

**An investigation to understand the role of key facets
involved in the coagulation/anticoagulation process by
employing computational and molecular approaches**

Thesis Submitted for the Degree of
Doctor of Philosophy (Science)

to



Jadavpur University

May 2022

by

Suparna Banerjee



School of Biological Sciences

Indian Association for the Cultivation of Science

Jadavpur, Kolkata 700 032

India



Indian Association for the Cultivation of Science

2A & 2B Raja S C Mullick Road, Jadavpur, Kolkata – 700 032,

West Bengal, India, e-mail: bcps@iacs.res.in

Fax: +91-33 2473-2805

Phone: +91-33 2473-4971, Extn: 1518.

Dr. Prosenjit Sen

Professor

School of Biological Sciences

Date: 10.05.2022

CERTIFICATE FROM THE SUPERVISOR

This is to certify that the thesis entitled “**An investigation to understand the role of key facets involved in the coagulation/anticoagulation process by employing computational and molecular approaches**” submitted by Smt. Suparna Banerjee, who got her name registered on **24.07.2017** for the award of **Ph. D. (Science) degree of Jadavpur University**, is absolutely based upon his own work under the supervision of Dr. **Prosenjit Sen** and that neither this thesis nor any part of it has been submitted for either any degree/diploma or any other academic award anywhere before.

DR. PROSENJIT SEN
Associate Professor
Indian Association For
The Cultivation of Science
2A & B Raja S. C. Mullick Road
Jadavpur, Kolkata - 700 032


Prosenjit Sen
(Signature of the Supervisor &
date with official seal)

Dedicated

to

My mother Smt. Lily Banerjee,

&

My husband Shri Soumajit Chatterjee

DECLARATION

The research work manifested in this thesis entitled “**An investigation to understand the role of key facets involved in the coagulation/anticoagulation process by employing computational and molecular approaches**” being submitted to Jadavpur University, Kolkata, has been carried out at the Indian Association for the Cultivation of Science, Jadavpur, under the supervision of Dr. Prosenjit Sen, Professor, School of Biological Sciences, Indian Association for the Cultivation of Science. This work is original and has not been submitted in part or in full, for any degree or diploma to this or any other university.

(Suparna Banerjee)

TABLE OF CONTENTS

	Page No.
• Preface	v
• Acknowledgment	vi
• Abstract	vii
• List of Tables	viii-x
• List of Figures	xi-xiv
• List of Abbreviations	xv-xvi
Chapter 1: General Introduction	1-22
1. Overview	1
2. Blood Coagulation Pathways	2
3. Intrinsic (contact activation) pathway	5
4. Extrinsic (tissue-factor mediated) pathway	7
5. Coagulation protein factor VII/factor VIIa	8
6. Tissue factor	11
7. Endothelial protein C receptor (EPCR)	12
8. The motivation for the current work	13
9. Significance of the present thesis work	15
References	17
Chapter 2 In silico mutation and binding studies of human FVIIa GLA-domain to endothelial protein C receptor: A molecular dynamics simulation approach	23-56
1. Introduction	23
2. Computational Section	25
2.1 System Setup	25
2.2 Protein modeling.	26
2.3 Molecular Dynamics Simulations Protocol.	28
2.4 Potential of mean force calculation using the ABF method	30
2.5 Protein-protein binding affinity change predictions by BindProfX	31
2.6 Residue-wise interaction energy calculation using gRINN.	32
3. Results and Discussion	32
3.1 Dynamics of FVIIa/FXa GLA-domain in EPCR complex.	33

3.2	Free binding energy profile of fIEPCR and Gla-domain complex.	37
3.3	Evaluation of the interactions formed during MD simulations.	42
3.4	Solvent Accessible Surface Area	48
4.	Conclusions	51
	References	52

Chapter 3 A molecular dynamics simulation study to elucidate the effect of Cholesterol and Tissue Factor Palmitoylation on TF-FVIIa-FXa ternary complex in different lipid environments. 57-100

1.	Introduction	57
2.	Methodology	60
2.1	System Set up	60
2.2	Modeling of protein target—TF-FVIIa-FXa (TERNARY COMPLEX)	62
2.3	MD simulation setup protocol	65
3.	Results and Discussion	66
3.1	Protein Dynamics Study	66
3.2	Conformation analysis of complexes	72
3.3	Evaluation of the Interactions Formed during MD Simulations	77
3.4	Contact Probability estimation between Protein and lipid	85
3.5	Protein stability predictions by FoldX	89
4.	Conclusion	91
	References	93

Chapter 4 Identification of potential antifibrinolytic compounds against kringle-1 and serine protease domain of Plasminogen and kringle-2 domain of tissue-type plasminogen activator using combined virtual screening, molecular docking, and molecular dynamics simulation approaches. 101-142

1.	Introduction	101
2.	Methodology	103
2.1.	Protein Preparation	103
2.2.	Structure-based virtual screening	104
2.3	Molecular Docking	104

2.3.1	Glide (Grid-based ligand docking with energetic)	104
2.3.2	Auto Dock Vina	105
2.3.3	ParDOCK & Bappl+	106
2.4.	ADMET and TOPKAT	106
2.5.	Molecular Dynamics Simulation (MD)	107
2.6.	Principal Component Analysis	107
2.7.	Free energy calculation	108
3.	Results and discussion	109
3.1	Virtual screening (VS)	109
3.2	Comparative Molecular Docking	110
3.3	Trajectory Analysis	120
3.3.1	Root-mean-square deviation analysis	120
3.3.2	Residue flexibility analysis	122
3.3.3	Compactness analysis	123
3.3.4	Solvent Accessible Surface Area analysis	125
3.3.5	Hydrogen Bond and Bond distribution analysis	127
3.3.6	Principal component analysis	129
3.3.7	Binding free energy calculation	130
4.	Conclusions	134
	References	136

Chapter 5 Evaluation of compounds having potential anti-cancer activity against programmed death-ligand 1 (PD-L1) using combined molecular docking, and molecular dynamics simulation approaches. 143-178

1.	Introduction	143
2.	Methodology	147
2.1	Selection of the PD-L1 protein structure and a reference ligand for docking purposes	147
2.2	Molecular Docking	147
2.2.1	Auto Dock Vina	148
2.2.2	ParDOCK & Bappl+	148
2.3	ADMET and TOPKAT	148
2.4	Molecular Dynamics Simulation (MD)	149

2.5	Principal Component Analysis	150
2.6	Free energy calculation	150
3.	Results and discussion	151
3.1	Comparative Molecular Docking	151
3.2	Trajectory Analysis	158
3.2.1	Root-mean-square deviation analysis	158
3.2.2	Residue flexibility analysis	160
3.2.3	Compactness analysis	161
3.2.4	Solvent Accessible Surface Area analysis	162
3.2.5	Hydrogen Bond and Bond distribution analysis	164
3.2.6	Principal component analysis	165
3.2.7	Binding free energy calculation	166
4.	Conclusions	170
	References	172
	List of Publications	179-180
	Paperwork included in the thesis	
	Paperwork not included in the thesis	
	Book chapters not included in the thesis	

PREFACE

I hereby declare that the matter manifested in this thesis, “**An investigation to understand the role of key facets involved in the coagulation/anticoagulation process by employing computational and molecular approaches**” is the outcome of research carried out by me in the School of Biological Sciences, Indian Association for the Cultivation of Science (IACS), Jadavpur, Kolkata, India under the supervision of Dr. Prosenjit Sen.

- **Chapter 1** covers the fundamental of the blood coagulation process and illustrates the significant contributory factors (TF, FVII, FX, and EPCR) that are responsible for maintaining hemostasis.
- **Chapter 2** provides detailed insight into differences in species-specific binding of FVIIa GLA-domain and FXa GLA-domain towards EPCR and helps in understanding the structure-function relationship of the protein complex upon mutating the key residues.
- **Chapter 3** demonstrates the dynamics and interactions of cholesterol with the TF-FVIIaFXa Ternary complex and also assesses the effect of post-translational modification (Tissue Factor Palmitoylation of CYS²⁴⁵ residue) on protein-lipid interactivity in presence of cholesterol.
- **Chapter 4** explores computational drug discovery or virtual screening process to identify potent antifibrinolytic agents that may inhibit fibrinolysis by targeting the kringle-1 and serine protease domain of plasminogen and the kringle-2 domain of tissue plasminogen activator.
- **Chapter 5** involves combined molecular docking and molecular dynamics simulation approaches for an evaluation of in-house synthesized small-molecule inhibitors having potent anti-cancer activity against the programmed death-ligand 1 (PD-L1) as a part of a collaborative project.

May 2022

School of Biological Sciences

(Suparna Banerjee)

Indian Association for the Cultivation of Science
Jadavpur, Kolkata – 700032 India.

ACKNOWLEDGMENT

I am highly grateful to God and to all those people who accompanied me at every phase of my Ph.D. tenure and inspired me for the successful completion of my Doctoral thesis work. First and foremost, I express my heartfelt gratitude to my Ph.D. supervisor, **Prof. Prosenjit Sen**, for being my guiding light and a strong pillar throughout this journey. It is because of his constant counsel, I have been able to proceed with the work in the right direction. He has always extended his full support to me and helped me to achieve my aim.

I am highly grateful to the **Indian Association for the Cultivation of Science (IACS)** for providing me the opportunity to pursue my research work. I would like to thank all faculty members of our department, Prof. Arindam Banerjee, Prof. Rupa Mukhopadhyay, Prof. Prasanta Kumar Das, Prof. Benu Brata Das, and Prof. Deepak Kr. Sinha, Prof. Siddhartha Sankar Jana for helping me during my Ph.D. course work. I express my sincere gratitude to Professor K. Sekar, Department of Computational and Data Sciences, Indian Institute of Science, Bangalore for his wholehearted support and cooperation.

I am highly obliged to the **Council of Scientific & Industrial Research (CSIR), India** for providing me with the fellowship.

I am extremely thankful to my senior Dr. Ramesh Prasad for helping me during the initial phase of my Ph.D. work. I am grateful to my labmates Dr. Kaushik, Dr. Abhishek, Arnab, Subhojit, Subhasis, Dr. Anindita, Arpana, Dr. Ruma, Akash, Avinanadan, and Dr. Suman for inspiring me throughout this phase and making my stay enjoyable. Words are not enough to acknowledge my brother **Yeshwanth M.** who has helped me immensely in the completion of my work.

Lastly and most importantly, I owe my debt of gratitude to my mother, **Smt. Lily Banerjee**, my Ma **Smt. Sumita Ganguly**, and my husband **Shri Soumajit Chatterjee** for motivating and helping me to achieve my goal. They have been the constant driving force for the successful accomplishment of my work.

Suparna Banerjee

Abstract

Index No. 69/17/Life Sc./25

Thesis title: “An investigation to understand the role of key facets involved in the coagulation/anticoagulation process by employing computational and molecular approaches”

Submitted by: Suparna Banerjee

School of Biological Sciences

Indian Association for the Cultivation of Science, Jadavpur, Kolkata-700032

Hemostasis is an intricate, highly regulated, and specialized process that not only prevents and arrests bleeding/hemorrhage but also eliminates blood clots after the restoration of vascular integrity. The equilibrium between the two processes namely blood coagulation and anticoagulation exclusively depends on a balanced interplay among a series of elements: the coagulation factors (mostly proteolytic enzymes), endothelium, and platelets. Factor VIIa (FVIIa), used in hemophilia treatment, is known to bind to procoagulant receptor Tissue Factor (TF) and initiate coagulation. Recent studies have revealed that FVIIa also binds to EPCR (endothelial protein C receptor), a receptor involved in the protein C/APC-anticoagulant pathway but with unclear hemostatic consequences. Being homologous to protein C, FVIIa also binds to EPCR through its GLA-domain. Factor Xa (FXa) has a GLA-domain like FVIIa and protein C; however, there is a variation in FVIIa/FXa GLA-domain binding with EPCR in different species such as humans and mice. This thesis work provides detailed insight into differences in species-specific binding of FVIIa GLA-domain and FXa GLA-domain towards EPCR, which may facilitate the designing of FVIIa-mutant molecules for improved therapeutic use associated with bleeding disorders and septic shock. In the next project, we have tried to characterize the dynamics and interactions of cholesterol with the TF-FVIIa-FXa Ternary complex. The effect of cholesterol content and post-translational modifications of TF on TF-FVIIa coagulant activity is still indeterminate. Previous studies suggest that the presence of cholesterol/LDL in atherosclerotic plaques confers high importance to atherogenesis. Aberrant palmitoylation is related to various diseases including neurological disorders like Huntington's disease, Parkinson's disease, Alzheimer's disease, metabolic disorders, and Cancer. Additionally, we also attempted to assess the effect of Palmitoylation, on protein-lipid interactivity in presence of cholesterol. Furthermore, we also undertook a project associated with the virtual screening, docking, and molecular dynamics simulation of antifibrinolytic agents which will prevent blood clot dissolution and abnormal blood loss. Herein, we attempt to find potential antifibrinolytic inhibitors having both maximum specificity and efficacy with minimum side effects. As a part of a collaborative project, an evaluation of in-house synthesized small-molecule inhibitors having potent anti-cancer activity against the programmed death-ligand 1 (PD-L1) was performed using *in silico* approaches.

(Signature of the supervisor)

(Signature of the Candidate)

DR. PROSENJIT SEN
Associate Professor
Indian Association For
The Cultivation of Science
2A & B Raja S. C. Mullick Road
Jadavpur, Kolkata - 700 032

List of Tables

- Table 1-1. List of coagulant/anticoagulant factors with their function, and the associated diseases with it.
- Table 2-1. Details of in silico mutation associated with human FVIIa GLA-domain.
- Table 2-2. Root mean square deviation (RMSD) of EPCR and GLA-domain, calculated from the last 100 ns of the production run for all the systems (membrane-bound flEPCR and GLA-domain complex)
- Table 2-3. The average potential of mean force with standard error is associated with the EPCR-GLA domain complexes system.
- Table 2-4. Summary of the binding profile of all mutant systems (mouse FVIIa and human FXa-based) using BindProfX.
- Table 3-1. Details of system size and the simulation description of all systems.
- Table 3-2. Average C α root mean square deviation (RMSDs) and standard deviation (SD) of various domains of TF-FVIIa-FXa Ternary Complex in POPC: POPS (CONTROL), TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CHOL), and Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CYSP)
- Table 3-3. Comparative analysis of interacting residues for H-Bond formation with a cut-off distance of 3.5 Å for all the three systems: TF in TF-FVIIa-FXa Ternary Complex (CONTROL), TF-FVIIa-FXa Ternary Complex (CHOL) and Palmitoylated TF-FVIIa-FXa Ternary Complex (CYSP) obtained from MD simulation study.
- Table 3-4.1. Contact probability between Chain B of FXa with lipid.
- Table 3-4.2. Contact Probability between Chain L of FVIIa with lipid.
- Table 3-4.3. Contact Probability data between Chain P of TF with lipid
- Table 3-5: Interaction energy between TF and FVIIa (light chain) using the FoldX algorithm.

- Table 4-1. AUTODOCK-VINA scores of the top 10 identified ligands docked with their respective protein targets.
- Table 4-2. PARDOCK/BAPPL+ scores of the top 10 identified ligands docked with their respective protein targets.
- Table 4-3. GLIDE scores of the top 10 identified ligands docked with their respective protein targets.
- Table 4-4. GLIDE scores of the top 3 identified ligands docked with their respective protein targets.
- Table 4-5. Molecular docking results of the top nine identified ligands (P27, P67, P76, C19, C90, C97 U19, U94, U97) and the reference ligands (ACA, XO3, and 89M) bound with their respective protein targets (1PK2, 4CIK and 5UGG) using three docking tools: Autodock vina, ParDOCK/BAPPL+, and Glide respectively.
- Table 4-6: ADMET properties of top 3 identified ligands for each protein target.
- Table 4-7: TOPKAT properties of top 3 identified ligands for each protein target.
- Table 4-8: RMSD values of all the protein-ligand complexes.
- Table 4-9: Rg values of all the protein-ligand complexes.
- Table 4-10: SASA values of all the protein-ligand complexes.
- Table 4-11: MMPBSA detailed energies (Binding energy and the individual energy components van der Waal energy (ΔE_{vdW}), electrostatic energy (ΔE_{elec}), polar solvation energy ($\Delta G_{polar-solv}$), SASA energy ($\Delta G_{non-polar}$) of all the protein-ligand complexes.
- Table 5-1: Molecular docking results of ten novel small-molecule inhibitors (M1-M10) and the reference ligand (6GX) bound with their respective protein target (5J89) using two docking tools: Autodock vina, ParDOCK/BAPPL+ respectively.
- Table 5-2: ADMET properties of novel identified ligands (M1-M10) and the reference ligand (6GX).

- Table 5-3: TOPKAT properties of novel identified ligands (M1-M10) and the reference ligand (6GX) TOPKAT properties: NC -Non-carcinogenic, C-Carcinogenic, NT-Non-Toxic, T-Toxic, NI-Non-Irritant, I-Irritant, NM-Non-Mutagenic, M-Mutagenic
- Table 5-4: RMSD profile of 5J89 protein-inhibitor complexes [novel small-molecule inhibitors (M1-M10) and reference ligand 6GX] along with APO.
- Table 5-5: Rg values of all the protein-inhibitor complexes [novel small-molecule inhibitors (M1-M10) and reference ligand 6GX] along with APO.
- Table 5-6: SASA values of all the protein-inhibitor complexes [novel small-molecule inhibitors (M1-M10) and reference ligand 6GX] along with APO.
- Table 5-7: MMPBSA detailed energies (Binding energy and the individual energy components van der Waal energy (ΔE_{vdw}), electrostatic energy (ΔE_{elec}), polar solvation energy ($\Delta G_{polar-solv}$), SASA energy ($\Delta G_{non-polar}$) of all the protein-ligand complexes [5J89 complexed with the novel small-molecule inhibitors (M1-M10) and reference ligand 6GX].

LIST OF FIGURES

- Figure 1-1. The blood coagulation pathway
- Figure 1-2. The intrinsic pathway of blood coagulation.
- Figure 1-3. The extrinsic pathway of blood coagulation.
- Figure 1-4. Schematic representation of the binary complex.
- Figure 1-5. Three-dimensional structure of FVIIa, membrane-bound full length(fl) TF, and membrane-bound full length (fl) TF -FVIIa complex.
- Figure 1-6. Vitamin K-dependent gamma-carboxylation of specific Glu (glutamic acid) residues
- Figure 1-7. Three-dimensional structure of full-length TF-FVIIa complex on the membrane lipid bilayer.
- Figure 1-8. Three-dimensional structure of EPCR with GLA domain of FVIIa/FXa on the membrane lipid bilayer.
- Figure 2-1. Overview of flEPCR-FVIIa/FXa GLA-domain structure.
- Figure 2-2. Root mean square deviations (RMSD) of GLA-domain and EPCR were obtained from simulation of membrane-bound flEPCR and GLA-domain complex.
- Figure 2-3. Root mean square deviations (RMSD) of GLA-domain and EPCR were obtained from 130 ns unbiased MD simulation trajectories of the EPCR-GLA complex for all the systems.
- Figure 2-4. Residue-wise fluctuation of GLA-domain and EPCR from membrane-bound flEPCR-GLA domain complex.
- Figure 2-5. Free energy profile (potential of mean force, PMF) obtained from ABF simulations along the reaction coordinate (RC) between GLA-domain and sEPCR.
- Figure 2-6. The conformational changes of EPCR-GLA domain for FVIIaWT system in presence of PCh-containing lipid within EPCR groove at center-of-mass (COM) distance of 32 Å, 34 Å, and 36 Å, obtained from the ABF simulation trajectories.

- Figure 2-7. Occupancy of interacting residue-pair between EPCR and GLA-domain obtained from MD simulation.
- Figure 2-8. Interaction energy (IE) calculation between the selected residue pair of GLA domain and EPCR, obtained from the occupancy plot.
- Figure 2-9. Distance distribution of unique interacting residue-pair between Gln150:NE2-Gla7:OE12 for all systems.
- Figure 2-10. Schematic representation of the solvent-accessible surface area (SASA).
- Figure 3-1. 3D representation of (A) TF-FVIIa-FXa Ternary Complex in POPC:POPS (CONTROL), (B)TF-FVIIa-FXa Ternary Complex in POPC:POPS:CHL (CHOL), and (C)Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC:POPS:CHL (CYSP) used in the present study.
- Figure 3-2. Root mean square deviations (RMSD) obtained from simulation of TF-FVIIa-FXa Ternary Complex in POPC: POPS (CONTROL), TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CHOL), and Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CYSP) for the total 200ns.
- Figure 3-3. Root mean square fluctuation (RMSF) obtained from simulation of TF-FVIIa-FXa Ternary Complex in POPC: POPS (CONTROL), TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CHOL), and Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CYSP) for the last 60 ns of the total simulation run.
- Figure 3-4. Structural analysis of TF-FVIIa-FXa Ternary Complex in POPC: POPS (CONTROL), TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL(CHOL), and Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC:POPS: CHL (CYSP).
- Figure 3-5. Insight into Catalytic triad (CT) dynamics in TF-FVIIa-FXa Ternary Complex in POPC: POPS(CONTROL), TF-FVIIa-FXa Ternary Complex in POPC:POPS: CHL (CHOL) and Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC:POPS: CHL (CYSP)
- Figure 3-6. Detailed interaction profile (H-bond forming pairs) between various domains of TF-FVIIa-FXa Ternary Complex in POPC: POPS

(CONTROL), TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CHOL) and Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CYSP) calculated from the 200ns simulation period of all the three complexes.

Figure 3-7. Detailed contact probability profile (lipid-protein interaction) in FVIIa-FXa Ternary Complex in POPC: POPS (CONTROL), TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CHOL), and Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CYSP) calculated from the 200ns simulation period of all the three complexes.

Figure 3-8. Schematic representation of the residues in contact with lipid (contact probability profile) of TF-FVIIa-FXa Ternary Complex in POPC: POPS(CONTROL), TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CHOL) and Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CYSP)

Figure 4-1. The interplay of enzymes in the process of fibrinolysis. Abbreviations used are FDPs, fibrin degradation products; PAI, plasminogen activator inhibitors; tPA, tissue plasminogen activator.

Figure 4-2. Flow chart of the virtual screening process using combined molecular docking and molecular dynamics simulation approaches.

Figure 4-3. The 2D structures of the lead-identified compounds.

Figure 4-4. The 2D interaction profile of lead molecules into the binding site of their respective protein targets 1PK2, 4CIK, and 5UGG.

Figure 4-5. RMSD profile of all the systems: 1PK2, 4CIK, and 5UGG.

Figure 4-6: Radius of gyration profile of all the systems: 1PK2, 4CIK, and 5UGG

Figure 4-6. Dynamics and interaction profile of GLA domain.

Figure 4-7: Hydrogen bonding profile of all the systems: 1PK2, 4CIK, and 5UGG.

Figure 4-8: Principal component analysis and MMPBSA analysis of all the systems: 1PK2, 4CIK, and 5UGG.

Figure 4-9: Schematic representation of the top three identified ligands docked to the active site residues of their respective protein targets using VMD.

- Figure 5-1. PD-1 and PD-L1 interaction of tumor cells and antigen-presenting cells (APC) with T cells that inhibit immune response. IFN γ help induces or maintains the expression of PD-L1. Anti-PD-L1 inhibits the interaction between PD-1 and PD-L1.
- Figure 5-2. The 2D structures of synthesized novel small-molecule inhibitors.
- Figure 5-3. The 2D interaction profile of novel identified ligands (M1-M10) docked into the binding site of their respective protein target 5J89.
- Figure 5-4: RMSD profile of 5J89 protein-inhibitor complexes [novel identified ligands (M1-M10) and reference ligand 6GX] along with APO.
- Figure 5-5. Figure 5-5: RMSF profile of 5J89 protein-inhibitor complexes [novel small-molecule inhibitors (M1-M10) and reference ligand 6GX] along with APO.
- Figure 5-6. Rg profile of 5J89 protein-inhibitor complexes [novel small-molecule inhibitors (M1-M10) and reference ligand 6GX] along with APO.
- Figure 5-7. SASA profile of 5J89 protein-inhibitor complexes [novel small-molecule inhibitors (M1-M10) and reference ligand 6GX] along with APO.
- Figure 5-8: Hydrogen bonding pattern of 5J89 protein-inhibitor complexes [novel identified ligands (M1-M10) and reference ligand 6GX] along with APO.
- Figure 5-9: Principal component analysis.
- Figure 5-10: Binding energy profile (ΔG) of 5J89 complexed with the novel small-molecule inhibitors (M1-M10) and reference ligand 6GX.
- Figure 5-11: Schematic representation of the top three novel small-molecule inhibitors bound to the active site residues of their respective protein target 5J89 using VMD.
- Figure 5-12: 3D representation of PDL1 protein dimer (chain C and chain D)

LIST OF ABBREVIATIONS

Activated factor VII (FVIIa)
Activated protein C (APC)
Adaptive biasing force (ABF)
Adsorption, Distribution, Metabolism and Excretion (ADME)
Bristol- Myers Squibb (BMS)
Catalytic triad (CT)
CHARMM (Chemistry at Harvard Macromolecular Mechanics)
Center-of-mass (COM)
Contact activation system (CAS)
Cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4)
Endothelial cell protein C receptor (EPCR)
Epidermal Growth Factor-like domain (EGF domain)
Factor VII (FVII)
Factor X (FX)
Factor XII (Hageman factor)
Full length TF (flTF)
Gamma-carboxyglutamic Acid-rich domain (GLA domain)
Gamma-aminobutyric acid antagonistic activity (GABAa)
Grid-based ligand docking with energetic (Glide)
Groingen Machine for Chemical Simulations (GROMACS)
GScore (GlideScore)
High Molecular Weight Kininogen (HMWK)
Hydrogen bonds (H-bonds)
Human protein C (hPC)
Interaction Energy (IE)
Kringle (KR)
Low-density lipoprotein (LDL)
Lysine-binding sites (LBSs)
Molecular Dynamics (MD)
Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA)
Multiple Sequence Alignment (MSA)
Optimized Potential Liquid Simulation (OPLS)

Partial thromboplastin time (PTT)
Particle Mesh Ewald (PME)
Pan-apple (PAp)
Phosphatidylcholine (PC)
Phosphatidylserine (PS)
Phospholipid (PL)
Plasminogen(Plg)
Plasmin (Plm)
Potential of mean force(PMF)
Programmed death-ligand 1 (PD-L1)
Prekallikrein (PK) Protein C (PC)
Principal component analysis (PCA)
Protein Data Bank (PDB)
Protein-Protein Interaction (PPI)
Radius of Gyration (Rg)
Root mean square deviation (RMSD)
Root mean square fluctuation (RMSF)
Solvent accessible surface area (SASA)
Serine Protease domain (SP domain)
Simple point-charge water model (SPCE)
Surface Plasmon Resonance (SPR)
Three-dimensional (3D)
Tissue Factor (TF)
Tissue-type plasminogen activator (tPA or PLAT)
Tranexamic acid (TXA)
Urokinase Plasminogen Activators (uPAs)
Virtual screening (VS)
Visual Molecular Dynamics (VMD)

Chapter 1

General Introduction

1. Overview

Hemostasis is an intricate physiological process that permits an organism to maintain blood fluidity [1]. It is regulated by a fine balance that exists between thrombogenic and anti-thrombogenic mechanisms present in the body. This highly conserved specialized process not only prevents and arrests bleeding/hemorrhage but also eliminates blood clots after the restoration of vascular integrity. Any imbalance of this process specifically in critical illness or diseased conditions might lead to problems associated with either thrombosis or bleeding. Blood coagulation is a host defense system that aids in maintaining the integrity of the closed, high-pressure mammalian circulatory system after blood vessel injury [2]. The equilibrium between the two processes namely blood coagulation and anticoagulation exclusively depends on an intricate interplay among a series of elements – the coagulation factors (mostly proteolytic enzymes), endothelium, and platelet. Certain plasma proenzymes are sequentially activated to their enzyme forms and thereafter proceed through either the intrinsic or extrinsic pathway of the blood coagulation process. These plasma glycoproteins, including factor XII, factor XI, factor IX, factor X, factor VII, and prothrombin are zymogens of serine proteases [3]. These proteins bear marked structural and functional homology to the digestive proteases' trypsin and chymotrypsin family. Each of these proteins is converted from an inactive enzyme to an active enzyme by limited proteolysis of one or two peptide bonds, however, this active form of the enzyme is short-lived. Most of the blood clotting enzymes are effective only when assembled in complexes on membrane surfaces with protein cofactors such as factor VIII and factor V [4]. Clot formation occurs when fibrinogen is converted by thrombin to fibrin, the structural protein that assembles into the fibrin polymer. The clot, formed after tissue injury, is composed of activated platelets and fibrin. The clot mechanically obstructs the flow of blood from the injured vessel thereby minimizing blood loss from the wound. The healing process is initiated after a stable clot has been formed. The clot is gradually dissolved by the enzymes of the fibrinolytic system [5]. Thus, coagulation is a dynamic process, linked with an enzymatic cascade reaction in

which cellular components and specific proteins play an important role in response to vascular injury. Nevertheless, this mechanism is firmly controlled, so it has a “protective effect”. The natural coagulation pathway strikes a delicate balance between the pro-coagulant pathway responsible for clot formation and its complement mechanism that inhibits the clot formation beyond the injury site. Abnormal blood coagulation may lead to several pathological conditions like ischemic stroke, unstable angina, deep-vein thrombosis, and other cardiovascular diseases [6].

2. Blood Coagulation Pathways

The coagulation cascade of hemostasis is split into two distinct pathways: the contact activation pathway (the intrinsic pathway) and the tissue factor pathway (the extrinsic pathway) Figure 1-1. Earlier research suggested that the two pathways of the coagulation cascade are equally important. Recent studies have established the prevalence of extrinsic pathways over intrinsic pathways [7].

The majority of clotting factors are generated inside the liver and present in blood in an inactive form (zymogenic form, precursors of proteolytic enzymes). Nomenclature for the activation of each zymogenic form is represented by the suffixing letter “a” to the particular zymogen Roman numeral. For example, if the factor VII (FVII) is converted into an activated form, the represented form is FVIIa. The principal function of coagulation factors is to trigger the formation of a blood clot when a blood vessel is injured. The clot is subsequently dissolved by another protease (known as a natural anticoagulant) [8]. These anticoagulants maintain the blood in the liquid fluid state within the circulation. An extremely delicate balance is required between these coagulation and fibrinolysis processes to maintain hemostasis. Any disturbance in either of these will cause problems associated with thrombosis or bleeding. A list of coagulant/anticoagulant factors and their function and disorders are mentioned in Table 1-1.

