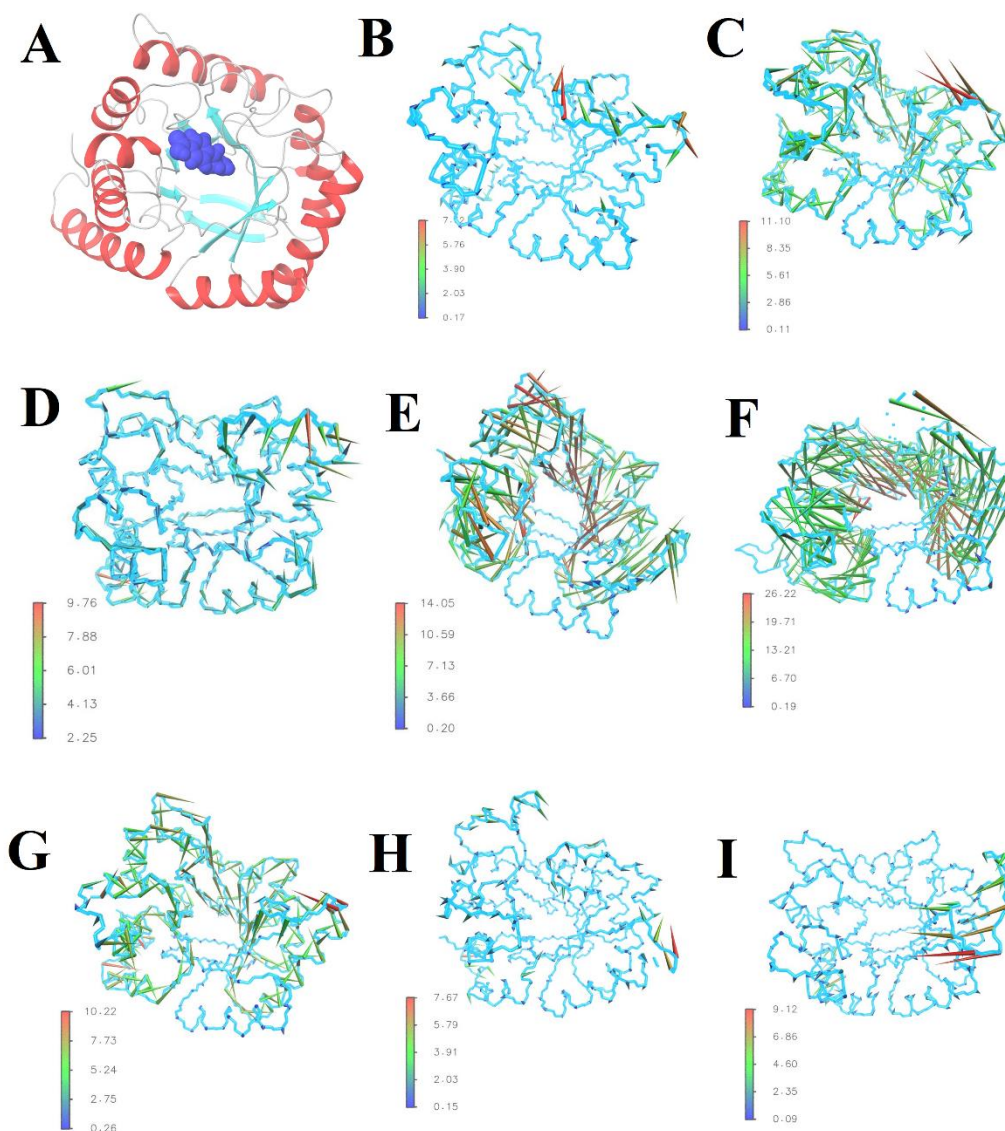


Figure 4.8. (A) The secondary structure of the DHPS. Porcupine plots of the first eigenvector for MD simulations of DHPS bound with (B) Compound 62, (C) Compound 61, (D) Compound 51, (E) Compound S1, (F) Compound S3, (G) Compound 6, (H) Pteric acid and (I) *apo*-DHPS.

However, the entire DHPS structure is stable in both cases. In contrast, the flexible loops of inhibitor bound DHPS (except compound 62) showed closing motion which induces enhanced movements in the rigid domains of the protein backbone. Especially, the amplitude of the motions in the rigid domain is significant when the DHPS is bound with compound S1 and S3. This observation suggested that the binding of highly active

8MG compounds induces the shrinking of catalytic pockets to restrict natural substrate binding and product release. The essential space of the overall dynamics of inhibitor bound DHPS and *apo*-DHPS can be visualized by the 2D-projection of the first two principal components such as PC1 and PC2. It is evident from the projection plot that the extent of sampling is remarkably different in each case (**Appendix XXX**). DHPS bound with compound S1 and S3 is found to span larger conformational space compare to other inhibitor bound states, indicating highly systematic movement of flexible loops located at the surface of the DHPS catalytic pocket. This can be related to the porcupine plot shown in **Figure 4.8-F, G**. In contrast, DHPS bound with highly active Compound 62 and 61 have confined conformational space which implies that intrinsic flexibility of DHPS reduces due to the presence of such compounds. A similar conformational landscape is found for the product-bound state of DHPS. Further, the larger essential space of *apo*-DHPS provides a hint regarding the recovery of the flexibility of protein backbone, gained due to the removal of the ligand from the catalytic pocket. Similar phenomena were found for the inactive ligand-bound state of DHPS as a result of non-specific inhibitor binding. Lastly, the cosine content of PC1 is measured to confirm the absolute convergence of our MD trajectory (Papaleo et al. 2009). The cosine content corresponds to PC1 from all MD trajectories is determined to be under 0.5 indicating adequate sampling of MD trajectories.

4.3.6 MM/PBSA binding free energy

The binding energies of highly active 8MG derivatives such as Compound **62**, **61**, **51**, and top-scored compounds **S1** & **S3** obtained from pharmacophore-based virtual screening are computed by MM/PBSA(Kumari et al. 2014a) algorithm and the results are given in **Table 4.7**. It is found that compound **51** has comparatively lower binding energy (-67.26 kJ/mol) among the highly active 8MG analogs. The free energy decomposition shown in **Table 4.7**, indicates that the electrostatic energy (-154.25 kJ/mol) is the principal contributor to the better stabilization of compound **51** at the DHPS catalytic pocket. The compound **62** is found to have a stronger van der Waals component (-179.56 kJ/mol) due to the presence of hydrophobic benzene ring. It can be observed that the Compound-S3 obtained from database screening has the lowest binding energy of -71.29 kJ/mol due to the most stable electrostatic contribution of -

260.65 kJ/mol. Further, the weak binding affinity of inactive Compound **6** is reflected by its comparatively higher binding energy of -28.19 kJ/mol. However, the binding affinity of the product i.e., 7,8-dihydropteroic acid is lower compare to the 8MG compounds which indicates better competitive inhibition of DHPS by such product analogues. Notably, in all the cases the polar solvation term is found to be unfavourable in the context of stabilizing the protein-inhibitor complex. The binding free energies of the compounds correspond to

Table 4.7. Binding Free energy (kJ/mol) of the 8MG and pterioic acid at the DHPS binding pocket by MM/PBSA method (Simulation batch _1)

Compounds	$\Delta G_{\text{bind-eff}}$ (kJ/mol)	ΔG_{vdw} (kJ/mol)	ΔG_{ele} (kJ/mol)	$\Delta G_{\text{sol-pol}}$ (kJ/mol)	ΔG_{SASA} (kJ/mol)
Compound 62	-50.94	-179.56	-130.57	275.83	-16.64
Compound 61	-60.19	-149.86	-146.22	250.65	-14.33
Compound 51	-67.26	-139.05	-154.25	243.10	-17.06
Compound 6	-28.19	-134.68	-136.83	257.65	-14.33
Compound_S1	-62.03	-134.03	-186.48	273.21	-14.73
Compound_S3	-71.29	-154.54	-260.65	362.75	-18.85
Product	-43.09	-181.88	-134.56	288.30	-14.95

$\Delta G_{\text{bind-eff}}$: The enthalpic contribution of binding free energy (without entropic term), ΔG_{vdw} : The van der walls energy component of binding free energy, ΔG_{ele} : The electrostatic energy component of binding free energy, $\Delta G_{\text{sol-pol}}$: Polar Solvation Energy, ΔG_{SASA} : The free energy corresponding to the solvent-accessible surface area.

simulation_batch 2 and simulation_batch 3 are given in **Appendix XXXI**, **Appendix XXXII** respectively. The trend is found to be like simulation_batch 1. Furthermore, the total binding energies in each protein-ligand complex are decomposed into the contribution made by each DHPS residue and shown in **Figure 4.9**. This enables a comparison between the relative contribution of the interacting residues towards the overall binding affinity.

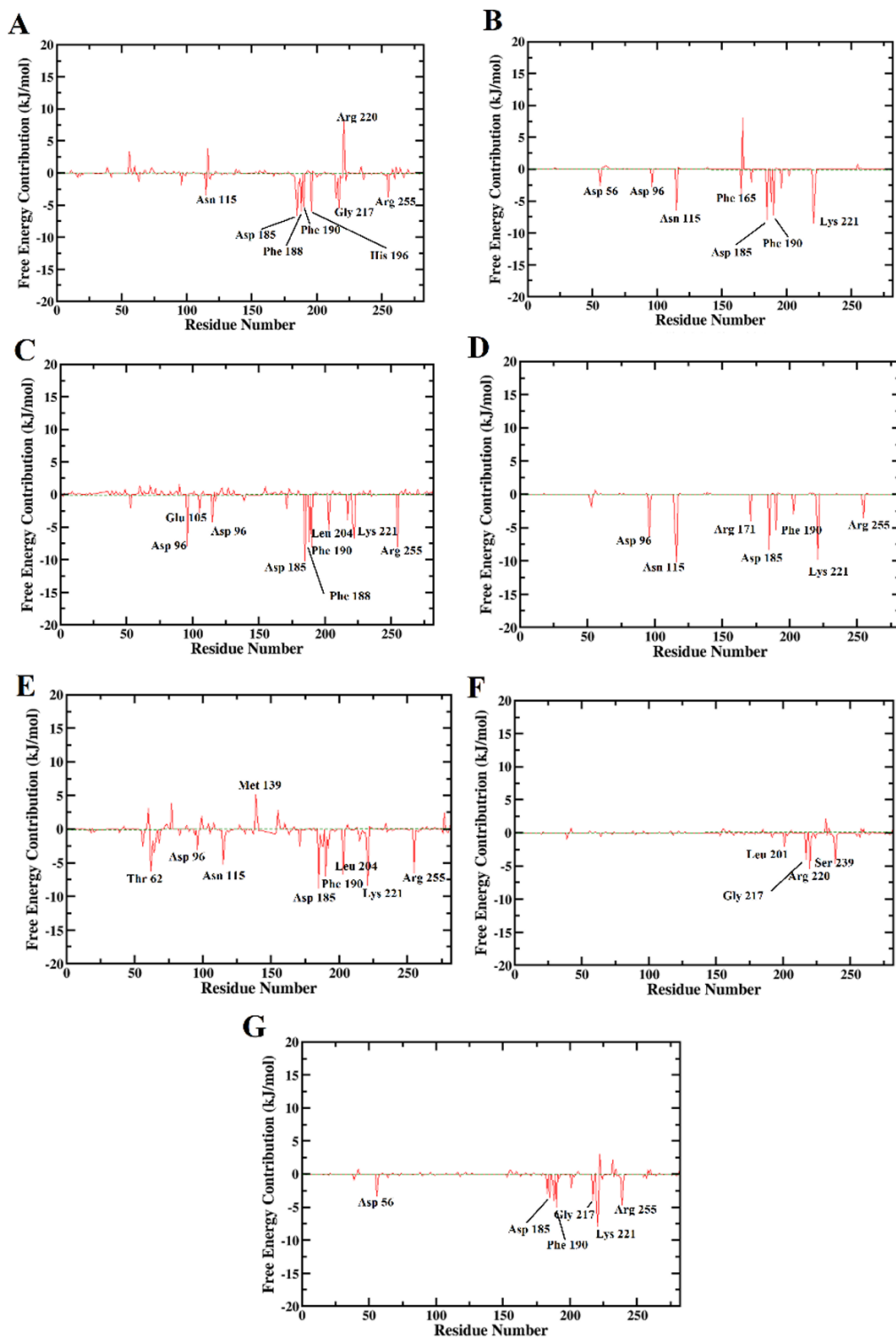


Figure 4.9. The per-residue decomposition of MM/PBSA free energy for (A) Compound 62, (B) Compound 61, (C) Compound 51, (D) Compound S1, (E) Compound S3, (F) Compound 6, (G) Pteric acid.

The residue-inhibitor interaction spectrum is found to be almost similar for the compounds which showed lower binding energy towards DHPS. It is found that interaction of Glu 105, Asn 115, Phe 188, Phe 190, Lys 221 and Arg 255 is common to the highly active 8-MG derivatives considered for free energy calculation. The amino acid region Thr53-Thr62 also stabilizes Compound **61**, **51**, and **S3** at the catalytic pocket of DHPS. In the case of compound 6, the non-specific residues such as Leu 201, Gly 217, Arg 220, and Ser 239 have contributed towards the binding free energy. It can be observed that amino acid residues with hydrophilic side chains have more contribution to stabilizing the inhibitors. However, compound 62 is mainly stabilized by hydrophobic residues.

4.3.7 Effect of Mutation

In order to verify the mutation-resistant activity of the 8-MG derivatives, single point mutations were introduced at residue number 64 (Pro64Ser) and 221 (Lys221Gln) of the DHPS backbone. These two mutations are reported to destabilize both the pterin, sulfa-based inhibitors against DHPS, and the strongest drug-resistant mutation ever found for DHPS (Yun et al., 2012). The effect of the mutation was tested on highly active Compound 51 and S3 and compared with the wild-type Protein-ligand complex. The RMSD of compound 51 bound DHPS backbone is found to be stabilized with an average value of 0.25 nm while the RMSD of compound 51 is stabilized with an average value of 0.15 (**Figure 4.10 A**). It can be seen that the single point mutation did not cause a significant change in ligand conformation. There are five hydrogen bonds are found between mutated DHPS and compound 51 throughout the simulation. Similarly, in the case of compound S3, the backbone RMSD is found to be stabilized at 0.3 nm for the last 30 ns of the simulation. The compound S3 is found to be stabilized with a higher RMSD value of 0.23 nm for the last 10 ns of the simulation (**Figure 4.10 B**). The total number of hydrogen bonds between compound S3 and mutated DHPS is found to be 4. Further, the RMSF values for the amino acid residues are slightly increased in the case of mutated DHPS. However, the change in RMSF is not much significant to destabilize the highly active 8-MG compounds from DHPS active site (**Figure 4.10 C & 4.10 D**).

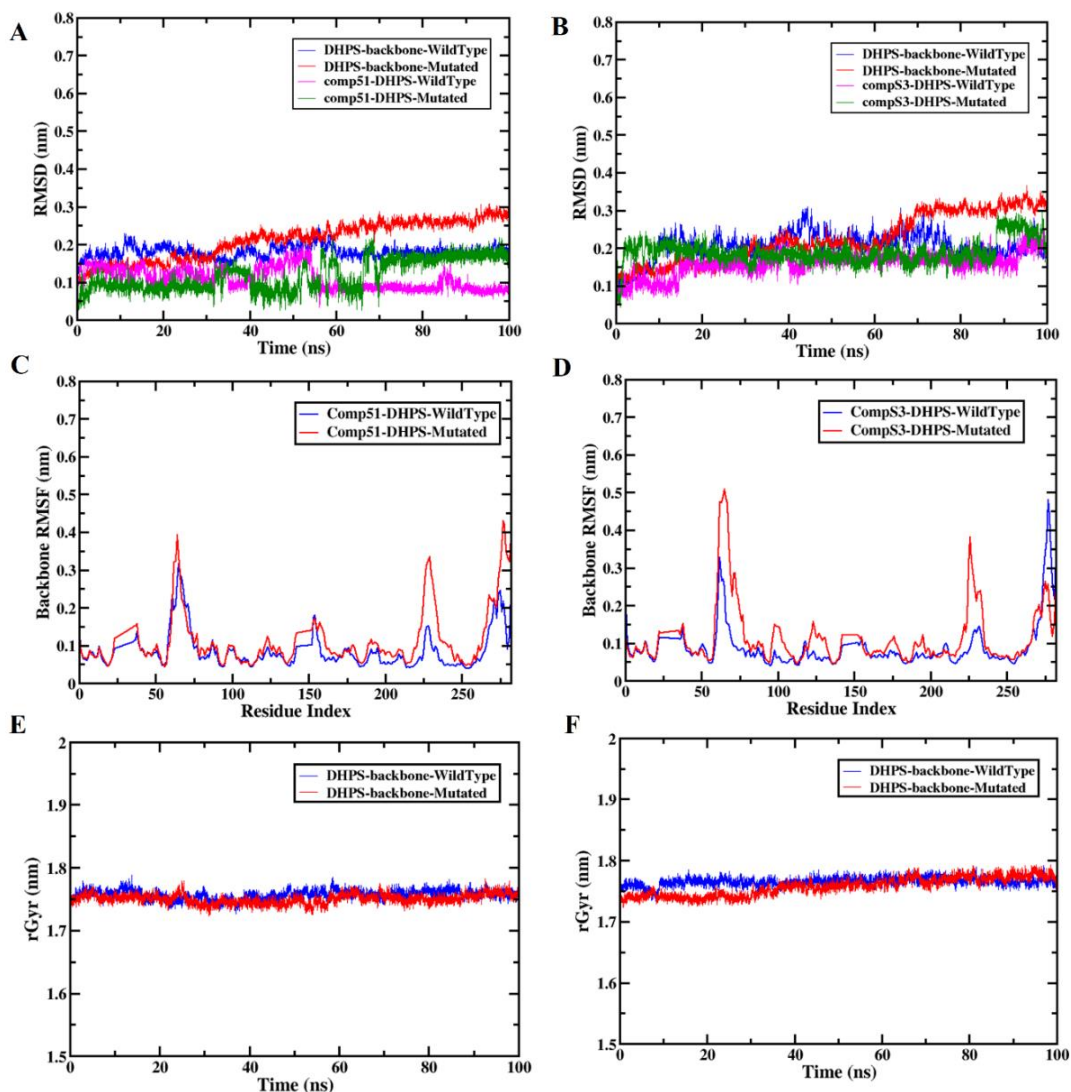


Figure 4.10. The RMSD profile for DHPS bound with (A) compound 51, (B) Compound S3. The RMSF profile for DHPS backbone bound with (C) Compound 51, (D) compound S3. The gyration radius profile for DHPS backbone bound with (E) Compound 51 and (F) Compound S3.

It is evident from the gyration radius profile that the point mutations are not responsible for changing the overall shape of DHPS (**Figure 4.10 E & 4.10 F**). Lastly, we calculated MM/PBSA calculation to assess the binding affinity of the compounds at the mutated catalytic cavity of DHPS. The binding energy of compound 51 and S3 is found to be -52.38 kJ/mol and -69.49 kJ/mol (**Table 4.8**) which indicates that 8-MG compounds have a promising ability to circumvent DHPS resistant mutations.

Table 4.8. Binding Free energy (kJ/mol) of the Compound 51 and S3 at the mutated DHPS binding pocket by MM/PBSA method (Sim-III)

Compounds	$\Delta G_{\text{bind-eff}}$ (kJ/mol)	ΔG_{vdw} (kJ/mol)	ΔG_{ele} (kJ/mol)	$\Delta G_{\text{sol-pol}}$ (kJ/mol)	ΔG_{SASA} (kJ/mol)
Compound 51	-52.38	-133.26	-140.84	238.19	-16.472
Compound_S3	-69.49	-165.88	-235.65	352.754	-20.71

4.3.8 Drug unbinding by Umbrella Sampling

In order to study the dissociation pathway of the 8MG-derivatives from the DHPS catalytic pocket and to determine their free energy of inhibitor dissociation, umbrella simulation studies have been employed. Due to the high computational cost, we have selected highly active compound 51, compound S1, and inactive compound 6 to study the dissociation process. The reaction coordinate is extended from 0 to ~ 5 nm to ensure complete dissociation of the ligand from the binding pocket. It can be observed from that substrate entry and product release can occur through narrow, solvent-exposed, and metastable unstructured loops which acts as the gateway to enter a comparatively stable pterin binding pocket of DHPS. Consequently, there is a major dissociation energy barrier to be crossed by product-analogs to escape from narrow catalytic pocket. Therefore, an external biasing potential has been applied to steer the ligand towards the bulk solvent. The structure of the DHPS binding pocket is restrained by applying harmonic potential on the C_{α} atoms of binding pocket residues. Application of pulling force allows the perturbation of dynamic equilibrium of the system, thus hindering the calculation of thermodynamic quantities directly from the steered molecular dynamics (SMD) trajectories with minimum errors. In such a scenario, the weighted histogram analysis method (WHAM) is commonly used to compute equilibrium data from nonequilibrium SMD trajectories.

The force acts on the ligand are found to be increasing sharply due to constant velocity pulling until a breaking point is reached (**Appendix XXXIII**). At this point, the critical interactions are disrupted which allows the inhibitors to dissociate from the catalytic pocket of DHPS.

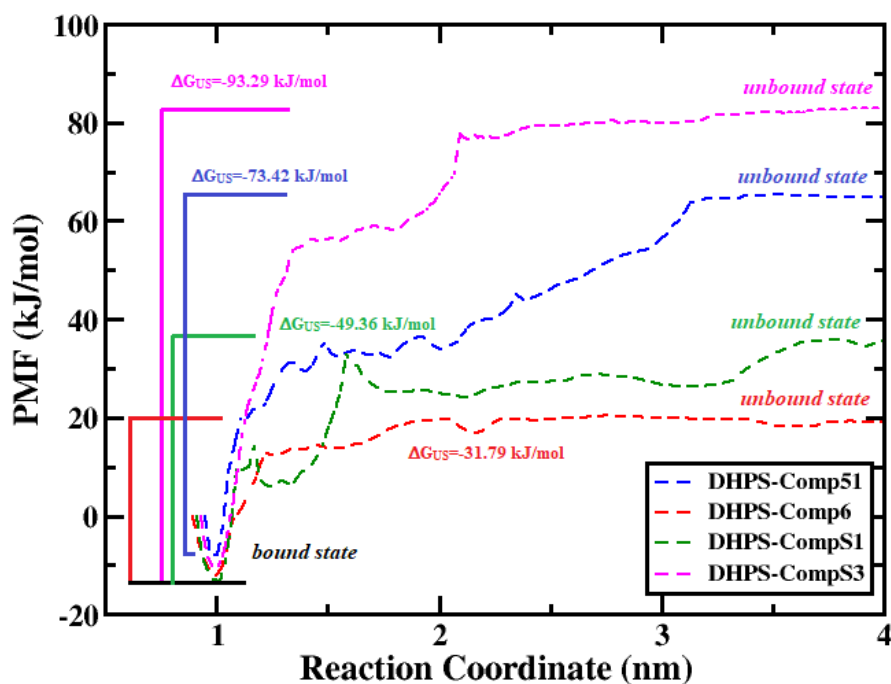


Figure 4.11. The Potential of Mean Force (PMF) profile of inhibitor unbinding from the catalytic pocket of DHPS.

The events leading up to the uncoupling of each ligand considered here are found to be differed significantly, although the direction of pulling is the same in each case. Therefore, the force vs time curves, computed from COM pulling is not essentially comparable due to the path dependency of the dissociation process. In the case of compound 51, the hydrogen bonds with Asp 96, Asn 115, Asp 185, and water-mediated hydrogen bonds with Cys 137, Phe 188, Gln 226 are found to be broken at the point of maximum force. Similarly, the hydrogen bond between Asn 115, Asp 185, and compound S1 are disrupted at the breaking point. It is evident from the above discussion that Asp 115 and Asp 185 are critical to the stability of the inhibitors inside the DHPS catalytic pocket. However, compound 6 is found to be formed only water-mediated hydrogen bonds with non-specific Arg 233 and Arg 235 during the dissociation process. Further, the free energy profile along the reaction coordinate (RC) is analyzed and the PMF profile for each studied system considered for umbrella sampling is shown in **Figure 4.11**. It can be seen from the graph that the PMF profile for the inhibitor dissociation follows a similar trend. The PMFs are found to drop to a minimum value and then increase to a stable state when RC reaches to ~3 nm- 4 nm. The flat region of

the PMF indicates that the inhibitor is in isotropic bulk or the interaction between the protein and ligand is completely disrupted. Free energy of dissociation is determined from the difference between the highest and smallest values of the PMF curve. It is evident from **Figure 4.11** that compound S3 has the lowest binding energy of -93.29 kJ/mol indicating the highest energy barrier to dissociate from the DHPS catalytic pocket. Further, compound 51 is found to show lower binding energy of -73.42 kJ/mol, while compound S1 has a comparatively lesser binding affinity of -49.36 kJ/mol. This result indicates that the stabilization of 8-MG compounds containing sulfonamide or sulfonyl fragments, at the flexible pABA binding pocket enhances the barrier for inhibitor dissociation from DHPS catalytic pocket. In the case of compound 6, the dissociation energy barrier is found to be less due to the highest binding energy of -31.79 kJ/mol. It can be noted here that the binding free energy (i.e dissociation energy) computed by umbrella sampling is significantly smaller compare to binding energy estimated by MM/PBSA or MM/GBSA method(Sun et al. 2015). This is due to the ignorance of conformational entropies and dependencies of polar solvation equation schemes (Genheden and Ryde 2015) in binding free energy calculation. However, the trend of the binding free energy of the 8MG derivatives is found to be correlated in both US and MM/PBSA methods.

4.4 CONCLUSION

It is extensively recognized that drug discovery with respect to DHPS, in general, is a time taking, expensive, and highly complex process. *In-silico* studies have become an indispensable tool for probing unknown chemical space to identify novel bioactive molecules with improve effectivity. In this article, *in-silico* techniques such as pharmacophore mapping, atom-based 3D-QSAR modeling, high throughput virtual screening, molecular docking, MD simulation with a pair of free energy techniques, namely MM/PBSA & umbrella sampling were employed to identify crucial counterparts of DHPS competitive inhibition with 8-MG derivatives. The contributions of various atomic features, like hydrogen bond donor, hydrophobic interaction, and electron-withdrawing groups on the 8MG analogs, towards DHPS inhibition were highlighted. 8 MG- compounds containing phenylacetamide (Compound S1) and phenyl sulfonyl (Compound S3) were identified as the highest-ranked hits among the

organic molecules obtained from database screening. Docking study revealed that hydrogen bonding, π - π stacking, and π -cation interactions of product analogs with active site residues provide the stability of the inhibitors. It is found that most of the non-bonded interactions are present between the guanine ring of the inhibitors and the active site residues located deep inside the active site. The active site residues such as Asp 96, Asn 115, Asp 185, Phe 188, and Ser 222 are mainly involved in hydrogen bonding while Phe 190, Lys 221, and Arg 255 are associated with π - π stacking and π -cation interactions. Further, the DFT study reveals that the overlapping of HOMO-LUMO orbitals on common guanine ring decreases the activity of 8MG analogs. Moreover, the highly active 8-MG compounds were found to have higher water solubility as predicted by ADMET calculations. The dynamic stability of the inhibitors at the DHPS catalytic pocket was confirmed by atomistic MD simulation studies. The active compounds are found to be most stable compared to the inactive one due to their highly specific binding in DHPS catalytic pocket similar to pteronic acid. The binding of highly active compounds such as Compound 51, 62, and S3 was found to reduce the flexibility of residues 62-77 which is crucial for ligand stability. The non-bonded interaction predicted by docking calculations was found to have similarities with the MD simulation studies. PCA studies indicate that binding of the top-scored compounds induces closing motion in the flexible loops thereby causes prolonged inhibitions of DHPS. Further, MM/PBSA free energy studies indicate that electrostatic energy is the principal contributor to the stability of the inhibitors. Lastly, the free energy of dissociation of the product analogs reveals that water-mediated hydrogen bonding is also important for the stability of the inhibitors. The overall study summarizes that the substitution of the sulphonamide class of fragments at the central Sulphur atom of 8-MG compounds is crucial to enhance the activity and stabilize them at the flexible *p*ABA binding loops. The information obtained from this study offers a basis for understanding the conformational dynamics of DHPS with highly active 8-MG analogs and inspire medicinal chemist to synthesize newer generation of antimicrobials.

CHAPTER 5

EPITOPE-BASED POTENTIAL VACCINE CANDIDATE FOR HUMORAL AND CELL MEDIATED IMMUNITY TO COMBAT SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS 2 PANDEMIC

Abstract: *The emergence of severe acute respiratory syndrome from Sever Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has put an immense pressure worldwide where vaccination is believed to be an efficient way for developing hard immunity. Herein, we employ immunoinformatic tools to identify B-cell, T-cell epitopes associated with Spike protein of SARS-CoV-2 which is important for genome release. The results showed that the highly immunogenic epitopes located at stalk part are mostly conserved compared to the receptor binding domain (RDB). Further, two vaccine candidates were computationally modelled from the linear B-cell, T-cell epitopes. Molecular Docking reveals the crucial interactions of the vaccines with immune-receptors and their stability is assessed by MD simulation studies. The chimeric vaccines showed remarkable binding affinity towards the immune cell receptors computed by MM/PBSA method. Van der Waals and electrostatic interactions are found to be the dominant factors for the stability of the complexes. The molecular-level interaction obtained from this study may provide deeper insight in the process of vaccine development against the pandemic of COVID-19.*

Keywords: *Severe-acute-respiratory-syndrome; β -corona virus; immunoinformatic; epitope-vaccine; MD simulation; MM/PBSA*

5.1 BACKGROUND

The sudden outbreak of febrile respiratory syndrome caused by a novel β -coronavirus (2019-nCoV) has created a global catastrophe where death rate is increasing every day (Lam et al. 2020 p. 201; Sanche et al. 2020; Zheng 2020). The rapid propagation of novel-coronavirus or SARS-CoV-2 outcompetes the past epidemics caused by Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), Middle East Respiratory Syndrome Coronavirus (MERS-CoV) (de Wit et al. 2016) and therefore is declared as a global pandemic by WHO. As of now, more than 29 million confirmed cases and 924 thousand deaths are caused by SARS-CoV-2.

Many efforts have been made for the discovery of antiviral drugs against SARS-CoV-2 (Ghosh et al. 2020a; b; Gutierrez-Villagomez et al. 2020; Mishra et al. 2020), but no such licensed therapeutic are available in the market till date. Therefore, the development of an effective treatment strategy for the pandemic is a research priority. Moreover, the design of a novel vaccine against viruses using kits and related antibodies is time-consuming and expensive (Chen et al. 2020; Tahir ul Qamar et al. 2018). Previously, numerous approaches including the whole virus, viral-DNA, subunit, and virus-like particles have been employed in developing epitope based vaccines against SARS and MERS virus (Liu et al. 2017; Prompetchara et al. 2020; Schindewolf and Menachery 2019; Shi et al. 2015b; Song et al. 2019; Yong et al. 2019). These epitopes can be prepared by chemical synthesis techniques and are easier in quality control (Alderete and Neace 2013; Geysen et al. 1984). There are evidences which supports that *In silico* predictions are helpful in successful production of commercially important vaccines (Zawawi et al. 2020). However, the structural modifications, delivery systems, and adjuvants are the additional requirements in the formulation due to low immunogenicity caused by their structural complexity and low molecular weight (Azmi et al. 2014). Recently, a set of B and T cell epitopes from the highly conserved region in SARS-CoV-2 were identified which may help in developing vaccine candidates (Kar et al. 2020; Lin et al. 2020; Lizbeth et al. 2020). However, very little is known about the dynamic stability and affinity of the predicted epitope towards the interacting domain of Antibody and T-cell Receptors (TCRs) which is crucial for validating and improving the efficacy of predicted vaccines. In this respect, we apply a

combination of immuno-informatic(Nelapati et al. 2020; Walls et al. 2020) approach to identify potent epitopes to design the vaccine candidates followed by computational chemistry analysis to check their effectiveness. With the help of molecular docking, MD simulations and free energy calculations an analysis of all the important interactions necessary to give stability to the immuno-receptor complexes have been performed.

T-cells are known to recognize and activate defence responses against viral infection; B-cells on the other hand can have antibody reactions which help in recovering extreme respiratory infection. Therefore, we have done detail analysis of the viral antigens to predict B-cell, T-cell linear epitopes located at the S-protein of SARS-CoV-2, evaluated their immunogenicity, and designed chimeric vaccines. The conservation of all B- and T- cell epitopes were assessed across most of the isolates and coronavirus species from different parts of the globe. Furthermore, we carried out *in-silico* cloning of the linear vaccine construct to design a recombinant plasmid that can help in expressing the vaccines in *E.coli* expression system.

5.2 COMPUTATIONAL DETAILS

5.1 Screening of Potential epitopes

To primarily identify the linear B-cell epitopes from the amino acid sequence of SARS Cov2 spike protein, a collection of methods such as (i) Emini Surface analysis, (ii) Parker Hydrophilicity (iii) Karplus Schulz Flexibility, (iv) Kolaskar & Tongaonkar Antigenicity and (v) Chou Fasman Beta-Turn has been adopted which predict linear epitopes by correlating different antigenic parameters such as hydrophilicity, surface exposure, flexibility and β -turn propensity etc with the location of continuous amino acid stretch. In general, all the methods compute antigenic score of a residue i within a window size of n ($n= 5$ to 7) and $i-(n-1)/2$ residues at each side of residue i . The highest score residue along with the amino acid stretch then consider as potential linear antigen. Nevertheless, the above-mentioned methods are different based on the amino acid scales and the method implemented to predict antigen. The detailed description of each method is given below.

Emini Surface Accessibility: In this method the surface exposure of an amino acid sequence can be determined by using the formulae $S_n = [\prod_{i=1}^6 \delta_{n+4-i}] * (0.37)^{-6}$ where δ is the fractional surface probability of one residue in a hexapeptide sequence (Emini et al. 1985a)

Parker Hydrophilicity: This method determines the hydrophilic scale based on the retention time of the amino acid residue. Further, a window size of seven residues is considered and the fourth atom of the peptide window is scored by calculating the arithmetic mean of the hydrophilicity scale introduced to each seven residues (Parker et al. 1986a)

Karplus Schulz Flexibility: In this method the flexibility scale of the amino acid residues is determined based on experimental B-factor of the alpha carbon. The calculation of antigenic scores is done by considering the first amino as the centre of six residue window length. Additionally, there are three flexibility scale instead of one (Karplus and Schulz 1985a).

Kolaskar & Tongaonkar Antigenicity: This a semi empirical method which determine antigenicity of a polypeptide sequence based on the physicochemical properties of amino acids and their frequencies of occurrence in an experimentally known epitope segment (Kolaskar and Tongaonkar 1990b).

Chou Fasman Beta-Turn: This method determines the possibility of a polypeptide sequence (turn scale) to adopt a β -turn based on Pellequer et al (Pellequer et al. 1993). In this method Chou-Fasman scale is used to predict the β -turn (Chou and Fasman 1978).

It is evident from the above-mentioned discussion that antigenic scores differ for each method due to the difference in amino acids scales correspond to each method and working equations. Notably, the antigenic scores are first normalized before making the amino acid scales in each method and later the scores of each residue are averaged within the given window length to predict the antigenicity of hotspot residues.

5.2 MD simulation Protocol

All protein molecules are protonated at the biological pH of 7.0. Then the chimeric vaccine and peptide epitopes are docked at the antigen binding domain of respective immune-receptors by Hex software. We carried out atomistic MD simulations of all the systems using GROMACS-2016.5(Abraham et al. 2015a; Berendsen et al. 1995; Spoel et al. 2005). Amber 99-SB (Schmid et al. 2011a) force-field was employed due to its better balance in β -sheet and helicity. propensity compared to other force fields (Best et al. 2008a; Chamachi and Chakrabarty 2016; Man et al. 2019a). SPC/E(“Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K - The Journal of Physical Chemistry A (ACS Publications)” n.d.) water molecules were used to solvate the receptor-peptide systems due to its compatibility with AMBER force field. All the systems underwent a 50000 step energy minimization by steepest-descent algorithm(Averill and Painter 1992; Meza 2010) to remove the steric clash. Leapfrog integrator algorithm was used to integrate the equation of motions with a time step of 2 fs. LINCS(Hess et al. 1997a) algorithm was applied to constrain all the bonds in the peptide molecule and SETTLE algorithm was employed to constrain the geometry of water molecules. The systems were equilibrated in canonical ensemble (NVT) followed by isothermal-isobaric ensemble (NPT) for 5 and 10 ns respectively by restraining the solute heavy atoms. Next, the restrain were removed and the protein molecules were allowed to move freely during the production run for 100 ns. The temperature and pressure of the system was maintained employing Velocity rescale (Bussi et al. 2007a) ($\tau_t=0.1$ ps) and Parrinello-Rahman coupling algorithm(Parrinello and Rahman 1981a) ($\tau_p=0.2$ ps). The cut-off for short range electrostatic and van der Waals interactions was assigned to 1.2 nm. Particle Mesh Ewald method(Darden et al. 1993; Essmann et al. 1995b).

5.3 RESULTS AND DISCUSSION

5.3.1 Screening of Epitopes

The amino-acid sequences of the spike-proteins for twenty-six different coronavirus species (SI-1) were considered for multiple sequence analysis (MSA) to figure out the conserved amino-acid region and the variable region which differentiate the SARS-CoV-2 from other classes of coronavirus. Multiple sequence analysis (MSA) reveals

that the C-terminal end is highly conserved compared to the N-terminal end of the amino acid sequence of spike protein retrieved from different coronavirus species. This C-terminal perfusion or shaft domain (S2-domian) is thus common in all coronavirus spices for genome transfer while the receptor binding domain (RBD or S1-domain) is unique in SARS-CoVid2 (**Figure 5.1-A**). The Phylogenic tree constructed from MSA indicates that the spike protein is of β -coronavirus species (SARS CoVid, MERS-CoVid) (**Appendix XXXIV**). The size of the predicted linear B-cell epitopes is found to be varied between 7mer- 15mer which is believed to be of optimum size. The hot-spot residues which are considered as the key to the antigenicity of the epitopes, are found to be has maximum structural similarity with γ -coronavirus. The antibody mediated defence responses or Humoral immunity against viral infection is believed to be guided by B-cell epitopes(Charles A Janeway et al. 2001). The linear B-cell epitopes of S-glycoprotein sequence were predicted by a set-of physicochemical parameters such as the exposed surface propensity by Emini et. al(Emini et al. 1985b), hydrophilicity by Parker et. al(Parker et al. 1986b), flexibility by Karplus et. al(Karplus and Schulz 1985b), antigenic propensity by Kolaskar et. al(Kolaskar and Tongaonkar 1990a) and β -

Table 5.1. The highly immunogenic Liner B-cell and T-cell epitope from Spike glycoprotein of SARS-CoVid2

Linear B-cell Epitope						
Methods	Peptides	Start_ Position	End_ Position	Hotspot Residues	Score	Length
Emini Surface Accessibility	DPSKPSKRS F	808	817	P, S	5.67	10
Parker Hydrophilicity	DSTECNS	745	751	E	6.34	7
	GTNTSNQ	601	607	T	6.09	7
Karplus & Schulz Flexibility	EGKQGNF	180	186	K, Q	1.1	7
	PGQTGKI	412	418	T	1.1	7
BepiPred Linear Epitope	ILPDPSKPS KRS	805	816	P, S, K, P, S	1.67, 2.06, 2.29, 2.16,	12
	QTQTNSPRR RARSV	675	687	T, N, S	1.51, 1.69, 1.54	13
Kolaskar & Tongaonkar Antigenicity	FLVLLPLVS SQCIVN	4	18	L, L, P, L, V	1.243, 1.239, 1.22, 1.261, 1.227	15

	PHGVVFLH VTYVPA	1057	1070	F, L	1.21, 1.21	14
Chou-Fasman Beta-Turn	PGTNTSN	600	606	N	1.36	7
	KGCCSCG	1245	1251	C	1.3	7
T-cell Epitopes						
MHC- Class	HLA-type	Peptide	Start_ Position	End_ Position	Percentil e Rank	Length
MHC-I	A*02 :01	VDIADTTDA V	567	576	0.1	10
	A*24:02	TKRFDNPVL PF	76	86	0.1	11
	B*40:01	VRFPNITNL	327	335	0.1	8
	B*58:01	VFAQVKQIY	781	789	0.1	9
MHC-II	DRB1*04:01	NTLVKQLSS NFGA	960	972	0.47	13
	DRB1*07:01	SLIVNNAT NVVIK	116	129	0.02	14

turn propensity by Chou et. al(Chou and Fasman 1977) etc. implemented in immune epitope database and analysis resource (IEDB) tool(Peters et al. 2005). The predicted B-cell and T-cell epitopes with high antigenicity scores are shown in **Table 5.1** and **Appendix XXXV**. mostly polar in nature which may contribute in stable hydrogen bonds to the residues present in the antigen binding region of the antibodies. In **Figure 5.1 B** we showed the location of linear B-cell epitopes on the single chain of SARS-CoV-2 spike protein. It is evident from **Figure 5.1 B** that most of the epitopes are located at the RDB domain and the shaft region of the S-protein. The peptide ⁸⁰⁸DPSKPSKRSF⁸¹⁷ and ⁶⁰⁰PGTNTSNQ⁶⁰⁷ are found to be common in Emini-surface accessibility, Bepipred linear epitope and Chou-Fasman methods indicating their greater importance as linear B-cell epitopes. The 3-D structure of the spike protein was generated by homology modelling implemented in Swiss Model(Waterhouse et al. 2018) using PDB ID : 6VIB and 6VXX as template. The conformational epitopes of the spike protein of SARS-CoV-2 in their closed and open state were determined by Disctope(Haste Andersen et al. 2006; Kringelum et al. 2012) (V-1.1) server. It can be seen that the conformational B-cell epitopes of spike protein are located in almost in the same position (RDB region) in their both closed and open state (**Appendix XXXVI**). The locations of the conformational epitopes on the tertiary structure of S-protein are mainly found at 405-427 and 439-505 residue stretches and in shown in **Figure 5.1 C, D**. Based on the location of the epitopes, the domains are marked as R1,

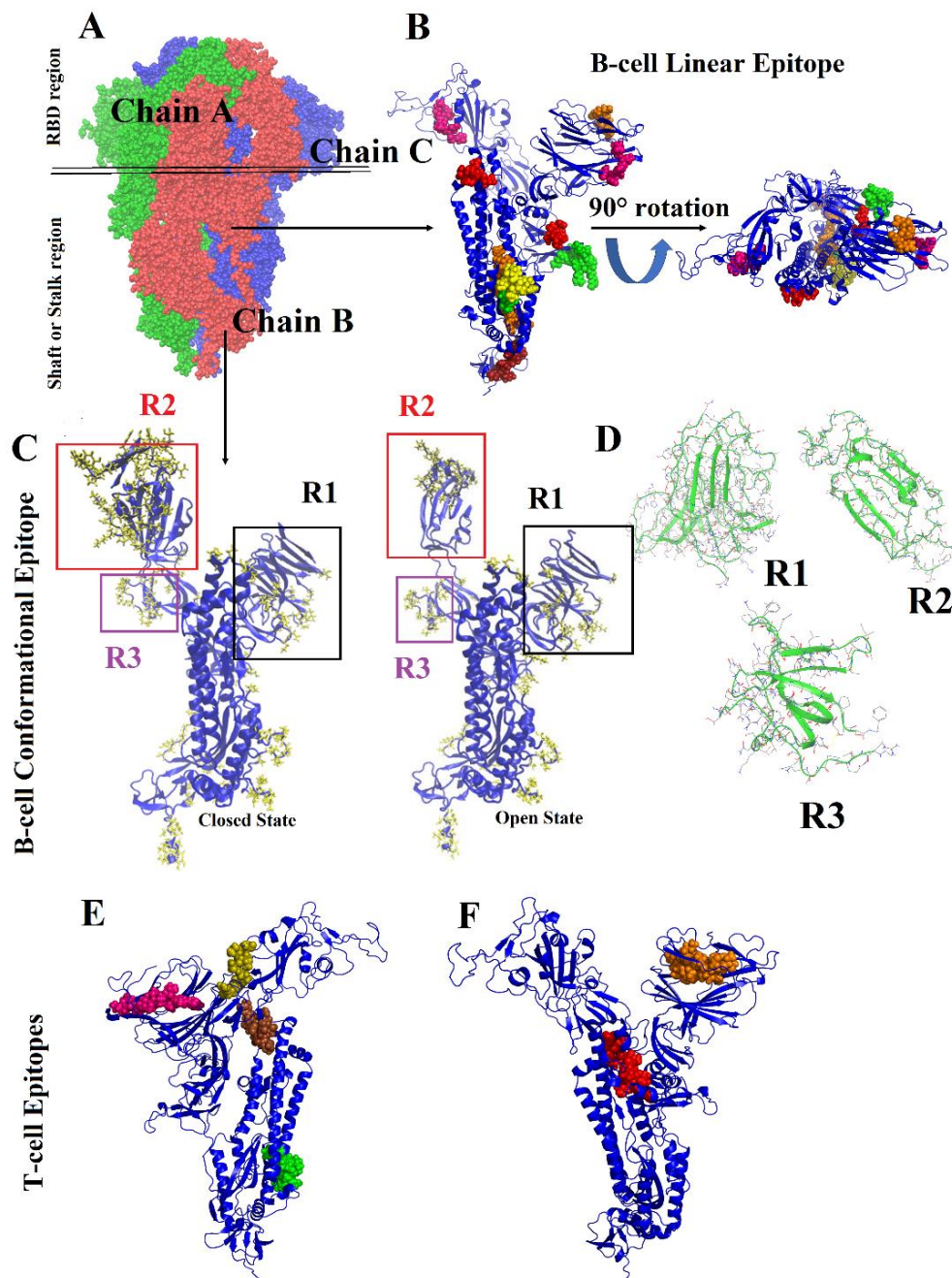


Figure 5.1. (A) The structure of spike glycoprotein in trimeric form. (B) The location of the linear B-cell epitope on the chain B of spike glycoprotein. The yellow, red, pink-, green-, orange- and firebrick-colored regions are predicted by Emini surface accessibility, Parker Hydrophilicity, Karplus & Schulz flexibility, Kolaskar & Tongaonkar Antigenicity and Chou Fasman method respectively. (C) The location of the conformational B-cell epitopes on the chain B of spike glycoprotein in their open state and closed state. The residues of conformational B-cell epitope are indicating in

yellow color. (D) The R1, R2 and R3 are indicating the three different conformational epitope regions in the single chain of spike protein. The location of T-cell peptide epitopes corresponds to (E) MHC-I and (F) MHC-II. The brown, pink, olive, green, red and orange color corresponds to the peptide associated with A*02 :01, A*24:02, B*40:01, B*58:01, DRB1*04:01 and DRB1*07:01 respectively.

R2 and R3. It can be noted that in both closed and open state, all antibody binding regions (R1, R2 and R3) have extended anti-parallel β -sheet or β -barrel structure which is an important factor for antigenicity (**Figure 5.1 D**). The identification of CD8+ and CD4+ T-cell epitopes are known to be important for eliciting cell-mediated immunity or generation of memory B-cell against viral infections (Punt et al. 2019). Such peptide epitopes are generally presented by the Major Histocompatibility complex (MHC) - class I & MHC-class II molecules expressed on the surface of helper T-cell (T_H) (Agudelo and Patarroyo 2010; Punt et al. 2019). We employed IEDB server to identify and validate the T-cell peptide epitopes of the S-protein of SARS-CoV-2 which can bind with MHC-I and MHC-II receptors. All MHC-I and MHC-II based epitope-peptides obtained from IEDB T-cell epitope prediction tools.

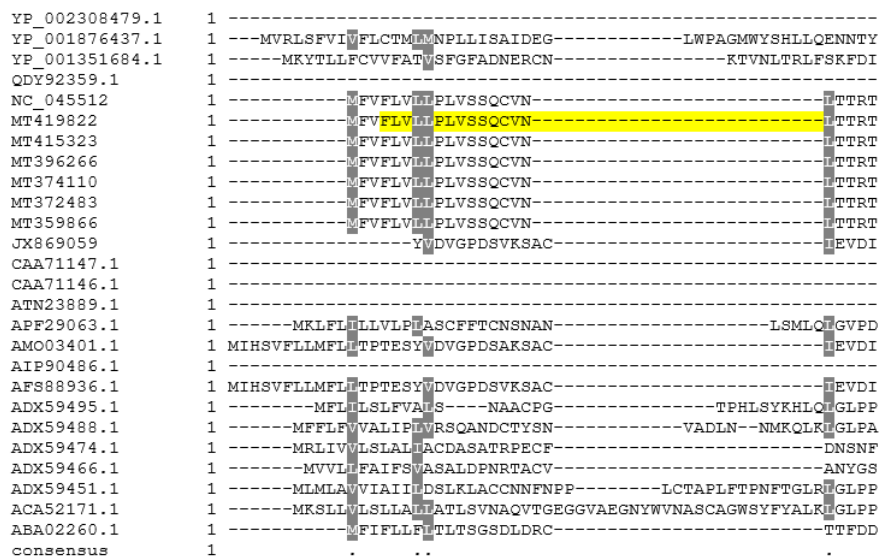
The T-cell epitope peptides with highest antigenicity and high affinity are listed in **Table 5.1** and their location at the S-protein of SARS-CoV-2 is depicted in **Figure 5.1-E, F**. The size of the peptides for MHC binding are found to vary from 8mer to 15 mer. It is evident from **Figure 5.1** that peptide epitope ⁷⁸¹VFAQVKQIY⁷⁸⁹ corresponding to MHC-I and ⁹⁶⁰NTLVKQLSSNFGA⁹⁷² associated with MHC-II are located at the stem part of the S-protein. Unlikely, the epitope ⁵⁶⁷RDIADTTDAV⁵⁷⁶, ⁷⁶TKRFDNPVLPF⁸⁶, ⁷⁴⁷VRFPNITNL⁷⁵⁴, (related to MHC-I) and ¹¹⁶SLLIVNNATNVVIK¹²⁹ (related to MHC-II) are found to be located at the surface i.e the receptor binding domain (RBD) of S-protein. The location of this peptides shares the binding domain of Angiotensin-converting enzyme 2 (ACE2) receptor which make them more interesting. These selected peptides are used to design multiepitope vaccine for cell mediated immunity.

5.3.2 Design of Multimeric Vaccine candidate

The knowledge regarding the linear-epitope-conservation is necessary to design vaccines capable of inducing adaptive immune response for all coronavirus species. To determine the conservation of selected epitopes multiple sequence alignment has been carried out and shown in **Figure 5.2**. It is found that all the epitopes considered here contain at least two conserved residues. Epitopes with three or more identical or similar residues are considered to be highly conserved. The conservation percentage was calculated by the following formulae

$$\text{Percentage Conservation} = \frac{\text{Number of Conserved or Identical Residues}}{\text{Length of the Epitope}} \quad (5.1)$$

Among all the selected B-cell epitope in **Table 5.1**, ⁴¹²PGQTGKI⁴¹⁸(42%), ⁷⁴⁵DSTECSN⁷⁵¹ (42%), ⁸⁰⁵ILPDPSKPSKRS⁸¹⁶ (41%), ⁸⁰⁸ DPSKPSKRSF ⁸¹⁷ (30%), ¹⁰⁵⁷PHGVVFLHVITYVPA¹⁰⁷⁰ (71%), ¹²⁴⁵KGCCSCG¹²⁵¹(42%) are found to be highly conserved. In case of T-cell epitope, ¹¹⁶SLIVNNATNVVIK¹²⁹ (28%) ⁷⁸¹VFAQVKQIY⁷⁸⁵(44%), and ⁹⁶⁰NTLVKQLSSNFGA⁹⁷² (76%) are highly conserved. It is evident from **Figure 5.1-B** and Figure 5.2 that most of the highly conserved B-cell epitopes are located at the stalk or shaft region of the S-protein except ⁴¹²PGQTGKI⁴¹⁸ and ¹⁰⁵⁷PHGVVFLHVITYVPA¹⁰⁷.



YP_002308479.1 1 -----MQRVVL
 YP_001876437.1 106 FTNRTLGEGLDIDYFFNCTNVTNEFRTNMGYFIQNMVVARCLSLCAVVIIPKGRVFFEK
 YP_001351684.1 98 PS--TYQLYLHRDTNGNSN-AFAYLRICKWPSKKWLQSTSNMDSGRFCLVNK-KIPAAF
 QDY92359.1 1 -----
 NC_045512 70 VSGTNGTKRNDNPVLPFND-----GVYF-----ASTEKS
 MT419822 70 VSGTNGTKRNDNPVLPFND-----GVYF-----ASTEKS
 MT415323 70 VSGTNGTKRNDNPVLPFND-----GVYF-----ASTEKS
 MT396266 70 VSGTNGTKRNDNPVLPFND-----GVYF-----ASTEKS
 MT374110 70 VSGTNGTKRNDNPVLPFND-----GVYF-----ASTEKS
 MT372483 70 VSGTNGTKRNDNPVLPFND-----GVYF-----ASTEKS
 MT359866 70 VSGTNGTKRNDNPVLPFND-----GVYF-----ASTEKS
 JX869059 79 TPQKLFVANISQDVKCFANGFVVRIG-----AAANSTG
 CAA71147.1 1 -----
 CAA71146.1 1 -----
 ATN23889.1 1 -----
 APF29063.1 82 YDVNQYIYVTEHGLNAS---VTLKICKRGINTTFDFLS-NSSSFDCIVNL----LF
 AMO03401.1 96 TPQKLFVANISQDVKCFANGFVVRIG-----AAANSTG
 AIP90486.1 1 -----MQRALL
 AFS88936.1 96 TPQKLFVANISQDVKCFANGFVVRIG-----AAANSTG
 ADX59495.1 88 DVKGQWALYLVRKPNGEK----IFRVCKWSTAKNVSSTV-TVSSADVCMYDR-VFKMNF
 ADX59488.1 92 KSNSSWGMYSCHCKNGDHR---AKIRICKWDRYYPVRYAN-SELNGRDCLVNK-DIEFTF
 ADX59474.1 85 AKDRNGGYINNTTILSVAS----GLFVHTWMYRQDS---ASTKCY
 ADX59466.1 82 A---GQLMNDTSAVAGD-----GIFVNVHMYRRLQGITDSNLKYC
 ADX59451.1 106 QNMDQWAMYVYCKNNDINTGRAMRIKVCWVHGLYTFVYEP-NKPVSQCLINK-QFNFI
 ACA52171.1 108 AGGEQYSAYMAOQNGKML---VLRICKWONGTLAAPTLL-QSTSGKDCIVNVKVDNHF
 ABA02260.1 74 HT-----HGNFVIFPKD-----GIYF-----AATEKS
 consensus 121

YP_002308479.1 7 ISFIHIAASVADRILDFITFPNGHKYLYPHR-----GDLGQLKSRVNS
 YP_001876437.1 166 NEVWWWQDRYQPGNYTFRHERFMSLSQNIETVIGLAGNFSLFGLGQTININSASYDWGF
 YP_001351684.1 154 TDHANMIVGTHWDQDRVTFYT-DKVIHFYVPPN----RWSRVVSWCSAADSCAMQYINS
 QDY92359.1 1 ---MLGLALVLLTTLVLFVPIV-----SLSGDNWVYVY
 NC_045512 99 NIIRGWVFGTTLDSKTQSLIVN-----NATN-----VVVKVCFQFCNDPFLGVVY
 MT419822 99 NIIRGWVFGTTLDSKTQSLIVN-----NATN-----VVVKVCFQFCNDPFLGVVY
 MT415323 99 NIIRGWVFGTTLDSKTQSLIVN-----NATN-----VVVKVCFQFCNDPFLGVVY
 MT396266 99 NIIRGWVFGTTLDSKTQSLIVN-----NATN-----VVVKVCFQFCNDPFLGVVY
 MT374110 99 NIIRGWVFGTTLDSKTQSLIVN-----NATN-----VVVKVCFQFCNDPFLGVVY
 MT372483 99 NIIRGWVFGTTLDSKTQSLIVN-----NATN-----VVVKVCFQFCNDPFLGVVY
 MT359866 99 NIIRGWVFGTTLDSKTQSLIVN-----NATN-----VVVKVCFQFCNDPFLGVVY
 JX869059 112 TVIISPSTSATIRKIYPAFMIIGSSVGNFSDGKM-----GRFFNHTLVLLPDGCGTILR
 CAA71147.1 1 -----
 CAA71146.1 1 -----
 ATN23889.1 1 ---MKLFTVETLNASIRVLYGCES-----VDFNLFNTIFST
 APF29063.1 133 TEQLGAPLCTTISGETVREHTLVNVRTFYVPAAY---KLTKEVSKCYFNYSCVFSVVA
 AMO03401.1 129 TVIISPSTSATIRKIYPAFMIIGSSVGNFSDGKM-----GRFFNHTLVLLPDGCGTILR
 AIP90486.1 7 IMTLFCVRAKFAADDLLDITFFGAHRFLHKPT-----RNSSSLYSRANN
 AFS88936.1 129 TVIISPSTSATIRKIYPAFMIIGSSVGNFSDGKM-----GRFFNHTLVLLPDGCGTILR
 ADX59495.1 141 KDNLSKIFGVTWAGDQVRYVYVENTYSFNPNVNS---WSRATFCEHSYACAMQVSD
 ADX59488.1 146 AHAALVVGATWAGDKVVLYTRSGITSLFVPGAT---LWDVVAISCKNKEACAHQVNS
 ADX59474.1 123 QEPFGVAFGTFIQDRIAEVVAIA-----PDLR-----GSWAQSTYRSVTVVHLVCS
 ADX59466.1 119 QEPFGVAFGTFIQEPIVVAIA-----PGNY-----GRPALQIPIATVTVAVCS
 ADX59451.1 164 RTGVTGTFIHTWSSNYVTYGRDDFYQFYLEHD-----WDNVQVRCRSECFYVARE
 ACA52171.1 162 YHAADVVGMSWGGDAVRYTQTDTKTYIIPNS-----WDRVVSIRCPDKFSCSSQIVTK

ABA02260.1 96 NVIRGWVFGTTLNKNKSQSLIVN-----NSTN-----VVVKRACNFELCDNPFPAVSK
 consensus 181
 YP_002308479.1 52 NYLRDTRITLKHGYLIED-----FHIFSP-----INSNTIPI
 YP_001876437.1 226 VIRGRLLDDRICELOMTGADLCNLTRVDTSSDAFYTAGLVFASAALIADGAKPIPTGVVF
 YP_001351684.1 208 TIYYNLNVTIPGGPGGIYSVCTKHCTGLADNVFTDQGGIIPPIFFYNNVFLITNTSTLV
 QDY92359.1 32 PSAFRPSNGWHKQEGAYSVVGIQSTNVTGNSAGCFYMFNTSOLIASDAVAIMSKFPTTW
 NC_045512 146 HKNNKSWMSEFRVYSS---ANNCTFEVVSQFFLDLEGRQGNFKNLRREIVKKNIDGYFK
 MT419822 146 HKNNKSWMSEFRVYSS---ANNCTFEVVSQFFLDLEGRQGNFKNLRREIVKKNIDGYFK
 MT415323 146 HKNNKSWMSEFRVYSS---ANNCTFEVVSQFFLDLEGRQGNFKNLRREIVKKNIDGYFK
 MT396266 146 HKNNKSWMSEFRVYSS---ANNCTFEVVSQFFLDLEGRQGNFKNLRREIVKKNIDGYFK
 MT374110 146 HKNNKSWMSEFRVYSS---ANNCTFEVVSQFFLDLEGRQGNFKNLRREIVKKNIDGYFK
 MT372483 146 HKNNKSWMSEFRVYSS---ANNCTFEVVSQFFLDLEGRQGNFKNLRREIVKKNIDGYFK
 MT359866 146 HKNNKSWMSEFRVYSS---ANNCTFEVVSQFFLDLEGRQGNFKNLRREIVKKNIDGYFK
 JX869059 165 AFYCILEPFRGNHCPAGNSYTSFATYHTPATDCSDGNYNENASLNSFKEFNFRNCTFMY
 CAA71147.1 1 -----
 CAA71146.1 1 -----
 ATN23889.1 34 HRGLSNTTSVITGAYPSTNKS DWS CNTRTGHLSGSGFGIGLYVQTPREOQGGSGAGGY
 APF29063.1 189 TVTV--NVTTHN-GRVVNYTVCDDCNGITDNIISVQQDGLIPNGFFNNVFLITNGSTLV
 AMO03401.1 182 AFYCILEPFRGNHCPAGNSYTSFATYHTPATDCSDGNYNENASLNSFKEFNFRNCTFMY
 AIP90486.1 52 NFD---VGVLPGYFVKN-----VNLISF-----ITNSTLPI
 AFS88936.1 182 AFYCILEPFRGNHCPAGNSYTSFATYHTPATDCSDGNYNENASLNSFKEFNFRNCTFMY
 ADX59495.1 195 VVTY--NVTDSKGAITSYNLCKSCNGVPTHVFSVPIDGSIPADFSFDNIFLITNTSSVV
 ADX59488.1 202 SITA--IVSIDSAGLISNYTVCSNCSGVVDHVVAVETGCTIPASFSFDNIFLITNTSSVP
 ADX59474.1 170 NATLCINPFGFNRWGPAGNIIASNALVDH-SNSCFVNNSNFNVNISTSRINLARIFDKDGLL
 ADX59466.1 166 NGTLCRNPVLPWRWGTSGNIVVNDALKCGSSITVFDQKNISYTEAASRFNIAIHDGNNLL
 ADX59451.1 218 QLTY--NVTGSHGQIISYNLCSQCNGVPAHVFSVQEGGYIPADFDNIFLITNTSTLV
 ACA52171.1 216 AITV--NVTTFANGTIDKVAICDNCNGVPAHIFVSEGGIIPADFNFSNIFLITNTSTIV
 ABA02260.1 143 PMG---TQHTMIFDN---AFNCTFEVVSQFFLDLEGRQGNFKNLRREIVKKNIDGFLY
 consensus 241

YP_002308479.1 189 ---NTWIC-----PLPANTD---FTIYHTGQLVANGIYLGVLPDTIKOVFLIQS---
 YP_001876437.1 406 WEDLRSLNMSRCVRFYDKLKKVMSVLEPMVKAFIIDFRQQLAIDGFPFGALIADEFFRFS
 YP_001351684.1 379 TNHIVRRLS-----VFPIIIRRE---FVITKYGNVYVNGIYLRTRRITAVHLNASHS
 QDY92359.1 212 FONGTALDL-----ILCDDTPRGIACQYGTGNFDGLYPFTHSSVINKTLEVYLTNS
 NC_045512 312 IYQTSNRRV-----QPTESTRV---FNITNLCFFGEVSNATRFASVYAMNRKRISN-C
 MT419822 312 IYQTSNRRV-----QPTESTRV---FNITNLCFFGEVSNATRFASVYAMNRKRISN-C
 MT415323 312 IYQTSNRRV-----QPTESTRV---FNITNLCFFGEVSNATRFASVYAMNRKRISN-C
 MT396266 312 IYQTSNRRV-----QPTESTRV---FNITNLCFFGEVSNATRFASVYAMNRKRISN-C
 MT374110 312 IYQTSNRRV-----QPTESTRV---FNITNLCFFGEVSNATRFASVYAMNRKRISN-C
 MT372483 312 IYQTSNRRV-----QPTESTRV---FNITNLCFFGEVSNATRFASVYAMNRKRISN-C
 MT359866 312 IYQTSNRRV-----QPTESTRV---FNITNLCFFGEVSNATRFASVYAMNRKRISN-C
 JX869059 343 VYSVSFEA-----KPSGSVVE---QAEG-VECDSPLLSGT-PPQVYNRRLVFN-C
 CAA71147.1 79 NETSARVQ-----ALPKTVRE---FVISRTGHFVINGIRYFTL-GNVEAVNFNVT--T
 CAA71146.1 79 NETSARVQ-----ALPKTVRE---FVISRTGHFVINGIRYFSL-GNVEAVNFNVT--N
 ATN23889.1 214 VKDISVVLVET-----NSGGSVFRYS--CADGFVNDLQCKLRFLDIPPGYNSSEVEYPTA
 AFP29063.1 359 THVSRFVE-----ILPPTVRE---IVVARTGQFVINGIKYFDL-GFVEAVNFNVT--T
 AMO03401.1 360 VYSVSFEA-----KPSGSVVE---QAEG-VECDSPLLSGT-PPQVYNRRLVFN-C
 AIP90486.1 185 ---NTWIC-----ILPPTDID---FTVYRTGQFVANGILLGTLPIITNYVRLYRG--
 AFS88936.1 360 VYSVSFEA-----KPSGSVVE---QAEG-VECDSPLLSGT-PPQVYNRRLVFN-C
 ADX59495.1 363 ---TSDFVE-----MLPPTVRE---FVITRAGAVVNGIRIFQY-VPDGVIFNVT--S
 ADX59488.1 378 QTRTSRFVE-----MLPPTVRE---FVVMRNGDFVINGIRIPSV-GTKSAREFNIS--T
 ADX59474.1 337 VYPLSRMRA-----QVD-GAVR---VTQRGAVCRP-PYNALLDPPQPVVWRFFMYD-C
 ADX59466.1 328 VYSLSNMRS-----QPS-ATVR---LSSADGICNV-NYTHFLRPLPFYVRRFQIKS-C
 ADX59451.1 385 NRRENRFVE-----ILPPTVRE---IVISRWGSFVNGVQLFES-PEVESVIFNVT--S
 ACA52171.1 385 ---TSDFVE-----MLPPTVRE---IVITRYGSIMLNGISIFQG-PPQGVLFNVT--N
 ABA02260.1 299 IYQTSNRRV-----VPSGDVVR---FNITNLCFFGEVSNATRFASVYAMNRKRISN-C
 consensus 421

YP_002308479.1 233 SIPSNYTFFALANLDTLITLNTTITNTVYCDR-STLDMACQISKHYDDEEYSDPQK
 YP_001876437.1 466 IVAMPQTFISTLNYTALVTFDRGNMRSWIDCNSEKPFDRRCDLQTDVWSDGEYFVNK
 YP_001351684.1 429 QDVAGFWTIAATNFIDVLEVNNTGICQRLLCD--TPENSVKCSQLSFELEDGEYSMTAD
 QDY92359.1 265 VVNTIVYQQLVNVTFYNQTNVQVGNISGSQLGNTL-NQLRLYQDKVFDAPYVYNNFSDS
 NC_045512 362 VADYSVLYNSASFSTFKCYGVSETRKNDLCFTNV-YADSFVIRGDEVROAPGQTERIAD
 MT419822 362 VADYSVLYNSASFSTFKCYGVSETRKNDLCFTNV-YADSFVIRGDEVROAPGQTERIAD
 MT415323 362 VADYSVLYNSASFSTFKCYGVSETRKNDLCFTNV-YADSFVIRGDEVROAPGQTERIAD
 MT396266 362 VADYSVLYNSASFSTFKCYGVSETRKNDLCFTNV-YADSFVIRGDEVROAPGQTERIAD
 MT374110 362 VADYSVLYNSASFSTFKCYGVSETRKNDLCFTNV-YADSFVIRGDEVROAPGQTERIAD
 MT372483 362 VADYSVLYNSASFSTFKCYGVSETRKNDLCFTNV-YADSFVIRGDEVROAPGQTERIAD
 MT359866 362 VADYSVLYNSASFSTFKCYGVSETRKNDLCFTNV-YADSFVIRGDEVROAPGQTERIAD
 JX869059 391 NYNLTKRLSLFVNDFTCSQISPAATASNCYSST-ILYFYSYPLSMKSDSVSSAGFTSQ
 CAA71147.1 127 AETTDFTVALASYADVLVNVSCTAHANLHGN--SVINRRCQDLSFDVDFGEYSTSPI
 CAA71146.1 127 AETT-CTVALASYADVLVNVSCTAHANLHGN--SVINRRCQDLSFDVDFGEYSTSPI
 ATN23889.1 269 LYTVVHNMSACPERPDSYCGSNSCPFRKRAVFN--CHVNYTWTWVNDQRDFQHLL
 AFP29063.1 407 ASATDFWTVAFATFVDVLVNVSATKIQMLYCD--SPEERKCEHLQFQDGEYSANFL
 AMO03401.1 408 NYNLTKRLSLFVNDFTCSQISPAATASNCYSST-ILYFYSYPLSMKSDSVSSAGFTSQ
 AIP90486.1 229 HLSANSAHFALANLDTLITLNTTITNTVYCDK-SVVDSEACQRSSHEVEDGEYSNPKS
 AFS88936.1 408 NYNLTKRLSLFVNDFTCSQISPAATASNCYSST-ILYFYSYPLSMKSDSVSSAGFTSQ
 ADX59495.1 408 TNGADWTVAFASDADVLVNVNATNORLYCD--NQLNIKKCQMQFTPDGEYVYSA
 ADX59488.1 426 VDGRDFWTVAFASVDMVNDNSTNORLYCD--NPLNQKCCQLRFNDDGEYSYAPE
 ADX59474.1 385 AFDFSVVIDNLPHQLCQYGISPRRNASNCYSST-TIDVMRINATHLNNLNRVPDSFSL
 ADX59466.1 376 KFDFNTVVSALPTYDLWCRGISPERRIGSMCGAV-TDPSMINTHYNDKSNVPDAFTR
 ADX59451.1 433 KVGSDFWTIAAQSVDVLLVNVSTSTNLYCD--SPNLIKCCQLSFRDDEGEYASSV
 ACA52171.1 430 RGATDWTVALSNFTVLAEVQSTAKKALYCD--DPLSQKCCQLQFSFDGEYATASL
 ABA02260.1 349 VADYSVLYNSTFFSTFKCYGVSETRKNDLCFTNV-YADSFVVKGDDVROAPGQTERIAD
 consensus 481

YP_002308479.1 383 STK--SGLSCTLVATKYWNYQIAQK--LYVTAIEGQSTGTTATSAQ--ADSSIVT
 YP_001876437.1 637 GVEYQGCQSDKACRFYFDNVCFQYGNKLYKPDLEFIDYQFGLFRKNSVQSNITQ
 YP_001351684.1 589 STSPADGACEIKYYVWNTIGAVSHLAGTLYVQHTKCDIITCTPKPLQ---GLNDISELHL
 QDY92359.1 388 YMYVNNQPLCAGIDPEVYNRTRCALAVVVMQSGSRICTSETVPIINTVSTLNNTFVL
 NC_045512 514 SFELHHPATVCGPKKSTNLVKNK---CVNENFNGLIGTGVLTESNKK--FPPFOQFGR
 MT419822 514 SFELHHPATVCGPKKSTNLVKNK---CVNENFNGLIGTGVLTESNKK--FPPFOQFGR
 MT415323 514 SFELHHPATVCGPKKSTNLVKNK---CVNENFNGLIGTGVLTESNKK--FPPFOQFGR
 MT396266 514 SFELHHPATVCGPKKSTNLVKNK---CVNENFNGLIGTGVLTESNKK--FPPFOQFGR
 MT374110 514 SFELHHPATVCGPKKSTNLVKNK---CVNENFNGLIGTGVLTESNKK--FPPFOQFGR
 MT372483 451 ---LHHPATVCGPKKSTNLVKNK---CVNENFNGLIGTGVLTESNKK--FPPFOQFGR
 MT359866 514 SFELHHPATVCGPKKSTNLVKNK---CVNENFNGLIGTGVLTESNKK--FPPFOQFGR
 JX869059 560 YGDTNSVCPKLEFANDTKIASGLGN--CVPMSLYGVSGRQVFNCTAV--GRRQRFVY
 CAA71147.1 291 SLKDIPEGCAMPIVANWAYSKYTYTIG-SLYVSWSDGDGIVTGPVPVE---GVSSFMNATL
 CAA71146.1 288 SLKDIPEGCAMPIMANLANLNHSHNIG-SLYVSWSDGDVITGVPKPE---GVSSFMNATL
 ATN23889.1 378 DNQDSYDWNWADFVSTGAFVCSNP---EAPKTGVCVTVTVEKVTFGQ-VLYESFTF
 AFP29063.1 569 STVEVPESCNFPLEATWHYTSYTVIG-ALYVWVSEGNITGVVYVPS---GIREFSNLVL
 AMO03401.1 577 YGDTNSVCPKLEFANDTKIASGLGN--CVPMSLYGVSGRQVFNCTAV--GRRQRFVY
 AIP90486.1 380 STE--SGLACEKLIIVTYVWNYLLRQR--LYVTAIEGQHTGT--TSVH---ADTSSVIT
 AFS88936.1 577 YGDTNSVCPKLEFANDTKIASGLGN--CVPMSLYGVSGRQVFNCTAV--GRRQRFVY
 ADX59495.1 581 SLEPQAC-ACTLQINKIWSVTFVFA-SLYVSYKTSAIMGVKTPSV---GIFDFSTLVL
 ADX59488.1 593 SLVPNGG-CGMSIITQGSYGEDPKFA-VLYVSYTNGDNVVGVEGKAPQPGVYDSEVHY
 ADX59474.1 509 TLKPASGNLVCPIANDTIDR---CVQMSLYGVIGTGVLSTNTSL--TVPKGV-F
 ADX59466.1 504 SLTIAS-NTMCPVGNDFNIVSGV---CVNENIYCYGCTGVIS-NSSA--VPPGSKV-F
 ADX59451.1 584 SVQPVGA-SCTSIHIGWMTYFIPWR-DIYVTRYKHSALIGVTRKAST---GIFDPSFLVL
 ACA52171.1 598 SLNGTICKGCTGIYKRASSQYIPW-NVWVAYSGDNIIGVREPVE---GVRDQSVVHQ
 ABA02260.1 500 SFELHHPATVCGPKKSTNLVKNK---CVNENFNGLIGTGVLTPSSKR--FPPFOQFGR
 consensus 661

YP_002308479.1 436 DVCTDYTYGVSGETIT---HSDLELHNGIAFTGP-----TGEYFAFKNTTSGKS
YP_001876437.1 697 EGCHDYDIYGKGTGWVKPFDYSYLOTGGILMDSGGFPPLFRYNFALYEVTQCTREVA
YP_001351684.1 646 DTCTTYTYGFRGDEWIR---LTNQTFLSGVYYTSE-----SGDILAFKRVTTGQT
QDY92359.1 448 DTCVNYTYSRYGGLVT---NVTDOVTFGTFLNDGGLVLLDSSGALIDFAVNEVVRGRHY
NC_045512 568 DIADTTDAVRDPQTLEIL---DITFCSPGGVSVITP-----GNNTSNQVAVLYQGV
MT419822 568 DIADTTDAVRDPQTLEIL---DITFCSPGGVSVITP-----GNNTSNQVAVLYQGV
MT415323 568 DIADTTDAVRDPQTLEIL---DITFCSPGGVSVITP-----GNNTSNQVAVLYQGV
MT396266 568 DIADTTDAVRDPQTLEIL---DITFCSPGGVSVITP-----GNNTSNQVAVLYQGV
MT374110 568 DIADTTDAVRDPQTLEIL---DITFCSPGGVSVITP-----GNNTSNQVAVLYQGV
MT372483 502 DIADTTDAVRDPQTLEIL---DITFCSPGGVSVITP-----GNNTSNQVAVLYQGV
MT359866 568 DIADTTDAVRDPQTLEIL---DITFCSPGGVSVITP-----GNNTSNQVAVLYQGV
JX869059 616 DAYQN-LVGGYSDDENYY---CLRACVSVFVSVIYD-----KETKTHALFLGVSAC
CAA71147.1 347 DKCTKYNLYDVSGVWIR---VSNDFLNGITYTST-----SCNLGFKDVTKGTI
CAA71146.1 344 NKCTKYNLYDVSGVWIR---ISNDTFLNGITYTST-----SCNLVGFKDVINGTI
ATN23889.1 433 AQYNNLLYVG-SQLRYVR---LGGVYEVSSCFEAS-----YDLVLRNQSFGILI
APF29063.1 625 NNCTKYNLYDYGTEHIR---SSNQSLAGGLTYVSN-----SGNLGFKNVTGNI
AM003401.1 633 DAYQN-LVGGYSDDENYY---CLRACVSVFVSVIYD-----KETKTHALFLGVSAC
AIP90486.1 431 DVCTDYTYGVSGETIT---PSDLLLHNGIAFTGP-----TGEYFAFKNTTSGKT
AF888936.1 633 DAYQN-LVGGYSDDENYY---CLRACVSVFVSVIYD-----KETKTHALFLGVSAC
ADX59495.1 636 DYCSEYVYGIAGRWIS---RTNLTVMAGLYYTS-----SGLLGFKNATTGVY
ADX59488.1 651 NVCTSYTYGHTGRGWLIS---EAPVDYITGLFYTS-----AGDLAYKNTATGRV
ADX59474.1 562 TASSTGATIGVFINCTMY---SITECVT---VPVSV-----GYHANFVTFENGL
ADX59466.1 556 SLSHGDVSLFKYNNRKY---ALYFCAY---APLTA-----APYAGYNSTLFDNNL
ADX59451.1 639 DECTDYSYVIGSGRGLVIR---HVNTSYISGLYYTST-----ACGLIGFKNATTGCV
ACA52171.1 654 NVCTSYTYGHTGRGWLIS---EAPVDYITGLFYTS-----AGDLAYKNTATGRV
ABA02260.1 554 DVSDFTDSDVRDPKTEIL---DITFCSPGGVSVITP-----GNNTSNQVAVLYQGV
consensus 721 .. .