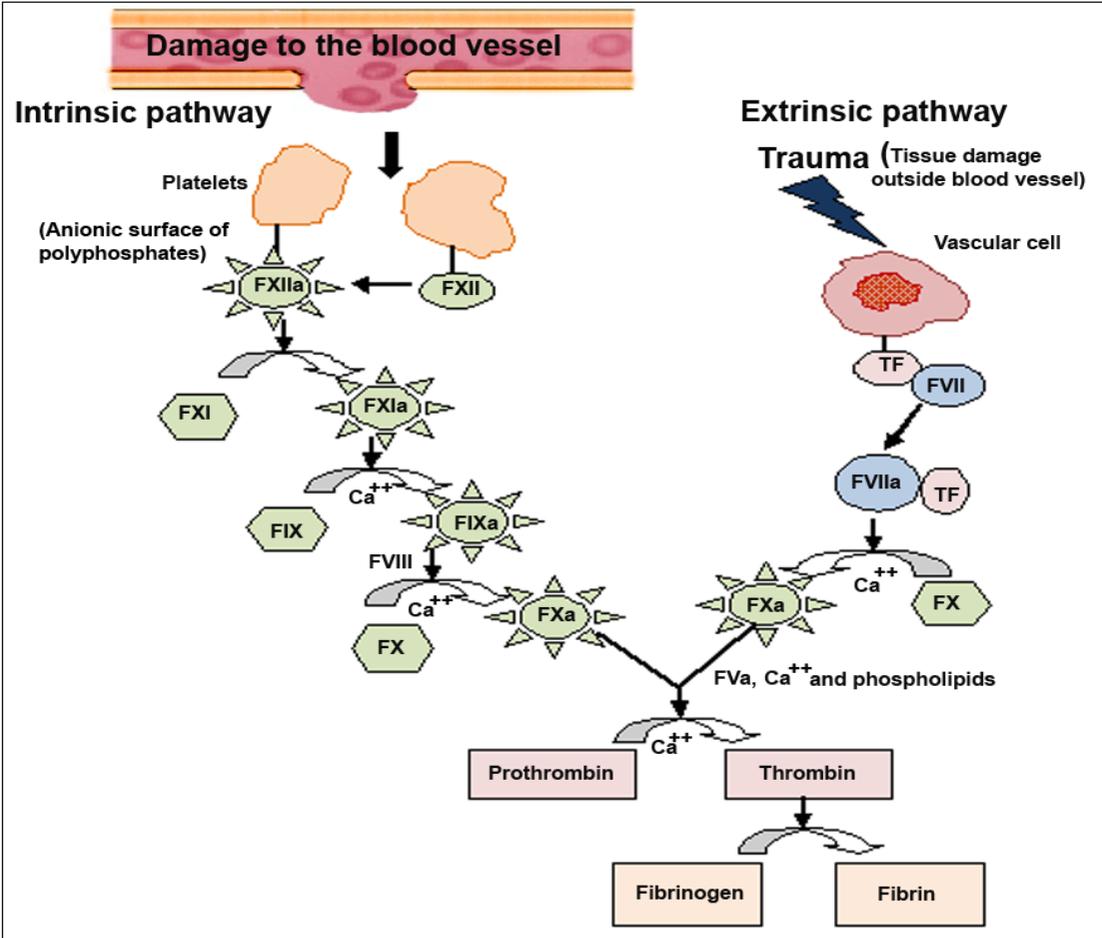


Figure 1-1. The blood coagulation pathway (Adapted from Bhattacharjee P, Bhattacharyya D. An insight into the abnormal fibrin clots—its pathophysiological roles. Fibrinolysis and thrombolysis. 2014 May 7:1-29.)

Table 1-1. List of coagulant/anticoagulant factors with their function, and the associated disease with it.

Number or Name	Function	Associated disease
I, Fibrinogen	Cleaved by thrombin to form fibrin monomers, which are then crosslinked by FXIIIa to form stable fibrin clots	Congenital Afibrinogenemia Familial Renal Amyloidosis
II, Prothrombin	Activated by the prothrombinase complex into thrombin, which then catalyzes the conversion of fibrinogen to fibrin	Thrombophilia
III, Tissue Factor	Cofactor or receptor for FVII/FVIIa initiates the the extrinsic pathway of the coagulation cascade.	Atherosclerosis
IV, Calcium	Required for coagulation factors to bind to phospholipid	Higher hematoma volume in patients with Intracerebral Hemorrhage.
V, Proaccelerin	Activated by thrombin into FVa which then participates in the formation of prothrombinase complex	Activated protein C resistance
VII, Proconvertin	Binds TF to form the TF- FVIIa complex to initiate the extrinsic pathway of coagulation, activates IX, X	Congenital proconvertin/factor VII deficiency
<u>VIII</u> , Antihemophilic factor A	cofactor of FIXa for the activation of FX, which it forms TENASE complex	Haemophilia A
IX, Antihemophilic factor B or Christmas factor	Activated by FXIa or FVIIa- TF into FIXa, which then activates FX in the presence of cofactor FVIIIa	Haemophilia B
X, Stuart-Prower factor	Activated by the TF-FVIIa complex or FIXa to form FXa, which then participates in the formation of the prothrombinase complex	Congenital Factor X deficiency
XI, Plasma thromboplastin antecedent (PTA)	Activated by FXIIa into FXIa, which then activates FIX	Haemophilia C

Number or Name	Function	Associated disease
XII, Hageman Factor	Activated by HMWK and PK to FXIIa, which then activates FXI to FXIa. It also binds to exposed collagen at the site of vessel injury	Hereditary Angioedema type III
XIII, Fibrin stabilizing factor (FSF)	Activated by thrombin into FXIIIa, which then crosslinks the fibrin monomers to form a more stable clot	Congenital Blood Coagulation Disorder
vWF, Von Willebrand factor	Binds to VIII, mediates platelet adhesion	Von Willebrand disease
PK, Prekallikrein (Fletcher factor)	Activates XII and prekallikrein, cleaves HMWK	Prekallikrein/Fletcher factor deficiency
HMWK, High Molecular Weight Kininogen	Supports reciprocal activation of XII, XI, and PK	Kininogen deficiency
Antithrombin III	Inhibits IIa, Xa, and other proteases	Antithrombin III deficiency
Protein C	Inactivates Va and VIIIa	Protein C deficiency
Protein S	Cofactor for activated protein C	Protein S deficiency
Protein Z	Mediates thrombin adhesion to phospholipids and stimulates degradation of FX by ZPI	Protein Z deficiency
Plasminogen	Converts to plasmin, lyses fibrin and other proteins	Plasminogen deficiency, type I (ligneous conjunctivitis)
tPA, tissue-plasminogen activator	Activates plasminogen	Thrombophilia

3. Intrinsic (contact activation) pathway

The contact activation pathway initiates from negatively charged surfaces with the formation of the primary complex by high molecular weight kininogen (HMWK), prekallikrein (PK), and factor XII (Hageman factor) and is subsequently activated when factor XII is transformed to factor XIIa (Figure 1-2). Firstly, factor XIIa activates factor XI and factor IX into factor XIa and factor IXa, respectively. Subsequently, factor IXa in presence of the factor VIIIa activates factor X (factor Xa). In the presence of the

cofactor factor Va, factor Xa triggers prothrombin to thrombin formation, and eventually, thrombin transforms soluble fibrinogen into insoluble fibrin meshwork [9].

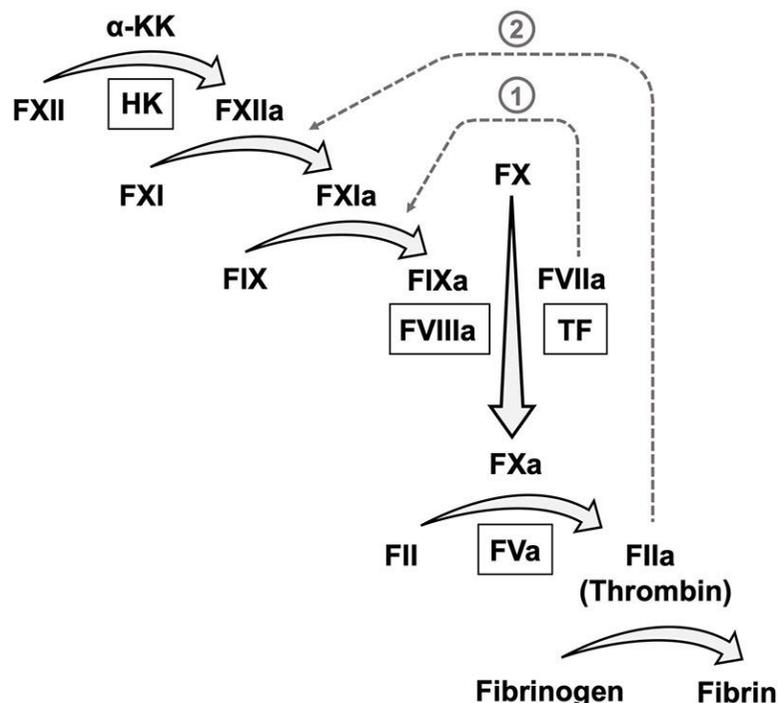


Figure 1-2. The intrinsic pathway of blood coagulation. (Adapted from Grover SP, Mackman N. *Intrinsic pathway of coagulation and thrombosis: Insights from animal models. Arteriosclerosis, Thrombosis, and Vascular Biology.* 2019 Mar;39(3):331-8)

The deficiency of coagulation components in the intrinsic pathway manifested hemorrhagic phenotypes in patients. The deficiency of factor VIII, and factor IX contributes to joints, muscles, and soft tissue bleeding abnormalities associated with hemophilia A and hemophilia B, respectively. Factor XI deficiency is related to a milder disorder like trauma or soft tissue-related hemorrhage, essentially involving tissues with high fibrinolytic activity [10]. Patients deficient in factor XII have noticeably extended partial thromboplastin time (PTT), however, such patients have no signs of bleeding abnormalities [11-12]. Modern shreds of evidence advocate that insufficiency of factor XII, prekallikrein (PK), and high-molecular-weight kininogen (HMWK) do not have a role in the hemostasis process. Nevertheless, a protein involved in the contact activation system (CAS) and kallikrein/kinin system (KKS) is associated with inflammation, blood pressure regulation, angiogenesis, and apoptosis processes [13].

4. Extrinsic (tissue-factor mediated) pathway

Initiation of extrinsic pathways requires an external factor (called tissue factor, TF) which is normally absent in the blood. Consequent to vascular injury, TF is exposed to circulatory zymogenic factor VII, eventually leading to the formation of the TF-FVIIa binary complex. This complex directly triggers consecutive activation of factor X and prothrombin into factor Xa and thrombin, respectively.

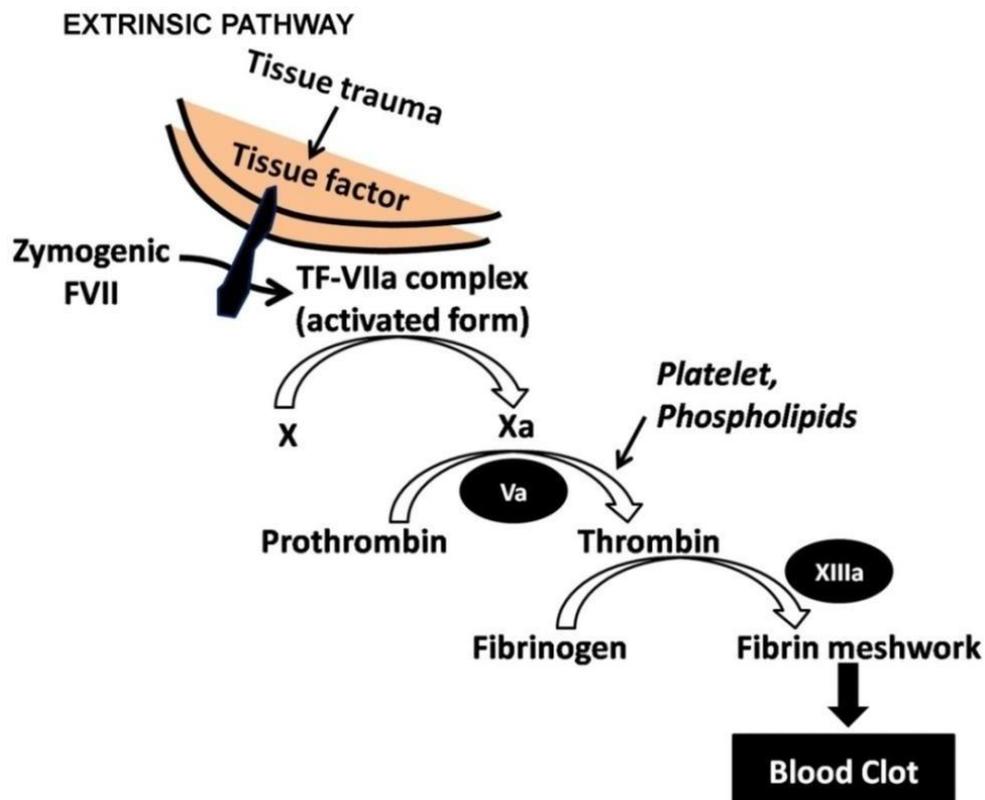


Figure 1-3. The extrinsic pathway of blood coagulation (Adapted from Grover SP, Mackman N. Intrinsic pathway of coagulation and thrombosis: Insights from animal models. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2019 Mar;39(3):331-8)

Concurrently, the small amounts of thrombin in presence of factor VIII (cofactor to factor IX) and factor V (cofactor to factor X), significantly increase the catalytic activity of factors IX and X on the platelet surface. Ultimately, thrombin (factor IIa) activation causes fibrin deposition which generates an impermeable platelet and fibrin plug at the site of injury (Figure 1-3). Both intrinsic and extrinsic pathways converge at the level of FXa generation [14-16].

5. Coagulation protein factor VII/factor VIIa

Factor VII is a member of the trypsin-like serine protease family, which is generated in the liver in zymogenic form (single-chain having 406 amino acids) and secreted into the bloodstream. As a result of proteolytic cleavage at position Arg152 and Ile153, FVII gets divided into two chains: heavy chain (254 residues) and light chain (152 residues), which are interlinked by a disulfide bond. The first active-site inhibitor bound crystal structure of sTF/FVIIa was reported by Banner et al. in 1996 (Figure 1-4) [17].

Structurally, FVIIa comprises four domains: GLA, EGF1, EGF2, and SP domain. The crystal structure of the binary complex has revealed that FVIIa binds seven Ca^{2+} ions in the GLA-domain, one in the EGF1 domain, and one in the protease domain (Figure 1-5).

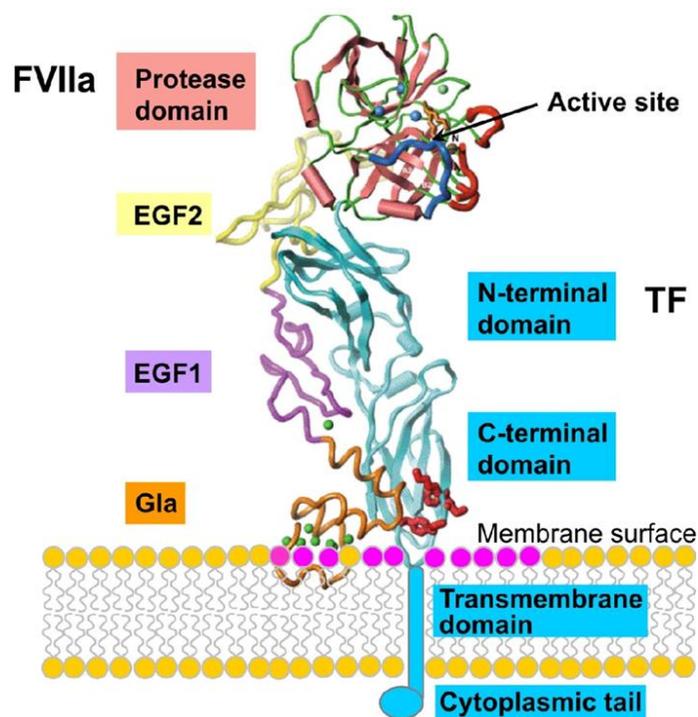


Figure 1-4 Schematic representation of the binary complex showing four domains of FVIIa: GLA, EGF1, EGF2, and SP domain in different colors with seven bound Ca^{2+} ions (shown in green) in the GLA-domain, one in the EGF1 domain, and one in the protease domain. Tissue Factor is shown in blue color ribbon representation with N-terminal and C-terminal domain, a transmembrane domain, and a cytoplasmic tail. (Adapted from Ke K. The tissue factor-factor vii (a) complex in blood coagulation. The University of Illinois at Urbana-Champaign; 2013.)

GLA domain: The ten Glu-residues within the first 45 residues of the GLA (gamma-carboxy glutamic acid-rich) domain of FVIIa are post-translationally modified to Gla-residue upon carboxylation by a vitamin K-dependent carboxylase [18-19]. Previous studies have shown that Ca^{2+} is required for the stability of the GLA domain thereby promoting its interaction with membrane and receptors. Three hydrophobic residues Phe4, Leu5, and Leu8 in the ω -loop of the GLA domain are referred to as the “keel”. The keel residues penetrate the lipid bilayer and also stabilize the GLA domain [20-21].

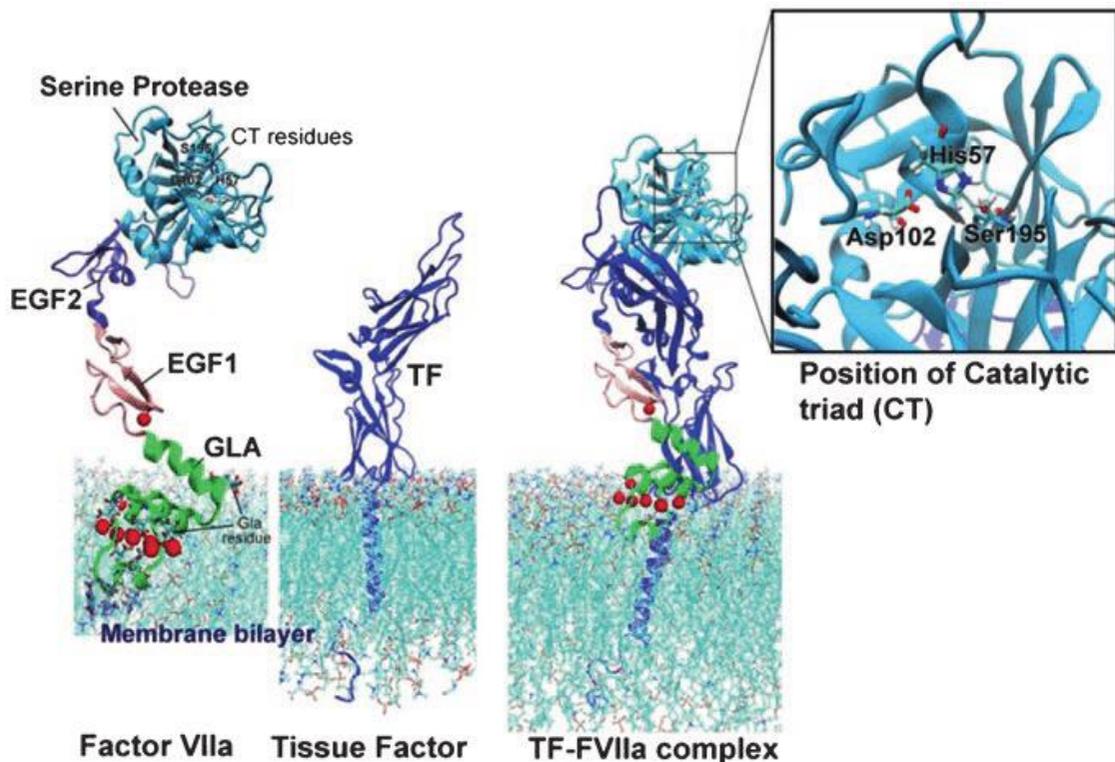


Figure 1-5. Three-dimensional structure of FVIIa, membrane-bound full length(fl) TF, and membrane-bound full length (fl) TF -FVIIa complex. Domains of FVIIa (GLA, EGF1, EGF2, and SP) are shown in different colors and fl TF in blue color new cartoon representation. Ca^{2+} ions bound to the FVIIa GLA domain and FVIIa EGF1 domain are shown in red color ball representation. The keel residues F4, L5, and L8 of FVIIa are shown in pink color stick representation. Positioning of CT (catalytic triad) residues, H57, S195, and D102 are shown in stick representation.

Vitamin K goes through oxidation and reduction reactions within the endoplasmic reticulum membrane and contributes electrons to activate specific proteins, carboxylase (Figure 1-6). The carboxylase enzyme performs gamma-carboxylation of glutamate groups in presence of CO_2 and molecular oxygen before getting reduced. Oxidized and diminished forms of vitamin K are the guiding factors for the gamma-carboxylation of

specific Glu residues. Vitamin K antagonists like warfarin hinder this post-translational modification of coagulation protein thereby preventing the formation of mature peptides. Consequently, it has medical importance in the use of anticoagulation therapy.

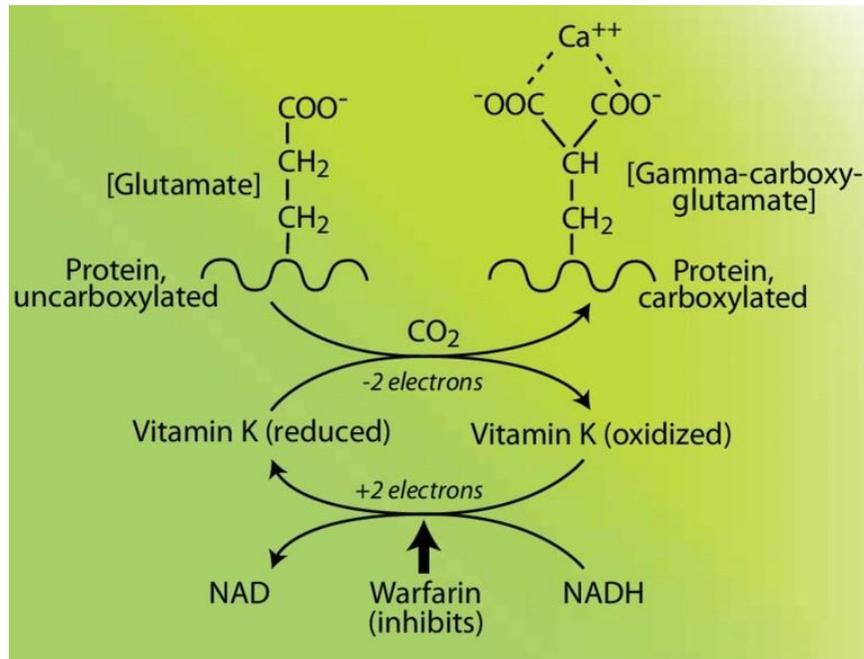


Figure 1-6. Vitamin K-dependent gamma-carboxylation of specific Glu (glutamic acid) residues [Adapted from Kidd PM. Vitamins D and K as pleiotropic nutrients: clinical importance to the skeletal and cardiovascular systems and preliminary evidence for synergy. Altern Med Rev. 2010 Sep 1;15(3):199-222].

EGF1 and EGF2 domain: Light chain of FVIIa is comprised of two epidermal growth factor-like domains. EGF domain is the most common domain found in the extracellular or membrane-embedded protein. It has been well established that these domains are rigid and are connected by disulfide bonds. In FVIIa, EGF1 and EGF2 domains comprise residues numbering from 47-83 and 87-128, respectively. EGF1 domain contains one high-affinity Ca^{2+} binding site and is responsible for TF binding. The EGF2 domain is located adjacent to SP (serine protease) domain.

SP domain: The core structure of the SP domain of FVIIa is similar to that of all trypsin-like serine proteases. It consists of 254 amino acids, containing specific regions like TF- binding site (170 loop region), catalytic triad (CT) region, activation loop 1, 2, 3, etc that are important for FVIIa function. Catalytic triad (CT) residues of serine protease include His57, Asp102, and Ser195 and are important to its catalytic function.

6. Tissue Factor

Tissue factor (TF) is a transmembrane glycoprotein that initiates coagulation in both physiological and pathological conditions and plays a pertinent role in maintaining homeostasis, thrombosis, and vascular development [22-23]. The Human TF gene is located on chromosome 1 (p21-22) spanning approximately 12.4 kilobases. Tissue factor is normally expressed on the surface of all non-vascular cells constitutively including vascular smooth muscle cells, pericytes, fibroblasts, etc [24]. Post-injury to the blood vessel, the endothelial layer disrupts, exposing TF to intravascular allosteric coagulation Factor VIIa (FVIIa) and forms a binary TF-FVIIa complex (Figure 1-7). Thereafter coagulation cascade gets triggered by inducing sequential activation of clotting proteins Factor IX and Factor X into factor IXa and FXa, respectively. This eventually causes thrombin burst, leading to the formation of clots followed by the generation of insoluble fibrin meshwork from soluble fibrinogen.

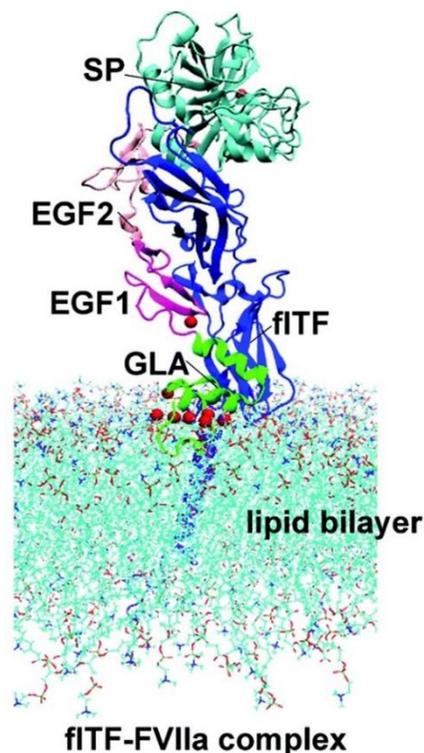


Figure 1-7. Three-dimensional structure of full-length TF-FVIIa complex on the membrane lipid bilayer.

Defective angiogenesis and embryonic lethality are traced to the absence of TF expression [25-28]. TF-FVIIa protease complex performs diverse functions associated with tumor cell progression and propagation apart from TF-mediated coagulation in the

extrinsic pathway. Metastasis is influenced by TF- FVIIa protease complex via multiple processes by inducing cellular signaling events. A broad range of cellular responses is regulated by such events, such as cell survival, gene transcription, and cytoskeletal changes which are crucial for a cell to adequately respond to its local environment and malignant transformation [29-32]. This event also induces the production of proteins that inhibits apoptosis thus creating a favorable environment for metastasis.

Structurally, TF consists of 263 amino acids having an extracellular part (1-219 amino acids), a transmembrane part (220-242), and a cytoplasmic tail part (243-263 amino acids) [33]. The extracellular part of TF consists of two immunoglobulin-like domains that share structural homology with fibronectin-type III protein. Two domains of TF are joined at an angle of 125 degrees and the cleft formed between the two domains constitutes the ligand-binding site [34]. TF contains two disulfide bonds at locations Cys⁴⁹-Cys⁵⁷ and Cys¹⁸⁶-Cys²⁰⁹ that help to maintain the structure of the extracellular domain. It is also known that the disulfide Cys¹⁸⁶-Cys²⁰⁹ bond located at the membrane-proximal region is important for TF cofactor activity [35].

7. Endothelial protein C receptor (EPCR)

Endothelial cell protein C receptor (EPCR) is generally located in the endothelium of large blood vessels. It is currently known as a multi-liganded and multifunctional receptor protein [36]. The primary ligand for EPCR is protein C /activated protein C (APC). The crystal structure of the human EPCR complex with human protein C (hPC) reveals that PC binds to hEPCR through its gamma-carboxyglutamic acid (GLA) domain with high affinity [37]. Other GLA-domain proteins, known to interact with EPCR are factor VIIa (activated FVII) and factor Xa (activated FX). FVIIa and FXa are vitamin K-dependent serine proteases that play a crucial role in blood coagulation [38-39].

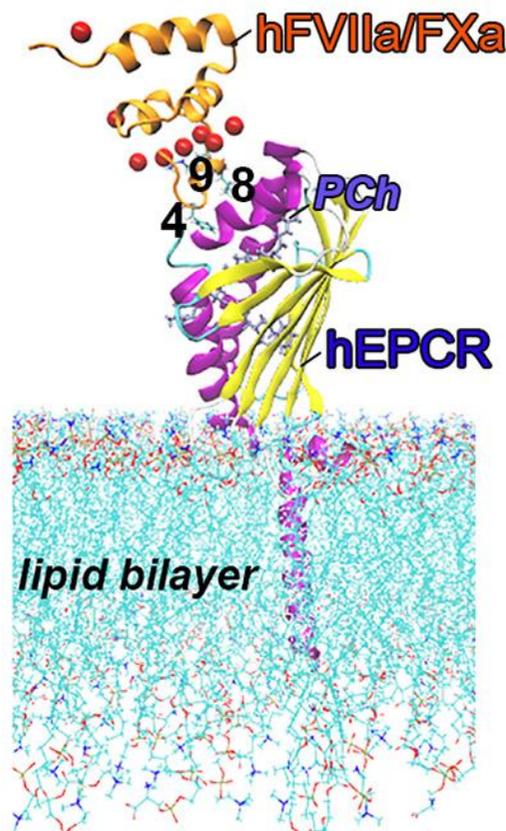


Figure 1-8. Three-dimensional structure of EPCR with GLA domain of FVIIa/FXa on the membrane lipid bilayer.

FVIIa GLA-domain has a high sequence identity with that of APC. FVIIa interacts with EPCR through its GLA-domain (Figure 1-8) [40]. The N-terminal region with residues 1-11 of the GLA-domain forms an omega-loop, which is known to contact the membrane surface. The keel residues 4, 9, and 8 of omega-loop form hydrophobic protrusions that play an essential role during the protein-lipid interaction [41]. Upon binding with EPCR, the coagulant activity of FVIIa gets reduced [42]. *In vivo* administration of FVIIa markedly provides a barrier protective effect in the presence of EPCR besides exhibiting anti-inflammatory and anti-apoptotic properties [43].

8. The motivation for the current work

It is well known for being homologous to protein C, Factor VIIa binds to EPCR through its GLA-domain. More importantly, the interaction study between EPCR-FVIIa has gained attention because of the usage of FVIIa in treating hemophilia patients [44]. The structures of the FVIIa and FXa GLA-domain reveal that there is a spatial disposition of the critical residues that are necessary for EPCR binding. Indeed, there is species specificity for FVIIa or FXa binding with EPCR [45]. Mouse FVIIa

(mFVIIa) cannot interact with EPCR; however, a single substitution of Leu4 with Phe4 in mFVIIa determines its interaction with EPCR [46]. In the case of hFXa, the residue present at position 4th of the GLA-domain is Phe; still, FXa cannot interact with EPCR. Unfortunately, the inability of hFXa to bind to hEPCR has not been explored to date. Therefore, conducting FVIIa-based EPCR studies could provide insight into understanding the structure-function relationship of these interacting proteins.