YP_002308479.1 539 EPLVLSYGPVSCSDGATIAETSFLQDRSRPISVLYDGE----VEVPSAFTNSVQTEYLQV
YP_001876437.1 816 PSLMMENSHCINIQGMVMNEPAITFNSTQYQPLLVSYG--DVQVPRITLMHHTTETETQIT
YP_001351684.1 746 SPRLVYSNIGVCTSGALGLLSPKAAQP-QVQPMFOGN-----ISIPNFNFTSVTEYLQI
QDY92359.1 565 CAYVSYGQYCKPDASVQIQPDIFDGFVTPLDNATEY---VLPDSFNDIVTDEYLOF
NC_045512 673 SYQTQTNERRRARSVASQSI IAYTMSLGAENSVAYSN---NSIAIPNFNFTSVTEYLLPV
MT419822 673 SYQTQTNERRRARSVASQSI IAYTMSLGAENSVAYSN---NSIAIPNFNFTSVTEYLLPV
MT415323 673 SYQTQTNERRRARSVASQSI IAYTMSLGAENSVAYSN---NSIAIPNFNFTSVTEYLLPV
MT396266 673 SYQTQTNERRRARSVASQSI IAYTMSLGAENSVAYSN---NSIAIPNFNFTSVTEYLLPV
MT374110 673 SYQTQTNERRRARSVASQSI IAYTMSLGAENSVAYSN---NSIAIPNFNFTSVTEYLLPV
MT372483 607 SYQTQTNERRRARSVASQSI IAYTMSLGAENSVAYSN---NSIAIPNFNFTSVTEYLLPV
MT359866 673 SYQTQTNERRRARSVASQSI IAYTMSLGAENSVAYSN---NSIAIPNFNFTSVTEYLLPV
JX869059 721 LPDTPSTLT PRSVRSVPGE MRLASIAFNHPIQVDQLNSSFYFKI SIPNFNFTSVQVEYLOI
CAA71147.1 448 DAVLTYSEFGWCADGSLI AVQPRNYSYDSVSAIVTAN-----LSIPSNFTSVQVEYLOI
CAA71146.1 445 DAVLTYSEFGWCADGSLI AVQPRNYSYDSVSAIVTAN-----LSIPSNFTSVQVEYLOI
ATN23889.1 528 NFLGDGFCDLGNVAVRRMTFEKHDITYVAPVTNER---YTEMPLDHCIIHTEQEQLQF
APF29063.1 726 TAVMTYSNFGWCADGSLI PVRPRNSSDNGISAIVTAN-----LSIPSNFTSVQVEYLOI
AM003401.1 738 LPDTPSTLT PRSVRSVPGE MRLASIAFNHPIQVDQLNSSFYFKI SIPNFNFTSVQVEYLOI
AIP90486.1 534 KPVLTSYGPISCSGDAIVGTSTLQNRSPISVLYDGE----VEVPSAFTNSVQTEYLQV
AF888936.1 738 LPDTPSTLT PRSVRSVPGE MRLASIAFNHPIQVDQLNSSFYFKI SIPNFNFTSVQVEYLOI
ADX59495.1 737 DEPVLTYSEFGLVCRNGATVNTARTVAAKPSVTVGVGN-----LSIPNFNFTSVQVEYLOI
ADX59488.1 757 VPSLTYGGGLICADGKLVNATRTVAATEPVSPVITGY----ISVPTNFNFTSVQVEYLOI
ADX59474.1 665 INDTTVAVARAAGLPRRYLVNVDPIVDNNSATPMTFV---YVWRIPNFNFTSVTEYLQV
ADX59466.1 659 R---PTVSARTTGGESMLVLYDPIVD---SLVPITPV---YCIDVPTNFNFTSVTEYLQV
ADX59451.1 740 DFIITYSYVGI CKDGTLVKVDPKPATSTFVSPITSTAN-----LVVPTNFNFTSVQVEYLOI
ACA52171.1 755 DPSVVYSYIGKCDGGLVFNSTRVGEFPDAISMGN-----ISVPTNFNFTSVQVEYLOI
ABA02260.1 659 SYHTVS---LRSYSQKSLVAYTMSLGAENSVAYSN---NTIIPNFNFTSVTEYLLPV
consensus 841 .

YP_002308479.1 594 QSEQVVDCAVYVCCENARCTRIISLQYTSACSNIESAALHSSAQDREITINMFQTSSQSL
YP_001876437.1 874 HSEKIIVDCKQVYVCCSESRCCRRIILVYGFCECSIMIRLNGVMADVSVTDLDFLDFNSFD
YP_001351684.1 800 FNKPVSVDCAMVYVCCNDRCCKOILSQYTSACKNIESAALQLSARIESMVNSMLTVSDEAL
QDY92359.1 621 RSEQVVDCAVYVCCSESRCCRRIILVYGFCECSIMIRLNGVMADVSVTDLDFLDFNSFD
NC_045512 720 SMNKTQVDCMYTCCGSTECSNLLIYOYGSFCCQLNRALTGHAIVECDKNTQEVFAQVQ--
MT419822 730 SMNKTQVDCMYTCCGSTECSNLLIYOYGSFCCQLNRALTGHAIVECDKNTQEVFAQVQ--
MT415323 730 SMNKTQVDCMYTCCGSTECSNLLIYOYGSFCCQLNRALTGHAIVECDKNTQEVFAQVQ--
MT396266 730 SMNKTQVDCMYTCCGSTECSNLLIYOYGSFCCQLNRALTGHAIVECDKNTQEVFAQVQ--
MT374110 730 SMNKTQVDCMYTCCGSTECSNLLIYOYGSFCCQLNRALTGHAIVECDKNTQEVFAQVQ--
MT372483 664 SMNKTQVDCMYTCCGSTECSNLLIYOYGSFCCQLNRALTGHAIVECDKNTQEVFAQVQ--
MT359866 730 SMNKTQVDCMYTCCGSTECSNLLIYOYGSFCCQLNRALTGHAIVECDKNTQEVFAQVQ--
JX869059 781 TTKQVTVDCQKQVYVCCNRFQCEQQLLREYGFCSKINCAHGANRCDSDVRNLFASFVRS-S
CAA71147.1 503 TSTPIVVDCAVYVCCGNVRCVVELLKOYTSACKTIEDALRNSARIESADVSEMLTFDKKAF
CAA71146.1 500 TSTPIVVDCAVYVCCGNVRCVVELLKOYTSACKTIEDALRNSARIESADVSEMLTFDKKAF
ATN23889.1 584 TMEKRSVSCETYTCDVSKACRNLFRYGFCECFCKQKVEADIRAGLITLDGVDSSLYSTDAKT
APF29063.1 781 TSTPIVVDCAVYVCCGNVRCVVELLKOYTSACKTIEDALRNSARIESADVSEMLTFDKKAF
AM003401.1 798 TTKQVTVDCQKQVYVCCNRFQCEQQLLREYGFCSKINCAHGANRCDSDVRNLFASFVRS-S
AIP90486.1 589 QAEQVTVDCQKQVYVCCNSRCLLOLLAQTYSRCSKIEAALHSSAQDREITINMFQTSSQSL
AF888936.1 798 TTKQVTVDCQKQVYVCCNRFQCEQQLLREYGFCSKINCAHGANRCDSDVRNLFASFVRS-S
ADX59495.1 792 AVTPVSVDCSTYVCCGNPCLRLLLSOYTMACKTIEDALQLSARIESGPEVPLDVSHTDM
ADX59488.1 812 MKRPVSVDCSTYVYVCCGNPCLRLLOLLAQTYSRCSKIEAALHSSAQDREITINMFQTSSQSL
ADX59474.1 712 NAKPVTVDCARTYCCGSSRCLNVLLHYGFCDINRPAKAVNTLTDGALSLSKELINT
ADX59466.1 712 YASKLSDCISKYI CCSTVCCKNVLLOYGFCDVNDVNTALTRVFSLIDTALVETVSTKFS--
ADX59451.1 795 YNKPVSVDCAVYVCCGNPCLRLLLSOYTMACKTIEDALQLSARIESVPELMLTVSDDAV
ACA52171.1 810 STRPVSDCAVYVCCGNPCLRLLOLLAQTYSRCSKIEAALHSSAQDREITINMFQTSSQSL
ABA02260.1 712 SMNKTQVDCMYTCCGSTECANLLIYOYGSFCCQLNRALTSGHAAECDKNTREVFAQVQ--
consensus 901 .

YP_002308479.1 654 TLANIS-----NFQS--DNNFNNTITSKS-----GGRSAIEDLLENKVVVTNGLCTVD
 YP_001876437.1 934 NISVSLQNLNASFGQFEGYGFQKFLPKSFDGSDPRDARSATEDILFDKVVTVGLCTVD
 YP_001351684.1 860 KLATIS-----QFPGGGNFNNTIPANP-----GARSVIEDILFDKVVTSGLCTVD
 QDY92359.1 681 NYPPFEN-----NLDGGFNISLILLPQNK---SGVQSRSEIEDLLEFKIESVGLF-TD
 NC_045512 788 IYKTPP-----IKDRGGFNFSQIIL----PDPSPKPKRSFIEDLLENKVVTLADAG-FI
 MT419822 788 IYKTPP-----IKDRGGFNFSQIIL----PDPSPKPKRSFIEDLLENKVVTLADAG-FI
 MT415323 788 IYKTPP-----IKDRGGFNFSQIIL----PDPSPKPKRSFIEDLLENKVVTLADAG-FI
 MT396266 788 IYKTPP-----IKDRGGFNFSQIIL----PDPSPKPKRSFIEDLLENKVVTLADAG-FI
 MT374110 788 IYKTPP-----IKDRGGFNFSQIIL----PDPSPKPKRSFIEDLLENKVVTLADAG-FI
 MT372483 722 IYKTPP-----IKDRGGFNFSQIIL----PDPSPKPKRSFIEDLLENKVVTLADAG-FI
 MT359866 788 IYKTPP-----IKDRGGFNFSQIIL----PDPSPKPKRSFIEDLLENKVVTLADAG-FI
 JX869059 840 QSSPIL-----PGFGDENLILTEP-VSISTGSRARSATEDLLEFKVVTIADG-FM
 CAA71147.1 563 TLANVS-----SFG--DNNLSSVLPSPFTSGSRVAGRSATEDILFSKLVTSGLCTVD
 CAA71146.1 560 TLANVS-----SFG--DNNLSSVLPSPFRSGSRVAGRSATEDILFSKLVTSGLCTVD
 ATN23889.1 644 SSVVPT-----DRFNVSQFELPKTRSSAN---KYESRSATEDLLEFKLVTTGGC-FY
 APF29063.1 841 SLANVT-----SFG--DNNLSSVLPQRNIRSSRIAGRSATEDLLEFSKVVTSGLCTVD
 AMO03401.1 857 QSSPIL-----PGFGDENLILTEP-VSISTGSRARSATEDLLEFKVVTIADG-FM
 AIP90486.1 649 QLANTT-----NFKG-DNFSSILITRI-----GGRSAIEDLLENKVVTSGLCTVD
 AF88936.1 857 QSSPIL-----PGFGDENLILTEP-VSISTGSRARSATEDLLEFKVVTIADG-FM
 ADX59495.1 852 HLANIT-----TFG--PNNLSSVVGVG----GKRSFIEDLLEFKVVVTNGLCTVD
 ADX59488.1 872 ALGVIS-----NFNN-TFNMSAVVPPNV-----GGRSAIEDLLEFKVVTSGLCTVD
 ADX59474.1 782 RDEIQS-----FAFGDENFTGMLGCLGPNCGATYRSALSLLYDKVKITDGC-FM
 ADX59466.1 770 -TAPAD-----LAYTGFNFSLVGLGTSCNONSYSALSLLYDKVKIVADGC-FM
 ADX59451.1 855 RLANTT-----YFDN--YNVSAITPRAS-----GRGSFIEDLLEFKVVTSGLCTVD
 ACA52171.1 870 DLVNTS-----TFGG-DNNLTALLPQGG----GKRSVIEDILFDKVVTSGLCTVD
 ABA02260.1 770 MYKTPP-----LKYFGFNFSQIIL----PDPLKPKRSFIEDLLENKVVTLADAG-FM
 consensus 961 * * * * *

YP_002308479.1 812 LLTWALAIKIKIIVVNSQCEALSHLTAQLSNFCAISISIODIINRLEVEANGQVDRLLI
 YP_001876437.1 1100 FSAVSSAITEVQNAVNACSCALTKLLGQLNINFGATISSIRELVERLAKLEADVQVDRLLI
 YP_001351684.1 1019 LNWVAPALSQVQDVVNGOGNALSQITVQIQONNFCAISSIGDIIYRLLQITADACVDRLLI
 QDY92359.1 833 --ATSCALQIQVQDVVNCQAAVLTETVMSHNRNFCAISSVIODIKCLDEL SANACVDRLLI
 NC_045512 940 --STASALCKLQDVVNONAQAALNTLVKQLSSNFGAISSVWLDILSRLLDKVEAEVQIDRLLI
 MT419822 940 --STASALCKLQDVVNONAQAALNTLVKQLSSNFGAISSVWLDILSRLLDKVEAEVQIDRLLI
 MT415323 940 --STASALCKLQDVVNONAQAALNTLVKQLSSNFGAISSVWLDILSRLLDKVEAEVQIDRLLI
 MT396266 940 --STASALCKLQDVVNONAQAALNTLVKQLSSNFGAISSVWLDILSRLLDKVEAEVQIDRLLI
 MT374110 940 --STASALCKLQDVVNONAQAALNTLVKQLSSNFGAISSVWLDILSRLLDKVEAEVQIDRLLI
 MT372483 874 --STASALCKLQDVVNONAQAALNTLVKQLSSNFGAISSVWLDILSRLLDKVEAEVQIDRLLI
 MT359866 940 --STASALCKLQDVVNONAQAALNTLVKQLSSNFGAISSVWLDILSRLLDKVEAEVQIDRLLI
 JX869059 997 --TTNEAFQKQVQDVVNNACALSKLASELSNIFGAISSIGDIIQRLDVLVQDQVDRLLI
 CAA71147.1 726 LQVWALAIKIKIIVVNSQCEALSHLTAQLSNFCAISISIODIINRLEVEANGQVDRLLI
 CAA71146.1 723 LQVWALAIKIKIIVVNSQCEALSHLTAQLSNFCAISISIODIINRLEVEANGQVDRLLI
 ATN23889.1 794 --TTSALCALIIVVNONAQAQVESLVSGLTENFGAISINFKRAISCRLLDKLEADVQVDRLLI
 APF29063.1 1004 IHWVTITAIKIKIIVVNSQCEALSHLTAQLSNFCAISISIODIINRLEVEANGQVDRLLI
 AMO03401.1 1014 --TTNEAFQKQVQDVVNNACALSKLASELSNIFGAISSIGDIIQRLDVLVQDQVDRLLI
 AIP90486.1 807 LNWVAPALSQVQDVVNCQAAVLTETVMSHNRNFCAISSVIODIKCLDEL SANACVDRLLI
 AF88936.1 1014 --TTNEAFQKQVQDVVNNACALSKLASELSNIFGAISSIGDIIQRLDVLVQDQVDRLLI
 ADX59495.1 1009 INWVASALNKVQSVVNCQOATHQITKOLSNFCAISSISIDIIWRLDVAADACVDRLLI
 ADX59488.1 1032 LSWVASALNKVQSVVNEQGTALSQITKOLSNFCAISSISIDIIWRLDVAADACVDRLLI
 ADX59474.1 941 --ETSTALSKMVDVINCRAACLQTLVTOIGNSFCAISSISINEIISRLLEHAAEAQVDRLLI
 ADX59466.1 928 --ATNCALNKIQTIVVNHNLQCLQALVQOLGNTFGAISSVNEIISRLDQLEANAQVDRLLI
 ADX59451.1 1015 INWVASALNKVQSVVNEQGHATSQITROIASNFCAISSISIODIINRLEVEANGQVDRLLI
 ACA52171.1 1028 IKTVASALNKVQSVVNSQOATHQITKOLSNFCAISSISIDIIWRLDVAADACVDRLLI
 ABA02260.1 922 --TTSALCKLQDVVNONAQAALNTLVKQLSSNFGAISSVWLDILSRLLDKVEAEVQIDRLLI
 consensus 1141* * * * *

YP_002308479.1 931 PNGIFEMHVLVLPQOTRNVASAGLCVDAI-----KGYSLQ-PQLILYQ
 YP_001876437.1 1219 PNGMLFVHFSLVFNETITVKGIIIGCLNNSIGIVPAKDRGMFIQCSNGTYCFEENSINE
 YP_001351684.1 1139 PGLIMFLHVLVLPNGIINVTAVTGLCVDET-----IAMTRQSGVLFV
 QDY92359.1 950 PNGIAFTHFTYLPQTKNVVAVVGCQVTTNG-----SGYGLVFNNGRGIIFY
 NC_045512 1057 PHGVVFLHVTVVPAACEKNFTTAPAI CHDG-----KAHFPRREGV FVS
 MT419822 1057 PHGVVFLHVTVVPAACEKNFTTAPAI CHDG-----KAHFPRREGV FVS
 MT415323 1057 PHGVVFLHVTVVPAACEKNFTTAPAI CHDG-----KAHFPRREGV FVS
 MT396266 1057 PHGVVFLHVTVVPAACEKNFTTAPAI CHDG-----KAHFPRREGV FVS
 MT374110 1057 PHGVVFLHVTVVPAACEKNFTTAPAI CHDG-----KAHFPRREGV FVS
 MT372483 991 PHGVVFLHVTVVPAACEKNFTTAPAI CHDG-----KAHFPRREGV FVS
 MT359866 1057 PHGVVFLHVTVVPAACEKNFTTAPAI CHDG-----KAHFPRREGV FVS
 JX869059 1114 PNGLIFMHWGYPFSNHIEVVSAYGLQDAFN-----PTNCLAFVNGVFTK
 CAA71147.1 845 PGLVFLHVLVLPQTKVDEAWNSGLCVDGT-----NGYVLRQPNLALYK
 CAA71146.1 842 PGLVFLHVLVLPQTKVDEAWNSGLCVDGR-----NGYVLRQPNLALYK
 ATN23889.1 911 PSGIMFHFHSLVFNNTITVKKTPGLCESDELG-----SKCVAKDGLVLS
 APF29063.1 1123 PGLLIFLHVLVLPQTKVKNVKAWSGLCVDTI-----YGYVLRQPNLVLS
 AMO03401.1 1131 PNGLIFMHWGYPFSNHIEVVSAYGLQDSFN-----PTNCLAFVNGVFTK
 AIP90486.1 926 PNGIFEMHVLVLPNKITRNVASAGLCVDET-----KGYSLQ-PQLILYQ
 AF88936.1 1131 PNGLIFMHWGYPFSNHIEVVSAYGLQDAFN-----PTNCLAFVNGVFTK
 ADX59495.1 1128 PNGIMLHVLVLPQTKVDEAWNSGLCVSD-----AAYVLRQPNLALYK
 ADX59488.1 1151 PGLVFLHVLVLPQTKVNVSAFAGLCVDCS-----RAYVLRQPNLALYK
 ADX59474.1 1058 PNGVLFTHYAVLPQTKVNVSAFAGLCVDCS-----TGYPVLRQPNLALYK
 ADX59466.1 1045 PNGMLFTHYAVLPQTKVNVSAFAGLCVDCS-----TGYPVLRQPNLALYK
 ADX59451.1 1134 PGLVFLHVLVLPQTKVNVSAFAGLCVDCS-----YGLVLRQPNLALYK
 ACA52171.1 1147 PGLVFLHVLVLPQTKVNVSAFAGLCVDCS-----YGLVLRQPNLALYK
 ABA02260.1 1039 PHGVVFLHVTVVPSQERNFTTAPAI CHDG-----KAYFPRREGV FVS
 consensus 1261 * * * * *

```

YP_002308479.1 1088 VDLQWLNRRVEEYIKWPWYVWLAIFLAIAAFACIIIVTIFLCTGCCGSCFGCCGGGCFG----
YP_001876437.1 1390 IDLEWLNRYERYSKWPWWVWLIIVLCFAAFCLLVFWIFICTGCCGGCCNCCGIPALFGYC
YP_001351684.1 1297 VDLQWLNRRVEEYIKWPWYVWLAIVVVALIVLVVSLVFCCLISTGCCG-CGCGGSCFSG--C
QDY92359.1 1096 VDLQWLNRRVEEYIKWPWYVWLAIAAFACIIIFILLICWIFEMTGCCGCCGCGFLIPLMSKC
NC_045512 1198 IDLQELGKYEQYIKWPWYVWLEFIAGLTAIVMVTIMLCMTSCCCLRGCCSCGSCCKFD
MT419822 1198 IDLQELGKYEQYIKWPWYVWLEFIAGLTAIVMVTIMLCMTSCCCLRGCCSCGSCCKFD
MT415323 1198 IDLQELGKYEQYIKWPWYVWLEFIAGLTAIVMVTIMLCMTSCCCLRGCCSCGSCCKFD
MT396266 1198 IDLQELGKYEQYIKWPWYVWLEFIAGLTAIVMVTIMLCMTSCCCLRGCCSCGSCCKFD
MT374110 1198 IDLQELGKYEQYIKWPWYVWLEFIAGLTAIVMVTIMLCMTSCCCLRGCCSCGSCCKFD
MT372483 1132 IDLQELGKYEQYIKWPWYVWLEFIAGLTAIVMVTIMLCMTSCCCLRGCCSCGSCCKFD
MT359866 1198 IDLQELGKYEQYIKWPWYVWLEFIAGLTAIVMVTIMLCMTSCCCLRGCCSCGSCCKFD
JX869059 1264 IDLQELGKYEQYIKWPWYVWLEFIAGLTAIVMVTIMLCMTSCCCLRGCCSCGSCCKFD
CAA71147.1 1003 VDLKWLNRVEEYIKR-----SGYIPEAPRD-----GQAYVRDGEWV
CAA71146.1 1000 VDLKWLNRVEEYIKR-----
ATN23889.1 1055 VEVEKFORVYNYVKWAWWQWLLIFIALTIFAGHMLWCCLATGCCG-MGCLAAATCASCCD
APF29063.1 1281 VDLKLLNRVEEYIKWPWYVWLIIVSVVFWLISLLVFCCLSTGCCG-CCNCLTSSMRGCCD
AMO03401.1 1281 IDLQELGKYEQYIKWPWYVWLEFIAGLVAIAACVFFILCCTGCCGTNCMGKLCNRCCDRY
AIP90486.1 1083 VDLQWLNRRVEEYIKWPWYVWLAIAAFACIIIVTIFLCTGCCGSCFGCCGGGCFG----
AF888936.1 1281 IDLQELGKYEQYIKWPWYVWLEFIAGLVAIAACVFFILCCTGCCGTNCMGKLCNRCCDRY
ADX59495.1 1285 VDLQWLNRRVEEYIKWPWYVWLIIVVVALTIFAGHMLYCCCLATGCCG-CCSCMNTL----D
ADX59488.1 1311 VNLQWLNRRVEEYIKWPWYVWLAIVVVALIVLVGMLWCCLATGCCG-CCSCLANSCG---D
ADX59474.1 1201 IDLQELGKYEQYIKWPWYVWLAIAAFACIIIVTIFLCTGCCGSCFGCCGGGCFG----
ADX59466.1 1189 IDLQELGKYEQYIKWPWYVWLEFIAGLVAIAACVFFILCCTGCCGTNCMGKLCNRCCDRY
ADX59451.1 1292 VDLQWLNRRVEEYIKWPWYVWLAIFLVVALFSLFIMLYCCCLATGCCG-CCSCLASSTL----D
ACA52171.1 1304 VDLQWLNRRVEEYIKWPWYVWLIIVVVALTIFAGHMLWCCLATGCCG-CCSCMASTL----D
ABA02260.1 1180 IDLQELGKYEQYIKWPWYVWLEFIAGLTAIVMVTIMLCMTSCCCLRGCCSCGSCCKFD
consensus 1441 ... ..*.....

```

Figure 5.2 Multiple Sequence alignment of the S-protein of different corona virus species. The highly conserved residues are marked in black and moderately conserved residues are shown in grey colour. The yellow and red highlights are showing the linear B-cell and T-cell epitopes.

Similarly, in case of T-cell epitopes, highly conserved ¹¹⁶SLLVNNTATNVVIK¹²⁹, ⁷⁸¹VFAQVKQIY⁷⁸⁵ epitopes are located at the shaft of S-protein. The rest of the epitopes are less conserved and mostly located at the RBD region of S-protein. This indicates that amino-acid composition of RDB of SARS-CoV-2 S-protein is different from other coronavirus species.

Further, two separate vaccines have been designed from the epitopes predicted in **Table 5.1**, Vac-COVID-B and Vac-COVID-T from B-cell and T-cell respectively, which can be used as a combination for both humoral and cell-mediated immunity. The linear B-cell epitopes and T-cell epitopes are linked by GPGPG linker to avoid the formation of junctional epitope (Kar et al. 2020) and Cholera Toxin B (CTB) adjuvant linked by EAAAK linker for immune regulation. The total number of amino-acids of the vaccines are 133 and 111 amino acid (aa) long with molecular weight of 13.11 kDa and 11.4 kDa respectively. The 3D-models of the chimeric vaccine shown in **Figure 5.3** are predicted to be non-allergen by AllerTOP server (Dimitrov et al. 2014).

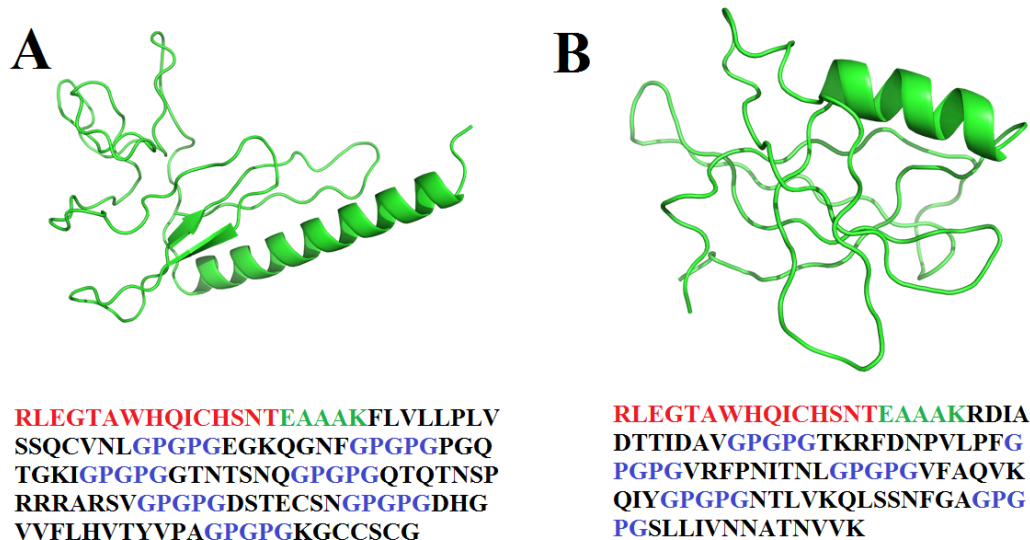


Figure 5.3. The tertiary structure of chimeric vaccines made of (A) B-cell linear epitopes (B) T-cell linear epitopes. The amino acid sequence of the corresponding vaccines is shown below. The Adjuvant is shown in red; the adjuvant linker is shown in green and the epitope linkers are indicated in blue.

The structural quality of the vaccines is assessed by Z-score and Ramachandran plot (**Appendix XXXVII**). The Z-score of the vaccines made of B-cell and T-cell epitopes are -3.43 and -4.85 respectively which confirm the reliability of our model. Additionally, 95.42 % and 92.7% amino acid residues are found to be in the favored region in the Ramachandran plot. Further, the antigenicity of the linear B-cell and T-cell epitopes and the multimeric vaccines have been assessed by VaxiJen (Gordon et al. 2020) webserver to confirm their ability to elicit protective immune response. The antigenicity scores of the liner epitopes obtained from VaxiJen webserver is documented in **Table 5.2**.

Table 5.2 The antigenicity score of selected linear B-cell and T-cell peptides epitopes.

<i>B-cell Epitopes</i>				
Method	Start	end	Peptide	Antigenicity score
Emini Surface Accessibility	808	817	DPSKPSKRSF	0.814

Parker Hydrophilicity	601	607	GTNTSNQ	0.940
	745	751	DSTECSN	0.698
Karplus & Schulz Flexibility				
B-epipred	180	186	EGKQGNF	1.933
	412	418	PGQTGKI	1.547
	675	687	QTQTNSPRRRARSV	0.530
	805	816	ILPDPSKPSKRS	0.532
Kolaskar & Tongaonkar Antigenicity	4	18	FLVLLPLVSSQCVNL	0.830
	1057	1070	PHGVVFLHVTYVPA	0.884
Chou & Fasman Antigenicity	600	606	PGTNTSN	0.917
	1245	1251	KGCCSCG	0.97
<i>T-cell Epitopes</i>				
HLA-allele	Start	End	Peptide	Antigenicity Score
A*02:01	567	576	RDIADTTDAV	1.75
A*24:02	76	86	TKRFDNPVLPF	1.305
B*40:01	327	335	VRFPNITNL	1.114
B*58:01	781	789	VFAQVKQIY	1.43
DRB1*04:01	960	972	NTLVKQLSSNFGA	1.15
DRB1*07:01	116	129	SLLVNNTATNVVIK	1.47

It is evident from **Table 5.2** that the antigenicity scores of all the linear epitopes are greater than the threshold value of 0.5 which indicates their antigenic nature and their ability to potentiate the immune memory against SARS-CoV2. Additionally, the antigenicity of the two vaccines candidates is found to be 0.73 and 0.58 (score > 0.5 is considered to be antigenic). In order to validate the antigenicity of vaccine candidates, the combinations of the linear epitopes have been altered. It is found that the alteration of linear epitope combination changes the antigenic score by 0-1-0.4 which is insignificant in the context of antigenicity. This is because the antigenic prediction by

Vaxigen is trained on the known antigen types. The antigenic score only changes significantly when the amino acid sequence of the linear epitope changes.

It is evident from current literatures that vaccine size of 20-200 nm is most effective in terms of antigen processing and presentation by antigen presenting cells. These larger antigens are mostly carried by dendritic cells (DCs) to the lymphatic system and takes 24 hrs for the immune response (Manolova et al. 2008) . However, there are reports which suggest that vaccine size less than ~5 nm is efficient in terms of direct transportation through specialized small antigen conduits to the lymph node for quick (1 hr) immune response (Carrasco and Batista 2007; Singh 2021). Thus, the size of the subunit and chimeric vaccines designed in the current chapter may elicit immune response more efficiently by diffusing through tight lymphatic capillaries to lymphoid organs.

5.3.2 Molecular Docking

It will be interesting to study the interaction of the multi-epitope vaccines designed with the immune-cell receptor for eliciting stable immune response. We considered the structure of the humanized antibody (7BZ5) as immune receptor for B-cell epitope, where the antigen binding region of 7BZ5 is formed by a variable region of light chain (V_L) and heavy chain (V_H) (**Appendix XXXVIII-A**). The antigen binding pocket of the MHC-I molecule (2GTZ, 5WWJ, 6IEX, 5IM7) are formed by the interaction of $\alpha 1$ and $\alpha 2$ domain of α -chain (**Appendix XXXVIII-B**) whereas; the antigen presenting domain of MHC-II (2SEB, 6BIY) molecules are formed by the association of $\alpha 1$ and $\beta 1$ domains (**Appendix XXXVIII-C**). However, the peptide presenting platform is similar for both the MHC class-I and class-II molecules.

The detailed interaction of the chimeric vaccine with the variable region of 7BZ5 is depicted in **Figure 5.4-A** by molecular docking. The linear vaccine from B-cell epitope exhibited twelve hydrogen bond interactions with the variable region of the antibody. The residues Gly38, Gly50, Glu 41, Gly76, Gly54, Thr56, Tyr32, Ser31 of Vac-COVID-B exhibited hydrogen bond with Ser56, Tyr58, Tyr52, Ser53, Tyr33, Tyr94 of V_H and Tyr97, Ala99, Ile59 and Lys58 of V_L respectively. The interaction of conformational epitopes with the 7BZ5 at the molecular level is depicted in **Figure 5.4-**

B, C, D. The conformational epitope R1 (Arg 102, Ser98, Phe140 and Leu242) was found to make hydrogen bond with Tyr100, Tyr32, Tyr 33 and Tyr94 of the antibody molecule (**Figure 5.4-B**).

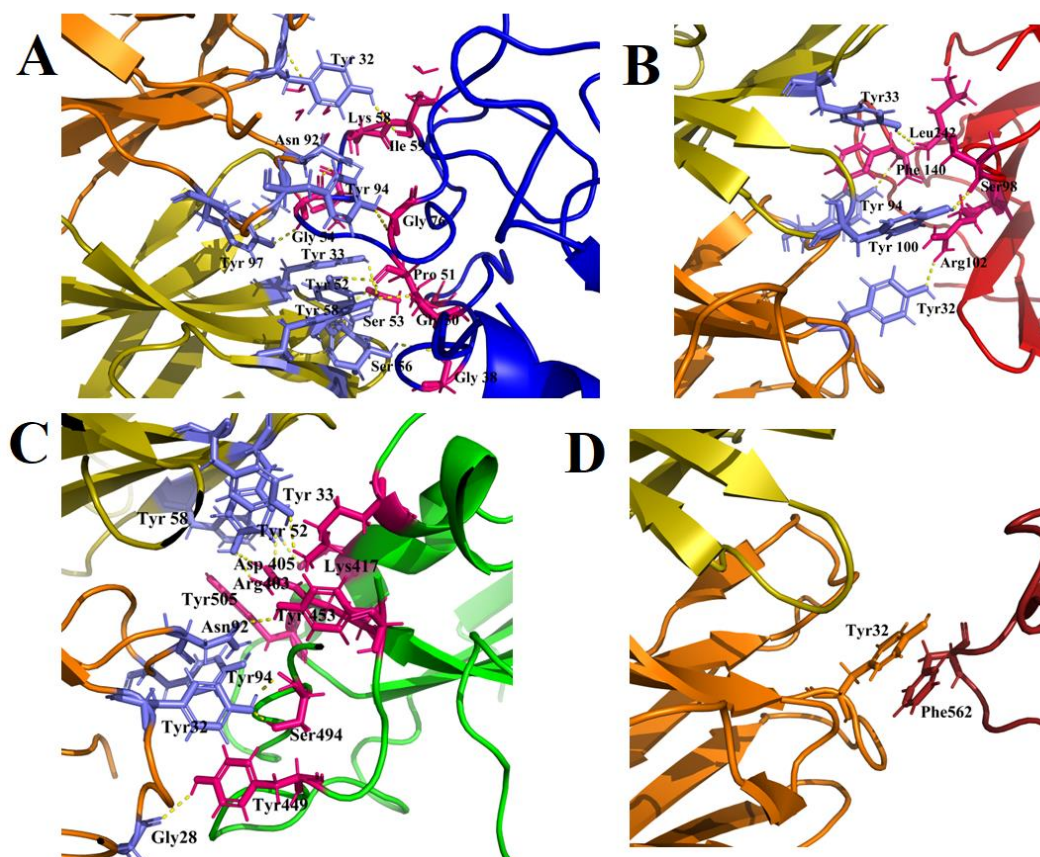


Figure 5.4 The non-bonded interaction of the vaccines and conformational epitopes with the immune receptor. (A) The residues of the linear vaccine involved in the formation of hydrogen bond with 7BZ5. The vaccine is shown in blue colour. The interaction of (B) R1 (red), (C) R2 (green) and (D) R3 (brown) with the variable region of 7BZ5.

Similarly, Gly28, Tyr32, Asn92, Tyr58, Tyr52, Tyr33 of 7BZ5 formed hydrogen bond with Tyr449, Ser494, Tyr453, Arg403, Asp405 and Lys417 of R2 (**Figure 5.4-C**). In addition, Tyr 94 of the antibody and Tyr 505 of R2 showed π - π interaction which is

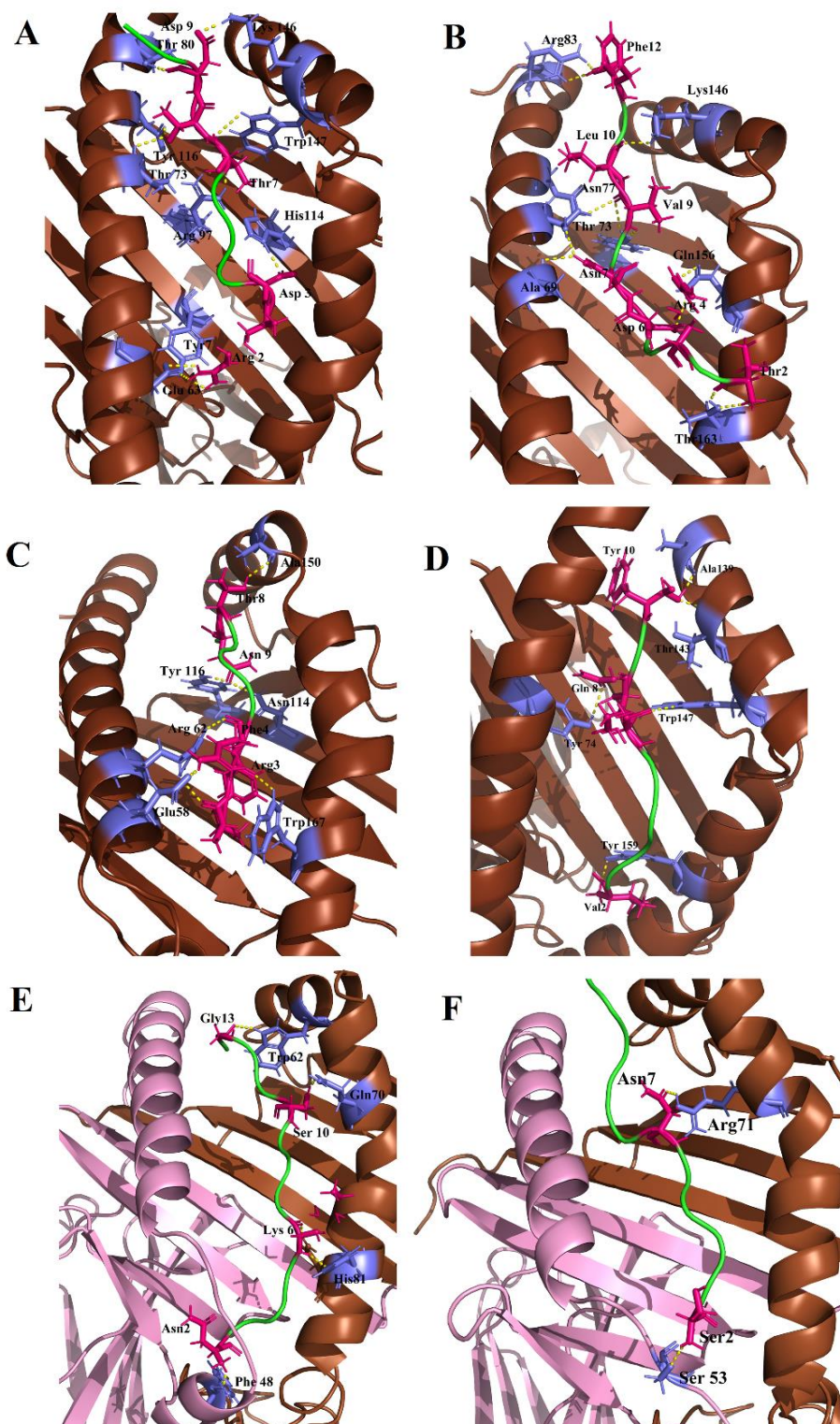


Figure 5.5. The T-cell epitope peptide presented at the peptide presenting groove of (E) HLA-A*02:01, (F) HLA-A*24:01, (G) HLA-B*40:01, (H) HLA-B*58:01, (I)

DRB1*04:01, (J) DRB1*07:01. The interacting residues of the receptor is shown in violate color whereas the residues of vaccine or discontinuous epitope are marked in pink color.

crucial for the antibody-epitope stabilization. In case of R3, no non-bonded interaction was found (**Figure 5.4-D**) except one π - π stacking interaction between Tyr 32 and Phe562. The T-cell peptide epitopes were also found to stabilized by the hydrogen bond formed with the peptide presenting groove of the MHC-receptors. The TCRs are known to recognize the antigens in pieces that are presented by MHC molecules(Punt et al. 2019). Therefore, the interaction of full-length Vac-COVID-T with MHC molecules is scientifically not required. The epitope ⁵⁶⁷RDIADTTDAV⁵⁷⁶ formed hydrogen bond with Thr80, Lys146, Trp147, Tyr116, Thr73, Arg97, Tyr7 and Glu63 of HLA-A*02:01 (**Figure 5.5-A**). ⁷⁶TKRFDNPVLPF⁸⁶ is found to form hydrogen bond with Arg83, Asn77, Thr73, Ala69, Thr163, Lys146 and Trp147 of the HLA-A*24:02 peptide presenting pocket (**Figure 5.5-B**) and in case of HLA*40:01, Ala150, Tyr116, Asn114, Arg62, Glu58 and Trp167 were involved in hydrogen bond formation with the ³²⁷VRFPNITNL³³⁵ peptide (**Figure 5.5-C**).

The epitope ⁷⁸¹VFAQVKQIY⁷⁸⁹ exhibited hydrogen bond with Ala139, Thr 143, Trp147, Tyr74, Tyr159 of HLA-B*58:01 (**Figure 5.5-D**). Further, the MHC-II based T-cell epitope ⁹⁶⁰NTLVKQLSSNFGA⁹⁷², ¹¹⁶SLIVNNATNVVIK¹²⁹ were found to be involved in hydrogen bond with Phe48, His81, Gln70, Trp62 of DRB1*-04:01 (**Figure 5.5-E**) and Arg71, Ser53 of DRB1*07:01 respectively (**Figure 5.5-F**).

5.3.3 MD Simulation Study

The docking predicted interaction of the vaccine-receptor complex in motion were assessed by all atom MD-simulation studies(Elcock et al. 2001; Lemkul and Bevan 2010b; Nazar et al. 2020). In order to check the reproducibility of our MD results we carried out another set of simulation and provided in Supplementary 2 (Figure S20, S21) and 3 (Table S21). The time evolution of the RMSD, RMSF, R_g of B-cell linear vaccine and epitopes located at R1 R2, R3 region are illustrated in **Figure 5.6**. The RMSD of the linear vaccine (Vac-COVID-B) was found to increased approximately 0.7 nm up to initial 8 ns and remain stabilize till 40 ns. The RMSD fluctuation and rGyr

profile of Vac-COVID-B near 60-90 ns fluctuation indicate minor secondary structure modification of Vac-COVID-B during the course of simulation (**Figure 5.6-A & B**).

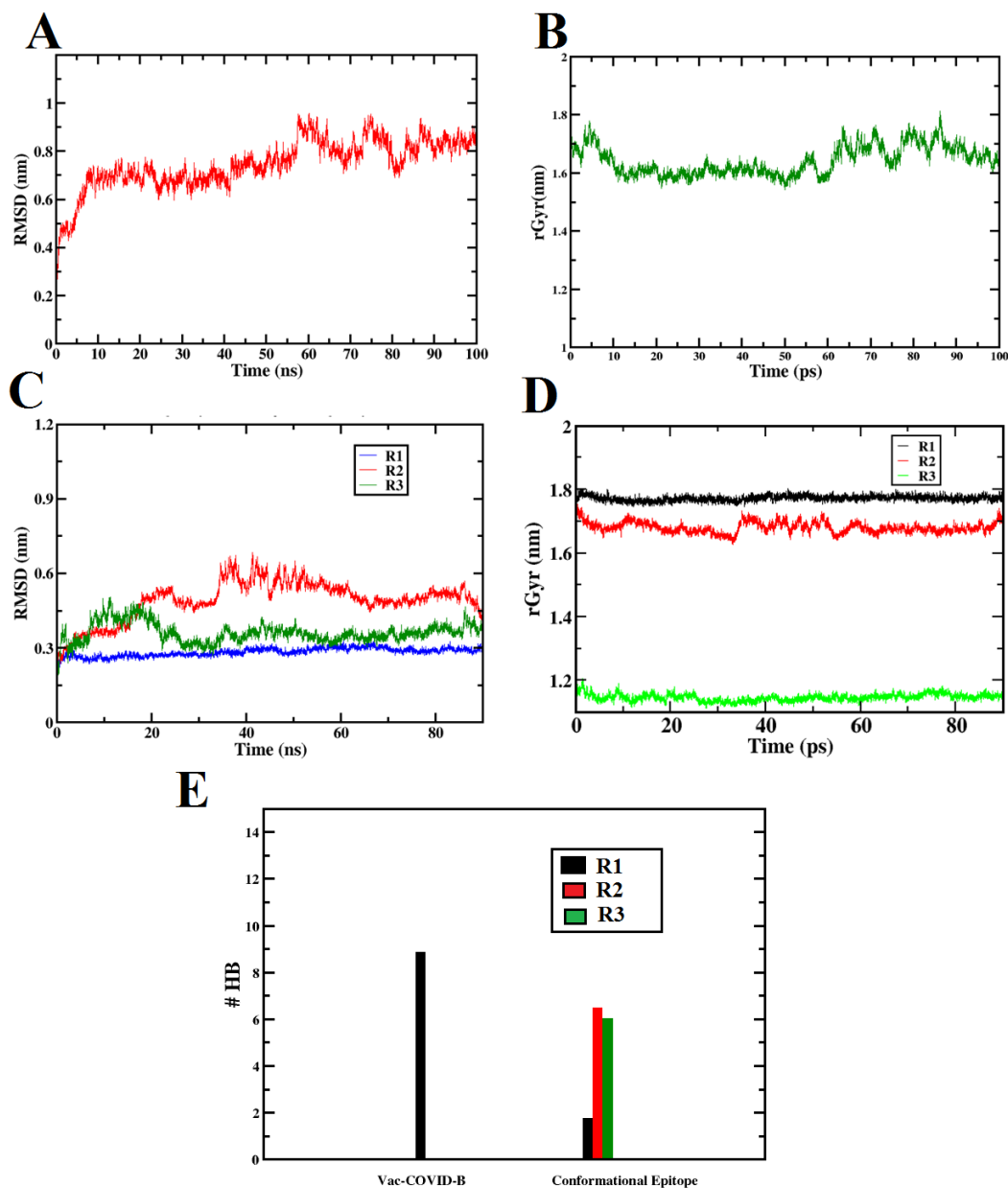


Figure 5.6 (A) The RMSD profile of the chimeric vaccine (Vac-COVID-B) (B) The gyration radius (rGyr) profile of Vac-COVID-B throughout the simulation. (C) The time evolution of RMSD of epitopes located at R1, R2 and R3. (D) The gyration radius of the conformational epitope located at R1, R2, R3. (E) The average number of hydrogen bond between the antibody and vaccine molecules.

The RMSF profile of the designed vaccine is shown in **Appendix XXXIX**. The residue stretches 39-48, 57-73 were found to highly fluctuate during the course of simulation due to the interaction with the antibody molecule. The RMSF value of the residues located at antigen binding domain of 7BZ5 are found to be less than 0.3 which helped the chimeric vaccine to get stabilized at the immune complex (**Appendix XXXX**). The average number of hydrogen bonds between the antibody and the vaccine is calculated to be 9 and remain intact throughout the simulation trajectory (**Figure 5.6-E, Appendix XXXXI-A**). The docking predicted residue pairs Gly54-Tyr97, Gly54-Tyr52, Thr56-Tyr33 have hydrogen bond occupancy of 59.4 %, 13.7%, 11.3%. The highest hydrogen bond occupancy of 69.7%, is found between Tyr58 of the linear vaccine and Pro51 of 7BZ5 which newly evolved during the course of simulation (**Appendix XXXXII-A**). The structural stability of the conformational epitopes located at R1, R2, R3 is assessed in **Figure 5.6 C & D**. The RMSD and radius (rGyr) profile of the conformational epitope located at R1 and R3 is found to be stable; whereas that of R2 is highly unstable which indicates the conformational change of the epitope during the course of the simulation. The C-terminal and the N-terminal end of all the epitopes have greater fluctuation due to solvent exposure (**Appendix XXXIX-B, C, D**). The average number of hydrogen bond between the epitope located at R1, R2, R3 are 2, 7 and 6 respectively. The docking predicted residues pairs were not found to make hydrogen bond during the course of simulation rather, the new hydrogen bonds evolved during the course of simulation with lesser hydrogen bond occupancy percentage for R1 (**Appendix XXXXII-B**). In case of R2, the docking predicted residue pair Asn92-Tyr453 showed highest hydrogen bond occupancy of 99.9% (**Appendix XXXXII-C**). The epitope located at R3, was found to make stable hydrogen bond with Asn32 and Asn92 with H-bond occupancy of 55.9%, 51.6% respectively (**Appendix XXXXII-D**). The hydrogen profile of R1 is fluctuating compared to other conformational epitopes whereas the number of hydrogen bond is found to be increasing between the antibody and R3 with respect to simulation time to stabilize the immune complex (**Appendix XXXXI -B, C, D**). The conformational free energy landscape of the antibody and antigen binding is depicted in **Appendix XXXXIII** to confirm the adequate sampling of the immune complexes. The Solvent accessible surface and RMSD of the chimeric vaccine, B-cell conformational epitopes are considered as reaction coordinate. It is evident from FEL

graphs that, R1 is trapped in a deep minimum which indicates the single binding conformation throughout the simulation. In case of other antigens, two prominent binding conformations along with others are present which evolved during the course of simulation.

Further, the MD simulation of individual peptide epitopes with their corresponding HLA compounds have been carried out. The dynamic nature of the T-cell epitopes with their respective HLA-complex are assessed in the **Figure 5.7**. The RMSD of the peptides correspond to MHC-I increased up ~ 20 ns and stabilized with an average value of 0.45 nm for the rest of the simulation (**Figure 5.7-A**). The peptide associated with MHC-II is found to stable for the last 30 ns of simulation (**Figure 5.7-B**). Further, we calculated the time evolution of the distance between the peptide and the surface of the binding pocket of MHC molecules and showed in **Appendix XXXXIV**. The distance between the surface of the peptide epitope and the platform of antigen presenting domain was found to be stable throughout the simulation indicating that the peptides didn't diffuse from the peptide presenting grooves. The solvent assessable surface area of the peptide epitopes is found to be stabilized with their initial value which confirms the constant solvent exposure at the peptide presenting groove (**Appendix XXXXV**). The peptide associated with HLA-B*40:01 and DRB1*04:01 is found to have less fluctuation in MHC-I and MHC-II based epitopes respectively. MHC-I based peptides mostly fluctuate at their N-terminal end whereas; MHC-II based peptides have higher fluctuation at the C-terminal end (**Figure 5.7 C & D**). The RMSF profile of the binding domain of MHC-I and MHC-II molecules are below 0.3 which helps to stabilize the peptide epitopes inside the peptide presenting pocket (**Appendix XXXXVI**). The average number of hydrogen bond formed between MHC-I based TCRs T-cell epitopes are found to form 4, 8, 6, 7 respectively for the four alleles during 100 ns simulation (**Figure 3.7E**). It is evident from **Appendix XXXXVII** that number of hydrogen bonds were almost constant throughout the simulation time. Hydrogen bond occupancy of 41.2% is found between Ala5 and Thr73 of HLA-A*02:01-RDIADTTDAV complex (**Appendix XXXXVIII -A**).

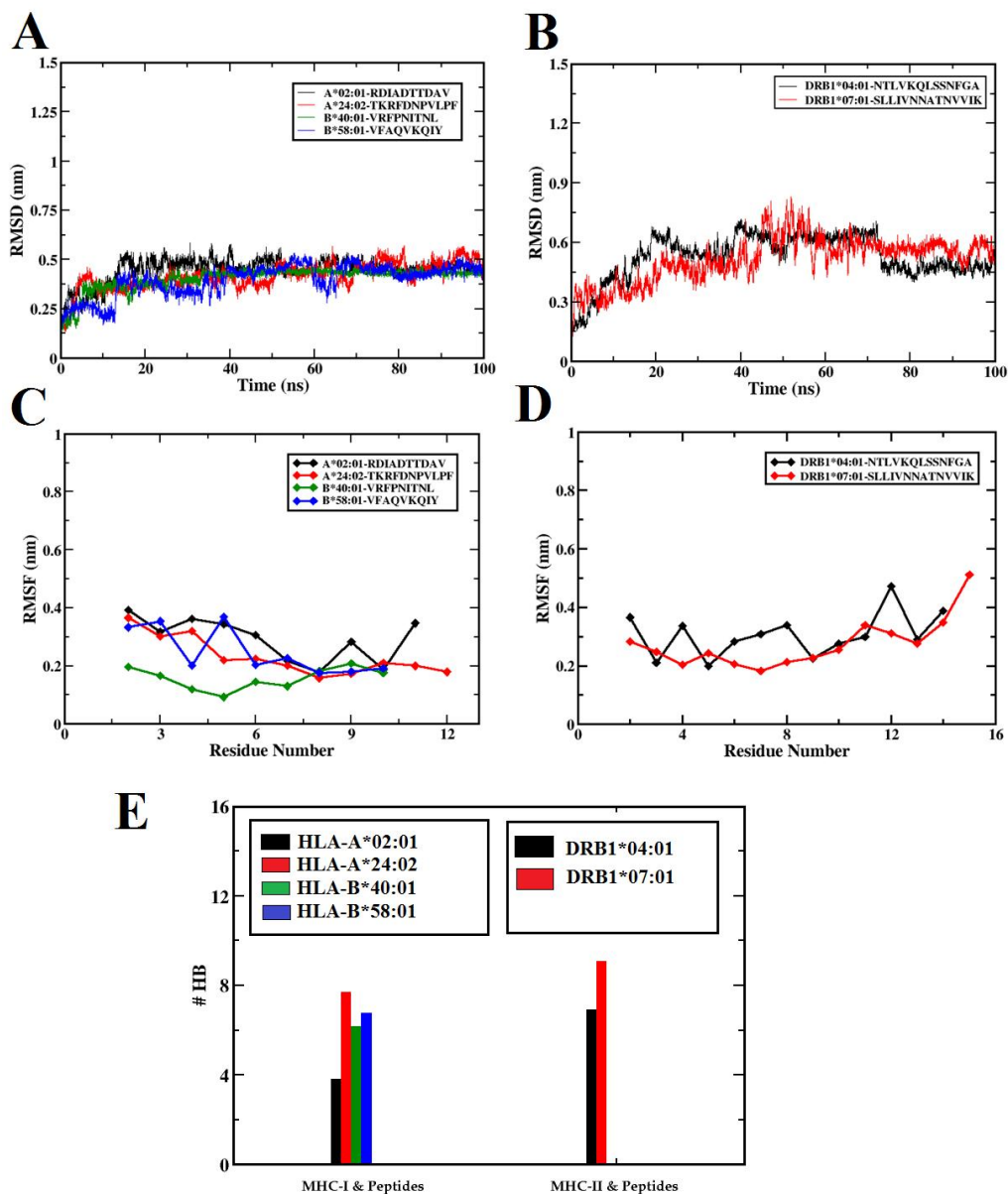


Figure 5.7 The RMSD profile of the T-cell epitope peptides corresponding to (A) A*-02:01, A*-24:02, B*-40:01 and B*-58:01. (B) DRB1*04:01, DRB1*07:01. The RMSF profile of (C) MHC-I and (D) MHC-II based peptide epitopes. (E) The average number of hydrogen bond formed between the peptide epitope and MHC molecules during the simulation.

The residue pair Leu10-Trp147, Phe12-Tyr84 of HLA-A*24:02- TKRFDNPVLPF complex and Pro5-Tyr159, Arg3-Glu63 of HLA-B*40:01- VRFPNITNL have high hydrogen bond occupancy which stabilizes the peptide-ligand at the peptide presenting domain (**Appendix XXXXVIII B, C**). The higher hydrogen bond occupancy of Gln8-

Asn77 (87.8%), Tyr10-Tyr80 (89.9%) Tyr10-Thr143 (83.0%) residue pair stabilized the peptide VFAQVKQIY at the binding domain of HLA-B*58:01 (**Appendix XXXXVIII -D**). MHC-II based peptides epitope were found to have higher number of hydrogen bond due to their higher size (**Figure 3.7-E**). The residue pair Lys6-Asn82, Asn2-Ala52 showed hydrogen bond occupancy of 95.0%, 87.7% for the peptide associated with DRB1*04:01 whereas; Leu4-Asn82, Ser2-His81, Leu3-Ser53 of DRB1*07:01-peptide complex showed high hydrogen bond occupancy (**Appendix XXXXVIII -E, F**). This above-mentioned hydrogen bond between the T-cell epitope peptide and HLA-molecules stabilized the complex to transduce stable immune response against SRAS-CoV-2. The free energy landscape of peptide epitope binding at the peptide presenting pocket of MHC molecules are shown in **Appendix XXXXIX**. It is evident that all the peptides crossed a higher free energy barrier and stabilized at deep minima at the higher RMSD values.

Lastly, the dynamic stability of the designed chimeric vaccines at aqueous solvent were assessed by MD simulation studies (**Appendix XXXXX**). The time evolution of RMSD, radius of gyration and number of native hydrogen bonds were calculated to check their conformation in motion. The RMSD of both the vaccines were found to be converged with an average value of 0.6 nm. It is evident from the gyration radius profile (rGyr) of Vac-COVID-B that compactness of the protein is decreased and stabilized with an average value of 1.5 nm. The compactness of Vac-COVID-T is found to be constant throughout the simulation. It is evident from current literatures that vaccine size of 20-200 nm is most effective in terms of antigen processing and presentation by antigen presenting cells(Bachmann and Jennings 2010). These larger antigens are mostly carried by dendritic cells (DCs) to the lymphatic system and takes 24 hrs for the immune response (Manolova et al. 2008) . However, there are reports which suggest that vaccine size less than ~5 nm is efficient in terms of direct transportation through specialized small antigen conduits to the lymph node for quick (1 hr) immune response (Carrasco and Batista 2007; Singh 2021). Thus, the size of the subunit and chimeric vaccines designed in the current chapter may elicit immune response more efficiently by diffusing through tight lymphatic capillaries to lymphoid organs. Further, the number of native hydrogen bond is found to be constant throughout

the simulation trajectory which indicates the conformation stability of the vaccines in polar solvent.

5.3.4 Free energy calculation by MM/PBSA approach

The free energy of binding (ΔG_{bind}) is believed to be an important thermodynamic quantity to assess the favourable protein-protein interaction as well as their affinity for accurate modelling of biological systems. In this regard, we calculated the binding free energy of the epitopes at the binding domain of immune cell receptors in implicit solvent model by end-point free energy method such as MM/PBSA (Kumari et al. 2014b) and summarized in **Table 5.3**. The designed linear vaccine showed remarkable high binding affinity of -453.59 kJ/mol with the variable region of 7BZ5 which confirms thermodynamically stability of the complex. The ΔG_{elec} term (-469.00 kJ/mol) between the residues of vaccine and humanized antibody found to have important contribution towards the stability of the complex (**Table 5.3**). The affinity of the epitope located at R1, is lowest compared to other discontinuous epitopes due to favourable van-der-Waals energy (-198.54 kJ/mol). The lower affinity of R2 and R3 is due to higher ΔG_{sol} penalty and positive electrostatic interactions respectively. It is found that both the MHC-II based peptides ⁹⁶⁰NLTKQLSSNFGA⁹⁷² and ¹¹⁶SLIVNNTNVVIK¹²⁹ exhibit lowest binding energy of -493.66 kJ/mol and -538.71 kJ/mol respectively. Among the MHC-I peptide epitopes, ⁷⁶TKRFDNPVLPF⁸⁶ associated with HLA-A*24:02 have lowest binding energy of -430.3 kJ/mol. All the peptide epitopes related to MHC-molecules have considerably low Gibbs free energy indicating stable epitope-TCR complex. It is evident that the electrostatic and van-der-Waals free energy terms of T-cell peptide epitopes are crucial for their stability. The hydrogen bond observed during our MD study is mainly responsible for favourable electrostatic energy contribution which is dominating the positive solvation free energy term for the stable interaction of the epitopes at the immune cell receptors.

Further, we calculated the free energy contribution of residues located at the antigen binding domain of immune cell receptors (**Appendix XXXXXI** and **Appendix XXXXXII**). The docking predicted residues and residues with higher hydrogen bond occupancy percentage was found to contribute lower energies which stabilize the immune complexes. In case of chimeric vaccine (Vac-COVID-B), Ile 51 of V_H and Pro

51 of V_L has maximum free energy contribution. The amino acid residues with benzene ring (Phe 27, Tyr 32, Tyr 33, Tyr 58, Tyr 94, Tyr 100) have higher energy contribution due to stacking interactions (**Appendix XXXXXI**). In case of T-cell epitopes, the hydrophilic or polar amino acids located at peptide presenting groove have higher contribution to stabilize the peptide epitopes (**Appendix XXXXXII**).

Table 5.3. Binding affinities (kJ/mol) of the vaccines towards the immune cell receptors by MM/PBSA method. The ΔG_{elec} , ΔG_{vdw} , ΔG_{sol} , and ΔG_{SASA} are indicating the electrostatic, van der Waals, polar solvation, solvent accessible surface energies respectively

Immune Receptor	Complexes	ΔG_{vdw}	ΔG_{elec}	ΔG_{sol}	ΔG_{SASA}	ΔG_{bind}
Antibody	Ab-VAC-COVID B	-145.07	-469.00	345.5	-158.02	- 453.59
	Ab- R1	-198.54	-80.89	73.3	-12.27	-218.4
	Ab-R2	-111.79	-41.15	103.12	-28.95	-78.77
	Ab-R3	-75.258	27.69	37.84	-8.76	-18.52
TCR (MHC-I)	HLA-A*-02:01-RDIADTTDAV	-199.89	-365.33	367.24	-27.98	- 225.95
	HLA-A*-24:02-TKRFDNPVLPF	-340.64	-826.49	779.80	-43.00	-430.3
	HLA-B*-40:01-VRFPNITNL	-328.14	-528.54	602.63	-40.70	- 294.76
	HLA-B*-58:01-VFAQVKQIY	-176.52	-384.67	368.43	-20.62	-213.39
TCR- (MHC-II)	HLA-DRB1*04:01-NTLVKQLSSNFGA	-272.99	-652.81	463.34	-31.19	- 493.66
	HLA-DRB1*07:01-SLLIVNNATNVVIK	-339.73	-608.54	450.9	-41.33	- 538.71

5.3.5 In-silico Design of Recombinant Plasmid

In order to design a recombinant .plasmid, we back translated the protein sequences of the vaccines and optimized the codons(Grote et al. 2005) (Figure S) in *E.Coli* system for successful expression of linear B-cell (Vac-COVID-B) and T-cell (Vac-COVID-T)

vaccines proposed in our immunoinformatic study. The size of c-DNA sequences of the vaccines made from linear B-cell, T-cell epitopes are 399 base pair (bp) and 333 bp long respectively. The codon adaptation index (CAI) value for both the vaccines were computed to be 1.0 and the percentage of GC content for Vac-COVID-B, Vac-COVID-T are 60.65%, 56.75% respectively which are in the permissible range(Sharp and Li 1987) and hence confirmed their proficient expression in E-coli K-12 strain. Finally, the c-DNA sequences of the vaccines were inserted computationally at the multiple cloning site (MCS) of pUC-19 plasmid vector. The restriction map of the recombinant vector is shown in **Figure 5.8**.

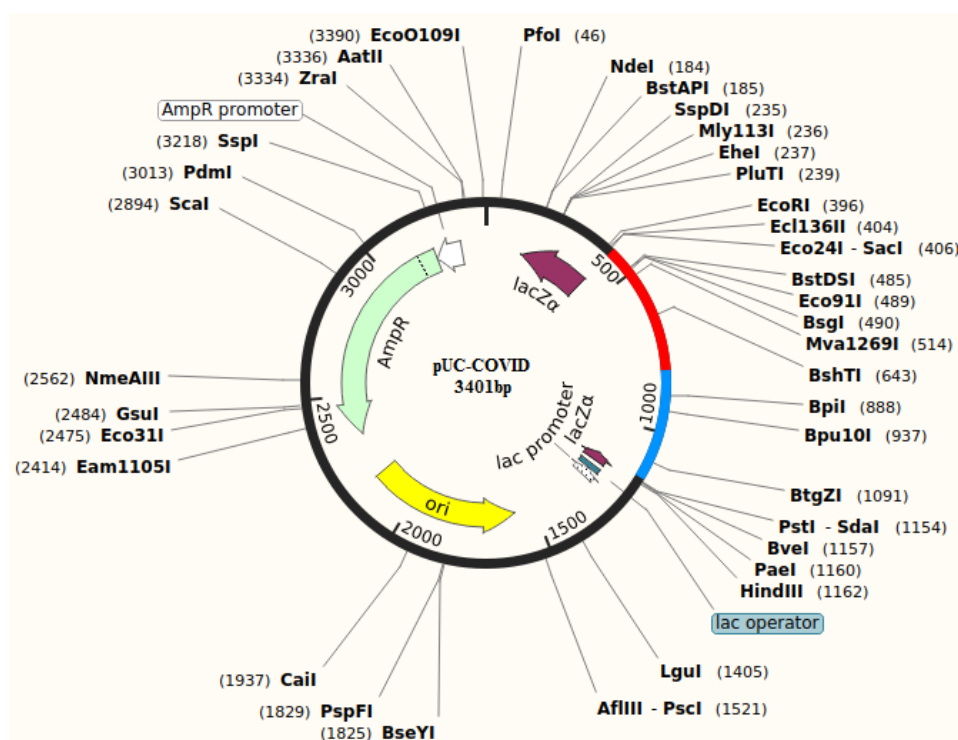


Figure 5.8 Restriction cloning of pUC19 vector. The red and blue coloured portion indicates the codons for the B-cell, T-cell linear vaccine respectively, inserted in pUC19 vector. The rest of the vector part is represented in black colour.

5.4 CONCLUSION

To conclude, in the present chapter we tried to design vaccine based on B-cell and T-cell epitopes present in spike glycoprotein of highly infective SARS-CoV-2. With the help of immunoinformatic studies, we identified the most promising epitopes which are

found to be scattered on the RBD and shaft region of the spike protein. The epitopes located in the shaft region are highly conserved. The hotspot residues that are considered as key to the antigenicity are mostly found to be polar in nature which contribute to stable electrostatic interaction with respective immune receptors. Docking calculations showed the major interaction between the immune-receptor complexes were hydrogen bond and π - π stacking interactions which were found to contribute maximum in the stability, as evident from free energy decomposition studies. The hydrogen bond evolved during MD simulation studies between the antigen and receptors was found to be main contributor of the electrostatic energy. Vaccine designed from linear B-cell epitopes were found to exhibit higher number of hydrogen bonds at the binding domain of antibody compared to the conformational epitopes. All peptide epitopes corresponding to MHC-I & MHC-II showed remarkable stability due to the van der Waals and the electrostatic energy terms. The present article therefore provides deeper biophysical insights towards the stabilization of predicted vaccine candidates with immune cell receptors which will be helpful in further experimental design of potential vaccine against SARS-CoV-2.

CHAPTER 6

EXPLORING THE BARRIERS IN THE AGGREGATION OF HEXADECAMERIC HUMAN PRION PEPTIDE THROUGH MARKOV STATE MODEL

Abstract: *The pre-fibrillar aggregation kinetics of the prion peptides are still an enigma. In this perspective, we employ atomistic MD simulations of the shortest human prion peptide (HPP) (¹²⁷GYMLGS¹³²) at various temperatures and peptide concentrations and apply the Markov state model to find out the various intermediates and lag phases. Our results reveal that the natural mechanism of prion-peptide self-assembly in the aqueous phase is impeded by two significant kinetic barriers with the oligomer size of 6-9 and 12-13 peptides, respectively. The first one is the aggregation of unstructured lower-order oligomers, and the second is fibril nucleation that impedes the further growth of prion aggregates. Among these two activation barriers, the second one is found to be dominant irrespective of increment in temperature and peptide concentration. These lag phases are captured in all three different force field parameters namely GROMOS-54a7, AMBER-99SB-ILDN, and CHARMM36m at different concentrations. GROMOS-54a7 and AMBER99SB-ILDN force fields showed a comparatively higher percentage of β -sheet formation in the metastable aggregate that evolved during the aggregation process. In contrast, the CHARMM-36m forcefield showed mostly coil or turn conformation. The addition of a novel catecholamine derivative (NQDA) arrests the aggregation process between the lag phases by increasing the activation barrier for the Lag1 and Lag2 phases in all the force fields which further validates the existence of these lag phases. The preferential binding of NQDA with the peptides increases the hydration of peptides and eventually disrupts the organized morphology of pre-fibrillar aggregates. It reduces the dimer dissociation energy by -24.34 kJ/mol.*

Keywords. *Prion-peptide aggregation; kinetic barriers; critical nucleus; rate-limiting stage; naphthoquinone dopamine*

6.1. BACKGROUND

Prion diseases or transmissible spongiform encephalopathies (TSE) are the most recent classes of fatal neurodegenerative disorders associated with the deposition of fibrillar, infectious protein plaques in the mammalian brain (Carrell and Lomas 1997; Collinge 2001; Horiuchi and Caughey 1999; Jucker and Walker 2013; Prusiner 1998). The precipitation of such misfolded protein or peptide aggregates gives rise to astrocytosis, synaptic loss, gliosis, progressive dementia, and consequently death within one year of the disease onset (Clinton et al. 1993; Fraser 1993; Gajdusek et al. 1966; Jeffrey et al. 1994). Two significant mechanisms are proposed for prion diseases: genetic (familial) or spontaneous (sporadic). Moreover, infectious prion particles are also reported to transmit from one mammalian body to another. Thus, investigating the molecular mechanisms and effective cure strategy for such severe diseases is an ongoing area of research. Despite significant progress in elucidating the structure and dynamic properties of mature fibrillar aggregates, little is understood about the origin of pathogenicity corresponding to prion disease.

In general, misfolded proteins or peptides are known to be aggregated by two models: (A) “template assembly” or “isodesmic” polymerization and (B) “nucleation polymerization.” In the template model, the misfolded protein or a peptide aggregate acts as a template for the aggregation process and converts normal cellular protein to misfolded form (Bemporad and Chiti 2012; Griffith 1967; Oosawa and Kasai 1962). Such nucleation is also called secondary nucleation (Michaels et al. 2018). Here the free energy for the nucleation is decreased by the use of template or the seeds. In the “nucleation polymerization” or the primary nucleation process, aggregation accelerates after the formation of a stable intermediate, known as the “critical nucleus” (Bishop and Ferrone 1984; Chakraborty and Patey 2013; Ferrone 1999; Jarrett and Lansbury 1993; Yamamoto et al. 2018). The critical nucleus's formation is therefore considered the primary stable energy state (Ferrone 1999). A new model has also been proposed where converting an amorphous nucleus to its ordered form is said to be the rate-determining step that accelerates further growth (Lomakin et al. 1996; Serio et al. 2000). Various experimental and computational studies show that the addition of a critical nucleus in a peptide polymerization reaction significantly reduces the nucleation lag phase of

peptide aggregation (Arosio et al. 2015; Dear et al. 2020; Luiken and Bolhuis 2015; Smit et al. 2017; Törnquist et al. 2018). Hence, the thermodynamics and energetics of the critical nuclei or fibril nucleation at the early aggregation stage are important to formulate strategies to decelerate the initial assembly. It is reported earlier that the formation of mature or organized aggregates is the principal cause of neurotoxicity (Hardy and Allsop 1991; Pike et al. 1991). However, recent evidence supports the fact that peptide conglomerates at the early aggregation stage are the main causative agents which damage the membrane fluidity of the neuronal cells. Therefore, studying the molecular events during the formation of a critical nucleus is of great interest and would eventually pave the path to designing drugs that can inhibit the aggregation process in the nucleation phase.

Experimental characterization of the factors affecting the structural and energetic requirements for the critical nucleus formation is not available till date due to the peptide's random polymeric distribution in solution (Burra et al. 2021). Hence, atomistic MD simulations can bridge the gap by overcoming the practical difficulties of the experimental setup and are used as an important tool to understand the transient nucleation phase in atomic resolution (Carballo-Pacheco and Strodel 2016). The main challenges in simulating the aggregation process of peptides in atomic resolution are associated with computationally expensive simulation time scale and its complexity. Despite the considerable progress in characterizing the pre-fibrillar aggregates for other neurodegenerative diseases, (Carballo-Pacheco and Strodel 2016) such as Alzheimer's, and Huntington's disease, the nucleation process for the prion peptide remains an unsolved topic. Recently, the crystal structure of human-prion protein revealed an important motif (¹²⁷-GYMLGS-¹³²) which has a high propensity to form a two-stranded anti-parallel β -sheet and is believed to be the initiation point of prion disease (Yu et al. 2015a). Although the aggregation of steric zipper-forming small peptides was previously reported, the pre-fibrillar events related to human prion peptides are far from reaching consensus. A recent study showed the formation of five different types of intermediates during the formation of the steric zipper conformation of this motif. Therefore, studying the aggregation process of this motif can serve as a good template for understanding the lag phases and the energetics of the prion aggregation (van der Wel et al. 2007).

In the present chapter, sub-microsecond all-atom MD simulation techniques have been employed to probe the formation of the critical nucleus or the rate-determining nucleation steps that lead to the formation of the pre-fibrillar aggregates of human prion peptide (¹²⁷-GYMLGS-¹³²) by varying the temperature, concentration of the peptides, and addition of NQDA. Changing conditions such as temperature and concentrations helped to accelerate the aggregation process while the use of small molecule inhibitor helped to retard the aggregation process. Different structural and dynamic properties of peptides were analysed to reveal the underlying mechanism of peptide aggregation in pure water in the presence of inhibitor molecules. The aggregation process has been studied for three different force field parameters, namely GROMOS-54a7, AMBER-99SB-ILDN, and CHARMM-36m to understand how the peptide-peptide and peptide-water interaction forces can modify the aggregation process. We also employed simulation studies with a potential inhibitor, naphthoquinone dopamine (NQDA) to confirm the lag phases and the kinetics. NQDA showed success in disintegrating the toxic aggregates of IDPs. Since amyloid disease and prion diseases are cross-related, NQDA can be a potential drug for disassembling the lower-order oligomers of prion peptides. Moreover, targeting the metastable oligomers formed during the aggregation Lag phases can be an alternative way to stop abnormal prion aggregation.

6.2. COMPUTATIONAL DETAILS

6.2.1 Procedure for System Setup:

The 3D crystal structure of the human prion peptide segment ¹²⁷GYLMGS¹³² (PDB ID: 4WBU, 1.15 Å) was obtained from the x-ray diffraction model provided by Yu et al. (Yu et al. 2015b). The monomeric structure of the peptide was made by deleting the extra peptide chain and the crystal water molecules. The aggregation propensity is known to be dependent on the monomer conformation. Therefore, the conformation of the peptide was initially assessed by simulating the monomeric state. The peptide monomers were sampled in the aqueous solution within a cubic box having a buffer space of 1.5 nm from the edge of the peptide. Further, two different simulation systems were prepared. The first setup was done to observe the effect of temperature and peptide

concentration on the oligomer formation during the aggregation of the hexadecameric peptide system. The peptide conformations obtained from the of monomer simulations. were randomly inserted in the simulation box with different initial velocities the by PACKMOL(Martínez et al. 2009) utility. To minimize the biasness of the peptides to aggregate, the inter-peptide distance of 2 nm was maintained in the starting coordinates of all the simulations. The peptide concentrations were maintained by varying the box length and water molecules while the number of peptides remained constant in all the systems. The second setup was done to explore the effect of dopamine-based small molecules i.e., naphthoquinone dopamine (NQDA), on the hexadecameric nucleation of prion peptide. The simulation setup for peptide+ inhibitor systems was like the control one, except for the random addition of required NQDA molecules in the cubic box at 1:2 molar ratio. Furthermore, the fibril formation propensity of the peptides was assessed by the atomistic simulation of mini-fibrils. The initial structure for the mini-fibril simulation was constructed by modelling sixteen peptides in a steric-zipper arrangement. Primarily, the monomers were translated at the x-direction by maintaining an inter-peptide distance of 0.48 nm to make a β -sheet layer which is mutually antiparallel. the two layers of β -sheet are obtained by translating one layer to their parallel manner with an inter-layer distance of 1.012 nm. The step-by-step construction of the mini-fibril is illustrated in Figure S25. In total we carried out 29 different types of simulations to elucidate the dynamism in the early stages of prion peptide aggregation. The detail of the simulation system is shown in **Table 6.1**.

6.2.2 MD protocol:

The MD simulation of all the system mention in Section 2.1 was carried out using the GROMACS-2018.4 package(Abraham et al. 2015b) with the combination of three force field such as GROMOS-54a7 force field (Man et al. 2019b; Schmid et al. 2011b), AMBER99SB-ILDN(Lindorff-Larsen et al. 2010b) and CHARMM-36m (Huang et al. 2017a) which were known to reproduce the spatial and temporal resolution of aggregation prone peptides to assess their motions . It is reported in the literature that the group of GROMOS force-fields favours beta-sheet formation in a nano-second time scale which helps to study the pre-fibrillar aggregation of prion peptides(Best et al. 2008b; Huang et al. 2011). Therefore, the usage of three different force-filed is essential

to assess the similarity or dissimilarity in the formation of the oligomers during aggregation process. The geometry of the small molecule inhibitor (NQDA) was optimized by Gaussian 09 software(Frisch et al. 2009) at B3LYP/6-31G (d, p) level(Sarkar et al. 2019), and the topology, force-field parameters of the inhibitor were generated using PRODRG 2.5 server(Schüttelkopf and van Aalten 2004), CGenFF(Vanommeslaeghe et al. 2010) and ACEPYPE(Sousa da Silva and Vranken 2012b). The partial charges obtained from DFT calculation are replaced by the charges obtained from the abovementioned web-servers.

Table 6.1 Summary of the Simulated Systems. The simulations done in GROMOS, AMBER, CHARMM are designated as G, A, C respectively.

System No.	Temperature(K)	C _{Pep} (mM)	N _{Pep}	N _{Inv}	N _w	Box Length (nm)	Simulation time (ns)
G _M	300	-	1	-	3100	4.56	(200) *5
A _M	300	-	1	-	3093	4.56	(200) *5
C _M	300	-	1	-	3127	4.56	(200) *5
G1	300	10	16	0	88203	13.95	(300) * 2
G2	350	10	16	0	88203	13.95	(300) * 2
G3	375	10	16	0	88203	13.95	(300) *2
G4	300	20	16	0	42659	10.99	(300) * 2
G5	300	50	16	0	16505	7.99	(300) * 2
G6	300	10	16	33 (1:2) NQDA	87712	13.95	(300) *2
G7	350	10	16	33(1:2) NQDA	87712	13.95	(300) *2
G8	300	20	16	32 (1:2) NQDA	42186	10.99	(300) *2
G9	375	10	16	32 (1:2) NQDA	88203	13.95	(300) *2
G10	300	19.21	20	0	56039	12.0	200
G11	300	20	26	0	71094	13.0	200
G12	300	22.72	30	0	70949	13.0	200
G13	300	50	52	0	55171	12.0	(250) *2

G14	300	-	2	0	1319	3.5	(1000) *2
G15	300	-	2	4 (1:2) NQDA	1184	3.5	(1000) *2
A1	300	25	16	0	28930	9.6	300
A2	320	25	16	0	28930	9.6	300
A3	330	25	16	0	28930	9.6	300
A4	300	25	16	0	34226	10.15	300
A5	300	40	16	0	21118	8.67	300
A6	300	50	16	0	16444	7.99	300
A7	300	30	16	32 (1:2) NQDA	28447	9.6	300
C1	300	45	16	0	18606	8.26	300
C2	320	45	16	0	18606	8.26	300
C3	330	45	16	0	18606	8.26	300
C4	300	30	16	0	28930	9.59	300
C5	300	40	16	0	21117	8.63	300
C6	300	50	16	0	16443	7.98	300
C7	300	50	16	32 (1:2) NQDA	15977	7.97	300
G _F	300	95.04	16	0	8614	6.53	300
A _F	300	95.04	16	0	8570	6.53	300
C _F	300	95.04	16	0	8564	6.53	300

The subscript M and F indicates monomer and fibril simulation respectively. 'C_{pep}, N_{pep}, N_{Inv}, N_w indicate the concentration of peptide and number of peptide monomers, inhibitor respectively. The number after * indicates the number of simulation replicas carried out for each system.

SPC/E water molecules(Mark and Nilsson 2001c) were used to solvate the prion peptide systems due to their excellent compatibility with the GROMOS force field(Moučka et al. 2013). Recent reports support that the modelling of protein-drug, protein-protein, and protein-membrane interactions in the aqueous environment gives a better correlation with experimental conditions when the SPC/E water model is

chosen for the GROMOS force field(Lima et al. 2019; Moučka et al. 2013) . In case of AMBER-99SB-ILDN and CHARMM-36m TIP3P water model was used. To remove bad contacts between solute-solute and solute-solvent molecules, all the systems underwent a 10,000-step energy minimization using the steepest descent algorithm(Jaidhan et al. 2014). The Leapfrog algorithm(Van Gunsteren and Berendsen 1988) was used to integrate the equation of motions with a time step of 2 fs. LINCS algorithm(Hess et al. 1997b) was applied to constrain all the bonds in the peptide molecule, and SETTLE algorithm(Miyamoto and Kollman 1992b) was employed to constrain the geometry of water molecules. The system's temperature was gradually heated from 0K to the required temperature limit for 500ps in canonical ensemble (NVT) and further equilibrated up to 5ns. Then the system condition was changed from constant volume to isothermal-isobaric ensemble at 1 atm pressure. In this condition, the cell volume is allowed to fluctuate isotropically to reach the appropriate density for 10 ns. The thermodynamic equilibrium of the systems was ensured by calculating the temperature, density, and the potential energy of the systems during the 10 ns NPT ensemble and depicted in **Appendix XXXXXIII**. During equilibration simulations, the heavy atoms of the solutes were restrained by applying external force (force constant 1000 kJ/mol/nm²). Next, the restraints were removed from solutes, and production simulations were carried out in an NPT ensemble for 300 ns. Short-range electrostatic and van der Waals interactions cut-off were set to 1.4 nm. The Particle Mesh Ewald method(Shamshirgar et al. 2017) was used to calculate the long-range electrostatic interactions. The temperature and pressure of the system were maintained by employing Velocity rescale (Bussi et al. 2007b)($\tau_t=0.1$ ps) and Parrinello-Rahman coupling algorithm(Parrinello and Rahman 1981b) ($\tau_p=2$ ps). The coordinates were saved at every 100 ps and were considered for analysis, and VMD was used for visualization purposes. Taken together, the current article provides the analysis of 9.6 μ s trajectory data in the production simulation phase.

6.2.3 Umbrella Sampling:

Lastly, to assess the stability of the smallest aggregate, i.e., dimer in the presence or absence of NQDA, we employed the umbrella sampling method(Kästner 2011b) and calculated the potential of mean force (PMF) for dimer dissociation. The dimers formed

in the peptide aggregation simulations (with or without NQDA) were taken as the starting configuration for umbrella sampling simulations. The distance between the peptide's centre of mass (COM) was considered the reaction coordinate (RC) for plotting the PMF. The relative distance between the monomers is known to govern the aggregation and non-bonded interactions; thus, the RC mentioned above can be considered a suitable measure to determine the energy barrier for the dissociation process (Lemkul and Bevan 2010c). The rate of pulling or pulling velocity for the dissociating peptide was assigned to 0.007 nm/ns. The entire pulling simulation was divided into symmetric windows with a distance interval of 0.1 nm between the COM of the peptides. A production simulation of 100 ns was carried out after 5 ns equilibration in each window. The unbiased PMFs were then reconstructed by Weighted Histogram Analysis Method (WHAM) (Kumar et al. 1992b; Meng and Roux 2015). The histograms related to each umbrella window are presented in **Appendix XXXXXIV** to ensure the overlap between the adjacent windows.

6.3. RESULTS AND DISCUSSION

The present study focuses on the dynamics related to the pre-fibrillar aggregation and the aggregation lag phases of the shortest fibrillogenic human prion peptide. The Lag phases can be defined as the rate-determining steps or the slowest steps that are crucial for the aggregation process. This involves metastable aggregation stages that take a longer time to achieve some favourable conformations during the successive addition of monomers or lower-order peptide clusters (dimer to tetramer) in the self-assembly process. The influence of different system parameters such as (i) temperature, (ii) peptide concentration, and the effect of dopamine-based inhibitor molecule (NQDA) was studied. The addition of NQDA in the prion peptide aggregation process will help us to study the rate-determining steps or the major aggregation barrier that affects the aggregation process. Further, the influence of atomistic force fields on pre-fibrillar amyloidosis pathways of prion peptides was carried out to check different oligomeric intermediates. In the subsequent sections, we first explored the conformations of the monomers to understand the propensity of the aggregation in each force field and further move forward to study the oligomerization of HPPs.

6.3.1 Conformational states of monomer and influence of force fields:

The distribution of torsion angles (ϕ and ψ) corresponding to each residue were calculated and plotted into two-dimensional distribution graphs and presented in **Appendix XXXXXV**. The distribution of these two conformational degrees of freedom (ϕ/ψ) related to the terminal (Gly-127, Gly-131, and Ser-132) and middle residues (Tyr-128, Met-129, and Leu-130) were separately plotted for understanding the conformations of prion peptide monomers in three different force fields. It is believed that the ϕ/ψ values between -180° , -100° / -0° , 45° related to alpha helix; the ϕ/ψ values populated at -121° / $+128^\circ$ or -66° / 137° correspond to beta-sheet and the torsion angles related to random coils -50° / -40° or -120° / 130° (Hovmöller et al. 2002). Additionally, the distribution of ϕ/ψ angle related to any residue parallel to the diagonal ($\psi=-\phi$) is believed to be in the unstructured state. It is evident from **Appendix XXXXXV** that the terminal residue Ser and middle residues such as Tyr, Met, and Leu have less probability of forming alpha-helix in the GROMOS-54a7 force field compared to the other two force fields. In contrast, all the force fields showed adequate distribution of torsional angles corresponding to the beta-sheet region for middle residues and Ser. It can be noted here that the distribution of torsion angles of prion peptide residues is more parallel to the $\psi=-\phi$ line in the case of the CHARMM-36m force field (**Appendix XXXXXV-B**). Therefore, the CHARMM-36m force-field has more propensity to sample random coil or twisted structures of human prion peptide compared to the other two force-fields considered herein. Further to have more insights into the transition of monomeric conformations in three different force fields, we carried out Markov State modelling. First, the sample densities from five independent simulations (each of 200 ns) have been plotted to describe the transition between discretized meta-stable states. It can be found that the number of basins in the case of GROMOS is less compared to the other two force-fields. However, in the case of AMBER force field, the basins are well separated in sample space. Further for constructing MSM, best-suited feature or variable was selected by calculating VAMP-2 scores (variational approaches towards Markov processes). It is evident from **Appendix XXXXXVI** that backbone torsion angles (ϕ/ψ) have more advantage in preserving the kinetic variance. Further K-means clustering and implied time scale (ITs) plots were performed on the selected feature to build MSM with optimum lag time (τ). It can be found in **Appendix XXXXXVII** that

the lag time of 2.0 ns is efficient in building MSM with five microstates for GROMOS-54a7, and six microstates for AMBER-99SB-ILDN and CHARMM-36m force field due to the convergence of ITs at this lag time. Apart from the convergence of ITs, the goodness of the model was validated by the CK (Chapman-Kolmogorov) test depicted in **Appendix XXXXXVIII-XXXXXX**.

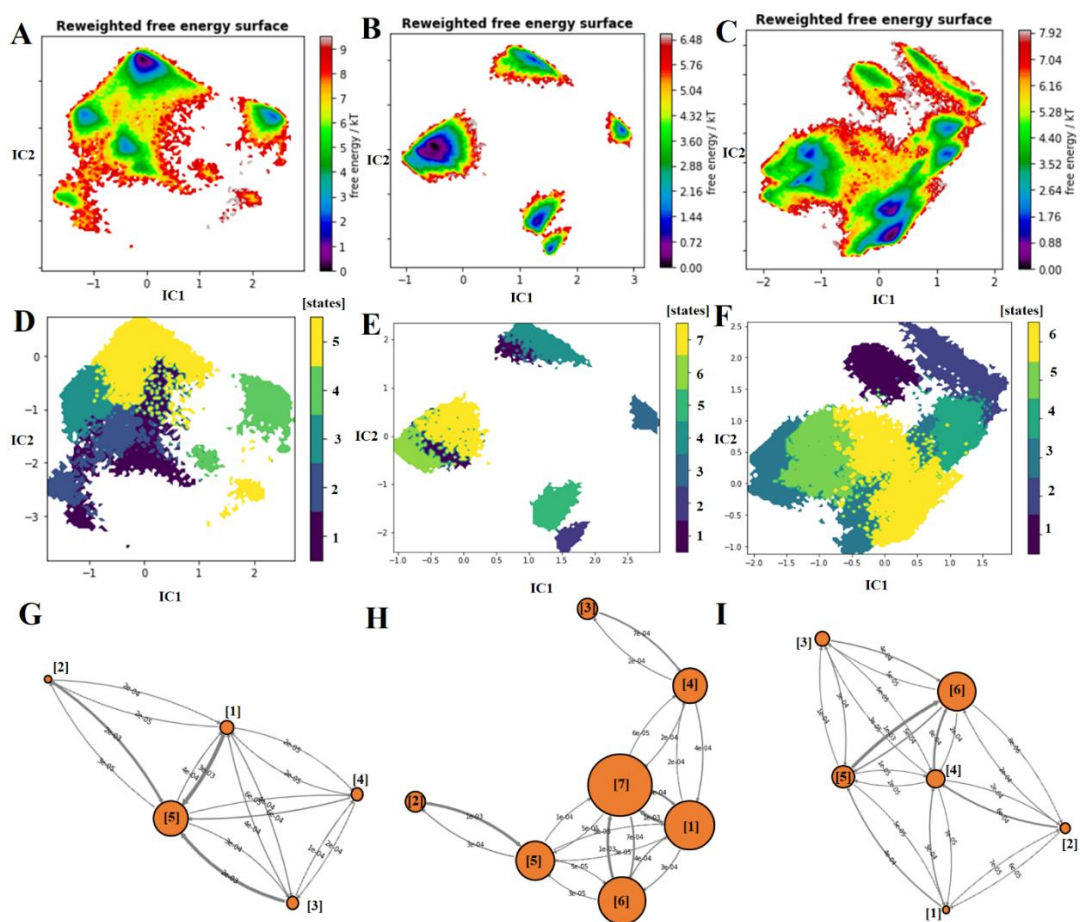


Figure 6.1. The free energy landscape of the prion peptide monomer based on the ϕ/ψ angles in (A) GROMOS-54a7, (B) AMBER99SB-ILDN, and (C) CHARMM-36m force field. The population of different conformational states on the free energy landscapes of prion peptide correspond to (D, E, F) three different force-filed. The different colors have been used for the discretization of microstates. (G, H, I) The transition network (TN) of the prion monomer using ϕ/ψ angles as descriptors.

The MSMs are represented as networks where nodes correspond to the population of discrete peptide conformers and edges represent the transitions respectively. It can be found from **Figure 6.1-G** that the State5 (74.88% population) in the GROMOS-54a7 force field is attractive compared to other states due to the inflow of peptide conformations from states 1, 2, and 3. The representative structure of each microstate and corresponding contact maps are shown in **Appendix XXXXXXI**. The intra-peptide residues are said to be in contact if the distance between them falls below 0.45 nm. It can be found that inter residue contacts in state 5 are parallel to the diagonal which indicates the extended state of the peptide favourable for making β -sheet in the presence of other copies of the peptides. Unlikely, the other two force-fields found to sample 3 to 4 different states with higher population of different peptide conformations. In the case of AMBER99SB-ILDN state 1, 5, 6, and 7 are found to be highly populated. In state 1 (population 32%) the peptide structure is found to adopt the hairpin type of structure with higher contact probability between Y128, M129 and G131 (**Appendix XXXXXXII**). Similarly in state 5, the higher contact between G127 and M129 is found which leads to turn peptide conformation. In states 6 and 7 the peptide is found to be in the extended state. Therefore, the AMBER force-field has the propensity to sample both twisted and extended state of prion peptide. In case of CHARMM-36m forcefield Y128, L130, M129, S132 (state 5) contacts and G127, Y128, L129, G130 (state 6) contacts are higher which stabilizes the peptide in twisted conformation (**Appendix XXXXXXIII**). It can be found that the CHARMM-36m force field yields high intra-peptide side chain interactions compared to the other two force-fields. Furthermore, the diffusion coefficient of the peptide monomers was calculated (**Appendix XXXXXXIV**) and the prion peptides are found to be highly diffusive in case of CHARMM force-field which may hinder the formation of higher oligomers during the self-assembly process.

6.3.2 Aggregate Size Distribution and Effect of NQDA in three different force-fields:

To quantitatively identify the intermediates of prion peptide aggregation, the time evolution of the largest aggregate size is monitored throughout the simulation time scale (**Figure 6.2**). A cluster or aggregate is said to be formed if the distance between the C_{α}

of two peptide copies is within 0.8 nm. The graphs in **Figure 6.2** illustrate the spontaneous formation of dimer or lower-order intermediates (dimer to tetramer) at the initial aggregation phase in all the three force-fields. It can be noted that such lower-order aggregates are readily formed and do not depend much on the variation of temperature and peptide concentration. Nevertheless, the peptide self-assembly process is found to be lagged in two long-lived oligomeric states in all the force-fields. Multiple associations and dissociation events between the peptides are found during each lag phase in all temperatures and peptide concentrations. This indicates the reordering of the largest oligomeric aggregate during the two lags to attain the stable stage before further nucleation. The addition of peptides to higher oligomers occurred initially by sidechain interaction and later they conformationally evolved by mainchain or backbone interaction. It can be seen in **Appendix XXXXXXV** that initially the side chain contacts increased and the gradually main chain contact happened. In the case of the GROMOS-54a7 force field, no dissociation events are found after reaching the final aggregation state (16 peptides in one cluster) except 375K (**Figure 6.2-A, B**). However, in case of AMBER-99SB (**Figure 6.2-C, D**) and CHARMM-36m (**Figure 6.2-E, F**) force-field the dissociation events are comparatively higher at higher oligomeric state in all temperatures as well as peptide concentration due to enhanced reordering of peptides within the cluster. It can be observed that the prion peptides are fully aggregated at lower concentration (20 mM at 300 K) for GROMOS-54a7 force field compared to the AMBER (30 mM, 300 K) which is again lower than CHARMM force field (50 mM, 300 K). This is due to the change in the intra-peptide and peptide-water interactions in different force-fields. In GROMOS, the peptide-peptide interactions are found to be more compared to the other two force fields (Samantray et al. 2020). This observation depicts that change in force-fields can elicit different pathways of peptide aggregation.

Firstly, the aggregation process is found to undergo a short lag phase (*Lag 1*) with an aggregate of 7-9 peptides in the GROMOS-54a7 force-field (**Figure 6.2-A, B**). In AMBER-99SB-ILDN and CHARMM-36m forcefields the aggregates size corresponds to *Lag1* is found to oscillate between 5-7 peptides and 6-8 peptides respectively. This state can be compared with the peptide-rich disordered phase that are

metastable intermediate which serves as a precursor for the nucleation process. A recent experimental report by Burra *et al* showed the formation of the critical nucleus occurs with a preferential population of aggregates containing ~ 7 members (Burra et al. 2021) for yeast prion peptide containing hydrophilic residues which have comparable length with the HPP. Therefore, the formation peptide aggregate size corresponding to *Lag1*

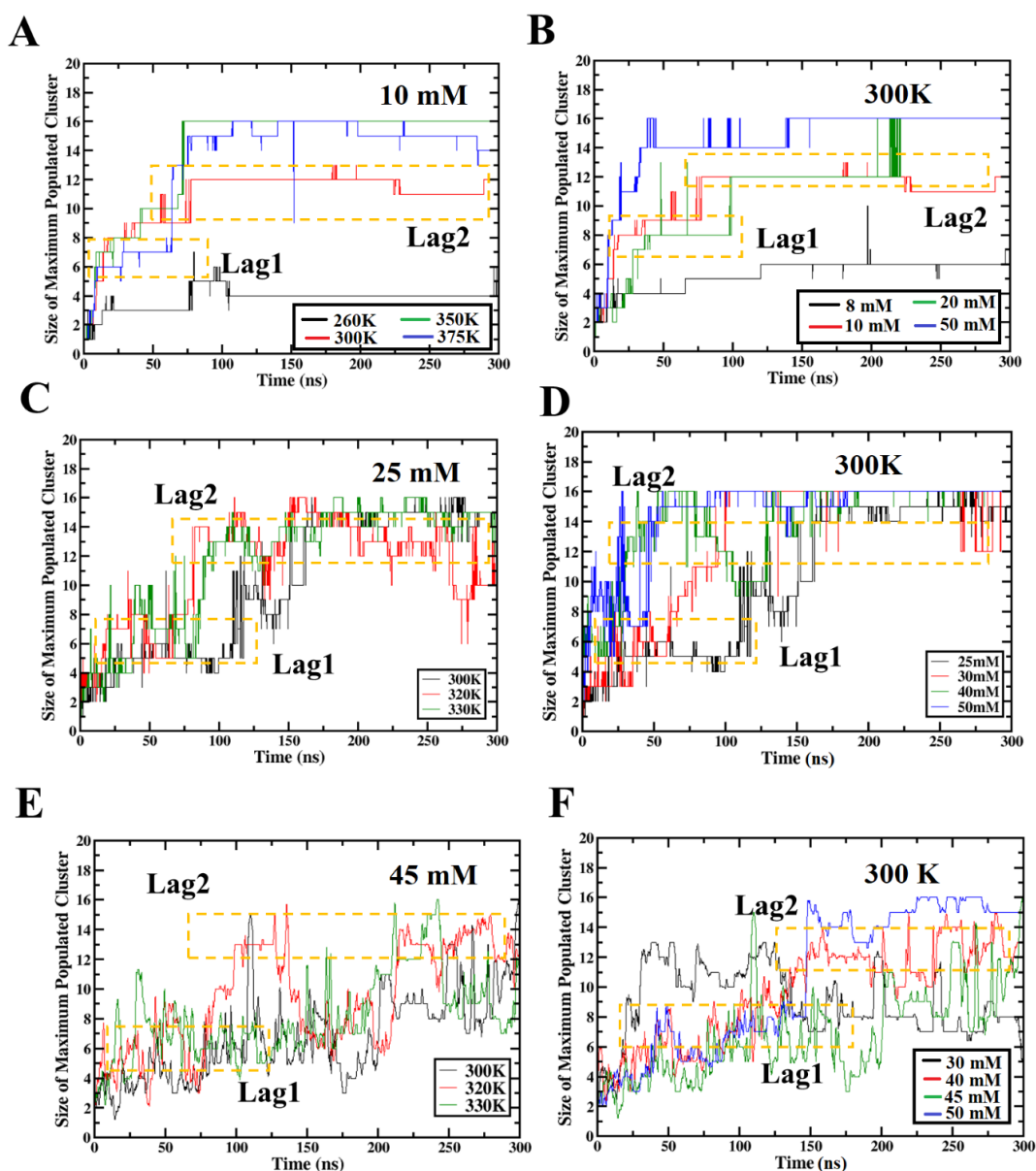


Figure 6.2. The time evolution of maximum populated cluster size at different temperature, peptide concentrations for (A, B) GROMOS-54a7 force-field, (C, D) AMBER-99SB-ILDN force-field, (E, F) and CHARMM-36m force-field.

in all the force fields can be considered as an important step for the primary aggregation barrier in self-assembly process. In the GROMOS-54a7 force-field the existence of *Lag1* is also noted at higher temperatures (such as 350K and 375K), but the time span is comparatively lesser, as seen at 300K (**Figure 6.2-A**). Similarly, at 20 mM peptide concentration (300K), the *Lag1* phase is also found (**Figure 6.2-B**) with an aggregate size of eight peptides. However, at a very high peptide- concentration (50 mM), the *Lag1* almost disappears. The disappearance of the *Lag1* at high concentration is mainly due to the increased probability of the effective collision of the peptides which helped them to aggregate faster. This shows the shifting of the two-step mechanism to one-step mechanism. A similar scenario is envisioned in case of the AMBER and CHARMM force-fields. The time span of *Lag1* is higher in AMBER and CHARMM force-field compared to the GROMOS force field.

The second lag phase (*Lag 2*) is observed with a cluster size of 12-13 peptides for GROMOS-54a7 and 12-14 peptides for AMBER, CHARMM force-field respectively. The *Lag2* is found to be successive to *Lag1* due to the cooperative addition of smaller oligomers like trimers and tetramers to the nucleus of *Lag1*. It can be noticed here that the increment in temperature and peptide concentration causes the fusion of two critical nucleus oligomers to give rise to such peptide cluster in all the force-fields. The time scale for *Lag2* is larger than *Lag1* (irrespective of temperatures and monomer concentrations) due to greater reorganization events occurring before further nucleation. This reorganization event is more prominent in AMBER and CHARMM force-field at different temperatures and peptide concentrations. Such reorganizational event is also prominent in the post-nucleation stage (Tang and Han 2022a). Therefore, the *Lag2* can be considered as the rate-determining stage for the present aggregation process. At 300K and 10 mM peptide concentration, the aggregation process is found to be trapped in the *Lag2* phase in case of GROMOS-54a7 force-field (**Figure 6.2-B**). This is further validated by the time evolution of the cluster size at 310K, where the aggregation process is trapped at the higher aggregate size containing 12 peptides (**Appendix XXXXXXVI**). Similar observation is found in the case of AMBER and CHARMM force-field at 25, 40 mM peptide concentration (**Figure 6.2-D, F**) respectively. Therefore, the addition of a small size oligomer (dimer or trimer) to the

critical nucleus or the fusion of the rate-limiting higher-order oligomer remains the main bottleneck for the further growth of the aggregates. The existence of *Lag* phases with the aggregate size of a similar number of peptides in all the force-fields shows the statistical robustness of our findings (**Figure 6.2**). Further, to validate the existence of lag phases we calculated the time evolution of peptide clusters during the simulation time scale (**Appendix XXXXXXVII**). It can be seen that in AMBER and CHARMM force field the peptide aggregate associated and dissociated between peptide number 12-14 and 13-15 respectively. The aggregate size in the second lag phase did not decrease lesser than 12 and 13 respectively. Therefore, the aggregate size for second lag phase can be taken as 12 -13 for all force fields. The dissociation of peptides from the oligomers correspond to *Lag2* is found in all the force field at higher temperature. It is found that AMBER and CHARMM force-field the dissociation of peptide is prominent in higher temperature compared to higher peptide concentration. It is observed that at moderately higher temperature, the peptide aggregated at lower concentration. For GROMOS, aggregation took place at 10mM, 350 K. For AMBER, the aggregation took place at 25mM at 330 K. For CHARMM force field, we found at 45mM aggregation took place at 320 K. No aggregation was achieved at this concentration for 300 K (**Figure 6.2-A, C, E**). The minimum concentration that was found to aggregate at 300 K was found to be 20 mM, 30mM and 50mM for GROMOS, AMBER and CHARMM respectively (**Figure 6.2-B, D, F**). At higher temperature, the peptide dissociation rate was found to be more. Therefore, at higher temperature, even though the kinetic barrier is overcome by higher collision rate, full aggregation is hampered due to the dissociation tendency of the peptide. This can be further clarified from **Figure 6.3**.

At moderate higher temperature such as 350 K for GROMOS, 330K for AMBER, and 320K for CHARMM, we found that the peptides aggregated at lower concentrations for all the three force fields. This can be explained by the increase in the rate of collision rate which increases the probability of aggregation. In **Figure 6.3**, the dependence of the interacting distance between two peptides with respect to temperature and peptide concentration for different force fields has been plotted.

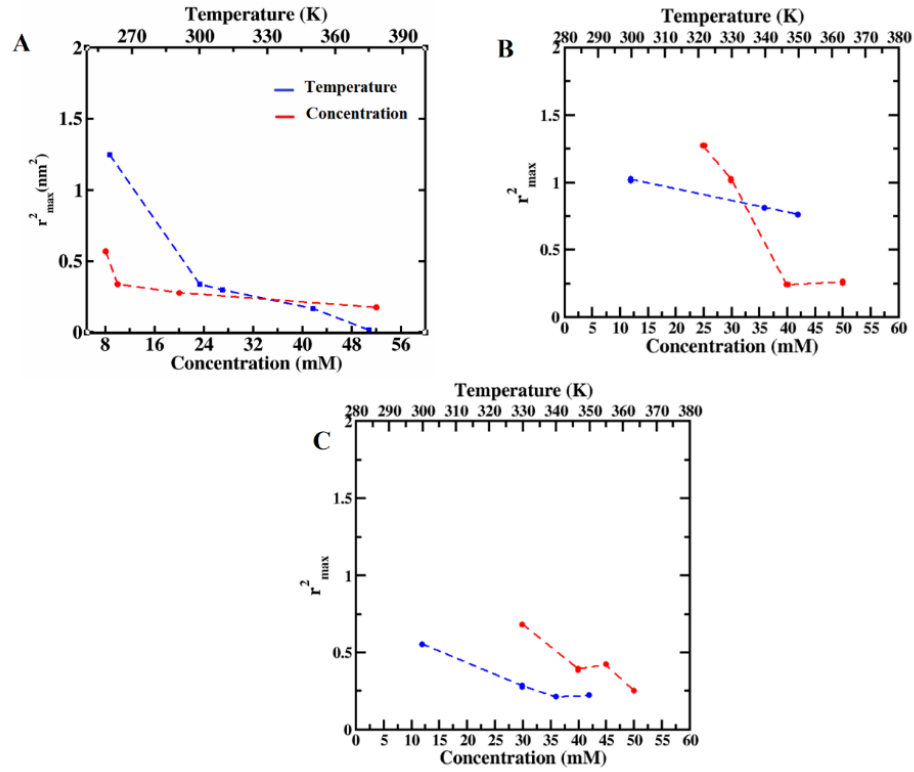


Figure 6.3. The dependency of the mean free path of the prion peptides with the temperature and peptide concentration in the (A)GROMOS-54a7, (B) AMBER-99SB-ILDN, (C) CHARMM-36m.

We have considered the trajectory after initial 1 ns of the production simulation until the formation of the first dimer to calculate the mean free path of the prion peptides. The peptide collision propensity is believed to be proportional to the r_{max}^2 or square of the mean free path defined as (Adebayo et al. 2010)

$$f = 4\pi\rho r_{max}^2 g(r_{max}) \left(\frac{\pi k_B T}{m}\right)^{1/2} \dots\dots\dots (6.1)$$

Where ρ is the number density of peptides, $g(r_{max})$ is the value of pair correlation function at r_{max} , m is the total mass of the peptides, k_B is the Boltzmann constant, and T is the absolute temperature.

It is found in GROMOS-54a7 force-field that the interacting distance (r_{max}^2) between the monomers decreases with increase in temperature and peptide concentration. A similar trend is found for AMBER and CHARMM force-field.

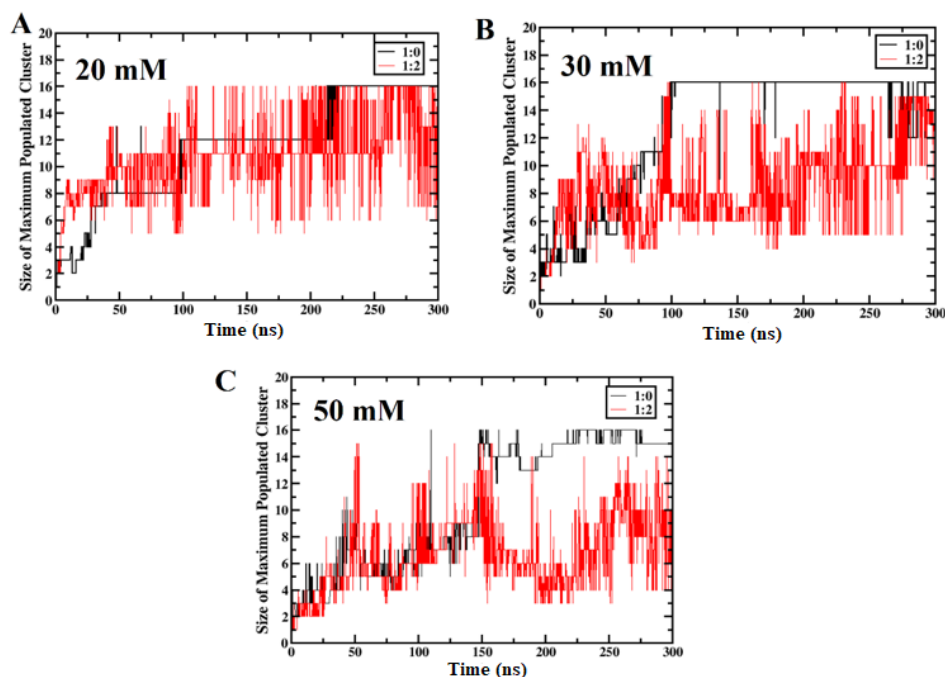


Figure 6.4. The time evolution of maximum populated cluster size at the absence or presence of NQDA at 1:2 Peptide to NQDA molar ratio in (A) GROMOS-54a7, (B) AMBER-99SB-ILDN, (C) CHARMM-36m at 300K.

In total, the whole observation suggests that the entire pre-fibrillar aggregation process corresponds to two major lag phases with relatively stable two oligomers. The activation barrier corresponding to *Lag2* is higher than the formation of the critical nucleus. More importantly, mutual reorganization between the peptides at the *Lag 2* phases is an inevitable event to attain β -sheet conformation essential for initiating fibrillar growth.

Our simulation study is mainly concentrated on the pre-nucleation stages. To check whether these lag phases still exist in the higher system size, we simulated additional simulations shown in **Appendix XXXXXXVIII** with more than 16 peptides (system G10, G11, G13, G14). The details of these simulations are given in **Table 6.1**. It is found that *Lag1* and *Lag2* still exist at this higher concentration. However, they are overcome easily in comparison to the lower concentration. Further, the growth of the aggregation took place by addition of the fibrillar oligomers (specifically 12-14 peptides). We have checked this case with multiple simulations (**Appendix XXXXXXVIII-B**). This step can be referred to the Ostwald ripening phase where the

smaller aggregates join to form the bigger aggregate. A study of the post-nucleation stage of Amyloid- β (16-21) have been reported where it is found that at lower concentration of monomer, the rate of aggregation is concentration dependent but at higher concentration, the rate becomes dependent on the structural conversion to β -sheet (Tang and Han 2022b).

The addition of NQDA in the prion peptide self-assembly process will further help us to reconfirm the activation barriers and the critical factors affecting the aggregation process. It can be observed that there is a remarkable impact on the time scale of the aggregation lag phases and the structural ordering of the aggregates in the presence of NQDA. It can be seen from **Figure 6.4-A** that the aggregation process is hindered with a maximum populated cluster size of 8-10 peptides and remains in the same stage up to 300ns. In this condition, the maximum populated cluster size is found to be highly fluctuating and even reduced to the oligomer size of five peptide members during the course of the simulation. This phenomenon is observed for 20mM 300K for the GROMOS force field (**Figure 6.4-A**), 30 mM 300K for AMBER force field (**Figure 6.4-B**) and 50 mM 300 K for CHARMM forcefield (**Figure 6.4-C**) in the presence of NQDA. Notably, the maximum populated cluster size is found to be similar in all the force-fields. Thus, the addition of NQDA increases the activation barrier for the formation of the critical nucleus or higher-order aggregates and eventually arrests the process of fibril nucleation. Therefore, it can be commented here that NQDA delayed the aggregation process between *Lag1* and *Lag2* despite increasing the system concentration. The inhibition of peptide aggregation by NQDA can be comparable with the aggregation process found in lower temperature (260K for GROMOS 54a7) and lower peptide concentration (8 mM for GROMOS-54a7). In both cases, the aggregation process is significantly slower due to the increment in the activation barrier at first *Lag1*.

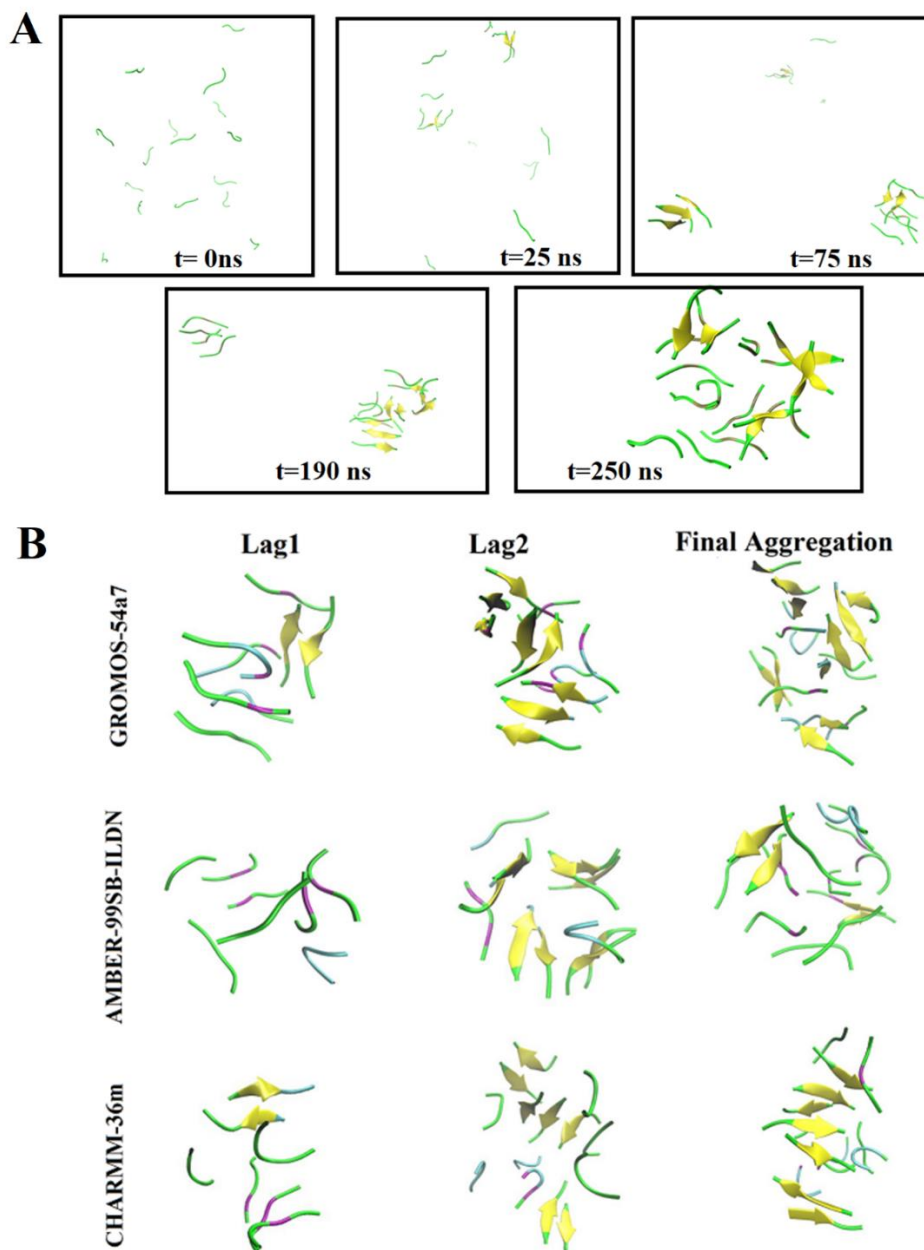


Figure 6.5. (A) MD snapshots of the prion peptide self-assembly at the prefibrillar stage in GROMOS54a7 force field [20 mM, 300K]. (B) The secondary structure arrangement of the metastable peptide aggregates. (B) The structure of aggregates in the presence of Inhibitor in three different force-fields. The secondary structures such as β -sheet, β -bridge, and coil structures are colored in yellow, olive green, and green color respectively.

6.3.3 Secondary Structure Analysis:

The appearance of β -sheet in the aggregation process is believed to be the hallmark of toxicity. In **Figure 6.5-A** and **6.5-B** we have shown the simulation snapshots of the peptide aggregation process and different secondary structural arrangements at two Lag phases and the final aggregation state correspond to three force-fields. Further, we assessed the percentage of secondary structure in three different force fields and their behavior with the temperature and concentration have been assessed in this section and depicted in **Figure 6.6** and **Appendix XXXXXXIX**. It is evident that the coil percentage or the percentage of residues involved in disordered state disordered are found to vary from 60-70% in all the force-field. However, there is a remarkable change in the β -sheet percentage for GROMOS, AMBER and CHARMM forcefields during the self-assembly of prion peptides. It is found that the propensity of β -sheet formation is more in the case of GROMOS force field in different temperature and peptide concentration. At 300K and 310K, the average percentage of β -sheet is 13.8% and 12% respectively in pure water. At higher temperature and peptide concentration, the β -sheet content is found to be gradually increased and reached to ~ 21% at higher temperature (375K) and peptide concentration (50 mM). In case of AMBER force-field the β -sheet content found to vary from 3-11% at different temperature. The increment of bend and turn content is found to be higher in AMBER force-field which can be the reason behind the lesser β -sheet content in peptide aggregates. Additionally, the appearance of α -helix is found at higher temperature in this force-field which causes a dip in the β -sheet content. The beta-sheet content was found to increase linearly with respect to peptide concentration. In CHARMM force-field the β -sheet content found to be increase to 14% at 330K and 18% at 50 mM in room temperature. This observation indicates that temperature increment helps in re-ordering of the metastable aggregates for the appearance of β -sheet during the aggregation process. Furthermore, the formation of the extended β -sheet and cross β -sheet (steric zipper) was found in 1 μ s simulation of 50 mM 300K for GROMOS force-field which confirms the fibrillar aggregation of these prion peptides and is further confirmed experimentally by Nelson et al (Nelson et al. 2005b) (**Figure S15-A,B**).

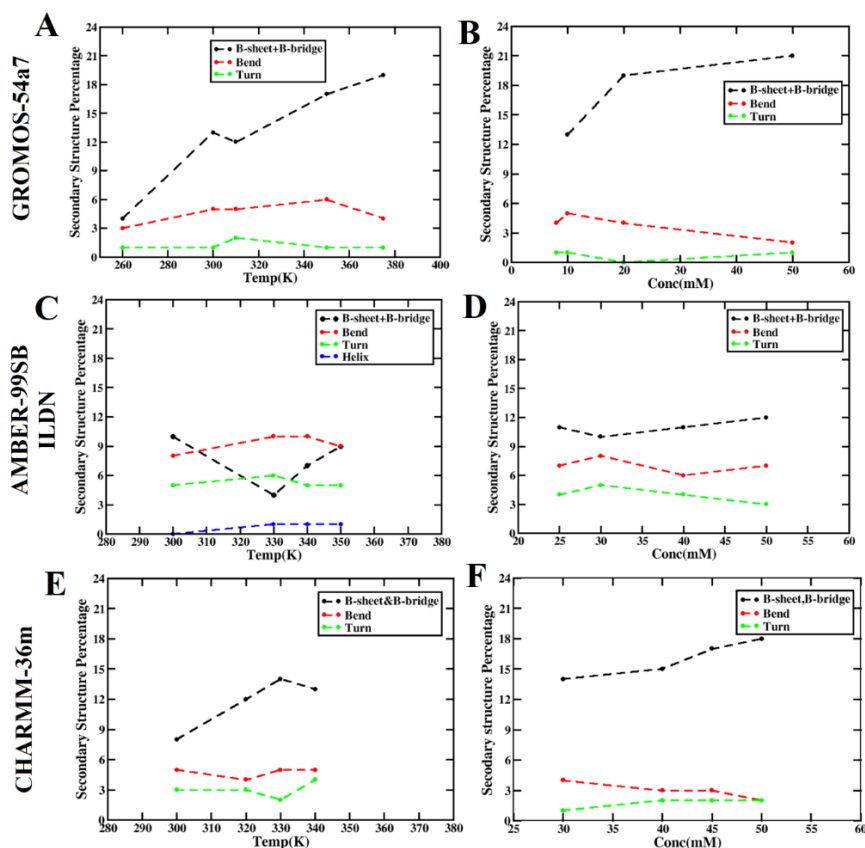


Figure 6.6. The secondary structure content of the peptide aggregates at different temperature and peptide concentration in (A, B) GROMOS-54a7, (C, D) AMBER-99SB-ILDN, (E, F) CHARMM-36m force-field.

It can be found from **Appendix XXXXXXXX A, B** that the distance between the parallel sheets of steric zipper is 10.5 \AA and the distance between two beta sheets is 4.67 \AA (perpendicular to the fibril axis) which matches with the literature. During $1 \mu\text{s}$ simulation the aggregate was found to be stable and did not dissociate into smaller aggregates as shown in **Appendix XXXXXXXX-D**. In **Figure 6.5-B**, the structure of the intermediate steps corresponding to *Lag1*, *Lag2* and the final aggregate has been provided for all the three force-fields. It can be found that the peptide cluster corresponding to *Lag1* is amorphous in nature and mostly contains bridge- β structure. In case of GROMOS and CHARMM force-field antiparallel β -sheet dimer is found in critical nucleus oligomer. Further, the appearance of ordered oligomers is found in *Lag2* phase for all the force-filed. This indicates that *Lag2* correspond to fibril nucleation state where fibrillization starts. In case of AMBER force-field the β -sheet content is

comparatively lesser in *Lag2* stage compared to other force-fields. It can be noted here that β -sheet content is comparatively similar in *Lag2* and final aggregation state for all the force-field.

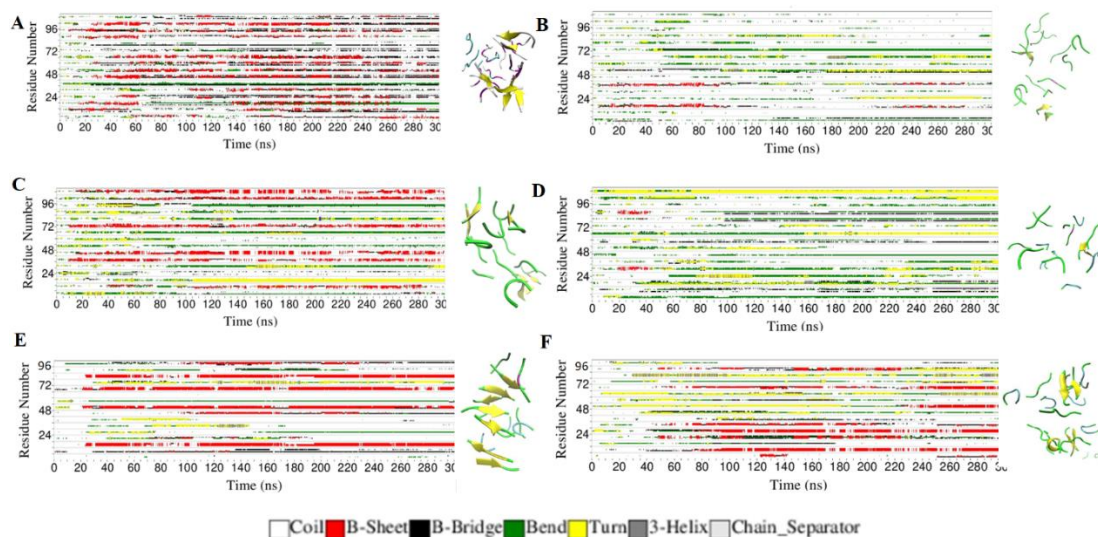


Figure 6.7. The time evolution of secondary structure elements throughout the simulation time scale in absence or presence of NQDA and the final peptide aggregated structure for (A, B) GROMOS 54a7, (C, D) AMBER 99SB ILDN, (E, F) CHARMM 36m force-field.

Further, to corroborate the formation of fibril in the *Lag2* phase or final aggregation state, stability of minifibrill is assessed and depicted in **Appendix XXXXXXXXI**. The details of Minifibrill construction for initial MD coordinates is provided in the methods section. It can be found that after 1 μ s the minifibrill conformation highly deviated from it's initial state in all three force-fields and some of the peptides are in unstructured state with reactive hydrogen atoms for further growth of the self-assembly process. However, the proto-fibrils are found to be stabilize more in GROMOS and AMBER force-field in barrel or zipper conformation compared to CHARMM due to higher interaction of inter peptide residues. This is evident from the time evolution of solvent accessible surface area and inter-peptide hydrogen bond shown in **Appendix XXXXXXXXII**. The SASA profile of the minifibrill is found to be more fluctuating in case of CHARMM 36m force-field which indicates instability of the minifibrill compared to other two force-fields. Further, we calculated the occupancy of inter

peptide hydrogen bond and documented in **Appendix XXXXXXXXIII**. It can be noted from **Appendix XXXXXXXXIII** that mini-fibril is highly stabilized by more backbone hydrogen bond in AMBER 99SB ILDN force-field compared to other two force-fields and the middle residues such as Leu, Met and Tyr are mostly contributing hydrogen bond that governs β -sheet conformation in the self-assembled peptide aggregates. The 2D contact matrix given in **Appendix XXXXXXXXI** validates the role of middle residues to form stable fibrillar aggregates. These interactions are more prominent in case of AMBER 99SB ILDN and GROMOS 54a7 forcefield compared to CHARMM 36m.

The addition of NQDA found to decrease the β -sheet content in the peptide aggregates remarkably for all the three force-fields. The secondary structure evolution and structure of the aggregates in presence of NQDA for the three chosen force-field is shown in **Figure 6.7**. In case of GROMOS force field the beta-sheet content reduced from 19% to 3% in 20 mM peptide concentration at 300K. Similarly, the β -sheet and β -bridge content reduce to 1% and 3% during peptide self-assembly for AMBER forcefield. In CHARMM force-field the β -sheet content of peptide aggregates is 6% in presence of NQDA. This reduction of β -sheet is prominent compared to the aggregates formed in pure water for CHARMM force-field. These results indicate that the incorporation of NQDA shifts the dynamic equilibrium of aggregates from extended β -sheet to random coil state essential for reducing the toxicity.

6.3.4 Aggregation free energy landscape:

To further visualize the effect of force-fields and inhibitor molecules on the lag phases and the energy barriers of the peptide aggregation process the free energy surface have been constructed. The number of inter-peptide hydrogen bonds and SASA are taken as the reaction coordinate (RC). In the present case, these two order parameters have been considered as RCs due to their suitability to explain the complete aggregation process. Boltzmann inversion of conformational distribution(Zhou et al. 2019b) was employed to compute the FES based on the following equation.

$$\Delta G_i = -k_B T [\ln P_i - \ln P_{max}] \quad (6.2)$$

In **Figure 6.8**, we compared the property-based aggregation free energy surface (FES) of prion peptides at 300K for different force fields such as GROMOS-54a7, AMBER-99SB-ILDN and CHARMM-36m.

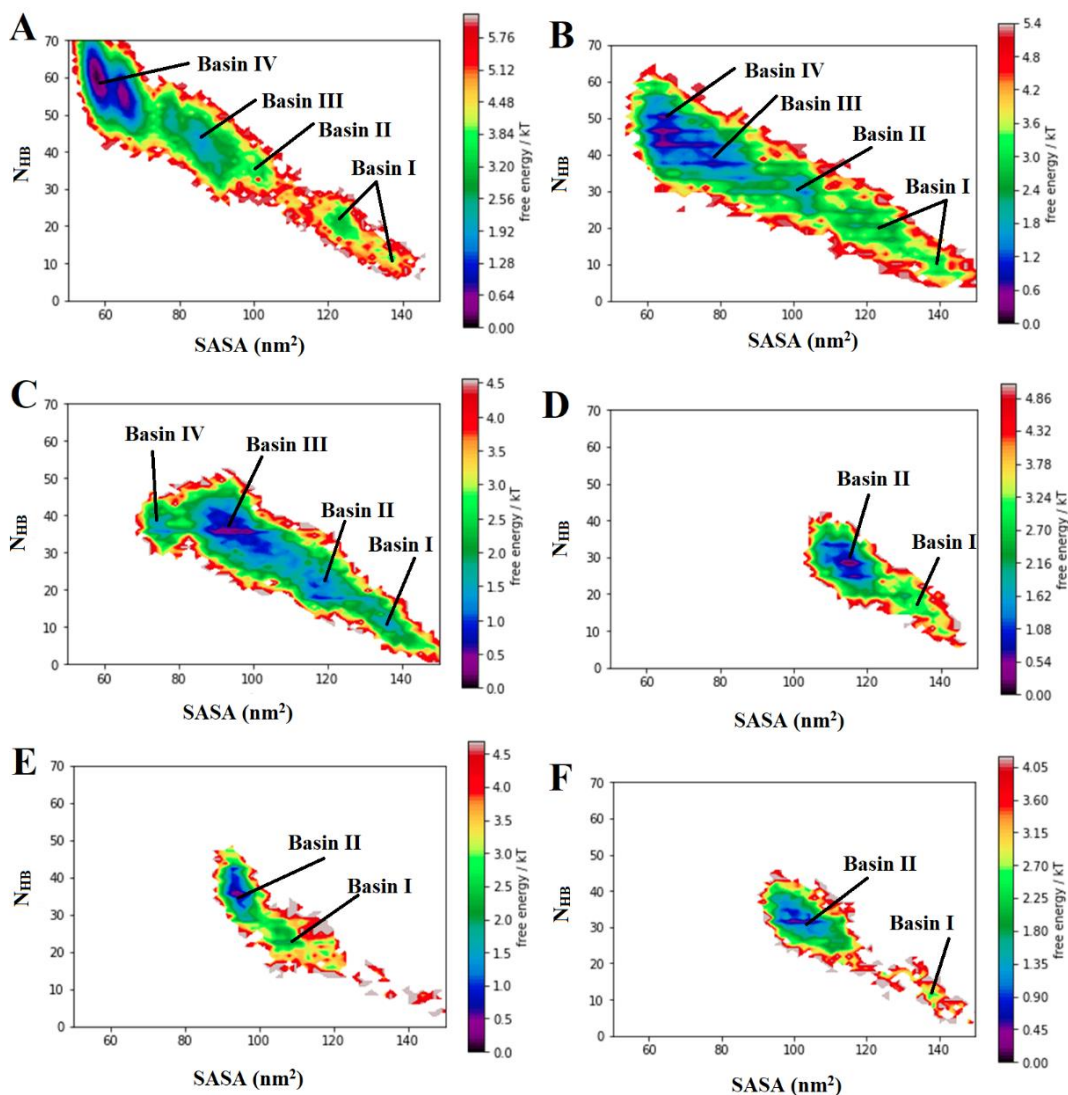


Figure 6.8. Free energy surface involving the number of interchain hydrogen bonds versus solvent accessible surface area in (A) GROMOS-54a7 [20 mM] (B) AMBER-99SB-ILDN [30 mM], (C) CHARMM-36m [50 mM] at 300K. The aggregation energy landscape in presence of NQDA for (D) GROMOS-54a7, (E) AMBER-99SB-ILDN and (F) CHARMM-36m force field.

Four major basins have been found for all the force-fields. The locations of four basins are marked on the FES and are shown in **Figure 6.8** to identify the metastable

intermediates between the two extremes (thoroughly segregated and fully bound) states of conformational space. In case of GROMOS-54a7 the four basins are found to be well separated with thermodynamic barrier. Basin I is mainly populated with lower-order intermediates like monomers or dimers, characterized by higher SASA value and a lower number of hydrogen bonds. A large population of peptide oligomers up to critical nucleus size is located at basin II. Further, Basin III is mainly populated with the oligomer size corresponding to the second lag phase where fibril nucleation occurs at 300K. At this temperature, in GROMOS-54a7 force-field basin II and III are separated by a small energy barrier of 1.4kT, which further reduced at AMBER-99SB-ILDN and CHARMM-36m force-fields. However, the aggregation state corresponding to basin III at 300K is again trapped within a deep minimum in case of CHARMM (~ energy barrier of 2.8kT), which restricts the formation of the final aggregation state. The final aggregation state is found to be separated by barrier height of 3.84kT. Further, the presence of NQDA is found to have a prominent effect on the free energy surface of prion peptide aggregation. It can be seen from **Figure 6.8-D, E, F** that the FES shifted to a higher SASA value and lower number of inter-peptide hydrogen bonds in all the force-fields, which indicates the population of lower-order intermediates. Moreover, in presence of NQDA, the lower-order peptide aggregates are found to be stabilized at a deep minimum in basin II with a barrier height of ~4KT. Consequently, the aggregation process is restricted to reaching the free energy basin corresponding to *Lag2* or the final aggregation state. Therefore, it can be inferred from the above results that a significant energy barrier exists at the *Lag2* phase, which determines the peptide aggregation rate. Additionally, the presence of NQDA increases the activation barrier of prion-peptide aggregation and the concentration of lower-order oligomers, which ultimately inhibit the overall aggregation process between the two lag phases.

6.3.5 Accessing aggregation kinetics by Markov State model:

Further to assess the aggregation kinetics of prion peptide at three different force-fields, Markov state modelling has been applied. This algorithm is efficient in identifying the metastable aggregates and the probable transition path from the initial state to final state. The collective variables extracted from the simulations of different peptide

concentrations at 300K were merged to ensure adequate population of different metastable peptide aggregates.

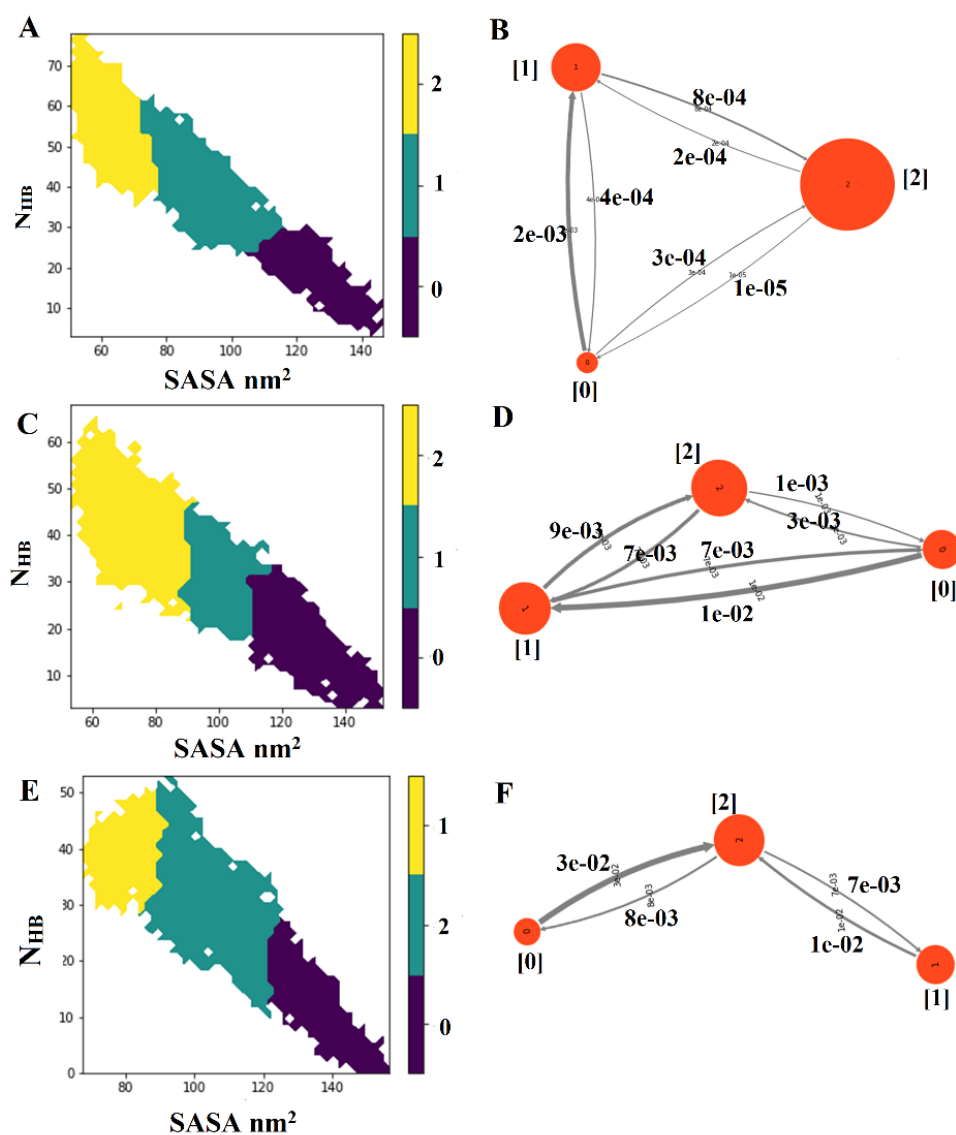


Figure 6.9. The decomposition of metastable states and corresponding transition paths of prion peptide aggregation for (A, B) GROMOS-54a7, (C, D) AMBER-99SB-ILDN and (E, F) CHARMM-36m force-fields. The values written on the edges depict the transition rate between the microstates.

This strategy helps in identifying metastable aggregates that co-exist in dynamic equilibrium. The first step toward finding the microstates is the application of K-means clustering on the free energy landscape of aggregation process. Further, the MSM based on the microstates is subjected to estimate the lag time for hidden Markov state model (HMM) depending on the convergence of implied time scales correspond to slowest process. It can be found from the **Appendix XXXXXXXXIV** that the implied time scale (IPS) plots are converged at 50-time step for all the force-fields. The mean free passage time (MFPT) analysis also carried out validates the existence of the three states as observed in **Figure 6.9**. The time step or lag time obtained from implied time scale is chosen to coarse-grain the MSM model to HMM. The transition matrix obtained from HMM is depicted as network diagram where the edges and their sizes are representing the population in each microstate. The thickness of the lines connecting the edges indicates the transition probability from one state to another. The Markovianity of the MSM was ensured with Chapman-Kolmogorov test (**Appendix XXXXXXXXV**). In the present case three different metastable state have been observed from the MSM calculation taking SASA and inter-peptide hydrogen bond as collective variable. The [0] state correspond to the lower-order oligomer state, [1] correspond to the critical nucleus and fibril nucleation state and [2] correspond to final aggregation state (**Figure 6.9**). It is evident from the MSM model of peptide aggregation that the transition of smaller order oligomer to the critical nucleus is faster compared to the transition of fibril nucleation to final aggregation state. This further validates the formation of fibril nucleation is the slow or rate-determining step in peptide aggregation. Further, we have carried out Markov state model for prion peptide aggregation in the presence of NQDA. The validation for the MSM is given in **Appendix XXXXXXXXVI** and **Appendix XXXXXXXXVII**. It can be seen from **Figure 6.10** that in the presence of NQDA the peptide aggregates have two states and the second state or the critical nucleus is mostly populated one. In case of AMBER99SB-ILDN force-field, the population of state1 is comparatively lesser than two other force-fields. Notably, the inflow of conformation is mostly from state1 to state2. However, in case of CHARMM force-field the inflow of conformations is same from two states. This indicates the continuous transition of critical nucleus to monomeric sate or vice versa in presence of NQDA.

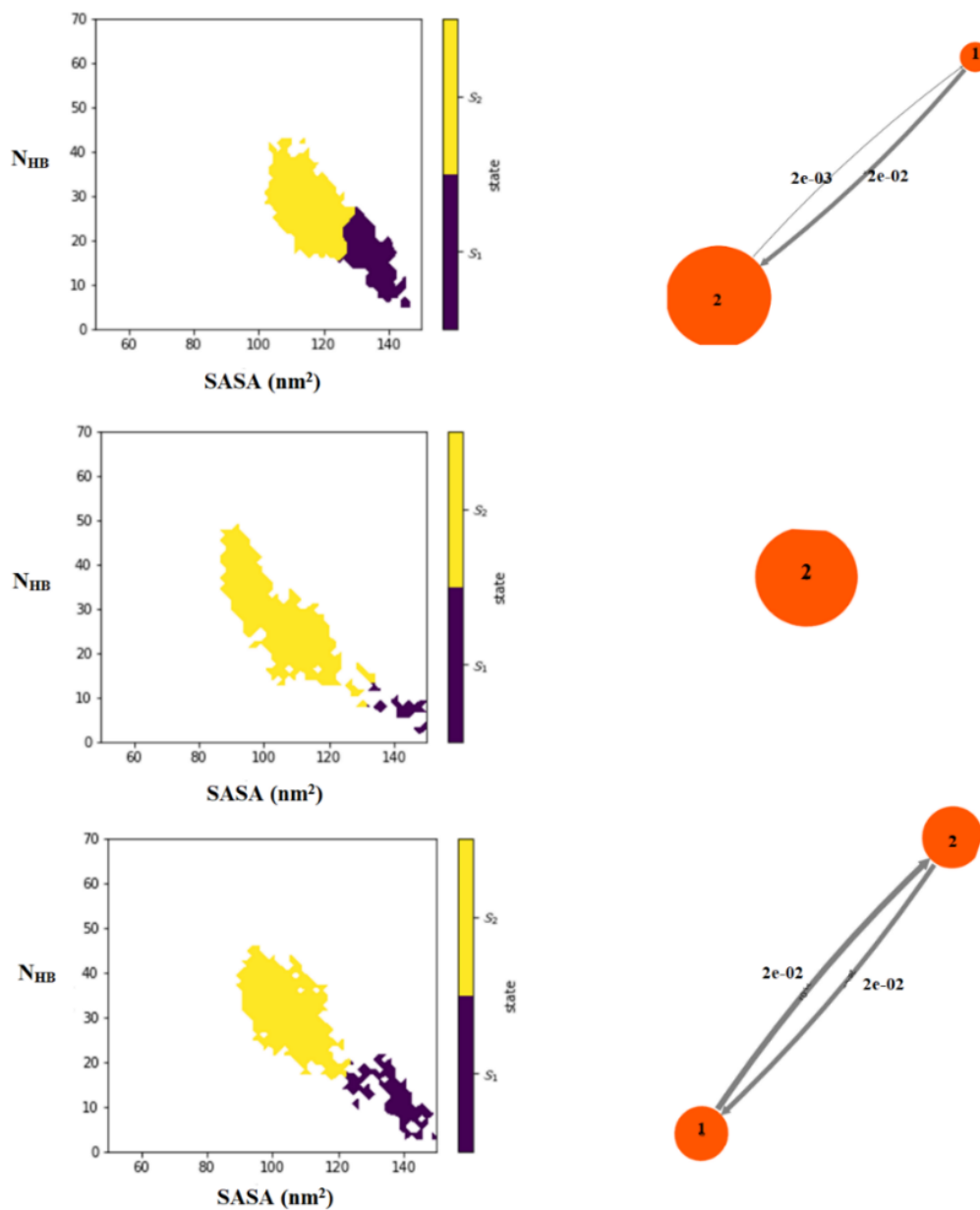


Figure 6.10. The decomposition of microstates regarding prion peptide aggregation in presence of NQDA and corresponding transition network obtained from Markov state modelling in case of (A) GROMOS-54a7, (B) AMBER99SB-ILDN and (C) CHARMM-36m force fields at 300K. The values written on the edges depict the transition rate between the microstates.

6.3.6 Mechanism of the peptide self-assembly inhibition by NQDA:

In this section, the underline mechanism of NQDA towards the inhibition of prion peptide self-assembly have been explored. The first step towards assessing the effect of NQDA in the peptide aggregation process is calculating the peptide-peptide radial distribution function and corresponding peptide-peptide, peptide-water coordination number in presence and absence of NQDA. It is evident from **Appendix XXXXXXXXVIII** that in the presence of NQDA, the position of the first and second peak of peptide-peptide RDF has not changed much, but there is a significant reduction in the peak height in all the force-fields. Moreover, there is a marked decrease in peptide-peptide coordination number in presence of NQDA (**Appendix XXXXXXXXIX**) shows it's inhibitory effect on prion peptide self-assembly. Further, to assesses the effect of NQDA on the energetics of peptide self-assembly, we calculate the short-range electrostatic and van der Waals interaction energy and provided in **Appendix XXXXXXXXX**. It is evident from **Appendix XXXXXXXXX** that coulomb or electrostatic interaction between peptide and water influences remarkably in controlling the aggregation and dissociation of peptides during the pre-fibrillar self-assembly. In previous sections, it has been noted that the GROMOS 54a7 force field shows a higher aggregation and β -sheet propensity followed by AMBER 99SB ILDN and CHARMM 36m forcefield. This is due to lesser peptide-water electrostatic interaction energy in GROMOS 54a7 compared to the other two force fields. The marked decrease in peptide-peptide coulomb interaction energy and increase in peptide-water interaction energy in the presence of NQDA potentiates the inhibition of prion-peptide pre-fibrillar assembly at its early stage. This trend is found to be similar in all the three force fields which supports the inhibitory role of NQDA. Further, we calculated the residue wise interaction energy in the aggregation process. The decomposition of interaction energy with respect to the peptide residues is depicted in **Figure 6.11**. It is evident from **Figure 6.11-A** and **Figure 6.11-C** that Tyr, Met, and Leu residues (Index 2, 3 &4) have maximum contribution towards the aggregation process in case of GROMOS and CHARMM force-field. The higher energy contribution of the residues located at the middle portion of the peptides validates the higher percentage of β -sheet in above two force-fields. In case of AMBER force-field the energy contribution of terminal residue is more compared to the middle residue. The higher interaction energy contribution of

terminal residues can be the region behind the lower β -sheet content in AMBER force-field. The presence of NQDA found to reduce these contributions in all the force-fields. Especially, the interactions of Tyr and Leu residues are highly affected by the inhibitor. In case of CHARMM force-field the reduction of residue wise energy contribution is comparatively lesser than other two force-fields considered in this study. This can be the reason behind the higher β -sheet content in presence of NQDA for CHARMM force-field.

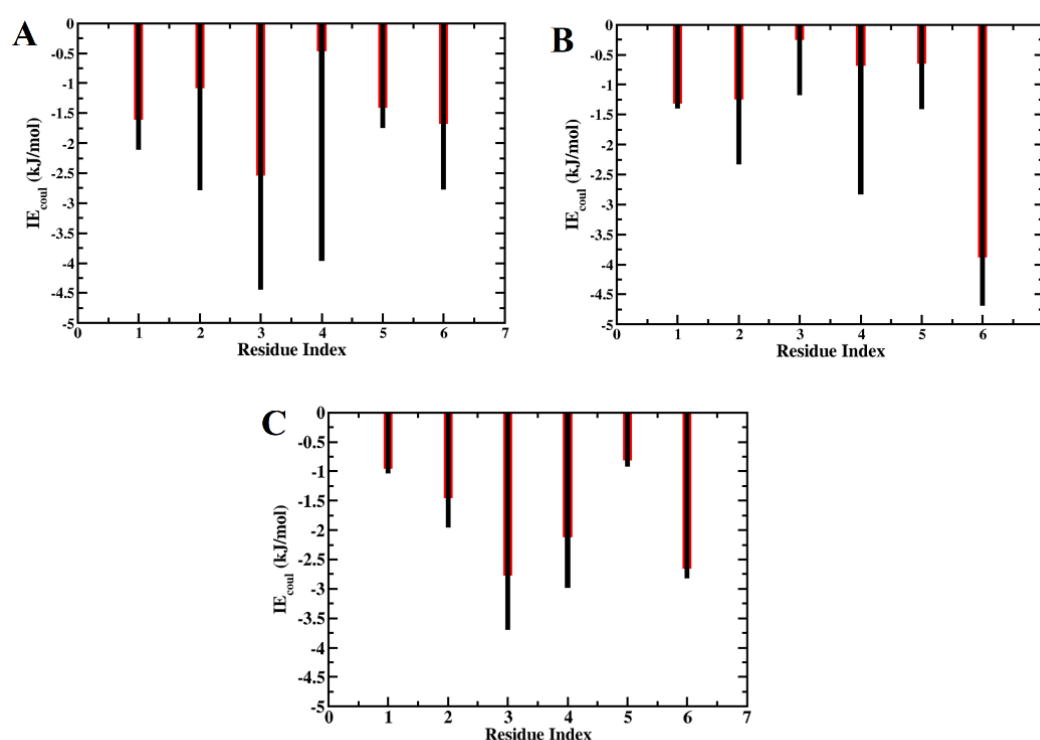


Figure 6.11. The residue-wise interaction energy at (A) GROMOS-54a7 [20 mM], (B) AMBER-99SB-ILDN [30 mM], and (C) CHARMM-36m [50 mM] in the absence (black) or presence (red) of NQDA at 300K.