Previous studies have proposed that TF contributes significantly to balancing of initiation and propagation of thrombus on the rupture of atherosclerotic plaque [47]. Various studies confirm the presence of cholesterol/ oxidatively modified low-density lipoprotein (LDL) in atherosclerotic plaques and it has been found that the agglomeration and activity of TF are more in the lesions collected from patients with unstable angina, myocardial infarction, and other acute coronary events [48]. Previous reports suggest that TF-FVIIa coagulant activity at the cell surface may be affected by various processes including changes in cholesterol content and post-translational modifications of TF. Numerous studies were conducted but yielded inconclusive results about the effect of cholesterol on TF expression [49-50]. Therefore, it is very important to unveil the underlying molecular mechanism of how cholesterol affects structural modulations on the TF-FVIIa-factor Xa(FXa) Ternary complex.

The zymogen protease Plasminogen (Plg) and its active form plasmin (Plm) carry out important functions in the blood clot disintegration (breakdown of fibrin fibers) process [51]. Inhibition of plasmin effectively reduces fibrinolysis to evade heavy bleeding. Irregular balance of the Plg/Plm system and immoderate fibrinolytic activity leads to fatal hemorrhagic disorders, thrombotic vascular injury, and severe complications during general surgery or major trauma [52]. Presently, available Plm inhibitor tranexamic acid (TXA) that is used to treat severe hemorrhages is associated with an increased incidence of seizures which in turn were traced to gamma-aminobutyric acid (GABA) antagonistic activity in addition to having multiple side effects like headaches, nasal symptoms, or back, abdominal, and muscle pain [53]. Therefore, the development of new antifibrinolytic agents against Plg or Plm may help in preventing the disintegration of the blood clot and rampant blood loss.

Immune checkpoint therapy is currently considered to be one of the most promising strategies in the field of cancer immunotherapy [54]. Cancer cells can exploit

immune checkpoints as a way to evade immune detection and elimination by distorting its co-stimulatory and coinhibitory molecules [55]. Previous studies suggest that the interaction of programmed death-ligand 1 (PD-L1)/PD-1 in the tumor microenvironment promotes cancer development and progression by enhancing tumor cell proliferation and survival [56]. Presently available treatments targeting PD-1/PD-L1 in the market like monoclonal antibodies (mAbs) exhibit high specificity and are ideal for blocking protein–protein interaction, but some shortcomings exist like small half-life, less oral bioavailability, low stability, high manufacturing cost, immune-related adverse effects, and incomplete response to cancer patients [57-58]. Therefore, the development of such novel small-molecule inhibitors as potent anti-cancer agents is beneficial for medical science.

9. Significance of the present thesis work

This thesis provides relevant information to understand the precise mechanisms of FVIIa binding to EPCR, which may facilitate the designing of FVIIa-mutant molecules for improved therapeutic use associated with bleeding disorders and septic shock. Our study can provide detailed insight into differences in species-specific binding of FVIIa GLA-domain and FXa GLA-domain towards EPCR. From our study, we can understand how mutation at the residue positions 4th, 8th, and 9th interfere with the energy profile of the FVIIa GLA-domain bound to EPCR. We explore how point mutations in GLA-domain affect the complex formation (FVIIa-EPCR) in the presence of phosphatidylcholine (PCh) in the EPCR groove. Lastly, our studies may provide valuable insights into the prediction and identification of protein-lipid interactions thus augmenting advances in understanding structural modulation and function of proteins.

The effect of cholesterol content and post-translational modifications of Tissue Factor (TF) on TF-FVIIa (Factor VIIa) coagulant activity is still indeterminate. Various studies suggest that the presence of cholesterol/LDL in atherosclerotic plaques confers high importance to atherogenesis. Aberrant palmitoylation is related to various diseases including neurological disorders like Huntington's disease, Parkinson's disease, Alzheimer's disease, metabolic disorders, and Cancer. In this thesis work, we have tried to characterize the dynamics and interactions of cholesterol with the TF-FVIIa-FXa Ternary complex. Additionally, we also attempted to assess the effect of Palmitoylation, on protein-lipid interactivity in presence of cholesterol. Overall, our

studies may throw light on the inconclusive results about the effect of cholesterol and TF palmitoylation on TF-FVIIa (Factor VIIa) coagulant activity.

The zymogen protease Plasminogen (Pgn) and its active form plasmin (Plm) carry out important functions in blood clot disintegration, tissue remodeling, cell migration, and bacterial pathogenesis. Irregular balance of plasminogen/plasmin system ensues in fatal hemorrhagic disorders or thrombotic vascular occlusion. The development of potent antifibrinolytic agents against plasmin will prevent blood clot dissolution and abnormal blood loss. In this thesis work, we attempt to find potential antifibrinolytic inhibitors having both maximum specificity and efficacy with minimum side effects.

An encouraging number of studies report that targeting the PD-L1/ PD-1 immune checkpoint promotes an effective immune response against cancer cells and may be of significant therapeutic value [16]. Therefore, designing and synthesizing small molecules is a promising approach to the development of novel and inexpensive immune checkpoint inhibitors for PD-L1/ PD-1 [25]. Furthermore, in combination with the mAbs, novel small-molecule inhibitors are contemplated to attain synergistically anticancer effects [26]. Hence, a novel class of small-molecule inhibitors was designed and synthesized to prevent the binding of PD-L1 to its receptor PD-1. In this thesis work, as a part of a collaborative project, an evaluation of these in-house synthesized small molecule inhibitors was carried out using *in silico* approaches.

References

1. LaPelusa A, Dave HD. Physiology, Hemostasis.
2. Furie B, Furie BC. The molecular basis of blood coagulation. *Cell*. 1988 May 20;53(4):505-18.
3. Mac Farlane RG. An enzyme cascade in the blood clotting mechanism, and its function as a biological amplifier. *Nature*. 1964:202-498.
4. Butenas S, Mann KG. Blood coagulation. *Biochemistry (Moscow)*. 2002 Jan;67(1):3-12.
5. Lijnen HR. Elements of the fibrinolytic system. *Annals of the New York Academy of Sciences*. 2001 Jun;936(1):226-36.
6. Palta S, Saroa R, Palta A. Overview of the coagulation system. *Indian journal of anesthesia*. 2014 Sep;58(5):515.
7. Mackman N, Tilley RE, Key NS. Role of the extrinsic pathway of blood coagulation in hemostasis and thrombosis. *Arteriosclerosis, thrombosis, and vascular biology*. 2007 Aug 1;27(8):1687-93.
8. Rosenberg RD, Rosenberg JS. Natural anticoagulant mechanisms. *The Journal of clinical investigation*. 1984 Jul 1;74(1):1-6.
9. Grover SP, Mackman N. Intrinsic pathway of coagulation and thrombosis: Insights from animal models. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2019 Mar;39(3):331-8.
10. Salomon O, Steinberg DM, Seligshon U. Variable bleeding manifestations characterize different types of surgery in patients with severe factor XI deficiency enabling parsimonious use of replacement therapy. *Haemophilia*. 2006 Sep;12(5):490-3.
11. Lämmle B, Wuillemin WA, Huber I, Krauskopf M, Zürcher C, Pflugshaupt R, Furlan M. Thromboembolism and bleeding tendency in congenital factor XII deficiency-a study on 74 subjects from 14 Swiss families. *Thrombosis and haemostasis*. 1991;65(02):117-21.
12. Kaplan, Allen P. Intrinsic coagulation, thrombosis, and bleeding. *Blood* 1996; 87: 2090-2090.

13. Loreto C, La Rocca G, Anzalone R, Caltabiano R, Vespasiani G, Castorina S, Ralph DJ, Cellek S, Musumeci G, Giunta S, Djinovic R. The role of intrinsic pathway in apoptosis activation and progression in Peyronie's disease. *BioMed research international*. 2014 Jan 1;2014.
14. Mann KG. Biochemistry and physiology of blood coagulation. *Thrombosis and haemostasis*. 1999;82(08):165-74.
15. Furie B, Furie BC. Molecular and cellular biology of blood coagulation. *New England Journal of Medicine*. 1992 Mar 19;326(12):800-6.
16. Dahlbäck B. Blood coagulation. *The Lancet*. 2000 May 6;355(9215):1627-32.
17. Banner DW, D'Arcy A, Chène C, Winkler FK, Guha A, Konigsberg WH, Nemerson Y, Kirchhofer D. The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature*. 1996 Mar;380(6569):41-6.
18. Gröber U, Reichrath J, Holick MF, Kisters K. Vitamin K: an old vitamin in a new perspective. *Dermato-endocrinology*. 2014 Jan 1;6(1):e968490.
19. Vermeer C. Gamma-carboxyglutamate-containing proteins and the vitamin K-dependent carboxylase. *Biochemical Journal*. 1990 Mar 15;266(3):625.
20. Ohkubo YZ, Tajkhorshid E. Distinct structural and adhesive roles of Ca²⁺ in membrane binding of blood coagulation factors. *Structure*. 2008 Jan 1;16(1):72-81.
21. Falls LA, Furie BC, Jacobs M, Furie B, Rigby AC. The ω -loop region of the human prothrombin γ -carboxyglutamic acid domain penetrates anionic phospholipid membranes. *Journal of Biological Chemistry*. 2001 Jun 29;276(26):23895-902.
22. Drake TA, Morrissey JH, Edgington TS. Selective cellular expression of tissue factor in human tissues. Implications for disorders of hemostasis and thrombosis. *The American journal of pathology*. 1989 May;134(5):1087.
23. Contrino J, Hair G, Kreutzer DL, Rickles FR. In situ detection of tissue factor in vascular endothelial cells: correlation with the malignant phenotype of human breast disease. *Nature medicine*. 1996 Feb;2(2):209-15.
24. Pawlinski R, Mackman N. Cellular sources of tissue factor in endotoxemia and sepsis. *Thrombosis research*. 2010 Apr 1;125:S70-3.

25. Bugge TH, Xiao Q, Kombrinck KW, Flick MJ, Holmbäck K, Danton MJ, Colbert MC, Witte DP, Fujikawa K, Davie EW, Degen JL. Fatal embryonic bleeding events in mice lacking tissue factor, the cell-associated initiator of blood coagulation. *Proceedings of the National Academy of Sciences*. 1996 Jun 25;93(13):6258-63.
26. Carmeliet P, Mackman N, Moons L, Luther T, Gressens P, Van Vlaenderen L, Demunck H, Kasper M, Breier G, Evrard P, Müller M. Role of tissue factor in embryonic blood vessel development. *Nature*. 1996 Sep;383(6595):73-5.
27. Mackman N. Role of tissue factor in hemostasis, thrombosis, and vascular development. *Arteriosclerosis, thrombosis, and vascular biology*. 2004 Jun 1;24(6):1015-22.
28. Mackman N. The role of tissue factor and factor VIIa in hemostasis. *Anesthesia and analgesia*. 2009 May;108(5):1447.
29. Versteeg HH, Spek CA, Slofstra SH, Diks SH, Richel DJ, Peppelenbosch MP. FVIIa: TF induces cell survival via G12/G13-dependent Jak/STAT activation and BclXL production. *Circulation research*. 2004 Apr 30;94(8):1032-40.
30. Åberg M, Eriksson O, Mokhtari D, Siegbahn A. Tissue factor/factor VIIa induces cell survival and gene transcription by transactivation of the insulin-like growth factor 1 receptor. *Thrombosis and haemostasis*. 2014;111(04):748-60.
31. Antoniak S, Sparkenbaugh E, Pawlinski R. Tissue factor, protease activated receptors and pathologic heart remodelling. *Thrombosis and haemostasis*. 2014 Nov;112(11):893-900.
32. Versteeg HH, Spek CA, Peppelenbosch MP, Richel DJ. Tissue factor and cancer metastasis: the role of intracellular and extracellular signaling pathways. *Molecular medicine*. 2004 Jan;10(1):6-11.
33. Harlos K, Martin DM, O'Brien DA, Jones EY, Stuart DI, Polikarpov I, Miller A, Tuddenham EG, Boys CW. Crystal structure of the extracellular region of human tissue factor. *Nature*. 1994 Aug;370(6491):662-6.
34. Rehemtulla A, Ruf W, Edgington TS. The integrity of the cysteine 186-cysteine 209 bond of the second disulfide loop of tissue factor is required for binding of factor VII. *Journal of Biological Chemistry*. 1991 Jun 5;266(16):10294-9.

35. Bach RR. Tissue factor encryption. *Arteriosclerosis, thrombosis, and vascular biology*. 2006 Mar 1;26(3):456-61.
36. Mohan Rao LV, Esmon CT, Pendurthi UR. Endothelial cell protein C receptor: a multiliganded and multifunctional receptor. *Blood, The Journal of the American Society of Hematology*. 2014 Sep 4;124(10):1553-62.
37. Fukudome K, Esmon CT. Identification, cloning, and regulation of a novel endothelial cell protein C/activated protein C receptor. *Journal of Biological Chemistry*. 1994 Oct 21;269(42):26486-91.
38. Huang M, Rigby AC, Morelli X, Grant MA, Huang G, Furie B, Seaton B, Furie BC. Structural basis of membrane binding by Gla domains of vitamin K–dependent proteins. *Nature Structural & Molecular Biology*. 2003 Sep;10(9):751-6.
39. Mizuno H, Fujimoto Z, Atoda H, Morita T. Crystal structure of an anticoagulant protein in complex with the Gla domain of factor X. *Proceedings of the National Academy of Sciences*. 2001 Jun 19;98(13):7230-4.
40. Esmon CT. The endothelial protein C receptor. *Current opinion in hematology*. 2006 Sep 1;13(5):382-5.
41. LÓPEZ- SAGASETA J, Montes R, Puy C, Diez N, Fukudome K, Hermida J. Binding of factor VIIa to the endothelial cell protein C receptor reduces its coagulant activity. *Journal of Thrombosis and Haemostasis*. 2007 Sep;5(9):1817-24.
42. Kondreddy V, Wang J, Keshava S, Esmon CT, Rao LV, Pendurthi UR. Factor VIIa induces anti-inflammatory signaling via EPCR and PAR1. *Blood, The Journal of the American Society of Hematology*. 2018 May 24;131(21):2379-92.
43. Ghosh, S.; Pendurthi, U. R.; Steinoe, A.; Esmon, C. T.; Rao, L. V. M. Endothelial Cell Protein C Receptor Acts as a Cellular Receptor for Factor VIIa on Endothelium. *J Biol Chem* 2007, 282 (16), 11849–11857.
44. Pendurthi UR, Rao LV. Factor VIIa interaction with endothelial cells and endothelial cell protein C receptor. *Thrombosis research*. 2010 Apr 1;125:S19-22.
45. Sen P, Clark CA, Gopalakrishnan R, Hedner U, Esmon CT, Pendurthi UR, Rao LV. Factor VIIa binding to endothelial cell protein C receptor: differences between mouse and human systems. *Thrombosis and haemostasis*. 2012;107(05):951-61.

46. Pavani G, Zintner SM, Ivanciu L, Small JC, Stafford KA, Szeto JH, Margaritis P. One amino acid in mouse activated factor VII defines its endothelial protein C receptor (EPCR) binding and modulates its EPCR- dependent hemostatic activity in vivo. *Journal of Thrombosis and Haemostasis*. 2017 Mar;15(3):507-12.
47. Ardissino D, Merlini PA, Arlens R, Coppola R, Bramucci E, Lucreziotti S, Repetto A, Fetiveau R, Mannucci PM. Tissue factor in human coronary atherosclerotic plaques. *Clinica chimica acta*. 2000 Feb 15;291(2):235-40.
48. Mandal SK, Iakhiaev A, Pendurthi UR, Rao LVM. Acute cholesterol depletion impairs functional expression of tissue factor in fibroblasts: Modulation of tissue factor activity by membrane cholesterol. *Blood* 2005; 105: 153–60.
49. Petit L, Lesnik P, Dachet C, Moreau M, Chapman MJ. Tissue factor pathway inhibitor is expressed by human monocyte-derived macrophages: Relationship to tissue factor induction by cholesterol and oxidized LDL. *Arterioscler Thromb Vasc Biol* Lippincott Williams and Wilkins; 1999; 19: 309–15.
50. Van Den Eijnden MMED, Van Noort JT, Hollaar L, Van Der Laarse A, Bertina RM. Cholesterol or triglyceride loading of human monocyte-derived macrophages by incubation with modified lipoproteins does not induce tissue factor expression. *Arterioscler Thromb Vasc Biol* Lippincott Williams and Wilkins; 1999; 19: 384–92.
51. Chapin JC, Hajjar KA. Fibrinolysis and the control of blood coagulation. *Blood reviews*. 1;29(1):17-24.
52. Colomina MJ, Méndez E, Sabate A. Altered fibrinolysis during and after surgery. In *Seminars in Thrombosis and Hemostasis* 2021 Apr 20. Thieme Medical Publishers, Inc.
53. Nishida T, Kinoshita T, Yamakawa K. Tranexamic acid and trauma-induced coagulopathy. *Journal of Intensive Care*. 2017 Dec;5(1):1-7.
54. Beatty GL, Gladney WL. Immune escape mechanisms as a guide for cancer immunotherapy. *Clinical cancer research*. 2015 Feb 15;21(4):687-92.
55. Zappasodi R, Merghoub T, Wolchok JD. Emerging concepts for immune checkpoint blockade-based combination therapies. *Cancer cell*. 2018 Apr 9;33(4):581-98.

56. Han Y, Liu D, Li L. PD-1/PD-L1 pathway: current researches in cancer. *American journal of cancer research*. 2020;10(3):727.
57. Konstantinidou M, Zarganes- Tzitzikas T, Magiera- Mularz K, Holak TA, Dömling A. Immune Checkpoint PD- 1/PD- L1: Is There Life Beyond Antibodies?. *Angewandte Chemie International Edition*. 2018 Apr 23;57(18):4840-8.
58. Wang T, Wu X, Guo C, Zhang K, Xu J, Li Z, Jiang S. Development of inhibitors of the programmed cell death-1/programmed cell death-ligand 1 signaling pathway. *Journal of Medicinal Chemistry*. 2018 Sep 24;62(4):1715-30.

Chapter 2

In silico mutation and binding studies of human FVIIa GLA-domain to endothelial protein C receptor: A molecular dynamics simulation approach

1. Introduction

Endothelial cell protein C receptor (EPCR) is generally located in the endothelium of large blood vessels. It is currently known as a multi-liganded and multifunctional receptor protein [1]. The primary ligand for EPCR is protein C/activated protein C (APC). The crystal structure of the human (h) EPCR complex with protein C (PC) reveals that PC binds to hEPCR through its gamma-carboxy glutamic acid-rich (GLA) domain with high affinity [2]. Other GLA-domain proteins, known to interact with EPCR are factors(F) VIIa, and Xa. FVII and FX are vitamin K-dependent serine proteases, produced by the liver and present in blood in zymogenic form. The activation of each zymogen is represented by the suffixing letter “a” to the corresponding Roman numeral, for example, FVIIa (activated FVII) and FXa (activated FX). After activation, these enzymes participate in the blood coagulation process [3-5]. FVIIa GLA-domain has a high sequence identity with that of APC. FVIIa also interacts with EPCR through its GLA domain [6]. The N-terminal region with residues 1-11 of the GLA-domain forms an omega-loop, which is also known to contact the membrane surface. The keel residues 4, 5, and 8 of omega-loop form hydrophobic protrusions that play an essential role during the protein-lipid interaction [4-5]. Upon binding with EPCR, the coagulant activity of FVIIa gets reduced [7]. *In vivo* administration of FVIIa markedly provides a barrier protective effect in the presence of EPCR besides exhibiting anti-inflammatory and anti-apoptotic properties [8]. More importantly, the interaction study between EPCR-FVIIa has gained attention because of the usage of FVIIa in treating hemophilia patients [9-10].

The binding affinity of both hPC and hFVIIa with hEPCR is similar. In contrast, mouse (m) FVII does not significantly interact with both mEPCR and hEPCR. Human FXa (hFXa) is highly homologous to the protein C and FVIIa GLA-domain; however, any noticeable binding of hFXa to hEPCR is not seen [11-13]. The molecular mechanism for this remains unclear.

Several groups have shown that FXa binds to endogenously expressed EPCR, and induces cellular signaling through cleavage of protease-activated receptor-1 (PAR-1) [14]. Cellular signaling by FXa has been associated with various conditions like wound healing, atherosclerosis, cancer dissemination, angiogenesis, etc. FXa also provides a barrier protective effect against VEGF-induced vascular leakage that is independent of EPCR [1,15]. However, the binding of hFXa to hEPCR is not well studied. The structures of the FVIIa and FXa GLA-domain reveal that there is a spatial disposition of the critical residues that are necessary for EPCR binding. Indeed, there is species specificity for FVIIa or FXa binding with EPCR. Mouse FVIIa (mFVIIa) cannot interact with EPCR; however, a single substitution of Leu4 to Phe4 in mFVIIa determines its interaction with EPCR [12,16].

In the case of hFXa, the residue present at position 4th of the GLA-domain is Phe; still, FXa cannot interact with EPCR. Unfortunately, the inability of FXa to bind to EPCR has not been explored to date. In this study, we hypothesized that the residue position of the GLA-domain at 8th and 9th might play a crucial role in determining EPCR binding; therefore we have considered six successive mutations of FVIIa GLA-domain based on mouse FVIIa and human FXa as provided in Table 2-1. *In silico* mutation modeling, molecular simulations, and the potential of mean force calculations can provide detailed insight into differences in species-specific binding of FVIIa GLA-domain and FXa GLA-domain towards EPCR. In this study, we have taken membrane-bound full-length EPCR complex either with wild-type hFVIIa/ hFXa GLA-domain or mutated GLA-domain (Table 2-1).

Table 2-1. Details of *in silico* mutation associated with human FVIIa GLA-domain.

S/N	Wild-type/mutation system	Mutating residue at the position	Mutation based on the species GLA-domain
1.	FVIIa (wild-type)	-	Human FVIIa
2.	FVIIa (F4L)	4	Mouse FVIIa
3.	FVIIa (R9W)	9	Mouse FVIIa
4.	FVIIa (F4L; R9W)	4, 9	Mouse FVIIa
5.	FXa (wild-type)	-	Human FXa
6.	FVIIa (L8M)	8	Human FXa
7.	FVIIa (R9K)	9	Human FXa
8.	FVIIa (L8M; R9K)	8, 9	Human FXa

Eventually, we performed molecular dynamics (MD) simulation to characterize the distinct binding of wild-type (WT) FVII_a_{WT}, all FVIIa mutants, and FX_a_{WT} towards EPCR and explored the changes in protein-protein interactions concerning FVII_a_{WT} in the presence of phosphatidylcholine (PCh) within EPCR groove. Human FVII_a_{WT} binds to hEPCR in a true ligand fashion; however when the hFVII_a_{WT} GLA-domain was mutated to FVII_a_{F4L; R9W} at positions 4th and 9th, the fluctuation in the GLA-domain is higher. The fluctuation in the GLA-domain is also higher for all FVII_a_{L8M}, FVII_a_{R9K}, and FVII_a_{L8M; R9K} systems. Mutation at the residue positions 4th, 8th, and 9th interfere with the energy profile of the FVIIa GLA-domain that is bound to the EPCR system. We discuss how point mutations in GLA-domain affect the complex formation (FVIIa-EPCR) in the presence of PCh in the EPCR groove. Our studies also reveal that lipid-mediated interaction with Phe4 residue (GLA-domain) further facilitates the intermolecular H-bonds between EPCR and GLA-domain for wild-type FVIIa system. These findings may have important implications in understanding the precise mechanisms of FVIIa binding to EPCR, which may facilitate the designing of FVIIa-mutant molecules for the improved therapeutic use associated with bleeding disorders and septic shock [8].

2. Computational Section

2.1 System Setup.

We have prepared membrane-bound full-length EPCR with FVIIa/FXa/all mutant GLA-domain systems. In this study, we constructed 1-palmitoyl-2-oleoyl-sn-glycero-3-PC (POPC) and 1-palmitoyl-2-oleoyl-sn-glycero-3-PS (POPS) lipid binary models in 4:1 concentrations. The rectangular lipid bilayer models, POPC: POPS were generated using the CHARMM-GUI membrane builder server, containing a total of 200 lipids (100 per leaflet), as described in the previous work [17-20]. During membrane generation, the system charge was neutralized by randomly placing Na⁺ and Cl⁻ ions in the bulk water. Equilibration of the bilayer was done using the following CHARMM input files in the NPT ensemble (constant pressure and temperature) for 8–10 ns. We found that our lipid bilayer is equilibrated with an average area per lipid bilayer (~61 Å²) and bilayer thickness (40 Å) at the end of the simulation. The last membrane coordinates were used for modeling membrane-bound fIEPCR with either hFVIIa/hFXa or all mutant GLA-domain in the complex system.

2.2 Protein modeling.

The initial coordinates of the extracellular domain of EPCR (residues 7-178) were taken from the X-ray crystallographic structure of the EPCR (PDB ID: 1LQV) [21]. Since full length (fl) of EPCR crystal structure is not available; therefore missing residues of EPCR (those are absent in the crystal structure) were constructed using template-based search in protein BLAST [17]. Initially, N-terminal amino acids of residues 1-6 (sequence SQDASD) of EPCR were modeled based on protein BLAST template search having PDB ID: 2Y3A from chain A (residue numbering 44-50) [22]. Similarly, modeling of the C-terminal of missing residues of the extracellular region from residues 179-193 (sequence AENTKGSQTSRSYTS) was carried out using protein BLAST template search (with >40% identity) having PDB ID: 1DDB, from chain A (residues 55-69, sequence ELQTDGSQASRSFNQ) [23]. Three polypeptides: residues 1-6, 7-178, and 179-193 were joined using discovery studio with a proper alignment that forms extracellular EPCR (193 amino acids).

Similarly, the transmembrane region (length 21 amino acids) from residue numbering 194-214 (sequence LVLGVLVGSFIIAGVAVGIFL) was constructed based on the template (PDB ID: 4O6Y, the crystal structure of cytochrome b561, chain A, residues numbering 198-218) having highest protein BLAST score using Modeller [24-25]. To construct membrane-bound flEPCR, we have accomplished the following criteria as follows: firstly, the extracellular region of EPCR (193 amino acids) was placed perpendicular to the equilibrated lipid bilayer model. After that, alignment of a transmembrane region of EPCR was implemented concerning the bilayer-normal (Z-axis). The C-terminal region of the extracellular portion (residue number 193) to the N-terminal portion of the EPCR transmembrane region (residues 194-214) were joined. The remaining residues (215-221, sequence CTGGRRC) comprising the C-terminal cytoplasmic tail region were joined using Pymol software [26]. Lastly, the cytoplasmic tail part (7 amino acids) was attached to the remaining part of EPCR (extracellular and transmembrane portion) using Discovery studio software.

For the mutant GLA-domain, following substitutions in the FVIIa GLA-domain based on mouse FVIIa (FVIIa_{F4L}, FVIIa_{R9W}, FVIIa_{F4L; R9W}) and human FXa (FVIIa_{L8M}, FVIIa_{R9K}, FVIIa_{L8M; R9K}) were performed as shown Table 2-1. Initial coordinates for the FVIIa GLA-domain (residues 1-50) were taken from the X-ray crystallographic structure of the sTF-FVIIa binary complex (PDB ID: 1DAN) [27]. The molecule (GLA-domain) was aligned and superimposed based on the crystal structure of protein

mutations were carried out based on mouse FVIIa and human FXa. Residues whose mutations are based on mouse FVIIa (FVIIa_{F4L}, FVIIa_{R9W}, and FVIIa_{F4L; R9W}) are shown in blue color. Other residues whose mutations are based on human FXa (FVIIa_{L8M}, FVIIa_{R9K}, and FVIIa_{L8M; R9K}) are shown in green color.

Recently, it was shown that the phosphatidylcholine (PCh) molecule is the major phospholipid that resides in the antigen-binding groove of EPCR [30]. Therefore, we considered this molecule in the antigen-binding groove of EPCR. Construction of the PCh molecule was done using CHARMM membrane builder, and the superimposition of the molecule was performed from the previously reported crystalline PTY molecule, which was done to fit the lipid molecule to its correct position, that was used in our previous simulations [31].

2.3 Molecular Dynamics Simulations Protocol

GLA-domain contains several (approx 10) post-translational modified groups known as gamma-carboxyglutamate residue (Gla residue), therefore both topology and parameter files of Gla residue were prepared based on the similar chemical groups present within topology and parameter files. After that, a protein structure file (PSF) was produced using the Psfgen package tool within VMD for each of the systems. Each system was solvated using the TIP3P water model in a periodic box having dimensions of 128x128x182 Å [32]. Autoionize plugin of VMD was employed to achieve 5mM calcium ionic concentration by randomly replacing water molecules with Ca²⁺ and Cl⁻ ions.

Atomistic simulations were performed by employing NAMD 2.9 dynamics software package with CHARMM36 force field parameter [33-34]. Total potential energy functions (U_{total}) used in NAMD simulations is the sum of the potential energy function associated with bonds, angles, dihedrals, improper and non-bonded interactions. These are as follows:

$$U_{total} = U_{bonds} + U_{angles} + U_{dihedrals} + U_{impropers} + U_{nonbonded} \dots\dots\dots (1)$$

$$\text{where, } U_{bonds} = \sum_{bonds} K_r (r - r_0)^2$$

$$U_{angles} = \sum_{angles} K_{\theta} (\theta - \theta_0)^2$$

$$U_{dihedrals} = \sum_{dihedrals} K_{\chi} (1 + \cos(n\chi - \delta))$$

$$U_{\text{improvers}} = \sum_{\text{improvers}} K_{\phi} (\phi - \phi_0)^2$$

$$U_{\text{nonbonded}} = \sum_{\text{nonbonded}} \epsilon_{ij} \left[\left(\frac{R_{ij}^{\text{min}}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^{\text{min}}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon r_{ij}}$$

K_r , K_{θ} , K_{χ} , and K_{ϕ} are the force constants for bond stretching, angular vibrational motion, dihedral, and improper angles. In nonbonded interaction, both Lennard-Jones (LJ) “6-12” and Coulombic terms were included. ϵ is the dielectric constant. q_i and q_j are the partial atomic charges, r_{ij} is the distance between the i^{th} and j^{th} atoms. R_{ij}^{min} and ϵ_{ij} are the LJ minimum distance and well depth, respectively.