Further, to assess the direct interaction of NQDA, the peptide-peptide hydrogen bonding, and π - π stacking interactions have been calculated that are crucial for peptide self-assembly. Considering the number of inter-peptide hydrogen bonds, a notable decrease in the hydrogen bond number between the peptides at different force-fields as found in the presence of NQDA (**Figure 6.12 A-C**). In case of CHARMM force-field the change in intra-peptide hydrogen bond is comparatively lesser in presence of

NQDA. This is due to the lesser number of peptide-NQDA hydrogen bond compared to other force-fields (**Figure 6.12-D**).

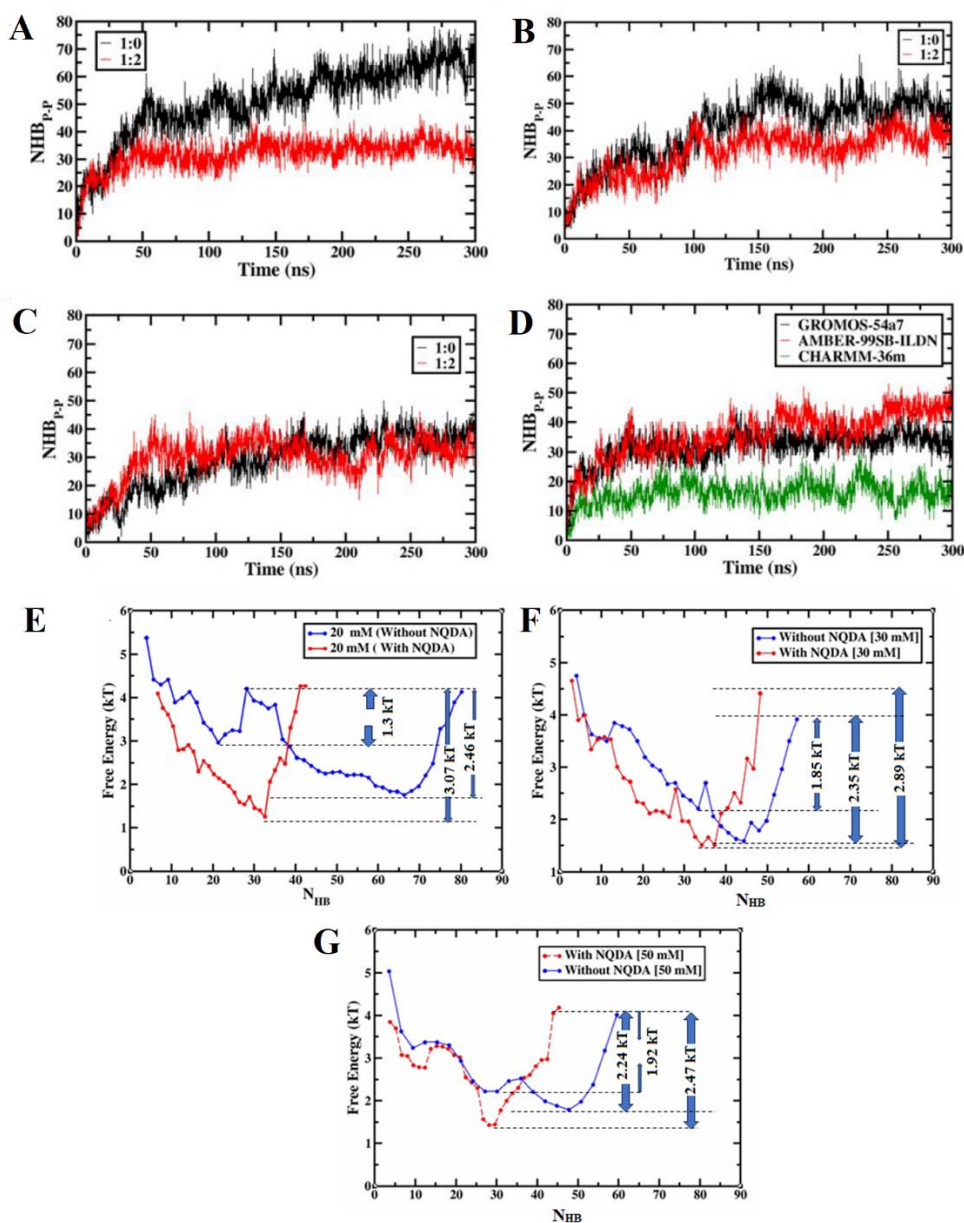


Figure 6.12. The time evolution of peptide-peptide hydrogen bond number in the presence or absence of NQDA (A) GROMOMS-54a7 [20 mM], (B) AMBER-99SB-ILDN [30 mM], (C) CHARMM-36m [50 mM] at 300K. The black curve is showing the number of intra-peptide hydrogen bonds in pure water, and the red curve is showing the number of intrapeptide hydrogen bonds in presence of NQDA. (D) The time evolution of peptide-NQDA hydrogen bonds at different force-field at 300K. The 1-

dimensional free energy landscape along the intrapeptide hydrogen bond as reaction coordinate for (E) GROMOS-54a7, (F) AMBER 99SB ILDN and (G) CHARMM 36m force-field.

Further we calculated the free energy profile along the inter peptide hydrogen bond as reaction coordinate using the normalized probability distribution of number of inter peptide hydrogen bond, $G(HB) = -kT \ln(p(HB))$. It can be noted from **Figure 6.12 (E-G)** that the peptide cluster related to Lag 1 can easily shift to peptide cluster correspond to stable lag phase 2 with a small energy barrier of 0.32 kT-0.50 kT as seen in the case of AMBER and CHARMM. This shows the transient nature of the peptide cluster related to Lag1. It is evident from **Figure 6.12 E-G** that NQDA increases the activation barrier towards the formation of higher number of intra-peptide hydrogen bond required for fibril nucleation. The barrier height is found to be increased to 3.07kT, 2.89kT and 2.47kT for the three force-fields respectively. Additionally, the π - π stacking interaction between the tyrosine residues that further stabilizes the prion-peptide oligomers (**Figure 6.13-A**) is found to be replaced by the π -stacking interaction between naphthalene-1,4-dione or the pyrocatechol rings of NQDA compounds. In the present scenario, the face-to-face and edge-to-face π -stacking interaction (**Figure 6.13-B**) is found. We calculated the angle between the phenol rings of the tyrosine residues to validate this interaction. The angle between the rings is highly populated at 95°, indicating majority of edge-to-face π -stacking conformation (**Appendix XXXXXXXXI**). Further, we calculated the pair correlation function between Tyr residues in the presence or absence of NQDA to assess the inhibition effect. The peak height in the RDF is found to be decreased in the presence of the inhibitor (**Appendix XXXXXXXXII**).

Furthermore, the life-time of those hydrogen bonds was evaluated and depicted in **Table 6.2** for more insights into the peptide aggregation process. The hydrogen bond can be defined as(Chandra 2000)

$$S_{HB}(t) = \frac{\langle h(0).H(t) \rangle}{\langle h(0)^2 \rangle} \dots\dots\dots (6.3)$$

The hydrogen bond between donor and acceptor is said to be formed if their interatomic distance is < 0.25 nm. Here, <...> indicates the average over all the pairs of given type.

$H(t)$ is indicating the population parameter which is 1 if a specific hydrogen bond exists from time $t = 0 \rightarrow 1$ or zero otherwise.

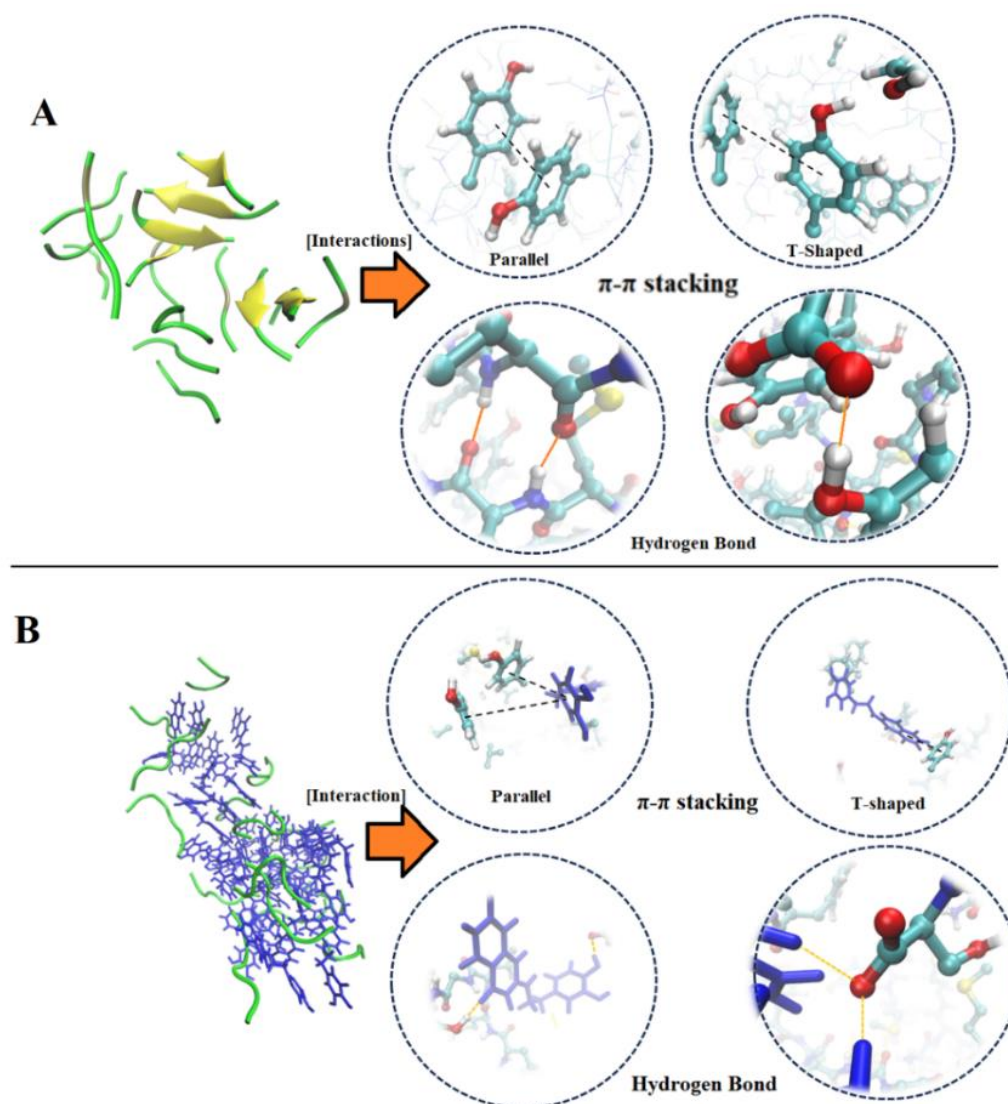


Figure 6.13. (A) The intrapeptide Hydrogen bond and π - π stacking interaction between Tyrosine residues within the final oligomeric stage of prion peptides in pure aqueous solution (20 mM, 300K). (B) Replacement of hydrogen bond and Tyrosine-Tyrosine π -stacking interaction by NQDA-molecules. The peptides are illustrated in cartoon format. The β -sheet, β -bridge, coil, and turns are represented in yellow, olive green, and green color respectively. The NQDA molecule is shown in green color.

We calculated the hydrogen bond between protein polar hydrogen (attached to nitrogen and oxygen) and oxygen of water, protein oxygen, and hydrogen of water. We have

considered interfacial water molecules up to a cut off value of 0.5 nm for the calculation of protein-water hydrogen bond lifetime. For the lifetime of the intra-peptide hydrogen bond the polar hydrogen and the oxygen and nitrogen atoms of the peptides are considered. In the case of peptide-NQDA hydrogen bond interaction, the polar hydrogen of NQDA and oxygen, and nitrogen of peptides are chosen. It can see from **Table 6.2** that peptide-peptide hydrogen bond life time decreases at higher temperature and increases in higher peptide concentration. In case of GROMOS-54a7 force-field there is a remarkable decrease in the intra-peptide hydrogen bond life time in presence of NQDA. Similar phenomena have been observed in case of AMBER-99SB-ILDN and CHARMM-36m force-field. The higher life-time of NQDA-peptide hydrogen bond compare to intra peptide hydrogen bond in all the force-field. (**Table 6.2**) suggest that NQDA has more propensity to make a hydrogen bond with peptides. This observation depicts that, in the presence of NQDA, peptides can only be able to self-aggregate into smaller aggregates (dimer or trimers) and fail to fuse into a critical nucleus size or pre-fibrillar aggregates, thereby promoting a significant reduction in toxicity.

Table 6.2. Average relaxation time or life time of continuous hydrogen bonds between peptide-peptide ($\langle \tau_c^{PP} \rangle$), peptide-water ($\langle \tau_c^{PW} \rangle$) and peptide-NQDA molecules. $\langle \dots \rangle$ indicates time averaging

System	$\langle \tau_c^{PW} \rangle$ (ps)	$\langle \tau_c^{PP} \rangle$ (ps)	$\langle \tau_c^{P-NQDA} \rangle$ (ps)
300K, 10 mM [G1]	4.04	17.62	-
350K, 10 mM [G2]	1.52	9.50	-
375K, 10 mM [G3]	1.18	6.30	-
300K, 20 mM [G4]	4.15	29.24	-
300K, 50 mM [G5]	4.72	30.12	-
300K, 30 mM [A1]	11.80	26.39	-
300K, 50mM [C1]	4.28	37.33	-
300K, 10 mM [G6], (1:2)	3.39	7.60	15.96
350K, 10 mM [G7], (1:2)	1.73	6.38	7.15
300K, 20 mM [G8], (1:2)	3.11	11.78	19.61
375K, 10 mM [G9], (1:2)	1.65	5.94	7.87
300K, 30 mM [A8], (1:2)	15.46	10.67	14.92
300K, 50 mM [C8], (1:2)	3.76	12.40	17.18

6.3.7 Role of NQDA on the solvation of peptides

Furthermore, we calculated the RDF of water molecules around the peptides during the aggregation process for GROMOS, CHARMM, AMBER force-fields and depicted in

Figure 6.14. Two well defined solvation shell is found around the peptide aggregates in the absence of NQDA.

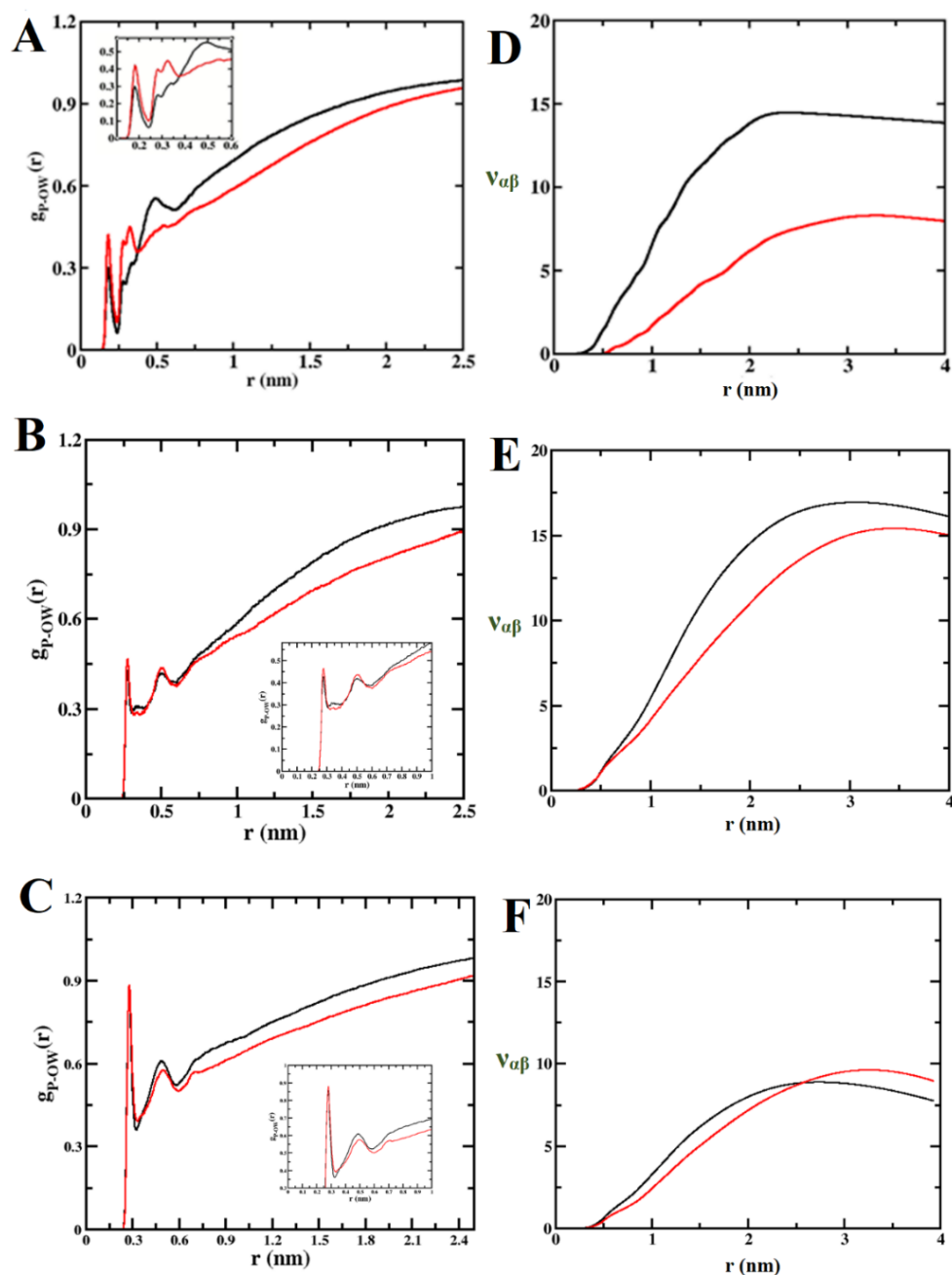


Figure 6.14. The pair correlation function of water molecules from the surface of the largest populated peptide cluster at (A) GROMOS-54a7 [20 mM], (B) AMBER99SB-ILDN [30 mM], (C) CHARMM-36m [50 mM] force-fields at 300K. (D-F) The preferential interaction parameter of peptides in the presence or absence of NQDA for different (D) GROMOS, (E) AMBER and (F) CHARMM force-field. The black curve

is showing for pure aqueous solution and the red curve corresponds to the presence of NQDA.

At 300K, the first peak of the peptide RDF is observed around ~0.15 nm whereas the second solvation shell located at ~0.5 nm in case of GROMOS force-field. Presence of NQDA shows significant perturbation in both the solvation shells of hydration water in GROMOS and AMBER force-fields (**Figure 6.14-A, B**). At 300K, in presence of NQDA, there is increment in peak height for both the solvation shells indicates the increase in water density in the hydration environment. In case of CHARMM force-field the reduction in the second solvation shell is due to lesser number of water molecules present at the higher concentration of the peptide (50 mM) (**Figure 6.14-C**). Additionally, the lifetime of water-peptide hydrogen bond is increased in presence of inhibitor for GROMOS and AMBER force-field (**Table 6.14**). The peptide-water hydrogen bond lifetime is decreased less in case of CHARMM force-field due to insufficient availability of the water molecules. The concentration used for CHARMM is 50 mM which contains only 15977 water molecules compared to 28447 and 42186 water molecules for AMBER and GROMOS respectively. Thus, it can be inferred that in presence of NQDA peptides are more hydrated which essentially destabilize the prion peptide aggregates by weakening the peptide-peptide non-bonded interactions.

Since the amphiphilic prion-peptides show a high aggregation tendency in the aqueous medium, the peptide fragments prefer to interact with other copies of peptides rather than water. Hence, we calculated the preferential binding coefficient (PBC) between the peptides to further assess the extent of aggregation. The PBC can be defined as (Ganguly et al. 2018; Singh and Chakraborty 2021),

$$v_{PP} = \rho_P (G_{PP} - G_{PW}) \quad (6.4)$$

$$G_{PP} = 4\pi \int_0^\infty [g_{PP}(r) - 1] r^2 dr \quad (6.5)$$

$$G_{PW} = 4\pi \int_0^\infty [g_{PW}(r) - 1] r^2 dr \quad (6.6)$$

Where, P, W, and ρ_P represent the peptide, water molecules, and number density of the peptide, respectively. Further, G_{PP} and G_{PW} are the KB (Kirkwood-Buff) integrals computed from the Peptide-peptide and peptide-water RDF. A positive value of v for peptides indicates the favourable inter-peptide self-assembly. In **Figure 6.14 D-F** we

have plotted preferential binding coefficient between the peptides at pure aqueous solution and in the presence of NQDA.

It is evident from Figure **6.14-D-F** that v_{PP} has positive value for the pure aqueous solution of prion peptide in all the force-fields. However, incorporating NQDA leads to a significant reduction of v_{PP} value even in the increment in temperature and peptide concentration in all the three force fields. This effect is prominent in GROMOS and AMBER force-fields. The lesser positive value of v_{PP} in the presence of NQDA indicates the peptide association as lower-order oligomers. Therefore, it can be summarized that NQDA not only inhibits peptide aggregation by direct interaction with the peptide, but it also modulates the hydration structure by increasing peptide-water interaction, which causes a higher population of peptides with random coil conformation.

6.3.8 Assessing dimer dissociation energy by umbrella sampling simulation

It is evident from the cluster size evolution of prion-peptide aggregation that NQDA has a visible effect on the peptide oligomer size equal to or higher than the critical nuclear size. However, the impact of NQDA is unclear for lower-order aggregates such as dimers. To see the effect of NQDA on dimer formation, we calculated the potential of mean force (PMFs) of the association between two prion peptide fragments by umbrella sampling simulations (Das and Chakraborty 2021; Roy and Paul 2021) in GROMOS 54a7 force field (G14, G15 in Table 2). Notably, the propensity of β -sheet formation is highest in GROMOS 54a7 force-field (**Figure 6.7-A**) compared to other two force-fields and therefore we selected this force-field to assess the effect of NQDA on dimerization and subsequent β -sheet formation. Since the calculation of PMF by Umbrella sampling method is computationally demanding process, therefore, we chose to perform Markov state model (MSM) (**Appendix XXXXXXXXXIII**) to find out the most populated conformations in presence or absence of NQDA and computed the PMF for the corresponding conformation. In absence of NQDA, we found one predominate conformation while in presence of NQDA, we found three major populations. The PMF is computed as the function of the distance between the COM of the peptide backbones and is presented in **Figure 6.15**. We have also plotted the force profile as a function of simulation time for the dissociation of monomer in the presence or absence of NQDA

and depicted in **Appendix XXXXXXXXXIV**. The inhibitory effect of NQDA on the primary aggregation state, i.e., dimerization is visible from the figures given below. It can be seen from **Appendix XXXXXXXXXIV** that the pull force for properly formed dimer increases up to 100 ps which indicates the breaking point for the crucial β -sheet hydrogen bond that stabilizes the dimer in pure aqueous solution. In contrast, the pull force is significantly less for monomer dissociation in the presence of NQDA. In the absence of NQDA, the energy minima for dimer formation are located at 0.29 nm, indicating their most stable interaction at this location. Notably, the location of the minimum increases to 0.40 nm in the presence of the NQDA compound. This suggests that the dimerization is fragile due to the interaction of NQDA. Further, we calculated the free energy of dissociation from the difference between the highest and lowest part of the PMF curve.

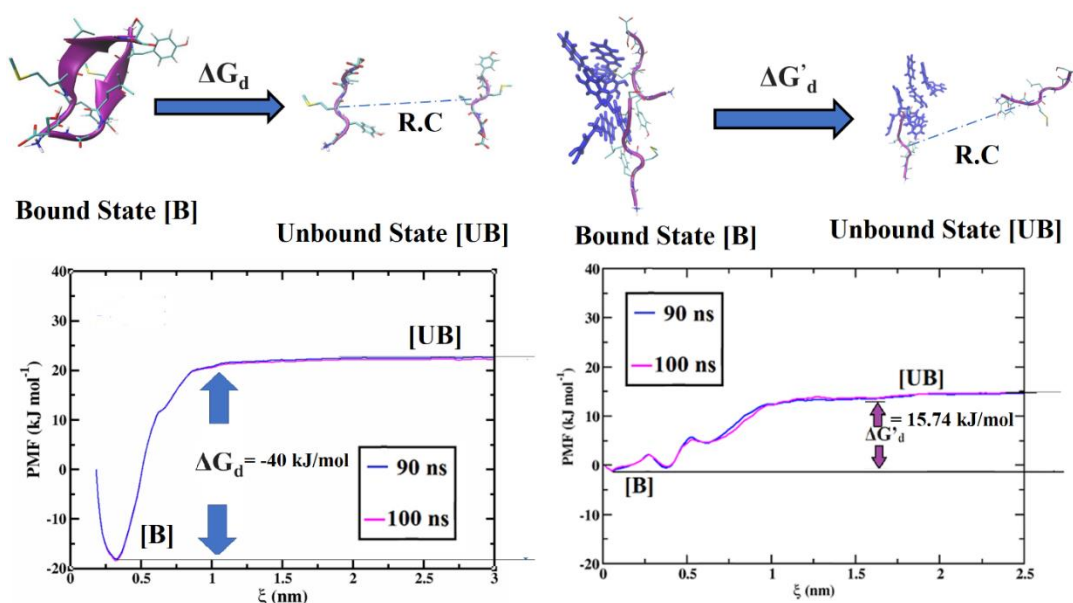


Figure 6.15. The representative potential of mean force (PMF) graph for dimer dissociation in the (A) absence or (B) presence of NQDA. The dissociation free energy corresponds to bound, and unbound states are denoted as ΔG_d and $\Delta G'_d$ respectively. The peptide is being pulled along the direction of the inter-peptide (dimer) hydrogen bond.

We found the PMF data is converged within 100 ns simulation at each umbrella window. In pure water, the dimer dissociation free energy is computed to be -40.08 kJ/mol, while the dissociation free energy is reduced by -24.34 kJ/mol due to the addition of NQDA correspond to maximum populated conformation i.e Conformation 1. We have carried out umbrella sampling of three different dimer conformations due to structural heterogeneity in presence of NQDA (**Appendix XXXXXXXXXV**). The dissociation energy is found to be differ by ~2-3 kJ/mol which validates the dimer dissociation energy in presence of NQDA. These results can be attributed to the fact that NQDA not only impacts the higher-order oligomers, it also weakens the dimerization process.

6.4 CONCLUSIONS

Finally, to conclude, in this investigation the effect of temperature, peptide concentration, and the presence of inhibitor molecules on the oligomer nucleation and fibrillar association of hexadecameric shortest human prion peptide in three different force-fields have been summarised. Results obtained from simulations showed that the aggregation behavior is more favourable in case of GROMOS-54a7 force-field due to lesser diffusion of the monomers. The diffusion of monomers was found to be more in case of CHARMM 36m force-field which delays the aggregation process compared to other two force-fields and final aggregated state is found at 50 mM peptide concentration. The aggregation process comes across two activation barriers with a preferential population of clusters containing 6-9 and 12-13 copies of peptides. The energy barrier for the second lag phase is higher compared to the first lag phase which traps the aggregation process. In contrast, the first lag phase is transient in nature. The existence of these lag phases is validated by using Markov state model and they were found to present in all the three force fields. The kinetic barrier corresponding to *Lag1*, *Lag2* for HPP aggregation are found to be decreased at higher temperature and peptide concentration. At higher temperatures dissociation events are also noticed which is more prominent in the case of AMBER99SB-ILDN and CHARMM 36m force field. Further, the incorporation of inhibitors freezes the aggregation process efficiently between the two well-defined lags by increasing the sampling and stability of lower-order peptide aggregates. In presence of NQDA, a marked decrease in toxic beta-sheet

confirmation is visible in the entire oligomerization process. The inhibitor affects the hydration environment by increasing the solvation of peptide molecules. Lastly, umbrella sampling simulation reveals that the dissociation energy for dimerization process weakens by -19.97 kJ/mol which indicates the inhibitory effect of NQDA on dimer formation. This study, therefore, provides insightful information on the mechanism of prion peptide aggregation, the rate-determining step, and the effect of small molecule inhibitors in the self-assembly process which might be useful for the rational design of newer generation anti-prion agents.

CHAPTER 7

SUMMARY AND CONCLUSIONS

***Abstract:** This chapter furnishes the brief summary of the entire research work and portrays a comparison between the various computational techniques employed for inspecting the crucial interactions of therapeutically relevant protein-ligand and protein-protein associations. Lastly, it also includes the major conclusions drawn from the present research work.*

7.1 SUMMARY

- Pharmacophore and 3D-QSAR model have been adopted to identify the molecular counterpart essential for the competitive inhibition of the flexible loops of DHPS catalytic pocket.
- Molecular Docking calculations are performed to assess the binding mode of sulfone compounds at the binding pocket of DHPS.
- Density functional theory have been applied to visualize the location of HOMO-LUMO orbitals at the ligand surface as well as the binding pocket of DHPS.
- The relative binding free energy on functional group substitution in sulfone compounds have been calculated by free energy perturbation (FEP) approach.
- The impact of sulfone compounds on human body have been assessed by ADME/toxicity and pharmacokinetics analysis.
- Molecular dynamics simulation has been employed to confer the stability of sulfone compounds and identify the interacting amino acids that are crucial for sulfonamide resistance.
- The interaction of 8-mercaptopguanine (8MG) compounds at the bi-substrate catalytic pocket of DHPS have been studied to assess their ability to circumvent the sulfonamide resistance.
- Both ligands based (Pharmacophore and 3D-QSAR) and structure based (docking) approaches have been carried out to identify the amino acid interaction at DHPS catalytic pocket.

- Molecular dynamics simulation has been performed to assess the dynamic stability and calculate the occupancy of hydrogen bonds that formed during simulation.
- Principle component analysis (PCA) have been employed to observe the motion of flexible loops of DHPS ligand binding pocket in the presence of different 8-MG compounds.
- End-point free energy calculation (MM/PBSA) and umbrella sampling simulations have been carried out to compute the free energy of binding or unbinding of the 8-MG compounds.
- The immunogenic region (linear B-cell, conformational B-cell, T-cell epitopes) of SARS-CoV2 and corresponding the hotspot residues have been identified by immunoinformatic approaches. Further, multiple sequence analysis has been carried out to confirm the homology of the epitopes with other coronavirus species.
- Multiepitope vaccines have been constructed by joining the linear epitopes.
- The interaction of B-cell epitopes with human antibody and T-cell epitopes with corresponding MHC molecules have been studied by molecular docking technique.
- The dynamic stability of the epitope-receptor complex has been assessed by molecular dynamics simulation and endpoint free energy techniques.
- In-silico plasmid design have been carried out by back translating the protein sequence of the vaccine candidates.
- Molecular dynamics simulation has been employed to identify the major kinetic barriers of the pre-fibrillar aggregation of human prion peptide (127-GYMLGS-132) and their dependence on temperature, peptide concentration and force-field.
- Property based potential energy landscape have been constructed to quantitatively estimate the energy barriers during the aggregation process.
- Markov State Modelling have been used to assess the aggregation kinetics.
- NQDA compounds are added to target the aggregation lag phases and inhibit the formation of β -sheet.

- The life-time of interpeptide hydrogen bonds have been calculated in the presence or absence of NQDA to assess the inhibitory activity. Additionally, the structural change in hydration waters of the peptide aggregates have been investigated in the presence of NQDA.
- The dimer dissociation energy of the prion peptides in presence or absence of NQDA have been assessed by umbrella sampling simulation.

7.2 CONCLUSIONS

The significant conclusions of the current research work are listed below

- The ligand-based drug discovery techniques showed that the presence of hydrogen bond donor and electrostatic groups at the 2' and 4' position of R9 ring increase the potency of sulfone compounds for the competitive inhibition of DHPS.
- The docking studies showed that sulfone compounds are stabilized by Thr62, Arg63, Pro145, Phe190, Lys221, Ser222 and His257. The salt-bridge interaction by Arg63 and His257 is the main contributor to stabilize the sulfone compounds at the flexible loop of DHPS and the mutation in the nearby residues the main source of drug resistance mutation.
- The DFT studies showed that the presence of LUMO orbitals at interacting amino acid residues and the presence of HOMO orbitals in R8, R9 rings is crucial for substrate-based competition.
- The FEP studies showed that the replacement of hydrophobic groups with the electron withdrawing groups (hydroxyl and phenoxide) favours the decrease in relative binding affinity that stabilize the compounds at the catalytic pocket of DHPS.
- The lower RMSD value of sulfone compounds (0.15 nm) with respect to backbone RMSD (0.164 nm) of DHPS catalytic pocket and the lower degree of RMSF profile at interacting residues further indicates the dynamic stability of sulfone compounds at binding site.

- The pharmacophore analysis of 8-MG compounds showed that the change in pharmacophoric site affects the potency of the compounds for the competitive inhibition of DHPS.
- 3D-QSAR analysis indicates that presence of hydrogen bond donor group at H11 increases the activity of 8-MG compounds. Additionally, the presence of bulky hydrophobic groups at R12 site decreases the activity of 8MG based inhibitor.
- Docking studies reveal that hydrogen bonding, π - π stacking and π -cation interactions by Asp96, Asn115, Asp185, Phe188, Phe190, Lys221, Arg255 contribute towards the stability of 8MG compounds.
- Frontier molecular orbital studies showed that the overlap of HOMO, LUMO orbitals decreases the activity of 8-MG compounds.
- It is evident from MD simulation studies that highly active 8-MG compounds significantly reduces the fluctuation of flexible loops due to the higher hydrogen bond occupancy of between interacting hydrogen and inhibitor molecules.
- The PCA analysis indicates the closing motion of the DHPS catalytic pocket which essentially restrict the substrate binding and product release.
- End point free energy techniques also validates the role of electrostatic interaction in stabilizing the DHPS antagonists. Whereas the polar solvation terms disfavour the stability of receptor-ligand complexes. The free energy decomposition validates the favourable contribution of docking predicted residues.
- In-silico mutagenesis studies reveal that affinity of 8MG compounds is not affected by drug resistant mutations like Pro64Ser and Lys221Gln which indicates the resilience of such inhibitors against sulfa-resistant mutations.
- Umbrella sampling simulation indicates the hydrogen bonding interaction of Asp96, Asn115, Asp185 and water mediated hydrogen bonds with Cys137, Phe188, Gln226 serve as the breaking point during the unbinding process of 8MG compounds from the DHPS catalytic pocket.
- The solubility of sulfone compounds is lesser compare to 8 MG compounds due to higher number of polar atoms.

- In case of SARS-CoV2 spike protein the C-terminal region is highly conserved compared to the N-terminal receptor binding domain (RBD). This indicates that the perfusion region related to genome transfer is mostly common to all coronavirus species.
- Immunoinformatics study reveals that the linear B-cell and T-cell epitopes are scatter throughout the RBD and shaft region. However, the epitopes located at shaft region is mostly conserved and inclusion of such regions in multiepitope vaccines may induce immunity against broad spectrum coronavirus species.
- The hotspot residues are mainly polar in nature which contributes to favourable electrostatic interaction towards the corresponding immune receptors.
- The linear B-cell epitopes ⁸⁰⁸DPSKPSKRSF⁸¹⁷ and ⁶⁰⁰PGTNTSNQ⁶⁰⁷ are found to be common in all physiochemical parameters which indicates it's importance to elicit humoral immunity.
- The conformational B-cell epitopes are mainly located at the RBD and mostly have extended β -sheet or β -barrel conformations.
- Docking study reveals that multimeric vaccines made up of linear B-cell epitopes are mainly stabilized by hydrogen bonding interactions. Additionally, the conformational epitopes are stabilized by hydrogen bonding and π - π stacking interaction. In case of T-cell epitopes, the peptides are mainly stabilized by hydrogen bond interaction inside the peptide presenting domain of MHC molecules.
- MD simulation studies showed the dynamic stability of both the linear and conformational epitopes with immune receptors.
- Endpoint free energy (MM/PBSA) calculation showed that multiepitope B-cell vaccine has higher affinity compared to the conformation B-cell epitopes towards the human antibody.
- In case of T-cell epitope peptides, the peptides correspond to MHC-II showed higher affinity compared to the peptides related to MHC-I.
- The MD simulation of prion peptide at aqueous solution showed two major lag-phases or aggregation barriers with cluster size containing 6-9 and 12-14 peptides for all the force-fields. The first barrier corresponds to the critical nucleus formation and the second barrier correspond to fibril nucleation.

- Markov state modelling indicates that there are three main metastable states during the aggregation process.
- The incorporation of NQDA were found to arrest the aggregation process between two barriers at ambient temperature in all three force-fields as well as higher temperature and peptide concentration which indicates the efficacy of NQDA to prevent prion amyloidosis. Moreover, there is a marked decrease in β -sheet conformation in the presence of NQDA that indicates the decrease in cytotoxicity at pre-fibrillar stage.
- The inhibition process is mainly governed by π - π stacking interaction between Tyr and the aromatic rings of NQDA, peptide-inhibitor hydrogen bond and the higher solvation of peptide oligomers.
- Umbrella sampling simulation suggests that NQDA not only inhibit the higher order oligomer formation but also reduces the propensity to dimer formation. The incorporation of NQDA reduces the dimer dissociation energy by -19.97 kJ/mol.

The current research work illustrates how the in-silico techniques such as cheminformatics, bioinformatics, molecular docking, and multiscale molecular dynamics simulation techniques can be employed to study biomolecular recognition and associated rare events at temporal and special resolution. The ligand-based drug discovery techniques are found to be efficient in identifying crucial molecular fragments or functional group essential for targeted enzyme inhibition without the structural information of catalytic pocket. Moreover, the activity of the drug like small organic molecules can be predicted by the HOMO, LUMO energy data. To understand the stabilizing forces for ligand-receptor interaction, molecular docking is found to be computationally economical. Especially in vaccine design approaches molecular docking is the only way to screen potent peptide epitopes. However, to validate the docking predicted binding pose of biomolecular complexes, unbiased molecular dynamics simulation techniques and free energy techniques are essential.

APPENDIX

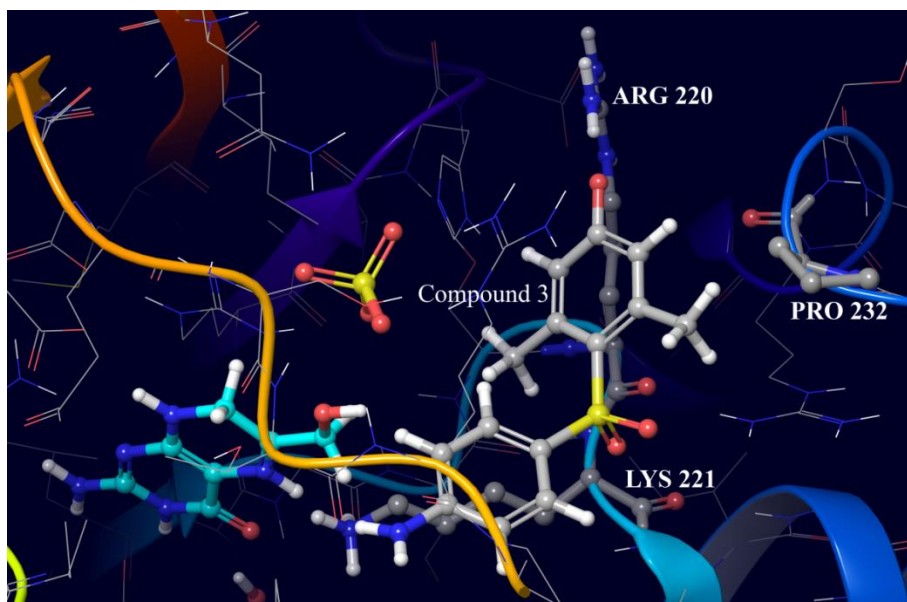
Appendix I. List of functional group present at 2', 4', 6' side of R9 ring. ✓ shows the presence of the group and × shows the absence of these groups

No	Anionic			Hydrogen Bond Donor Group			Hydrophobic Groups			Electron-Withdrawing Group		
	2'	4'	6'	2'	4'	6'	2'	4'	6'	2'	4'	6'
1	✓	✓	×	×	×	×	×	×	×	×	×	×
2	×	✓	×	×	×	×	✓	×	×	×	×	×
3	×	✓	×	×	×	×	✓	×	✓	×	×	×
4	×	×	×	×	✓	×	×	×	×	✓	×	×
5	×	✓	×	×	×	×	×	×	×	✓	×	×
6	×	×	×	×	✓	×	✓	×	×	×	×	×
7	×	×	✓	×	×	×	✓	✓	×	×	×	×
8	✓	×	×	×	✓	×	×	×	×	×	×	×
9	×	×	×	×	✓	×	✓	×	✓	×	×	×
10	×	×	×	✓	✓	×	×	×	×	×	×	×
11	✓	×	×	×	✓	✓	×	×	×	×	×	×
12	×	✓	×	✓	×	✓	×	×	×	×	×	×
13	×	✓	×	×	×	×	×	×	×	×	×	×
14	×	×	×	✓	✓	×	×	×	×	×	×	×
15	×	×	×	×	×	×	✓	✓	✓	×	×	×
16	×	×	×	×	×	×	✓	✓	×	×	×	×
17	×	×	×	×	×	✓	✓	✓	×	×	×	×
18	×	×	×	×	×	×	✓	✓	✓	×	×	×

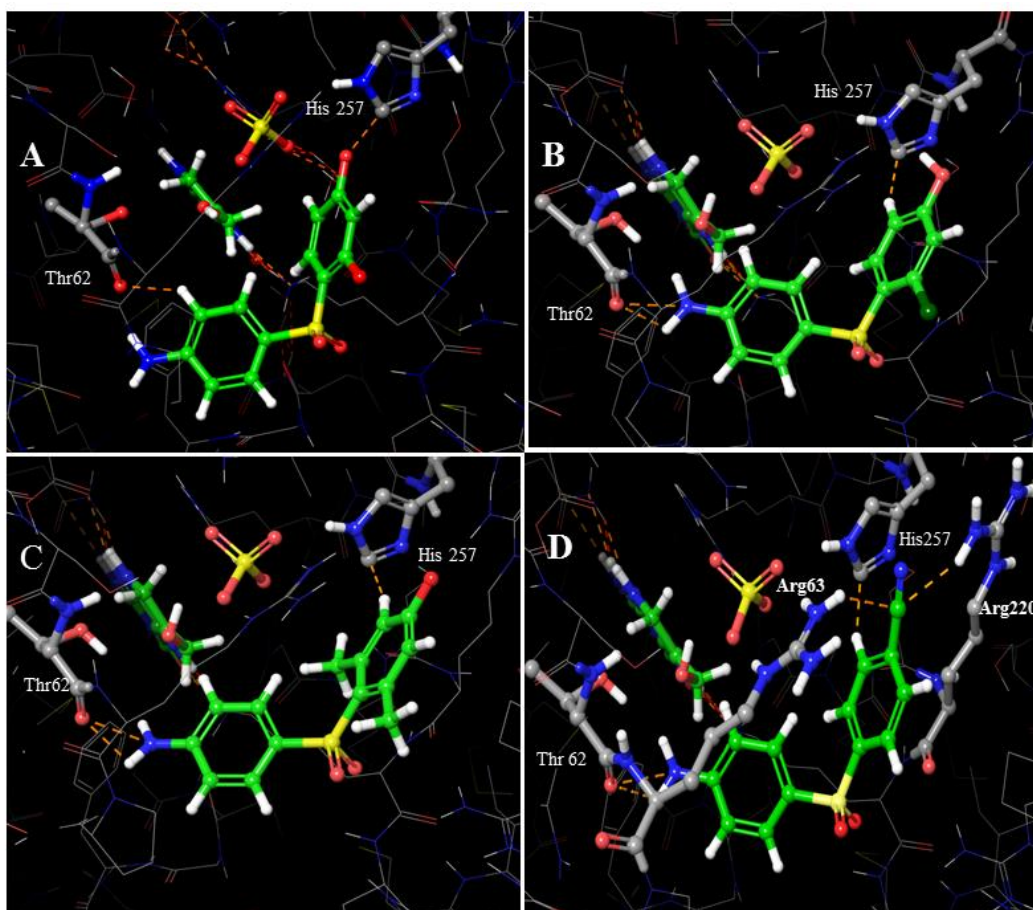
19	×	×	×	×	×	×	×	✓	✓	×	×	×	×
20	×	×	×	✓	✓	✓	×	×	×	×	×	×	×
21	×	×	×	×	×	×	×	×	×	×	✓	✓	×
22	×	×	×	✓	×	×	×	×	×	×	×	×	×
23	×	×	×	×	✓	×	×	×	×	×	×	×	×
24	×	×	×	×	✓	×	×	×	×	×	×	×	×
25	×	×	×	×	×	×	×	×	✓	×	×	×	×
26	×	×	×	×	×	×	×	✓	✓	✓	×	×	×
27	×	×	×	×	×	×	×	✓	✓	×	×	×	×
28	×	×	×	×	×	×	×	×	×	×	✓	✓	✓
29	×	×	×	×	×	×	×	×	✓	×	×	×	×
30	×	×	×	×	✓	×	×	×	×	×	×	×	×
31	×	×	×	×	×	×	×	×	✓	×	×	×	×
32	×	×	×	×	×	×	×	×	×	×	×	✓	×
33	×	✓	×	×	×	×	×	×	×	×	×	×	×
34	×	×	×	×	✓	×	×	×	×	×	×	×	×
35	×	×	×	×	✓	×	×	×	×	×	×	×	×
36	×	×	×	×	×	×	×	×	×	×	×	×	×
37	×	×	×	×	×	×	×	×	✓	×	×	×	×
38	×	×	×	×	×	×	×	×	✓	×	×	×	×
39	×	×	×	×	×	×	×	×	×	×	×	×	×
40	×	×	×	×	×	×	×	×	×	×	×	✓	×

41	×	×	×	×	×	×	×	✓	✓	✓	×	×	×
42	×	×	×	×	×	×	×	×	×	×	×	✓	×
43	×	×	×	×	×	×	×	×	×	×	×	✓	×
44	×	×	×	×	✓	×	×	×	×	×	×	×	×
45	×	×	×	×	×	×	×	×	×	×	×	✓	×
46	×	×	×	×	×	×	×	×	×	×	✓	✓	×
47	×	×	×	×	×	×	×	×	×	×	×	✓	×
48	×	×	×	×	×	×	×	×	×	×	×	✓	×
49	×	×	×	×	×	×	×	×	×	×	×	✓	×
50	×	×	×	×	×	×	×	×	×	×	×	✓	×

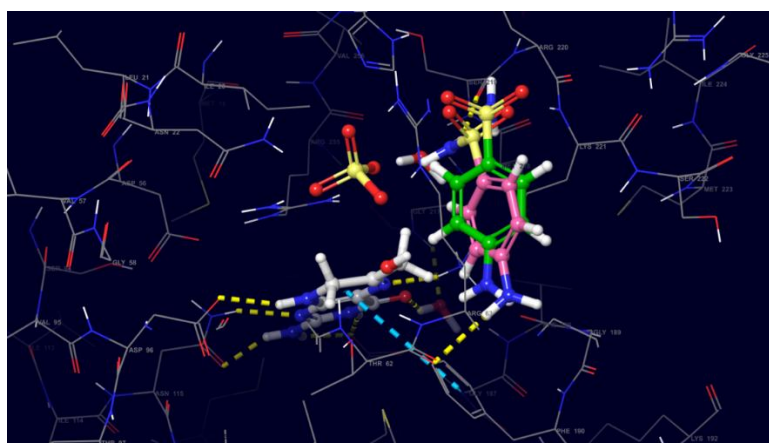
Appendix II. Hydrophobic interaction of compound 3 with Lys 221, Arg220 and Pro 232.



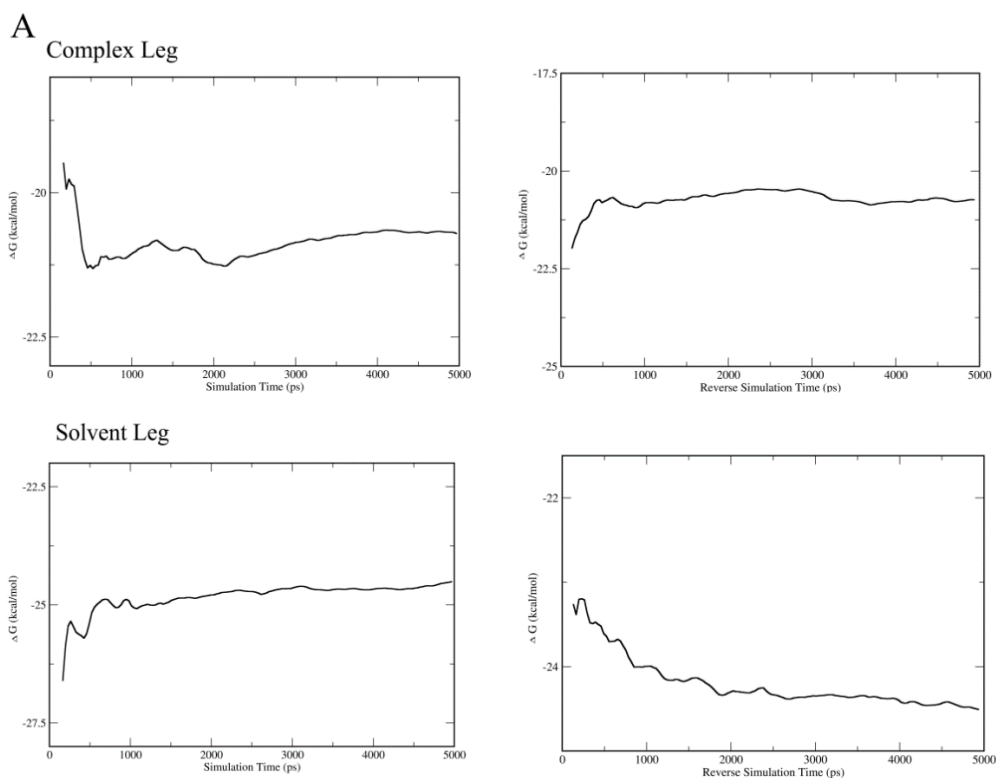
Appendix III. Binding pose and steric clashes between sulfone compounds and the neighboring amino acid residues in DHPS catalytic pocket (A)Compound 1, (B) Compound 3, (C) compound 4, (D) compound 50. The protonation of His257 presented in this picture is in Ne-H form.



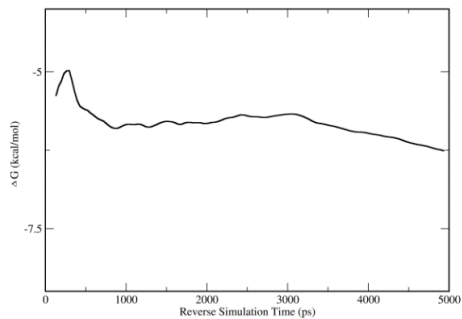
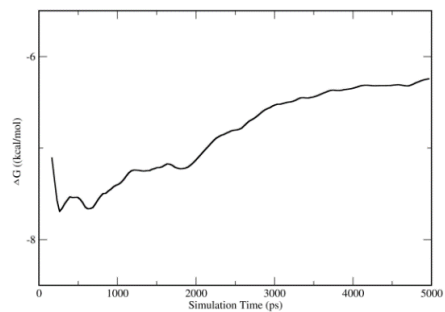
Appendix IV. Overlay of docking pose (pink) of sulphonamide with its crystal structure conformation (green) (RMSD= 1.21Å)



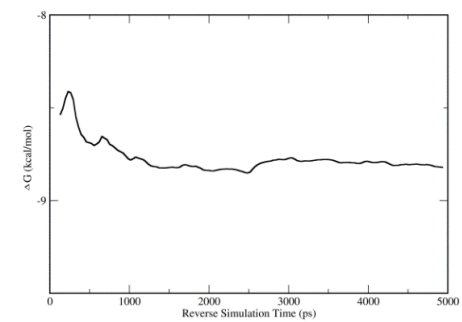
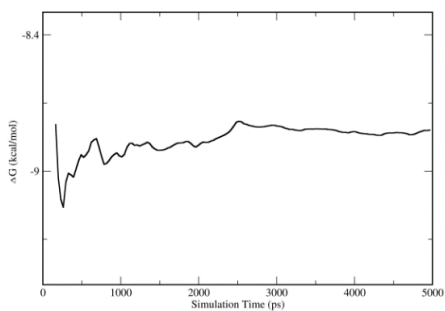
Appendix V. The relative free energy between two ligands in complex and solvent legs are plotted as a function of time. (A) Represents the convergence of calculated free energy difference when compound 50 is mutated to compound 1. (B) Represents when compound 50 mutated to compound 4 and (C) represents when compound 41 mutated to compound 1. The solvent leg and the complex leg describe the ligand atoms that are sampled more extensively in solvent (unbound) and complex (bound) respectively.



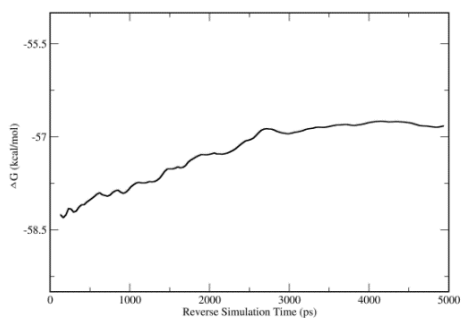
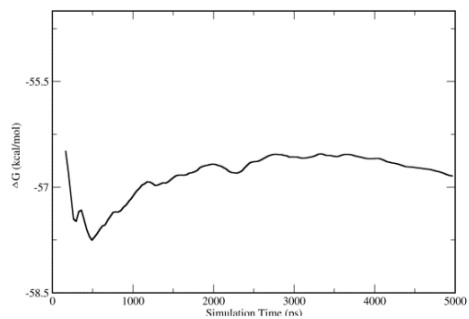
B Complex Leg



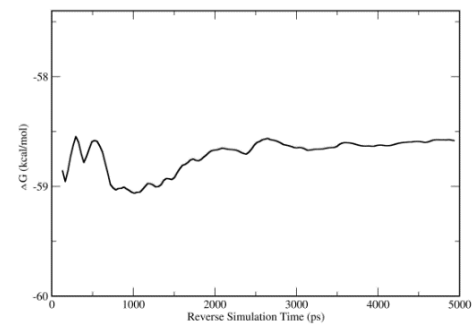
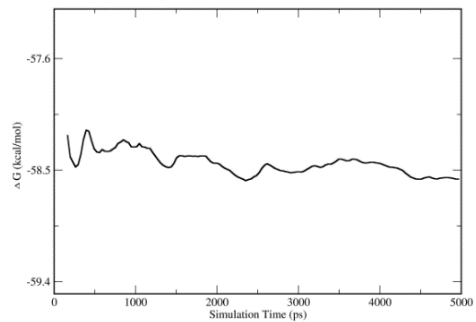
Solvent Leg



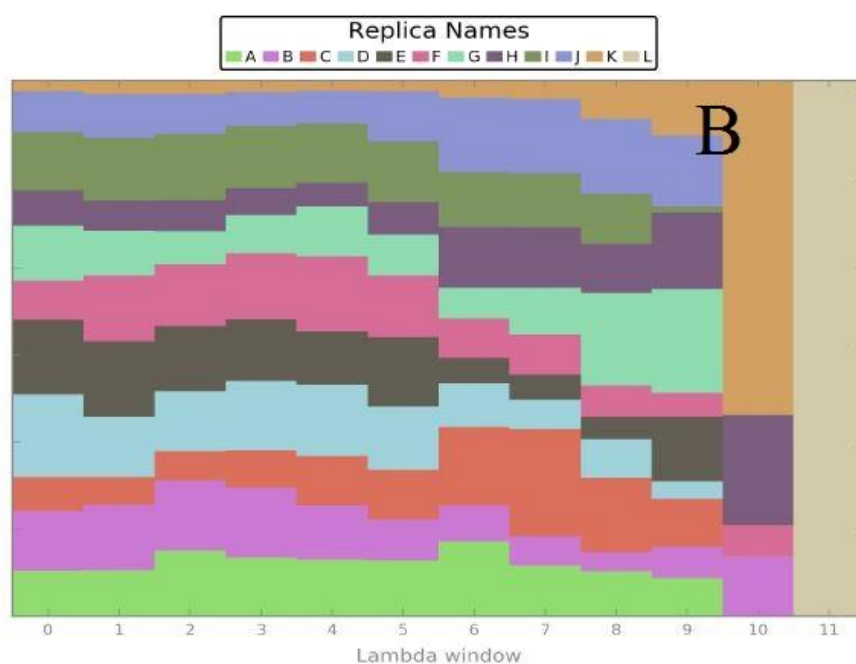
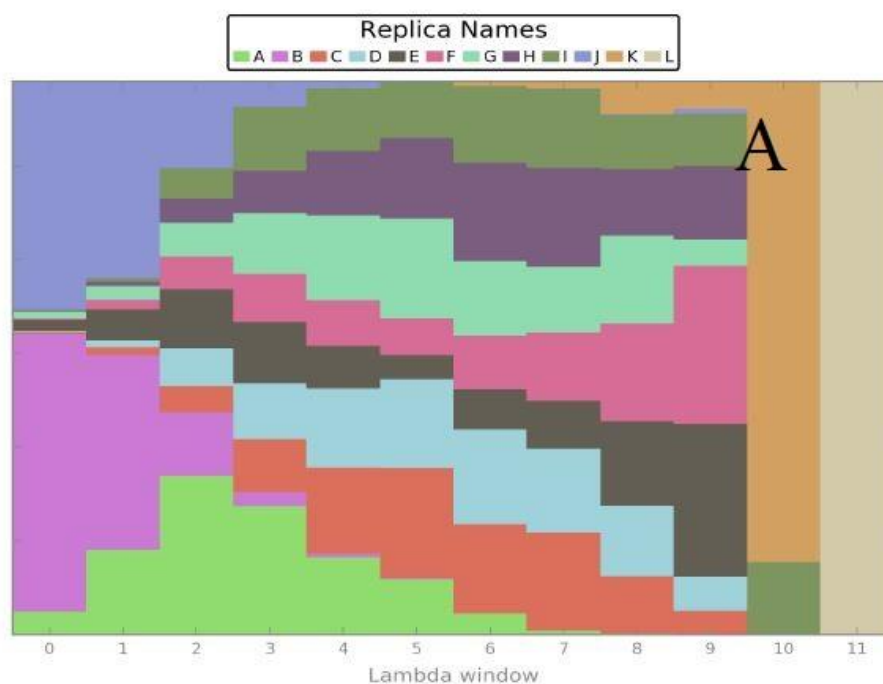
C Complex Leg

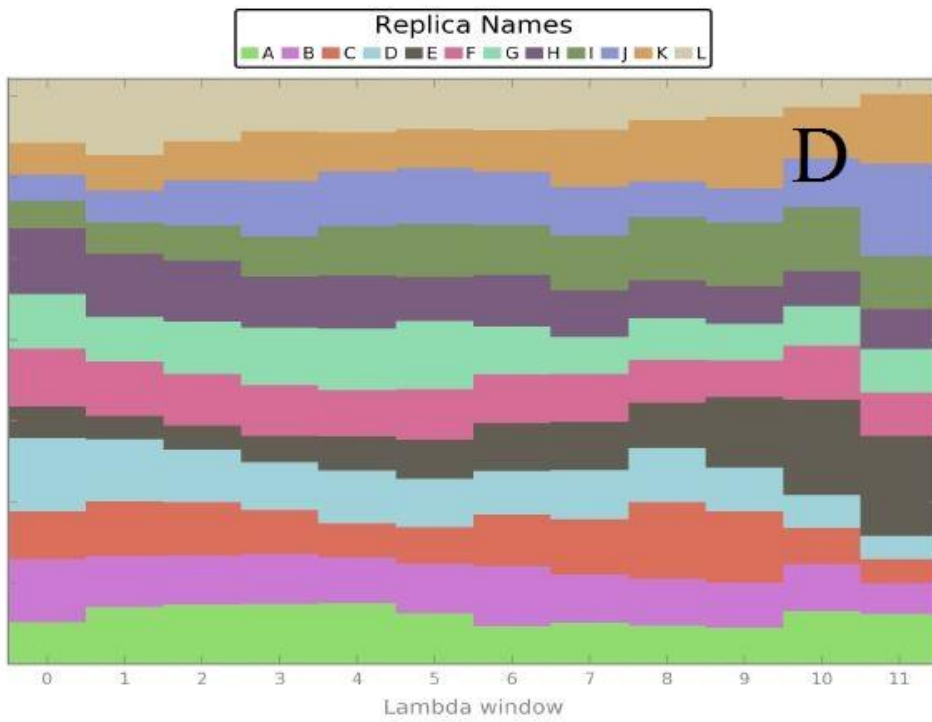
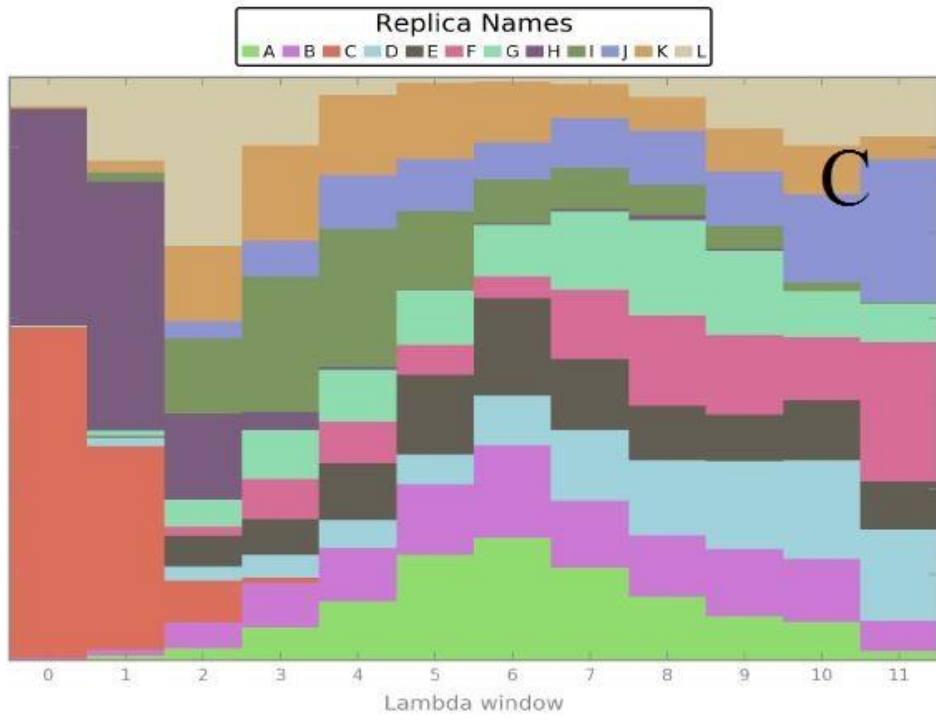


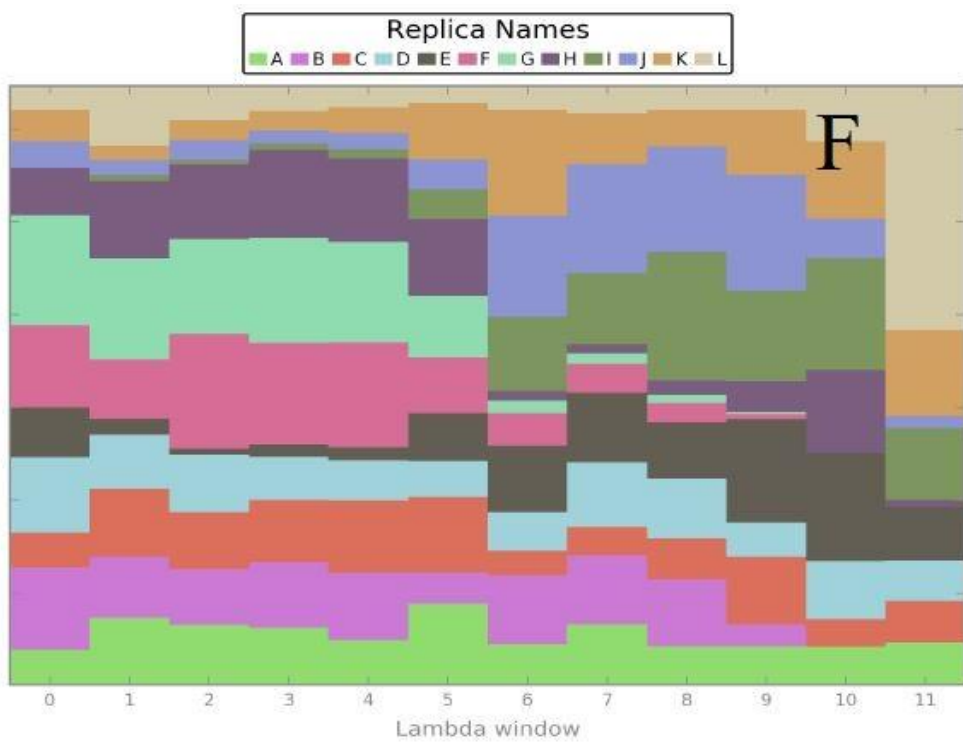
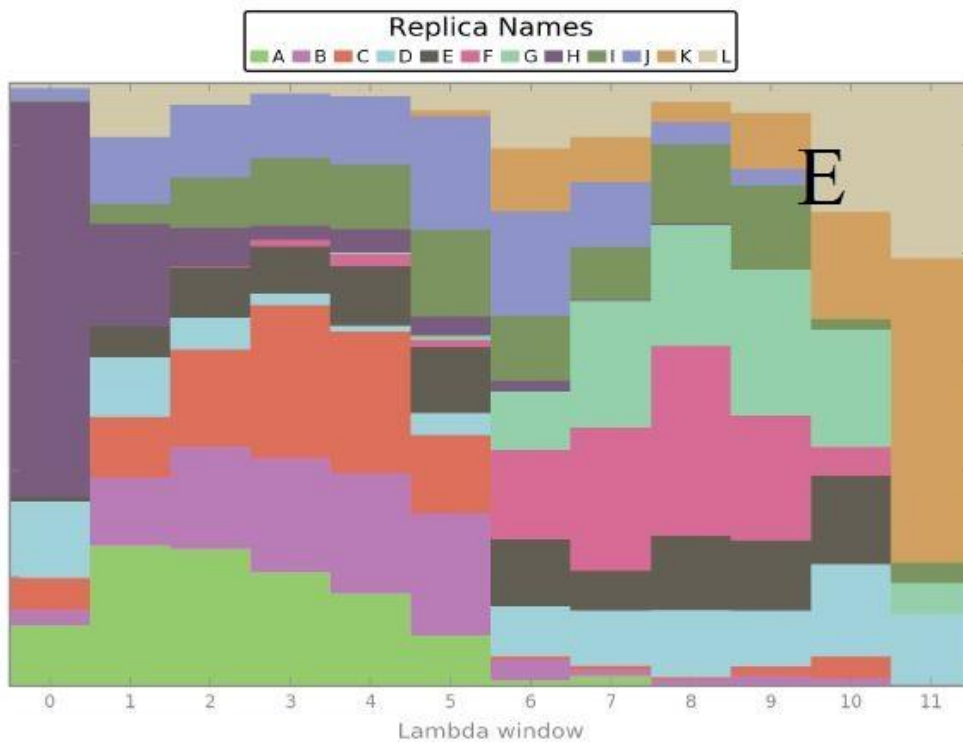
Solvent Leg



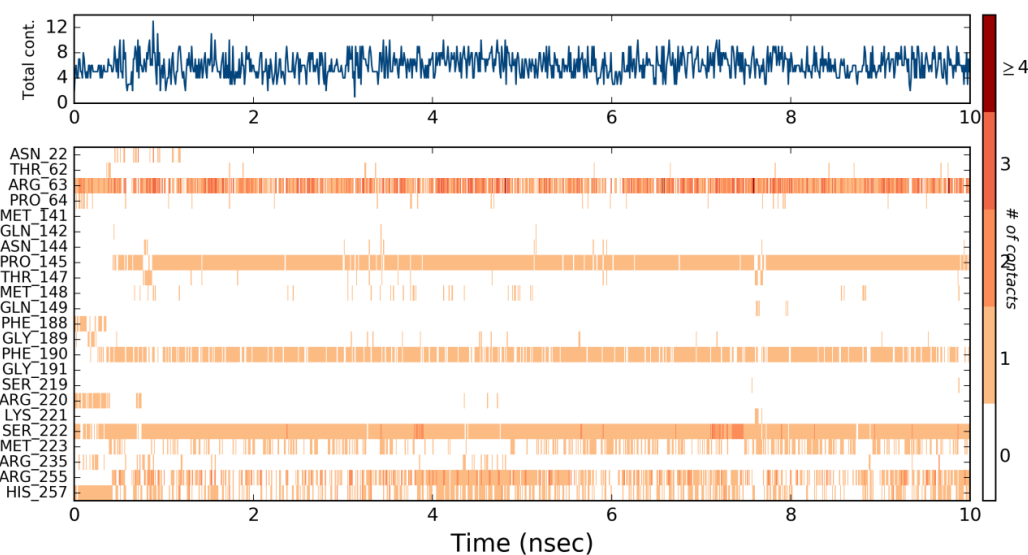
Appendix VI. The representation of the replica exchange windows of FEP simulations. Each replica is colour coded and plot shows how it occupy different λ window during the course of the simulation. (A) and (B) show the replica exchange windows of **compound 50** to **compound 1** mutation. (C) and (D) show the replica exchange windows of **compound 50** to **compound 4** mutation. (E) and (F) show the replica exchange windows of **compound 41** to **compound 4** mutation.



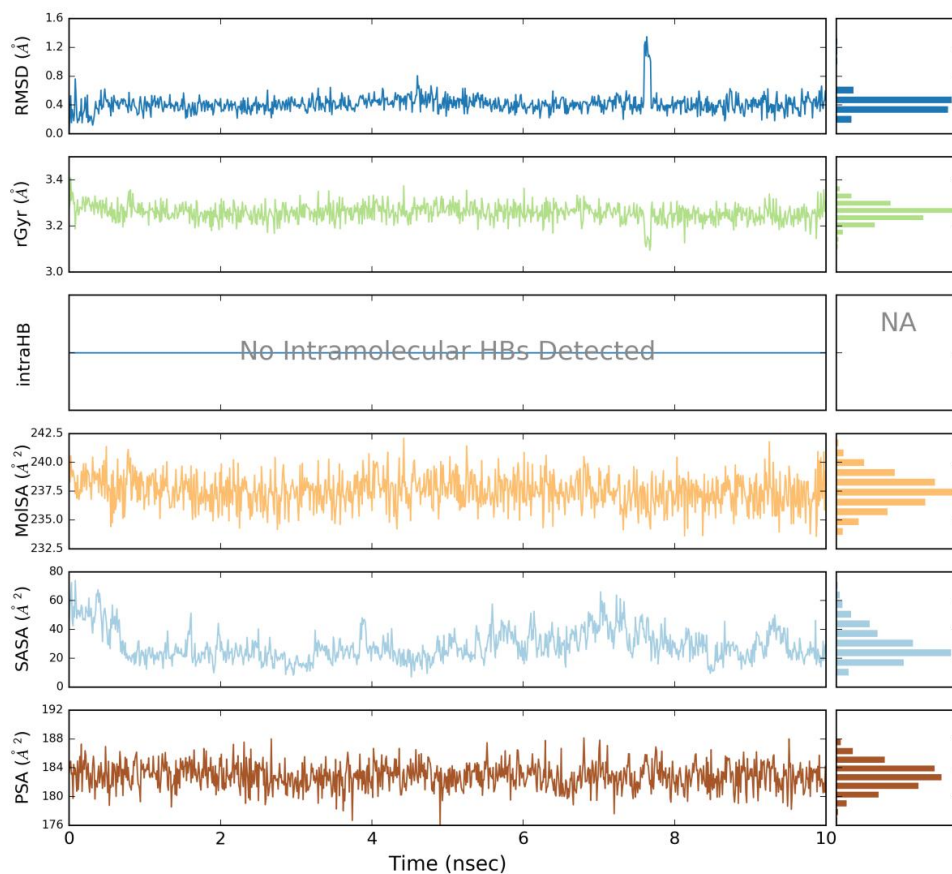




Appendix VII Timeline representation of protein ligand contact



Appendix VIII RMSD: Root mean square deviation of a ligand with respect to the reference conformation; rGyr: the extendedness of a ligand is measured through radius of gyration; intra HB: Intramolecular Hydrogen Bonds; MolSA: Molecular Surface Area; PSA: Polar Solvent area; SASA: Solvent accessible surface area.



Appendix IX. Grouping the 8-MG compounds based on factor scores of descriptors by k-means clustering

Cluster	Compounds	Number of compounds in cluster	Number of compounds in test set	Test Set Compounds
1	6, 13, 14, 16, 17, 19, 20, 22, 24, 26, 27, 29, 30, 31, 55, 56, 59	17	4	17, 26, 31, 56
2	1, 3, 37, 38, 43, 44, 53, 58, 61	9	2	37, 44, 53
3	15, 32, 33, 36, 39, 40, 51, 54, 57, 62	10	2	15, 54
4	2, 18, 21, 23, 25, 28, 34, 35, 41, 42, 45, 46, 47, 48, 49, 50, 52, 60	18	5	2, 18, 34, 48, 49
5	3, 4, 7, 8, 9, 10, 11, 12	8	2	4, 8

Appendix IX Statistical details of all pharmacophore hypothesis.