During the simulation, the damping coefficient (γ) of 1ps^{-1} by Langevin dynamics was used to keep the temperature of the system constant at 310 K. Particle Mesh Ewald (PME) method was applied to calculate long-range electrostatic forces without truncation [35]. The pair-list distance cut-off was set to 2 Å with a direct non-bonded potential cut-off of 12 Å and a scaling factor in the range 1-4 was employed in this system. Constant pressure was maintained at 1 atm using the periodic boundary conditions by applying the *Langevin piston Nosé-Hoover* method [36-37]. Preceding the structural equilibration, the model complex systems were put through several energy minimizations and relaxation cycles so that the bad contacts (steric clashes) present during the system preparation are abolished. Initially, water and counterions were equilibrated for 0.5 ns with a 1 fs time interval using a short “constant-pressure” NPT to pack lipid and water against the modeled protein-lipid. After that, the modeled protein was fixed. The counterions were allowed to relax their positions. System energy minimization was carried out for 40,000 steps using the conjugate gradient method to prevent steric clashes. Next, a short equilibration was performed using NPT for a period of 2 ns with 2 fs timesteps with protein constrained, in which hydrogen atoms were subjected to the SHAKE algorithm [38]. For achieving a free and stable simulation system, a short equilibration for this system was carried out with a time interval of 2 fs for 2ns of NPT equilibration with positional restraints on the heavy atoms of the protein complex. Herein, a small harmonic potential with a spring constant ($1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$) was applied to keep the protein atoms constrained. Thereupon, the entire system was switched to NPNAT (constant membrane-normal pressure, P_n , temperature, and membrane area) from the NPT ensemble without restraints on the

proteins for 200 ns simulation time, in which a time step of 2 fs was employed for integrating the equations of motions.

2.4 Potential of mean force calculation using ABF method.

After the equilibration of each system, the free energy profile concerning the potential of mean force (PMF) was calculated [39-40]. According to the reverse Boltzmann relationship, free energy for the protein-protein interaction system can be written as

$$A_{ij} = -kT \ln [f_{ij}(r)/Z_{ij}] \dots \dots \dots (2)$$

where, k = Boltzmann constant

T= Absolute temperature

$f_{ij}(r)$ = frequency of contacts between receptor protein atom type i^{th} and ligand-protein atom type j^{th} , occurring at the center of mass distance (r).

Here, we have used the adaptive biasing force (ABF) module implemented in the NAMD simulation. An external biasing force was applied to the groups of atoms during PMF calculations, with respect to proper reaction coordinates (RC). In this ABF simulation, RC was calculated based on the center of mass (COM) distance between EPCR and GLA- domain (FVIIa/FXa/all mutants), in the range of 31.3 to 36 Å with a bin width $d\zeta$ of 0.1 Å, in which 5000 force samples were collected in each bin. The biasing force was evaluated locally from the sampled conformations of the system at each step and updated continuously during simulation. Each RC was divided into equally spaced windows (width). Subsequently, for each window, ABF simulations were executed, and both upper and lower boundaries of simulation input files were subjected to a harmonic force of 100 kcal/mol/Å. The standard errors were estimated based on calculated PMFs from three individual runs (different starting coordinates) of simulation length of 5 ns for each system (EPCR and wildtype hFVIIa/hFXa/all mutants GLA-domain) using the NPT ensemble. NAMD ‘colvars’ module was employed to restrain the transmembrane portion and cytoplasmic region of EPCR for each simulation with a force constant of 100 kcal/mol/Å². Langevin dynamics was utilized to maintain the system temperature at 310 K. To tackle the long-range electrostatic forces, particle-mesh Ewald (PME) algorithm was used. The time step

integration was set to 1fs to calculate the equations of the motion and the rigid bonds in the configuration file were set to all.

2.5 Protein-protein binding affinity change predictions by BindProfX.

To predict protein-protein binding affinity change upon mutation of residues at the interface, BindProfX was employed [41]. In case of multiple mutations, BindProfX uses iAlign to align the Protein-Protein Interaction (PPI) interface structure of the target protein to a set of PPI interface structures from the PIFACE database to predict $\Delta\Delta G$ of interface residue mutations [42].

$\Delta\Delta G$ is calculated for the difference between the energies of wild-type and mutant complexes:

$$\Delta\Delta G_{WT \rightarrow Mut} = E_{WT}(\text{complex}) - E_{Mut}(\text{complex})$$

In the case of single point mutation for predicting the binding affinity change, FoldX physics potential $\Delta\Delta G_{\text{foldx}}$ can be optionally combined with the profile conservation score which is calculated from the interface alignment [43].

BindProfX is a refurbished technique to calculate the binding affinity of protein-protein interface based on gene mutation through structure profiling. Free energy change on binding ($\Delta\Delta G$) is counted as the logarithm of the relative probability of mutant amino acids over wild-type ones in the interface alignment matrix, wherein three pseudo-counts are incorporated to diminish the limit of the current interface library. The interface comparison is performed by the iAlign program, where all interfaces with a high interface similarity score (IS score) are used to construct a matrix. The binding free-energy change upon mutation is calculated by

$$\begin{aligned} \Delta\Delta G_{\text{evo}}(i) &= -\lambda \ln \{P(A_{Mut},i)/P(A_{WT},i)\} \\ &= -\lambda \ln \{N_{\text{obs}}(A_{Mut},i) + N_{\text{pseudo}}(A_{Mut},i) / N_{\text{obs}}(A_{WT},i) + N_{\text{pseudo}}(A_{WT},i)\} \end{aligned}$$

where, $P(A_{Mut},i)$ and $P(A_{WT},i)$ are the possibility of mutant and wild-type amino acids, respectively, appearing at the i th position of matrix.

$N_{\text{obs}}(\text{Mut}, i)$ and $N_{\text{obs}}(\text{WT}, i)$ are the number of the corresponding amino acids observed in the iMSA matrix, where $N_{\text{pseudo}}(A, i)$ is the corresponding pseudo-count number introduced to offset the limitations of statistics. Interface alignment and

structural analog search are performed by the I-align program, which is built on IS score:

$$\text{IS-score} = \frac{S + S_0}{1 + S_0}$$

$$\text{where, raw interface similarity score } S = \frac{1}{L_Q} \sum_{i=1}^{N_a} \frac{f_i}{1 + \left(\frac{d_i}{d_0}\right)^2}$$

$$s_0 = 0.18 - \frac{0.35}{L_Q^{0.3}} \text{ is a scaling factor to normalize the interface size.}$$

In the raw score S , L_Q is the average number of interface residues, N_a is the number of aligned interface residues, f_i is the fraction of conserved interface contacts at the i^{th} aligned position, and d_i/d_0 is the normalized C_α distance at the i^{th} aligned position. It is not required to follow the sequential order for the alignment of interfacial residues while running I-align.

2.6 Residue-wise interaction energy calculation using gRINN.

To find the non-bonded interaction energies for the residue pairs (between EPCR and GLA domain) gRINN tool was used [44]. gRINN attributes graphical user interfaces and a command-line interface to generate and analyze pair-wise residue interaction energies from MD simulation trajectories. For running this standalone program, PDB, PSF, and DCD files of all the systems were provided as input. A solute dielectric constant of value 1.0 was used while computing the electrostatic component of the interaction energy. Percent cutoff value and filtering distance cutoff values were 60% and 12 Å, respectively meaning that only residue pairs whose centers-of-mass come closer than 12 Å in at least 60 percent of trajectory frames were included.

3. Results and Discussion

Administration of hFVIIa is used in hemophilia treatment because hFVIIa binds to EPCR and generates hemostasis [45]. The importance of residue position 4th was highlighted by Pavani *et al.*, who demonstrated that a single substitution of Leu4 to Phe4 in mouse FVIIa determines its endothelial protein C receptor (EPCR) binding. Human FXa GLA-domain already has Phe at the 4th position; however, hFXa poorly interacts with hEPCR. The reason for this discriminate nature remains unclear. The key differences between hFXa and hFVIIa GLA-domain residues are at positions 8th and 9th. We hypothesized that these residues might be responsible for showing different

consequences in binding. To address these problems, we did *in silico* mutation of the FVIIa GLA-domain as shown in Table 2-1 and Figure 2-1b. We simulated the following systems: FVIIa, FXa, all FVIIa GLA-domain mutants (single, double) complex with membrane-bound flEPCR.

3.1 Dynamics of FVIIa/FXa GLA-domain in EPCR complex.

As shown in the crystallographic structures of EPCR-protein C (PDB ID: 1LQV), protein C binds to EPCR through its GLA-domain [46]. As the GLA-domain of FVIIa and FXa is homologous to that of the protein C GLA-domain, therefore, we have superimposed all the GLA-domains from protein C crystallographic structures bounded with EPCR for the model preparation. To predict the stability of the system, root means square deviations (RMSDs) of all GLA-domain for each system were calculated over time, relative to the initial structure (Figure 2-2). In the case of the FVIIa_{WT} GLA-domain, the least fluctuation was observed in RMSD, indicating that the wild-type system is more stable. However, all mutants exhibit higher deviations from the initial structure. We observe that the relative stability of FVIIa and mutants GLA-domain with respect to positions 4th and 9th are in the following order: FVIIa_{WT}>FVIIa_{R9W}>FVIIa_{F4L}> FVIIa_{F4L; R9W} (Figure 2-2a). The average RMSDs with standard deviation (SD) of FVIIa, all mutants, and FXa GLA-domain for the last 100 ns are mentioned in Table 2-2. However, we found that mutation at position 9th (FVIIa_{R9K}) induces a higher fluctuation in GLA-domain as compared to mutation at position 8th (FVIIa_{L8M}) (Figure 2-2b). Average RMSD of the GLA-domain for these systems are in the following order: FVIIa_{WT}>FVIIa_{L8M}>FVIIa_{R9K}>FVIIa_{L8M; R9K}>FXa_{WT}, indicating that mutations induce significant conformational alterations in the GLA-domain. Further, we calculated the RMSD of EPCR for all systems to check whether mutations in FVIIa GLA-domain alter the stability of the whole complex (Figure 2-2c and 2-2d). Data reveal that the RMSD of EPCR from the wild-type FVIIa system is much lower (2.8 ± 0.1 Å) compared to other systems (Table 2-2), indicating that the stability of EPCR in the complex is much higher. However, both double mutants (FVIIa_{F4L; R9W} and FVIIa_{L8M; R9K}) affect the stability of the whole EPCR in the complex. We have also run another set of simulations (around 130 ns) with different initial velocity coordinates to check the robustness of the simulation data (Figure 2-3). We found that both RMSD of GLA and EPCR for all the systems follow a similar pattern as observed with the previous data. Therefore, we conclude that the dynamics of the overall system are also dependent upon the mutation of the GLA-domain.

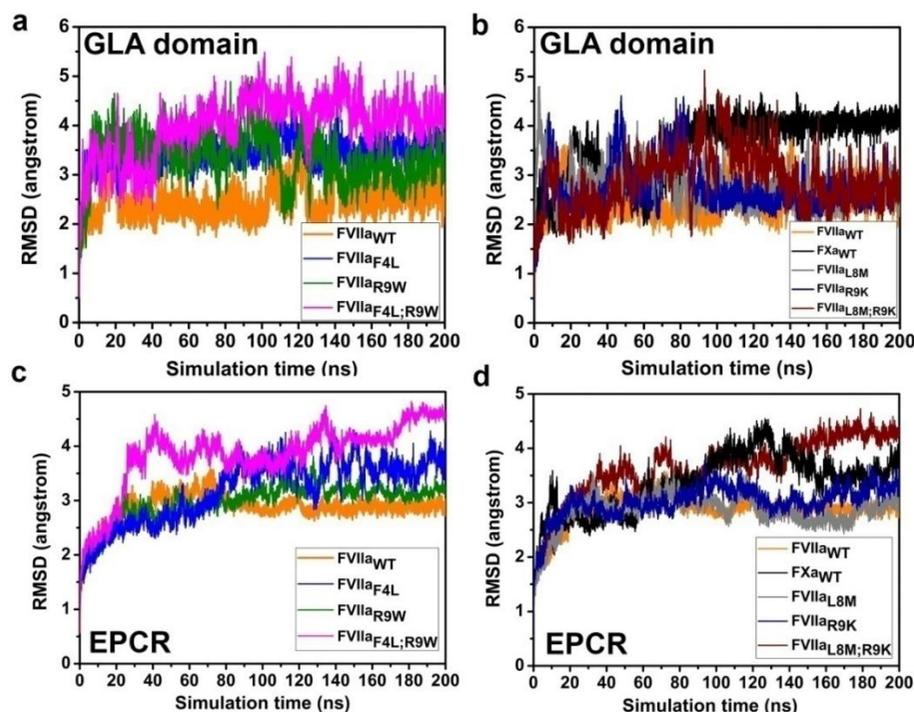


Figure 2-2. Root mean square deviations (RMSD) of GLA-domain and EPCR were obtained from simulation of membrane-bound flEPCR and GLA-domain complex. Backbone RMSD of (a) GLA-domain of FVIIa_{WT}, FVIIa_{F4L}, FVIIa_{R9W}, and FVIIa_{F4L; R9W} (b) GLA-domain of FXa_{WT}, FVIIa_{L8M}, FVIIa_{R9K}, and FVIIa_{L8M; R9K}, (c) EPCR of FVIIa_{WT}, FVIIa_{F4L}, FVIIa_{R9W}, and FVIIa_{F4L; R9W}, (d) EPCR of FXa_{WT}, FVIIa_{L8M}, FVIIa_{R9K}, and FVIIa_{L8M; R9K} for 200 ns simulation run.

Table 2-2. Root mean square deviation (RMSD) of EPCR and GLA-domain, calculated from the last 100 ns of the production run for all the systems (membrane-bound flEPCR and GLA-domain complex)

System	EPCR	GLA-domain
	RMSD \pm S.D. (\AA)	RMSD \pm S.D. (\AA)
<i>Mutation based on mouse FVIIa system</i>		
FVIIa _{WT} system	2.8 \pm 0.1	2.6 \pm 0.3
FVIIa _{F4L} system	3.6 \pm 0.2	3.6 \pm 0.2
FVIIa _{R9W} system	3.2 \pm 0.1	3.1 \pm 0.4
FVIIa _{F4L; R9W} system	4.2 \pm 0.3	4.3 \pm 0.3
<i>Mutation based on human FXa system</i>		
FXa _{WT} system	3.7 \pm 0.3	4.0 \pm 0.2
FVIIa _{L8M} system	2.8 \pm 0.2	2.6 \pm 0.2
FVIIa _{R9K} system	3.1 \pm 0.2	2.7 \pm 0.2
FVIIa _{L8M; R9K} system	4.0 \pm 0.3	3.0 \pm 0.5

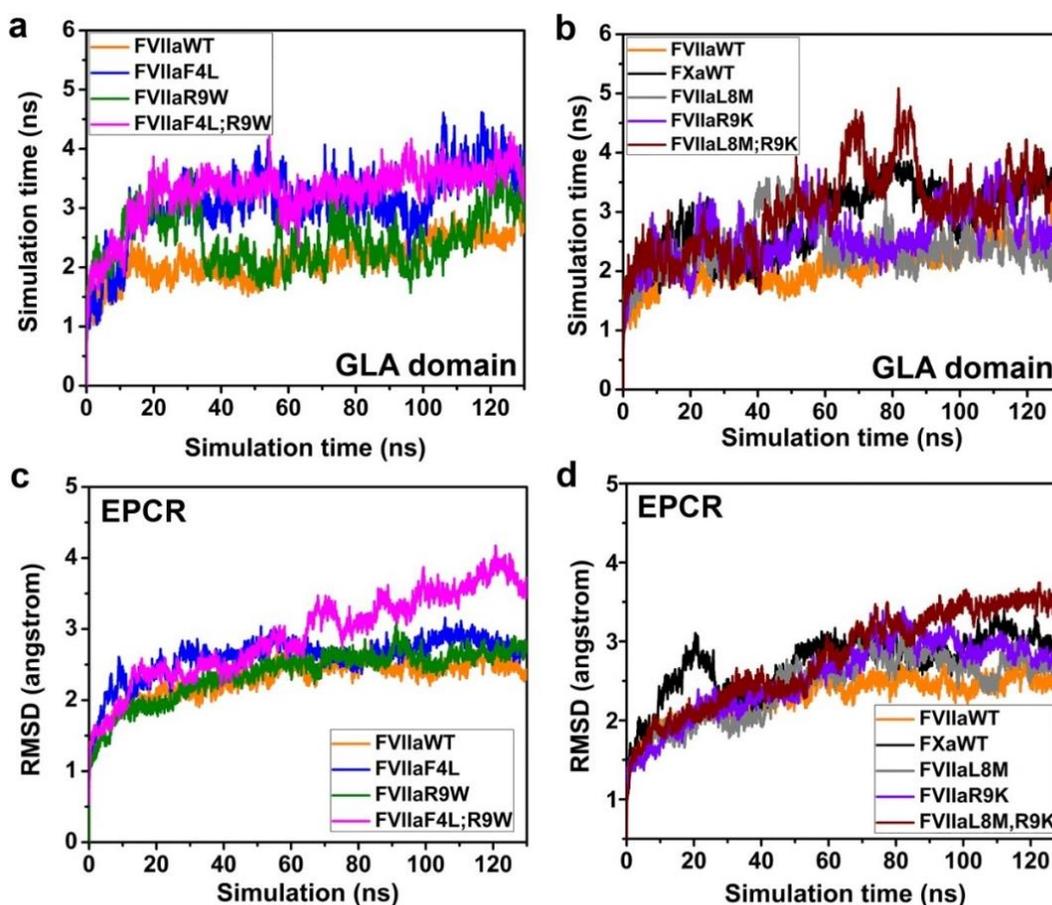


Figure 2-3. Root mean square deviations (RMSD) of GLA-domain and EPCR were obtained from 130 ns unbiased MD simulation trajectories of the EPCR-GLA complex for all the systems. (a) GLA domain of FVIIa_{WT}, FVIIa_{F4L}, FVIIa_{R9W}, and FVIIa_{F4L; R9W} system, (b) GLA domain of FXa_{WT}, FVIIa_{L8M}, FVIIa_{R9K}, and FVIIa_{L8M; R9K} system, (c) EPCR of FVIIa_{WT}, FVIIa_{F4L}, FVIIa_{R9W}, and FVIIa_{F4L; R9W} system, (d) EPCR of FXa_{WT}, FVIIa_{L8M}, FVIIa_{R9K}, and FVIIa_{L8M; R9K} system.

Evaluation of root mean square fluctuations (RMSF) is insightful to understand the overall residue-wise protein dynamics in the protein complex. To explore the local dynamical variations in a protein complex, we have computed RMSF of both GLA-domain as well as EPCR by taking the average structure of the last 60 ns of the simulation (2 ns interval) (Figure 2-4). In the case of the FVIIa_{WT} GLA-domain, the residue-wise fluctuation is much lower, and the variation in the anchoring residue is in the lower range below 2 Å (Figure 2-4a). In the case of the point mutation at residue position 4th (FVIIa_{F4L}), fluctuation of GLA-domain residues was much higher in the range 2.8-3.2 Å; whereas mutation at position 9th is not significantly associated with the residue-wise fluctuation of GLA-domain and RMSF data is similar with that of FVIIa_{WT}. In contrast, the double mutation at positions 4th and 9th (FVIIa_{F4L; R9W})

significantly alters the relative fluctuations of the GLA-domain specifically in the region from residues 5 to 18. This may be due to a weaker binding affinity of the GLA-domain to EPCR. Overall, our results indicate that the mutation at residue position 4th is the driving factor in the regulatory movement of the GLA-domain. Our finding is also consistent with the recent experimental findings.

Despite having the same residue (Phe4) located at the 4th position, FXa GLA-domain exhibits higher instability. The probable reason for this alteration is still unknown. Therefore, we evaluated the dynamics of the GLA-domain by calculating RMSF for the last 60 ns of the total simulation (Figure 2-4b). We observe that both single and double mutants contributed to excessive fluctuations in FVIIa GLA-domain. The RMSF of the GLA-domain for the given system is in the following orders: FVIIa_{WT} > FVIIa_{L8M} > FVIIa_{R9K} > FVIIa_{L8M; R9K} > FXa_{WT}.

Similarly, we analyzed the RMSF of EPCR in the complex for all systems (Figure 2-4c). We found that residue-wise fluctuation of EPCR from the FVIIa_{WT} system in the complex is least as compared to both double (FVIIa_{F4L; R9W}) and single mutant (FVIIa_{F4L}) systems, indicating that the wild-type FVIIa system is most stable. Residue-wise stabilities for the EPCR system are in the following order: FVIIa_{WT} > FVIIa_{R9W} > FVIIa_{F4L} > FVIIa_{F4L; R9W}. Fluctuations of EPCR residue are thought to be related to the dynamics of the GLA domain.

A similar trend was observed in the RMSF of EPCR in FXa-bound form (Figure 2-4d). The fluctuation of EPCR in the FXa system is much higher as compared to the other system, which indicates that the overall stabilization of the bound complex is dependent upon the dynamics of the GLA domain. It is also observed that the relative fluctuation of EPCR in the region nearer to H2 and H3 (residues 140-180) is much higher for the system FVIIa_{L8M} and FVIIa_{R9K}. The probable explanation for the dramatic increase in the overall flexibility might be due to the disruption of the native H-bonding network.

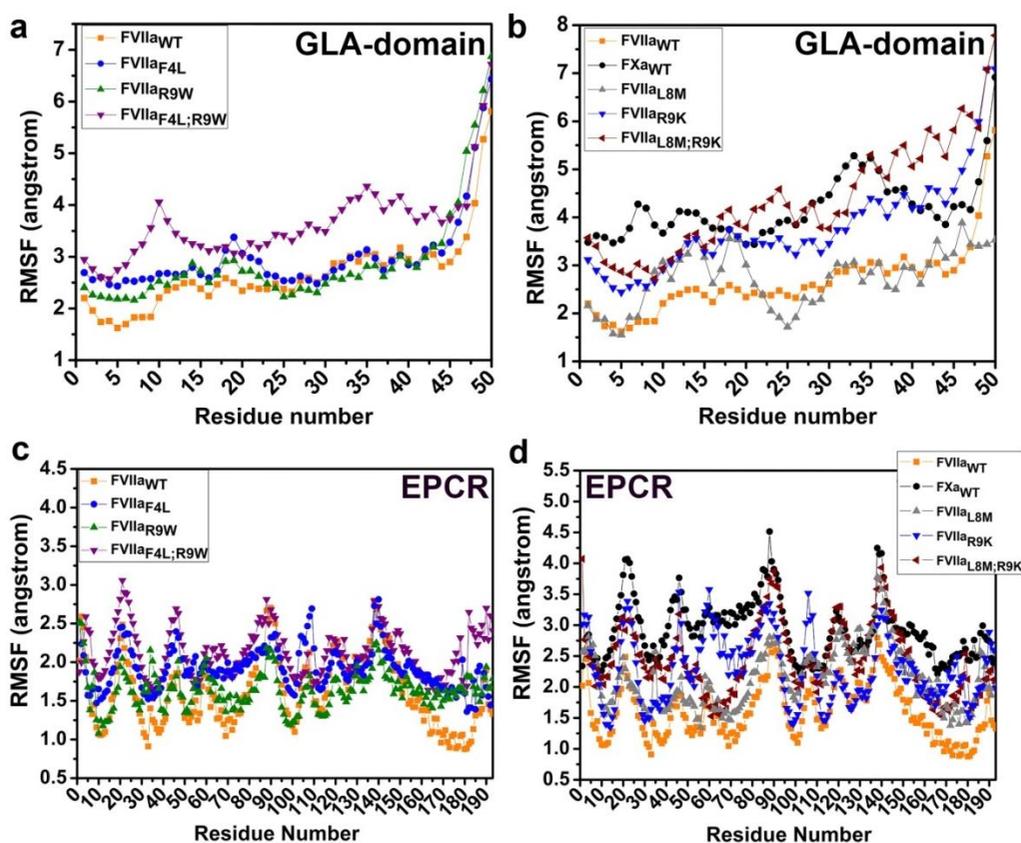


Figure 2-4. Residue-wise fluctuation of GLA-domain and EPCR from membrane-bound fIEPCR-GLA domain complex. Root mean square fluctuation (RMSF) of (a) GLA-domain of FVIIa_{WT}, FVIIa_{F4L}, FVIIa_{R9W}, and FVIIa_{F4L;R9W}, (b) GLA-domain of FXa_{WT}, FVIIa_{L8M}, FVIIa_{R9K}, and FVIIa_{L8M;R9K}, (c) EPCR of FVIIa_{WT}, FVIIa_{F4L}, FVIIa_{R9W}, and FVIIa_{F4L;R9W}, (d) EPCR of FXa_{WT}, FVIIa_{L8M}, FVIIa_{R9K}, and FVIIa_{L8M;R9K}.

3.2 Free binding energy profile of fIEPCR and Gla-domain complex.

In recent years, the impact of mutations in regulating protein-protein binding has begun to be understood [47]. Recently, Pavani *et al.* showed that Phe4 in mouse FVIIa is the determining factor for mEPCR binding. Sequence homology studies among hFVIIa, hFXa, and mFVIIa GLA-domain (Figure 2-1a) reveal that Phe4 is also present in hFXa; however, hFXa has a weaker binding affinity towards hEPCR. There are many possibilities of residues that may have an impact on the protein complex assembly. Therefore, we considered specific point mutations of key residues at positions 8th and 9th of the hFVIIa GLA-domain, as mentioned in Figure 2-1. To decipher the role of the GLA-domain for EPCR binding, the potential of mean force (PMF) was determined individually for the following permutations of the FVIIa/FXa with EPCR complex: FVIIa_{WT}-EPCR, FVIIa_{F4L}-EPCR, FVIIa_{R9W}-EPCR, FVIIa_{F4L;R9W}-EPCR, FVIIa_{L8M}-EPCR, FVIIa_{R9K}-EPCR, FVIIa_{L8M;R9K}-EPCR and FXa_{WT}-EPCR (Figure 2-5).

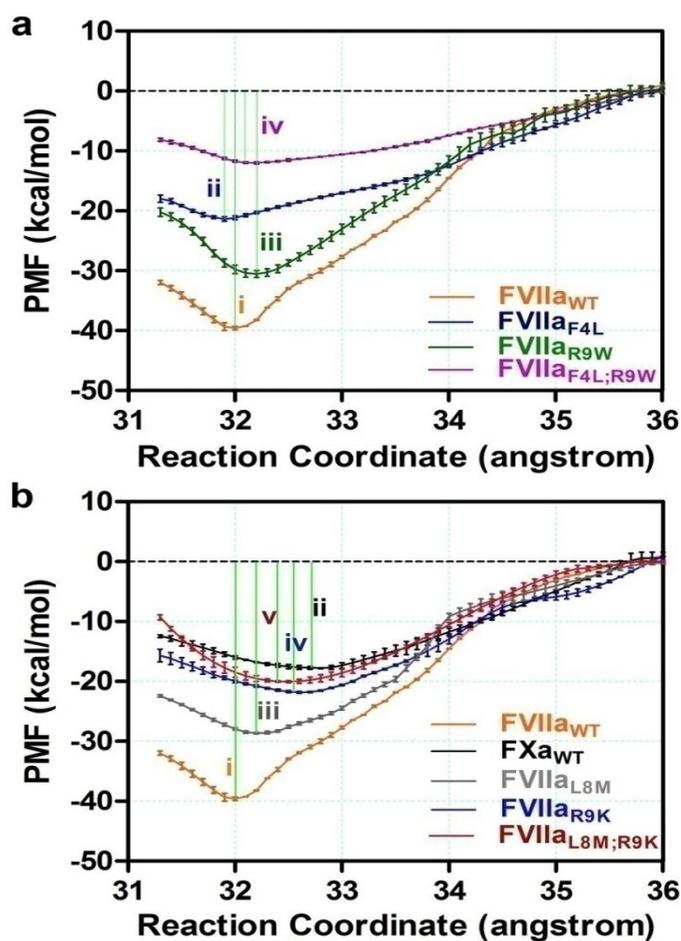


Figure 2-5. Free energy profile (potential of mean force, PMF) obtained from ABF simulations along the reaction coordinate (RC) between GLA-domain and sEPCR. PMF profile for a complex system (a) FVIIa_{WT}, FVIIa_{F4L}, FVIIa_{R9W}, and FVIIa_{F4L;R9W} mutation based on mouse FVIIa and (b) FXa_{WT}, FVIIa_{L8M}, FVIIa_{R9K}, and FVIIa_{L8M;R9K} mutation based on human FXa. Well-equilibrated membrane-bound flEPCR complex with GLA domain for all systems were used for PMF calculation. The reaction coordinate here is the center of mass (COM) distance between two bulk proteins: the extracellular portion of EPCR and the GLA-domain (FVIIa, FXa, and all mutants). Reaction Coordinate (RC) was calculated based on the center of mass (COM) distance between EPCR and GLA-domain (wildtype FVIIa/ wildtype FXa/all mutants), in the range of 31.3 to 36 Å. For all the systems, error estimates were obtained from the bootstrap analysis of all the sets, individually. The Bulk PMF value was shifted to 0 kcal/mol.

Mutations of FVIIa were performed using Pymol. Estimation of the free energy profile concerning the PMF provides an accurate description of the thermodynamic feasibility of the process. We have used the adaptive biasing force (ABF) simulation methods implemented in NAMD to compare the energy barriers along with the reaction coordinates (RC). Here, RC was measured based on the COM distance between the GLA-domain and EPCR (Figure 2-6).

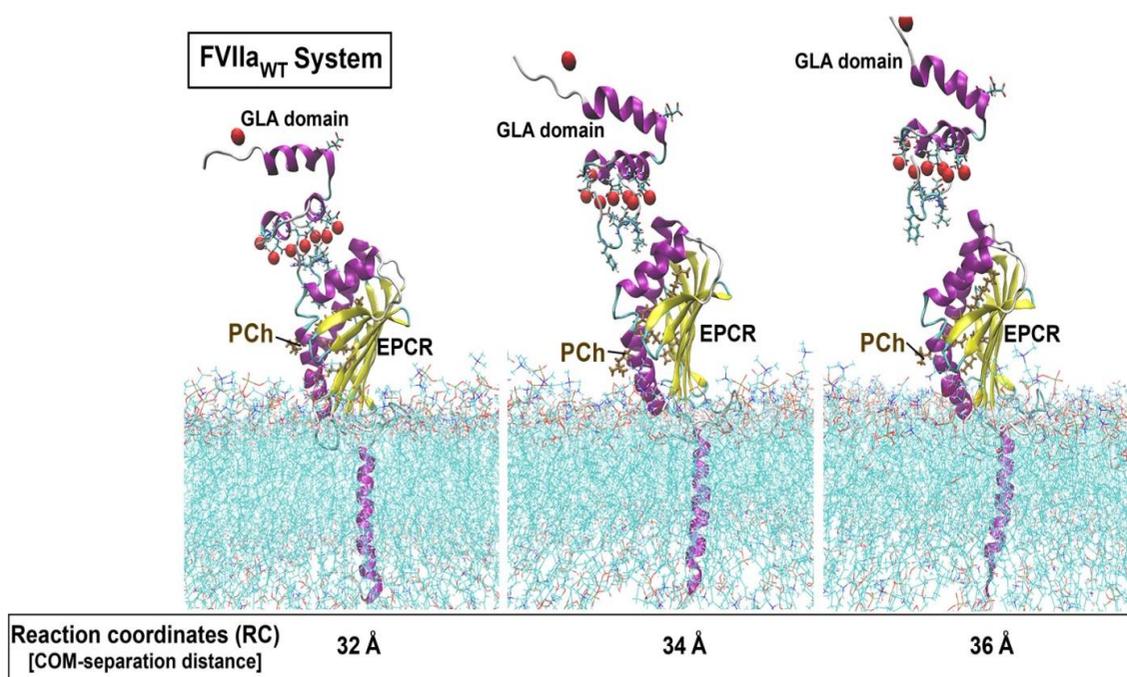


Figure 2-6. The conformational changes of EPCR-GLA domain for FVIIa_{WT} system in presence of PCh-containing lipid within EPCR groove at center-of-mass (COM) distance of 32 Å, 34 Å, and 36 Å, obtained from the ABF simulation trajectories. For the calculation of free energies, COM-separation distances between the GLA domain and EPCR were chosen as the reaction coordinates. Antigenic PCh-containing lipid within the EPCR groove is represented by brown licorice color.

As provided in Table 2-3, we found that each of the systems is associated with single energy minima (depth of the well). An interesting trend observed in the PMF calculation is that binding happens to be near -39.57 ± 0.3 kcal/mol for wild-type FVIIa GLA-domain/EPCR system around RC=32 Å (Figure 2-5a). The decrease in the PMF suggests the strength of interaction of the GLA domain with EPCR is higher. This can be understood through the formation of a stable complex at the given RC, as similar to the experimental binding affinity. On the other hand, an increase in the PMF also indicates the instability of the complex system.

Table 2-3. The average potential of mean force with standard error is associated with the EPCR-GLA domain complexes system.

System	PMF (kcal/mol)
Mutation based on mouse FVIIa system	
FVIIa _{WT}	-39.57±0.3
FVIIa _{F4L}	-21.35±0.4
FVIIa _{R9W}	-30.54±0.6
FVIIa _{F4L;R9W}	-11.97±0.2
Mutation based on human FXa system	
FXa _{WT}	-17.79±0.1
FVIIa _{L8M}	-28.64±0.1
FVIIa _{R9K}	-21.82±0.1
FVIIa _{L8M;R9K}	-20.04±0.2

Due to the mutation at the residue, position 4th (FVIIa_{F4L}), PMF dramatically increases to a value -21.35±0.4 kcal/mol, which suggests that protein-ligand interaction is weaker for EPCR and GLA domain for FVIIa_{F4L} system. The differences in the minimum of PMF are also observed, i.e., -9.03 kcal/mol for the mutant system at position 9th (FVIIa_{R9W}) when compared with the wild-type FVIIa_{WT} system. However, the differences in the minimum of PMF are much more significant (-27.6 kcal/mol) for the double mutant system (FVIIa_{F4L;R9W}) than that of the wild-type system. The PMF for the mutant system with respect to residue position 4th and 9th follows the order: FVIIa_{WT} (-39.57 kcal/mol) < FVIIa_{R9W} (-30.54 kcal/mol) < FVIIa_{F4L} (-21.35 kcal/mol) < FVIIa_{F4L;R9W} (-11.97 kcal/mol). Our results show that mutation at residue position 4th alters the interaction for EPCR and GLA-domain, however double mutation further impairs the complex formation.

Mutations at positions 8th (L8M) and 9th (R9W) of the hFVIIa GLA-domain also show similar trends (Figure 2-5b). The magnitude of the binding energy of the FVIIa_{L8M} system is similar to those found for FVIIa_{R9W}. However, the mutation at position 9th (R9K) drastically alters the energy profile of the EPCR and GLA-domain system more than that of the wild-type system (FVIIa_{WT}). Interestingly, we also found that the positions of minima are changing upon point mutation. A maximum difference

is observed between FVIIa_{WT} and FVIIa_{R9K} system and between FVIIa_{WT} and FVIIa_{L8M; R9K}. The probable explanation for this change may be due to mutation in residue position somehow alters the global minima of the complex.

We found that mutation at both residue positions 8th and 9th (FVIIa_{L8M; R9K}) guides the energy profile of the whole system. The order of PMF for the given system is as follows: FVIIa_{WT} (-39.57 kcal/mol) < FVIIa_{L8M} (-28.64 kcal/mol) < FVIIa_{R9K} (-21.82 kcal/mol) < FVIIa_{L8M; R9K} (-20.04 kcal/mol) < FXa_{WT} (-17.79 kcal/mol). The alterations of the energy profile in the mutants compared with the wild-type system might have implications for complex formation, which may be due to the difference in the H-bonding profile. To check further, how the mutation might affect the free energy profile of the complex system, we explored the intermolecular interaction between EPCR and GLA-domain for all the systems.

The binding profile of EPCR and GLA-domain of wild-type FVIIa, wild-type FXa, and all mutants were also examined using BindProfX. According to the data obtained from BindProfX (Table 2-4), it is observed that on mutating the wildtype FVIIa GLA-domain to mutant FVIIa at the position 4th, $\Delta\Delta G$ of FVIIa_{F4L} is 1.34 kcal/mol, which is much higher compared to the $\Delta\Delta G$ of FVIIa_{R9W}, implicating that the total free energy binding affinity profile for FVIIa_{R9W} is more feasible than that of the FVIIa_{F4L}. Similarly, $\Delta\Delta G$ of FVIIa_{F4L; R9W} is 1.46 kcal/mol, which is even higher compared to the $\Delta\Delta G$ of FVIIa_{F4L} and FVIIa_{R9W}. We suggest that the double mutation imparts a greater contribution toward a decrease in binding affinity, which is similar to the observed data from the PMF calculation. As shown in Table 2-4, it can be inferred that $\Delta\Delta G$ of FVIIa_{L8M; R9K} is 1.77 kcal/mol which is greater compared to FVIIa_{L8M} and FVIIa_{R9K}, indicating that the binding affinity profile for FVIIa_{L8M; R9K} is less favorable. Based on PMF and BindProfX, we conclude that mutation at position 9th (FVIIa_{R9K}) may play a significant role in affecting stability as well as causing the change in the binding affinity of EPCR and GLA-domain.

Table 2-4. Summary of the binding profile of all mutant systems (mouse FVIIa and human FXa-based) using BindProfX.

Mutating residue	Position	$\Delta\Delta G$ (kcal/mol)
<i>Mutation based on mouse FVIIa</i>		
F 4 L	4	1.34
R 9 W	9	0.01
F 4 L ; R 9W	4,9	1.46
<i>Mutation based on human Xa</i>		
L 8 M	8	0.01
R9K	9	1.75
L8 M; R9K	8,9	1.77

3.3 Evaluation of the interactions formed during MD simulations.

The number of H-bond formations in the complex system explains the overall stability of the protein. Thus, H-bond analysis of the whole systems in both wild-type and all mutants were evaluated and compared. Intermolecular interactions (both H-bond and hydrophobic) were analyzed, which are formed with the side chain-side chain or side chain-main chain of GLA-domain and EPCR within a cutoff distance of 3.5 Å. To ascertain the binding strength, we evaluated the interaction occupancy (%) between FVIIa GLA-domain and EPCR for all the systems (Figure 2-7). H-bond segment of the GLA-domain (residues 1-50) mainly interacts with EPCR regions 87-154. The interaction between Arg87 (EPCR) and Gla25 (protein C GLA-domain) is reported in the crystal structure of protein C GLA-domain and EPCR (PDBID:1LQV). This interaction pair is also prominent for EPCR and wild-type FVIIa GLA-domain. However, there is a difference in occupancy level (%) for the interacting residue of EPCR and GLA-domain for other systems. Interestingly, we found that intermolecular H-bond forming tendency is much higher for the wild-type hFVIIa GLA-domain EPCR system compared to the mutant systems. It is evident from the plot that a newly interacting H-bond is visible between residues Gln150-side-NE2 (EPCR) and Gla7-side-OE21 (FVIIa) for wild-type (Figure 2-7a). However, this interaction is weakly observed for the mutants FVIIa_{R9W}-EPCR and FVIIa_{F4L; R9W}-EPCR. It is noteworthy to mention that this interaction is negligibly observed in the mutant system FVIIa_{F4L}-EPCR. Also, it is very interesting to mention another interaction between Leu5:HN and Tyr154:OH, which is comparably higher in the wild-type FVIIa system than the

remaining ones. Other interacting residue pairs formed between GLA-domain and EPCR are Gln85-Leu8, and Tyr154-Asn2, and are prominently higher in the FVII_a^{WT} system. A most important type of hydrophobic interaction between aromatic amino acids phenylalanine and tyrosine is observed between EPCR and FVII_a^{WT} GLA-domain within a distance cutoff of 3-4.2 Å. Occupancy (%) between residues Phe4 (GLA-domain) and Tyr154 (EPCR) is approx 28.47% for the wild-type system, which is significantly higher than FVII_a^{F4L} (5.58%), FVII_a^{R9W} (15.36%) and FVII_a^{F4L; R9W} (7.60%), as shown by colored GLA-domain residues sphere (Figure 2-7b). As reported in the crystal structure of protein C GLA-domain and EPCR bounded with antigenic lipid, the omega-loop region of protein C appears to interact with phospholipid. Therefore, we have investigated the protein-lipid interaction for all the systems. While analyzing the protein-lipid interaction, we found that lipid-dependent protein interaction is prominent in the wild-type FVII_a system, shown by the black arrow (Figure 2-7c). However, in the case of mutants, lipid-based interaction (between phosphatidylcholine and Phe4) gets hindered. We suggest that the numbers of the H-bonds between the GLA-domain and EPCR appear to decrease when lipid-based protein interaction is not formed, hence impairing the stability of complex formation.

In the FXa-based mutant system, Phe is located at position 4th; however, the stability of the whole system is less compared with the wild-type system. The probable answer for this fluctuation might be either due to a weaker interaction between GLA-domain and EPCR, protein-lipid (intrinsic bound lipid), or both. To evaluate further, we check the interacting residue between EPCR and GLA-domain for the systems FX_a^{WT}, FVII_a^{L8M}, FVII_a^{R9K}, and FVII_a^{L8M; R9K}. Based on Figure 2-7, it is clear that the occupancy of interacting residue between EPCR and FVII_a GLA-domain mutant is much lesser than in the FVII_a^{WT} system. There is hardly any interaction between Gln85 and Leu8 for FX_a^{WT}, FVII_a^{L8M}, and FVII_a^{L8M; R9K}. The relevant difference in occupancy exists for the residue-pair Gln150-Gln7, Gln85-Leu8, and Leu5-Tyr154 for the FVII_a^{WT} system in comparison with other mutant systems. We found that occupancy for the interacting residue pair Tyr 154-Asn2 is significantly higher for wild-type FVII_a system over mutant system FVII_a^{R9K}. However, the mutation at position 8th (FVII_a^{L8M}) does not significantly alter the occupancy for this residue pair. Additionally, we analyzed the lipid-protein interactions for all the mutant systems (FVII_a^{L8M}, FVII_a^{R9K}, and FVII_a^{L8M; R9K}) and surprisingly, we found that lipid-based interaction is either lost or significantly reduced for mutant FVII_a^{L8M}, FVII_a^{R9K}, and FVII_a^{L8M; R9K}. In the case of the FX_a system, we found that lipid-based protein interaction is completely lost. For further analysis, the same hydrophobic interaction between aromatic amino acids Phe and Tyr was measured as mentioned in the previous

section. Occupancy (%) for the interacting residues Phe4 and Tyr154 for the whole system was obtained from 143 ns simulation. We obtained the occupancy in the following order: FVIIa_{WT} (28.47%) > FVIIa_{L8M} (19.41%) > FVIIa_{R9K} (14.93%) > FXa_{WT} (14.07%) > FVIIa_{L8M;R9K} (10.81%). As a whole, our data suggest that the wild-type FVIIa system, in which phenylalanine rings closely pack in the interior of EPCR and interact with Tyr154 in the presence of PCh lipid molecule, is the determining factor for the complex formation.

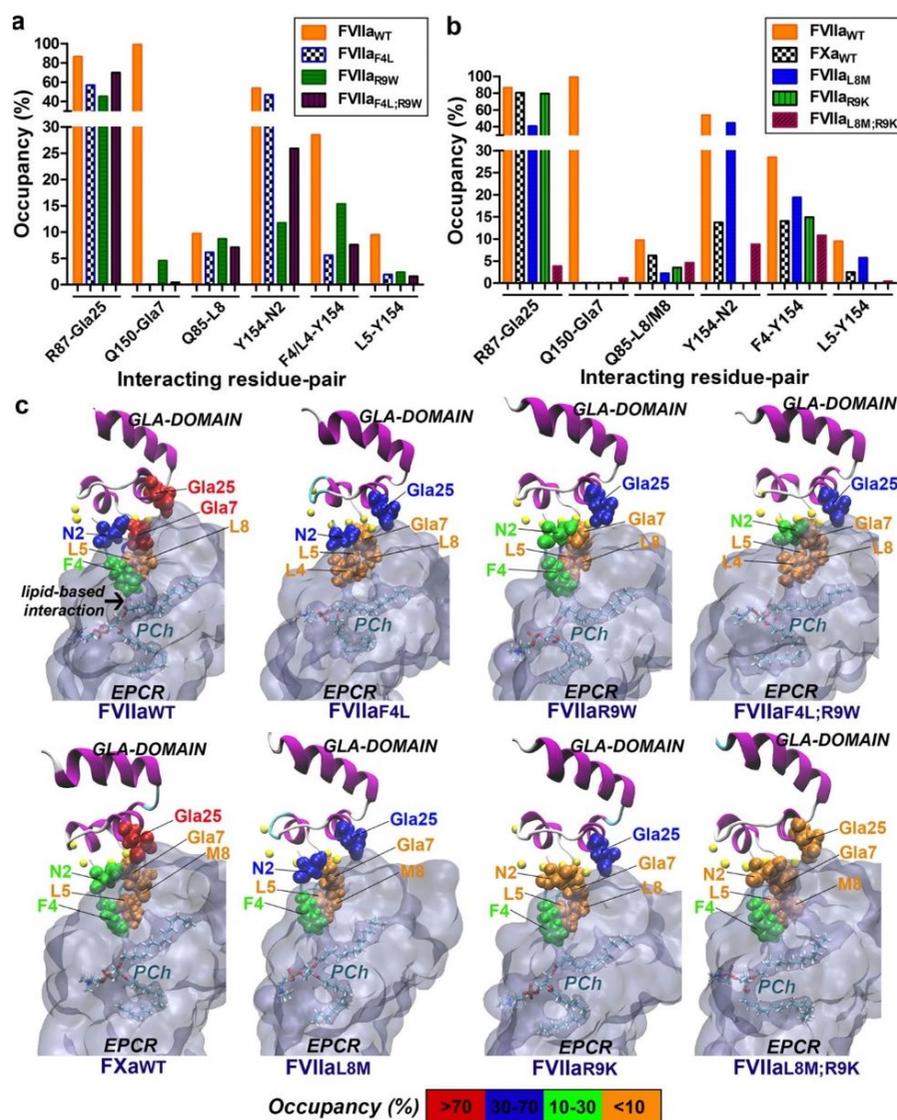


Figure 2-7. Occupancy of interacting residue-pair between EPCR and GLA-domain obtained from MD simulation. H-bond occupancy (%) plot between the residues of the GLA-domain and EPCR, calculated from 143 ns simulation time of (a) FVIIa_{WT}, FVIIa_{F4L}, FVIIa_{R9W}, and FVIIa_{F4L;R9W}, (b) FXa_{WT}, FVIIa_{L8M}, FVIIa_{R9K}, and FVIIa_{L8M;R9K} systems. (c) EPCR interacting residues of the GLA-domain are defined using a cut-off distance of 3.5 Å, and colored based on occupancy during the simulations are defined as follows: red for >70%, blue for 30-70%, green for 10-30%, and orange for <10%.

To find the non-bonded interaction energies (IE) for the residue pairs (between EPCR and GLA domain) gRINN (get Residue Interaction eNergies and Networks) tool was used (Figure 2-8). Mean IE among the selected residue pairs Gln85-Leu8, Try154-Asn2, Phe/Leu4-Tyr154, Leu5-Tyr154, Arg87-Gla25, and Gln150-Gla7 for all the systems were calculated by taking MD trajectories for the last 40 ns of the total run. Based on Figure 2-6, we found that the FVIIa_{WT} system has stronger interaction for the residue pairs Tyr154-Asn2, Arg87-Gla25, and Gln150-Gla7 with a value of mean IE of -11.89, -10.85, and -7.53 kcal/mol, respectively as compared to FVIIa_{F4L} system as well as other mutants.

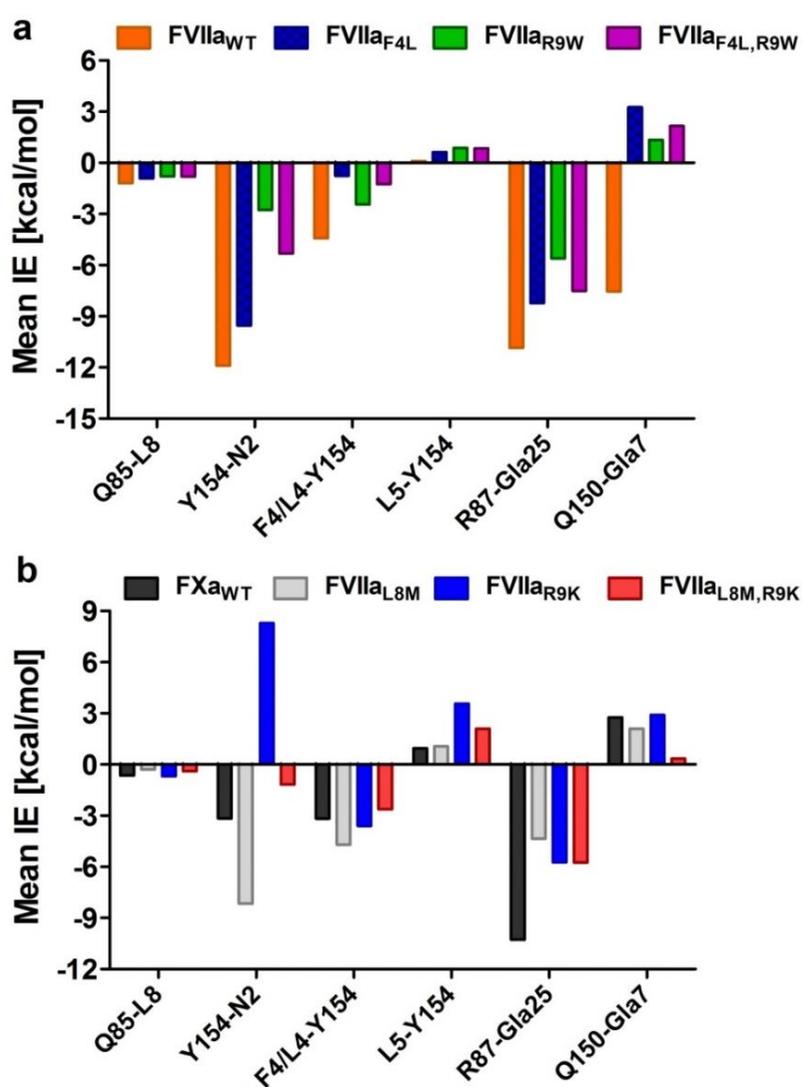


Figure 2-8. Interaction energy (IE) calculation between the selected residue pair of GLA domain and EPCR, obtained from the occupancy plot. Mean IE (kcal/mol) for the residue pairs Gln85-Leu8, Try154-Asn2, Phe4/Leu4-Try154, Leu5-Try154, Arg87-Gla25, and Gln150-Gla7 for all the systems (a) FVIIa_{WT}, FVIIa_{F4L}, FVIIa_{R9W}, FVIIa_{F4L; R9W} and (b) FXa_{WT}, FVIIa_{L8M}, FVIIa_{R9K}, FVIIa_{L8M; R9K}.

However, a substantial change in IE for the residue-pair Try154-Asn2 is also observed for the mutant system at the residue position 9th (FVIIa_{R9K}) with a value of ~8.29 kcal/mol, suggesting that mutations at the successive position 4th and 9th guide the interaction pattern within EPCR-GLA domain.

We have further deciphered the residue-pair distance by measuring distance distribution among those residues to estimate the H-bond propensity throughout the simulation period (Figure 2-9). We selected two interacting residue pairs, Gln150:NE2-Gla7:OE12 and Leu5:HN-Tyr154:OH to generate a distance distribution plot. As inferred from the plot, it is apparent that H-bonds formed for both the residue pair: Gln150-Gla7 and Leu5-Tyr154 contribute toward the extra-stability of the whole system. For the wild-type FVIIa system, we observed that the Gln150-Gla7 distance in the complex fluctuates in the range of 2.5-3.7 Å with a peak at 2.8 Å, indicating that binding between EPCR and GLA-domain is much higher (Figure 2-9a). In contrast, in the mutant FVIIa_{F4L} system, the Gln150-Gla7 distance fluctuates maximum in the range 6-8 Å with the main peak around 7 Å. The distance distribution for mutant systems FVIIa_{R9W} and FVIIa_{F4L; R9W} also have a similar pattern to that of the mutant FVIIa_{F4L} system. In addition, a significant difference in distance distribution was observed for the residue-pair Gln150:NE2-Gla7:OE12 for FX_{awT}, FVIIa_{L8M}, FVIIa_{R9K}, and FVIIa_{L8M; R9K} than that of wild-type FVIIa system (Figure 2-9b). The residue-pair distance for Leu5:HN-Tyr154: OH fluctuates between 2Å and 4.7 Å with a peak at 3 Å, suggesting that this interaction is also stronger for the FVIIa_{WT} system. However, for all mutant systems (FVIIa_{F4L}, FVIIa_{R9W}, and FVIIa_{F4L; R9W}), the distance fluctuates from 3 to 7.2 Å (Figure 2-9c), clearly indicating that the strength of interaction between EPCR and FVIIa GLA-domain mutant is weaker. Based on these results, we suggest that mutation at the residue 4th (FVIIa_{F4L}) drastically alters the strength of H-bond formation between EPCR and GLA-domain. Mutation at positions 4th and 9th (FVIIa_{F4L; R9W}) of the GLA-domain also exhibits lower affinity to EPCR. It is also noted that Leu5:HN-Tyr154:OH distance for the FVIIa_{L8M} system fluctuates maximum extent in the range 2.5-5 Å with the main peak around 4 Å; however, the distances for FX_{awT}, FVIIa_{R9K} (single) and FVIIa_{L8M; R9K} (double) fluctuate in the range 4-7 Å with a peak around 5-5.5 Å (Figure 2-9d). Thus, these results indicate that the H-bond forming tendency for the residue pair Leu5-Tyr154 for the mutant systems is in the following

order: $FVIIa_{WT} > FVIIa_{L8M} > FXa_{WT} > FVIIa_{R9K} > FVIIa_{L8M; R9K}$. We conclude that the incorporation of lysine mutation at the 9th position of the GLA-domain seems to be the consequence of the weakening of the EPCR-FVIIa GLA-domain complex formation, which is in accordance with the observed PMF result.

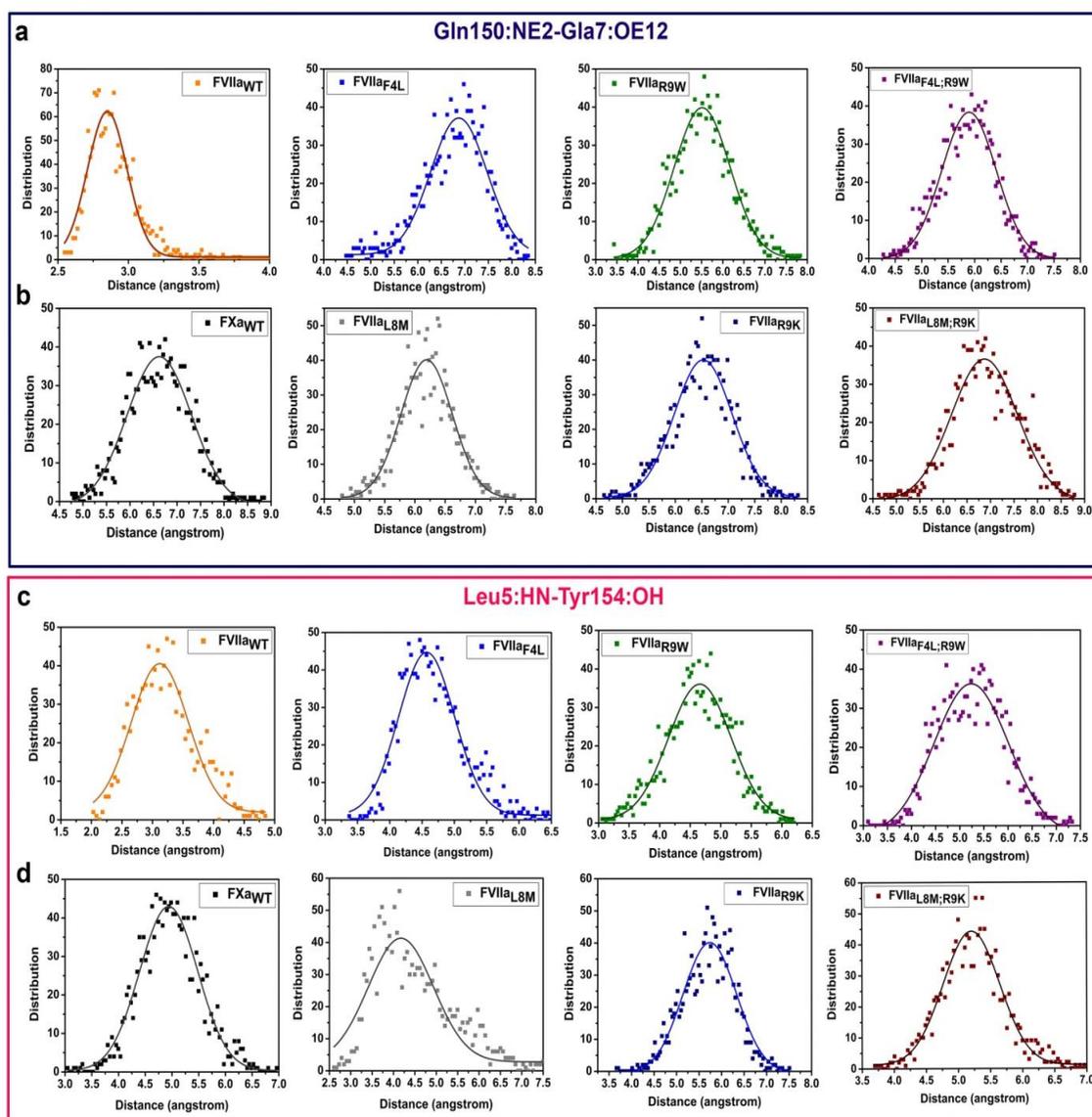


Figure 2-9. Distance distribution of unique interacting residue-pair between Gln150:NE2-Gla7:OE12 for all systems. (a) $FVIIa_{WT}$, $FVIIa_{F4L}$, $FVIIa_{R9W}$ and $FVIIa_{F4L;R9W}$, (b) FXa_{WT} , $FVIIa_{L8M}$, $FVIIa_{R9K}$ and $FVIIa_{L8M;R9K}$. Distribution of unique interacting residue-pair between Leu5:HN-Tyr154:OH for systems (c) $FVIIa_{WT}$, $FVIIa_{F4L}$, $FVIIa_{R9W}$ and $FVIIa_{F4L;R9W}$, and (d) FXa_{WT} , $FVIIa_{L8M}$, $FVIIa_{R9K}$, and $FVIIa_{L8M;R9K}$. The distribution plot was prepared based on the last 60 ns simulation of the production run.

3.4 Solvent Accessible Surface Area

Solvent accessible surface area (SASA) is used to predict the hydration of the hydrophobic core that is also associated with the protein-protein interaction. The lower in SASA indicates that the hydrophobic core is more buried and the binding association is stronger between receptor and ligand. To elucidate the SASA of the hydrophobic region at the position 4th, 8th, and 9th within FVIIa GLA-domain for all systems (FVIIa_{WT}, FVIIa_{F4L}, FVIIa_{R9W}, and FVIIa_{F4L; R9W}, FXa_{WT}, FVIIa_{L8M}, FVIIa_{R9K}, and FVIIa_{L8M; R9K}), SASA distribution was carried out for the last 60 ns trajectory period having 1.4 Å probe radius (Figure 2-10). In the case of the wild-type FVIIa system, SASA distribution for the residue position 4th is in the range 20-90 Å with a maximum peak around ~35 Å (Figure 2-10a), which indicates that Phe4 residue is more buried in the groove of EPCR. However, the distribution for the mutant systems FVIIa_{F4L} and FVIIa_{F4L; R9W} are in the range 40-90 Å with the maximum peak around ~60 Å, which suggests that mutations at the GLA-domain position 4th cause the hydrophobic core to be exposed more, leading to the loss of protein-lipid as well as protein-protein interactions. The SASA for the FVIIa_{L8M}, on the other hand, is relatively similar to that of wild-type FVIIa. From the graph, we found that FXa_{WT}, FVIIa_{L8M}, FVIIa_{R9K}, and FVIIa_{L8M; R9K} system have increased in SASA distribution compared to wild-type FVIIa. These observations are in agreement with the protein-protein interaction profile, where mutation causes the destabilization of the hydrophobic region of the GLA-domain, causing an interruption in the hydrophobic interactions among nonpolar residues.