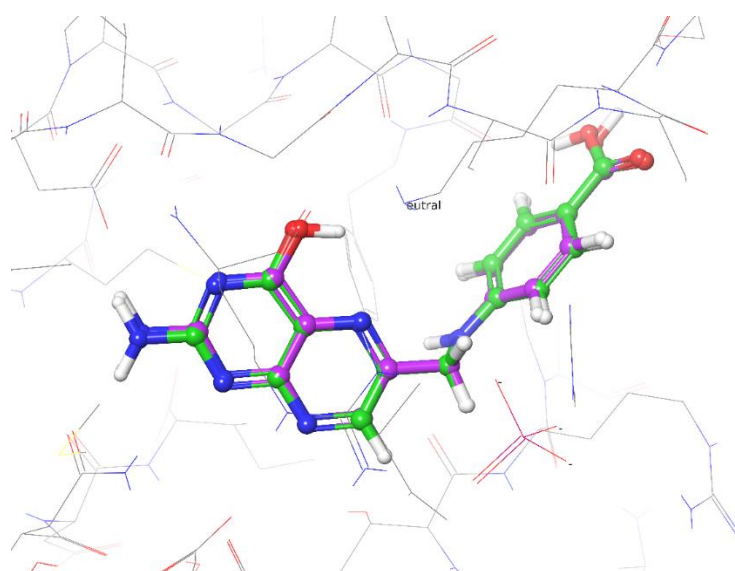
Hypothesis	Survival score	Site Score	Vector Score	Volume Score	BEDROC Score	phaseHypo Score
DDHRR.1	5.8125	0.999949	1.000	0.7988	1.000	1.3487
DDHRR.2	5.8062	0.999956	1.000	0.7988	1.000	1.3483
DDDHR.1	5.7900	0.999959	1.000	0.8052	1.000	1.3474
DDHRR.5	5.7878	0.999960	1.000	0.7736	1.000	1.3472
DDDHR.2	5.7866	0.999959	1.000	0.8052	1.000	1.3471
DDHRR.6	5.7814	0.999950	1.000	0.7961	1.000	1.3468
DDHRR.7	5.7809	0.999960	1.000	0.7736	1.000	1.3468
DDHRR.8	5.7806	0.999946	1.000	0.7961	1.000	1.3468
DDDHR.3	5.7789	0.999945	1.000	0.7961	1.000	1.3467
DDHRR.9	5.7768	0.999954	1.000	0.8052	1.000	1.3466
DDDHR.4	5.7750	0.999947	1.000	0.7961	1.000	1.3465
DDHRR.10	5.7668	0.999945	1.000	0.7961	1.000	1.3460
DDHRR.11	5.7590	0.999961	1.000	0.7736	1.000	1.3455
DDDHR.5	5.7582	0.999952	1.000	0.8052	1.000	1.3454
DDHRR.12	5.7581	0.999962	1.000	0.7736	1.000	1.3454
DDDHR.6	5.7565	0.999956	1.000	0.7736	1.000	1.3453
DDHRR.13	5.7555	0.999949	1.000	0.7849	1.000	1.3453
DDDHR.7	5.7533	0.999951	1.000	0.7988	1.000	1.3452
DDDHR.8	5.7525	0.999959	1.000	0.7736	1.000	1.3451
DDDHR.9	5.7467	0.999945	1.000	0.7961	1.000	1.3448
DDDHR.10	5.7355	0.999948	1.000	0.7849	1.000	1.3441
DDHRR.3	5.7929	0.999956	1.000	0.8052	1.000	1.3440
DDHRR.4	5.7914	0.999962	1.000	0.8052	1.000	1.3439
DDDHR.11	5.7264	0.999958	1.000	0.7736	1.000	1.3435
ADHRR.2	5.6284	0.999950	1.000	0.7961	1.000	1.3377
ADHRR.3	5.6139	0.999952	1.000	0.7961	1.000	1.3368

ADHRR.5	5.5996	0.999947	1.000	0.7961	1.000	1.3359
ADHRR.6	5.5938	0.999963	1.000	0.7616	1.000	1.3356
ADHRR.4	5.6035	0.999958	1.000	0.8052	0.998	1.3344
ADHRR.1	5.6314	0.999959	1.000	0.8052	0.994	1.3324

Appendix X. Grouping the 8-MG compounds based on factor scores of descriptors by k-means clustering

Cluster	Compounds	Number of compounds in cluster	Number of compounds in test set	Test Set Compounds
1	6, 13, 14, 16, 17, 19, 20, 22, 24, 26, 27, 29, 30, 31, 55, 56, 59	17	4	17, 26, 31, 56
2	1, 3, 37, 38, 43, 44, 53, 58, 61	9	2	37, 44, 53
3	15, 32, 33, 36, 39, 40, 51, 54, 57, 62	10	2	15, 54
4	2, 18, 21, 23, 25, 28, 34, 35, 41, 42, 45, 46, 47, 48, 49, 50, 52, 60	18	5	2, 18, 34, 48, 49
5	3, 4, 7, 8, 9, 10, 11, 12	8	2	4, 8

Appendix XI. The superposition of co-crystal (violet) and redocked (green) ligand (pteroic acid) at the catalytic pocket of DHPS.

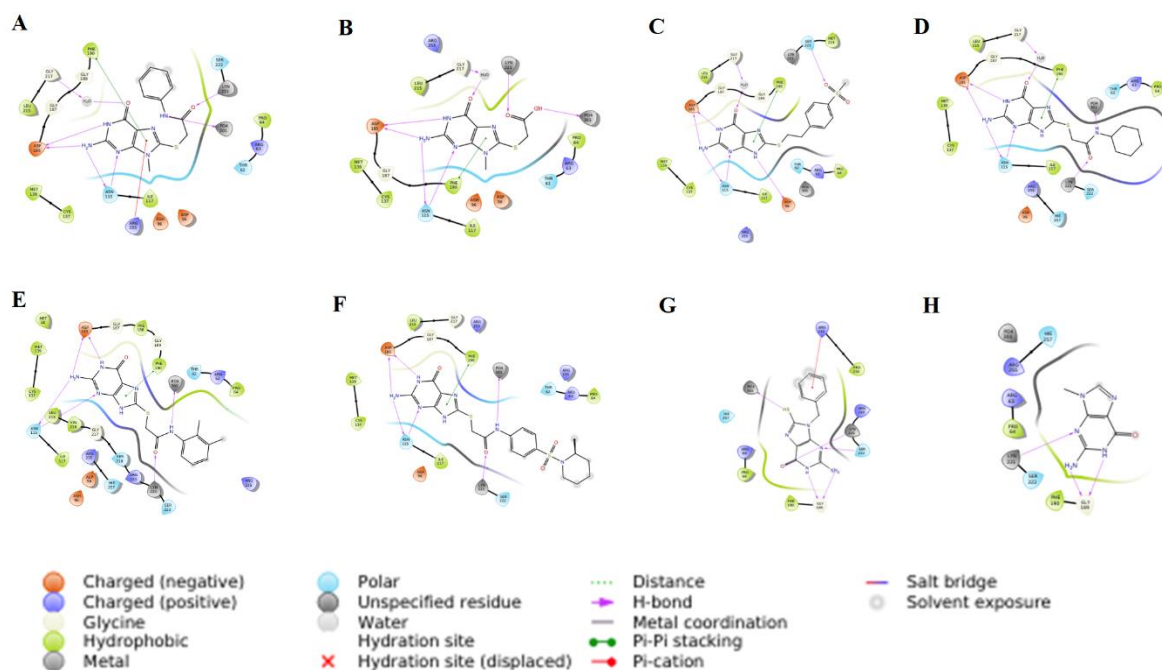


Appendix XII . The docking scores and binding free energy (Glide Emodel) predicted by Glide for 8MG derivatives at the binding pocket of DHPS.

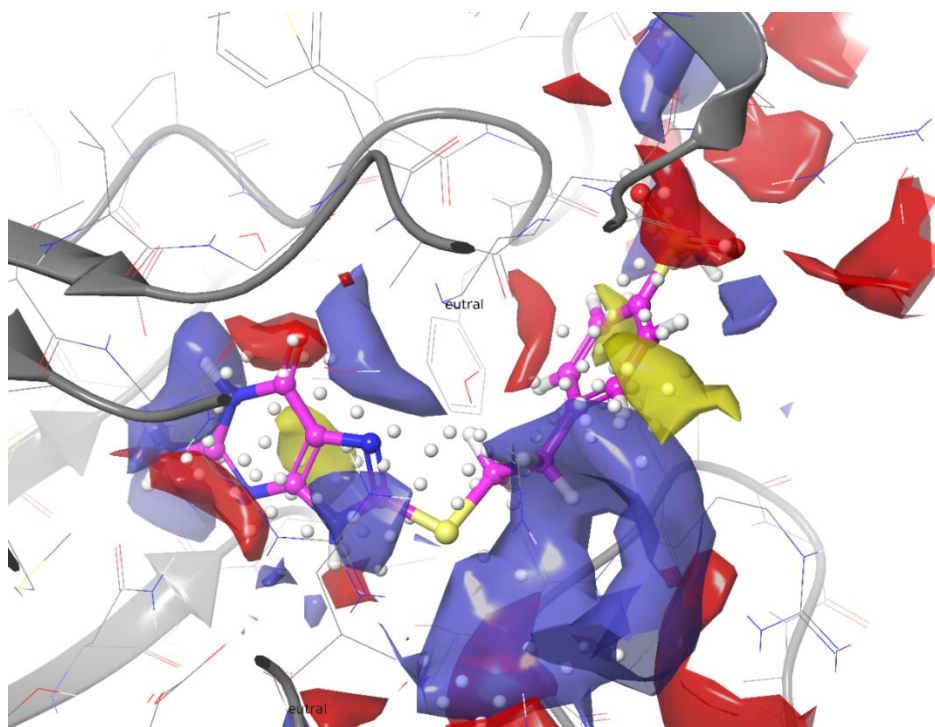
Compound	Docking Score(kJ/mol)	Glide Emodel (kJ/mol)
1	-19.2673	-192.351
2	-12.7654	-195.878
3	-26.9199	-247.747
4	-37.0284	-357.795
5	-17.9661	-235.785
6	-14.5101	-187.749
7	-25.5601	-232.367
8	-34.7649	-234.099
9	-26.9199	-247.747
10	-18.192	-192.795
11	-13.4976	-169.967
12	-24.6563	-231.312
13	-26.924	-306.482
14	-26.6186	-282.21
15	-36.6393	-283.797
16	-26.7358	-265.429
17	-27.673	-252.341
18	-22.0622	-213.652
19	-26.7107	-417.676
20	-26.8571	-290.533
21	-27.3968	-270.855
22	-28.5767	-271.914
23	-35.8276	-302.432
24	-29.5307	-477.018
25	-34.8276	-276.872
26	-36.4761	-285.453
27	-27.7357	-284.926
28	-26.8362	-274.918
29	-36.8066	-294.198
30	-37.2627	-288.512
31	-25.8027	-283.621
32	-37.0451	-324.741
33	-28.4679	-481.022
34	-35.5724	-276.244
35	-19.5686	-206.522
36	-21.6396	-218.539
37	-36.0284	-304.988
38	-39.3965	-417.764
39	-39.2041	-354.272
40	-40.8777	-370.343
41	-37.7397	-299.336
42	-34.0536	-286.993
43	-20.3928	-326.653
44	-34.359	-275.186
45	-37.2669	-304.558
46	-34.4218	-291.834
47	-37.5096	-304.595
48	-37.0912	-326.452

49	-37.2794	-304.558
50	-37.6142	-320.637
51	-43.7102	-455.985
52	-25.4052	-290.219
53	-35.6435	-344.357
54	-32.2169	-345.293
55	-29.0955	-324.54
56	-37.7522	-351.803
57	-35.3422	-469.298
58	-37.8401	-323.666
59	-28.1332	-270.659
60	-26.6019	-279.792
61	-37.8526	-324.189
62	-38.6476	-365.364

Appendix XIII. The two-dimensional representation of docking pose of highly active (A) Compound 62, (B) Compound 61, (C) Compound 51, (D) Compound 55, (E) Compound S1, (F) Compound S3 and inactive (G) Compound 2, (H) Compound 11.

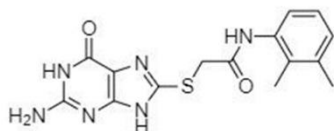


Appendix XIV The hydrophilic (blue and red colour) and hydrophobic fields (yellow colour) at the catalytic pocket of DHPS (5U10) in the presence of highly active compound 51.



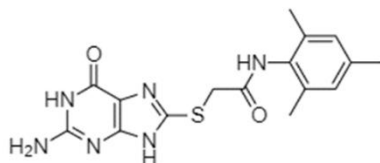
Appendix XV. The geometry of the top docked compound (A) S1, (B) S2, (C) S3, (D) S4 and (E) S5, obtained from database screening.

A



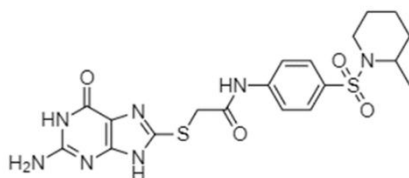
2-((2-amino-6-oxo-6,9-dihydro-1H-purin-8-yl)thio)-N-(2,3-dimethylphenyl)acetamide

B



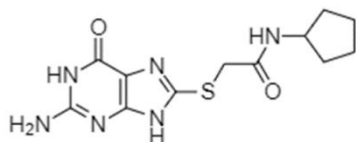
2-((2-amino-6-oxo-6,9-dihydro-1H-purin-8-yl)thio)-N-mesitylacetamide

C



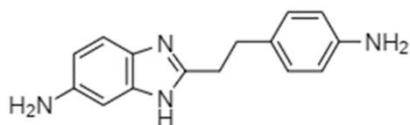
2-((2-amino-6-oxo-6,9-dihydro-1H-purin-8-yl)thio)-N-(4-((2-methylpiperidin-1-yl)sulfonyl)phenyl)acetamide

D



2-((2-amino-6-oxo-6,9-dihydro-1H-purin-8-yl)thio)-N-cyclopentylacetamide

E



2-(4-aminophenethyl)-1H-benzo[d]imidazol-6-amine

Appendix XVI. Docking results of top five Compounds obtained from database screening with respect to DDHRR_1 pharmacophore model.

Compound Name	Docking Score	Glide Emodel	Predicted Activity
S_324943 (S1)	-32.0913	-331.62	5.87
S_324945 (S2)	-30.9198	-321.553	5.63
S_325339 (S3)	-30.1666	-374.481	5.92
S_325321 (S4)	-29.8738	-274.203	5.08
S_196914 (S5)	-22.4681	-256.278	4.75

Table XVII-A. DFT based HOMO, LUMO, energy gap of the active, inactive 8MG analogues and pterioic acid for 6-31G**(d,p)

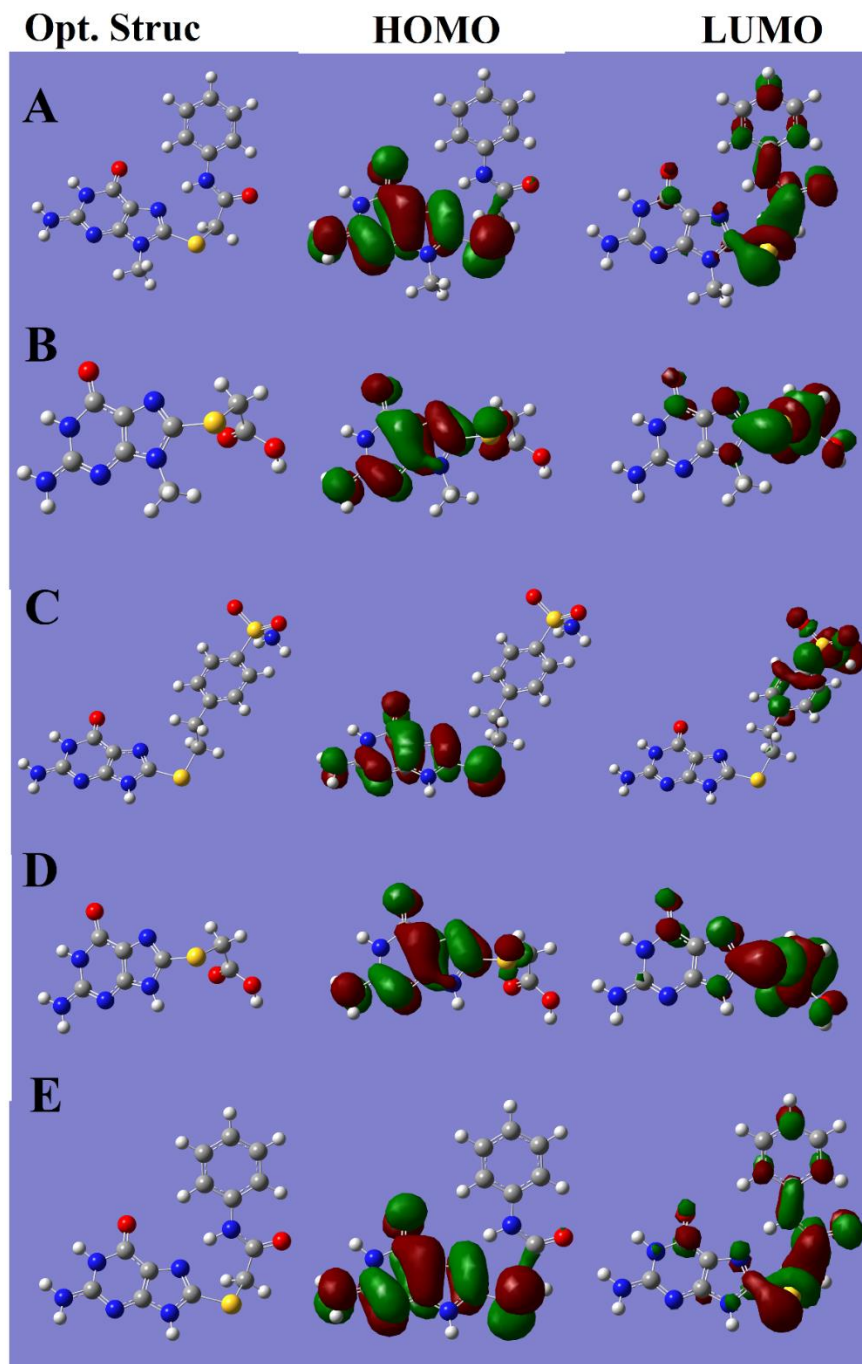
Compounds	6-31G**(d,p)				
	Vacuum			CPCM-aqueous solvent	
	E _{HOMO} (eV)	E _{LUMO} (eV)	E _{gap} (eV)	E _{HOMO} (eV)	E _{LUMO} (eV)
Compound 62	-5.45	-0.554	4.89	-5.447	-0.595
Compound 61	-5.36	-0.270	5.09	-5.476	-0.262
Compound 51	-5.57	-0.916	4.65	-5.472	-1.131
Compound 56	-5.48	-0.646	4.83	-5.476	-0.596
Compound 58	-5.43	-0.325	5.10	-5.505	-0.276
Compound 6	-5.38	-0.31	5.07	-5.797	-0.564
Compound 2	-5.58	-0.538	5.04	-5.678	-0.481
Compound 11	-5.58	-0.10	5.48	-5.702	-0.245
Compound S1	-5.48	-0.595	4.88	-5.566	-0.534
Compound S3	-5.85	-0.975	4.89	-5.705	-1.261
Pterioic Acid	-5.83	-2.934	2.90	-5.94	-2.811

Table XVII-B. DFT based HOMO, LUMO, energy gap of the active, inactive 8MG analogues and pterioic acid for 6-311G**(d,p) basis set

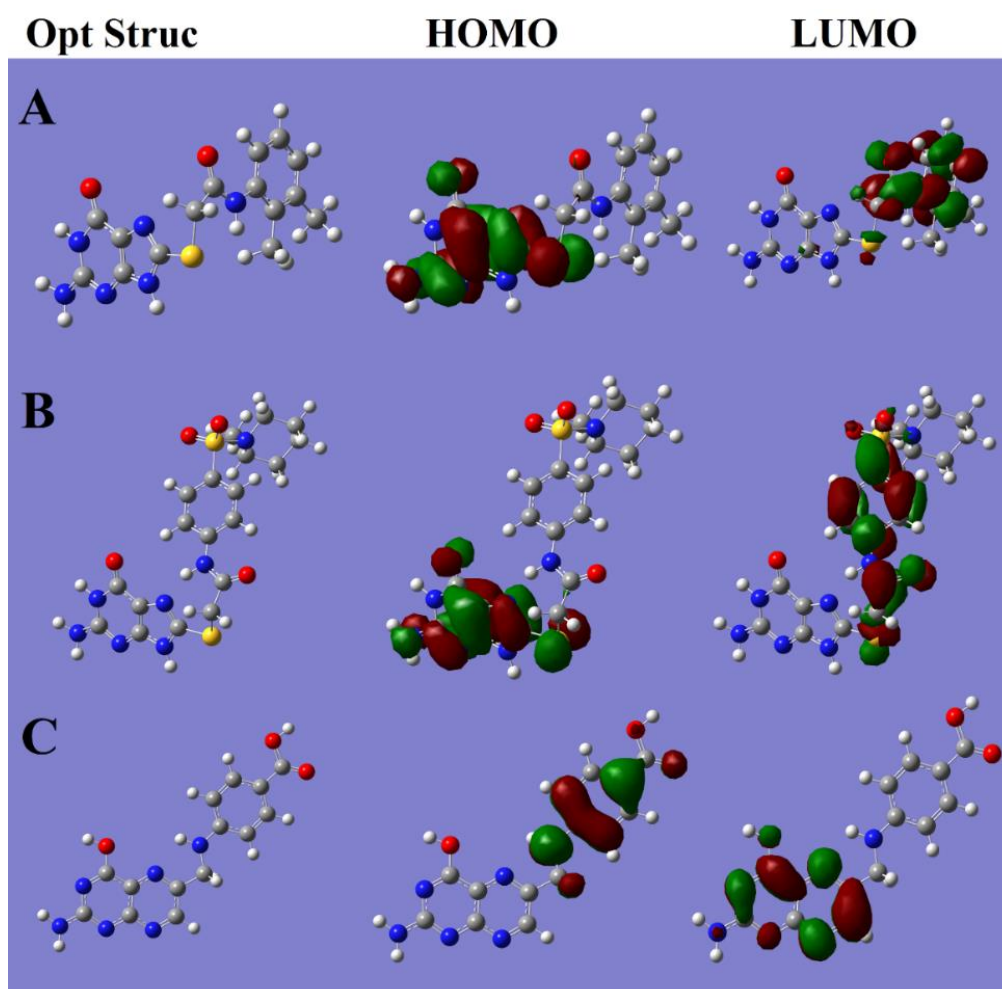
6-311G**(d,p)						
Vacuum				CPCM-aqueous solvent		
E _{gap} (eV)	E _{HOMO} (eV)	E _{LUMO} (eV)	E _{gap} (eV)	E _{HOMO} (eV)	E _{LUMO} (eV)	E _{gap} (eV)
4.85	-5.916	-1.780	4.13	-6.119	-1.512	4.60
5.20	-5.994	-1.794	4.2	-6.236	-2.001	4.23
4.34	-6.133	-1.760	4.37	-6.001	-2.151	3.85
4.87	-5.876	-1.397	4.47	-6.185	-1.544	4.64
5.22	-6.176	-1.885	4.29	-6.345	-2.022	4.33
5.23	-5.885	-0.825	5.06	-6.216	-0.903	5.31
5.18	-6.122	-0.979	5.14	-6.312	-0.915	5.39
5.45	-6.055	-0.879	5.17	-6.220	-1.070	5.15

5.03	-5.973	-1.788	4.18	-6.497	-2.245	4.25
4.43	-6.18	-1.740	4.44	-6.672	-2.317	4.35
3.12	-5.872	-2.965	2.90	-5.971	-3.125	2.84

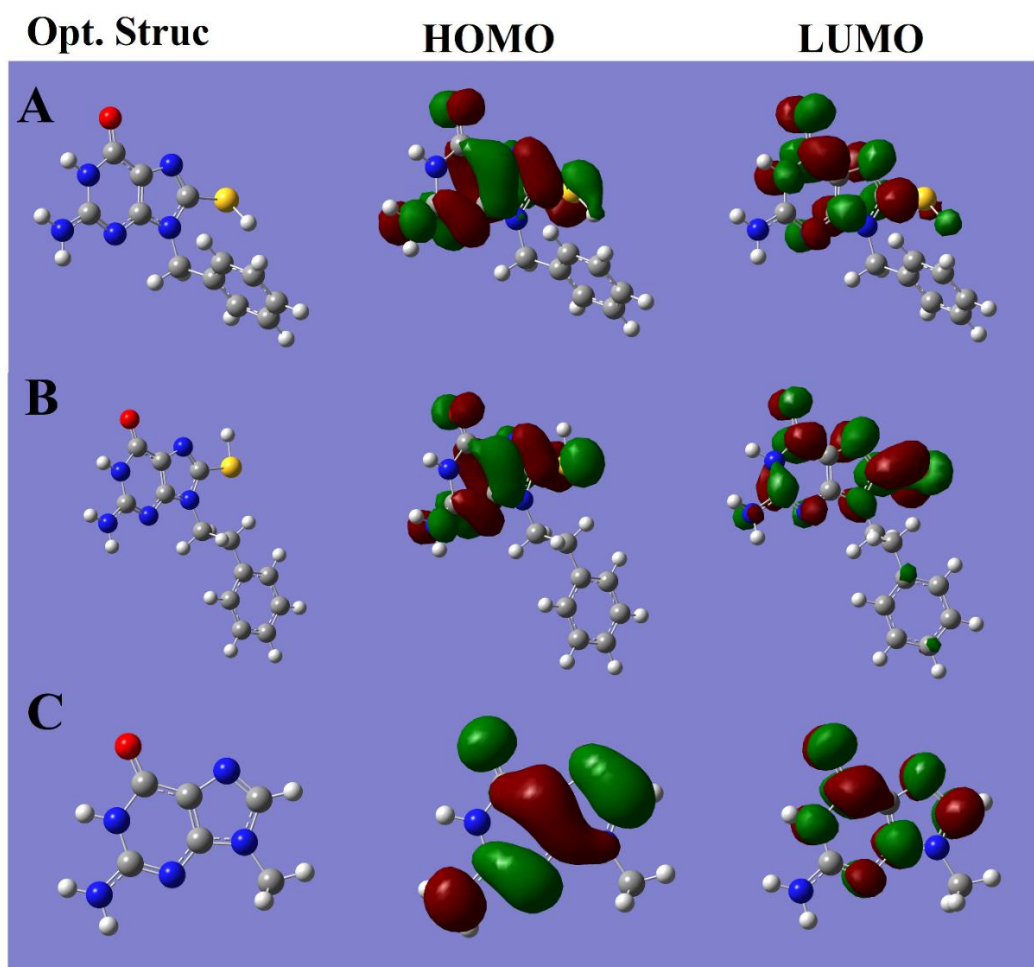
Appendix XVIII. The location of HOMO-LUMO on highly active (A) Compound 62, (B) Compound 61, (C) Compound 51, (D) Compound 58 and (E) Compound 56.



Appendix XIX. The location of HOMO-LUMO on highly active (A) Compound S1, (B) Compound S3 and (C) Pterioic acid.



Appendix XX. The location of HOMO-LUMO on inactive (A) Compound 2, (B) Compound 6 and (C) Pteric acid.



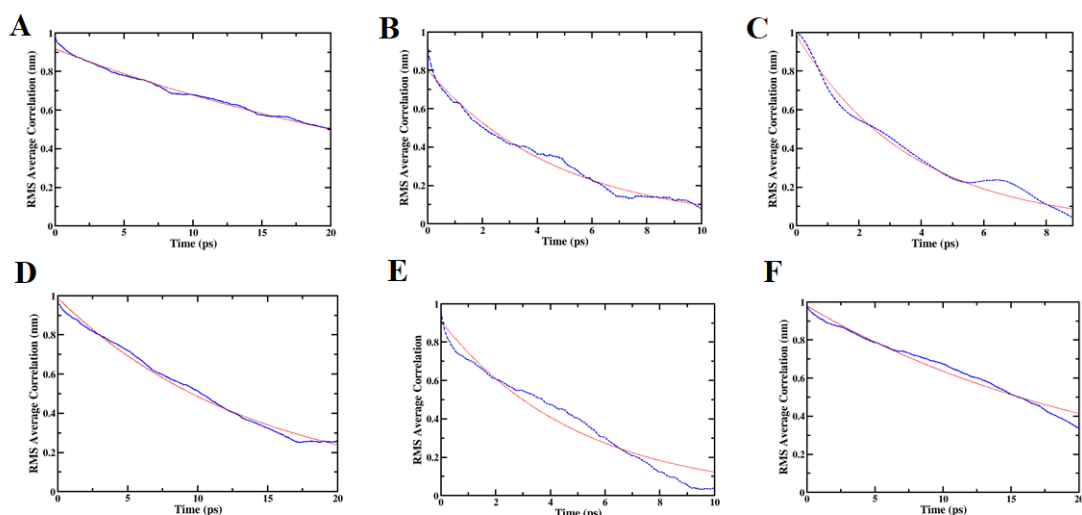
Appendix XXI. Chemical reactivity descriptors of 8MG derivatives calculated correspond to B3LYP/6-311G*(d, p) basis set.

Compound	(χ , eV)	(μ)	(η , eV)	(s , eV ⁻¹)	(ω , eV ²)	Dipole moment (Debye)
62	3.815	-3.815	2.30	0.21	3.163	4.807
61	4.118	-4.118	2.11	0.23	4.018	6.498
51	4.076	-4.076	1.92	0.26	4.326	13.384
56	3.864	-3.864	2.32	0.21	3.217	11.134
58	4.178	-4.178	2.16	0.23	4.040	6.185
S1	4.371	-4.371	2.12	0.23	4.506	11.777
S3	4.494	-4.494	2.17	0.23	4.650	15.520
6	3.559	-3.559	2.65	0.18	2.389	11.319
2	3.613	-3.613	2.69	0.18	2.426	12.272
11	3.645	-3.645	2.57	0.19	2.584	10.924
Pterioic Acid	4.548	-4.548	1.42	0.35	7.283	7.232

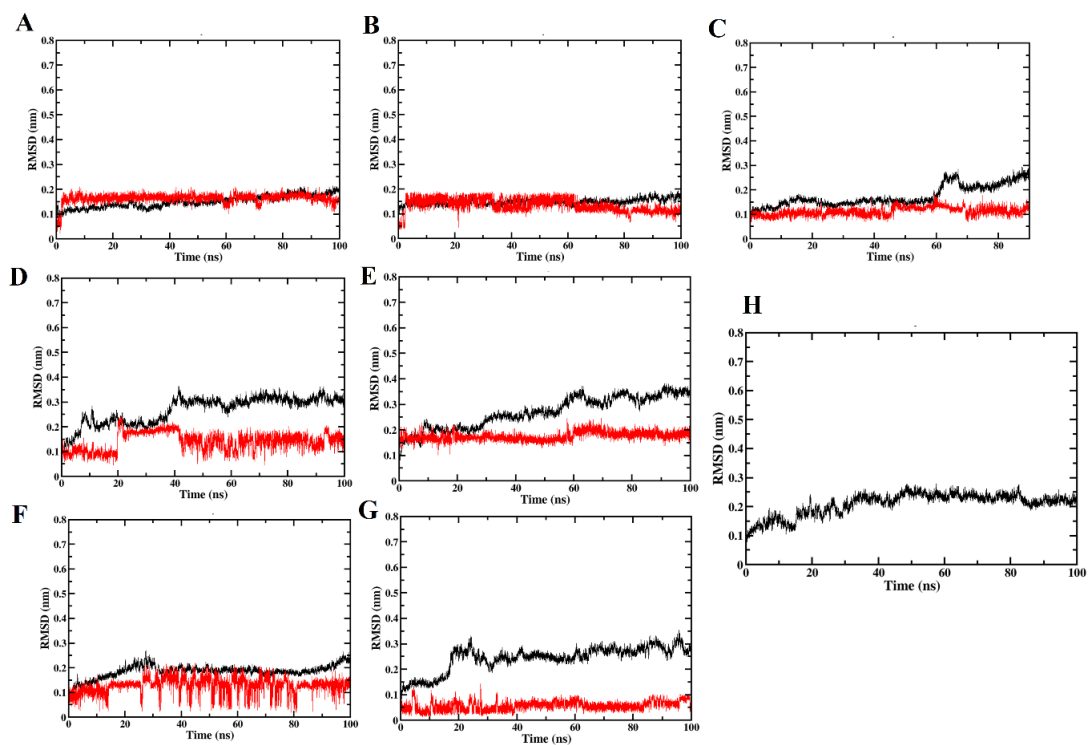
Appendix XXII. The ADME/TOX properties of active 8MG derivatives

Compounds	Molecular Weight	HBDDHBA	Rotatable Bond	QPLogPo/w	QPLogS	QPlogHERG	QPPCaco	QPlogBB	QPlogKhsa	QPPMDCK	Dipole	PSA	HumanOral Absorption
Compound 51	366.4	5	9	0.12	-3.51	-5.44	45.55	-2.37	-0.74	2.47	15.67	143.16	47.56
Compound 62	330.3	4	8	0.73	-3.72	-5.65	62.03	-1.99	-0.39	37.37	14.54	134.38	63.38
Compound 61	255.2	4	7	-0.63	-2.09	-5.95	30.9	-2.2	-0.87	2.39	9.97	148.8	33.84
Compound 58	241.2	5	7	-1.22	-1.88	-4.99	1.95	-2.47	-0.96	1.21	10.47	161.71	24.97
Compound 57	380.2	4	7	0.69	-3.87	-5.31	33.71	-2.04	-0.42	54.58	12.06	137.71	58.38
Compound 56	316.3	5	8	0.11	-3.42	-5.78	33.84	-2.29	-0.55	20.67	11.8	144.78	54.99
Compound 55	322.3	5	8	-0.47	-2.61	-3.41	19.75	-2.26	0.76	21.33	11.4	145.47	47.51
Compound 54	331.3	4	8	0.39	-3.2	-5.06	53.27	-2	-0.51	36.1	5.61	141.32	60.13
Compound 53	254.2	5	8	-1.76	-1.16	-2.87	12.66	-2.15	-1.09	14.86	11.91	148.18	36.34
Compound 52	297.3	4	7	0.14	-3.35	-4.6	40.07	2.55	-0.47	23.26	10.19	140.05	56.46
Compound 50	280.3	5	8	-1.22	-1.87	-3.36	16.48	-2.32	-1.02	18.313	8.59	146.41	41.53
Compound 49	296.3	5	8	-0.79	-2.19	-3.2	20.91	-2.26	-0.91	20.92	7.76	145.44	45.91
Compound 48	296.3	5	8	-0.9	-1.83	-3.24	15	-2.47	-0.96	15.08	14.73	147.96	42.76
Compound 47	296.3	4	8	-0.76	-1.19	-1.55	57.45	-1.39	-0.92	57.45	9.72	133.8	53.94
Compound 46	310.3	4	9	-1.3	-1.06	-1.24	59.37	-1.16	-0.96	59.37	6.75	140.63	51.07
Compound 45	296.3	5	8	-0.77	-2	-3.15	22.73	-2.17	-0.92	22.73	10.37	144.55	46.67
Compound 44	239.25	4	7	-1	-2.11	-3.9	30.55	-1.98	-0.75	30.55	10.61	136.98	47.66
Compound 43	255.25	5	7	-0.89	-2.06	-2.52	2.38	-2.55	-0.95	2.38	10.27	158.42	28.5
Compound 42	268.29	4	8	-1.38	-1.6	-2.82	23	-1.94	-1.03	23	7.53	138.02	43.22
Compound 41	296.34	5	8	-0.76	-2.48	-3.28	23.65	-2.19	-0.88	23.65	11.45	142.24	47.02
Compound 40	364.8	5	8	0.1	-2.88	-4.03	17.24	-2.23	-0.73	17.24	11.33	144.98	49.7
Compound_S1	344.39	5	8	0.53	-3.94	-5.39	34.57	-2.31	-0.35	18.65	12.52	142.99	57.61
Compound_S2	358.41	5	8	0.82	-4.25	-5.19	43.2	-2.23	-0.24	22.97	11.77	141.16	61.06
Compound_S3	477.55	5	11	0.15	-4.52	-5.58	43.2	-3.03	-0.43	3.79	16.96	183.05	30.09
Compound_S4	308.35	5	8	-0.706	-2.304	-5.36	72.22	-2.24	-0.85	21.23	9.91	145.2	45.86
Compound_S5	252.31	4	3	1.93	-3.34	-4.6	19.4	-1.3	-0.14	122.56	5.48	80.79	81.91

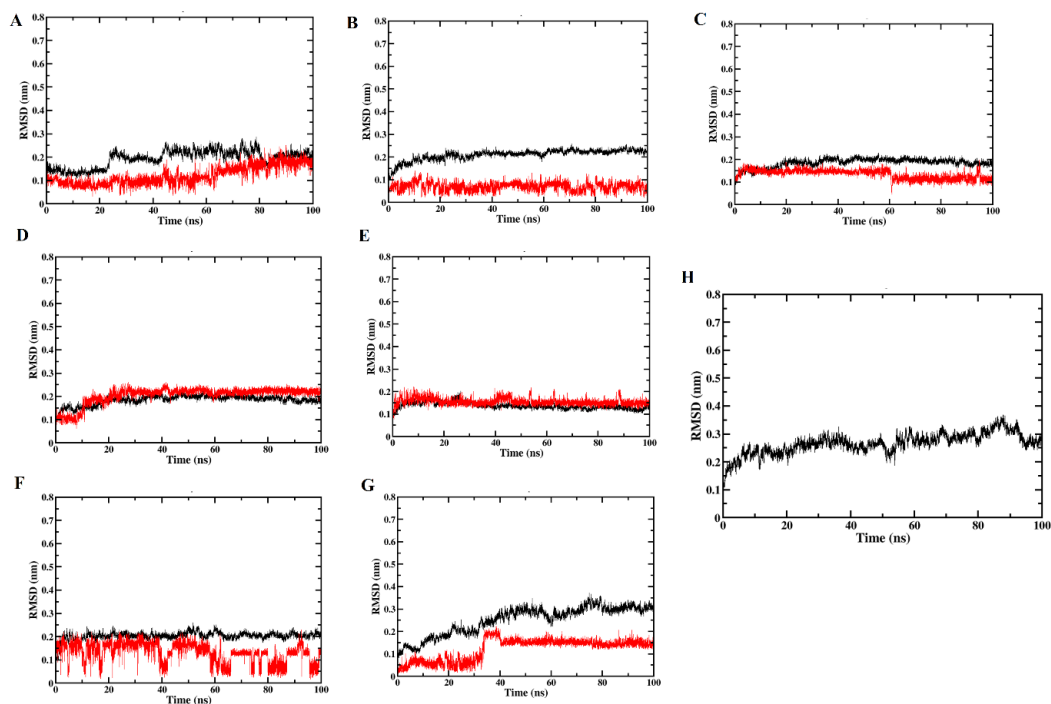
Appendix XXIII. The RMS average correlation for DHPS backbone bound with (A) Comp62, (B) Comp61, (C) Comp51, (D) CompS1, (E) CompS3 and (F) Comp 6.



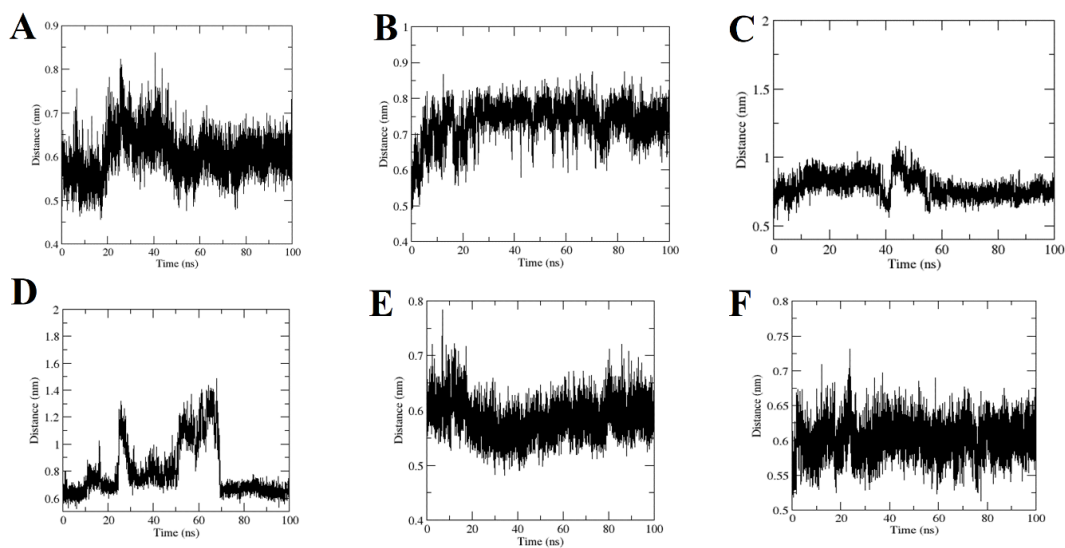
Appendix XXIV. The time evolution of RMSD profile for DHPS bound with (A) Compound 62, (B) Compound 61, (C) Compound 51, (D) Compound S1, (E) Compound S3, (F) Compound 6, (G) Pteric acid and (H) Apo-DHPS. The black color represents the backbone RMSD and red color represents the ligand RMSD. (Simulation_batch 2)



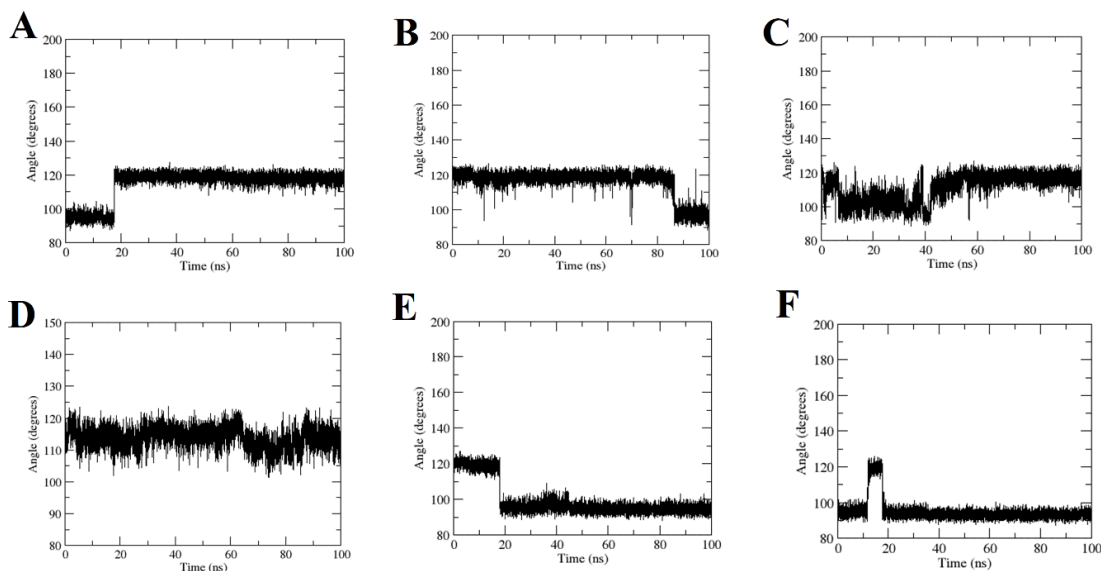
Appendix XXV. The time evolution of RMSD profile for DHPS bound with (A) Compound 62, (B) Compound 61, (C) Compound 51, (D) Compound S1, (E) Compound S3, (F) Compound 6, (G) Pteric acid and (H) Apo-DHPS. The black color represents the backbone RMSD and red color represents the ligand RMSD. (Simulation_batch 2)



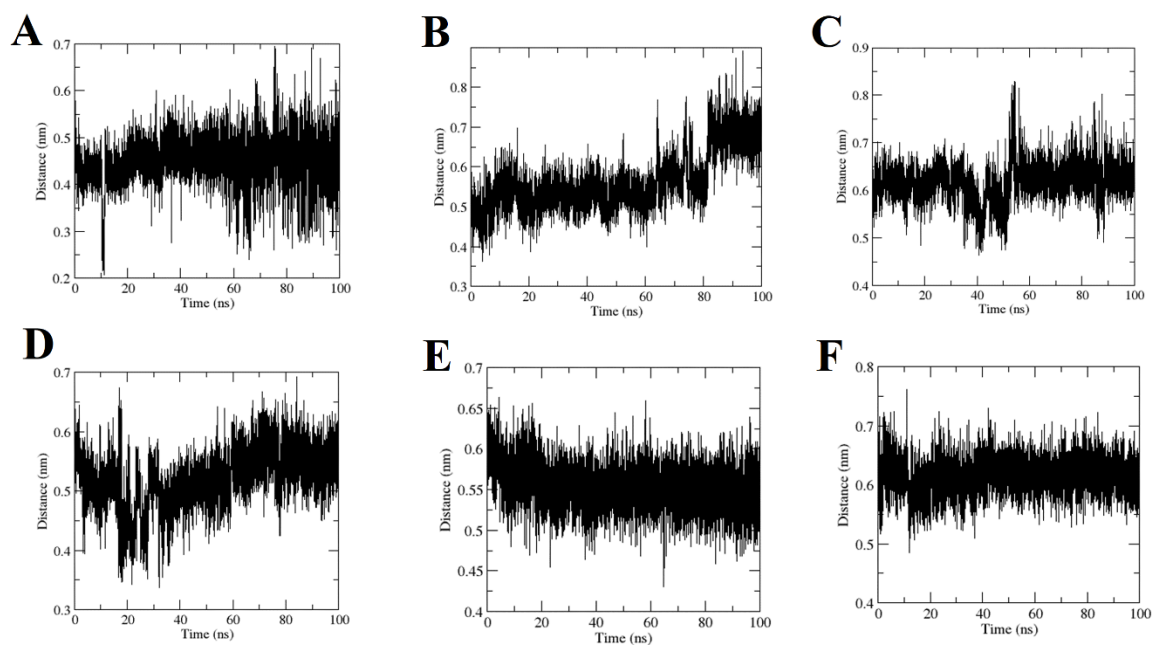
Appendix XXVI. The time evolution of distance between the surfaces of Phenylalanine and the aromatic rings of (A) compound 62, (B) Compound 61, (C) compound 51, (D) Compound S1, (E) Compound S3 and (F) Pteric acid.



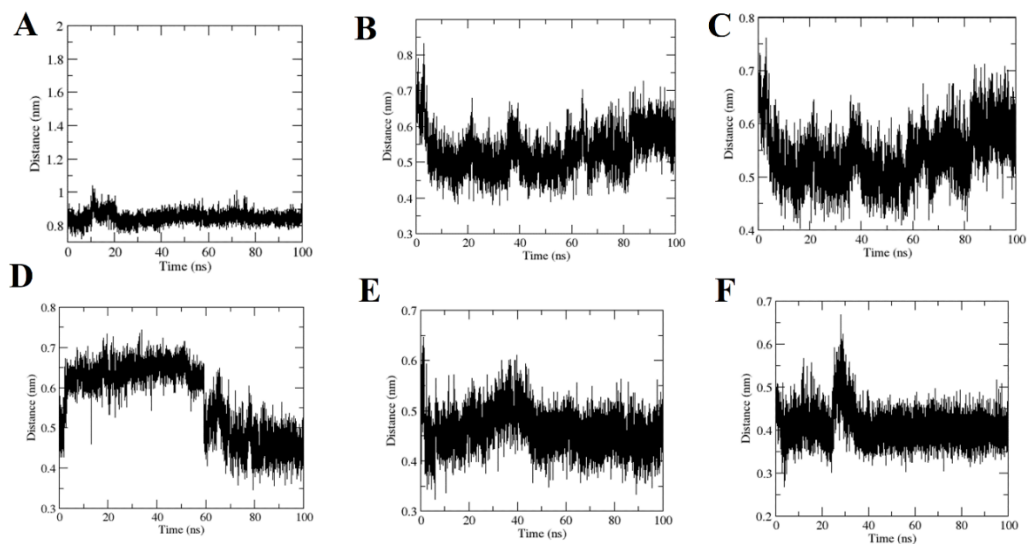
Appendix XXVII. The time evolution of angle between the surfaces of Phenylalanine 190 and the aromatic rings of (A) compound 62, (B) Compound 61, (C) compound 51, (D) Compound S1, (E) Compound S3 and (F) Pteric acid.



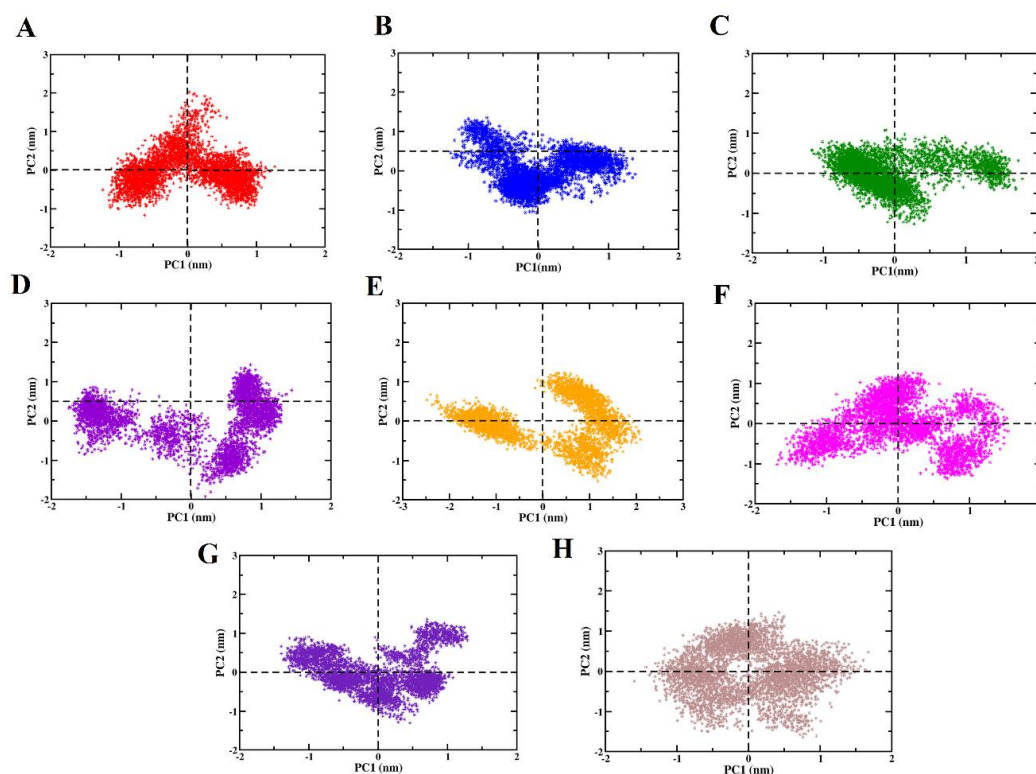
Appendix XXVIII. The time evolution of distance between the guanidium of Lys 221 and the aromatic guanine rings of (A) compound 62, (B) Compound 61, (C) compound 51, (D) Compound S1, (E) Compound S3 and (F) Pteric acid.



Appendix XXIX. The time evolution of distance between the guanidium of Arg 255 and the aromatic guanine rings of (A) compound 62, (B) Compound 61, (C) compound 51, (D) Compound S1, (E) Compound S3 and (F) Pteric acid.



Appendix XXX. The two-dimensional projection of principal component 1 (PC1) and principal component 2 (PC2) of DHPS backbone, as the horizontal axis and vertical axis respectively for (A) Compound 62, (B) Compound 61, (C) Compound 51, (D) Compound S1, (E) Compound S3, (F) Compound 6, (G) Pteric acid and (H) Apo-state.



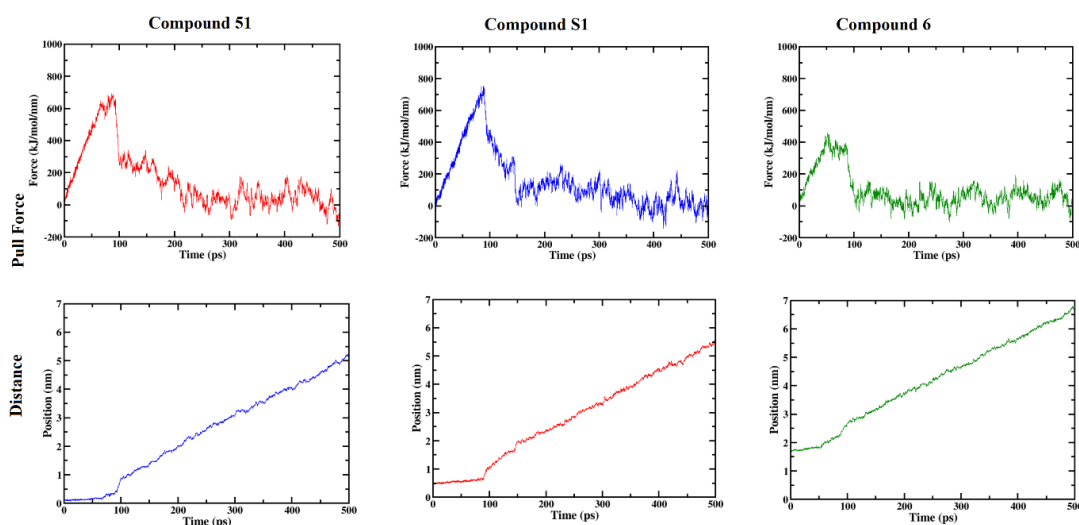
Appendix XXXI. Binding Free energy (kJ/mol) of the 8MG and pteric acid at the DHPS binding pocket by MM/PBSA method (Sim-II)

Compounds	$\Delta G_{\text{bind-eff}}$ (kJ/mol)	ΔG_{vdw} (kJ/mol)	ΔG_{ele} (kJ/mol)	$\Delta G_{\text{sol-pol}}$ (kJ/mol)	ΔG_{SASA} (kJ/mol)
Compound 62	-49.04	-152.56	-121.85	246.83	-17.64
Compound 61	-58.56	-132.92	-139.91	228.60	-14.33
Compound 51	-65.92	-131.66	-159.04	237.92	-17.06
Compound 6	-25.23	-113.27	-77.83	176.78	-9.915
Compound_S1	-61.76	-140.69	-180.11	273.21	-14.17
Compound_S3	-76.65	-162.041	-226.65	332.754	-20.71
Product	-43.084	-114.48	-116.79	199.31	-13.02

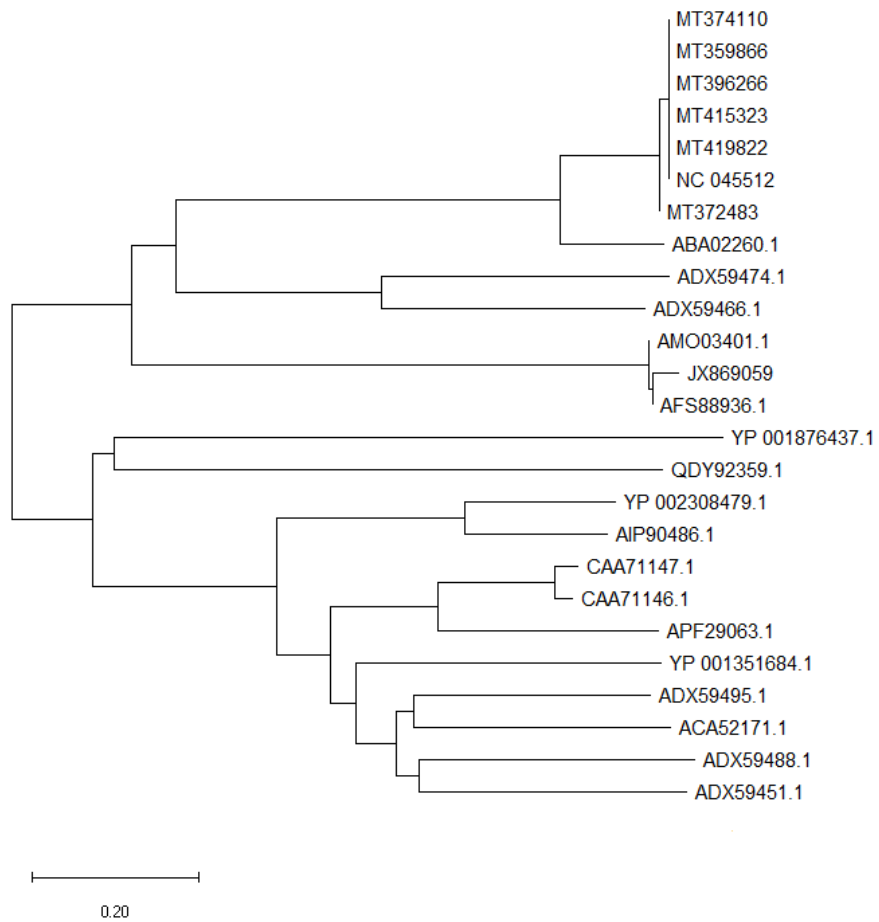
Appendix XXXII. Binding Free energy (kJ/mol) of the 8MG and pterotic acid at the DHPS binding pocket by MM/PBSA method (Sim-III)

Compounds	$\Delta G_{\text{bind-eff}}$ (kJ/mol)	ΔG_{vdw} (kJ/mol)	ΔG_{ele} (kJ/mol)	$\Delta G_{\text{sol-pol}}$ (kJ/mol)	ΔG_{SASA} (kJ/mol)
Compound 62	-51.21	-182.93	-131.22	276.65	-15.71
Compound 61	-61.72	-145.54	-151.78	248.60	-13.00
Compound 51	-68.31	-155.723	-169.04	274.92	-18.472
Compound 6	-27.66	-125.99	-62.83	171.48	-10.329
Compound_S1	-63.67	-162.69	-176.72	294.21	-18.47
Compound_S3	-70.17	-218.94	-240.65	410.41	-20.99
Product	-46.82	-147.60	-144.22	259.81	-14.81

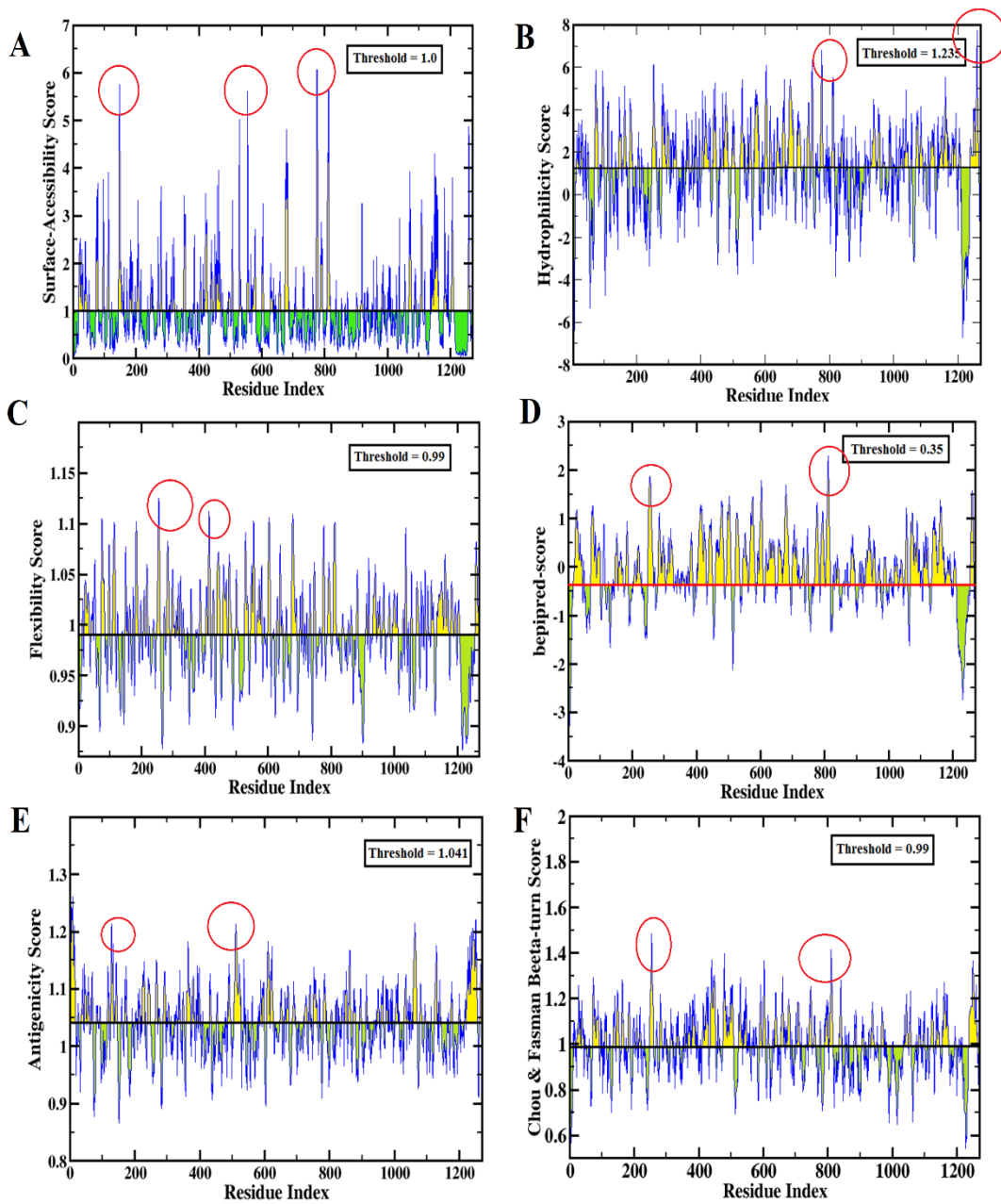
Appendix XXXIII. The time evolution of pulling force and the distance between the selected vectors for initial definition of reaction coordinate.



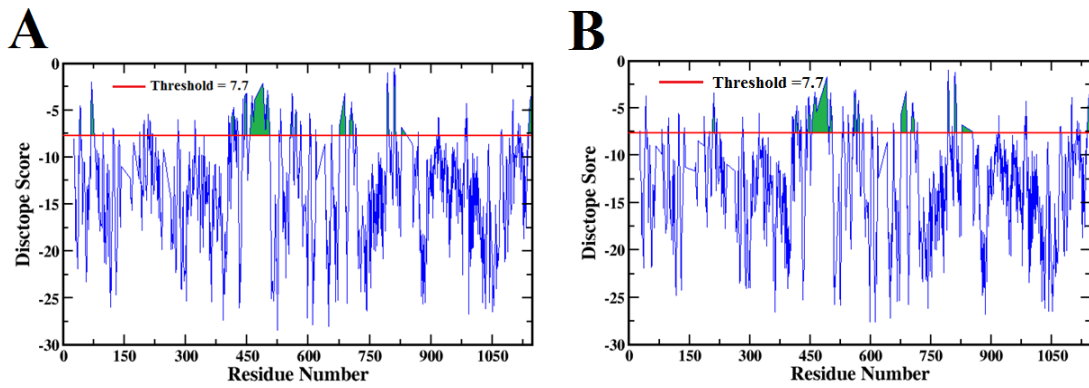
Appendix XXXIV. Phylogenetic tree based on the protein sequence of spike-glycoprotein



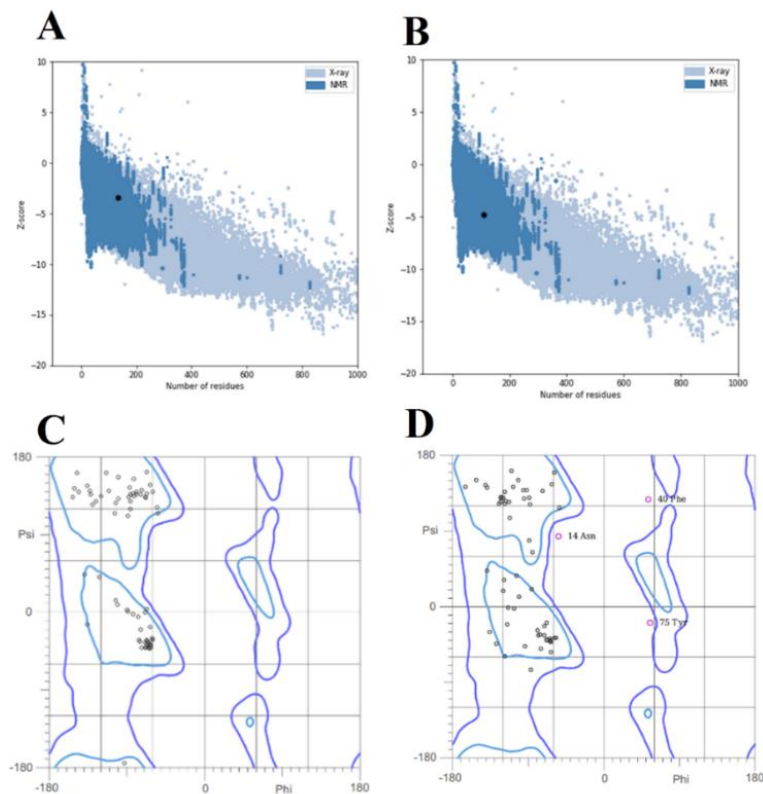
Appendix XXXV. The residue wise antigenicity score predicted by (A) Emini surface accessibility (B) parker hydrophilicity (C) Karplus Flexibility (D) BepiPred Linear Epitope (E) Kolaskar & Tongaonkar Antigenicity (F) Chou-Fasman Beta-Turn of SARS-CoVid-2 spike protein.



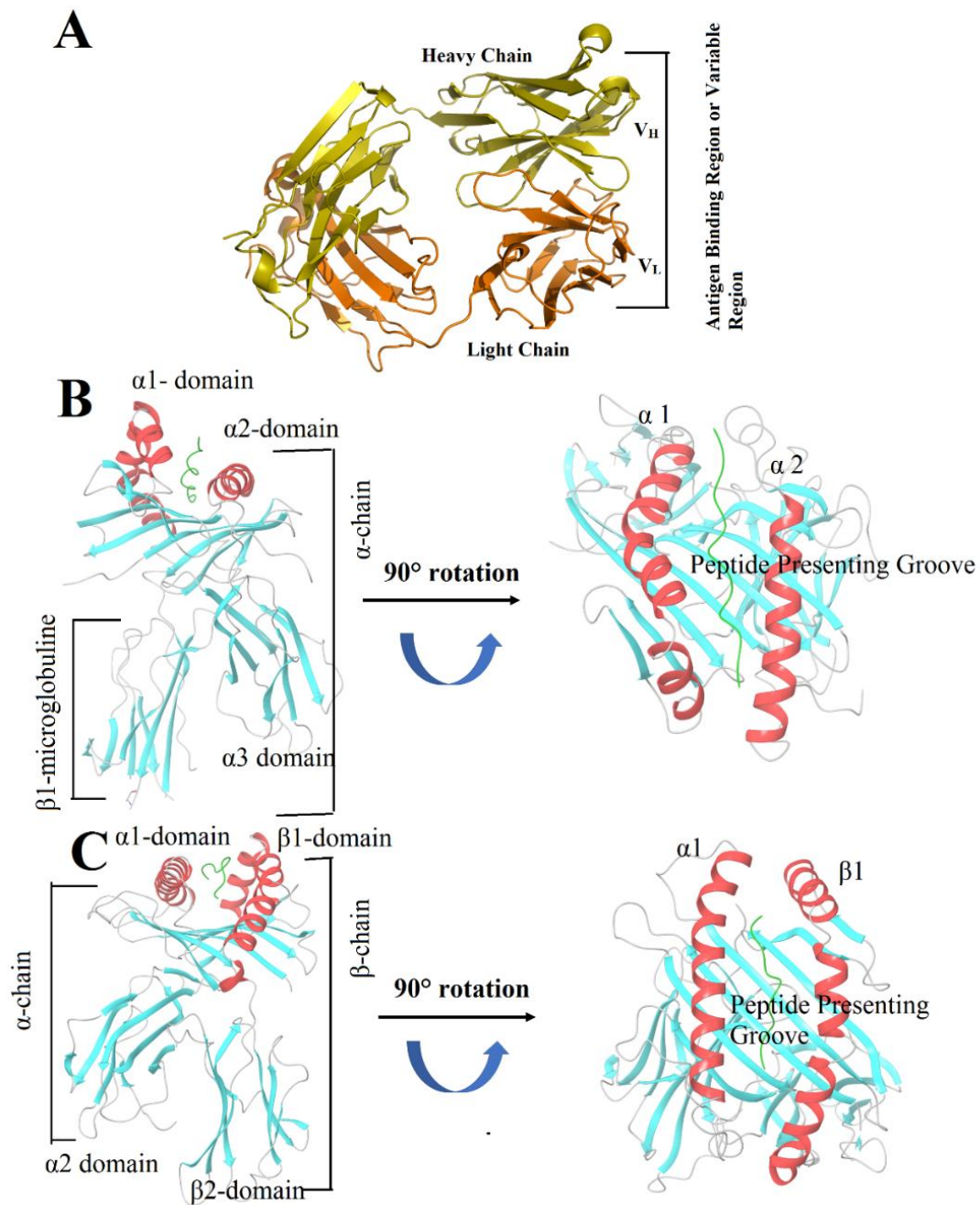
Appendix XXXVI. The location of conformational B-cell epitopes (A) open state, (B) closed state.



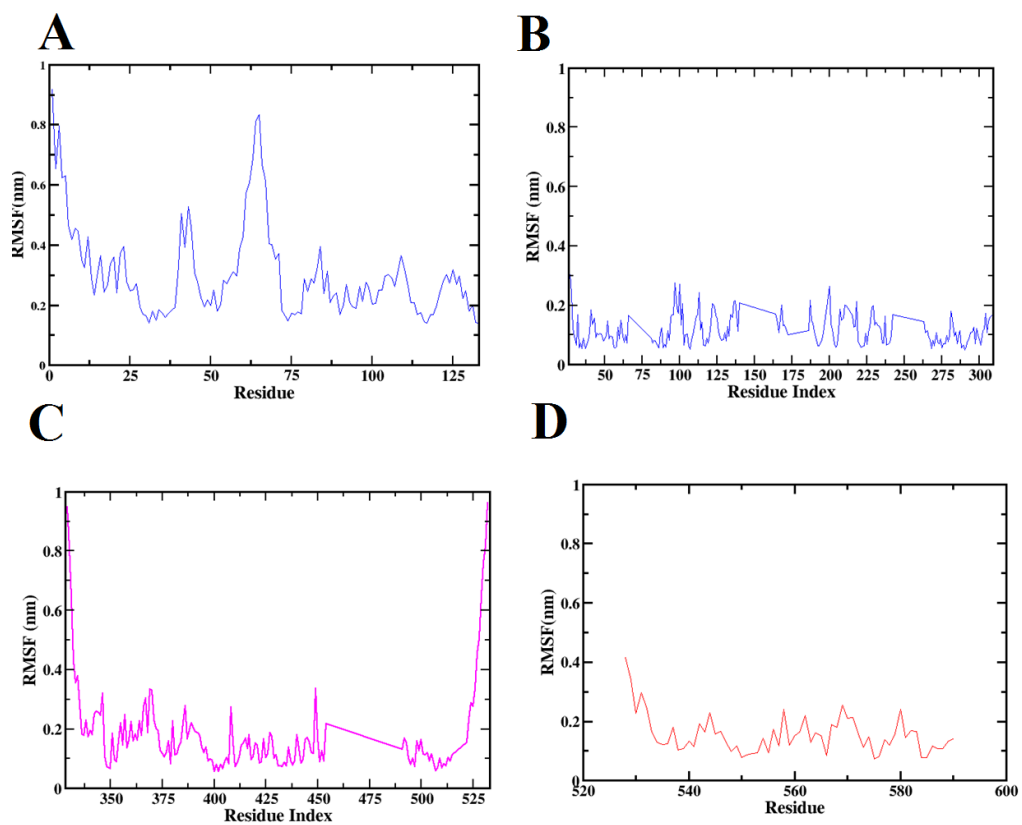
Appendix XXXVII. The Z score of (A) Vac-COVID-B, (B) Vac-COVID-T. The Ramachandran plot of (C) Vac-COVID-B, (D) Vac-COVID-T.



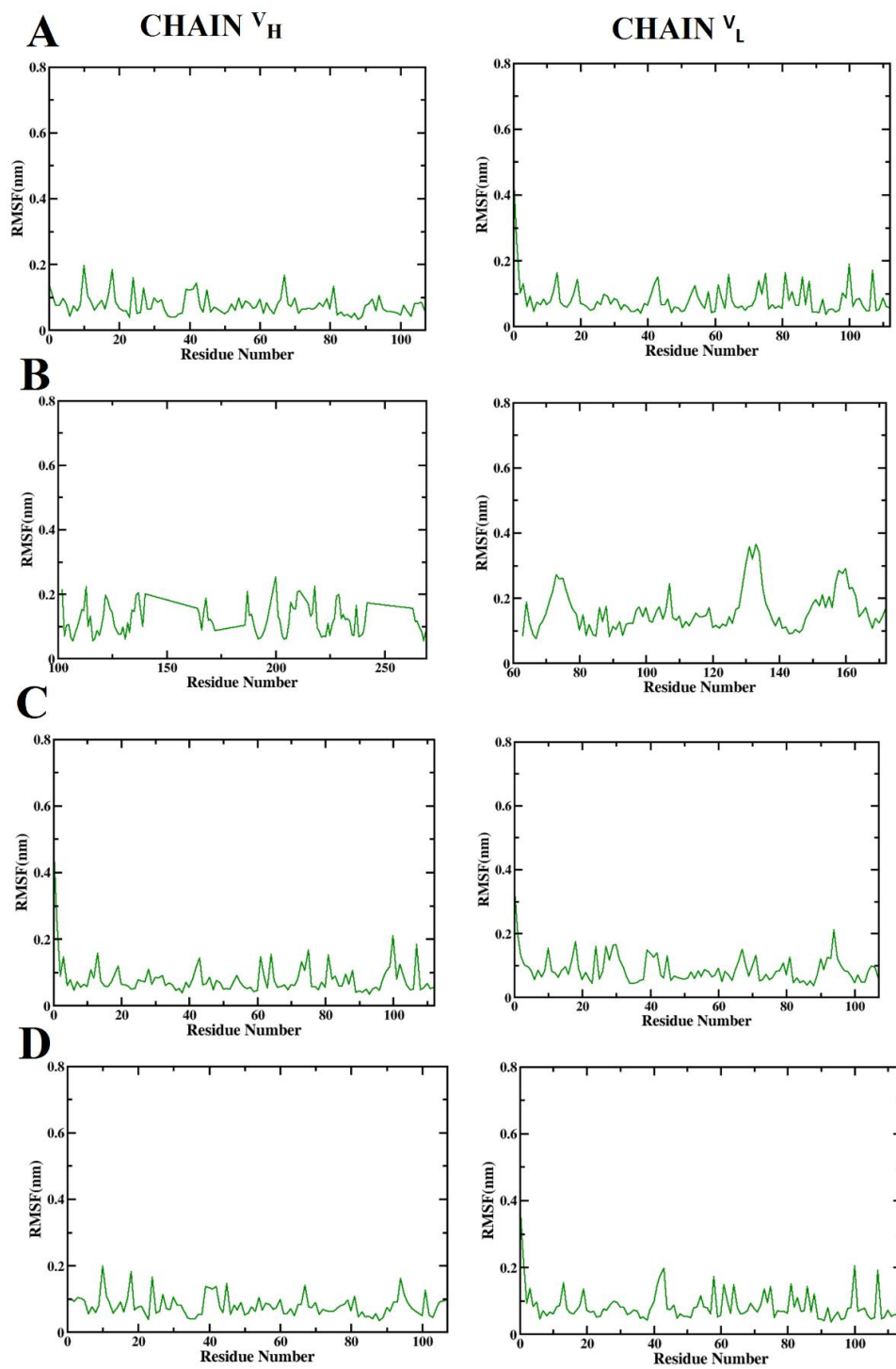
Appendix XXXVIII. The structural illustration of (A) antibody (PDB ID: 7B5) (B) MHC-I molecule (PDB ID: 2GTZ) (C) MHC-II molecule (PDB ID: 2SEB).