SASA distributions at position 8th also have similar trends with that of mutation at position 4th (Figure 2-10b). The distribution peak for the systems is in the following order: FVIIa_{WT} < FVIIa_{R9W} < FVIIa_{F4L; R9W} < FVIIa_{F4L}. The distribution curve upon L8M mutation is in the range of 40-90 Å, as demonstrated by the smaller peak value compared to that of the wild-type. It is evident from the plot that SASA distribution for the mutant systems FVIIa_{R9K} and FVIIa_{L8M; R9K} is also higher, which suggests that the hydrophobic region becomes more accessible to the aqueous environment. SASA

distribution curves for the GLA-domain at residue position 9th also show similar trends as observed for the SASA distribution for the residue positions 4th and 9th. As position 9th of the GLA-domain is not directly associated with the protein and lipid-mediated interaction,¹⁸ however, we found a significant difference in the SASA distribution plot among wild-type FVIIa, wild-type FXa, and all mutants. In the case of the wild-type system, the hydrophobic region at position 9th is more protected from the external environment, which ultimately leads to a decrease in SASA distribution. In the case of mutant systems FVIIa_{F4L} and FVIIa_{R9W}, however, there is a sharp increase in SASA distribution range from 130-200 Å (Figure 2-10c). In the case of double mutation systems FVIIa_{F4L; R9W} and FVIIa_{L8M; R9K}, there is a decrease in SASA peak compared to wild-type system FVIIa. We found a similar pattern of the plot for wild-type FXa, FVIIa_{L8M}, and FVIIa_{R9K} systems. Only the difference among them is the relative peak of SASA distribution. The probable explanation for this change might be due to the weaker interaction between receptor (EPCR) and ligand (GLA-domain). Mutations at both positions 4th and 8th drastically change the interaction pattern between EPCR and GLA-domain, which in turn also affects the SASA distribution at position 9th.

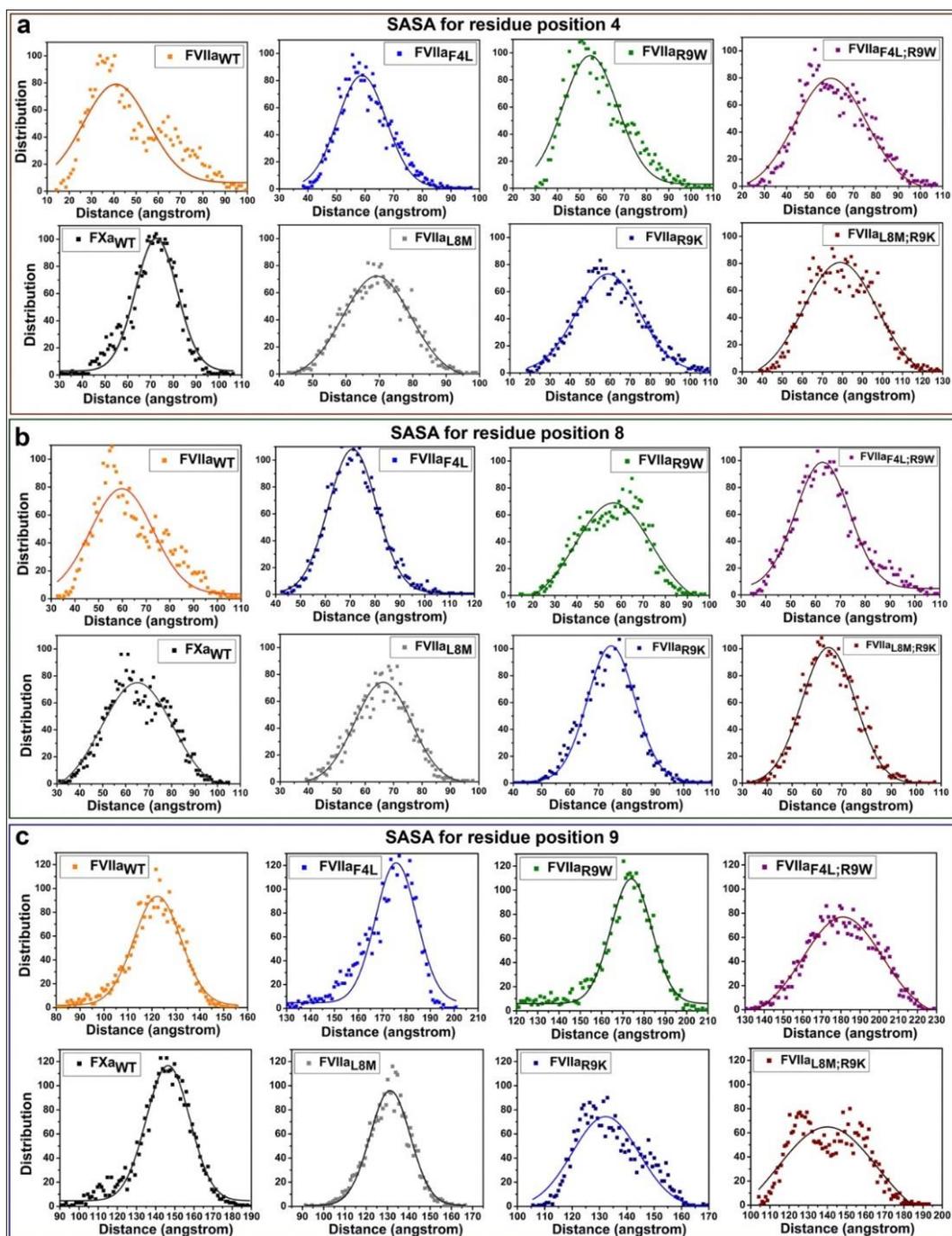


Figure 2-10. Schematic representation of the solvent-accessible surface area (SASA). (a) Distribution of SASA of position 4 for systems FVIIa_{WT} (orange), FVIIa_{F4L} (dark blue), FVIIa_{R9W} (green) and FVIIa_{F4L;R9W} (purple), FXa_{WT} (black), FVIIa_{L8M} (gray), FVIIa_{R9K} (light blue) and FVIIa_{L8M;R9K} (brown). (b) Distribution of SASA of position 8 for systems FVIIa_{WT} (orange), FVIIa_{F4L} (dark blue), FVIIa_{R9W} (green) and FVIIa_{F4L;R9W} (purple), FXa_{WT} (black), FVIIa_{L8M} (gray), FVIIa_{R9K} (light blue) and FVIIa_{L8M;R9K} (brown). (c) Distribution of SASA of position 9 for systems FVIIa_{WT} (orange), FVIIa_{F4L} (dark blue), FVIIa_{R9W} (green) and FVIIa_{F4L;R9W} (purple), FXa_{WT} (black), FVIIa_{L8M} (gray), FVIIa_{R9K} (light blue) and FVIIa_{L8M;R9K} (brown). The distribution plot was prepared based on the last 60 ns simulation of the production run.

4. Conclusions

Briefly, we attempted to envisage the probable reason for the difference in binding affinity of wild-type hFVIIa and hFXa with hEPCR using molecular dynamics simulation and performing free energy calculations. On calculating the average RMSD and RMSF of the GLA-domain, we found that the mutations induce significant conformational alterations in the complex. We suggest that the interacting residue pairs: Leu5 (FVIIa)-Tyr154 (EPCR) and Gla7 (FVIIa)-Gln150 (EPCR) are more prominent in the wild-type FVIIa system and hence may play an important role in imparting higher stability to the complex. On evaluating the protein-lipid interaction, we observe that lipid-dependent protein interaction (between PCh and Phe4) is also enhanced only in wild-type FVIIa system, however in the case of mutants; lipid-based interaction gets hampered. Based on the SASA distribution plot, it can be deduced that when lipid-based protein interaction and protein-protein interaction are absent or weaker, the stability of complex formation is altered. Based on PMF, BindProfX, and interaction energy analyses, we found that mutation at position 9th (FVIIa_{R9K}) based on FXa plays a relevant role in affecting the stability of the whole complex by significantly altering the binding affinity between EPCR and GLA domain and thereby somehow emphasizes the stability of the whole complex. Murine FVIIa contains Try at the 9th position; however, the possible causative effect for the weaker binding affinity may be the spatial position of two or more residues, which in turn induces conformational changes, hence regulating FVIIa binding to the EPCR. We believe that our results provide *in silico* evidence for understanding the structure-function relationship of the protein complex on mutating the key residues.

References

1. Mohan Rao LV, Esmon CT, Pendurthi UR. Endothelial cell protein C receptor: a multi-liganded and multifunctional receptor. *Blood, The Journal of the American Society of Hematology*. 2014 Sep 4; 124(10):1553-62.
2. Fukudome K, Esmon CT. Identification, cloning, and regulation of a novel endothelial cell protein C/activated protein C receptor. *Journal of Biological Chemistry*. 1994 Oct 21; 269(42):26486-91.
3. Huang M, Rigby AC, Morelli X, Grant MA, Huang G, Furie B, Seaton B, Furie BC. Structural basis of membrane binding by Gla domains of vitamin K-dependent proteins. *Nature Structural & Molecular Biology*. 2003 Sep;10(9):751-6.
4. Mizuno H, Fujimoto Z, Atoda H, Morita T. Crystal structure of an anticoagulant protein in complex with the Gla domain of factor X. *Proceedings of the National Academy of Sciences*. 2001 Jun 19;98(13):7230-4..
5. Falls LA, Furie BC, Jacobs M, Furie B, Rigby AC. The ω -loop region of the human prothrombin γ -carboxyglutamic acid domain penetrates anionic phospholipid membranes. *Journal of Biological Chemistry*. 2001 Jun 29;276(26):23895-902.
6. Esmon CT. The endothelial protein C receptor. *Current opinion in hematology*. 2006 Sep 1;13(5):382-5.
7. LÓPEZ- SAGASETA J, Montes R, Puy C, Diez N, Fukudome K, Hermida J. Binding of factor VIIa to the endothelial cell protein C receptor reduces its coagulant activity. *Journal of Thrombosis and Haemostasis*. 2007 Sep;5(9):1817-24.
8. Kondreddy V, Wang J, Keshava S, Esmon CT, Rao LV, Pendurthi UR. Factor VIIa induces anti-inflammatory signaling via EPCR and PAR1. *Blood, The Journal of the American Society of Hematology*. 2018 May 24;131(21):2379-92.
9. Ghosh S, Pendurthi UR, Steinoe A, Esmon CT, Rao LV. Endothelial cell protein C receptor acts as a cellular receptor for factor VIIa on endothelium. *Journal of Biological Chemistry*. 2007 Apr 20;282(16):11849-57.

10. Pendurthi UR, Rao LV. Factor VIIa interaction with endothelial cells and endothelial cell protein C receptor. *Thrombosis research*. 2010 Apr 1;125:S19-22.
11. Sen P, Nayak R, Clark CA, Gopalakrishnan R, Esmon CT, Pendurthi UR, Rao LV. Factor X binding to endothelial cell protein C receptor: comparison with factor VIIa and activated protein C. *Blood, The Journal of the American Society of Hematology*. 2011 Sep 1;118(9):2635-6.
12. Sen P, Clark CA, Gopalakrishnan R, Hedner U, Esmon CT, Pendurthi UR, Rao LV. Factor VIIa binding to endothelial cell protein C receptor: differences between mouse and human systems. *Thrombosis and haemostasis*. 2012;107(05):951-61.
13. Puy C, Hermida J, Montes R. Factor X and factor VII binding to endothelial protein C receptor differs between species. *Journal of Thrombosis and Haemostasis*. 2011 Jun;9(6):1255-7.
14. Schuepbach RA, Riewald M. Coagulation factor Xa cleaves protease- activated receptor- 1 and mediates signaling dependent on binding to the endothelial protein C receptor. *Journal of Thrombosis and Haemostasis*. 2010 Feb;8(2):379-88.
15. Burnier L, Mosnier LO. Novel mechanisms for activated protein C cytoprotective activities involving noncanonical activation of protease-activated receptor 3. *Blood, The Journal of the American Society of Hematology*. 2013 Aug 1;122(5):807-16.
16. Pavani G, Zintner SM, Ivanciu L, Small JC, Stafford KA, Szeto JH, Margaritis P. One amino acid in mouse activated factor VII defines its endothelial protein C receptor (EPCR) binding and modulates its EPCR- dependent hemostatic activity in vivo. *Journal of Thrombosis and Haemostasis*. 2017 Mar;15(3):507-12.
17. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular biology*. 1990 Oct 5;215(3):403-10.
18. Jo S, Lim JB, Klauda JB, Im W. CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophysical journal*. 2009 Jul 8;97(1):50-8.
19. Prasad R, Sen P. Molecular determinants involved in differential behaviour between soluble tissue factor and full-length tissue factor towards factor VIIa. *Physical Chemistry Chemical Physics*. 2017;19(33):22230-42.

20. Prasad R, Sen P. Structural modulation of factor VIIa by full-length tissue factor (TF1-263): Implication of novel interactions between EGF2 domain and TF. *Journal of Biomolecular Structure and Dynamics*. 2018 Feb 17;36(3):621-33.
21. Oganessian V, Oganessian N, Terzyan S, Qu D, Dauter Z, Esmon NL, Esmon CT. The crystal structure of the endothelial protein C receptor and a bound phospholipid. *Journal of Biological Chemistry*. 2002 Jul 12;277(28):24851-4.
22. Zhang X, Vadas O, Perisic O, Anderson KE, Clark J, Hawkins PT, Stephens LR, Williams RL. Structure of lipid kinase p110 β /p85 β elucidates an unusual SH2-domain-mediated inhibitory mechanism. *Molecular cell*. 2011 Mar 4;41(5):567-78.
23. McDonnell JM, Fushman D, Milliman CL, Korsmeyer SJ, Cowburn D. Solution structure of the proapoptotic molecule BID: a structural basis for apoptotic agonists and antagonists. *Cell*. 1999 Mar 5;96(5):625-34.
24. Lu P, Ma D, Yan C, Gong X, Du M, Shi Y. Structure and mechanism of a eukaryotic transmembrane ascorbate-dependent oxidoreductase. *Proceedings of the National Academy of Sciences*. 2014 Feb 4;111(5):1813-8.
25. Eswar N, Webb B, Marti- Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A. Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics*. 2006 Sep;15(1):5-6.
26. DeLano, W. *The PyMOL Molecular Graphics System*; 2003.
27. Banner DW, D'Arcy A, Chène C, Winkler FK, Guha A, Konigsberg WH, Nemerson Y, Kirchhofer D. The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature*. 1996 Mar;380(6569):41-6.
28. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *Journal of molecular graphics*. 1996 Feb 1;14(1):33-8.
29. Villoutreix BO, Blom AM, Dahlbäck B. Structural prediction and analysis of endothelial cell protein C/activated protein C receptor. *Protein engineering*. 1999 Oct 1;12(10):833-40.
30. López-Sagasetta J, Puy C, Tamayo I, Allende M, Cerveró J, Velasco SE, Esmon CT, Montes R, Hermida J. sPLA2-V inhibits EPCR anticoagulant and antiapoptotic properties by accommodating lysophosphatidylcholine or PAF in the

- hydrophobic groove. *Blood*, The Journal of the American Society of Hematology. 2012 Mar 22;119(12):2914-21.
31. Prasad R, Sen P. Phosphatidylcholine in the groove of endothelial cell protein C receptor (EPCR) regulates EPCR conformation and protein C recognition. *Integrative Biology*. 2018 Nov 1;10(11):696-704.
 32. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*. 1983 Jul 15;79(2):926-35.
 33. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. *Journal of computational chemistry*. 2005 Dec;26(16):1781-802.
 34. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan SA, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*. 1983 Jun;4(2):187-217.
 35. Darden T, York D, Pedersen L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of chemical physics*. 1993 Jun 15;98(12):10089-92.
 36. Martyna GJ, Tobias DJ, Klein ML. Constant pressure molecular dynamics algorithms. *The Journal of chemical physics*. 1994 Sep 1;101(5):4177-89.
 37. Feller SE, Zhang Y, Pastor RW, Brooks BR. Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of chemical physics*. 1995 Sep 15;103(11):4613-21.
 38. Andersen HC. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *Journal of computational Physics*. 1983 Oct 1;52(1):24-34.
 39. Darve E, Rodríguez-Gómez D, Pohorille A. Adaptive biasing force method for scalar and vector free energy calculations. *The Journal of chemical physics*. 2008 Apr 14;128(14):144120.
 40. Allen TW, Andersen OS, Roux B. Molecular dynamics—potential of mean force calculations as a tool for understanding ion permeation and selectivity in narrow channels. *Biophysical chemistry*. 2006 Dec 1;124(3):251-67.

41. Xiong P, Zhang C, Zheng W, Zhang Y. BindProfX: assessing mutation-induced binding affinity change by protein interface profiles with pseudo-counts. *Journal of molecular biology*. 2017 Feb 3;429(3):426-34.
42. Cukuroglu E, GURSOY A, NUSSINOV R, KESKIN O. Non-redundant unique interface structures as templates for modeling protein interactions. *PloS one*. 2014 Jan 27;9(1):e86738.
43. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic acids research*. 2005 Jul 1;33(suppl_2):W382-8.
44. Serçinoğlu, O.; Ozbek, P. GRINN: A Tool for Calculation of Residue Interaction Energies and Protein Energy Network Analysis of Molecular Dynamics Simulations. *Nucleic acids research* 2018, 46 (W1), W554–W562.
45. Pavani G, Ivanciu L, Faella A, Marcos-Contreras OA, Margaritis P. The endothelial protein C receptor enhances hemostasis of FVIIa administration in hemophilic mice in vivo. *Blood, The Journal of the American Society of Hematology*. 2014 Aug 14;124(7):1157-65.
46. Vermeer C. Gamma-carboxyglutamate-containing proteins and the vitamin K-dependent carboxylase. *Biochemical Journal*. 1990 Mar 15;266(3):625.
47. Kortemme T, Joachimiak LA, Bullock AN, Schuler AD, Stoddard BL, Baker D. Computational redesign of protein-protein interaction specificity. *Nature structural & molecular biology*. 2004 Apr;11(4):371-9.

Chapter 3

A molecular dynamics simulation study to elucidate the effect of Cholesterol and Tissue Factor Palmitoylation on TF-FVIIa-FXa ternary complex in different lipid environments.

1. Introduction

Tissue factor (TF), a transmembrane glycoprotein, primarily initiates coagulation in both physiological and pathological conditions and plays a pertinent role in maintaining homeostasis, thrombosis, and vascular development [1]. Tissue factor is normally expressed on the surface of all non-vascular cells constitutively. Once there is an injury to the blood vessel, the endothelial layer disrupts, and TF is exposed to intravascular allosteric coagulation Factor VIIa (FVIIa) and forms a binary TF-FVIIa complex. Thereafter coagulation cascade gets triggered by inducing sequential activation of clotting proteins Factor IX and Factor X into factor IXa and FXa, respectively. This eventually causes thrombin burst, thus leading to the formation of clots followed by the generation of insoluble fibrin meshwork from soluble fibrinogen [2].

Previous studies have put forward the proposition that TF contributes significantly to balancing of initiation and propagation of thrombus on the rupture of atherosclerotic plaque. It has been found that the agglomeration and activity of TF are more in the lesions collected from patients with unstable angina, myocardial infarction, and other acute coronary events [3]. Various studies confirm the presence of cholesterol/ oxidatively modified low-density lipoprotein (LDL) in atherosclerotic plaques and are considered to confer importantly to atherogenesis. Many studies were conducted to investigate the effect of cholesterol on TF expression [4]. Cholesterol consists of four hydrocarbon steroid rings placed in between the hydroxyl group and the hydrocarbon chain. Though mostly hydrophobic, the presence of the hydroxyl group imparts an amphiphilic property to the compound, characteristic of the lipid components of cellular membranes. Studies show that cholesterol can strongly influence the affinity state, binding capacity, and signal transduction property of membrane receptors by highly specific molecular interactions or by changing

membrane fluidity [5]. Membrane cholesterol modulates the flexibility and mechanical stability of the membrane and contributes to the regulation of various cellular processes such as endocytosis, and intracellular trafficking of receptors. Cholesterol also imparts a role in the differentiation and maintenance of cell surface microdomains of varying lipid composition, predominantly consisting of sphingolipid rafts. It is reported that Cholesterol and sphingolipid-rich rafts associate with caveolin, a structural protein, to form caveolae, which are present in the plasma membrane of many cell types, including smooth muscle cells and endothelial cells. These caveolae are flask-shaped invaginations that are 50-nm to 100-nm in diameter and their structural integrity is largely dependent on cholesterol. Upon removal of cholesterol, the structure of the caveolae gets drastically altered [6]. Recent studies suggest the linkage of TF with caveolae in smooth muscle cells and it is postulated that once the integrity of the vessel wall is lost, caveolae-associated TF is rapidly activated thus functioning as a latent pool of procoagulant activity [7]. Few studies highlighted that TF redistributes into caveolae following a sequence of events, that include TF binding to VIIa, generation of FXa, and subsequent formation of a transient ternary complex of TF-VIIa with FXa and consequent translocation and assembly with TFPI (tissue factor pathway inhibitor) in glycosphingolipid-rich microdomains. This provides evidence for the significance of caveolae in the control of cell surface proteolytic activity of TF. Previous studies also reveal that Caveolae/lipid rafts contribute to promoting protein-protein and protein-lipid interactions. It is also demonstrated that lipid rafts/caveolae-associated TF represents the encrypted form, and when such structures are disintegrated, TF can be activated rapidly [8]. Several groups of studies have shown that when monocytes/macrophages are exposed to modified LDL cholesterol, TF expression was induced [9-14]. However, few studies are inconsistent with these findings on the effect of cholesterol on TF expression [15-18]. In one such study, it is reported that as a result of treating HEK293 cells with methyl β -cyclodextrin (m β CD), lipid rafts get disrupted, which causes subsequent removal of cholesterol from the membrane, followed by an increase in TF activity at the cell surface [19]. Howbeit, while investigating how cholesterol plays a part in the control of TF receptor function in fibroblasts and cancer cells, it was found that depletion of membrane cholesterol by using methyl- β -cyclodextrin weakens the functional activity of TF by reducing its affinity to FVIIa and

subsequent activation of FX [20,21]. Moreover, these studies have proposed that the reduced cholesterol content at the cell surface accounts for lowered TF activity in cholesterol-depleted cells, and not because of the disruption of caveolar structure, thus implying the role of cholesterol in modulating TF affinity to FVIIa. Accordant with this finding, dietary lipid-lowering was found to reduce TF expression in rabbit atheroma [22]. Recently, to explore the coagulation-inducing potential of cholesterol crystals (CC), a detailed study was performed in lepirudin-based human whole blood and plasma models. Their findings signify the role of CC in thrombosis following plaque rupture and activating coagulation through thrombo-inflammation.

In our present study, we have tried to characterize the dynamics and interactions of Cholesterol with the TF-FVIIa-FXa Ternary complex using the molecular dynamics (MD) simulation approach. The present study may throw light on the contradictory experimental results and help in unveiling the underlying molecular mechanism of how cholesterol affects structural modulations on the TF-FVIIa-FXa Ternary complex. MD simulations permit membrane protein structures to be re-embedded into lipid bilayers *in silico* so that they can provide valuable insights into the prediction and identification of protein-lipid interactions thus augmenting advances in understanding the structural modulation and function of such proteins [23].

Our study also attempts to assess the effect of Palmitoylation on protein-lipid interactivity in presence of cholesterol. Palmitoylation is a reversible post-translational modification that can dynamically regulate protein structure and function and thereby influencing membrane binding and targeting. Palmitoylation consists of the attachment of the fatty acid palmitate to the cysteine residues often located near the transmembrane region of the protein via a thioester bond. Aberrant palmitoylation is related to various diseases including neurological disorders like Huntington's disease, Parkinson's disease, Alzheimer's disease, metabolic disorders, and Cancer [24-26]. Recent studies report that palmitoylation facilitates both protein-protein and protein-lipid interactions, and does influence the structure and function of membrane proteins by targeting them to cholesterol and sphingolipid-rich microdomains and caveolae [27-29]. One previous study suggests that palmitoylation may modulate TF procoagulant activity by regulating TF cytoplasmic domain phosphorylation and maintaining TF in a monomeric

form [30-31]. However, in a recent study, it is reported that the increase in the procoagulant activity is initiated by depalmitoylation and appears to be independent of TF phosphorylation [32]. It is important to consider that the association between TF palmitoylation and the aforementioned contributory factors remains unclear [33]. Therefore, computational investigation using the MD simulation approach may provide some clues to understanding the influence of TF palmitoylation on the structure and dynamics of the TF-FVIIa-FXa Ternary complex.

2. Methodology

2.1 System Set up:

We have built two mixed lipid bilayer systems, one having a composition of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylcholine (POPC) and 1-palmitoyl-2-oleoyl-sn-glycero-3-PhosphatidylSerine (POPS) in 4:1 ratio and the other having POPC, POPS, and Cholesterol (CHL) in 3:1:1 ratio. These rectangular lipid bilayer models containing a total of 200 lipids (100 per leaflet) are created using the CHARMM-GUI membrane builder server [34]. During membrane construction, the system charge was neutralized by randomly placing Na^+ and Cl^- ions in the bulk water. Both the systems are subjected to equilibration using consecutive CHARMM input files in the NPT ensemble (constant pressure and temperature) for 8–10 ns. We found that the POPC: POPS lipid bilayer is equilibrated with an average area per lipid bilayer of $\sim 61 \text{ \AA}^2$ and a bilayer thickness of 40 \AA at the end of the simulation. We also found POPC: POPS: CHL lipid bilayer system is equilibrated with an average area per lipid bilayer of $\sim 51 \text{ \AA}^2$ and a bilayer thickness of 44 \AA at the end of the simulation. This observation is in line with previous studies as the presence of cholesterol in lipid bilayers is known to cause a decrease in the area per lipid [35]. It is known that the lipid membrane behaves as an incompressible fluid, as a result when the area per lipid decreases, it causes the bilayer thickness to increase [36]. The last membrane coordinates were used for modeling membrane-bound TF-FVIIa-FXa (TERNARY COMPLEX). Detailed system size and the simulation description of all systems are provided in Table 3-1.

Table 3-1: Details of system size and the simulation description of all systems.

System	TF-FVIIa-FXa Ternary Complex in POPC: POPS (CONTROL)	TF-FVIIaFXa Ternary Complex in POPC: POPS: CHL (CHOL)	PalmitoylatedTF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CYSP)
Lipid Composition	160 (POPC): 40 (POPS) (100 in upper leaflet, 100 in lower leaflet) (Initial composition from CHARMM membrane builder)	120 (POPC): 40 (POPS): 40(CHL) (100 in the upper leaflet, 100 in the lower leaflet) (Initial composition from CHARMM membrane builder)	120 (POPC): 40 (POPS): 40(CHL) (100 in the upper leaflet, 100 in the lower leaflet) (Initial composition from CHARMM membrane builder)
Initial box size (Å³)	123 (x) x 123 (y) x 123 (z) = 1860867 Å ³	150 (x) x 150 (y) x 200 (z) = 4500000 Å ³	150 (x) x 150 (y) x 200 (z) = 4500000 Å ³
Number of Ca²⁺ ions	16 (9 bound + 7 additional ions)	16 (9 bound + 7 additional ions)	16 (9 bound + 7 additional ions)
Number of Cl⁻ ions	17 (1 bound + 16 additional ions)	17(1 bound + 16 additional ions)	17 (1 bound + 16 additional ions)
Water model	TIP3P	TIP3P	TIP3P
Number of water molecules	65972	63775	63716
Number of total atoms	266427	253292	253166
Integration time step size	2fs	2fs	2fs
Simulation time for production run in NPT/NPnTA ensemble	200ns	200ns	200ns

2.2 Modeling of protein target—TF-FVIIa-FXa (TERNARY COMPLEX)

The starting coordinates of the sTF-FVIIa complex with 7Ca²⁺ ions and Cl⁻ ion were extracted from the X-ray crystal structure (PDB code: 1dan), followed by the removal of the co-crystallized inhibitor [37]. Deriving knowledge from correlated residues from another crystal structure of FVIIa (PDB code: 1qfk), the missing C-terminal residues, 143-144 of FVIIa (light chain) were modeled [38]. Template segment (97-104) of chain L from FXa (PDB code: 2h9e) was used to model remaining absent residues, 145-152 [39]. To model the full-length TF (residue, 1-263), the X-ray crystal structure of sTF (PDB code: 1boy) was used to model the initial coordinates [40]. In the beginning, the two missing residues, Ser1 and Gly2 of TF were modeled on knowledge derived from the template structure of sTF (PDB code: 2hft) and were connected with the crystal structure of sTF (PDB code: 1boy) having residues 3-213. The transmembrane region of TF from Ile223-Leu242 was built as described by Lee et al [41-42]. Building on a clue from the similarity search (having the highest protein BLAST score) the connective loop region (residues 214-222) was constructed using the template chain A (residues 199-207) of carbonic anhydrase 2 (PDB code: 2w3n) [43,44]. For building the binary complex model, we have considered similar criteria as mentioned by Lee et al. Firstly, alignment of the equilibrated membrane was done with the z-axis. Secondly, insertion of TF-FVIIa complex into the phospholipid bilayer in a manner such that hydrophobic residues of the GLA domain ω -loop are put into the interstitial region of the phospholipid bilayer. Thirdly, penetration of the transmembrane helix into the membrane lipid. Accelrys Discovery Studio software was used to join the connective loop region (214-222) to the TF with the transmembrane part (TM) along with the visual examination. We measured the CT (catalytic triad) distance of the FVIIa modeled structure and found it to be around ~78-80 Å from the head group of membrane lipid. Finally, the cytoplasmic tail region (245-263) of TF was extracted from the NMR structure of the cytoplasmic part of TF (PDB code: 2ceh) [45]. PyMol software was used to construct the coordinates of the two missing residues (243-244) and to join them to the N-terminal region of the cytoplasmic part. Thereafter, the TM portion was joined with manual adjustment of the orientation of side chains of connected residues. After

minimization and equilibration of TF-FVIIa binary complex, we have modeled ternary complex TF-FVIIa-FX by superposition of putative ternary complex sTF-FVIIa-FX. The putative sTF-FVIIa-FXa ternary complex was taken from MD refined structure (PDB code: 1NL8) for initial model building [46]. To generate membrane-bound full-length TF-FVIIa-FX ternary complex, the coordinates of FXa along with bounded calcium ions from the ternary complex (TF-FVIIa-FXa) were taken after the superposition with equilibrated membrane-bound full-length TF-FVIIa complex structure. Missing residues of the FX-light chain having numbering Val144-Glu172 were modeled from the mouse HP1 (M31), (PDB code: 1DZ1, chain A, residues 21-49) as a template based on a 45% similarity search in protein BLAST [47]. For modeling, the missing residues (Asn173-Thr193), and the X-Ray crystallographic structure APA protein (PDB code: 3DMQ, chain A, residues 721-741) were used as templates [48]. The linker region (Arg140-Ser143) was joined with Arg139 of FX (light chain) and linked with the N-terminus of Val144. Discovery studio was used to join the missing residues (173-193) with the C-terminus of Glu172 along with the visual examination and manual adjustment of the side chains of the residues.

Previous studies reported the acylation of cysteine (residue number 245) in the cytoplasmic region of TF by both palmitic acid and stearic acid [49]. So, we have modified the CYS residue of TF at position 245 to CYSP (palmitoylated) using the “Add lipid tail option” in the PDB Manipulation section available in CHARMM GUI Server. Finally, three membrane-bound modeled structures were built and were subjected to MD simulation to generate the all-atom solvent-equilibrated model. These three systems are : a) TF-FVIIa-FXa Ternary Complex in POPC:POPS (CONTROL) b) TF-FVIIa-FXa Ternary Complex in POPC:POPS:CHL(CHOL) c) PalmitoylatedTF-FVIIa-FXa Ternary Complex in POPC:POPS:CHL (CYSP) (Figure 3-1).

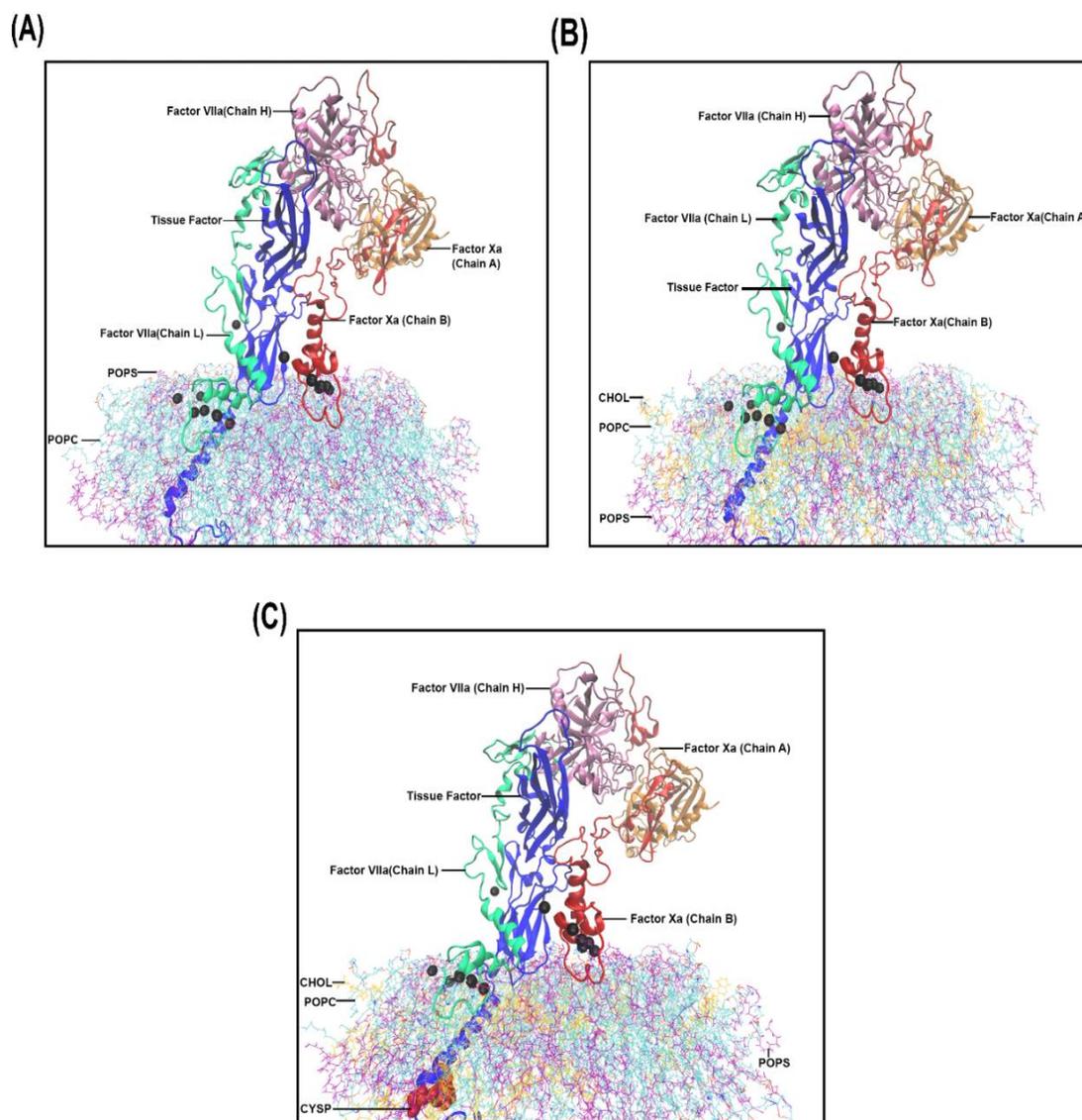


Figure 3-1. 3D representation of (A)TF-FVIIa-FXa Ternary Complex in POPC: POPS (CONTROL), (B)TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CHOL), and (C)Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CYSP) used in the present study. Tissue Factor, Factor VII(chain), Factor VII(chain), Factor X (chain A), and Factor X (chain B), are shown in the ribbon representation. FVIIa-bound and FX-bound Ca^{2+} ions in the GLA domain are represented as black spheres. The tissue Factor is shown by a light yellow color ribbon representation. The light chain of FVIIa (chain L), is represented as a green color ribbon. The heavy chain of FVII (chain H) is shown by the mauve color ribbon representation. Factor X (chain A), is shown as a dark yellow color ribbon representation. Factor X (chain B) is represented as a red color ribbon. In the bottom panel, phospholipid bilayer is shown in lines representation wherein POPC is shown as purple color lines, POPS shown as ice blue color lines, and CHOL in blue color lines. In the bottom middle panel of (C), CYSP is represented by a VDW sphere in brown color.

2.3 MD simulation setup protocol:

Atomistic simulations were performed using the NAMD-2.9 dynamics software package with CHARMM36 force field Parameter [50-52]. Topology and parameter files for Gla residue (γ -carboxyglutamate residue, post-translational modified groups) of FVIIa were constructed by adoption from similar chemical groups present in topology and parameter files. Thereafter, Protein Structure File (PSF) was generated using the Psfgen package tool of VMD for all three systems [53]. The systems were solvated using the TIP3P water model in a periodic box having a dimension of $128 \times 128 \times 194 \text{ \AA}^3$ with a spacing of at least 12 \AA from any given atom [54]. The total number of water molecules was 65,972. 12 Cl^- ions and 7 Ca^{2+} were added by replacing water molecules randomly to achieve the desired 5mM CaCl_2 concentration using AUTOIONIZE module in the VMD. The temperature of all the systems was kept constant at 310 K using the damping coefficient (γ) of 1 ps^{-1} by Langevin dynamics. The Particle Mesh Ewald (PME) method was used to calculate long-range electrostatic interactions without truncation [55]. The direct non-bonded potential cut-off was set to 12 \AA with a 2 \AA pair-list distance cut off and a 1-4 scaling factor was used for all three systems. The pressure was kept constant at 1 atm by using the Langevin pistonNosé-Hoover method using the periodic boundary conditions [56,57]. Before structural equilibration, the model complexes were subjected to several cycles of energy minimization and relaxation to eliminate the bad contacts (steric clashes) present during the model building of the complex. This was employed to ensure the simulation systems are stable and free. In the first step, dynamics were performed on all lipid molecules, water molecules, and counterions for 0.5 ns simulation with a 1fs time interval using a short “constant pressure and Temperature” (NPT ensemble) for packing of lipid and against the modeled protein and the counterions to relax their positions; while the modeled protein was kept fixed. In this condition, the energy of the system was minimized using the conjugate gradient method for 40,000 steps to remove steric clashes. In the second step, for a period of 2 ns with 2 fs timestep, short equilibration was done using NPT with protein constrained, in which the SHAKE algorithm was

applied for hydrogen atoms [58]. Constraining of protein atoms was done with a small harmonic potential (spring constant of 1.0 kcal/mol/Å²). In successive cycles, the whole system was switched to NPnAT ensemble (constant membrane-normal pressure, Pn, temperature, and membrane area) and simulation was performed without any constraint on the proteins with a 2fs time interval for 200 ns production run.

3. Results and Discussion

To understand the conformational dynamics of TF-FVIIa-FXa Ternary Complex in presence of cholesterol and also to explore the effect of Tissue Factor palmitoylation on the TF-FVIIa-FXa Ternary Complex, we had set up three systems: a) TF-FVIIa-FXa Ternary Complex in POPC: POPS (CONTROL) b) TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CHOL) c) PalmitoylatedTF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CYSP) as shown in (Figure 3-1). Thereafter, we have subjected all these systems individually to MD simulation for 200ns.

3.1 Protein Dynamics Study

Structurally, FVIIa comprises four domains: GLA (gamma- carboxy glutamic acid-rich) domain, Epidermal growth factor-like domains: EGF1 and EGF2, and Serine Protease (SP) domain. Crystal structure of binary complex revealed that FVIIa binds seven Ca²⁺ ions in the GLA-domain, one in the EGF1 domain, and one in the protease domain [59]. Based on folding pattern and conformation, the SP domain of FVIIa consists of two structural motifs, designated as A1E1 (residues 30-97) and A2F2 (residues 134-228). A1E1 comprises the 60s and Ca²⁺ binding loop and the A2F2 region is constituted of the 170s loop, activation loop 1, loop 2, and loop 3 [60]. To predict the stability of the system, root means square deviations (RMSDs) of various domains were calculated over time relative to the initial structure for each system. We observed that all the systems have attained equilibration and we computed average RMSDs for the key domains individually (Figure 3-2) (Table 3-2). We investigated Ca-RMSDs for both FVIIa_60s Domain (A1E1) and FVIIa_170 Loop (A2F2) regions separately. The average deviation in region FVIIa_60s Domain (1.89 ± 0.11 Å) in

CONTROL was only slightly higher than found in CHOL($1.86 \pm 0.12 \text{ \AA}$) and CYSP ($1.84 \pm 0.08 \text{ \AA}$). Similar differences in average RMSDs were observed in the FVIIa_170 Loop among CONTROL ($2.77 \pm 0.37 \text{ \AA}$), CHOL ($2.60 \pm 0.24 \text{ \AA}$), and CYSP ($2.44 \pm 0.19 \text{ \AA}$).

The previous report suggests that when TF is bound to FVIIa, the interaction of the GLA(gamma-carboxy glutamic acid-rich) domain of FVIIa with the lipid membrane is increased thus enhancing the catalytic activity of FVIIa and giving rise to stronger membrane anchoring ability. So, we investigated C α -RMSDs for the FVIIa GLA domain and thereafter obtained similar variations, i.e. ($2.71 \pm 0.38 \text{ \AA}$) in CONTROL, ($2.42 \pm 0.35 \text{ \AA}$) in CHOL, and ($2.41 \pm 0.33 \text{ \AA}$) in CYSP. The light chain of FVIIa contains two epidermal growth factor domains EGF1 and EGF2 comprising residues numbering from 47-83 and 87-128, respectively. We observed similar variations while investigating C α -RMSDs for the FVIIa EGF1 domain i.e. ($2.17 \pm 0.76 \text{ \AA}$) in CONTROL, ($1.87 \pm 0.47 \text{ \AA}$) in CHOL, and ($1.57 \pm 0.23 \text{ \AA}$) in CYSP and for FVIIa EGF2 domain i.e. ($2.51 \pm 0.55 \text{ \AA}$) in CONTROL, ($1.48 \pm 0.44 \text{ \AA}$) in CHOL and ($2.48 \pm 0.40 \text{ \AA}$) in CYSP. Further, we found that the average RMSDs of SP domain in TF-FVIIa-FXa Ternary Complex(CONTROL) in POPC:POPS ($6.01 \pm 1.59 \text{ \AA}$) is comparatively higher than in TF-FVIIa-FXa Ternary Complex(CHOL) in POPC:POPS:CHL ($5.31 \pm 0.89 \text{ \AA}$) and Palmitoylated TF-FVIIa-FXa Ternary Complex(CYSP) in POPC:POPS:CHL ($4.62 \pm 0.46 \text{ \AA}$) suggesting that the presence of Cholesterol does have an impact on the stability of the SP domain. Next, we examined C α -RMSDs for FXa GLA domain and observed variations i.e. ($2.16 \pm 0.26 \text{ \AA}$) in CONTROL, ($2.08 \pm 0.21 \text{ \AA}$) in CHOL and ($2.01 \pm 0.22 \text{ \AA}$) in CYSP respectively. Similarly, we observed variations for the FXa SP domain i.e. ($1.81 \pm 0.19 \text{ \AA}$) in CONTROL, ($1.80 \pm 0.17 \text{ \AA}$) in CHOL, and ($1.81 \pm 0.19 \text{ \AA}$) in CYSP respectively. Finally, we checked C α -RMSDs for full Tissue Factor and we noted variations i.e. ($3.16 \pm 0.43 \text{ \AA}$) in CONTROL, ($3.14 \pm 0.35 \text{ \AA}$) in CHOL, and ($3.04 \pm 0.28 \text{ \AA}$) in CYSP and for Tissue Factor Domain1 i.e. ($1.88 \pm 0.16 \text{ \AA}$) in CONTROL, ($1.78 \pm 0.14 \text{ \AA}$) in CHOL and ($1.69 \pm 0.12 \text{ \AA}$) in CYSP.

To understand the overall residue-wise protein dynamics in the protein complex, the evaluation of root mean square fluctuations (RMSF) plays a very important role. Therefore, to check residue-wise stability we have calculated the root mean square fluctuations (RMSFs) for the last 60 ns of the total simulation run (Figure 3-3).

As per recent studies, the FVIIa 60s domain and FVIIa_170 Loop might play a pertinent role in the process of allosteric modulation by forming a hinge-like structure that stabilizes the catalytic triad at an optimal position [61]. Therefore, we have computed RMSF for both the domains and we observed the residue-wise fluctuation in CHOL and CYSP is much lower than CONTROL (Figure 3-3A, 3-3B). A similar trend is observed in the RMSF profile of the FVIIa EGF1 domain comprising residues 47-83 and the FVIIa EGF2 domain comprising residues 87-128 (Figure 3-3C, 3-3D). Likewise, the RMSF profile of the FVIIa GLA domain (residues 1-38), FVIIa SP domain (residues 16-269), and FXa GLA domain (residues 1-46) indicate greater stability and minimum fluctuation in CHOL and CYSP systems as compared to CONTROL. (Figure 3-3E, 3-3F, 3-3G). As per previous findings, residues 142-154 (contact region) at the A2F2 region of the SP domain of FVIIa are responsible for substrate (FIX, FX) recognition and binding [59]. Greater fluctuation at these residues may lessen substrate recognition and retention capability. So, we also checked the RMSF profile of the same and found that CHOL and CYSP systems exhibit maximum stability than CONTROL (Figure 3-3H). Lastly, we also examined the RMSF profile of Tissue Factor comprising residues (1-219) and domain 1 of Tissue Factor comprising residues 157-167 responsible for efficient proteolytic activity of FVIIa and FXa and we obtained similar observations in the RMSF profile as earlier (Figure 3-3I, 3-3J).

Our findings suggest that the presence of Cholesterol allosterically affects the structural integrity of the complex and the presence of Palmitoylation of the Tissue factor additionally influences the stability of the Ternary complex and membrane association which is also substantiated by the experimental studies as mentioned earlier.

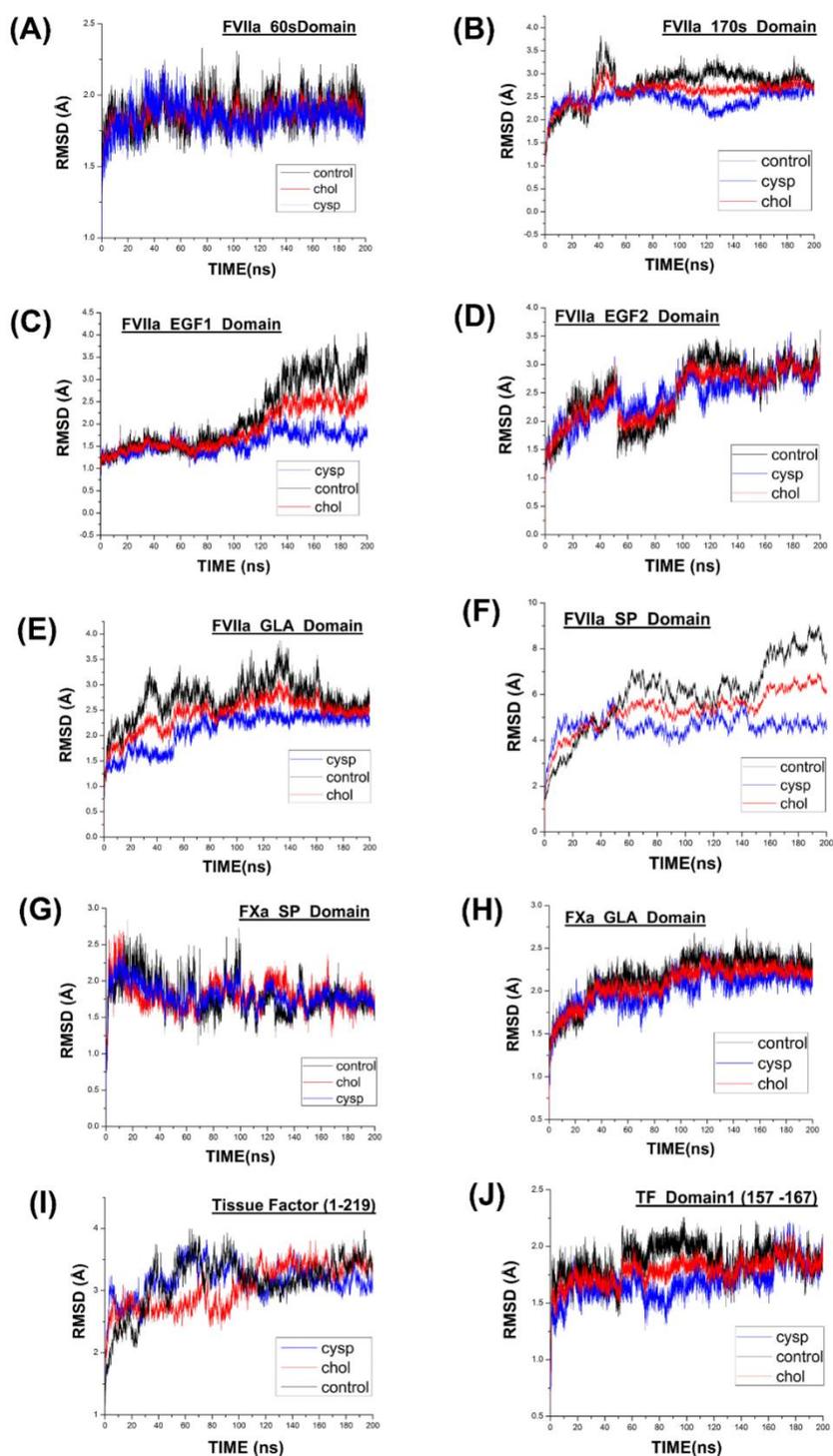


Figure 3-2. Root mean square deviations (RMSD) obtained from simulation of TF-FVIIa-FXa Ternary Complex in POPC: POPS (CONTROL), TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CHOL), and Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CYSP) for total 200ns. (A) RMSD of FVIIa_60s Domain (A1E1 region, Ca^{2+} binding loop, comprising residues 30-97) (B) RMSD of FVIIa_170 Loop (A2F2 region, TF binding domain, activation loop 1,2,3 comprising residues 134-228) (C) RMSD of FVIIa EGF1 Domain comprising residues 47-83 (D) RMSD of FVIIa EGF2 Domain comprising residues 87-128 (E) RMSD of FVIIa GLA Domain

comprising residues 1-38 (F) RMSD of FVIIa SP Domain comprising residues 16-269 (G) RMSD of FXa GLA Domain 1-46 (H) RMSD of A2F2 region of SP domain responsible for binding to FX comprising residues 142-154 (I) RMSD of Tissue Factor comprising residues 1-219 (J) RMSD of Domain 1 Tissue Factor comprising residues 157-167.

Table 3-2: Average C α root mean square deviation (RMSDs) and standard deviation (SD) of various domains of TF-FVIIa-FXa Ternary Complex in POPC: POPS (CONTROL), TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CHOL), and Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CYSP)

Domain/Residues	System		
	TF-FVIIa-FXa Ternary Complex in POPC: POPS (CONTROL) C α RMSDs (Å) Mean \pm S.D.	TF-FVIIaFXa Ternary Complex in POPC:POPS: CHL (CHOL) C α RMSDs (Å) Mean \pm S.D.	PalmitoylatedTF-FVIIa-FXa Ternary Complex in POPC:POPS: CHL (CYSP) C α RMSDs (Å) Mean \pm S.D.
FVIIa GLA domain (residues 1-38)	2.71 \pm 0.38	2.42 \pm 0.35	2.41 \pm 0.33
EGF1 (residues 47-83)	2.17 \pm 0.76	1.87 \pm 0.47	1.57 \pm 0.23
EGF2 (residues 87-128)	2.51 \pm 0.55	a. \pm 0.44	2.48 \pm 0.40
A1E1 region (residues 30-97)	1.89 \pm 0.11	1.86 \pm 0.12	1.84 \pm 0.08
A2F2 region (residues 134-228)	2.77 \pm 0.37	2.60 \pm 0.24	2.44 \pm 0.19
FXa GLA domain (residues 1-46)	2.16 \pm 0.26	2.08 \pm 0.21	2.01 \pm 0.22
FVIIa SP domain (residues 1-13)	6.01 \pm 1.59	5.31 \pm 0.89	4.62 \pm 0.46
FXa SP domain (residues 16-269)	1.81 \pm 0.19	1.80 \pm 0.17	1.79 \pm 0.23
Tissue Factor (residues 1-219)	3.16 \pm 0.43	3.14 \pm 0.35	3.04 \pm 0.28
Tissue Factor Domain 1 (residues 157-167)	1.88 \pm 0.16	1.78 \pm 0.14	1.69 \pm 0.12

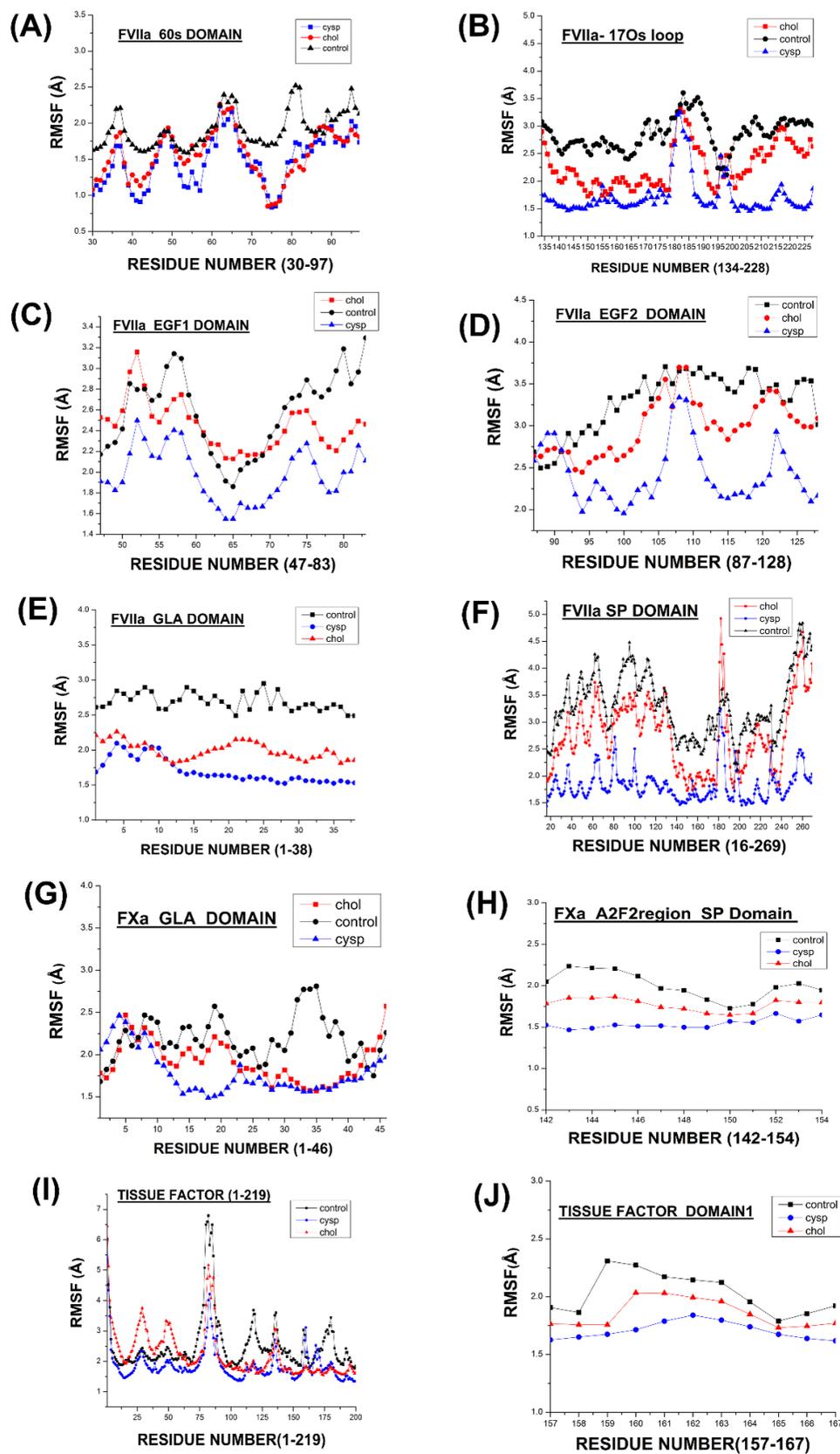


Figure 3-3. Root mean square fluctuation (RMSF) obtained from simulation of TF-FVIIa-FXa Ternary Complex in POPS: POPS (CONTROL), TF-FVIIa-FXa Ternary Complex in POPS:POPS: CHL (CHOL), and Palmitoylated TF-

FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CYSP) for the last 60 ns of the total simulation run. (A) RMSF of FVIIa_60s Domain (A1E1 region, Ca²⁺ binding loop, (residues 30-97) (B) RMSF of FVIIa_170 Loop (A2F2 region, TF binding domain, activation loop 1,2,3 (residues 134-228) (C) RMSF of FVIIa EGF1 Domain (residues 47-83) (D) RMSF of FVIIa EGF2 Domain (residues 87-128) (E) RMSF of FVIIa GLA Domain (residues 1-38) (F) RMSF of FVIIa SP Domain (residues 16-269) (G) RMSF of FXa GLA Domain (residues 1-46) (H)RMSF of A2F2 region of SP domain responsible for binding to FX (residues 142-154)(I) RMSF of Tissue Factor (residues1-219) (J) RMSF of Domain 1 Tissue Factor (residues 157-167).

3.2 Conformation analysis of complexes

We have measured the angle (acute angle formed between domains 1 and 2) among C α -atom of residues Thr86, Phe19, and Ser162 and plotted the distribution as mentioned previously to check the structural integrity of the complexes [62]. We have observed varying distribution patterns for all the systems. TF-FVIIa-FXa Ternary Complex (CONTROL) showed wider angular fluctuation having a maximal population at $\sim 127^\circ$, while TF-FVIIa-FXa Ternary Complex (CHOL) and Palmitoylated TF-FVIIa-FXa Ternary Complex (CYSP), showed narrow fluctuation having maxima at $\sim 136^\circ$ and $\sim 133^\circ$ respectively. [Figure 3-4(A-D)] Comparative wider angular fluctuation in TF-FVIIa-FXa Ternary Complex (CONTROL) indicates lesser structural integrity of the subdomains of TF concerning TF-FVIIa-FXa Ternary Complex (CHOL) and Palmitoylated TF-FVIIa-FXa Ternary Complex (CYSP).

Previous studies demonstrated that residues 157-167 of TF are essential for efficient proteolytic activation of FVII and FX [63]. Therefore, we calculated SASA (Solvent accessible surface area) for the residues 157-167 of TF with probe radius 1.4 Å for the last 60 ns simulation time. SASA will help us to predict the hydration of the hydrophobic core that is also associated with the PPI (protein-protein interaction) indirectly portraying the 3D structure of the protein [64]. From our SASA analysis, it can be inferred that SASA distribution is higher (963 \AA^2) for CONTROL as compared to CHOL (707 \AA^2) and CYSP (645.32 \AA^2). (Figure 3-4E, 3-4F). The lower SASA

indicates that the hydrophobic core (157-167 region consists of hydrophobic residues) is more buried and the difference in SASA values indicates the presence of cholesterol and Tissue Factor palmitoylation strongly guides the 3D structure of the 157-167 region of TF. Additionally, we checked SASA distribution for the cytoplasmic tail part (residues 243-263) of the Tissue Factor which is known to be involved in the signaling function [65]. We found that SASA distribution is lower for (997.407 Å²) for CYSP and CHOL (1063.14 Å²) as compared to CONTROL (1194.52 Å²) which suggests that palmitoylation of CYS²⁴⁵ residue in the cytoplasmic tail of TF decreases the SASA values, enhances hydrophobicity of proteins thus indicating more buried structure with increased protein-protein interaction.

Previous reports suggest that the GLA domain plays an essential role during protein-lipid interaction [66]. Hence, for quantifying the rigidity of the GLA domain and to predict the compactness of the protein structure, we calculated the radius of gyration (R_{gyr}) for GLA domains of both FVIIa and FXa. We found that the R_{gyr} of the GLA domain of FVIIa is much higher for CONTROL (0.44 ± 0.22 nm) than CHOL (0.33 ± 0.12 nm) and CYSP (0.28 ± 0.09 nm). A similar trend is observed in the Radius of gyration (R_{gyr}) for GLA domains of FXa i.e., R_{gyr} of GLA domain of FXa is much higher for CONTROL (0.29 ± 0.11 nm) than CHOL (0.19 ± 0.07 nm) and CYSP (0.09 ± 0.05 nm) (Figure 3-4 G-H). These observations suggest that the presence of cholesterol and Palmitoylation of the Tissue Factor contributes to the structural rigidity and compactness of the GLA domains of both FVIIa and FXa respectively.