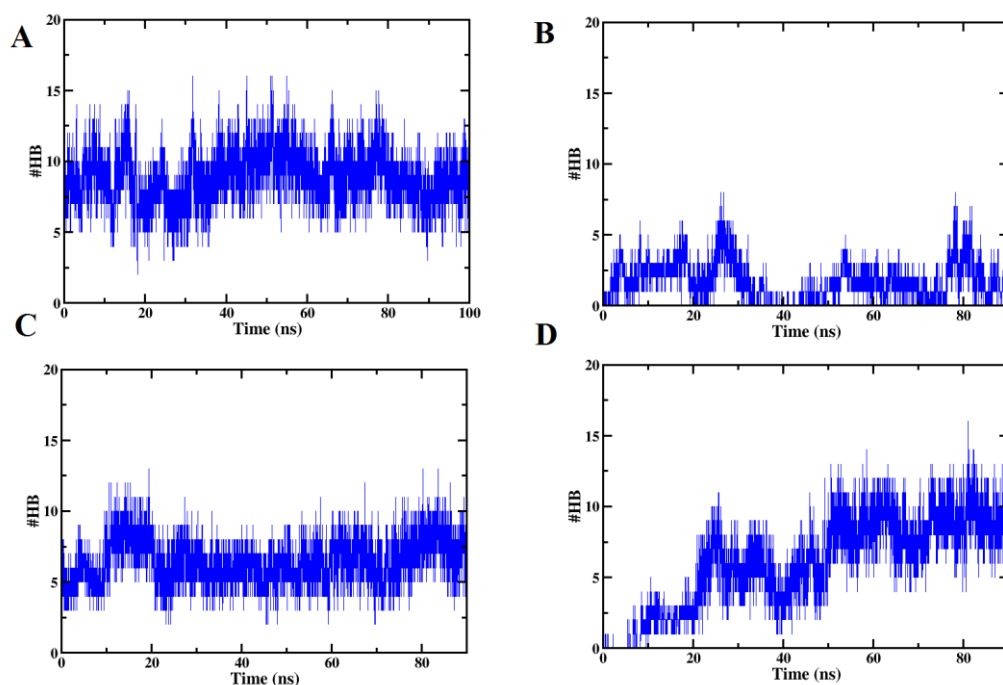
Appendix XXXIX. The RMSF profile of (A) Vac-COVID-B, (B) R1, (C) R2, (D) R3 during MD simulation



Appendix XXXX. The RMSF profile of 7BZ5 associated with (A) Vac-COVID-B, (B) R1, (C) R2, (D) R3. In first column we showed RMSF profile of V_H and in second column the RMSF profile of V_L is depicted.



Appendix XXXXI. The time evolution of hydrogen bond between 7BZ5 and (A) Vac-COVID-B, (B) R1, (C) R2 (D) R3.



Appendix XXXXII-A. The occupancy of the hydrogen bonds formed between the Vac-COVID-B and 7BZ5 during MD.

Pair ID	Donor	Acceptor	Occupancy (%)
1	92ASN(D21)	71GLN(O)	17.9
2	92ASN(D21)	53PRO(O)	11.3
3	52TYR(HH)	54GLY(O)	13.1
4	0ASP(H1)	76GLY(O)	63.9
5	58TYR(HH)	51PRO(O)	69.7
6	33TYR(HH)	56THR(O)	11.3
7	97TYR(HH)	54GLY(O)	59.4
8	88ARG(H21)	0ASP(OD2)	65.1
9	88ARG(H21)	0ASP(OD1)	25.3
10	88ARG(H11)	0ASP(OD1)	63.9
11	86ARG(H11)	0ASP(OD2)	16.1
12	86ARG(H11)	0ASP(OD1)	10
13	86ARG(H)	0ASP(OD1)	12.1

14	80THR(HG1)	28GLY(O)	10.5
15	80THR(HG1)	27GLN(OE1)	24.2
16	79GLN(E21)	30SER(OG)	16
17	79GLN(E21)	28GLY(O)	13.1
18	78THR(HG1)	27GLN(OE1)	10.7

Appendix XXXXII-B. The occupancy of the hydrogen bonds formed between the R1 and 7BZ5 during MD.

Pair ID	Donor	Acceptor	Occupancy (%)
1	53SER(HG)	81ASN(OD1)	12
2	33TYR(HH)	138ASP(OD2)	10.3
3	33TYR(HH)	81ASN(OD1)	11.8
4	137ASN(D21)	33 TYR(OH)	32.6

Appendix XXXXII -C. The occupancy of the hydrogen bonds formed between the R2 and 7BZ5 during MD.

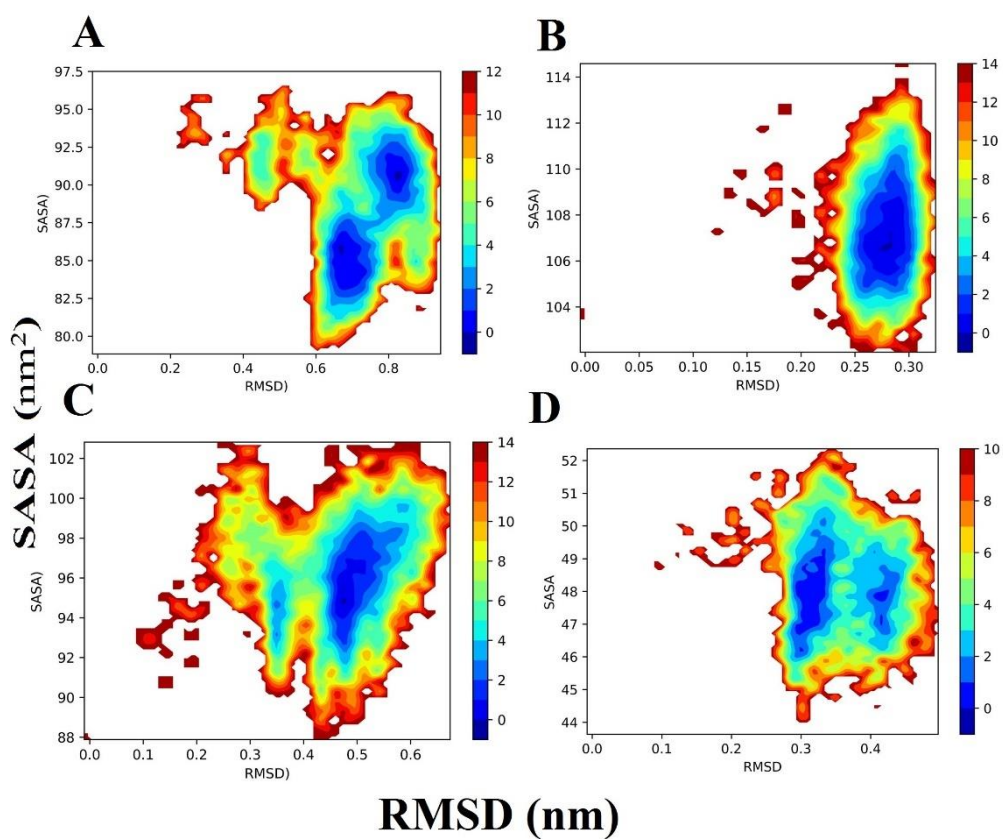
Pair ID	Donor	Acceptor	Occupancy (%)
1	Gly28	Tyr449	20.2
2	Tyr32	Ser494	51.6
3	Asn92	Tyr453	99.4
4	Tyr58	Arg403	69.3
5	Tyr52	Asp405	18.9
6	Tyr33	Lys417	21.4

Appendix XXXXII -D. The occupancy of the hydrogen bonds formed between the R3 and 7BZ5 during MD.

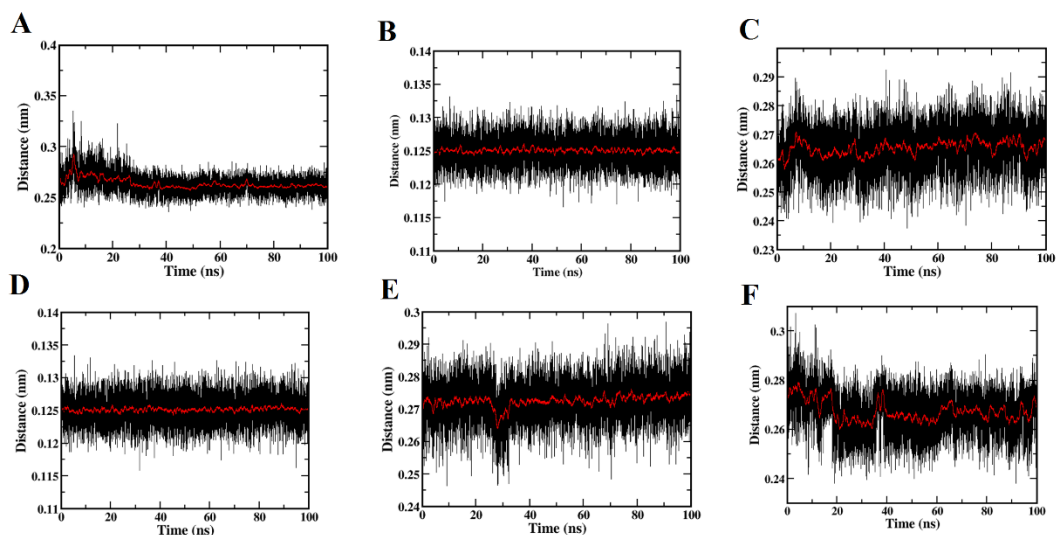
Pair ID	Donor	Acceptor	Occupancy (%)
1	94TYR(HH)	543PHE(O)	14.2
2	56SER(HG)	564GLN(OE1)	55.9
3	56SER(H)	571ASP(OD1)	14.7
4	55GLY(H)	571ASP(OD1)	26.3
5	54GLY(H)	571ASP(OD1)	11.7
6	54GLY(H)	571ASP(OD2)	41.1

7	33TYR(HH)	571ASP(OD1)	31.9
8	567ARG(H21)	571ASP(OD2)	21.7
9	567ARG(H21)	98GLU(OE2)	31
10	567ARG(H11)	98GLU(OE1)	32.8
11	567ARG(H11)	98GLU(OE2)	22.6
12	564GLN(E21)	98GLU(OE1)	17.6
13	547THR(HG1)	31SER(O)	38.3
14	544ASN(H)	94TYR(OH)	18.7
15	528LYS(H1)	92ASN(O)	51.6
16	528LYS(H1)	27GLN(OE1)	20.8
17	528LYS(H1)	0ASP(O)	14.6
18	56SER(HG)	0ASP(OD1)	11.7

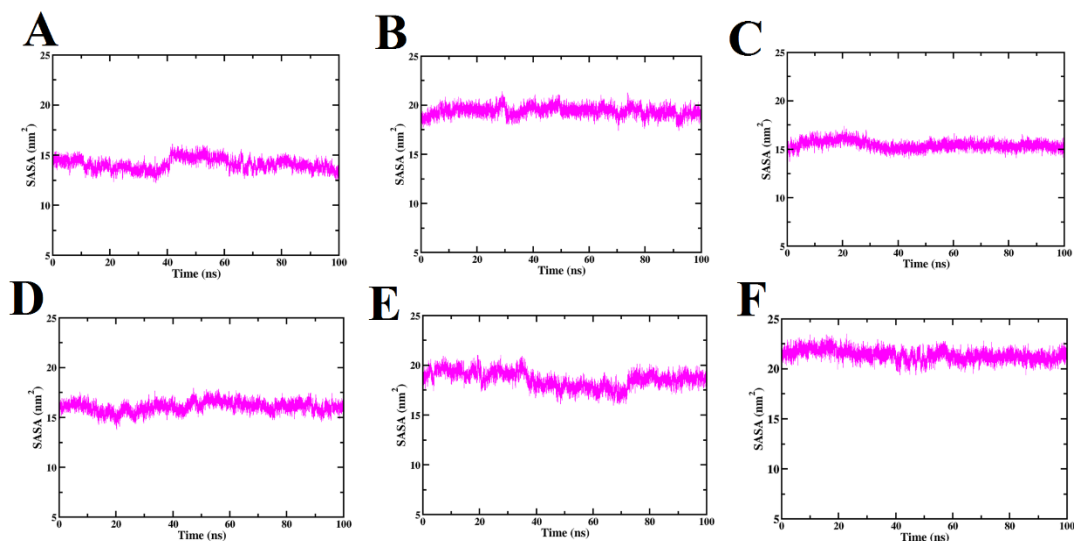
Appendix XXXXIII. The free energy landscape of (A) Vac-COVID-B, (B) R1, (C) R2, (D)R3



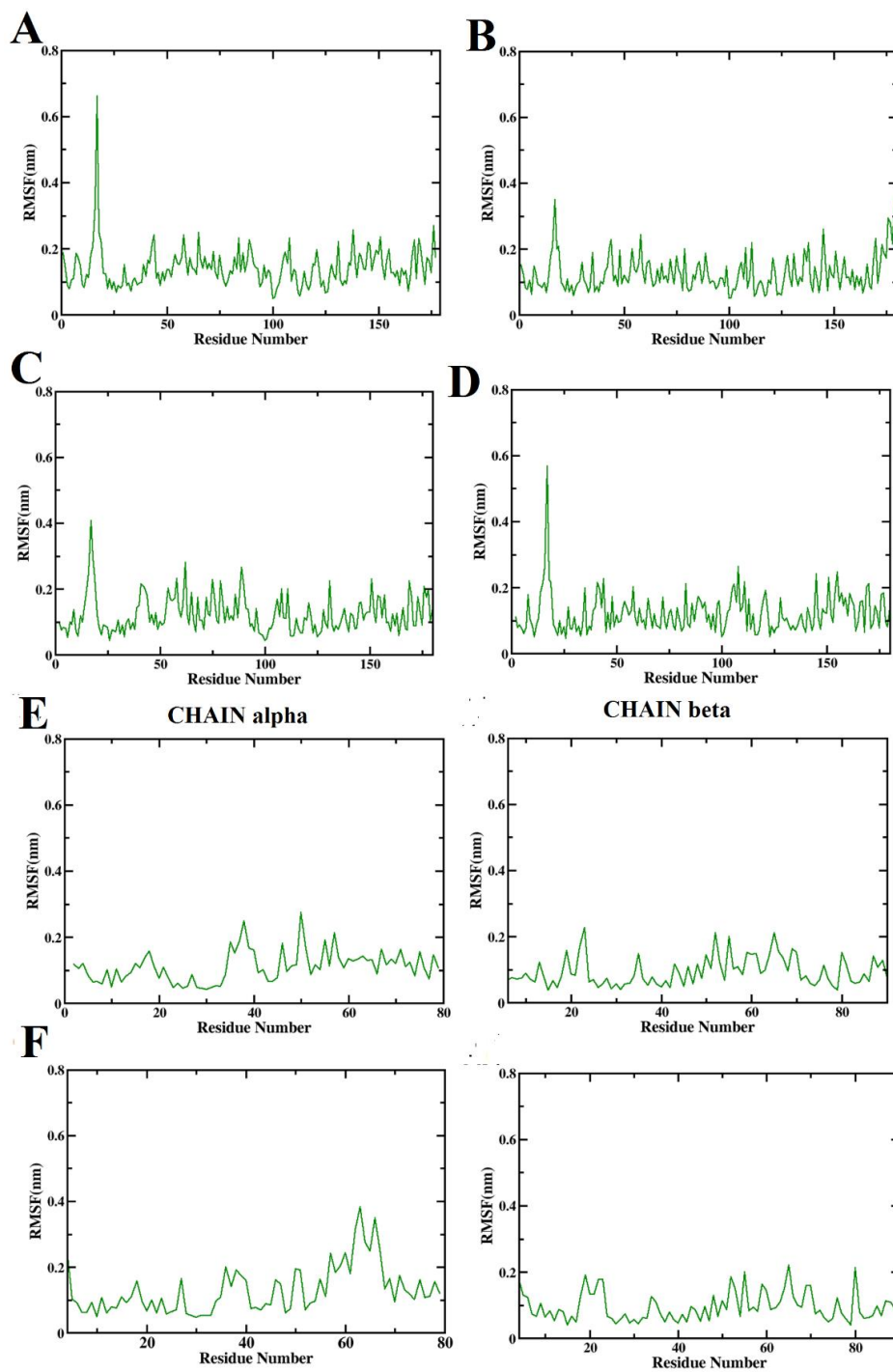
Appendix XXXXIV. The time evolution of the distance between the surface of the peptides and the platform formed by antiparallel β -sheet of peptide presenting groove of (A) HLA-A*02:01, (B) HLA-A*24:01, (C) HLA-B*40:01, (D) HLA-B*58:01 (E) DRB1*04:01, (F) DRB1*07:01



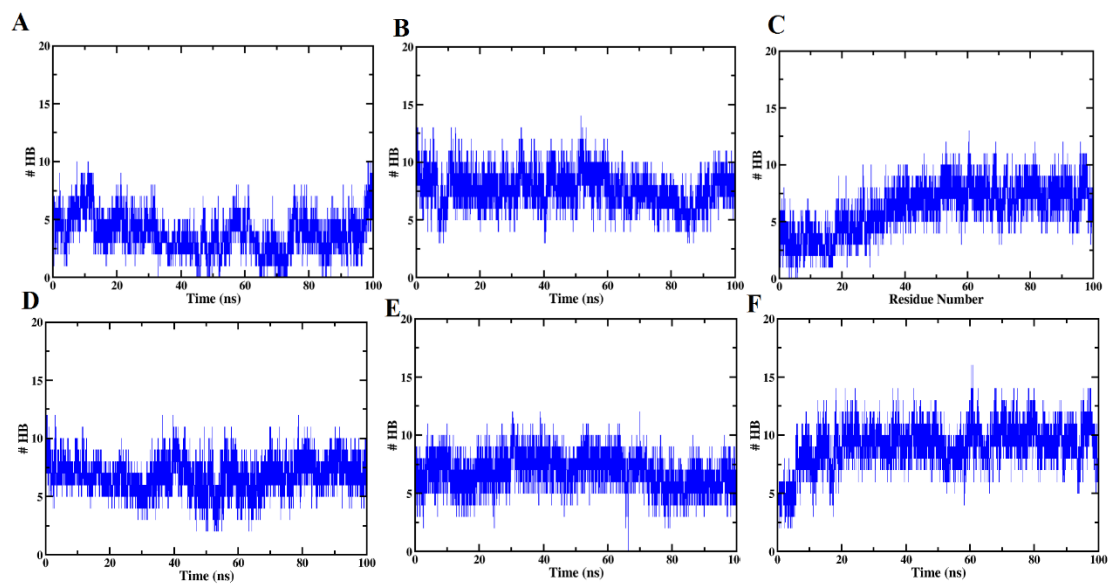
Appendix XXXXV. The time evolution of solvent accessible surface area (SASA) of peptide epitopes associated with (A) HLA-A*02:01, (B) HLA-A*24:01, (C) HLA-B*40:01, (D) HLA-B*58:01 (E) DRB1*04:01, (F) DRB1*07:01.



Appendix XXXXVI. The RMSF profile of the peptide presenting groove of (A) HLA-A*02:01, (B) HLA-A*24:01, (C) HLA-B*40:01, (D) HLA-B*58:01 (E) DRB1*04:01, (F) DRB1*07:01.



Appendix XXXXVII. The time evolution of hydrogen bond between the peptide epitopes and (A) HLA-A*02:01, (B) HLA-A*24:01, (C) HLA-B*40:01, (D) HLA-B*58:01 (E) DRB1*04:01, (F) DRB1*07:01.



Appendix XXXXVIII-A. he occupancy of the hydrogen bonds formed between peptide epitope RDIADTTDAV and HLA*-02:01 during MD.

Pair ID	Donor	Acceptor	Occupancy (%)
1	163THR(HG1)	2 ARG(O)	20.7
2	147TRP(HE1)	9 ASP(OD2)	14.7
3	147TRP(HE1)	9 ASP(OD1)	12.4
4	146LYS(HZ1)	9 ASP(O)	10.1
5	146LYS(HZ1)	9 ASP(OD2)	15.3
6	146LYS(HZ1)	9 ASP(OD1)	12.4
7	143THR(HG1)	9 ASP(OD1)	10
8	99TYR(HH)	3ASP(O)	38
9	73THR(HG1)	5 ALA(O)	41.2
10	66LYS(HZ1)	3 ASP(OD2)	19.1
11	8THR(HG1)	77 ASP(OD1)	28.5
12	2ARG(H21)	166GLU(OE2)	17.2
13	2ARG(H21)	166GLU(OE1)	11.9
14	2ARG(H11)	166GLU(OE1)	13

Appendix XXXXVIII -B. he occupancy of the hydrogen bonds formed between peptide epitope TKRFDNPVLPF and HLA*-24:02 during MD.

Pair ID	Donor	Acceptor	Occupancy (%)
1	147TRP(HE1)	10LEU(O)	91.6
2	146LYS(HZ1)	11PRO(O)	30.1
3	146LYS(HZ1)	10LEU(O)	28.1
4	84TYR(HH)	12PHE(OC1)	91
5	83ARG(H21)	12PHE(OC2)	46.1
6	83ARG(H21)	12PHE(OC1)	78
7	83ARG(H11)	12PHE(OC2)	58.8
8	83ARG(H11)	12PHE(OC1)	44.1
9	77ASN(D21)	9VAL(O)	15.2
10	12PHE(H)	143THR(OG1)	87.6
11	7ASN(D21)	69ALA(O)	12.6
12	3LYS(HZ1)	62GLU(OE2)	26.1
13	3LYS(HZ1)	62GLU(OE1)	26

14	3LYS(H)	62GLU(OE2)	11.8
----	----------	------------	------

Appendix XXXXVIII -C. The occupancy of the hydrogen bonds formed between peptide epitope VRFPNITNL and HLA-B*40:01 during MD.

Pair ID	Donor	Acceptor	Occupancy (%)
1	159TYR(HH)	5PRO(O)	80.8
2	146LYS(HZ1)	10 LEU(OC2)	18.9
3	146LYS(HZ1)	10 LEU(OC1)	32.2
4	116TYR(HH)	9ASN(OD1)	18.8
5	97ARG(H21)	7 ILE(O)	28.3
6	97ARG(H11)	7 ILE(O)	51.7
7	77SER(HG)	10LEU(OC2)	43.1
8	77SER(HG)	10LEU(OC2)	29.6
9	73THR(HG1)	9 ASN(OD1)	20.9
10	59TYR(HH)	2VAL(O)	25.6
11	8THR(H)	155GLN(OE1)	20.9
12	3ARG(H21)	63 GLU(OE2)	40.1
13	3ARG(H21)	63 GLU(OE1)	54.2
14	3ARG(HE)	63 GLU(OE1)	56.3
15	3ARG(HE)	63 GLU(OE1)	23.3

Appendix XXXXVIII -D. The occupancy of the hydrogen bonds formed between peptide epitope VFAQVKQIY and HLA-B*58:01 during MD.

Pair ID	Donor	Acceptor	Occupancy (%)
1	147TRP(HE1)	9ILE(O)	84.8
2	146LYS(HZ1)	10TYR(OC2)	59.9
3	146LYS(HZ1)	10TYR(OC2)	56
4	143THR(HG1)	10TYR(OC2)	83
5	97ARG(H21)	5GLN(OE1)	13.7
6	84TYR(HH)	10TYR(OC2)	89.9
7	77ASN(D21)	8GLN(O)	87.8
8	66ASN(D21)	5GLN(OE1)	11.6

9	10TYR(H)	116SER(OG)	26
10	8GLN(E21)	116SER(OG)	14.7
11	8GLN(E21)	114ASP(OD2)	11
12	8GLN(E21)	114ASP(OD1)	12
13	8GLN(E21)	77ASN(OD1)	23.2
14	5GLN(E21)	9TYR(OH)	11.8

Appendix XXXXVIII -E. The occupancy of the hydrogen bonds formed between peptide epitope NTLVKQLSSNFGA and DRB1*04:01 during MD.

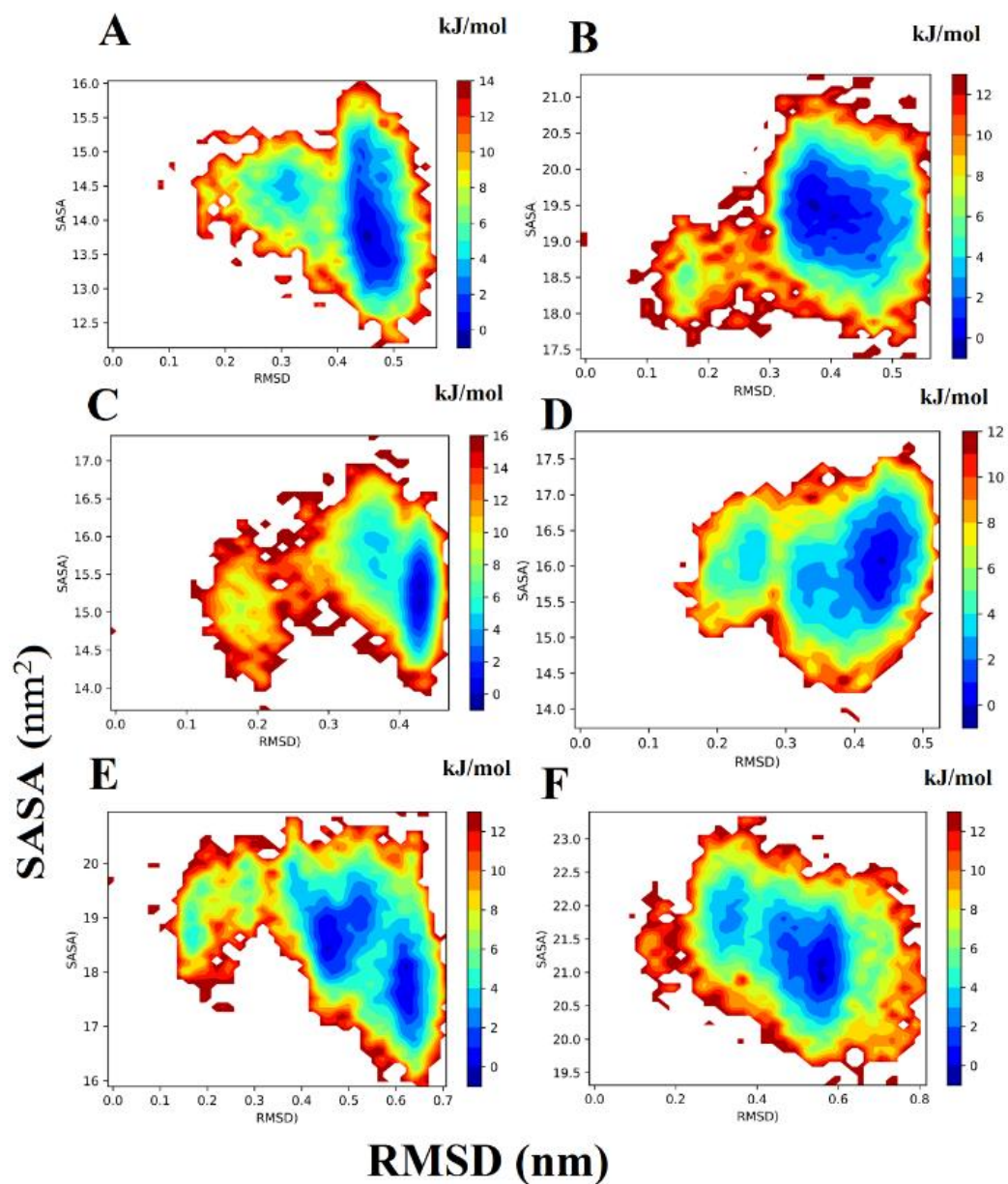
Pair ID	Donor	Acceptor	Occupancy (%)
1	82ASN(D21)	6 LYS(O)	95.8
2	52ALA(H)	2ASN(O)	86.3
3	51PHE(H)	2ASN(O)	32.8
4	9GLN(E21)	7 GLN(OE1)	43.5
5	7GLN(E21)	55 GLU(O)	13.3
6	7GLN(E21)	9 GLN(OE1)	52.3
7	6LYS(HZ1)	77 THR(O)	27.8
8	6LYS(H)	82ASN(OD1)	96
9	3THR(HG1)	48 PHE(O)	15
10	2ASN(H1)	52ALA(O)	87.7
11	2ASN(H1)	49GLY(O)	67.7

Appendix XXXXVIII -F. The occupancy of the hydrogen bonds formed between peptide epitope SLLIVNNATNVVIK and DRB1*07:01 during MD.

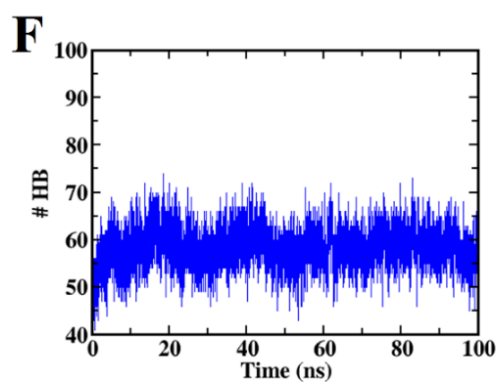
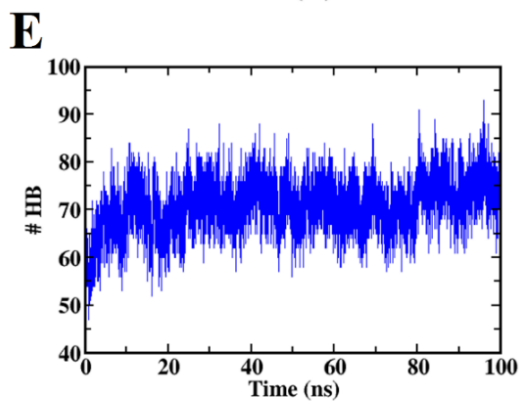
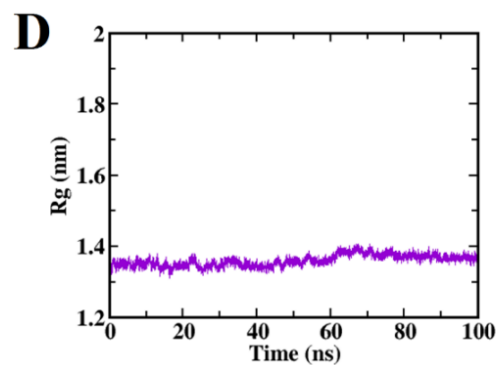
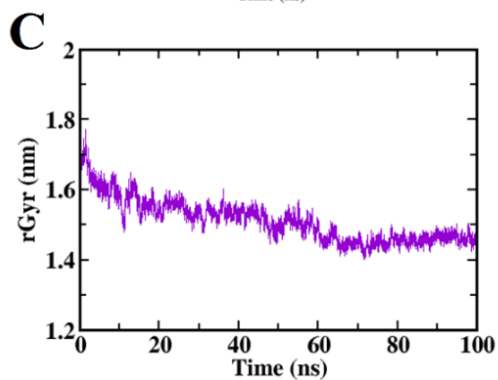
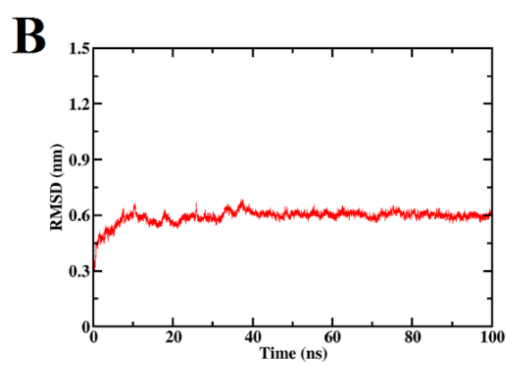
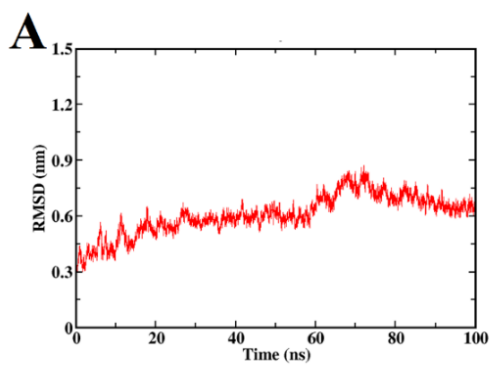
Pair ID	Donor	Acceptor	Occupancy (%)
1	82ASN(D21)	4 LEU(O)	94.1
2	81HIS(HE2)	2 SER(O)	82.6
3	71ARG(H21)	8 ASN(OD1)	19.6
4	71ARG(H21)	7 ASN(O)	57.4
5	71ARG(HE)	7 ASN(O)	10.1
6	62ASN(D21)	8 ASN(O)	19.9
7	62ASN(D21)	7 ASN(OD1)	12.1
8	62ASN(D21)	6 VAL(O)	17.5

9	10THR(HG1)	70 GLN(OE1)	39.6
10	10THR(H)	70 GLN(OE1)	37.1
11	8ASN(D21)	62ASN(O)	30
12	8ASN(H)	62 ASN(OD1)	37
13	7ASN(D21)	62 ASN(OD1)	14.1
14	4LEU(H)	82 ASN(OD1)	98.5
15	3LEU(H)	53SER(O)	98
16	2SER(HG)	55GLU(OE2)	36.3
17	2SER(HG)	55GLU(OE1)	30.2

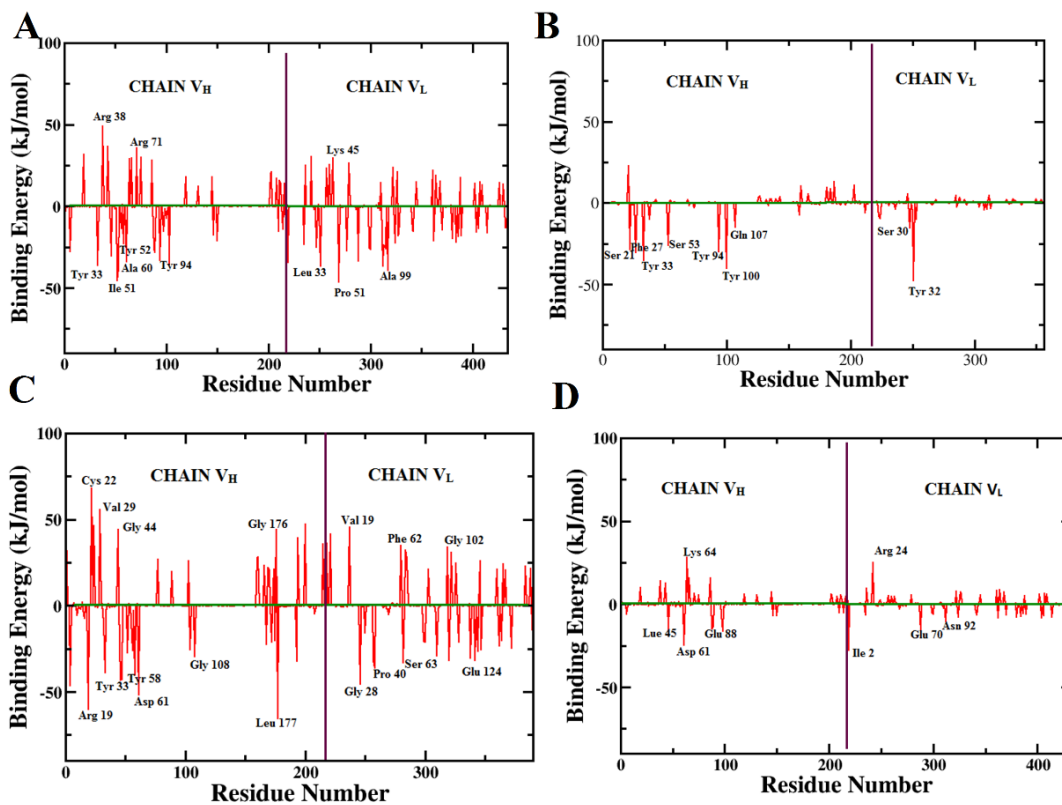
Appendix XXXIX. The free energy landscape of the peptide epitopes associated with (A) HLA-A*02:01, (B) HLA-A*24:01, (C) HLA-B*40:01, (D) HLA-B*58:01 (E) DRB1*04:01, (F) DRB1*07:01



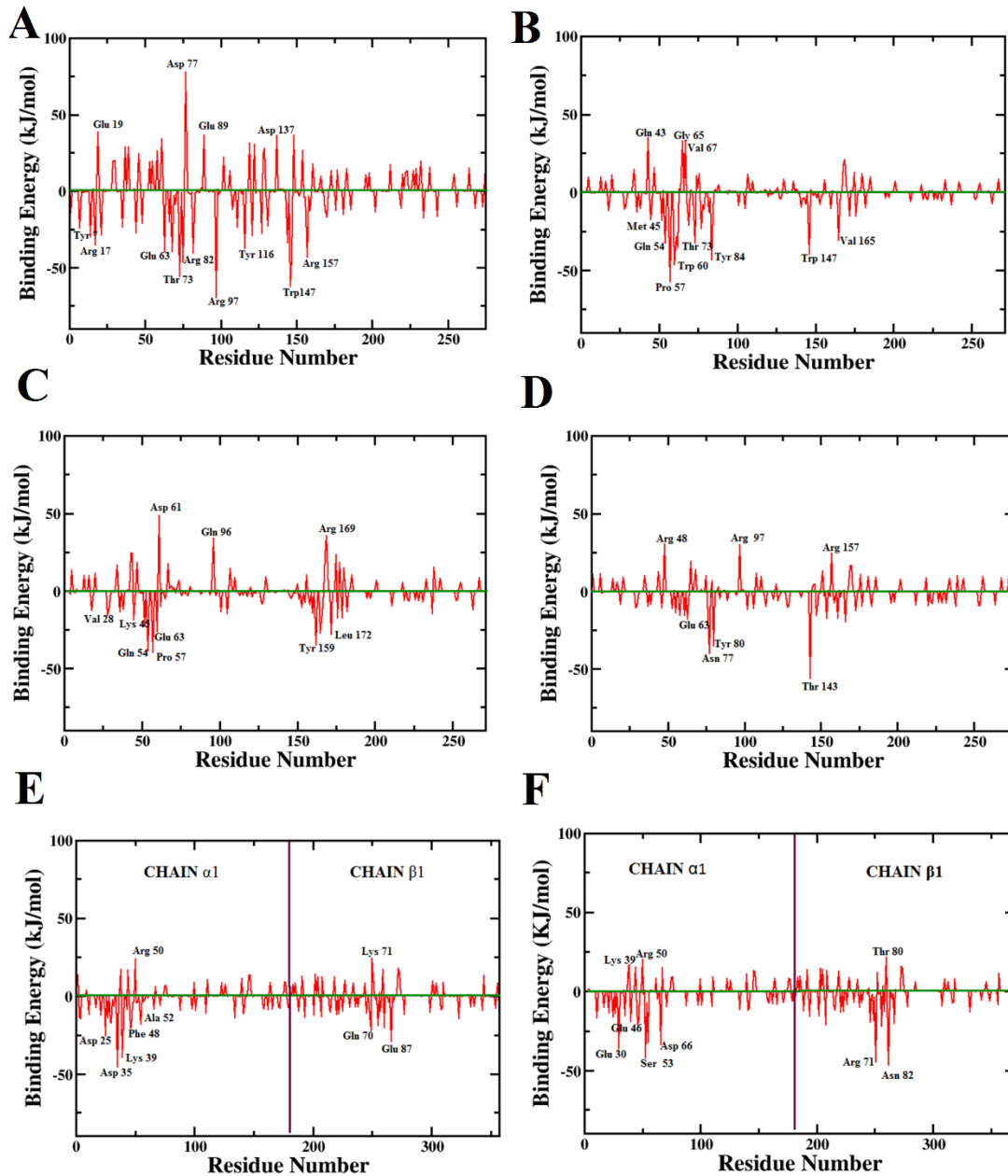
Appendix XXXXX. The RMSD profile of (A) Vac-COVID-B, (B) Vac-COVID-T. The radius of gyration profile of (C) Vac-COVID-B, (D) Vac-COVID-T. The time evolution of native hydrogen bond for (E) Vac-COVID-B, (F) Vac-COVID-T.



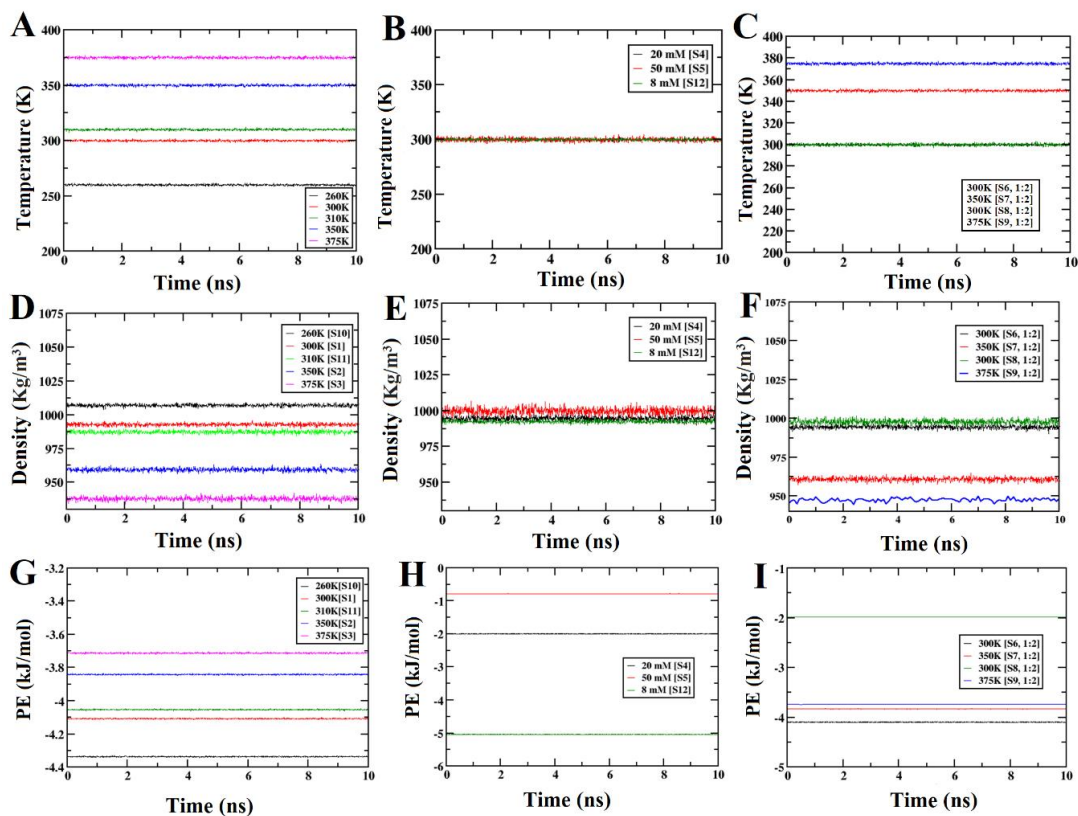
Appendix XXXXXI. The residue wise decomposition of free energy between 7BZ5 and (A) Vac-COVID-B, (B) R1, (C) R2, (D) R3.



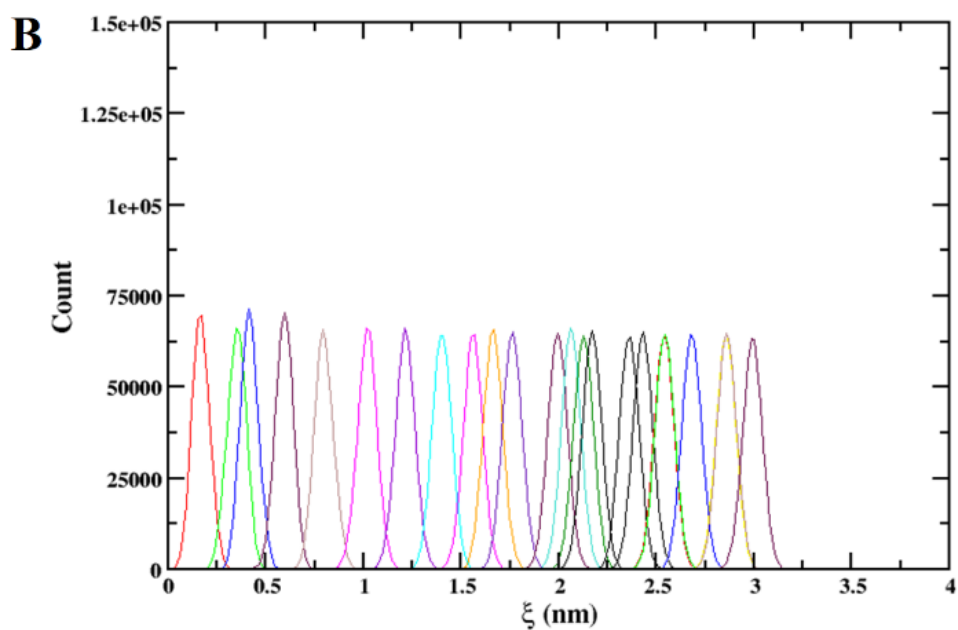
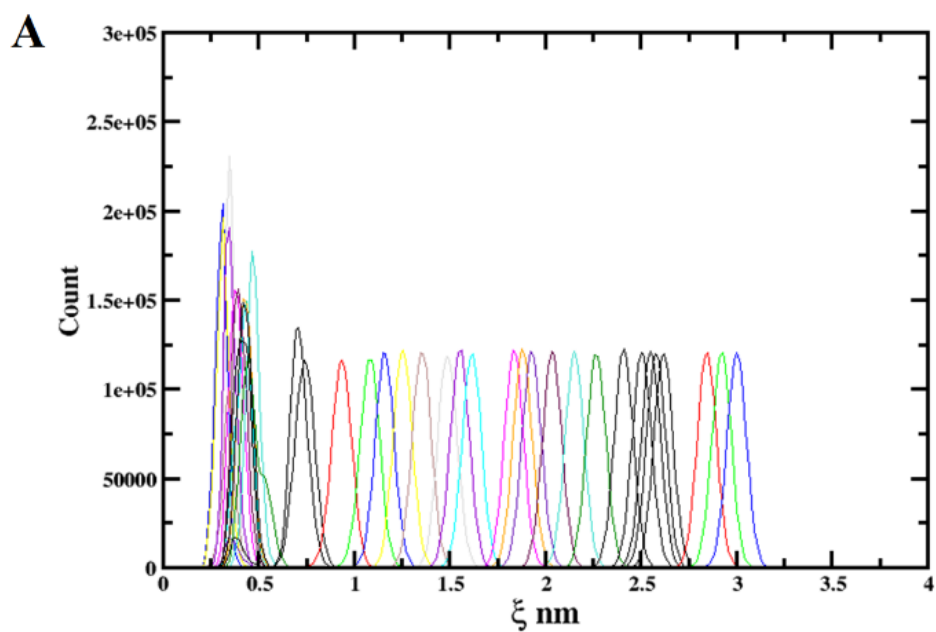
Appendix XXXXXII. The residue wise decomposition of free energy between the peptide epitope and (A) HLA-A*02:01, (B) HLA-A*24:01, (C) HLA-B*40:01, (D) HLA-B*58:01 (E) DRB1*04:01, (F) DRB1*07:01



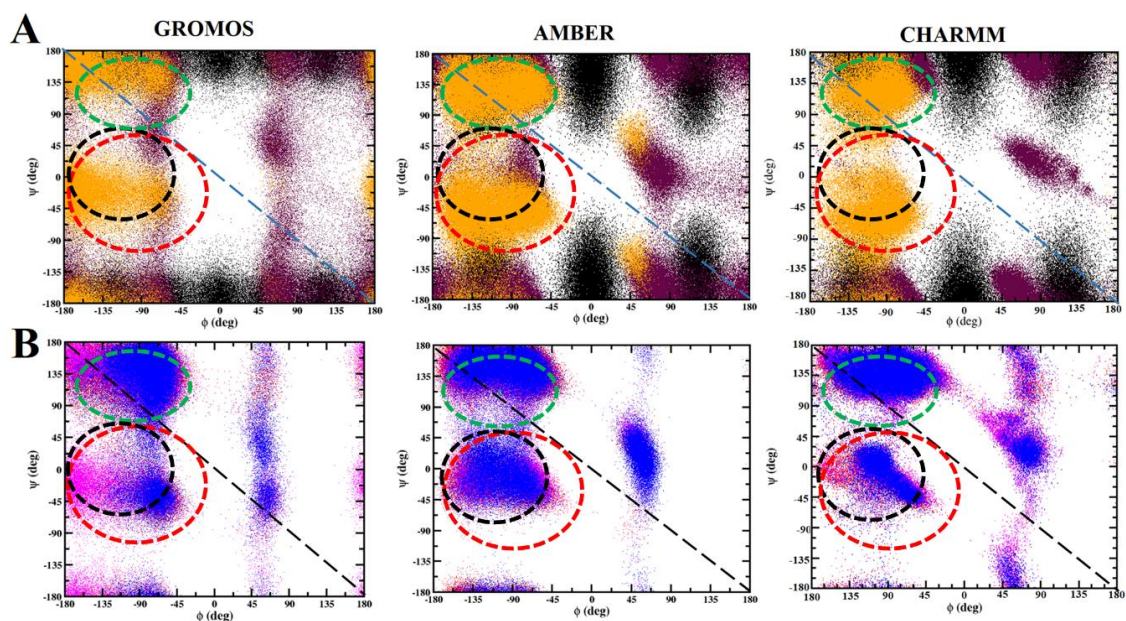
Appendix XXXXXIII. The graph shows the thermodynamic equilibrium of the systems. The graphs in (A, B, C) show the time evolution of Temperature, (D,E,F) show the time evolution of density, (GHI) shows the time evolution of the potential energy of the system.



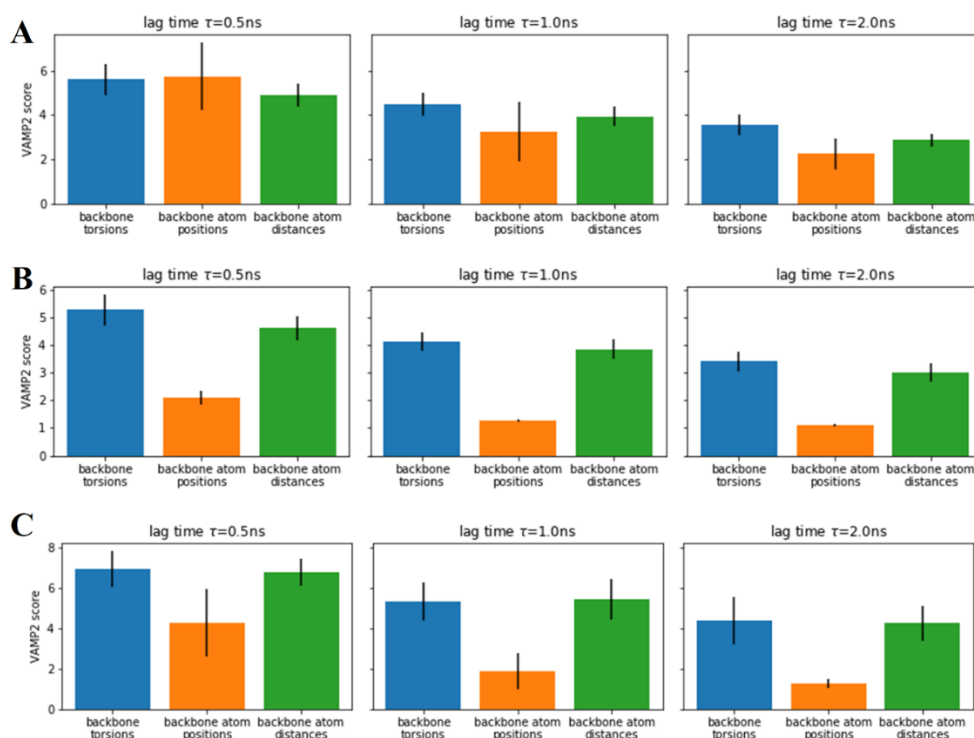
Appendix XXXXXIV. The histograms of each umbrella sampling window to generate the PMF for dimer dissociation at the (A) absence and (B) presence of NQDA.



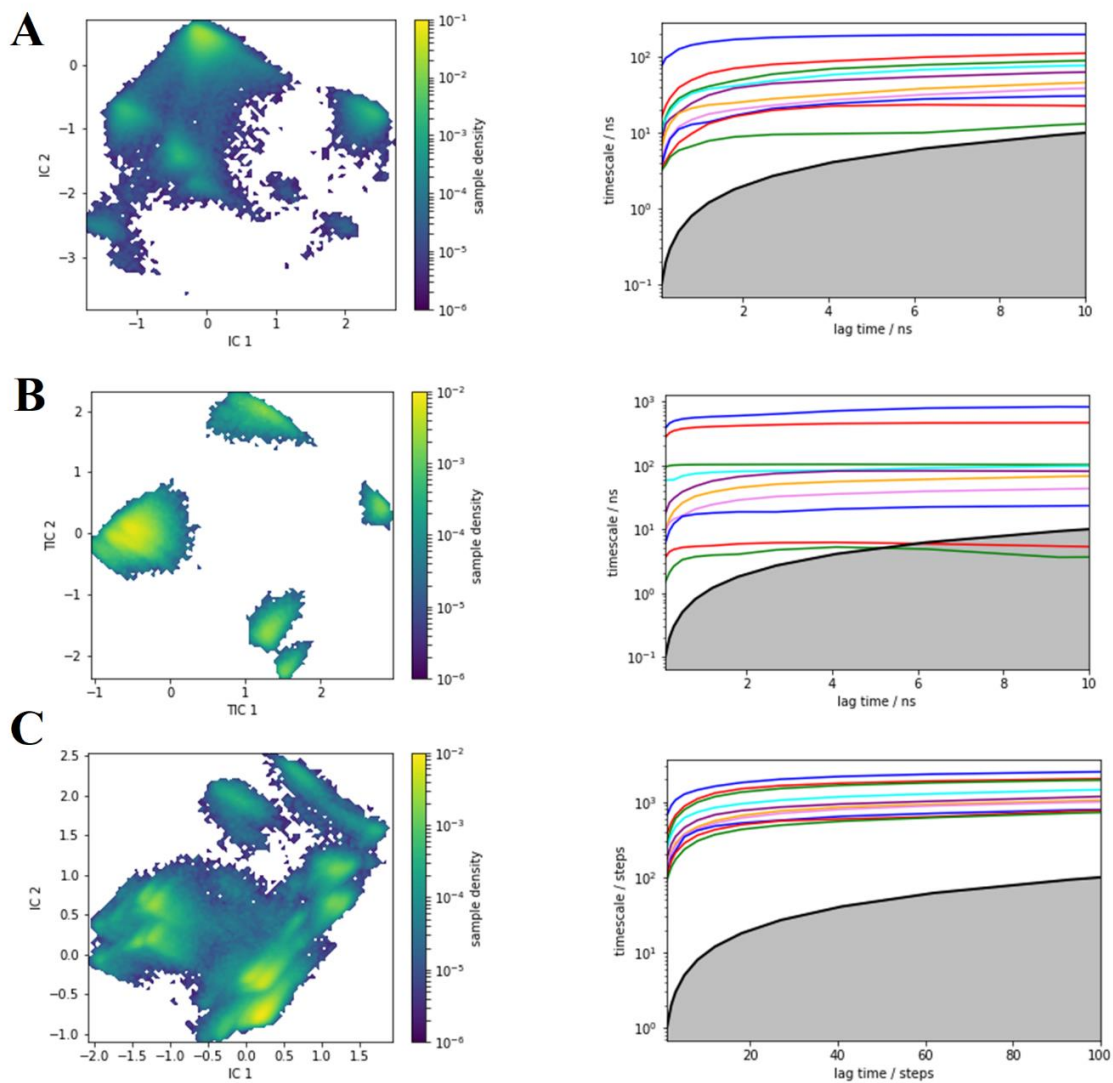
Appendix XXXXXV. The 2D distribution of torsional angles (ϕ/ψ) for (A) terminal residues (Gly127 [black], Gly 131 [maroon], Ser132 [orange]) and (B) the middle residues (Tyr128 [red], Leu129 [pink], Met130[blue]) of human prion peptide. The region for helix, sheet, and random coils are colored in red, green, and black respectively.



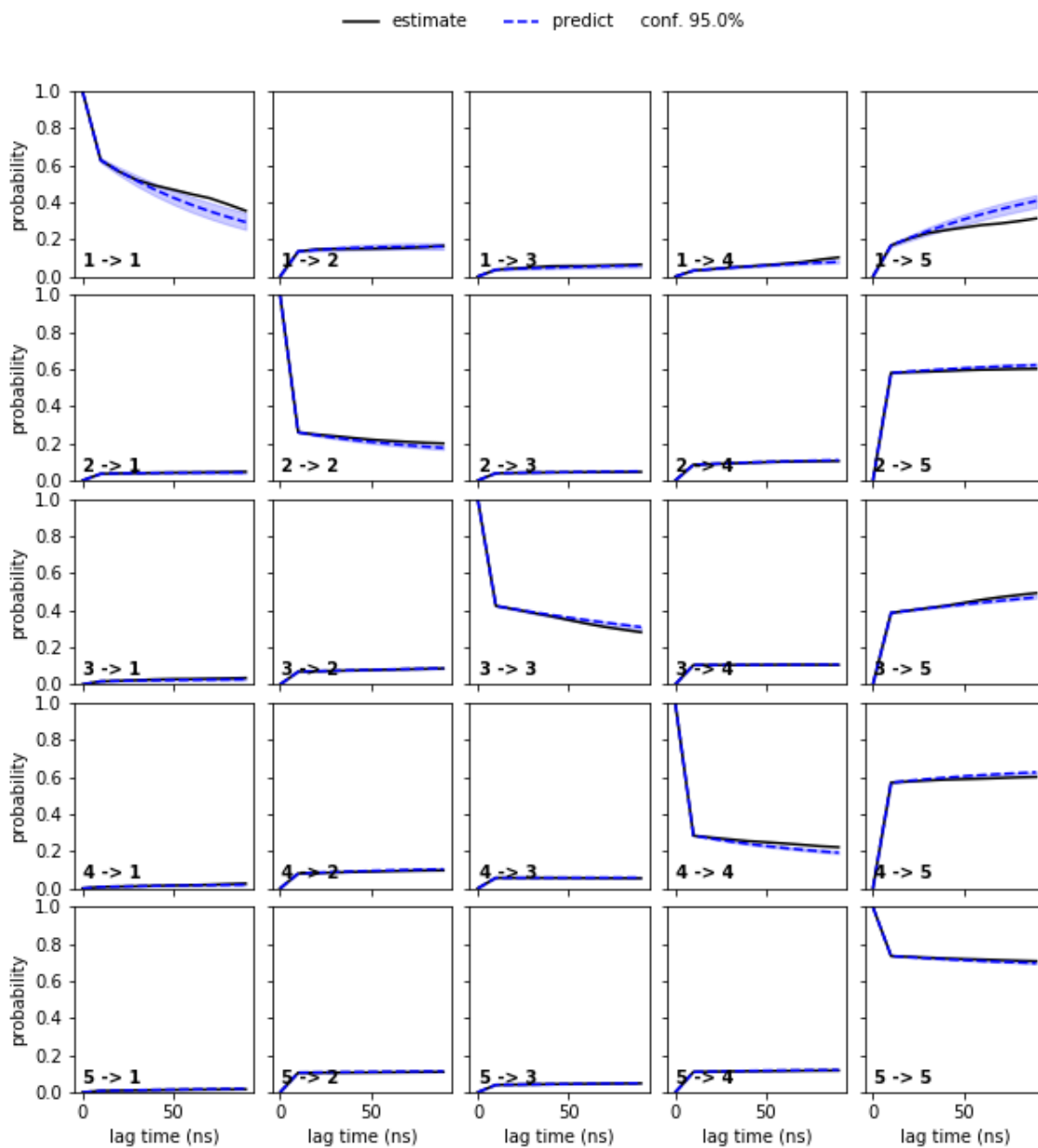
Appendix XXXXXVI. The VAMP-2 scores for three different molecular descriptors such as backbone torsion angles, atom position, and atom distances for human prion peptide at the different lag times (τ) for (A) GROMOS-54a7, (B) AMBER99SB-ILDN and (C) CHARMM-36m force-fields. The standard deviation is calculated to estimate the errors and is represented in a straight line.



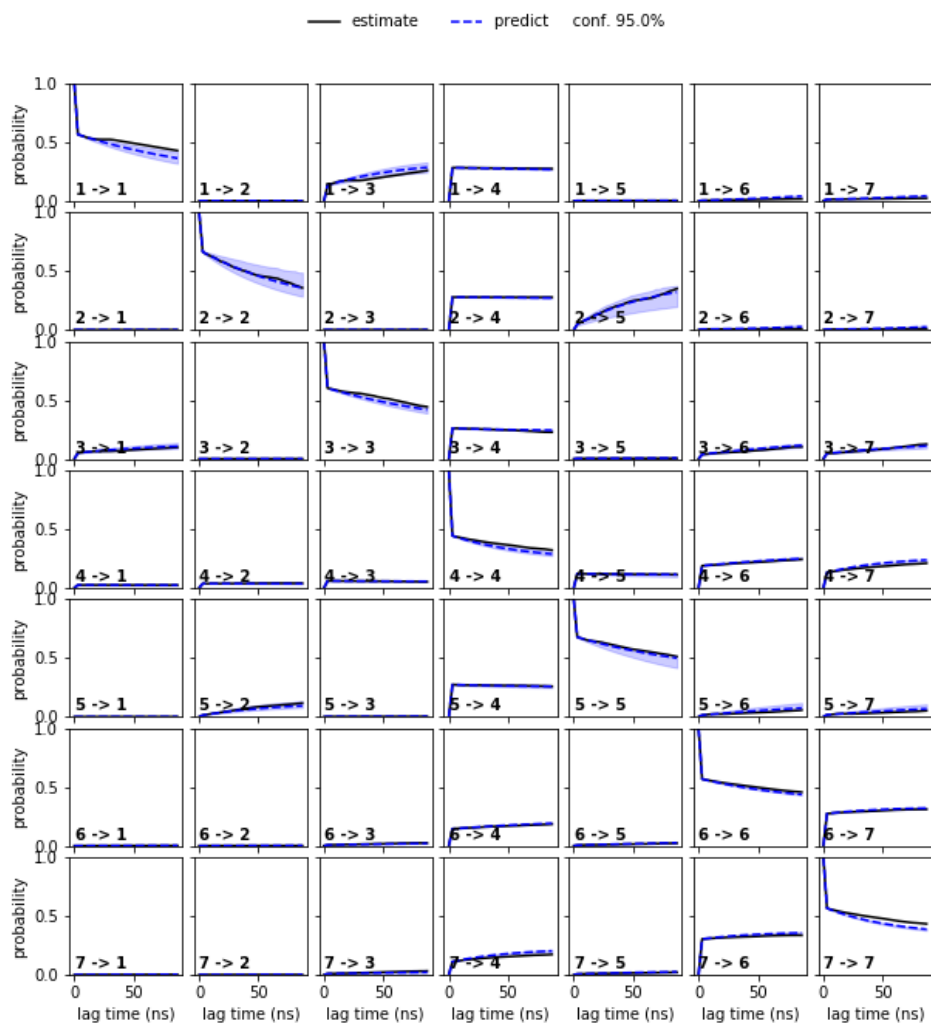
Appendix XXXXXVII. The sample densities (left) of the cumulative trajectories of 1 μ s correspond to the first two TICA components of prion peptide and the implied time scale plots (right) for different MSMs at different lag times. The slow and faster dynamics in the ITs plot are distinguished by the black line.



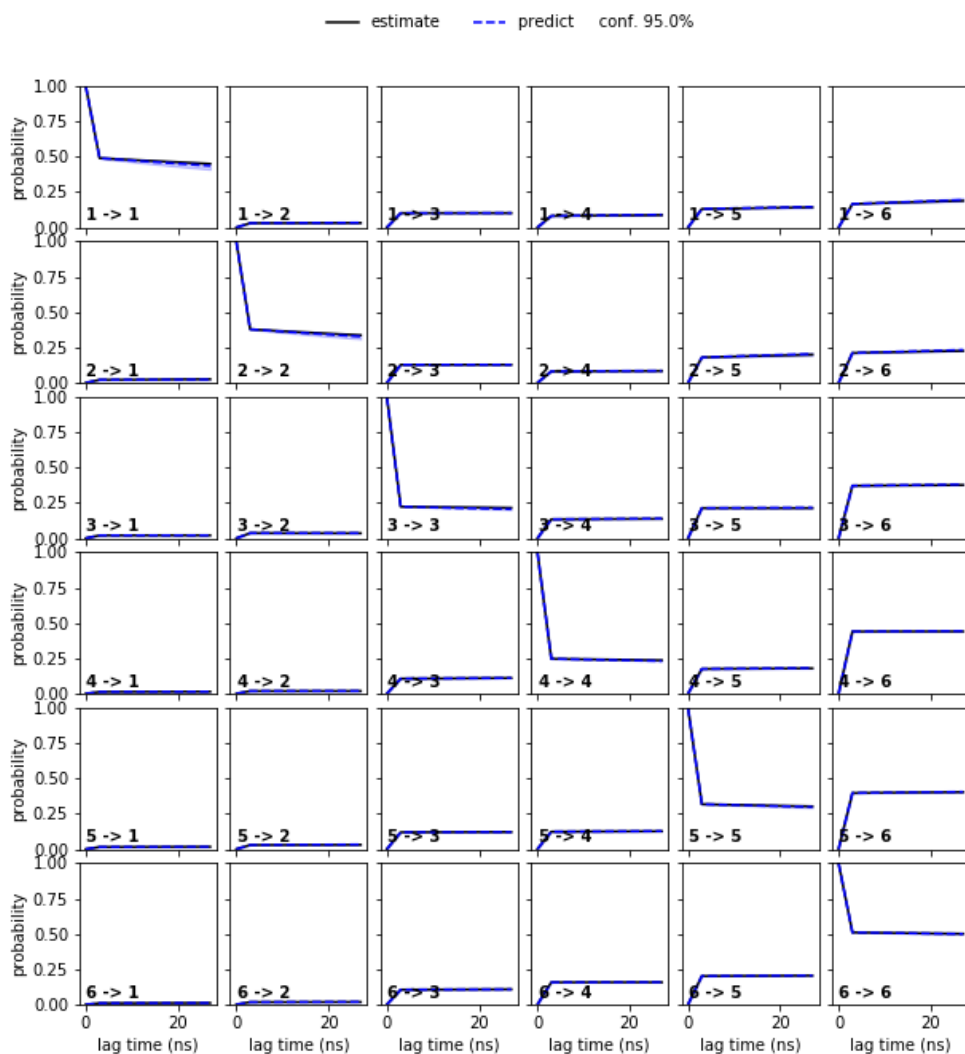
Appendix XXXXXVIII. The Chapman-Kolmogov validation for MSM is built for the prion peptide monomer using the lag time of 2 ns and five microstates for the GROMOS-54a7 force field.



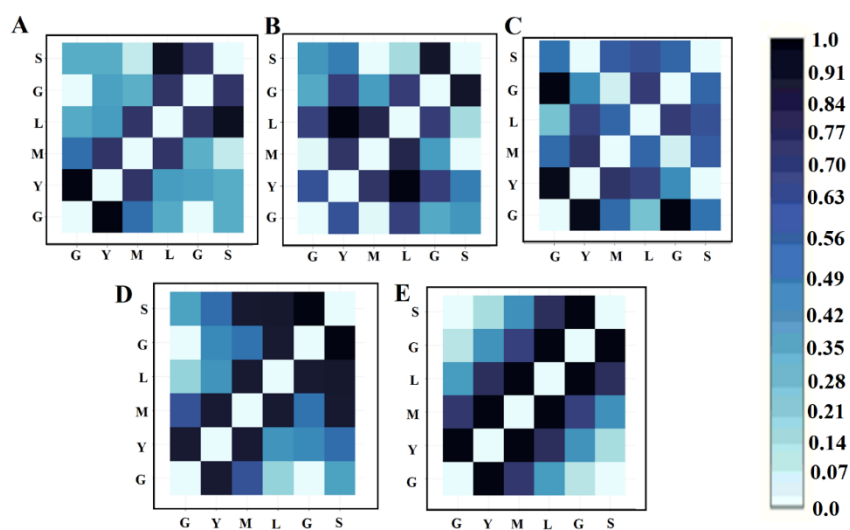
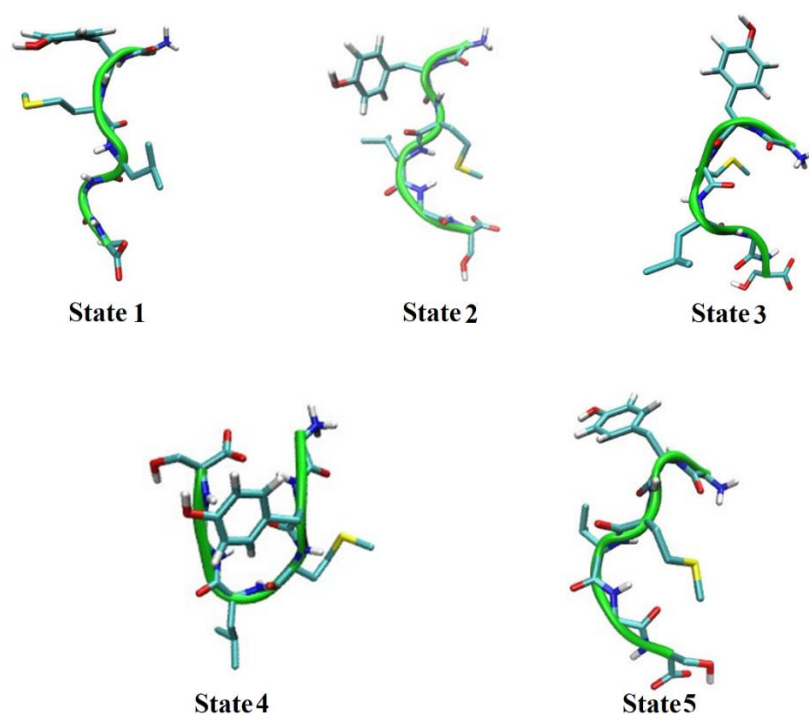
Appendix XXXXXIX. The Chapman-Kolmogov validation for MSM is built for the prion peptide monomer using the lag time of 2 ns and five microstates for the AMBER99SB-ILDN force field.



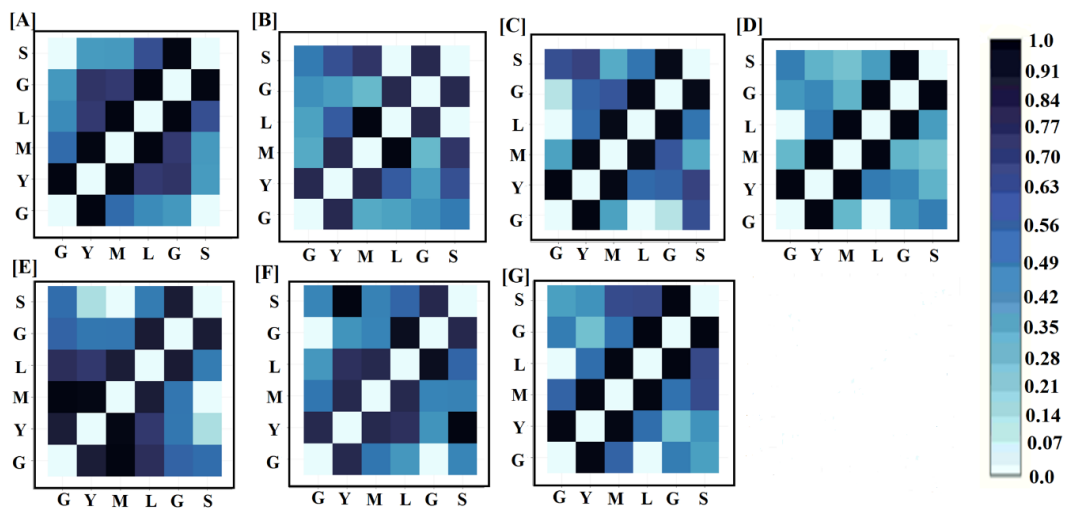
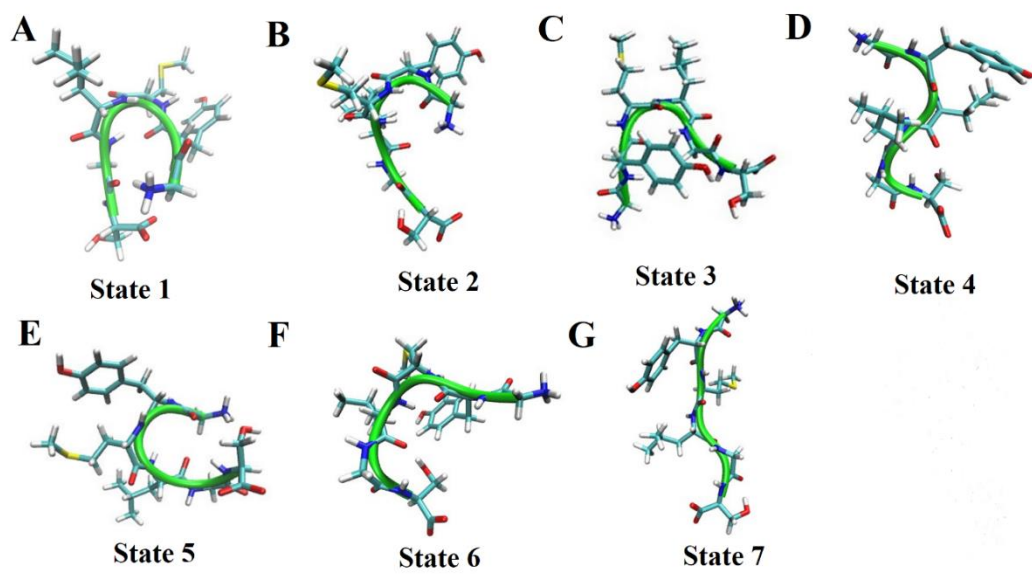
Appendix XXXXXX. The Chapman-Kolmogov validation for MSM is built for the prion peptide monomer using the lag time of 2 ns and five microstates for the CHARMM-36m force field.



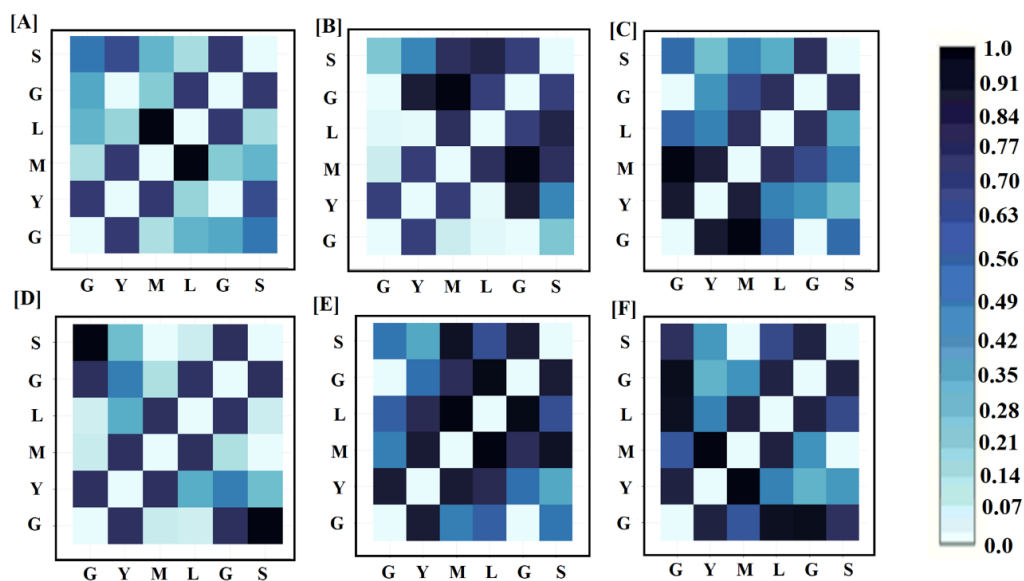
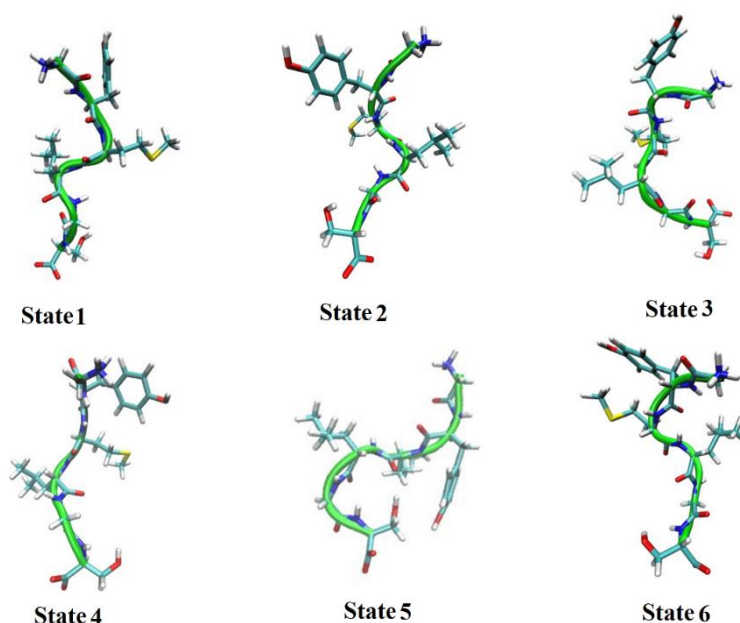
Appendix XXXXXXI. The representative peptide structure corresponds to every five Markov states in the GROMOS-54a7 force field (top). The carbon, nitrogen, oxygen, sulfur, and hydrogen atoms are colored in cyan, blue, red, yellow, and white. The 2D map for intrapeptide contacts of human prion peptide corresponds to each Markov state (bottom). The extent of the contract or the contact probability between the residues is represented by color code. The diagonal line of the contact map has light color to represent self-contact.



Appendix XXXXXXII. The representative peptide structure corresponds to every five Markov states in the AMBER-99sb-ILDN force field (top). The carbon, nitrogen, oxygen, sulfur, and hydrogen atoms are colored in cyan, blue, red, yellow, and white. The 2D map for intrapeptide contacts of human prion peptide corresponds to each Markov state (bottom). The extent of the contract or the contact probability between the residues is represented by color code. The diagonal line of the contact map has light color to represent self-contact.



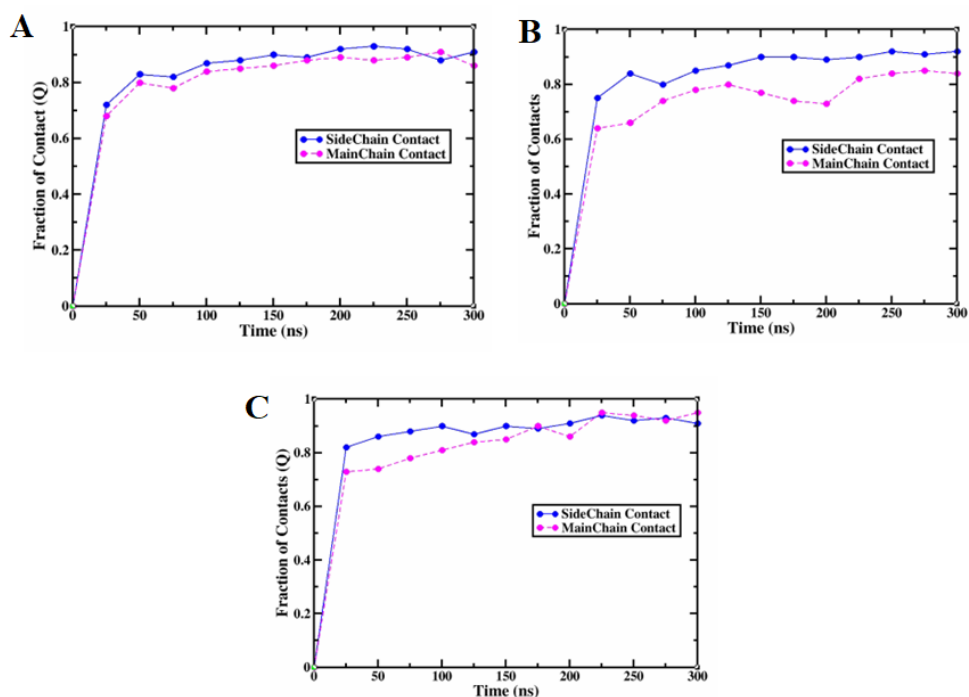
Appendix XXXXXXIII. The representative peptide structure corresponds to every five Markov states in the CHARMM-36m force field (top). The carbon, nitrogen, oxygen, sulfur, and hydrogen atoms are colored in cyan, blue, red, yellow, and white. The 2D map for intrapeptide contacts of human prion peptide corresponds to each Markov state (bottom). The extent of the contract or the contact probability between the residues is represented by color code. The diagonal line of the contact map has light color to represent self-contact.



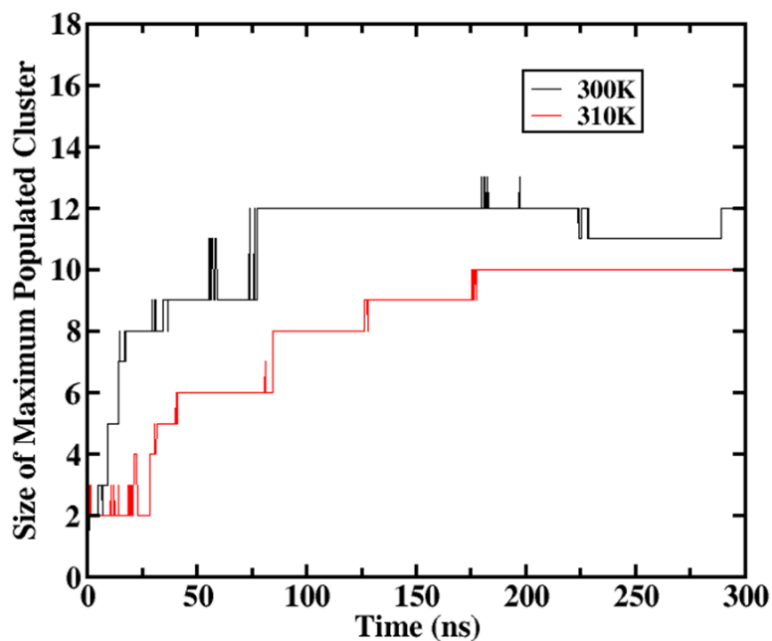
Appendix XXXXXXIV. The average Diffusion Coefficient (D) of prion peptide monomers in three different force fields with corresponding errors from five independent replicas.

Force-Field	D (cm ² /s) × 10 ⁵
GROMOS-54a7	0.1326±0.04
AMBER99SB-ILDN	0.2695±0.09
CHARMM-36m	0.4022±0.02

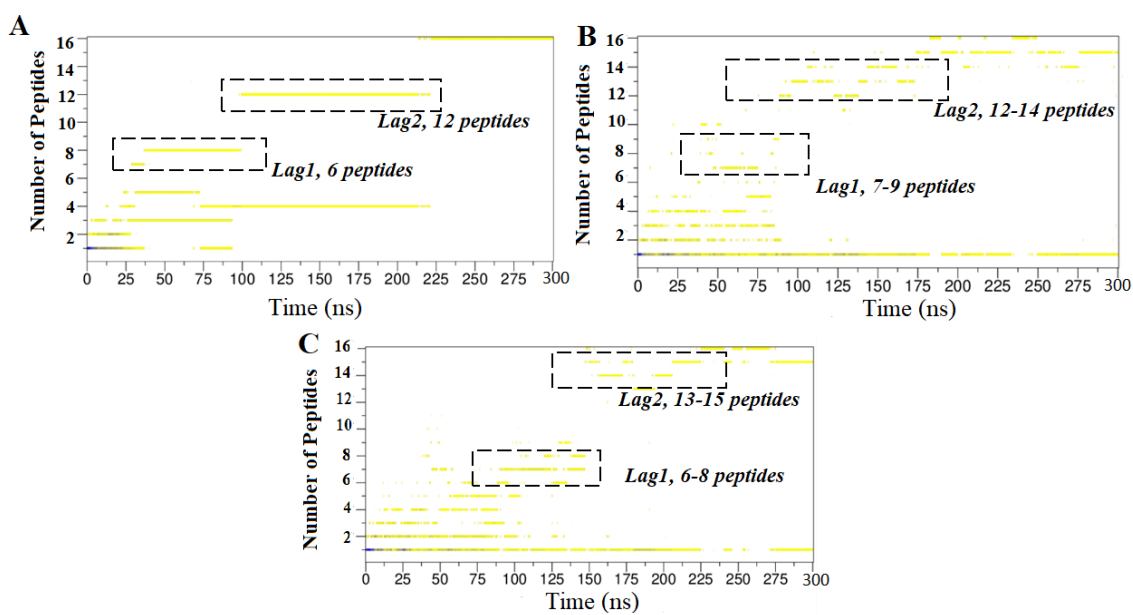
Appendix XXXXXXV. Time evolution of inter-peptide side-chain and main-chain or backbone contact at 20 mM peptide concentration.



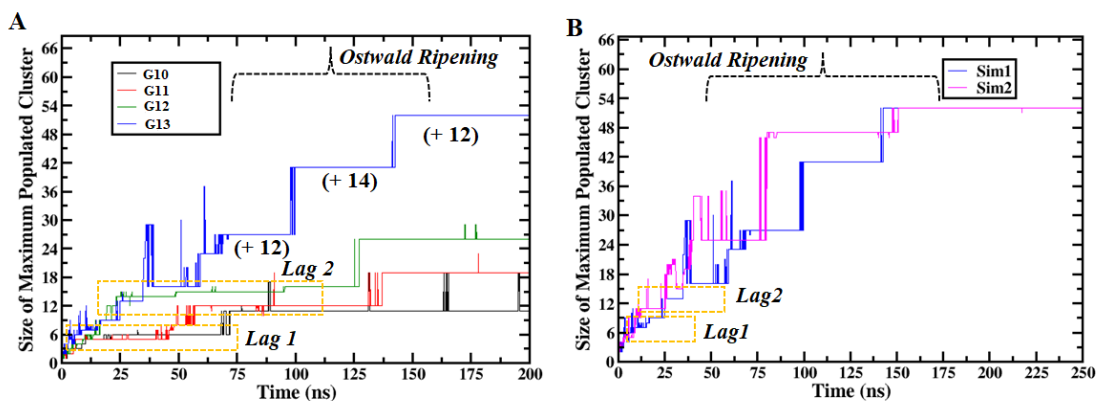
Appendix XXXXXXVI. Time evolution of maximum populated cluster at (A) 300K (black), (B) 310K (red) in GROMOS-54a7 force field.



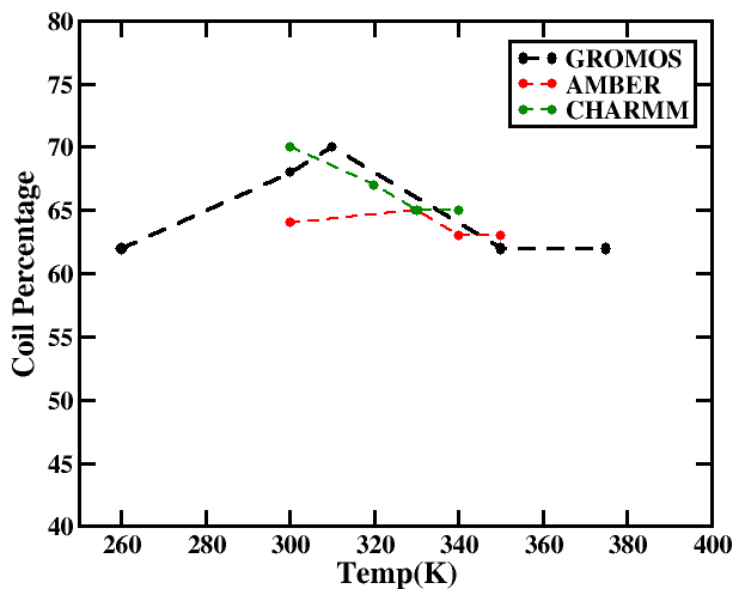
Appendix XXXXXXVII. The time evolution of peptide clusters in (A) GROMOS 54a7 [20 mM, 300K], (B) AMBER 99SB ILDN [30 mM, 300K] and (C) CHARMM 36m [50 mM, 300K].



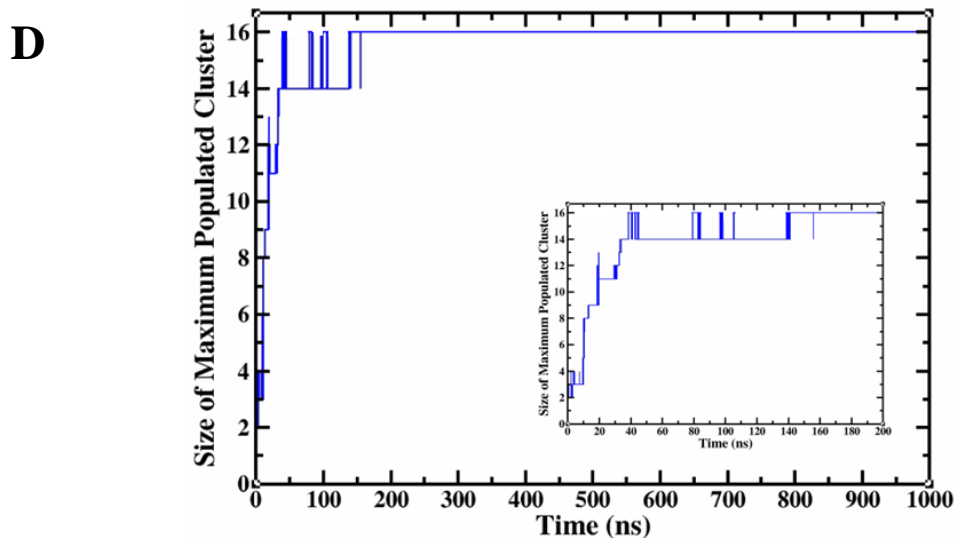
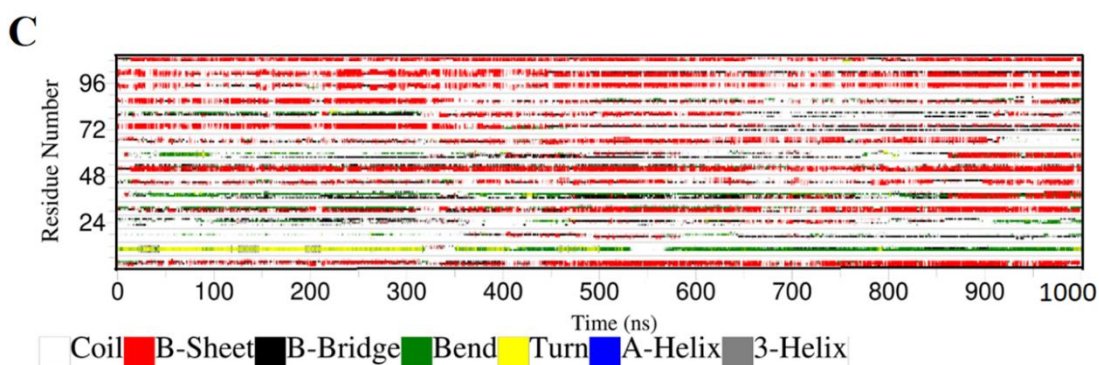
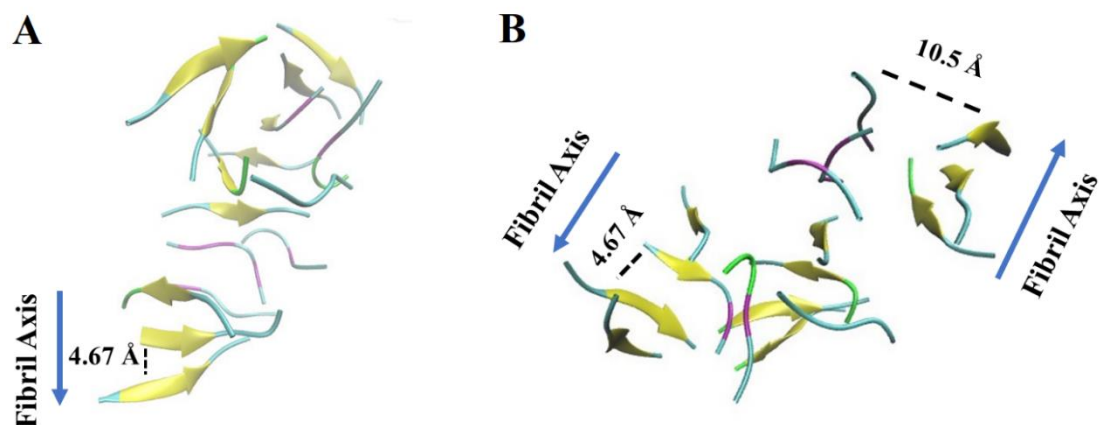
Appendix XXXXXXVIII. The cluster size distribution of peptide self-assembly simulation containing (A) more than 16 prion peptide monomers. (B) the second simulation replica for G13 system.



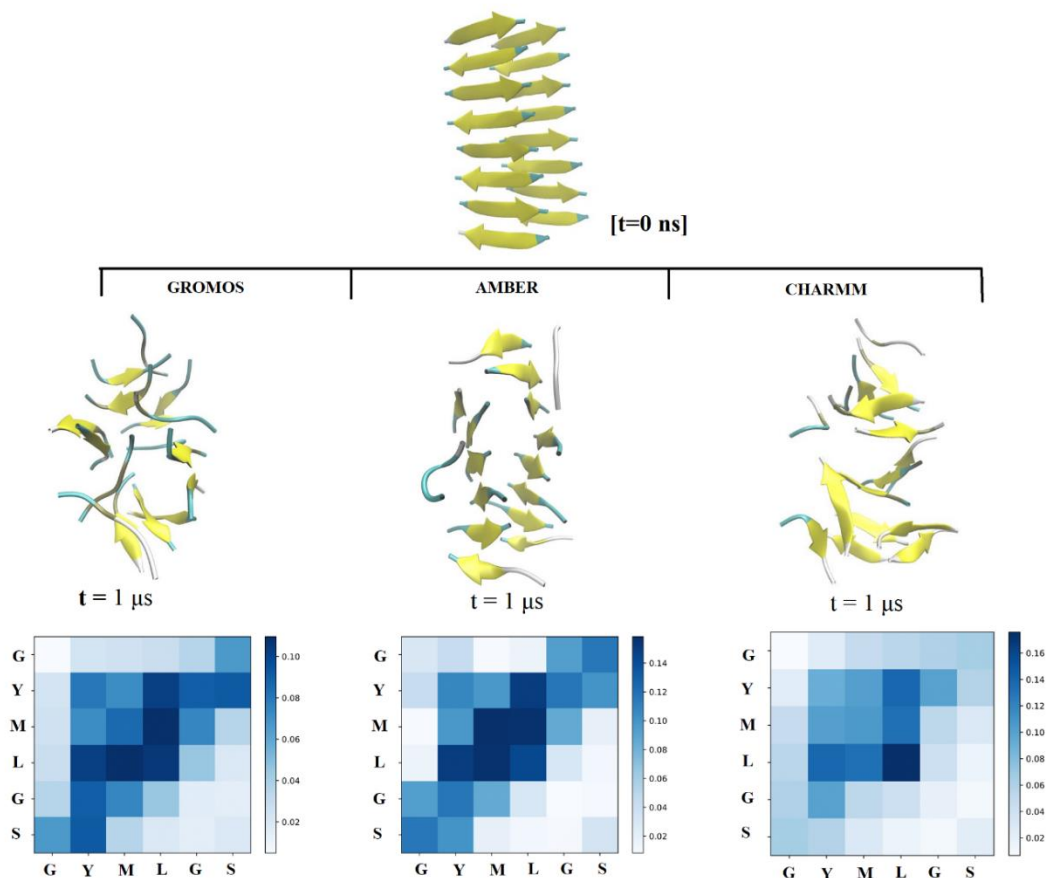
Appendix XXXXXXIX. The coil percentage in three different force fields during self-assembly of prion peptide.



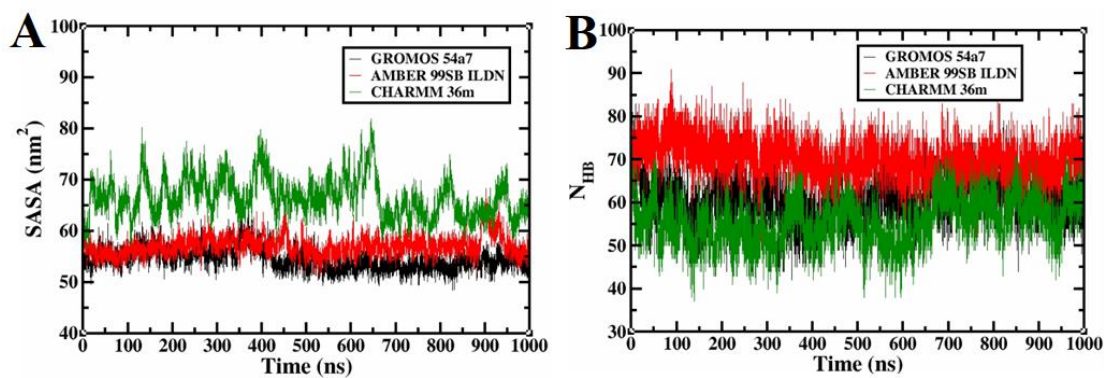
Appendix XXXXXXXX. (A & B) The steric zipper (C) secondary structure profile of the prion peptide aggregate [50 mM, 300] after 1 μ s simulation in GROMOS-54a7 force-field. (D) The Cluster Size distribution of maximum populated cluster size at 50 mM peptide concentration.



Appendix XXXXXXXXI. The secondary structure of the prion peptide mini-fibril after 1 μ s simulation.



Appendix XXXXXXXXII. The time evolution of (A) solvent accessible surface area and (B) inter peptide hydrogen bond of mini-fibril in three force-fields.



Appendix XXXXXXXXIII. Inter-peptide Hydrogen bonds that have more than 70% Occupancy during the minifibrill simulations in three force fields.

GROMOS 54a7

Pair ID	donor-acceptor	Atom Number	Occupancy (%)
1	130LEU(H) - 129MET(O)	777 - 718	78.7
2	130LEU(H) - 129MET(O)	606 - 547	70.3
3	130LEU(H) - 129MET(O)	549 - 490	90.1
4	129MET(H) - 130LEU(O)	540 - 613	86.9
5	130LEU(H) - 129MET(O)	378 - 319	88.9
6	131GLY(H) - 128TYR(O)	330 - 367	90.0
7	130LEU(H) - 129MET(O)	321 - 262	77.2
8	129MET(H) - 130LEU(O)	255 - 328	82.9
9	130LEU(H) - 129MET(O)	207 - 148	86.0
10	131GLY(H) - 128TYR(O)	159 - 196	70.9
11	130LEU(H) - 129MET(O)	150 - 91	95.9
12	129MET(H) - 130LEU(O)	141 - 214	76.3
13	131GLY(H) - 128TYR(O)	102 - 139	74.2
14	130LEU(H) - 129MET(O)	93 - 34	83.0
15	131GLY(H) - 128TYR(O)	45 - 82	72.7

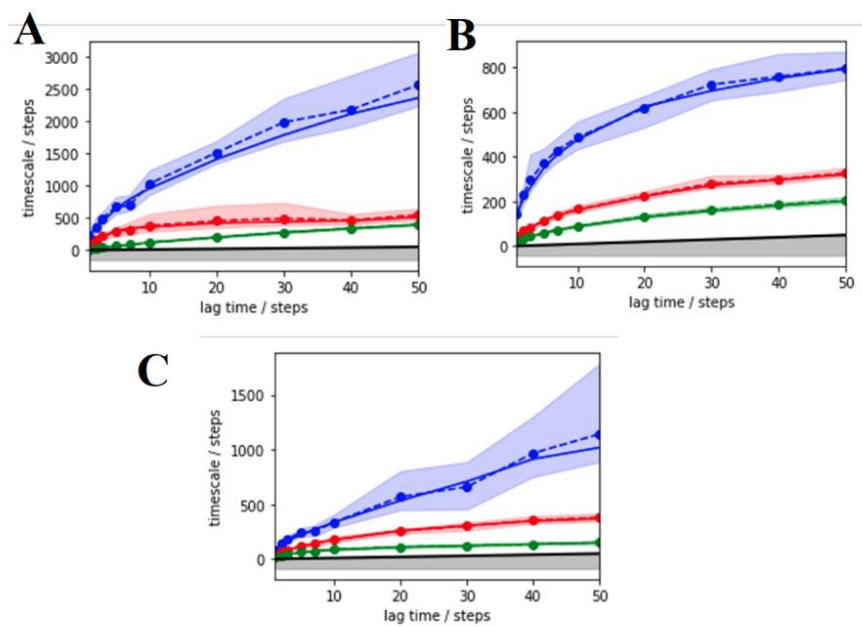
AMBER 99SB ILDN

Pair ID	donor-acceptor	Atom Number	Occupancy (%)
1	130LEU(H) - 129MET(O)	1239 - 1152	96.0
2	128TYR(H) - 131GLY(O)	1201 - 1178	91.2
3	131GLY(H) - 128TYR(O)	1173 - 1220	87.6
4	129MET(H) - 130LEU(O)	1137 - 1256	90.1
5	128TYR(H) - 131GLY(O)	1116 - 1093	89.6
6	131GLY(H) - 128TYR(O)	1088 - 1135	85.9
7	130LEU(H) - 129MET(O)	1069 - 982	96.0
8	128TYR(H) - 131GLY(O)	1031 - 1008	71.6
9	131GLY(H) - 128TYR(O)	1003 - 1050	74.2
10	130LEU(H) - 129MET(O)	984 - 897	98.0
11	129MET(H) - 130LEU(O)	967 - 1086	93.4
12	128TYR(H) - 131GLY(O)	946 - 923	88.9
13	131GLY(H) - 128TYR(O)	918 - 965	95.0
14	130LEU(H) - 129MET(O)	899 - 812	97.4
15	129MET(H) - 130LEU(O)	882 - 1001	91.0
16	131GLY(H) - 128TYR(O)	833 - 880	83.2
17	130LEU(H) - 129MET(O)	814 - 727	98.2
18	129MET(H) - 130LEU(O)	797 - 916	95.8
19	131GLY(H) - 128TYR(O)	748 - 795	86.5
20	129MET(H) - 130LEU(O)	712 - 831	81.7
21	130LEU(H) - 129MET(O)	474 - 387	98.1
22	128TYR(H) - 131GLY(O)	436 - 413	84.3
23	132SER(H) - 127GLY(O)	415 - 264	72.3
24	131GLY(H) - 128TYR(O)	408 - 455	90.3
25	130LEU(H) - 129MET(O)	389 - 302	96.1
26	129MET(H) - 130LEU(O)	372 - 491	90.7
27	128TYR(H) - 131GLY(O)	351 - 328	93.5
28	131GLY(H) - 128TYR(O)	323 - 370	95.1
29	130LEU(H) - 129MET(O)	304 - 217	97.1
30	129MET(H) - 130LEU(O)	287 - 406	94.1
31	128TYR(H) - 131GLY(O)	266 - 243	95.5
32	132SER(HG) - 131GLY(O)	252 - 243	78.1
33	132SER(H) - 127GLY(O)	245 - 94	93.5
34	131GLY(H) - 128TYR(O)	238 - 285	94.8
35	130LEU(H) - 129MET(O)	219 - 132	93.9
36	129MET(H) - 130LEU(O)	202 - 321	92.5
37	128TYR(H) - 131GLY(O)	181 - 158	78.7
38	131GLY(H) - 128TYR(O)	153 - 200	92.1
39	130LEU(H) - 129MET(O)	134 - 47	95.3
40	129MET(H) - 130LEU(O)	117 - 236	98.9
41	129MET(H) - 130LEU(O)	32 - 151	76.5

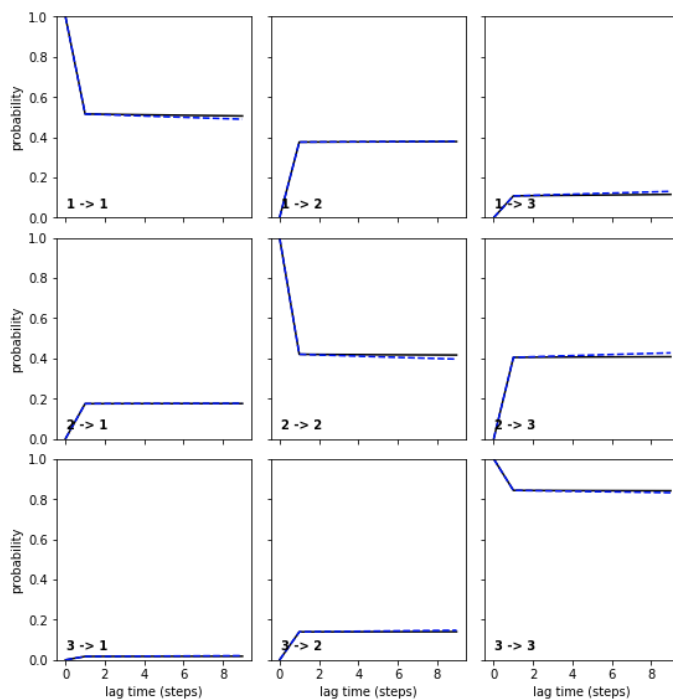
CHARMM 36m

Pair ID	donor-acceptor	Atom Number	Occupancy (%)
1	130LEU(HN) - 127GLY(O)	1239 - 1114	70.6
2	130LEU(HN) - 129MET(O)	1154 - 1067	95.7
3	129MET(HN) - 128TYR(O)	1137 - 1220	73.9
4	131GLY(HN) - 128TYR(O)	1088 - 1135	86.8
5	130LEU(HN) - 129MET(O)	1069 - 982	91.3
6	129MET(HN) - 130LEU(O)	1052 - 1171	72.1
7	129MET(HN) - 130LEU(O)	967 - 1086	80.5
8	130LEU(HN) - 129MET(O)	474 - 387	84.9
9	131GLY(HN) - 128TYR(O)	408 - 455	74.5
10	130LEU(HN) - 129MET(O)	389 - 302	96.8
11	131GLY(HN) - 128TYR(O)	323 - 370	87.4
12	130LEU(HN) - 128TYR(O)	219 - 30	72.1
13	128TYR(HN) - 130LEU(O)	181 - 66	76.7
14	130LEU(HN) - 128TYR(O)	49 - 200	84.9

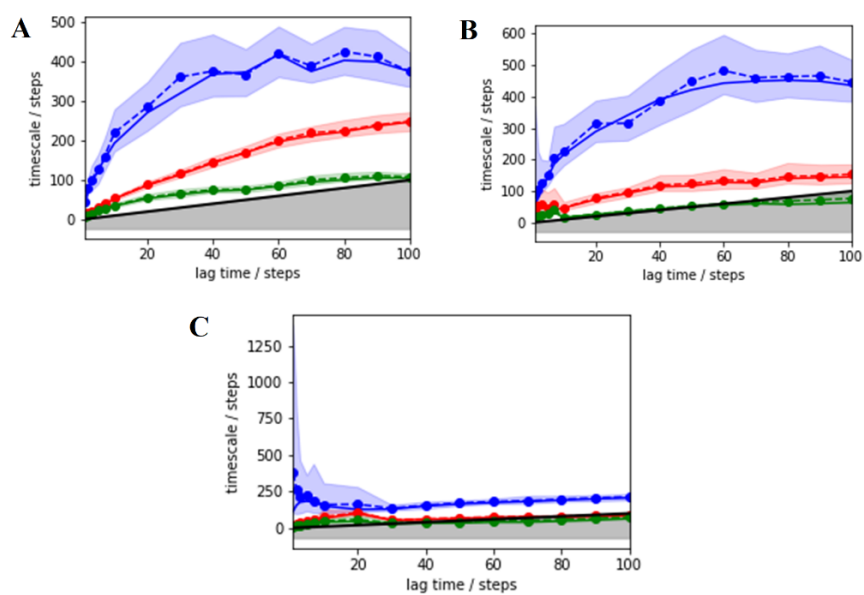
Appendix XXXXXXXXIV. The Implied time scale plot of MSM mode corresponds to oligomer simulation (A) GROMOS-54a7, (B) AMBER-99SB-ILDN, (C) CHARMM-36m.



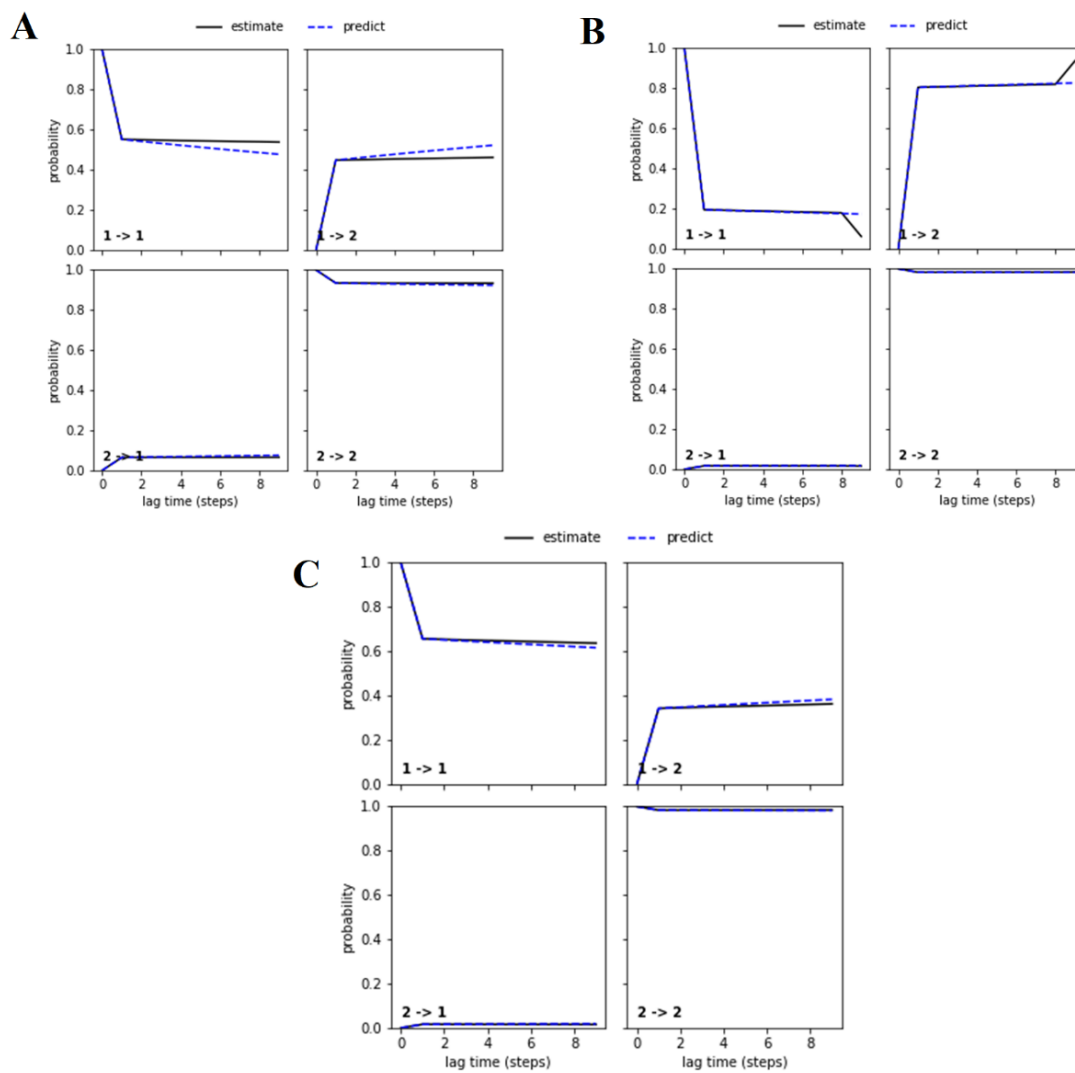
Appendix XXXXXXXXV. Chapman-Kolmogorov validation for Markov State Model for three microstates in the absence of NQDA.



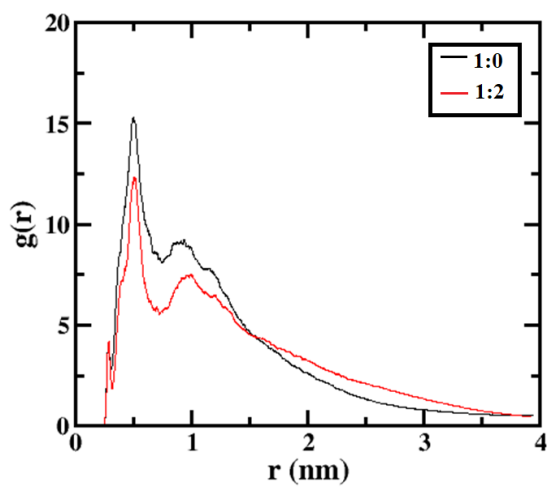
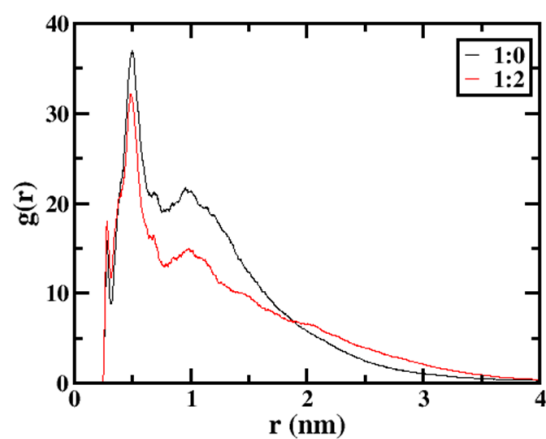
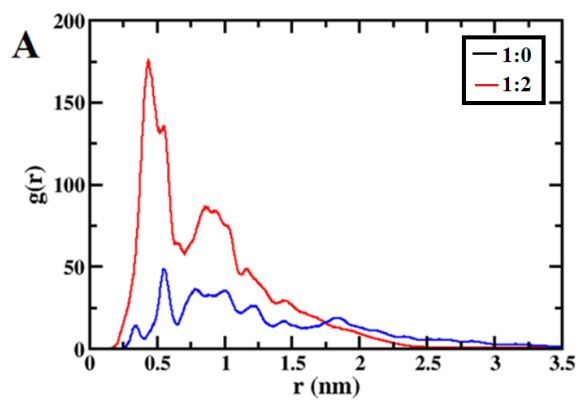
Appendix XXXXXXXXVI. The Implied time scale plot of MSM mode corresponds to oligomer simulation for (A) GROMOS-54a7, (B) AMBER-99SB-ILDN, (C) CHARMM-36m.



Appendix XXXXXXXXVII. The Implied time scale plot of the MSM model corresponds to oligomer simulation for (A) GROMOS-54a7, (B) AMBER-99SB-ILDN, (C) CHARMM-36m in the presence of NQDA.



Appendix XXXXXXVIII. The RDF of peptide heavy-atoms in the presence or absence of NQDA for (A) GROMOS-54a7 [20 mM], (B) AMBER-99SB-ILDN [30 mM], (C) CHARMM-36m [50 mM] at 300K.



Appendix XXXXXXXXIX. The coordination number (CN) of peptides and NQDA around the COM of peptides. The notations P-P, (P-P) NQDA, (P-NQDA), P-W, (P-W)NQDA represents the peptide-peptide and peptide water coordination number in the absence and presence of NQDA respectively.

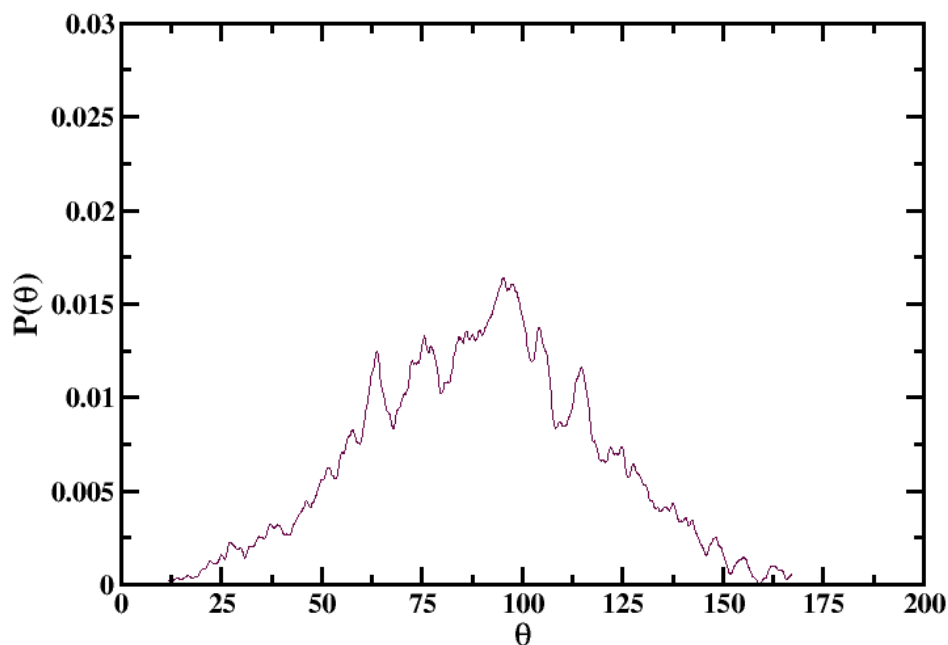
Force-field	CN _{P-P}	CN _{(P-P)NQDA}	CN _(P-NQDA)
GROMOS 54a7 [20 mM]	4	1	5
AMBER 99SB ILDN [30 mM]	3	1	4
CHARMM 36m [50 mM]	2	1	4

Appendix XXXXXXXXX. The decomposition of peptide-peptide and peptide-sol interaction energy with standard error (kJ/mol) for different systems corresponds to the last 100 ns of each simulation.

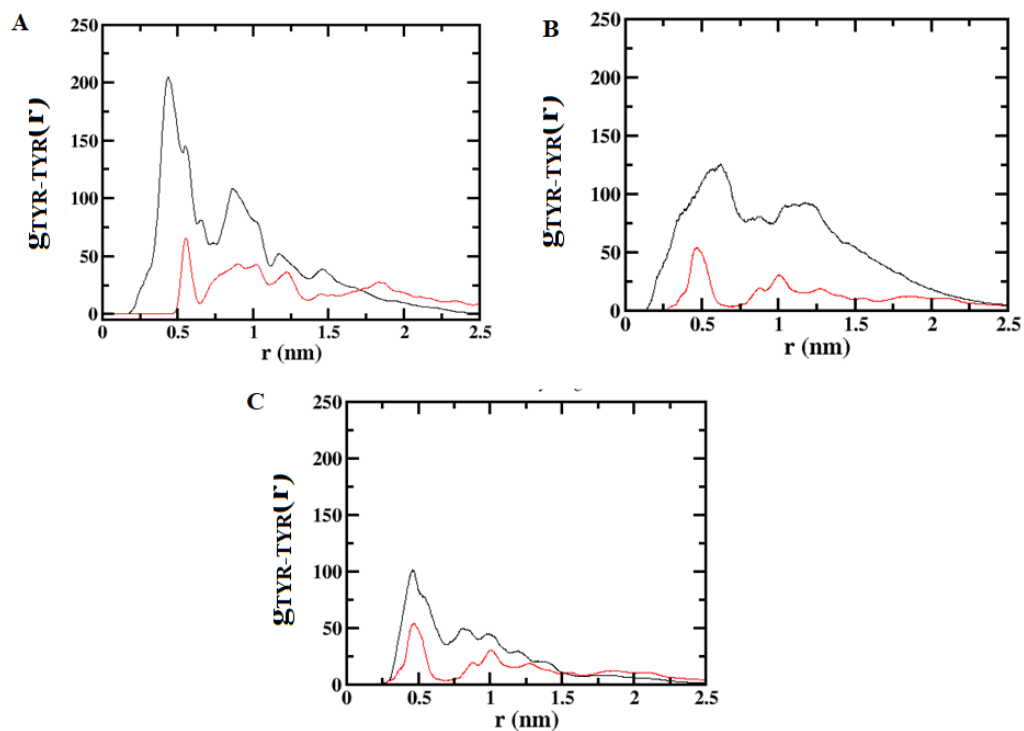
System	Peptide-peptide electrostatic	Peptide-peptide van der Waals	Peptide-water electrostatic	Peptide-water van der Waals
<i>GROMOS 54a7</i>				
G1 [10 mM, 300K]	-26497.8±160	-1713.31±70	-10969.9±380	-519.84±58
G2 [10 mM, 350K]	-26998.7±210	-1943.22±120	-9169.3±490	-263.38±100
G3 [10 mM, 375K]	-26664.8±270	-1804.98±110	-9283.4±490	-356.95±82
G4 [20 mM, 300K]	-26924.7±270	-1880.87±120	-10093.9±620	-398.24±86
G5 [50 mM, 300K]	-27177.6±160	-1985.36±80	-9616.78±350	-305.12±70
G6 [10 mM, 300K]	-25744.1±120	-1136.24±20	-10739.7±540	-481.66±57
(1:2 NQDA) G7 [10 mM, 350K]	-25358.4±60	-920.62 ± 30	-10156.7±250	-167.62±37
(1:2 NQDA) G8 [20 mM, 300K]	-25883.0±20	-1238.17±10	-10239.9±200	-218.61±46
(1:2 NQDA) G9 [10 mM, 375K]	-25915.9±140	-1186.23±50	-10009.5±200	-342.53±41

<i>AMBER 99SB ILDN</i>				
A1[25 mM, 300K]	-	-1818.54±140	-12634.2±320	-964.03±56
A2[25 mM, 320K]	23783.37±210	-1854.48±48	-11795.8±350	-907.09±78
A3[25 mM, 330K]	-24033.4±57	-1871.23±21	-11450.4±260	-861.28±63
A4 [30 mM, 300K]	-23752.1±220	-1813.71±150	-12639±530	-953.072±52
A5[40 mM, 300K]	-24213.9±310	-2080.89±75	-11662.2±340	-753.14±05
A6 [50 mM, 300K]	-24974.72±10	-2197.6±67	-11021.45±05	-698.64±42
A7 [30 mM, 300K] (1:2 NQDA)	-23678.78±30	-1587.41±08	-13789.89±200	-656.89±
<i>CHARMM 36m</i>				
C1[45 mM, 300K]	-22643.6±180	-1296.59±82	-13655.6±530	-1218.18±80
C2[45 mM, 320K]	-22735.2±300	-1465.89±89	-12190.4±580	-1216.07±82
C3 [45 mM, 330K]	-22698.3±210	-1446.48±78	-12076.7±480	-1195.69±74
C4 [30 mM, 300K]	-22340.3±190	-1415.86±47	-12808.3±390	-1237.95±45
C5 [40 mM, 300K]	-22503.5±290	-1457.3±96	-13058.3±560	-1282.28±98
C6 [50mM, 300K]	-22555.6±260	-1568.81±120	-13680.7±560	-706.199±210
C7[50 mM, 300K] (1:2 NQDA)	-21524.9±250	-1296.59±82	-13655.6±530	-1463.93±80

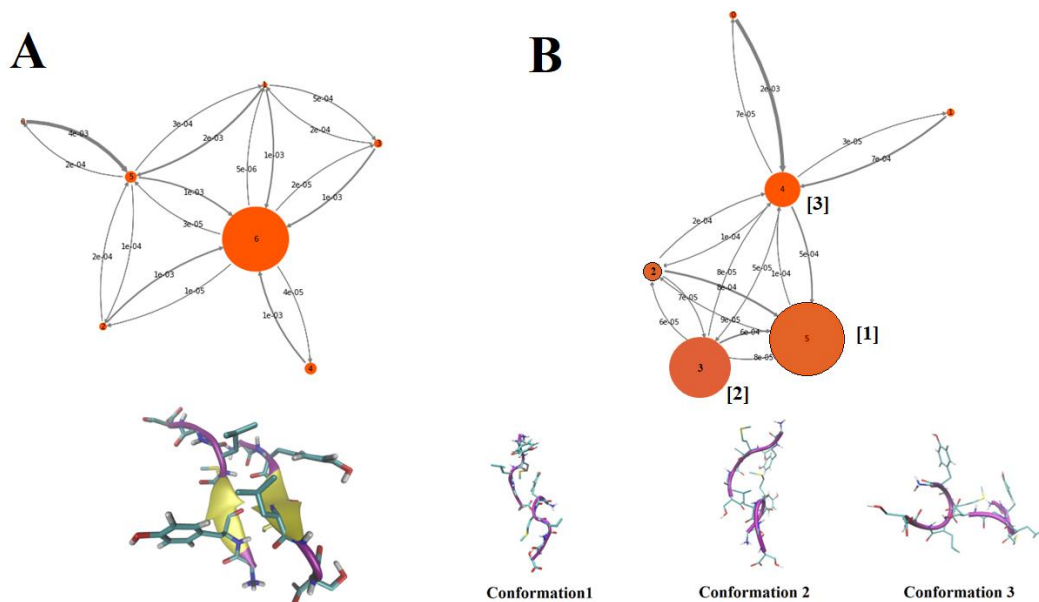
Appendix XXXXXXXXXI. The distribution of angles between the tyrosine rings during the aggregation process.



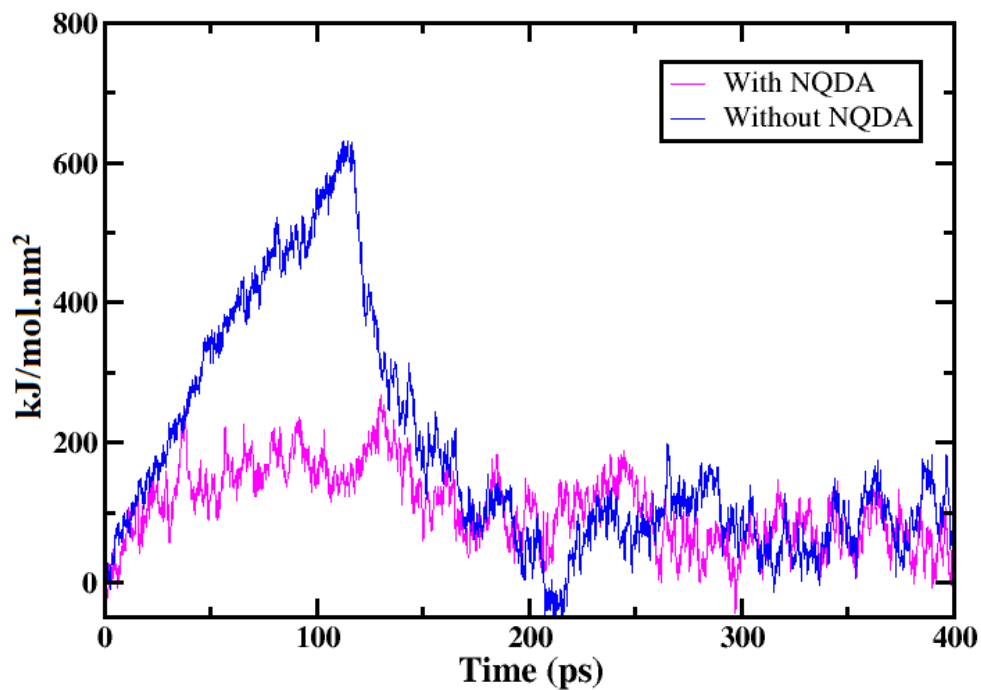
Appendix XXXXXXXXXII. The radial distribution of phenol rings of Tyrosine residues at (A) 300K, (B) 350 K, and (C) 20 mM peptide concentration in the presence (red) and absence of NQDA (black) in GROMOS force-field.



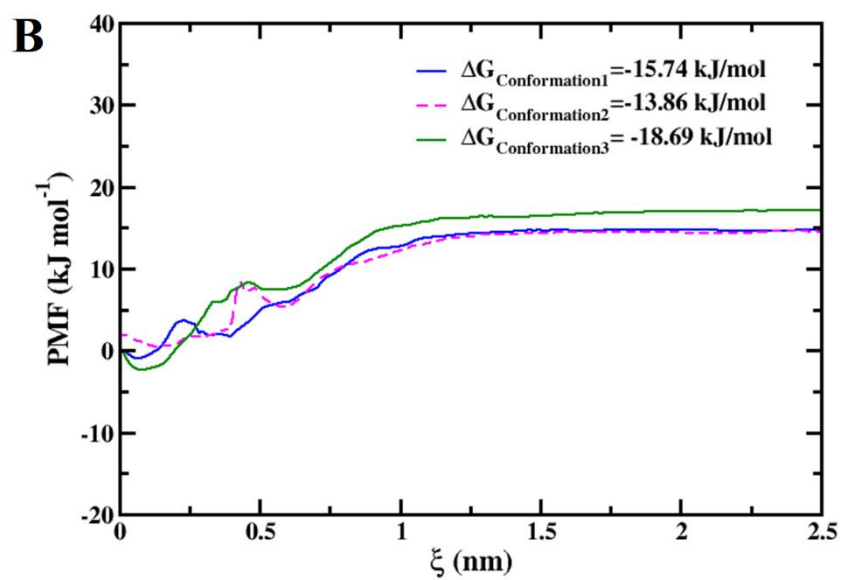
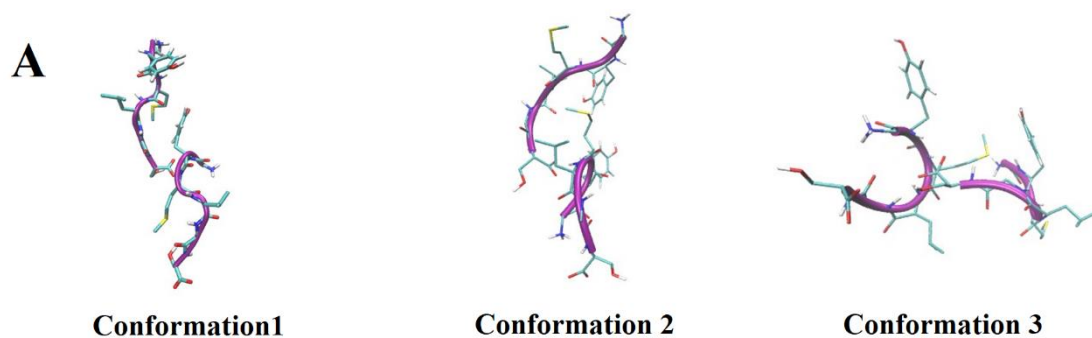
Appendix XXXXXXXXXXIII. The transition path between metastable states of dimer simulation in (A) absence or (B) presence of NQDA.



Appendix XXXXXXXXXXIV. The average force profile with respect to the simulation time scale for monomer dissociation from peptide dimer in the presence or absence of NQDA.



Appendix XXXXXXXXXV. The representative potential of mean force (PMF) graph for dimer dissociation in the presence of NQDA for the conformation1, conformation2, and conformation3.



REFERENCES

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015a). "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers." *SoftwareX*, 1–2, 19–25.
- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015b). "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers." *SoftwareX*, 1–2, 19–25.
- Achari, A., Somers, D. O., Champness, J. N., Bryant, P. K., Rosemond, J., and Stammers, D. K. (1997a). "Crystal structure of the anti-bacterial sulfonamide drug target dihydropteroate synthase." *Nat Struct Biol*, 4(6), 490–497.
- Achari, A., Somers, D. O., Champness, J. N., Bryant, P. K., Rosemond, J., and Stammers, D. K. (1997b). "Crystal structure of the anti-bacterial sulfonamide drug target dihydropteroate synthase." *Nat. Struct. Biol.*, 4(6), 490–497.
- Adebayo, G. A., Anusionwu, B. C., Njah, A. N., Adeniran, O. J., Mathew, B., and Sunmonu, R. S. (2010). "Collision frequency of Lennard-Jones fluids at high densities by equilibrium molecular dynamics simulation." *Pramana - J Phys*, 75(3), 523–536.
- Agudelo, W. A., and Patarroyo, M. E. (2010). "Quantum Chemical Analysis of MHC-Peptide Interactions for Vaccine Design." *Mini Rev Med Chem*, 10(8), 746–758.
- Ajmani, S., Jadhav, K., and Kulkarni, S. A. (2006). "Three-Dimensional QSAR Using the k-Nearest Neighbor Method and Its Interpretation." *J. Chem. Inf. Model.*, 46(1), 24–31.
- Akkaya, M., Kwak, K., and Pierce, S. K. (2020). "B cell memory: building two walls of protection against pathogens." *Nature Reviews Immunology*, 20(4), 229–238.
- 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." *Journal of Biomolecular Structure and Dynamics*, 39(18), 7263–7273.
- Alam, S., and Khan, F. (2019). "3D-QSAR, Docking, ADME/Tox studies on Flavone analogs reveal anticancer activity through Tankyrase inhibition." *Sci Rep*, 9(1), 5414.
- Alderete, J., and Neace, C. J. (2013). "Identification, characterization, and synthesis of peptide epitopes and a recombinant six-epitope protein for *Trichomonas vaginalis* serodiagnosis." *Immunotargets Ther*, 2, 91–103.
- Almeida, J., Volpini, J., Poiani, J., and Silva, C. (2014). "Ligand-Based Drug Design of Novel MARK-3 Inhibitors in Cancer." *CBC*, 10(2), 112–123.
- Andrada, M. F., Vega-Hissi, E. G., Estrada, M. R., and Martinez, J. C. G. (2017). "Impact assessment of the rational selection of training and test sets on the predictive ability of QSAR models." *SAR and QSAR in Environmental Research*, 28(12), 1011–1023.
- Andrade-Ochoa, S., García-Machorro, J., Bello, M., Rodríguez-Valdez, L. M., Flores-Sandoval, C. A., and Correa-Basurto, J. (2018). "QSAR, DFT and molecular modeling studies of peptides from HIV-1 to describe their recognition properties by MHC-I." *J Biomol Struct Dyn*, 36(9), 2312–2330.
- Andreatta, M., and Nielsen, M. (2016). "Gapped sequence alignment using artificial neural networks: application to the MHC class I system." *Bioinformatics*, 32(4), 511–517.
- Apostol, M. I., Perry, K., and Surewicz, W. K. (2013). "Crystal Structure of a Human Prion Protein Fragment Reveals a Motif for Oligomer Formation." *J. Am. Chem. Soc.*, 135(28), 10202–10205.

Arooj, M., Sakkiah, S., Cao, G. ping, and Lee, K. W. (2013). “An innovative strategy for dual inhibitor design and its application in dual inhibition of human thymidylate synthase and dihydrofolate reductase enzymes.” *PLoS One*, 8(4), e60470.

Arosio, P., Knowles, T. P. J., and Linse, S. (2015). “On the lag phase in amyloid fibril formation.” *Phys. Chem. Chem. Phys.*, 17(12), 7606–7618.

Averill, F. W., and Painter, G. S. (1992). “Steepest-descent determination of occupation numbers and energy minimization in the local-density approximation.” *Phys. Rev. B*, 46(4), 2498–2502.

Azmi, F., Ahmad Fuaad, A. A. H., Skwarczynski, M., and Toth, I. (2014). “Recent progress in adjuvant discovery for peptide-based subunit vaccines.” *Human Vaccines & Immunotherapeutics*, 10(3), 778–796.

Azzam, R. A., Elsayed, R. E., and Elgemeie, G. H. (2020). “Design, Synthesis, and Antimicrobial Evaluation of a New Series of *N*-Sulfonamide 2-Pyridones as Dual Inhibitors of DHPS and DHFR Enzymes.” *ACS Omega*, 5(18), 10401–10414.

Babaoglu, K., Qi, J., Lee, R. E., and White, S. W. (2004). “Crystal Structure of 7,8-Dihydropteroate Synthase from *Bacillus anthracis*: Mechanism and Novel Inhibitor Design.” *Structure*, 12(9), 1705–1717.

Baca, A. M., Sirawaraporn, R., Turley, S., Sirawaraporn, W., and Hol, W. G. (2000). “Crystal structure of *Mycobacterium tuberculosis* 7,8-dihydropteroate synthase in complex with pterin monophosphate: new insight into the enzymatic mechanism and sulfa-drug action.” *J Mol Biol*, 302(5), 1193–1212.

Bachmann, M. F., and Jennings, G. T. (2010). “Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns.” *Nat Rev Immunol*, 10(11), 787–796.

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.” *Biochemistry Research International*, 2016, e7264080.

Balbirnie, M., Grothe, R., and Eisenberg, D. S. (2001). “An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated β -sheet structure for amyloid.” *PNAS*, 98(5), 2375–2380.

Baral, P., Pavadai, E., Gerstman, B. S., and Chapagain, P. P. (2020). “In-silico identification of the vaccine candidate epitopes against the Lassa virus hemorrhagic fever.” *Sci Rep*, 10(1), 7667.

Bash, P. A., Singh, U. C., Brown, F. K., Langridge, R., and Kollman, P. A. (1987). “Calculation of the relative change in binding free energy of a protein-inhibitor complex.” *Science*, 235(4788), 574–576.

Becke, A. D. (1993). “Density-functional thermochemistry. III. The role of exact exchange.” *J. Chem. Phys.*, 98(7), 5648–5652.

Bemis, G. W., and Murcko, M. A. (1996). “The properties of known drugs. 1. Molecular frameworks.” *J. Med. Chem.*, 39(15), 2887–2893.

Bemporad, F., and Chiti, F. (2012). “Protein Misfolded Oligomers: Experimental Approaches, Mechanism of Formation, and Structure-Toxicity Relationships.” *Chemistry & Biology*, 19(3), 315–327.

Bendheim, P. E., Brown, H. R., Rudelli, R. D., Scala, L. J., Goller, N. L., Wen, G. Y., Kascak, R. J., Cashman, N. R., and Bolton, D. C. (1992). “Nearly ubiquitous tissue distribution of the scrapie agent precursor protein.” *Neurology*, 42(1), 149–156.

- Bennett, C. H. (1976). "Efficient estimation of free energy differences from Monte Carlo data." *Journal of Computational Physics*, 22, 245–268.
- Berendsen, H. J. C., Spoel, D. van der, and Drunen, R. van. (1995). "GROMACS: A message-passing parallel molecular dynamics implementation." *Computer Physics Communications*, 91(1), 43–56.
- Birmingham Alun and Derrick Jeremy P. (2002). "The folic acid biosynthesis pathway in bacteria: evaluation of potential for antibacterial drug discovery." *BioEssays*, 24(7), 637–648.
- Best, R. B., Buchete, N.-V., and Hummer, G. (2008a). "Are Current Molecular Dynamics Force Fields too Helical?" *Biophysical Journal*, 95(1), L07–L09.
- Best, R. B., Buchete, N.-V., and Hummer, G. (2008b). "Are Current Molecular Dynamics Force Fields too Helical?" *Biophysical Journal*, 95(1), L07–L09.
- 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.
- Bhole, R. P., Bonde, C. G., Bonde, S. C., Chikhale, R. V., and Wavhale, R. D. (2021). "Pharmacophore model and atom-based 3D quantitative structure activity relationship (QSAR) of human immunodeficiency virus-1 (HIV-1) capsid assembly inhibitors." *Journal of Biomolecular Structure and Dynamics*, 39(2), 718–727.
- Bishop, M. F., and Ferrone, F. A. (1984). "Kinetics of nucleation-controlled polymerization. A perturbation treatment for use with a secondary pathway." *Biophysical Journal*, 46(5), 631–644.
- Bosch, B. J., Zee, R. van der, Haan, C. A. M. de, and Rottier, P. J. M. (2003). "The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex." *J. Virol.*, 77(16), 8801–8811.
- Bourne, C. R. (2014). "Utility of the Biosynthetic Folate Pathway for Targets in Antimicrobial Discovery." *Antibiotics (Basel)*, 3(1), 1–28.
- Bowers, K. J., Chow, E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., Klepeis, J. L., Kolossvary, I., Moraes, M. A., Sacerdoti, F. D., Salmon, J. K., Shan, Y., and Shaw, D. E. (2006). "Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters." *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, SC '06, New York, NY, USA: ACM.
- 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." *Phytotherapy Research*, 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.
- Burra, G., Maina, M. B., Serpell, L. C., and Thakur, A. K. (2021). "Nucleation-dependent Aggregation Kinetics of Yeast Sup35 Fragment GNNQQNY." *Journal of Molecular Biology*, 433(3), 166732.
- Bussi, G., Donadio, D., and Parrinello, M. (2007a). "Canonical sampling through velocity rescaling." *J. Chem. Phys.*, 126(1), 014101.
- Bussi, G., Donadio, D., and Parrinello, M. (2007b). "Canonical sampling through velocity rescaling." *J. Chem. Phys.*, 126(1), 014101.

Carballo-Pacheco, M., and Strodel, B. (2016). “Advances in the Simulation of Protein Aggregation at the Atomistic Scale.” *J. Phys. Chem. B*, 120(12), 2991–2999.

Carrasco, Y. R., and Batista, F. D. (2007). “B Cells Acquire Particulate Antigen in a Macrophage-Rich Area at the Boundary between the Follicle and the Subcapsular Sinus of the Lymph Node.” *Immunity*, 27(1), 160–171.

Carrell, R. W., and Lomas, D. A. (1997). “Conformational disease.” *The Lancet*, 350(9071), 134–138.

Carugo, O., and Pongor, S. (2001). “A normalized root-mean-square distance for comparing protein three-dimensional structures.” *Protein Science*, 10(7), 1470–1473.

Cereto-Massagué, A., Guasch, L., Valls, C., Mulero, M., Pujadas, G., and Garcia-Vallvé, S. (2012). “DecoyFinder: an easy-to-use python GUI application for building target-specific decoy sets.” *Bioinformatics*, 28(12), 1661–1662.

Chakraborty, D., and Patey, G. N. (2013). “How Crystals Nucleate and Grow in Aqueous NaCl Solution.” *J. Phys. Chem. Lett.*, 4(4), 573–578.

Chamachi, N. G., and Chakraborty, S. (2016). “Replica Exchange Molecular Dynamics Study of Dimerization in Prion Protein: Multiple Modes of Interaction and Stabilization.” *J. Phys. Chem. B*, 120(30), 7332–7345.

Chandra, A. (2000). “Effects of Ion Atmosphere on Hydrogen-Bond Dynamics in Aqueous Electrolyte Solutions.” *Phys. Rev. Lett.*, 85(4), 768–771.

Charles A Janeway, J., Travers, P., Walport, M., and Shlomchik, M. J. (2001). “The Humoral Immune Response.” *Immunobiology: The Immune System in Health and Disease. 5th edition.*

Chatani, E., Imamura, H., Yamamoto, N., and Kato, M. (2014). “Stepwise Organization of the β -Structure Identifies Key Regions Essential for the Propagation and Cytotoxicity of Insulin Amyloid Fibrils.” *J Biol Chem*, 289(15), 10399–10410.

Chen, D., Ranganathan, A., IJzerman, A. P., Siegal, G., and Carlsson, J. (2013). “Complementarity between in Silico and Biophysical Screening Approaches in Fragment-Based Lead Discovery against the A2A Adenosine Receptor.” *J. Chem. Inf. Model.*, 53(10), 2701–2714.

Chen, T., Li, M., and Liu, J. (2018). “ π - π Stacking Interaction: A Nondestructive and Facile Means in Material Engineering for Bioapplications.” *Crystal Growth & Design*, 18(5), 2765–2783.

Chen, W.-H., Strych, U., Hotez, P. J., and Bottazzi, M. E. (2020). “The SARS-CoV-2 Vaccine Pipeline: an Overview.” *Curr Trop Med Rep*, 1–4.

Chen, Z., Li, H., Zhang, Q., Bao, X., Yu, K., Luo, X., Zhu, W., and Jiang, H. (2009). “Pharmacophore-based virtual screening versus docking-based virtual screening: a benchmark comparison against eight targets.” *Acta Pharmacol Sin*, 30(12), 1694–1708.

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.

Chou, P. Y., and Fasman, G. D. (1977). “Secondary structural prediction of proteins from their amino acid sequence.” *Trends in Biochemical Sciences*, 2(6), 128–131.

Chou, P. Y., and Fasman, G. D. (1978). “Prediction of the secondary structure of proteins from their amino acid sequence.” *Adv Enzymol Relat Areas Mol Biol*, 47, 45–148.

- Christ, C. D., Mark, A. E., and Gunsteren, W. F. van. (2010). "Basic ingredients of free energy calculations: A review." *Journal of Computational Chemistry*, 31(8), 1569–1582.
- Clinton, J., Forsyth, C., Royston, M. C., and Roberts, G. W. (1993). "Synaptic degeneration is the primary neuropathological feature in prion disease: a preliminary study." *Neuroreport*, 4(1), 65–68.
- Collinge, J. (2001). "Prion diseases of humans and animals: their causes and molecular basis." *Annu. Rev. Neurosci.*, 24, 519–550.
- Correa-Basurto, J., Ramos-Morales, F. R., Matus, M. H., Rosales-Hernández, M. C., Mancilla-Percino, T., Trujillo-Ferrara, J., and Ilizaliturri-Flores, I. (2012). "Docking and DFT Studies to explore the Topoisomerase II ATP Pocket employing 3-Substituted 2,6-Piperazindiones for drug design." *Molecular Simulation*, 38(13), 1072–1084.
- Csermely, P., Palotai, R., and Nussinov, R. (2010). "Induced fit, conformational selection and independent dynamic segments: an extended view of binding events." *Trends in Biochemical Sciences*, 35(10), 539–546.
- Cui, J., Li, F., and Shi, Z.-L. (2019). "Origin and evolution of pathogenic coronaviruses." *Nature Reviews Microbiology*, 17(3), 181–192.
- Cusinato, J., Cau, Y., Calvani, A. M., and Mori, M. (2021). "Repurposing drugs for the management of COVID-19." *Expert Opinion on Therapeutic Patents*, 31(4), 295–307.
- Dar, H., Zaheer, T., Rehman, M. T., Ali, A., Javed, A., Khan, G. A., Babar, M. M., and Waheed, Y. (2016). "Prediction of promiscuous T-cell epitopes in the Zika virus polyprotein: An in silico approach." *Asian Pacific Journal of Tropical Medicine*, 9(9), 844–850.
- Darden, T., York, D., and Pedersen, L. (1993). "Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems." *J. Chem. Phys.*, 98(12), 10089–10092.
- Das, B. K., and Chakraborty, D. (2020). "Epitope-Based Potential Vaccine Candidate for Humoral and Cell-Mediated Immunity to Combat Severe Acute Respiratory Syndrome Coronavirus 2 Pandemic." *J. Phys. Chem. Lett.*, 11(22), 9920–9930.
- Das, B. K., and Chakraborty, D. (2021). "Deciphering the competitive inhibition of dihydropteroate synthase by 8 marcapoguanine analogs: enhanced potency in phenylsulfonyl fragments." *Journal of Biomolecular Structure and Dynamics*, 0(0), 1–20.
- Das, B. K., Pv, 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.
- Dauber-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, J., Genest, M., and Hagler, A. T. (1988). "Structure and energetics of ligand binding to proteins: Escherichia coli dihydrofolate reductase-trimethoprim, a drug-receptor system." *Proteins*, 4(1), 31–47.
- De Benedetti, P. G., Iarossi, D., Folli, U., Frassinetti, C., Menziani, M. C., and Cennamo, C. (1989). "Quantitative structure-activity relationships in dihydropteroate synthase inhibition by multisubstituted sulfones. Design and synthesis of some new derivatives with improved potency." *J. Med. Chem.*, 32(10), 2396–2399.
- De Benedetti, P. G., Iarossi, D., Menziani, C., Caiolfa, V., Frassinetti, C., and Cennamo, C. (1987). "Quantitative structure-activity analysis in dihydropteroate synthase inhibition of sulfones. Comparison with sulfanilamides." *J. Med. Chem.*, 30(3), 459–464.

Dear, A. J., Michaels, T. C. T., Meisl, G., Klenerman, D., Wu, S., Perrett, S., Linse, S., Dobson, C. M., and Knowles, T. P. J. (2020). “Kinetic diversity of amyloid oligomers.” *Proceedings of the National Academy of Sciences*, 117(22), 12087–12094.

Dennis, M. L., Chhabra, S., Wang, Z.-C., Debono, A., Dolezal, O., Newman, J., Pitcher, N. P., Rahmani, R., Cleary, B., Barlow, N., Hattarki, M., Graham, B., Peat, T. S., Baell, J. B., and Swarbrick, J. D. (2014). “Structure-Based Design and Development of Functionalized Mercaptoguanine Derivatives as Inhibitors of the Folate Biosynthesis Pathway Enzyme 6-Hydroxymethyl-7,8-dihydropterin Pyrophosphokinase from *Staphylococcus aureus*.” *J. Med. Chem.*, 57(22), 9612–9626.

Dennis, M. L., Lee, M. D., Harjani, J. R., Ahmed, M., DeBono, A. J., Pitcher, N. P., Wang, Z.-C., Chhabra, S., Barlow, N., Rahmani, R., Cleary, B., Dolezal, O., Hattarki, M., Aurelio, L., Shonberg, J., Graham, B., Peat, T. S., Baell, J. B., and Swarbrick, J. D. (2018). “8-Mercaptoguanine Derivatives as Inhibitors of Dihydropteroate Synthase.” *Chemistry – A European Journal*, 24(8), 1922–1930.

Dennis, M. L., Pitcher, N. P., Lee, M. D., DeBono, A. J., Wang, Z.-C., Harjani, J. R., Rahmani, R., Cleary, B., Peat, T. S., Baell, J. B., and Swarbrick, J. D. (2016). “Structural Basis for the Selective Binding of Inhibitors to 6-Hydroxymethyl-7,8-dihydropterin Pyrophosphokinase from *Staphylococcus aureus* and *Escherichia coli*.” *J Med Chem*, 59(11), 5248–5263.

Dimitrov, I., Bangov, I., Flower, D. R., and Doytchinova, I. (2014). “AllerTOP v.2—a server for in silico prediction of allergens.” *J Mol Model*, 20(6), 2278.

Dixon, S. L., Smondyrev, 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. Comput. Aided Mol. Des.*, 20(10–11), 647–671.

Domagk, G. (1935). “Ein Beitrag zur Chemotherapie der bakteriellen Infektionen.” *Dtsch med Wochenschr*, 61(7), 250–253.

Doytchinova, I. A., and Flower, D. R. (2007). “VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines.” *BMC Bioinformatics*, 8, 4.

Duan, L. L., Tong, Y., Mei, Y., Zhang, Q. G., and Zhang, J. Z. H. (2007). “Quantum study of HIV-1 protease-bridge water interaction.” *J Chem Phys*, 127(14), 145101.

Duffy, E. M., and Jorgensen, W. L. (2000). “Prediction of Properties from Simulations: Free Energies of Solvation in Hexadecane, Octanol, and Water.” *Journal of the American Chemical Society*, 122(12), 2878–2888.

Dunn, M. F. (2010). “Protein–Ligand Interactions: General Description.” *eLS*, John Wiley & Sons, Ltd.

Ekins, S., Boulanger, B., Swaan, P. W., and Hupcey, M. A. Z. (2002). “Towards a new age of virtual ADME/TOX and multidimensional drug discovery.” *Mol. Divers.*, 5(4), 255–275.

Elcock, A. H., Sept, D., and McCammon, J. A. (2001). “Computer Simulation of Protein–Protein Interactions.” *J. Phys. Chem. B*, 105(8), 1504–1518.

Emini, E. A., Hughes, J. V., Perlow, D. S., and Boger, J. (1985a). “Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide.” *Journal of Virology*, 55(3), 836–839.

Emini, E., Hughes, J., Perlow, D., and Boger, J. (1985b). “Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide.” *Journal of virology*.

- Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995a). "A smooth particle mesh Ewald method." *J. Chem. Phys.*, 103(19), 8577–8593.
- Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995b). "A smooth particle mesh Ewald method." *J. Chem. Phys.*, 103(19), 8577–8593.
- Everitt, B. (n.d.). "Cluster Analysis." 348.
- Fehr, A. R., and Perlman, S. (2015). "Coronaviruses: An Overview of Their Replication and Pathogenesis." *Coronaviruses: Methods and Protocols*, Methods in Molecular Biology, H. J. Maier, E. Bickerton, and P. Britton, eds., New York, NY: Springer, 1–23.
- Ferguson David M. (1995). "Parameterization and evaluation of a flexible water model." *Journal of Computational Chemistry*, 16(4), 501–511.
- Ferrone, F. (1999). "Analysis of protein aggregation kinetics." *Meth. Enzymol.*, 309, 256–274.
- Foloppe, N., and Hubbard, R. (2006). "Towards predictive ligand design with free-energy based computational methods?" *Curr. Med. Chem.*, 13(29), 3583–3608.
- Fraser, H. (1993). "Diversity in the neuropathology of scrapie-like diseases in animals." *Br Med Bull*, 49(4), 792–809.
- Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004). "Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy." *J. Med. Chem.*, 47(7), 1739–1749.
- Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., and Mainz, D. T. (2006). "Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes." *J. Med. Chem.*, 49(21), 6177–6196.
- 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.
- Fukunishi, H., Watanabe, O., and Takada, S. (2002). "On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction." *The Journal of Chemical Physics*, 116(20), 9058–9067.
- Gajdusek, D. C., Gibbs, C. J., and Alpers, M. (1966). "Experimental Transmission of a Kuru-like Syndrome to Chimpanzees." *Nature*, 209(5025), 794–796.
- Galindo-Murillo, R., Roe, D. R., and Cheatham, T. E. (2015). "Convergence and reproducibility in molecular dynamics simulations of the DNA duplex

d(GCACGAACGAACGAACGC).” *Biochimica et Biophysica Acta (BBA) - General Subjects*, Recent developments of molecular dynamics, 1850(5), 1041–1058.

Gallagher-Jones, M., Glynn, C., Boyer, D. R., Martynowycz, M. W., Hernandez, E., Miao, J., Zee, C.-T., Novikova, I. V., Goldschmidt, L., McFarlane, H. T., Helguera, G. F., Evans, J. E., Sawaya, M. R., Cascio, D., Eisenberg, D. S., Gonen, T., and Rodriguez, J. A. (2018). “Sub-ångström cryo-EM structure of a prion protofibril reveals a polar clasp.” *Nature Structural & Molecular Biology*, 25(2), 131–134.

Gallivan, J. P., and Dougherty, D. A. (1999). “Cation- π interactions in structural biology.” *PNAS*, 96(17), 9459–9464.

Ganguly, P., Boserman, P., Vegt, N. F. A. van der, and Shea, J.-E. (2018). “Trimethylamine N-oxide Counteracts Urea Denaturation by Inhibiting Protein–Urea Preferential Interaction.” *J. Am. Chem. Soc.*, 140(1), 483–492.

Ganoth, A., Friedman, R., Nachliel, E., and Gutman, M. (2006). “A Molecular Dynamics Study and Free Energy Analysis of Complexes between the Mlc1p Protein and Two IQ Motif Peptides.” *Biophys J*, 91(7), 2436–2450.

Garro Martinez, J. C., Andrada, M. F., Vega-Hissi, E. G., Garibotto, F. M., Nogueras, M., Rodríguez, R., Cobo, J., Enriz, R. D., and Estrada, M. R. (2017). “Dihydrofolate reductase inhibitors: a quantitative structure–activity relationship study using 2D-QSAR and 3D-QSAR methods.” *Med Chem Res*, 26(1), 247–261.

Gasteiger, J., and Marsili, M. (1980). “Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges.” *Tetrahedron*, 36(22), 3219–3228. “Gaussian 09 Citation | Gaussian.com.” (n.d.). <<http://gaussian.com/g09citation/>> (Apr. 1, 2018).

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

Geysen, H. M., Meloen, R. H., and Barteling, S. J. (1984). “Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid.” *PNAS*, 81(13), 3998–4002.

Ghosh, R., Chakraborty, A., Biswas, A., and Chowdhuri, S. (2020a). “Evaluation of green tea polyphenols as novel corona virus (SARS CoV-2) main protease (Mpro) inhibitors – an in silico docking and molecular dynamics simulation study.” *Journal of Biomolecular Structure and Dynamics*, 0(0), 1–13.

Ghosh, R., Chakraborty, A., Biswas, A., and Chowdhuri, S. (2020b). “Identification of polyphenols from *Broussonetia papyrifera* as SARS CoV-2 main protease inhibitors using in silico docking and molecular dynamics simulation approaches.” *Journal of Biomolecular Structure and Dynamics*, 0(0), 1–14.

Gilson, M. K., and Honig, B. (1988a). “Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis.” *Proteins: Structure, Function, and Bioinformatics*, 4(1), 7–18.

Gilson, M. K., and Honig, B. (1988b). “Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis.” *Proteins: Structure, Function, and Bioinformatics*, 4(1), 7–18.

Gohlke, H., Kiel, C., and Case, D. A. (2003). “Insights into Protein–Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras–Raf and Ras–RalGDS Complexes.” *Journal of Molecular Biology*, 330(4), 891–913.

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. (2000a). "Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection." *Mol Divers*, 5(4), 231–243.

Golbraikh, A., and Tropsha, A. (2000b). "Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection." *Mol Divers*, 5(4), 231–243.

Golbraikh, A., and Tropsha, A. (2000c). "Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection." *Mol Divers*, 5(4), 231–243.

Goldfeld, D. A., Murphy, R., Kim, B., Wang, L., Beuming, T., Abel, R., and Friesner, R. A. (2015). "Docking and Free Energy Perturbation Studies of Ligand Binding in the Kappa Opioid Receptor." *J. Phys. Chem. B*, 119(3), 824–835.

Gordon, D. E., Hiatt, J., Bouhaddou, M., Rezelj, V. V., Ulferts, S., Braberg, H., Jureka, A. S., Obernier, K., Guo, J. Z., Batra, J., Kaake, R. M., Weckstein, A. R., Owens, T. W., Gupta, M., Pourmal, S., Titus, E. W., Cakir, M., Soucheray, M., McGregor, M., Cakir, Z., Jang, G., O'Meara, M. J., Tummino, T. A., Zhang, Z., Foussard, H., Rojce, A., Zhou, Y., Kuchenov, D., Hüttenhain, R., Xu, J., Eckhardt, M., Swaney, D. L., Fabius, J. M., Ummadi, M., Tutuncuoglu, B., Rathore, U., Modak, M., Haas, P., Haas, K. M., Naing, Z. Z. C., Pulido, E. H., Shi, Y., Barrio-Hernandez, I., Memon, D., Petsalaki, E., Dunham, A., Marrero, M. C., Burke, D., Koh, C., Vallet, T., Silvas, J. A., Azumaya, C. M., Billesbølle, C., Brilot, A. F., Campbell, M. G., Diallo, A., Dickinson, M. S., Diwanji, D., Herrera, N., Hoppe, N., Kratochvil, H. T., Liu, Y., Merz, G. E., Moritz, M., Nguyen, H. C., Nowotny, C., Puchades, C., Rizo, A. N., Schulze-Gahmen, U., Smith, A. M., Sun, M., Young, I. D., Zhao, J., Asarnow, D., Biel, J., Bowen, A., Braxton, J. R., Chen, J., Chio, C. M., Chio, U. S., Deshpande, I., Doan, L., Faust, B., Flores, S., Jin, M., Kim, K., Lam, V. L., Li, F., Li, J., Li, Y.-L., Li, Y., Liu, X., Lo, M., Lopez, K. E., Melo, A. A., Moss, F. R., Nguyen, P., Paulino, J., Pawar, K. I., Peters, J. K., Pospiech, T. H., Safari, M., Sangwan, S., Schaefer, K., Thomas, P. V., Thwin, A. C., Trenker, R., Tse, E., Tsui, T. K. M., Wang, F., Whitis, N., Yu, Z., Zhang, K., Zhang, Y., Zhou, F., Saltzberg, D., Consortium12[†], Q. S. B., Hodder, A. J., Shun-Shion, A. S., Williams, D. M., White, K. M., Rosales, R., Kehrer, T., Miorin, L., Moreno, E., Patel, A. H., Rihn, S., Khalid, M. M., Vallejo-Gracia, A., Fozouni, P., Simoneau, C. R., Roth, T. L., Wu, D., Karim, M. A., Ghoussaini, M., Dunham, I., Berardi, F., Weigang, S., Chazal, M., Park, J., Logue, J., McGrath, M., Weston, S., Haupt, R., Hastie, C. J., Elliott, M., Brown, F., Burness, K. A., Reid, E., Dorward, M., Johnson, C., Wilkinson, S. G., Geyer, A., Giesel, D. M., Baillie, C., Raggett, S., Leech, H., Toth, R., Goodman, N., Keough, K. C., Lind, A. L., Consortium[‡], Z., Klesh, R. J., Hemphill, K. R., Carlson-Stevermer, J., Oki, J., Holden, K., Maures, T., Pollard, K. S., Sali, A., Agard, D. A., Cheng, Y., Fraser, J. S., Frost, A., Jura, N., Kortemme, T., Manglik, A., Southworth, D. R., Stroud, R. M., Alessi, D. R., Davies, P., Frieman, M. B., Ideker, T., Abate, C., Jouvenet, N., Kochs, G., Shoichet, B., Ott, M., Palmarini, M., Shokat, K. M., García-Sastre, A., Rassen, J. A., Grosse, R., Rosenberg, O. S., Verba, K. A., Basler, C. F., Vignuzzi, M., Peden, A. A., Beltrao, P., and Krogan, N. J. (2020). "Comparative host-

coronavirus protein interaction networks reveal pan-viral disease mechanisms.” *Science*, 370(6521).

Griffin, M. J., and Brown, G. M. (1964). “THE BIOSYNTHESIS OF FOLIC ACID. III. ENZYMATIC FORMATION OF DIHYDROFOLIC ACID FROM DIHYDROPTEROIC ACID AND OF TETRAHYDROPTEROYLPOLYGLUTAMIC ACID COMPOUNDS FROM TETRAHYDROFOLIC ACID.” *J. Biol. Chem.*, 239, 310–316.

Griffith, E. C., Wallace, M. J., Wu, Y., Kumar, G., Gajewski, S., Jackson, P., Phelps, G. A., Zheng, Z., Rock, C. O., Lee, R. E., and White, S. W. (2018). “The Structural and Functional Basis for Recurring Sulfa Drug Resistance Mutations in *Staphylococcus aureus* Dihydropteroate Synthase.” *Front Microbiol*, 9.

Griffith, J. S. (1967). “Nature of the Scrapie Agent: Self-replication and Scrapie.” *Nature*, 215(5105), 1043–1044.

Grifoni, A., Sidney, J., Zhang, Y., Scheuermann, R. H., Peters, B., and Sette, A. (2020). “A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2.” *Cell Host & Microbe*, 27(4), 671-680.e2.

Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D. C., and Jahn, D. (2005). “JCat: a novel tool to adapt codon usage of a target gene to its potential expression host.” *Nucleic Acids Res*, 33(suppl_2), W526–W531.

Gui, M., Song, W., Zhou, H., Xu, J., Chen, S., Xiang, Y., and Wang, X. (2017). “Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding.” *Cell Research*, 27(1), 119–129.

Guillot-Sestier, M.-V., Sunyach, C., Druon, C., Scarzello, S., and Checler, F. (2009). “The alpha-secretase-derived N-terminal product of cellular prion, N1, displays neuroprotective function in vitro and in vivo.” *J Biol Chem*, 284(51), 35973–35986.

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.*, 8008–8016.

Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., and Banks, J. L. (2004). “Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening.” *J. Med. Chem.*, 47(7), 1750–1759.

Hammoudeh, D. I., Zhao, Y., White, S. W., and Lee, R. E. (2013). “Replacing sulfa drugs with novel DHPS inhibitors.” *Future Med Chem*, 5(11).

Hampele, I. C., D’Arcy, A., Dale, G. E., Kostrewa, D., Nielsen, J., Oefner, C., Page, M. G. P., Schönfeld, H.-J., Stüber, D., and Then, R. L. (1997). “Structure and function of the dihydropteroate synthase from *staphylococcus aureus*11Edited by R. Huber.” *Journal of Molecular Biology*, 268(1), 21–30.

Hardy, J., and Allsop, D. (1991). “Amyloid deposition as the central event in the aetiology of Alzheimer’s disease.” *Trends Pharmacol Sci*, 12(10), 383–388.

Haste Andersen, P., Nielsen, M., and Lund, O. (2006). “Prediction of residues in discontinuous B-cell epitopes using protein 3D structures.” *Protein Sci.*, 15(11), 2558–2567.

- He, S., Zhao, J., Elfalleh, W., Jemaà, M., Sun, H., Sun, X., Tang, M., He, Q., Wu, Z., and Lang, F. (2018). “In Silico Identification and in Vitro Analysis of B and T-Cell Epitopes of the Black Turtle Bean (*Phaseolus Vulgaris* L.) Lectin.” *CPB*, 49(4), 1600–1614.
- Henzler-Wildman, K., and Kern, D. (2007). “Dynamic personalities of proteins.” *Nature*, 450(7172), 964–972.
- Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997a). “LINCS: A linear constraint solver for molecular simulations.” *Journal of Computational Chemistry*, 18(12), 1463–1472.
- Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997b). “LINCS: A linear constraint solver for molecular simulations.” *Journal of Computational Chemistry*, 18(12), 1463–1472.
- Hevener, K. E., Yun, M.-K., Qi, J., Kerr, I. D., Babaoglu, K., Hurdle, J. G., Balakrishna, K., White, S. W., and Lee, R. E. (2010a). “Structural Studies of Pterin-Based Inhibitors of Dihydropteroate Synthase.” *J Med Chem*, 53(1), 166–177.
- Hevener, K. E., Yun, M.-K., Qi, J., Kerr, I. D., Babaoglu, K., Hurdle, J. G., Balakrishna, K., White, S. W., and Lee, R. E. (2010b). “Structural studies of pterin-based inhibitors of dihydropteroate synthase.” *J Med Chem*, 53(1), 166–177.
- Ho, J. M.-W., and Juurlink, D. N. (2011). “Considerations when prescribing trimethoprim-sulfamethoxazole.” *CMAJ*, 183(16), 1851–1858.
- Hoekman, D. (1996). “Exploring QSAR Fundamentals and Applications in Chemistry and Biology, Volume 1. Hydrophobic, Electronic and Steric Constants, Volume 2 J. Am. Chem. Soc. 1995, 117, 9782.” *J. Am. Chem. Soc.*, 118(43), 10678–10678.
- Honig, B., and Nicholls, A. (1995). “Classical electrostatics in biology and chemistry.” *Science*, 268(5214), 1144–1149.
- Horiuchi, M., and Caughey, B. (1999). “Prion protein interconversions and the transmissible spongiform encephalopathies.” *Structure*, 7(10), R231–R240.
- Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006a). “Comparison of multiple AMBER force fields and development of improved protein backbone parameters.” *Proteins*, 65(3), 712–725.
- Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006b). “Comparison of multiple Amber force fields and development of improved protein backbone parameters.” *Proteins: Structure, Function, and Bioinformatics*, 65(3), 712–725.
- Hovmöller, S., Zhou, T., and Ohlson, T. (2002). “Conformations of amino acids in proteins.” *Acta Cryst D*, 58(5), 768–776.
- Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., Groot, B. L. de, Grubmüller, H., and Alexander D. MacKerell, J. (2017a). “CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins.” *Nature methods*, 14(1), 71.
- Huang, M., Huang, W., Wen, F., and Larson, R. G. (2017b). “Efficient estimation of binding free energies between peptides and an MHC class II molecule using coarse-grained molecular dynamics simulations with a weighted histogram analysis method.” *Journal of Computational Chemistry*, 38(23), 2007–2019.
- Huang, W., Lin, Z., and Gunsteren, W. F. van. (2011). “Validation of the GROMOS 54A7 Force Field with Respect to β -Peptide Folding.” *J. Chem. Theory Comput.*, 7(5), 1237–1243.

Hundal, J., Carreno, B. M., Petti, A. A., Linette, G. P., Griffith, O. L., Mardis, E. R., and Griffith, M. (2016). “pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens.” *Genome Medicine*, 8(1), 11.

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.” *Journal of Molecular Graphics and Modelling*, 28(4), 371–381.

Jaidhan, B. J., Rao, P. S., and Apparao, A. (2014). “Energy Minimization and Conformation Analysis of Molecules using Steepest Descent Method.” 5, 4.

Jain, S., and Baranwal, M. (2019). “Computational analysis in designing T cell epitopes enriched peptides of Ebola glycoprotein exhibiting strong binding interaction with HLA molecules.” *Journal of Theoretical Biology*, 465, 34–44.

Jang, H., Connelly, L., Teran Arce, F., Ramachandran, S., Kagan, B. L., Lal, R., and Nussinov, R. (2013). “Mechanisms for the Insertion of Toxic, Fibril-like β -Amyloid Oligomers into the Membrane.” *J. Chem. Theory Comput.*, 9(1), 822–833.

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.

Jarrett, J. T., and Lansbury, P. T. (1993). “Seeding ‘one-dimensional crystallization’ of amyloid: a pathogenic mechanism in Alzheimer’s disease and scrapie?” *Cell*, 73(6), 1055–1058.

Jeffrey, M., Goodsir, C. M., Bruce, M. E., McBride, P. A., and Farquhar, C. (1994). “Morphogenesis of amyloid plaques in 87V murine scrapie.” *Neuropathol Appl Neurobiol*, 20(6), 535–542.

Jorgensen, W. L., and Duffy, E. M. (2000). “Prediction of drug solubility from Monte Carlo simulations.” *Bioorganic & Medicinal Chemistry Letters*, 10(11), 1155–1158.

Jorgensen, W. L., and Ravimohan, C. (1998). “Monte Carlo simulation of differences in free energies of hydration.” *The Journal of Chemical Physics*, 83(6), 3050.

Jorgensen, W. L., Ruiz-Caro, J., Tirado-Rives, J., Basavapathruni, A., Anderson, K. S., and Hamilton, A. D. (2006). “Computer-aided design of non-nucleoside inhibitors of HIV-1 reverse transcriptase.” *Bioorg Med Chem Lett*, 16(3), 663–667.

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.

Jucker, M., and Walker, L. C. (2013). “Self-propagation of pathogenic protein aggregates in neurodegenerative diseases.” *Nature*, 501(7465), 45–51.

Kalipillai, P., and Mani, E. (2021). “Adsorption of the amyloid β 40 monomer on charged gold nanoparticles and slabs: a molecular dynamics study.” *Phys. Chem. Chem. Phys.*, 23(34), 18618–18627.

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.

Kar, T., Narsaria, U., Basak, S., Deb, D., Castiglione, F., Mueller, D. M., and Srivastava, A. P. (2020). “A candidate multi-epitope vaccine against SARS-CoV-2.” *Sci Rep*, 10(1), 10895.

- Karplus, P. A., and Schulz, G. E. (1985a). "Prediction of chain flexibility in proteins: A tool for the selection of peptide antigens." *Naturwissenschaften*, 72(4), 212–213.
- Karplus, P. A., and Schulz, G. E. (1985b). "Prediction of chain flexibility in proteins." *Naturwissenschaften*, 72(4), 212–213.
- Kästner, J. (2011a). "Umbrella sampling." *WIREs Computational Molecular Science*, 1(6), 932–942.
- Kästner, J. (2011b). "Umbrella sampling." *WIREs Computational Molecular Science*, 1(6), 932–942.
- Katyal, N., and Deep, S. (2017). "Inhibition of GNNQQNY prion peptide aggregation by trehalose: a mechanistic view." *Phys. Chem. Chem. Phys.*, 19(29), 19120–19138.
- Keränen, H., Gutiérrez-de-Terán, H., and Åqvist, J. (2014). "Structural and Energetic Effects of A2A Adenosine Receptor Mutations on Agonist and Antagonist Binding." *PLOS ONE*, 9(10), e108492.
- Kirchmair, J., Markt, P., Distinto, S., Wolber, G., and Langer, T. (2008). "Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—What can we learn from earlier mistakes?" *Journal of Computer-Aided Molecular Design*, 22(3–4), 213–228.
- Kolaskar, A. S., and Tongaonkar, P. C. (1990a). "A semi-empirical method for prediction of antigenic determinants on protein antigens." *FEBS Lett.*, 276(1–2), 172–174.
- Kolaskar, A. S., and Tongaonkar, P. C. (1990b). "A semi-empirical method for prediction of antigenic determinants on protein antigens." *FEBS Lett.*, 276(1–2), 172–174.
- 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.
- Koneru, J. K., Sinha, S., and Mondal, J. (2019). "In Silico Reoptimization of Binding Affinity and Drug-Resistance Circumvention Ability in Kinase Inhibitors: A Case Study with RL-45 and Src Kinase." *J. Phys. Chem. B*, 123(31), 6664–6672.
- Kramer, B., Rarey, M., and Lengauer, T. (1999). "Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking." *Proteins: Structure, Function, and Bioinformatics*, 37(2), 228–241.
- Kringelum, J. V., Lundegaard, C., Lund, O., and Nielsen, M. (2012). "Reliable B Cell Epitope Predictions: Impacts of Method Development and Improved Benchmarking." *PLOS Computational Biology*, 8(12), e1002829.
- Kuhn, B., and Kollman, P. A. (2000). "Binding of a Diverse Set of Ligands to Avidin and Streptavidin: An Accurate Quantitative Prediction of Their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models." *J. Med. Chem.*, 43(20), 3786–3791.
- Kumar, A., Yadav, I. S., Hussain, S., Das, B. C., and Bharadwaj, M. (2015). "Identification of immunotherapeutic epitope of E5 protein of human papillomavirus-16: An in silico approach." *Biologicals*, 43(5), 344–348.

Kumar, K., M. Woo, S., Siu, T., A. Cortopassi, W., Duarte, F., and S. Paton, R. (2018). “Cation– π interactions in protein–ligand binding: theory and data-mining reveal different roles for lysine and arginine.” *Chemical Science*, 9(10), 2655–2665.

Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., and Kollman, P. A. (1992a). “THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method.” *Journal of Computational Chemistry*, 13(8), 1011–1021.

Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., and Kollman, P. A. (1992b). “THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method.” *Journal of Computational Chemistry*, 13(8), 1011–1021.

Kumar, S., Stecher, G., and Tamura, K. (2016). “MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.” *Molecular Biology and Evolution*, 33(7), 1870–1874.

Kumar, S., Tamura, K., Jakobsen, I. B., and Nei, M. (2001). “MEGA2: molecular evolutionary genetics analysis software.” *Bioinformatics*, 17(12), 1244–1245.

Kumari, R., Kumar, R., and Lynn, A. (2014a). “g_mmpbsa—A GROMACS Tool for High-Throughput MM-PBSA Calculations.” *J. Chem. Inf. Model.*, 54(7), 1951–1962.

Kumari, R., Kumar, R., and Lynn, A. (2014b). “g_mmpbsa—A GROMACS Tool for High-Throughput MM-PBSA Calculations.” *J. Chem. Inf. Model.*, 54(7), 1951–1962.

Laganowsky, A., Liu, C., Sawaya, M. R., Whitelegge, J. P., Park, J., Zhao, M., Pensalfini, A., Soriaga, A. B., Landau, M., Teng, P. K., Cascio, D., Glabe, C., and Eisenberg, D. (2012). “Atomic view of a toxic amyloid small oligomer.” *Science*, 335(6073), 1228–1231.

Lam, T. T.-Y., Shum, M. H.-H., Zhu, H.-C., Tong, Y.-G., Ni, X.-B., Liao, Y.-S., Wei, W., Cheung, W. Y.-M., Li, W.-J., Li, L.-F., Leung, G. M., Holmes, E. C., Hu, Y.-L., and Guan, Y. (2020). *Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China*. Microbiology.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007). “Clustal W and Clustal X version 2.0.” *Bioinformatics*, 23(21), 2947–2948.

Larsen, J. E. P., Lund, O., and Nielsen, M. (2006). “Improved method for predicting linear B-cell epitopes.” *Immunome Res*, 2, 2.

Larson, M. E., and Lesné, S. E. (2012). “Soluble A β oligomer production and toxicity.” *J Neurochem*, 120(Suppl 1), 125–139.

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

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

Lee, M. R., Duan, Y., and Kollman, P. A. (2000). “Use of MM-PB/SA in estimating the free energies of proteins: application to native, intermediates, and unfolded villin headpiece.” *Proteins*, 39(4), 309–316.

Lemkul, J. A., and Bevan, D. R. (2010a). “Assessing the Stability of Alzheimer’s Amyloid Protofibrils Using Molecular Dynamics.” *J. Phys. Chem. B*, 114(4), 1652–1660.

- Lemkul, J. A., and Bevan, D. R. (2010b). "Assessing the Stability of Alzheimer's Amyloid Protofibrils Using Molecular Dynamics." *J. Phys. Chem. B*, 114(4), 1652–1660.
- Lemkul, J. A., and Bevan, D. R. (2010c). "Assessing the Stability of Alzheimer's Amyloid Protofibrils Using Molecular Dynamics." *J. Phys. Chem. B*, 114(4), 1652–1660.
- Lenselink, E. B., Louvel, J., Forti, A. F., Veldhoven, J. P. D. van, Vries, H. de, Mulder-Krieger, T., McRobb, F. M., Negri, A., Goose, J., Abel, R., Vlijmen, H. W. T. van, Wang, L., Harder, E., Sherman, W., IJzerman, A. P., and Beuming, T. (2016). "Predicting Binding Affinities for GPCR Ligands Using Free-Energy Perturbation." *ACS Omega*, 1(2), 293–304.
- Leonard, J. T., and Roy, K. (2006a). "On Selection of Training and Test Sets for the Development of Predictive QSAR models." *QSAR & Combinatorial Science*, 25(3), 235–251.
- Leonard, J. T., and Roy, K. (2006b). "On Selection of Training and Test Sets for the Development of Predictive QSAR models." *QSAR & Combinatorial Science*, 25(3), 235–251.
- Lesné, S., Koh, M. T., Kotilinek, L., Kaye, R., Glabe, C. G., Yang, A., Gallagher, M., and Ashe, K. H. (2006). "A specific amyloid- β protein assembly in the brain impairs memory." *Nature*, 440(7082), 352.
- Levin, I., Giladi, M., Altman-Price, N., Ortenberg, R., and Mevarech, M. (2004). "An alternative pathway for reduced folate biosynthesis in bacteria and halophilic archaea." *Molecular Microbiology*, 54(5), 1307–1318.
- Levy, C., Minnis, D., and Derrick, J. P. (2008). "Dihydropteroate synthase from *Streptococcus pneumoniae*: structure, ligand recognition and mechanism of sulfonamide resistance." *Biochem. J.*, 412(2), 379–388.
- Lewis, V., and Hooper, N. M. (2011). "The role of lipid rafts in prion protein biology." *Front Biosci (Landmark Ed)*, 16(1), 151–168.
- Li, F. (2016). "Structure, Function, and Evolution of Coronavirus Spike Proteins." *Annu. Rev. Virol.*, 3(1), 237–261.
- Lima, M. P. M., Nader, M., Santos, D. E. S., and Soares, T. A. (2019). "Compatibility of GROMOS-Derived Atomic Parameters for Lipopolysaccharide Membranes with the SPC/E Water Model and Alternative Long-Range Electrostatic Treatments Using Single Nonbonded Cutoff and Atom-Based Charge Schemes." *J. Braz. Chem. Soc.*, 30, 2219–2230.
- Lin, L., Ting, S., Yufei, H., Wendong, L., Yubo, F., and Jing, Z. (2020). "Epitope-based peptide vaccines predicted against novel coronavirus disease caused by SARS-CoV-2." *Virus Research*, 288, 198082.
- Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010a). "Improved side-chain torsion potentials for the Amber ff99SB protein force field." *Proteins*, 78(8), 1950–1958.
- Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010b). "Improved side-chain torsion potentials for the Amber ff99SB protein force field." *Proteins*, 78(8), 1950–1958.
- Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (1997). "Experimental and computational approaches to estimate solubility and permeability in drug discovery

and development settings.” *Advanced Drug Delivery Reviews*, In Vitro Models for Selection of Development Candidates, 23(1), 3–25.

Liu, P., Kim, B., Friesner, R. A., and Berne, B. J. (2005). “Replica exchange with solute tempering: A method for sampling biological systems in explicit water.” *PNAS*, 102(39), 13749–13754.

Liu, W. J., Zhao, M., Liu, K., Xu, K., Wong, G., Tan, W., and Gao, G. F. (2017). “T-cell immunity of SARS-CoV: Implications for vaccine development against MERS-CoV.” *Antiviral Research*, 137, 82–92.

Lizbeth, R.-S. G., Jazmín, G.-M., José, C.-B., and Marlet, M.-A. (2020). “Immunoinformatics study to search epitopes of spike glycoprotein from SARS-CoV-2 as potential vaccine.” *J. Biomol. Struct. Dyn.*, 1–15.

Llorach-Pares, L., Nonell-Canals, A., Avila, C., and Sanchez-Martinez, M. (2018). “Kororamides, Convolutamines, and Indole Derivatives as Possible Tau and Dual-Specificity Kinase Inhibitors for Alzheimer’s Disease: A Computational Study.” *Mar Drugs*, 16(10).

Lomakin, A., Chung, D. S., Benedek, G. B., Kirschner, D. A., and Teplow, D. B. (1996). “On the nucleation and growth of amyloid beta-protein fibrils: detection of nuclei and quantitation of rate constants.” *Proc. Natl. Acad. Sci. U.S.A.*, 93(3), 1125–1129.

Lopez de Compadre, R. L., Pearlstein, R. A., Hopfinger, A. J., and Seydel, J. K. (1987). “A quantitative structure-activity relationship analysis of some 4-aminodiphenyl sulfone antibacterial agents using linear free energy and molecular modeling methods.” *J. Med. Chem.*, 30(5), 900–906.

Lu, X., Lv, M., Huang, K., Ding, K., and You, Q. (2012). “Pharmacophore and molecular docking guided 3D-QSAR study of bacterial enoyl-ACP reductase (FabI) Inhibitors.” *Int J Mol Sci*, 13(6), 6620–6638.

Luiken, J. A., and Bolhuis, P. G. (2015). “Primary Nucleation Kinetics of Short Fibril-Forming Amyloidogenic Peptides.” *J. Phys. Chem. B*, 119(39), 12568–12579.

Ma, B., Kumar, S., Tsai, C.-J., and Nussinov, R. (1999). “Folding funnels and binding mechanisms.” *Protein Engineering, Design and Selection*, 12(9), 713–720.

Ma, J. C., and Dougherty, D. A. (1997). “The Cation- π Interaction.” *Chem. Rev.*, 97(5), 1303–1324.

Malkhasian, A. Y. S., and Howlin, B. J. (2016). “Docking and DFT studies on ligand binding to Quercetin 2,3-dioxygenase.” *Journal of Biomolecular Structure and Dynamics*, 34(11), 2453–2461.

Man, V. H., He, X., Derreumaux, P., Ji, B., Xie, X.-Q., Nguyen, P. H., and Wang, J. (2019a). “Effects of All-Atom Molecular Mechanics Force Fields on Amyloid Peptide Assembly: The Case of A β 16–22 Dimer.” *J. Chem. Theory Comput.*, 15(2), 1440–1452.

Man, V. H., He, X., Derreumaux, P., Ji, B., Xie, X.-Q., Nguyen, P. H., and Wang, J. (2019b). “Effects of All-Atom Molecular Mechanics Force Fields on Amyloid Peptide Assembly: The Case of A β 16–22 Dimer.” *J Chem Theory Comput*, 15(2), 1440–1452.

Manolova, V., Flace, A., Bauer, M., Schwarz, K., Saudan, P., and Bachmann, M. F. (2008). “Nanoparticles target distinct dendritic cell populations according to their size.” *European Journal of Immunology*, 38(5), 1404–1413.

Mark, P., and Nilsson, L. (2001a). “Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K.” *J. Phys. Chem. A*, 105(43), 9954–9960.

Mark, P., and Nilsson, L. (2001b). "Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K." *J. Phys. Chem. A*, 105(43), 9954–9960.

Mark, P., and Nilsson, L. (2001c). "Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K." *J. Phys. Chem. A*, 105(43), 9954–9960.

Martin, T. M., Harten, P., Young, D. M., Muratov, E. N., Golbraikh, A., Zhu, H., and Tropsha, A. (2012a). "Does Rational Selection of Training and Test Sets Improve the Outcome of QSAR Modeling?" *J. Chem. Inf. Model.*, 52(10), 2570–2578.

Martin, T. M., Harten, P., Young, D. M., Muratov, E. N., Golbraikh, A., Zhu, H., and Tropsha, A. (2012b). "Does rational selection of training and test sets improve the outcome of QSAR modeling?" *J Chem Inf Model*, 52(10), 2570–2578.

Martinez, J. C. G., Andrada, M. F., Vega-Hissi, E. G., Garibotto, F. M., Nogueras, M., Rodríguez, R., Cobo, J., Enriz, R. D., and Estrada, M. R. (2016). "Dihydrofolate reductase inhibitors: a quantitative structure–activity relationship study using 2D-QSAR and 3D-QSAR methods." *Medicinal Chemistry Research*, 1(26), 247–261.

Martínez, L., Andrade, R., Birgin, E. G., and Martínez, J. M. (2009). "PACKMOL: A package for building initial configurations for molecular dynamics simulations." *Journal of Computational Chemistry*, 30(13), 2157–2164.

Martyna, G. J., Klein, M. L., and Tuckerman, M. (1992). "Nosé–Hoover chains: The canonical ensemble via continuous dynamics." *The Journal of Chemical Physics*, 97(4), 2635–2643.

Martyna, G. J., Tobias, D. J., and Klein, M. L. (1994). "Constant pressure molecular dynamics algorithms." *The Journal of Chemical Physics*, 101(5), 4177–4189.

Matherly, L. H., and Goldman, D. I. (2003). "Membrane transport of folates." *Vitam. Horm.*, 66, 403–456.

Matysiak, J. (2007). "Evaluation of electronic, lipophilic and membrane affinity effects on antiproliferative activity of 5-substituted-2-(2,4-dihydroxyphenyl)-1,3,4-thiadiazoles against various human cancer cells." *European Journal of Medicinal Chemistry*, 42(7), 940–947.

Meirovitch, H. (2007). "Recent developments in methodologies for calculating the entropy and free energy of biological systems by computer simulation." *Current Opinion in Structural Biology*, 17(2), 181–186.

Meng, Y., and Roux, B. (2015). "Efficient Determination of Free Energy Landscapes in Multiple Dimensions from Biased Umbrella Sampling Simulations Using Linear Regression." *J. Chem. Theory Comput.*, 11(8), 3523–3529.

Meza, J. C. (2010). "Steepest descent." *WIREs Computational Statistics*, 2(6), 719–722.

Miao, Y., Huang, Y. M., Walker, R. C., McCammon, J. A., and Chang, C. A. (2018). "Ligand Binding Pathways and Conformational Transitions of the HIV Protease." *Biochemistry*, 57(9), 1533–1541.

Michaels, T. C. T., Šarić, A., Habchi, J., Chia, S., Meisl, G., Vendruscolo, M., Dobson, C. M., and Knowles, T. P. J. (2018). "Chemical Kinetics for Bridging Molecular Mechanisms and Macroscopic Measurements of Amyloid Fibril Formation." *Annual Review of Physical Chemistry*, 69(1), 273–298.

Mishra, S. S., Ranjan, S., Sharma, C. S., Singh, H. P., Kalra, S., and Kumar, N. (2020). "Computational investigation of potential inhibitors of novel coronavirus 2019 through structure-based virtual screening, molecular dynamics and density functional theory studies." *Journal of Biomolecular Structure and Dynamics*, 0(0), 1–13.

- Miyamoto, S., and Kollman, P. A. (1992a). "SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models." *J. Comput. Chem.*, 13(8), 952–962.
- Miyamoto, S., and Kollman, P. A. (1992b). "SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models." *J. Comput. Chem.*, 13(8), 952–962.
- Mondal, S., Mandal, S. M., Mondal, T. K., and Sinha, C. (2015). "Structural characterization of new Schiff bases of sulfamethoxazole and sulfathiazole, their antibacterial activity and docking computation with DHPS protein structure." *Spectrochim Acta A Mol Biomol Spectrosc*, 150, 268–279.
- Morgan, R. E., Batot, G. O., Dement, J. M., Rao, V. A., Eadsforth, T. C., and Hunter, W. N. (2011). "Crystal structures of Burkholderia cenocepacia dihydropteroate synthase in the apo-form and complexed with the product 7,8-dihydropteroate." *BMC Structural Biology*, 11(1), 21.
- MORRIS, G. M., GOODSELL, D. S., HALLIDAY, R. S., HUEY, R., HART, W. E., BELEW, R. K., and OLSON, A. J. (n.d.). "Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function." *JOURNAL OF COMPUTATIONAL CHEMISTRY*, 19(14), 24.
- 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.
- Moučka, F., Nezbeda, I., and Smith, W. R. (2013). "Molecular force fields for aqueous electrolytes: SPC/E-compatible charged LJ sphere models and their limitations." *J. Chem. Phys.*, 138(15), 154102.
- 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.
- Nazar, A., Abbas, G., and Azam, S. S. (2020). "Deciphering the Inhibition Mechanism of under Trial Hsp90 Inhibitors and Their Analogues: A Comparative Molecular Dynamics Simulation." *J. Chem. Inf. Model.*, 60(8), 3812–3830.
- Nelapati, A. K., Das, B. K., Ponnann Ettiappan, J. B., and Chakraborty, D. (2020). "In-silico epitope identification and design of Uricase mutein with reduced immunogenicity." *Process Biochemistry*, 92, 288–302.
- Nelson, R., and Eisenberg, D. (2006). "Recent atomic models of amyloid fibril structure." *Current Opinion in Structural Biology, Theory and simulation/Macromolecular assemblages*, 16(2), 260–265.
- Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. Ø., Riek, C., Grothe, R., and Eisenberg, D. (2005a). "Structure of the cross- β spine of amyloid-like fibrils." *Nature*, 435(7043), 773–778.
- Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. Ø., Riek, C., Grothe, R., and Eisenberg, D. (2005b). "Structure of the cross- β spine of amyloid-like fibrils." *Nature*, 435(7043), 773–778.
- Nguyen, T. T., Viet, M. H., and Li, M. S. (2014). "Effects of Water Models on Binding Affinity: Evidence from All-Atom Simulation of Binding of Tamiflu to A/H5N1 Neuraminidase." *The Scientific World Journal*, 2014, e536084.

- Nie, R., Huo, Y., Yu, B., Liu, C., Zhou, R., Bao, H., and Tang, S. (2020). "Molecular insights into the inhibitory mechanisms of gallate moiety on the A β 1–40 amyloid aggregation: A molecular dynamics simulation study." *International Journal of Biological Macromolecules*, 156, 40–50.
- Nielsen, M., and Andreatta, M. (2016). "NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets." *Genome Medicine*, 8(1), 33.
- Nolte, R. T., Wisely, G. B., Westin, S., Cobb, J. E., Lambert, M. H., Kurokawa, R., Rosenfeld, M. G., Willson, T. M., Glass, C. K., and Milburn, M. V. (1998). "Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor- γ ." *Nature*, 395(6698), 137–143.
- Oosawa, F., and Kasai, M. (1962). "A theory of linear and helical aggregations of macromolecules." *Journal of Molecular Biology*, 4(1), 10–21.
- Otsuka Takao, Okimoto Noriaki, and Taiji Makoto. (2015). "Assessment and acceleration of binding energy calculations for protein–ligand complexes by the fragment molecular orbital method." *Journal of Computational Chemistry*, 36(30), 2209–2218.
- Pal, S., Kumar, V., Kundu, B., Bhattacharya, D., Preethy, N., Reddy, M. P., and Talukdar, A. (2019). "Ligand-based Pharmacophore Modeling, Virtual Screening and Molecular Docking Studies for Discovery of Potential Topoisomerase I Inhibitors." *Computational and Structural Biotechnology Journal*, 17, 291–310.
- Pal, S., and Paul, S. (2020). "ATP Controls the Aggregation of A β 16–22 Peptides." *J. Phys. Chem. B*, 124(1), 210–223.
- Paliwal, H., and Shirts, M. R. (2011). "A Benchmark Test Set for Alchemical Free Energy Transformations and Its Use to Quantify Error in Common Free Energy Methods." *J. Chem. Theory Comput.*, 7(12), 4115–4134.
- Papaleo, E., Mereghetti, P., Fantucci, P., Grandori, R., and De Gioia, L. (2009). "Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: The myoglobin case." *Journal of Molecular Graphics and Modelling*, 27(8), 889–899.
- Parenti, M. D., and Rastelli, G. (2012). "Advances and applications of binding affinity prediction methods in drug discovery." *Biotechnology Advances, Systems Biology for Biomedical Innovation*, 30(1), 244–250.
- Parker, J. M., Guo, D., and Hodges, R. S. (1986a). "New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites." *Biochemistry*, 25(19), 5425–5432.
- Parker, J. M. R., Guo, D., and Hodges, R. S. (1986b). "New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and x-ray-derived accessible sites." *Biochemistry*, 25(19), 5425–5432.
- Parrinello, M., and Rahman, A. (1981a). "Polymorphic transitions in single crystals: A new molecular dynamics method." *Journal of Applied Physics*, 52(12), 7182–7190.
- Parrinello, M., and Rahman, A. (1981b). "Polymorphic transitions in single crystals: A new molecular dynamics method." *Journal of Applied Physics*, 52(12), 7182–7190.

- Pellequer, J.-L., Westhof, E., and Van Regenmortel, M. H. V. (1993). “Correlation between the location of antigenic sites and the prediction of turns in proteins.” *Immunology Letters*, 36(1), 83–99.
- Peng, Y., Dong, H., and Welsh, W. J. (2019a). “Comprehensive 3D-QSAR Model Predicts Binding Affinity of Structurally Diverse Sigma 1 Receptor Ligands.” *J. Chem. Inf. Model.*, 59(1), 486–497.
- Peng, Y., Dong, H., and Welsh, W. J. (2019b). “Comprehensive 3D-QSAR Model Predicts Binding Affinity of Structurally Diverse Sigma 1 Receptor Ligands.” *J. Chem. Inf. Model.*, 59(1), 486–497.
- Peters, B., Sidney, J., Bourne, P., Bui, H.-H., Buus, S., Doh, G., Fleri, W., Kronenberg, M., Kubo, R., Lund, O., Nemazee, D., Ponomarenko, J. V., Sathiamurthy, M., Schoenberger, S., Stewart, S., Surko, P., Way, S., Wilson, S., and Sette, A. (2005). “The Immune Epitope Database and Analysis Resource: From Vision to Blueprint.” *PLOS Biology*, 3(3), e91.
- Pike, C. J., Walencewicz, A. J., Glabe, C. G., and Cotman, C. W. (1991). “In vitro aging of beta-amyloid protein causes peptide aggregation and neurotoxicity.” *Brain Res*, 563(1–2), 311–314.
- Pohorille, A., Jarzynski, C., and Chipot, C. (2010). “Good Practices in Free-Energy Calculations.” *J. Phys. Chem. B*, 114(32), 10235–10253.
- 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 pKa Values of Benzoic and Alkanoic Acids.” *J. Chem. Inf. Comput. Sci.*, 42(2), 184–191.
- Promptchara, E., Ketloy, C., and Palaga, T. (2020). “Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic.” *Asian Pac. J. Allergy Immunol.*, 38(1), 1–9.
- Prusiner, S. B. (1998). “Prions.” *PNAS*, 95(23), 13363–13383.
- Punt, J., Owen, J. A., Stranford, S. A., Jones, P. P., and Kuby, J. (2019). *Kuby immunology*.
- Puzyn, T., Leszczynski, J., and Cronin, M. T. (Eds.). (2010). *Recent Advances in QSAR Studies: Methods and Applications*. Challenges and Advances in Computational Chemistry and Physics, Springer Netherlands.
- Qi, J., Virga, K. G., Das, S., Zhao, Y., Yun, M.-K., White, S. W., and Lee, R. E. (2011). “Synthesis of bi-substrate state mimics of dihydropteroate synthase as potential inhibitors and molecular probes.” *Bioorganic & Medicinal Chemistry, Imaging Probes*, 19(3), 1298–1305.
- Richey, D. P., and Brown, G. M. (1969). “The biosynthesis of folic acid. IX. Purification and properties of the enzymes required for the formation of dihydropteroic acid.” *J. Biol. Chem.*, 244(6), 1582–1592.
- Rossi, M., Amaretti, A., and Raimondi, S. (2011). “Folate Production by Probiotic Bacteria.” *Nutrients*, 3(1), 118–134.
- Roucou, X., Gains, M., and LeBlanc, A. C. (2004). “Neuroprotective functions of prion protein.” *J Neurosci Res*, 75(2), 153–161.
- Roy, R., and Paul, S. (2020). “Theoretical Investigation of the Inhibitory Mechanism of Norepinephrine on hIAPP Amyloid Aggregation and the Destabilization of Protofibrils.” *J. Phys. Chem. B*, 124(48), 10913–10929.

Roy, R., and Paul, S. (2021). “Potential of ATP toward Prevention of hIAPP Oligomerization and Destabilization of hIAPP Protofibrils: An In Silico Perspective.” *J. Phys. Chem. B*, 125(14), 3510–3526.

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*, 8(1), e51429.

Samantray, S., Yin, F., Kav, B., and Strodel, B. (2020). “Different Force Fields Give Rise to Different Amyloid Aggregation Pathways in Molecular Dynamics Simulations.” *J. Chem. Inf. Model.*, 60(12), 6462–6475.

Sanche, S., Lin, Y. T., Xu, C., Romero-Severson, E., Hengartner, N., and Ke, R. (2020). *The Novel Coronavirus, 2019-nCoV, is Highly Contagious and More Infectious Than Initially Estimated*. *Epidemiology*.

Sánchez-Osuna, M., Cortés, P., Barbé, J., and Erill, I. (2019). “Origin of the Mobile Dihydro-Pterate Synthase Gene Determining Sulfonamide Resistance in Clinical Isolates.” *Front Microbiol*, 9.

Santuccione, A., Sytnyk, V., Leshchyn'ska, I., and Schachner, M. (2005). “Prion protein recruits its neuronal receptor NCAM to lipid rafts to activate p59fyn and to enhance neurite outgrowth.” *J Cell Biol*, 169(2), 341–354.

Sarkar, P., Das, B. K., Chakraborty, D., and Muthamma, K. (2019). “Carbohelicenes and thiahelicene from phthalaldehydes through Perkin approach.” *Journal of Molecular Structure*, 1195, 309–314.

Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013). “Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments.” *J. Comput. Aided Mol. Des.*, 27(3), 221–234.

Schindewolf, C., and Menachery, V. D. (2019). “Middle East Respiratory Syndrome Vaccine Candidates: Cautious Optimism.” *Viruses*, 11(1).

Schmid, N., Eichenberger, A. P., Choutko, A., Riniker, S., Winger, M., Mark, A. E., and Gunsteren, W. F. van. (2011a). “Definition and testing of the GROMOS force-field versions 54A7 and 54B7.” *Eur Biophys J*, 40(7), 843.

Schmid, N., Eichenberger, A. P., Choutko, A., Riniker, S., Winger, M., Mark, A. E., and Gunsteren, W. F. van. (2011b). “Definition and testing of the GROMOS force-field versions 54A7 and 54B7.” *Eur Biophys J*, 40(7), 843–856.

Schüttelkopf, A. W., and Aalten, D. M. F. van. (2004). “PRODRG: a tool for high-throughput crystallography of protein–ligand complexes.” *Acta Cryst D*, 60(8), 1355–1363.

Scrutton, N. S., and Raine, A. R. (1996). “Cation-pi bonding and amino-aromatic interactions in the biomolecular recognition of substituted ammonium ligands.” *Biochem J*, 319 (Pt 1), 1–8.

Seabrook, S. A., and Newman, J. (2013). “High-Throughput Thermal Scanning for Protein Stability: Making a Good Technique More Robust.” *ACS Comb. Sci.*, 15(8), 387–392.

Serio, T. R., Cashikar, A. G., Kowal, A. S., Sawicki, G. J., Moslehi, J. J., Serpell, L., Arnsdorf, M. F., and Lindquist, S. L. (2000). “Nucleated conformational conversion and the replication of conformational information by a prion determinant.” *Science*, 289(5483), 1317–1321.

Shah, U. A., Deokar, H. S., Kadam, S. S., and Kulkarni, V. M. (2010). "Pharmacophore generation and atom-based 3D-QSAR of novel 2-(4-methylsulfonylphenyl)pyrimidines as COX-2 inhibitors." *Mol Divers*, 14(3), 559–568.

Shahzad, S., Qadir, M. A., Ahmed, M., Ahmad, S., Khan, M. J., Gulzar, A., and Muddassar, M. (2020). "Folic acid-sulfonamide conjugates as antibacterial agents: design, synthesis and molecular docking studies." *RSC Adv.*, 10(70), 42983–42992.

Shamshirgar, D. S., Hess, B., and Tornberg, A.-K. (2017). "A comparison of the Spectral Ewald and Smooth Particle Mesh Ewald methods in GROMACS." *arXiv:1712.04718 [math]*.

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., and Li, F. (2020). "Structural basis of receptor recognition by SARS-CoV-2." *Nature*, 581(7807), 221–224.

Sharp, P. M., and Li, W.-H. (1987). "The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications." *Nucleic Acids Research*, 15(3), 1281–1295.

Shaw, G. X., Li, Y., Shi, G., Wu, Y., Cherry, S., Needle, D., Zhang, D., Tropea, J. E., Waugh, D. S., Yan, H., and Ji, X. (2014). "Structural enzymology and inhibition of the bifunctional folate pathway enzyme HPPK-DHPS from the biowarfare agent *Francisella tularensis*." *FEBS J*, 281(18), 4123–4137.

Shi, J., Zhang, J., Li, S., Sun, J., Teng, Y., Wu, M., Li, J., Li, Y., Hu, N., Wang, H., and Hu, Y. (2015a). "Epitope-Based Vaccine Target Screening against Highly Pathogenic MERS-CoV: An In Silico Approach Applied to Emerging Infectious Diseases." *PLOS ONE*, 10(12), e0144475.

Shi, J., Zhang, J., Li, S., Sun, J., Teng, Y., Wu, M., Li, J., Li, Y., Hu, N., Wang, H., and Hu, Y. (2015b). "Epitope-Based Vaccine Target Screening against Highly Pathogenic MERS-CoV: An In Silico Approach Applied to Emerging Infectious Diseases." *PLOS ONE*, 10(12), e0144475.

Shirts, M. R., and Chodera, J. D. (2008). "Statistically optimal analysis of samples from multiple equilibrium states." *J. Chem. Phys.*, 129(12), 124105.

Shivakumar, D., Deng, Y., and Roux, B. (2009). "Computations of Absolute Solvation Free Energies of Small Molecules Using Explicit and Implicit Solvent Model." *J. Chem. Theory Comput.*, 5(4), 919–930.

Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J., and Sherman, W. (2010). "Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field." *J. Chem. Theory Comput.*, 6(5), 1509–1519.

Singh, A. (2021). "Eliciting B cell immunity against infectious diseases using nanovaccines." *Nat. Nanotechnol.*, 16(1), 16–24.

Singh, O., and Chakraborty, D. (2021). "Preferential binding affinity of ions and their effect on structure and dynamics of water near antimicrobial peptide." *Journal of Molecular Liquids*, 344, 117789.

Sköld, O. (2000). "Sulfonamide resistance: mechanisms and trends." *Drug Resistance Updates*, 3(3), 155–160.

Smit, F. X., Luiken, J. A., and Bolhuis, P. G. (2017). "Primary Fibril Nucleation of Aggregation Prone Tau Fragments PHF6 and PHF6*." *J. Phys. Chem. B*, 121(15), 3250–3261.

Song, Z., Xu, Y., Bao, L., Zhang, L., Yu, P., Qu, Y., Zhu, H., Zhao, W., Han, Y., and Qin, C. (2019). “From SARS to MERS, Thrusting Coronaviruses into the Spotlight.” *Viruses*, 11(1).

Sorout, N., and Chandra, A. (2021). “Interactions of the A β (1–42) Peptide with Boron Nitride Nanoparticles of Varying Curvature in an Aqueous Medium: Different Pathways to Inhibit β -Sheet Formation.” *J. Phys. Chem. B*, 125(40), 11159–11178.

Sousa da Silva, A. W., and Vranken, W. F. (2012a). “ACPYPE - AnteChamber PYthon Parser interface.” *BMC Research Notes*, 5(1), 367.

Sousa da Silva, A. W., and Vranken, W. F. (2012b). “ACPYPE - AnteChamber PYthon Parser interface.” *BMC Research Notes*, 5(1), 367.

Spoel, D. V. D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005). “GROMACS: Fast, flexible, and free.” *Journal of Computational Chemistry*, 26(16), 1701–1718.

Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., and Case, D. A. (1998a). “Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate–DNA Helices.” *J. Am. Chem. Soc.*, 120(37), 9401–9409.

Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., and Case, D. A. (1998b). “Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate–DNA Helices.” *J. Am. Chem. Soc.*, 120(37), 9401–9409.

Srivastava, P., and Tiwari, A. (2017). “Critical Role of Computer Simulations in Drug Discovery and Development.” *Curr Top Med Chem*, 17(21), 2422–2432.

Stickle, D. F., Presta, L. G., Dill, K. A., and Rose, G. D. (1992). “Hydrogen bonding in globular proteins.” *J Mol Biol*, 226(4), 1143–1159.

Still, W. C., Tempczyk, A., Hawley, R. C., and Hendrickson, T. (1990). “Semianalytical treatment of solvation for molecular mechanics and dynamics.” *J. Am. Chem. Soc.*, 112(16), 6127–6129.

“Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K - The Journal of Physical Chemistry A (ACS Publications).” (n.d.). <<https://pubs.acs.org/doi/10.1021/jp003020w>> (Sep. 24, 2018).

Sun, H., Tian, S., Zhou, S., Li, Y., Li, D., Xu, L., Shen, M., Pan, P., and Hou, T. (2015). “Revealing the favorable dissociation pathway of type II kinase inhibitors via enhanced sampling simulations and two-end-state calculations.” *Scientific Reports*, 5(1), 8457.

Suthar, A. B., Vitoria, M. A., Nagata, J. M., Anglaret, X., Mbori-Ngacha, D., Sued, O., Kaplan, J. E., and Doherty, M. C. (2015). “Co-trimoxazole prophylaxis in adults, including pregnant women, with HIV: a systematic review and meta-analysis.” *Lancet HIV*, 2(4), e137-150.

Swegat, W., Schlitter, J., Krüger, P., and Wollmer, A. (2003). “MD Simulation of Protein-Ligand Interaction: Formation and Dissociation of an Insulin-Phenol Complex.” *Biophysical Journal*, 84(3), 1493–1506.

Tahir ul Qamar, M., Bari, A., Adeel, M. M., Maryam, A., Ashfaq, U. A., Du, X., Muneer, I., Ahmad, H. I., and Wang, J. (2018). “Peptide vaccine against chikungunya virus: immuno-informatics combined with molecular docking approach.” *J Transl Med*, 16.

Takano, Y., and Houk, K. N. (2005). “Benchmarking the Conductor-like Polarizable Continuum Model (CPCM) for Aqueous Solvation Free Energies of Neutral and Ionic Organic Molecules.” *J. Chem. Theory Comput.*, 1(1), 70–77.

- Tan, C., Tan, Y.-H., and Luo, R. (2007). “Implicit Nonpolar Solvent Models.” *J. Phys. Chem. B*, 111(42), 12263–12274.
- Tang, X., and Han, W. (2022a). “Multiscale Exploration of Concentration-Dependent Amyloid- β (16-21) Amyloid Nucleation.” *J. Phys. Chem. Lett.*, 13(22), 5009–5016.
- Tang, X., and Han, W. (2022b). “Multiscale Exploration of Concentration-Dependent Amyloid- β (16-21) Amyloid Nucleation.” *J. Phys. Chem. Lett.*, 13(22), 5009–5016.
- Tao, P., Fisher, J. F., Shi, Q., Vreven, T., Mobashery, S., and Schlegel, H. B. (2009). “Matrix Metalloproteinase 2 (MMP2) Inhibition: QM/MM Studies of the Inhibition Mechanism of SB-3CT and its Analog.” *Biochemistry*, 48(41), 9839.
- Tari, L. W. (2012). “The Utility of Structural Biology in Drug Discovery.” *Structure-Based Drug Discovery*, Methods in Molecular Biology, L. W. Tari, ed., Totowa, NJ: Humana Press, 1–27.
- Thakuria, R., Nath, N. K., and Saha, B. K. (2019). “The Nature and Applications of π - π Interactions: A Perspective.” *Crystal Growth & Design*, 19(2), 523–528.
- Tobi, D., and Bahar, I. (2005). “Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state.” *Proceedings of the National Academy of Sciences*, 102(52), 18908–18913.
- Tobler, I., Gaus, S. E., Deboer, T., Achermann, P., Fischer, M., Rüdliche, T., Moser, M., Oesch, B., McBride, P. A., and Manson, J. C. (1996). “Altered circadian activity rhythms and sleep in mice devoid of prion protein.” *Nature*, 380(6575), 639–642.
- Tomasi, J., Mennucci, B., and Cammi, R. (2005). “Quantum Mechanical Continuum Solvation Models.” *Chem. Rev.*, 105(8), 2999–3094.
- Törnquist, M., Michaels, T. C. T., Sanagavarapu, K., Yang, X., Meisl, G., Cohen, S. I. A., Knowles, T. P. J., and Linse, S. (2018). “Secondary nucleation in amyloid formation.” *Chem. Commun.*, 54(63), 8667–8684.
- Tripuraneni, N. S., and Azam, M. A. (2016). “Pharmacophore modeling, 3D-QSAR and docking study of 2-phenylpyrimidine analogues as selective PDE4B inhibitors.” *Journal of Theoretical Biology*, 394, 117–126.
- Tropsha, A., Golbraikh, A., and Cho, W.-J. (2011). “Development of kNN QSAR Models for 3-Arylisoquinoline Antitumor Agents.” *Bulletin of the Korean Chemical Society*, 32(7), 2397–2404.
- Tsai, C. J., Kumar, S., Ma, B., and Nussinov, R. (1999). “Folding funnels, binding funnels, and protein function.” *Protein Sci*, 8(6), 1181–1190.
- Umamatheswari, S., Balaji, B., Ramanathan, M., and Kabilan, S. (2010). “Synthesis, antimicrobial evaluation and QSAR studies of novel piperidin-4-yl-5-spirothiadiazoline derivatives.” *Bioorganic & Medicinal Chemistry Letters*, 20(23), 6909–6914.
- Van Gunsteren, W. F., and Berendsen, H. J. C. (1988). “A Leap-frog Algorithm for Stochastic Dynamics.” *Molecular Simulation*, 1(3), 173–185.
- Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., and MacKerell, A. D. (2010). “CHARMM General Force Field (CGenFF): A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields.” *J Comput Chem*, 31(4), 671–690.
- 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*, 0(0), 1–16.

Venugopal, P. P., Das, B. K., Soorya, E., and Chakraborty, D. (2020). “Effect of hydrophobic and hydrogen bonding interactions on the potency of β -alanine analogs of G-protein coupled glucagon receptor inhibitors.” *Proteins: Structure, Function, and Bioinformatics*, 88(2), 327–344.

Venugopal, P. P., M, S., and Chakraborty, D. (n.d.). “Theoretical insights into molecular mechanism and energy criteria of PARP-2 enzyme inhibition by benzimidazole analogues.” *Proteins: Structure, Function, and Bioinformatics*, n/a(n/a).

Verma, J., Khedkar, V. M., and Coutinho, E. C. (2010). “3D-QSAR in Drug Design - A Review.” *Current Topics in Medicinal Chemistry*, 10(1), 95–115.

Wagoner, J. A., and Baker, N. A. (2006). “Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms.” *PNAS*, 103(22), 8331–8336.

Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., and Velesler, D. (2020). “Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein.” *Cell*, 181(2), 281-292.e6.

Walsh, D. M., Klyubin, I., Fadeeva, J. V., Cullen, W. K., Anwyl, R., Wolfe, M. S., Rowan, M. J., and Selkoe, D. J. (2002). “Naturally secreted oligomers of amyloid β protein potently inhibit hippocampal long-term potentiation in vivo.” *Nature*, 416(6880), 535–539.

Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z. H., and Hou, T. (2019). “End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design.” *Chem. Rev.*, 119(16), 9478–9508.

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). “Development and testing of a general amber force field.” *J Comput Chem*, 25(9), 1157–1174.

Wang, L., Berne, B. J., and Friesner, R. A. (2012). “On achieving high accuracy and reliability in the calculation of relative protein–ligand binding affinities.” *PNAS*, 109(6), 1937–1942.

Wang, L., Friesner, R. A., and Berne, B. J. (2011). “Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2).” *J Phys Chem B*, 115(30), 9431–9438.

Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., Lupyan, D., Robinson, S., Dahlgren, M. K., Greenwood, J., Romero, D. L., Masse, C., Knight, J. L., Steinbrecher, T., Beuming, T., Damm, W., Harder, E., Sherman, W., Brewer, M., Wester, R., Murcko, M., Frye, L., Farid, R., Lin, T., Mobley, D. L., Jorgensen, W. L., Berne, B. J., Friesner, R. A., and Abel, R. (2015). “Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field.” *J. Am. Chem. Soc.*, 137(7), 2695–2703.

Wang, W., and Kollman, P. A. (2000). “Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model” Edited by B. Honig.” *Journal of Molecular Biology*, 303(4), 567–582.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., Beer, T. A. P. de, Rempfer, C., Bordoli, L., Lepore, R., and Schwede, T. (2018). “SWISS-MODEL: homology modelling of protein structures and complexes.” *Nucleic Acids Res.*, 46(W1), W296–W303.

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.

Wel, P. C. A. van der, Lewandowski, J. R., and Griffin, R. G. (2007). "Solid-State NMR Study of Amyloid Nanocrystals and Fibrils Formed by the Peptide GNNQQNY from Yeast Prion Protein Sup35p." *J. Am. Chem. Soc.*, 129(16), 5117–5130.

Wells, B. A., and Chaffee, A. L. (2015). "Ewald Summation for Molecular Simulations." *J. Chem. Theory Comput.*, 11(8), 3684–3695.

Westhof, E., Altschuh, D., Moras, D., Bloomer, A. C., Mondragon, A., Klug, A., and Van Regenmortel, M. H. V. (1984). "Correlation between segmental mobility and the location of antigenic determinants in proteins." *Nature*, 311(5982), 123–126.

Whetstine, J. R., Flatley, R. M., and Matherly, L. H. (2002). "The human reduced folate carrier gene is ubiquitously and differentially expressed in normal human tissues: identification of seven non-coding exons and characterization of a novel promoter." *Biochem J*, 367(Pt 3), 629–640.

Wiederstein, M., and Sippl, M. J. (2007). "ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins." *Nucleic Acids Res*, 35(suppl_2), W407–W410.

William L. Jorgensen, *, David S. Maxwell, and Tirado-Rives, J. (1996). "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids." <<https://pubs.acs.org/doi/abs/10.1021/ja9621760>> (May 15, 2018).

Wit, E. de, Doremalen, N. van, Falzarano, D., and Munster, V. J. (2016). "SARS and MERS: recent insights into emerging coronaviruses." *Nature Reviews Microbiology*, 14(8), 523–534.

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., Graham, B. S., and McLellan, J. S. (2020). "Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation." *Science*, 367(6483), 1260–1263.

Wright, D. W., Hall, B. A., Kenway, O. A., Jha, S., and Coveney, P. V. (2014). "Computing Clinically Relevant Binding Free Energies of HIV-1 Protease Inhibitors." *J. Chem. Theory Comput.*, 10(3), 1228–1241.

Xiu, X., Puskar, N. L., Shanata, J. A. P., Lester, H. A., and Dougherty, D. A. (2009). "Nicotine binding to brain receptors requires a strong cation- π interaction." *Nature*, 458(7237), 534–537.

Yamamoto, N., Tsuchida, S., Tamura, A., and Chatani, E. (2018). "A specific form of prefibrillar aggregates that functions as a precursor of amyloid nucleation." *Scientific Reports*, 8(1).

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020). "Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2." *Science*, 367(6485), 1444–1448.

Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., and Baker, D. (2020). "Improved protein structure prediction using predicted interresidue orientations." *PNAS*, 117(3), 1496–1503.

Yap, C. W. (2011). "PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints." *Journal of Computational Chemistry*, 32(7), 1466–1474.

Yogavel, M., Nettleship, J. E., Sharma, A., Harlos, K., Jamwal, A., Chaturvedi, R., Sharma, M., Jain, V., Chhibber-Goel, J., and Sharma, A. (2018). "Structure of 6-

hydroxymethyl-7,8-dihydropterin pyrophosphokinase–dihydropteroate synthase from *Plasmodium vivax* sheds light on drug resistance.” *Journal of Biological Chemistry*, 293(39), 14962–14972.

Yong, C. Y., Ong, H. K., Yeap, S. K., Ho, K. L., and Tan, W. S. (2019). “Recent Advances in the Vaccine Development Against Middle East Respiratory Syndrome-Coronavirus.” *Front Microbiol*, 10, 1781.

You, W., Tang, Z., and Chang, C. A. (2019). “Potential Mean Force from Umbrella Sampling Simulations: What Can We Learn and What Is Missed?” *J. Chem. Theory Comput.*, 15(4), 2433–2443.

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.” *Biochemical Pharmacology*, 183, 114296.

Ytreberg, F. M., Swendsen, R. H., and Zuckerman, D. M. (2006). “Comparison of free energy methods for molecular systems.” *J. Chem. Phys.*, 125(18), 184114.

Yu, L., Lee, S.-J., and Yee, V. C. (2015a). “Crystal Structures of Polymorphic Prion Protein β 1 Peptides Reveal Variable Steric Zipper Conformations.” *Biochemistry*, 54(23), 3640–3648.

Yu, L., Lee, S.-J., and Yee, V. C. (2015b). “Crystal Structures of Polymorphic Prion Protein β 1 Peptides Reveal Variable Steric Zipper Conformations.” *Biochemistry*, 54(23), 3640–3648.

Yuan, M., Wu, N. C., Zhu, X., Lee, C.-C. D., So, R. T. Y., Lv, H., Mok, C. K. P., and Wilson, I. A. (2020). “A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV.” *Science*, 368(6491), 630–633.

Yun, M.-K., Wu, Y., Li, Z., Zhao, Y., Waddell, M. B., Ferreira, A. M., Lee, R. E., Bashford, D., and White, S. W. (2012a). “Catalysis and Sulfa Drug Resistance in Dihydropteroate Synthase.” *Science*, 335(6072), 1110–1114.

Yun, M.-K., Wu, Y., Li, Z., Zhao, Y., Waddell, M. B., Ferreira, A. M., Lee, R. E., Bashford, D., and White, S. W. (2012b). “Catalysis and Sulfa Drug Resistance in Dihydropteroate Synthase.” *Science*, 335(6072), 1110–1114.

Zawawi, A., Forman, R., Smith, H., Mair, I., Jibril, M., Albaqshi, M. H., Brass, A., Derrick, J. P., and Else, K. J. (2020). “In silico design of a T-cell epitope vaccine candidate for parasitic helminth infection.” *PLOS Pathogens*, 16(3), e1008243.

Zhan, C.-G., Nichols, J. A., and Dixon, D. A. (2003). “Ionization Potential, Electron Affinity, Electronegativity, Hardness, and Electron Excitation Energy: Molecular Properties from Density Functional Theory Orbital Energies.” *J. Phys. Chem. A*, 107(20), 4184–4195.

Zhang, Y., Guo, R., Kim, S. H., Shah, H., Zhang, S., Liang, J. H., Fang, Y., Gentili, M., Leary, C. N. O., Elledge, S. J., Hung, D. T., Mootha, V. K., and Gewurz, B. E. (2021). “SARS-CoV-2 hijacks folate and one-carbon metabolism for viral replication.” *Nature Communications*, 12(1), 1676.

Zhao, Y., Hammoudeh, D., Yun, M.-K., Qi, J., White, S. W., and Lee, R. E. (2012). “Structure-based design of novel pyrimido[4,5-c]pyridazine derivatives as dihydropteroate synthase inhibitors with increased affinity.” *ChemMedChem*, 7(5), 861–870.

Zhao, Y., Shadrack, W. R., Wallace, M. J., Wu, Y., Griffith, E. C., Qi, J., Yun, M.-K., White, S. W., and Lee, R. E. (2016a). “Pterin-sulfa conjugates as dihydropteroate

synthase inhibitors and antibacterial agents.” *Bioorg. Med. Chem. Lett.*, 26(16), 3950–3954.

Zhao, Y., Shadrick, W. R., Wallace, M. J., Wu, Y., Griffith, E. C., Qi, J., Yun, M.-K., White, S. W., and Lee, R. E. (2016b). “Pterin-sulfa conjugates as dihydropteroate synthase inhibitors and antibacterial agents.” *Bioorg Med Chem Lett*, 26(16), 3950–3954.

Zheng, J. (2020). “SARS-CoV-2: an Emerging Coronavirus that Causes a Global Threat.” *Int. J. Biol. Sci.*, 16(10), 1678–1685.

Zheng, Y., Zheng, M., Ling, X., Liu, Y., Xue, Y., An, L., Gu, N., and Jin, M. (2013). “Design, synthesis, quantum chemical studies and biological activity evaluation of pyrazole–benzimidazole derivatives as potent Aurora A/B kinase inhibitors.” *Bioorganic & Medicinal Chemistry Letters*, 23(12), 3523–3530.

Zhou, H., Yang, Z., Tian, X., Chen, L., Lee, S., Huynh, T., Ge, C., and Zhou, R. (2019a). “Lanosterol Disrupts the Aggregation of Amyloid- β Peptides.” *ACS Chem. Neurosci.*, 10(9), 4051–4060.

Zhou, H., Yang, Z., Tian, X., Chen, L., Lee, S., Huynh, T., Ge, C., and Zhou, R. (2019b). “Lanosterol Disrupts the Aggregation of Amyloid- β Peptides.” *ACS Chem. Neurosci.*, 10(9), 4051–4060.

Zhou, J.-K., Yang, D.-Y., and Sheu, S.-Y. (2015a). “The molecular mechanism of ligand unbinding from the human telomeric G-quadruplex by steered molecular dynamics and umbrella sampling simulations.” *Phys. Chem. Chem. Phys.*, 17(19), 12857–12869.

Zhou, J.-K., Yang, D.-Y., and Sheu, S.-Y. (2015b). “The molecular mechanism of ligand unbinding from the human telomeric G-quadruplex by steered molecular dynamics and umbrella sampling simulations.” *Phys. Chem. Chem. Phys.*, 17(19), 12857–12869.

LIST OF PUBLICATIONS RELATER TO THE THESIS

1. **Das B. K.**, Pushyaraga P. V., Chakraborty D. (2018). “Computational Insights into Factors Affecting the Potency of Diaryl Sulfone Analogs as *Escherichia coli* Dihydropteroate Synthase Inhibitors.” *Computational Biology and Chemistry*, 78, pp. 37-52.
2. **Das B. K.**, Chakraborty D. (2020). “Epitope-Based Potential Vaccine Candidate for Humoral and Cell-Mediated Immunity to Combat Severe Acute Respiratory Syndrome Coronavirus 2 Pandemic.” *Journal of Physical Chemistry Letters*. 11, pp. 9920-9930.
3. **Das B. K.**, Chakraborty D. (2021). “Deciphering the Competitive Inhibition of Dihydropteroate Synthase by 8 Marcaptoguanine Analogues: Enhanced Potency in Phenylsulfonyl fragments.” *Journal of Biomolecular Structure and Dynamics*, pp. 1-20
4. **Das B. K.**, Singh O., Chakraborty D. (2023). “Exploring the barriers in the aggregation of hexadecameric human prion peptide through Markov state model” *ACS Chemical Neuroscience (Accepted)*

OTHER PUBLICATIONS

1. Sarkar P., **Das B. K.**, Chakraborty D., Muthamma K. (2019). “Carbohelicenes and thiahelicene from phthalaldehydes through Perkin approach.” *Journal of Molecular Structure*. 1195, pp. 309-314.
2. Pushyaraga P. V., **Das B. K.**, Soorya E., Chakraborty D. (2020). “Effect of hydrophobic and hydrogen bonding interactions on the potency of β -alanine analogues of G-protein coupled glucagon receptor inhibitors” *Proteins: Structure, Function and Bioinformatics*. 88, pp. 327-344.
3. Nelapati A. K., **Das B. K.**, Ettiyanpan J. P. B., Chakraborty D. (2020). “In-silico epitope identification epitope identification and design of uricase mutein with reduced immunogenicity.” *Process Biochemistry*. 92, pp. 288-302.
4. Dalal S., **Das B. K.**, Saini M., Chakraborty D., Sadhu K. K. (2021). “Diverse interactions of aggregated insulin with selected coumarin dyes: Time dependent fluorogenicity, simulation studies and comparison with thioflavin T. ” *Dyes and Pigments*. 184, pp. 108796.
5. Singh O., **Das B. K.**, Chakraborty D. “Influence of ion specificity and concentration on the conformational transition of intrinsically disordered sheep prion peptide”. *Chem phys chem (Accepted)* [**Equal Contribution**]

CONFERENCE PRESENTATION

1. **B.K. Das** and D. Chakraborty “Optimization of Structural Stability of Dihydropteroate Synthase through Molecular Dynamics Simulation” *Emerging Trends in Drug Discovery and Development-2018*.
2. **B.K. Das** and D. Chakraborty “In-Silico Discovery of 8-Mercaptoguanine Compounds as Potential Dihydropteroate Synthase Antagonist” *16th Theoretical Chemistry Symposium 2019*.
3. **B.K. Das** and D. Chakraborty “Potential peptide vaccines for Humoral and Cell Mediated Immunity against SARS-CoV2 pandemic” *17th Theoretical Chemistry Symposium 2021*.

BIODATA

Name: Bratin Kumar Das

Registration Number: 177074/177CY006

Email id: bratind145@gmail.com

Education:

Course	Institution	Duration
Ph. D (Computational Chemistry and Biology)	National Institute of Technology Karnataka, Surathkal	2017-2023
Master of Technology (Biochemical Engineering)	National Institute of Technology Karnataka, Surathkal	2015-2017
Bachelor of Technology (Biotechnology)	Heritage Institute of Technology Kolkata	2012-2015

Place: NITK Surathkal

Date: 11/09/2023