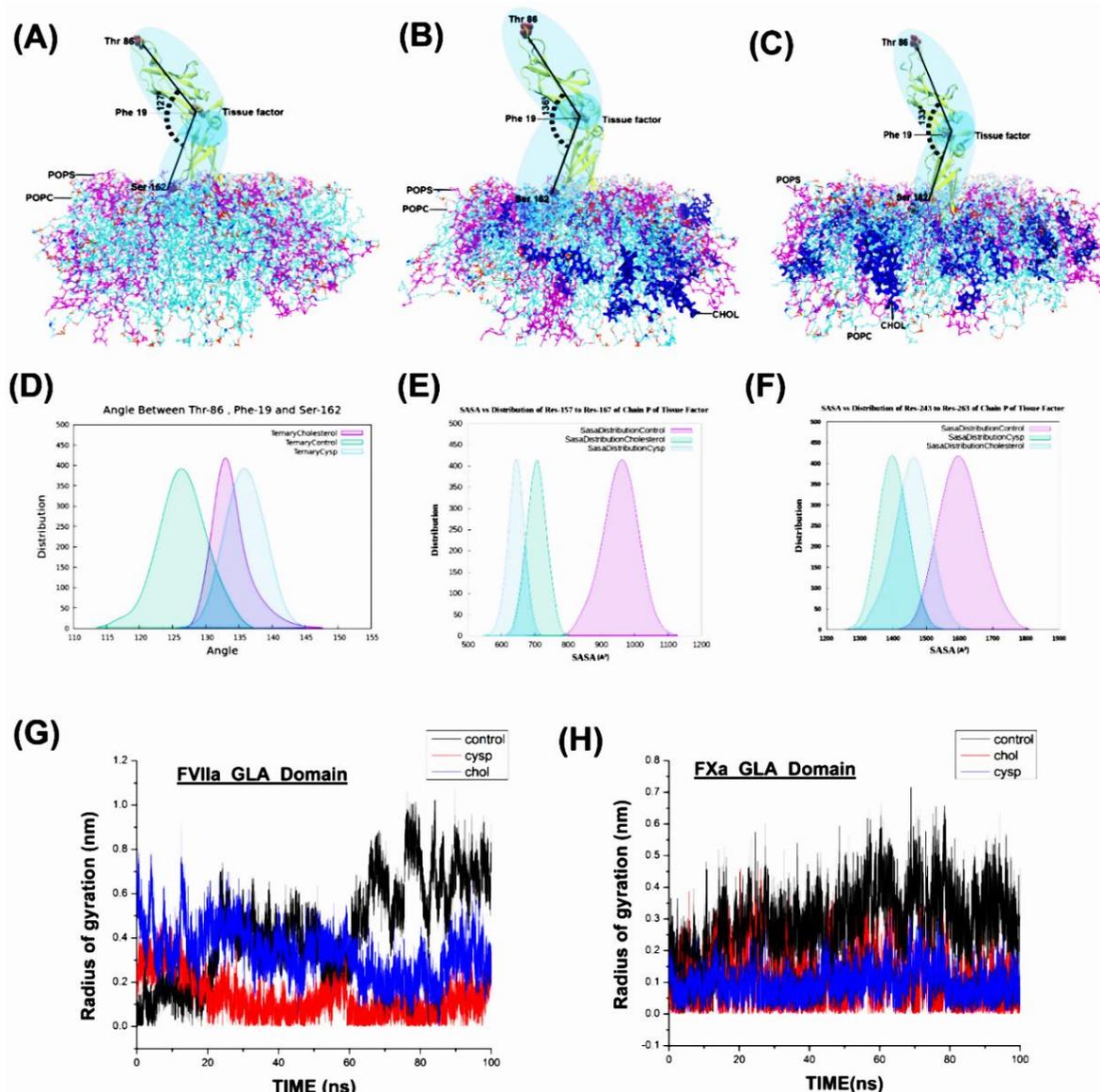


Figure 3-4. Structural analysis of TF-FVIIa-FXa Ternary Complex in POPC: POPS (CONTROL), TF-FVIIa-FXa Ternary Complex in POPC:POPS: CHL (CHOL), and Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC:POPS: CHL (CYSP). (A) Schematic representation of the Angle formed among α -atom of residues Thr86, Phe19, and Ser162 of TF in CONTROL (B) Schematic representation of the Angle formed among α -atom of residues Thr86, Phe19, and Ser162 of TF in CHOL (C) Schematic representation of the Angle formed among α -atom of residues Thr86, Phe19 and Ser162 of TF in CYSP (D) Angle distribution formed among α -atom of residues Thr86, Phe19 and Ser162 of TF in all the three systems, calculated from the last 40 ns of the simulation run. (E) Distribution of solvent accessible surface area (SASA) of residues 157-167 of chain P of TF for all the systems obtained from the last 40 ns simulation of the total run. (F) Distribution of solvent accessible surface area (SASA) of residues 243-263 of chain P of TF for all the systems obtained from the last 40 ns simulation of the total run. (G) The radius of gyration (Rgyr) of FVIIaGla-domain (residues 1-38) for all the systems obtained from the last 100 ns simulation of the total run. (H) The radius of gyration (Rgyr) of FXaGla-domain (residues 1-46) for all the systems obtained from the last 100 ns simulation of the total run.

It has been recently shown, that active conformation or the catalytic activity of the SP domain of FVIIa can be envisaged by measuring distance among catalytic triad (CT) residues, in which positioning of CT in space should remain confined and localized to the specific region [67,68]. In this context, we have computed the CT distance distribution between residues C β /Ser195-C β /His57 and simultaneously C β /Asp102-C β /His57 (Figure 3-5 A-D). We found that distances among CT residues remain confined within ~5.5-6.8 Å for residues C β /Ser195-C β /His57 and ~6.5-10.8 Å for residues C β /Asp102-C β /His57 in the case of TF-FVIIa-FXa Ternary Complex(CHOL). For Palmitoylated TF-FVIIa-FXa Ternary Complex (CYSP) distances among CT residues remain confined within ~6-6.7 Å for residues C β /Ser195-C β /His57 and ~7.5-10.4 Å for residues C β /Asp102-C β /His57 respectively. However, for TF-FVIIa-FXa Ternary Complex (CONTROL), the distances among CT residues remain distinctly scattered throughout the region specifically from ~5.5-7.5 Å and ~5-11.8 Å for residues C β /Ser195-C β /His57 and C β /Asp102-C β /His57, respectively. Thus, the presence of cholesterol and palmitoylation may play a critical role in the alteration of FVIIa activity by maintaining optimal distances among catalytic triad (CT) residues and providing extra stability to the SP domain.

Experimental FRET studies suggest that for exhibiting optimal enzymatic activity, proper positioning of substrate and enzyme is the most important criterion. Proteolytic cleavage by TF-FVIIa binary complex is an important phenomenon as it leads to the generation of activated FX (FXa), a key regulator in blood coagulation. To attain an optimal level of FX activation, FX needs to be properly aligned with the binary complex, which can only be achieved if the height of both binary complex and FX is compatible [69-70]. Therefore, we have computed the relative height plot between the O-atom of Arg139 (light chain of FXa) and the nearest P-atom of phospholipid (POPC or POPS) obtained from the simulation of all the ternary complex systems for total 200ns. In close agreement with previous reports we obtained the average heights for FVIIa-FXa Ternary Complex (CONTROL), TF-FVIIa-FXa Ternary Complex(CHOL) and Palmitoylated TF-FVIIa-FXa Ternary Complex (CYSP) are 81.79 ± 2.57 Å, 85.36 ± 2.17 Å and 88.508 ± 3.061442 respectively (Figure 3-5E). This indicates that the presence of cholesterol and palmitoylation has contributed to the positioning of active residue (Arg139) in a more upright, stable, and optimal conformation allowing for maximal substrate recognition and interaction.

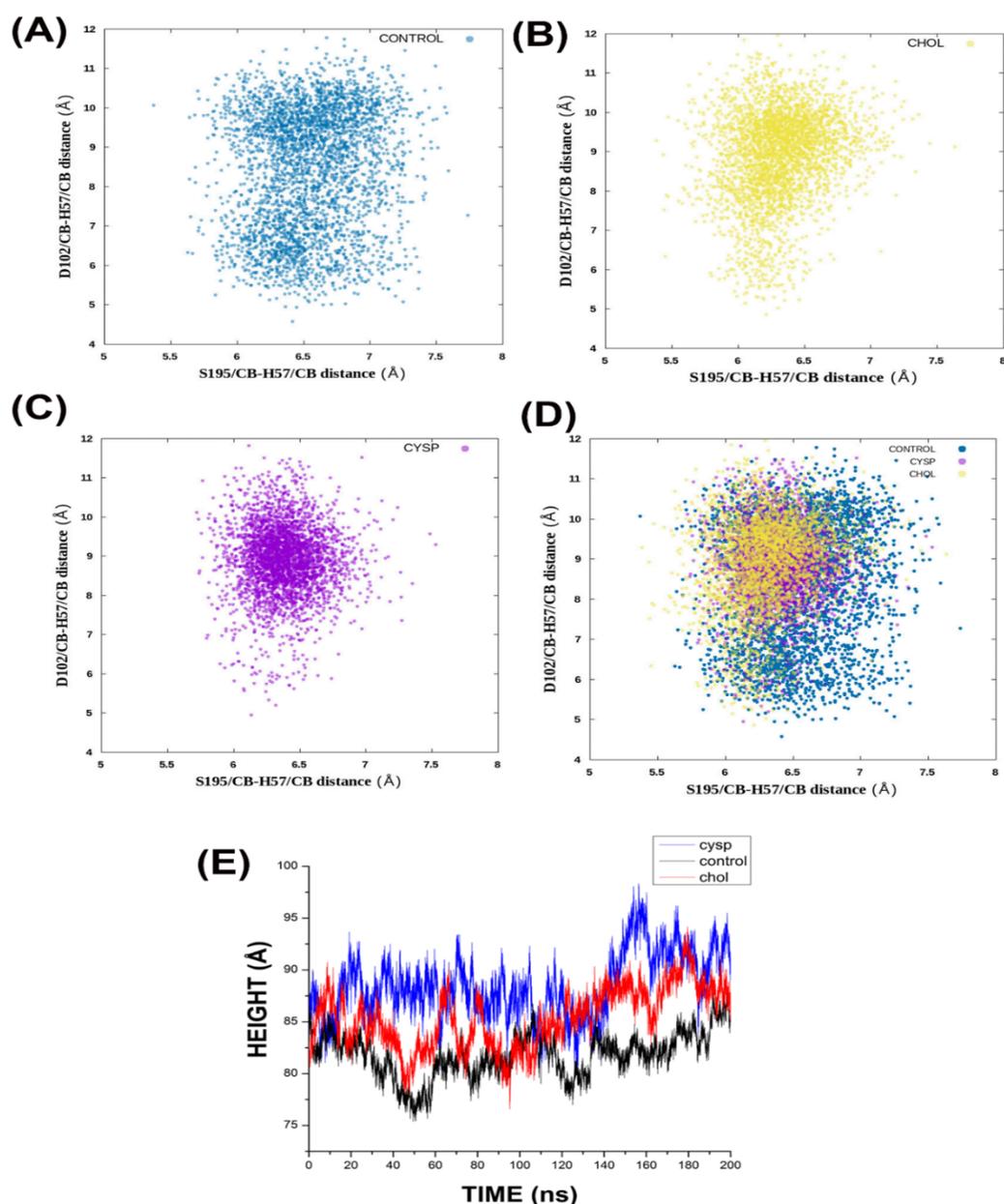


Figure 3-5: Insight into Catalytic triad (CT) dynamics in TF-FVIIa-FXa Ternary Complex in POPC: POPS(CONTROL), TF-FVIIa-FXa Ternary Complex in POPC:POPS: CHL (CHOL) and Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC:POPS: CHL (CYSP) Scattered plot for CT (catalytic triad) residues between C β atom of Ser195-H57 and C β atom of Asp102-H57 residues obtained from simulation of (A) TF-FVIIa-FXa Ternary Complex (CONTROL), (B)TF-FVIIa-FXa Ternary Complex (CHOL)and (C) Palmitoylated TF-FVIIa-FXa Ternary Complex(CYSP) (D) Superimposed scatter plot of all the three systems. The reference values for distance between C β atom (Asp 102) and C β atom (His 57) is 9 Å and between C β atom (Ser 195) and C β atom (His 57) is 6.5 Å respectively. The distance was measured and plotted in the interval of 20 ps for all the three systems for a total 100 ns simulation run. Each point represents the distance of CT residues among S195/C β -H57/C β and D102/C β -H57/C β . (H) Relative height plot between the O-atom of Arg139 (light chain of FXa) and the nearest P-atom of

phospholipid (POPC or POPS) was obtained from the simulation of the ternary complex for 200ns. The average height between residue Arg139 and the nearest P-atom of phospholipid is found to be 81.79 ± 2.57 Å, 85.36 ± 2.17 Å, and 88.508 ± 3.061442 for CONTROL, CHOL, and CYSP respectively.

3.3 Evaluation of the Interactions Formed during MD Simulations.

The number of H-bond formations in the complex system also explains the overall stability of the protein. Thus, to predict the H-bond propensity throughout the simulation period, H-bond analysis was conducted, evaluated, and compared for all three systems: CONTROL, CHOL, and CYSP. Intermolecular interactions (both H-bond and hydrophobic) were analyzed, which are formed with the side chain–side chain or side chain–main chain of the respective domains within a cutoff distance of 3.5 Å. To ascertain the binding strength, we evaluated the interaction occupancy (%) between various key domains for all systems (Figure 3-6). During simulation of the interaction between Tissue Factor and EGF1 Domain of FVIIa, we found a significant difference in the occupancy (%) between the most predominant H-bond forming pairs: GLY78-ASP58, LYS20-GLY78, GLN110-GLN64, LYS20-CYS70, ARG79-GLU24, LYS48-GLU77, ARG79-GLU56, GLN64-GLU130, ARG135-CYS72, LYS62-GLU130, LYS62-GLU128, and LEU65-GLN110 (Figure 3-6A). Many of these interactions are reported in the crystal structure of the sTF-FVIIa complex (PDB ID: 1DAN). Interestingly, for each interaction, we found that the intermolecular H-bond forming tendency is much higher for the CHOL and CYSP systems as compared to CONTROL. While analyzing H-bond interactions between Tissue Factor and EGF2 Domain of FVIIa, we found a noteworthy difference in the occupancy (%) between H-bond forming pairs: LYS46-GLU94, SER47-GLU94, SER47-ASN93, and ASN93-LYS48 (Figure 3-6B). It is evident from the plot that a newly interacting H-bond is visible between residues ASN93 (FVIIa) and LYS48 (TF) for CHOL and CYSP respectively (Figure 3-6B). However, it is worthy to mention that this interaction is absent in CONTROL throughout the simulation period. To evaluate further, we checked the interacting residues between the EGF2 domain of FVIIa and the SP domain of FVIIa, and we noted H-bond forming pairs: ARG113-GLU128, TYR215-CYS98, ARG216-GLU99, ARG134-ASP104, ARG113-GLU128, ARG216-GLU94 to be consistent throughout the simulation period (Figure 3-6C). While investigating H-bond interactions between FVIIa and FXa, we observed that H-bond forming pair LYS20-GLU19 is present only in CHOL and CYSP systems. Notable H-bond

interactions include GLN135- ARG153, ARG178- ASP93, ARG178- ASP95, ARG178- ASP97, LYS10- GLA35, ARG153- GLN135 (Figure 3-6D). We also checked the occupancy (%) of the GLA domain of FVIIa and Lipid, and we noted significant H-bond forming pairs: ARG9-POPC43, ARG36-POPS21, ASN2-POPC64, POPC27-GLA14, POPC64-GLA7, ASN2-POPC64. It is important to mention here that the LEU5-CHL84 interaction pair is absent in CONTROL but present in CHOL and CYSP (Figure 3-6E). We further obtained H-bond forming pairs between the GLA domain of FVIIa and TF: ARG36- SER163, GLN110- SER43, SER162- ASP33, LYS18- GLU216, LYS18- GLU208, LYS165- GLA35, GLU216- GLA14, VAL207- SER43, SER161- ASP33 (Figure 3-6F). While computing the occupancy plot of the GLA domain of FXa and Lipid we found prominent H-bond forming pairs: LYS9-POPS88, POPC33- GLA32, LYS10- POPS21, LEU5- POPS124 in all three systems. It is interesting to note that H-bond forming pair ARG28-POPS21 is present in CHOL and CYSP systems only (Figure 3-6G). We also computed occupancy plots between TF-chain P and FVIIa-chain L for all the three systems and we reported H-bond forming pairs: GLY78-ASP58, ARG79-GLU24, LYS85-ASP61, ARG79-GLU56, LYS20-CYS70, GLN110-GLN64 (Figure 3-6H). Occupancy plot of GLA domain of FXa and TF reveals novel interaction i.e between ASN199 and CYS50 which is only present in CHOL and CYSP. Notable bond-forming pairs present in all the three systems include LYS201- ASP35, LYS166-GLA32, ARG200-ASP48, and ARG200-GLU39. (Figure 3-6I). Similarly, while analyzing the Occupancy plot of TF and Lipid, we observed new interacting residue pairs.i.e between TYR222 and CHL6 and between GLY251 and POPC201 present only in CHOL and CYSP (Figure 3-6J). Important H-bond forming pairs present throughout the simulation period: CYS245-POPC224, LYS247-POPS247, ARG246-POPS230, LYS244-POPC263. Overall, the data from the occupancy plot suggests that the increased number of interactions between the various domains make CHOL and CYSP systems more structurally stable as compared to CONTROL. Comparative analysis of interacting residues for H-Bond formation with a cut-off distance of 3.5 Å has been performed for all the three systems and is provided in detail in Table 3-3.

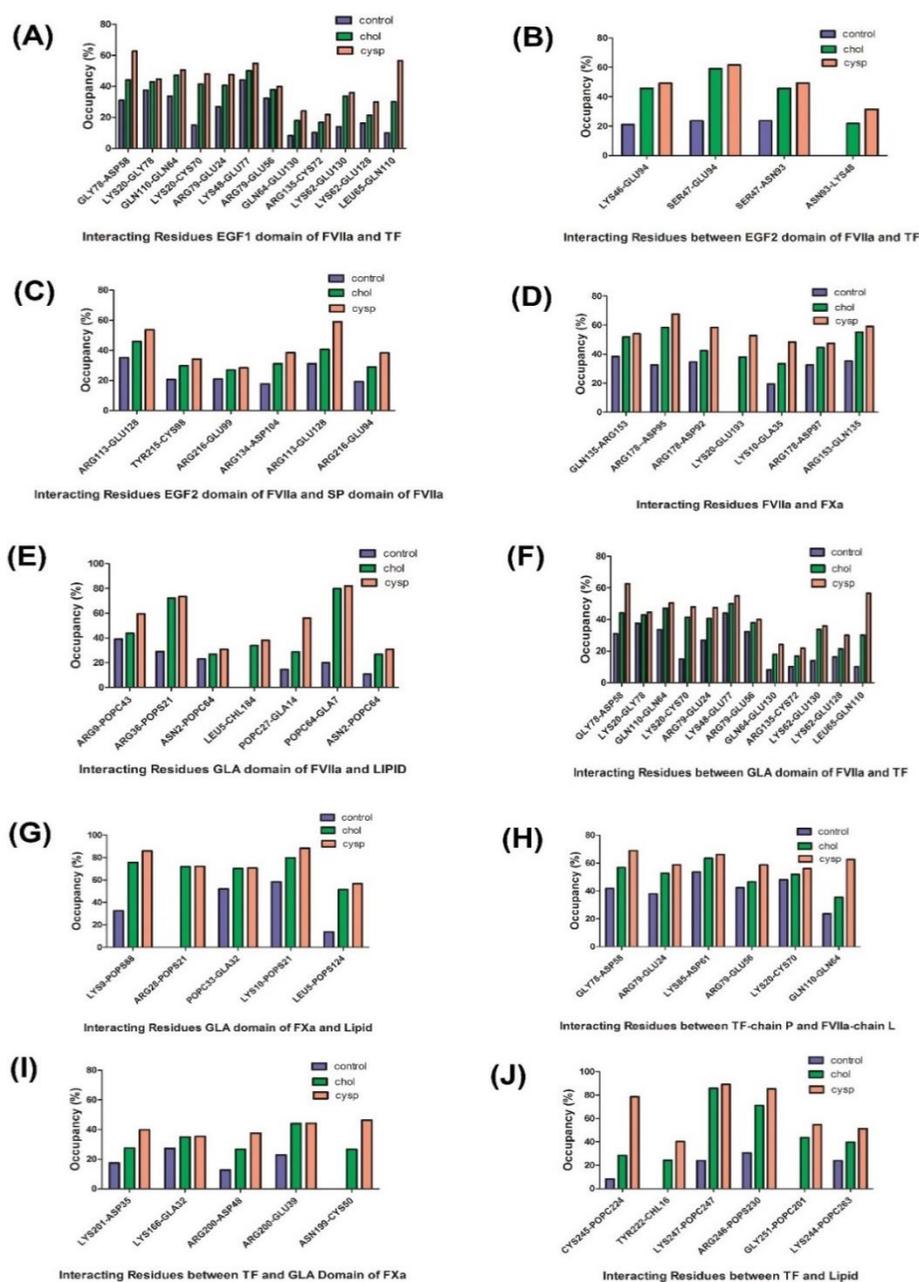


Figure 3-6: Detailed interaction profile (H-bond forming pairs) between various domains of TF-FVIIa-FXa Ternary Complex in POPC:POPS(CONTROL), TF-FVIIa-FXa Ternary Complex in POPC:POPS:CHL (CHOL) and Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC:POPS:CHL (CYSP) calculated from the 200ns simulation period of all the three complexes. (A) H-bond occupancy (%) plot between the residues of EGF1 Domain of FVIIa and TF (B) H-bond occupancy (%) plot between the residues of EGF2 Domain of FVIIa and TF (C) H-bond occupancy (%) plot between the residues of EGF2 Domain of FVIIa and SP Domain of FVIIa (D) H-bond occupancy (%) plot between the residues of FVIIa and FXa (E) H-bond occupancy (%) plot between the residues of FVIIa GLA domain and Lipid (F) H-bond occupancy (%) plot between the residues of GLA domain of FVIIa and TF (G) H-bond occupancy (%) plot between the residues of GLA domain of FXa and Lipid (H) H-bond occupancy (%) plot between the

residues of TF-chain P and FVIIa-chain L (I) H-bond occupancy (%) plot between the residues of TF and GLA domain of FXa(J) H-bond occupancy (%) plot between the residues of TF and Lipid.

Table 3-3: Comparative analysis of interacting residues for H-Bond formation with a cut-off distance of 3.5 Å for all the three systems: TF in TF-FVIIa-FXa Ternary Complex (CONTROL), TF-FVIIa-FXa Ternary Complex(CHOL) and PalmitoylatedTF-FVIIa-FXa Ternary Complex (CYSP) obtained from MD simulation study.

Domain/Residues	Systems		
	TF-FVIIa-FXa Ternary Complex in POPC:POPS (CONTROL)	TF-FVIIaFXa Ternary Complex in POPC:POPS: CHL (CHOL)	PalmitoylatedTF-FVIIa-FXa Ternary Complex in POPC:POPS: CHL (CYSP)
EGF1-TF	ARG79-GLU56, GLY78-ASP58, GLN110-GLN64, LYS20-CYS70, LYS48-GLU77, LYS65-GLN110, GLN64-GLU130, ARG135-CYS72, LYS62-GLU130, LYS62-GLU128, ARG79-GLU24	ARG79-GLU56, GLY78-ASP58, GLN110-GLN64, LYS20-CYS70, LYS48-GLU77, LYS65-GLN110, GLN64-GLU130, ARG135-CYS72, LYS62-GLU130, LYS62-GLU128, ARG79-GLU24, THR17-GLN64, PHE140-CYS72, PHE71-THR132, PHE71-ARG131	ARG79-GLU56, GLY78-ASP58, GLN110-GLN64, LYS20-CYS70, LYS48-GLU77, LYS65-GLN110, GLN64-GLU130, ARG135-CYS72, LYS62-GLU130, LYS62-GLU128, ARG79-GLU24, THR17-GLN64, PHE140-CYS72, PHE71-THR132, PHE71-ARG131, ARG131-SER60, ALA75-ASP61
EGF2-TF	LYS46- GLU94, SER47-GLU94, SER47-ASN93,	LYS46- GLU94, SER47-GLU94, SER47-ASN93, ASN93- LYS48, THR106-SER85, GLY107-TYR51	LYS46- GLU94, SER47-GLU94, SER47-ASN93, ASN93- LYS48, THR106-SER85, GLY107-TYR51
EGF2-SP	TYR215- CYS98, ARG255- GLU116, ARG113- GLU128, THR218- TYR118, ARG216- GLU94, ARG255- GLU116, ARG216- GLU99, ARG134- ASP104, ASN95- THR135, TYR215- GLU94, ARG113- GLU128, ARG255- GLU116, TYR215-CYS98	ASN95-THR135, ASN95-TYR215, THR135-ASN95, PRO127-TYR101, ARG216-GLU94, THR218-TYR118, ARG113-GLU128, ARG134-ASP104, ARG255-GLU116, TYR215-CYS98, ARG113-GLU128, ARG216-GLU99, ASN47-GLU116,	ASN95-THR135, ASN95-TYR215, THR135-ASN95, PRO127-TYR101, ARG216-GLU94, THR218-TYR118, ARG113-GLU128, ARG134-ASP104, ARG255-GLU116, TYR215-CYS98, ARG113-GLU128, ARG216-GLU99, ASN47-GLU116,

Domain/Residues	Systems		
	TF-FVIIa-FXa Ternary Complex in POPC:POPS (CONTROL)	TF-FVIIaFXa Ternary Complex in POPC:POPS: CHL (CHOL)	PalmitoylatedTF-FVIIa-FXa Ternary Complex in POPC:POPS: CHL (CYSP)
		THR130-ASP104, GLU128-TYR101, TYR215-CYS98	THR130-ASP104, GLU128-TYR101, GLN100-THR218, ASN95-PHE131, TYR215-CYS98
FVII-FX	ALA155-GLU26, LYS10-ASP33, LYS10-GLA35, ARG153-GLU138, ARG153-GLU138, TYR163-LEU40, LYS23-GLU159, ARG153-GLU138, ARG178-ASP95, ARG28-GLA14, ARG153-GLN135, LYS20-GLU193, ASP189-SER77, LYS23-GLU159, LYS20-GLU162, LYS43-GLA35, ARG87-ARG194, LYS32-GLA14, LYS62-THR148	ARG207- ALA232, ARG207-SER199, GLN135-ARG153, GLY18-GLU162, ALA155-GLU26, LYS191-GLU159, LYS138-SER199, ARG207-ALA232. ARG207-GLY198, GLN20-ALA155, GLN20-THR156, GLN137-GLY198, ARG178-ASP95, ARG178-ASP93, SER199-GLN137, HSE12-GLA29, ARG153-GLN135, ARG28-GLA14, LYS10-ASP33, LYS10-GLA35, GLN146-ASN36, ARG178-ASP97, HSE12-GLA29, ARG28-GLA14, LYS10-GLA35, LYS179-ASP95	ARG207- ALA232, ARG207-SER199, GLN135-ARG153, GLY18-GLU162, ALA155-GLU26, LYS191-GLU159, LYS138-SER199, ARG207-ALA232. ARG207-GLY198, GLN20-ALA155, GLN20-THR156, GLN137-GLY198, ARG178-ASP95, ARG178-ASP93, SER199-GLN137, HSE12-GLA29, ARG153-GLN135, ARG28-GLA14, LYS10-ASP33, LYS10-GLA35, GLN146-ASN36, ARG178-ASP97, HSE12-GLA29, ARG28-GLA14, LYS10-GLA35, LYS179-ASP95
FVIIaGLA domain - Lipid	ARG9-POPC43, POPS17-ALA34, POPS17-GLA35, POPS17-LYS32, LYS32-POPC16, LYS32-POPC35, POPS38-GLA29, PHE4-POPC67, LYS38-POPS17, ARG36-POPS17, ASN2-POPC64, LEU5-POPC73, THR37-POPS17, POPC27-GLA25, POPC27-ALA27,	ARG9-POPC43, POPS17-ALA34, POPS17-GLA35, POPS17-LYS32, LYS32-POPC16, LYS32-POPC35, POPS38-GLA29, PHE4-POPC67, LYS38-POPS17, ARG36-POPS21, ASN2-POPC64, LEU5-POPC73, THR37-POPS17, POPC27-GLA25, POPC27-ALA27,	ARG9-POPC43, POPS17-ALA34, POPS17-GLA35, POPS17-LYS32, LYS32-POPC16, LYS32-POPC35, POPS38-GLA29, PHE4-POPC67, LYS38-POPS17, ARG36-POPS21, ASN2-POPC64, LEU5-POPC73, THR37-POPS17, POPC27-GLA25, POPC27-ALA27,

Domain/Residues	Systems		
	TF-FVIIa-FXa Ternary Complex in POPC:POPS (CONTROL)	TF-FVIIaFXa Ternary Complex in POPC:POPS: CHL (CHOL)	PalmitoylatedTF-FVIIa-FXa Ternary Complex in POPC:POPS: CHL (CYSP)
	ARG36-POPS21, GLA7-POPC73, THR37-POPS17, POPS38-ASP33, PHE4-POPC67, LYS32-POPS17, ASN2-POPC64, ILE30-POPC27, ALA34-POPS17, THR37-POPS17, LYS38-POPS17, POPC64-GLA7	ARG36-POPS21, GLA7-POPC73, THR37-POPS17, POPS38-ASP33, PHE4-POPC67, LYS32-POPS17, ASN2-POPC64, ILE30-POPC27, ALA34-POPS17, THR37-POPS17, LYS38-POPS17, POPC64-GLA7, ARG9-CHL1223, ARG9-CHL153, GLY11-CHL1240, ARG9-CHL1391, POPC51-ASP33, POPC27-GLA14	ARG36-POPS21, GLA7-POPC73, THR37-POPS17, POPS38-ASP33, PHE4-POPC67, LYS32-POPS17, ASN2-POPC64, ILE30-POPC27, ALA34-POPS17, THR37-POPS17, LYS38-POPS17, POPC64-GLA7, ARG9-CHL1223, ARG9-CHL153, GLY11-CHL1240, ARG9-CHL1391, POPC51-ASP33, ARG9-CHL12, LYS32-CHL167, ARG9-CHL1249, CHL12-GLA29, POPC27-GLA14
FVIIaGLAdomain-TF	TRP158- THR37, LYS165- GLA35, ARG36- SER163, ARG36- SER160, TRP158- THR37, VAL207- SER43, SER161- ASP33, ARG36-SER163, GLN110-SER43, LYS18-GLU216, LYS18-GLU208, GLU216-GLA14	LYS18-GLU208, GLN110-SER43, LYS214-GLA14, TRP158-PHE31, LYS165-GLA35, LYS32-SER160, LYS18-GLU216, LYS165-GLA35, LYS214-GLA14, LYS165-GLA35, ARG36-SER163, LEU39-SER205, VAL207-SER43, SER161-ILE30, SER161-ASP33, GLU216-GLA14	LYS18-GLU208, GLN110-SER43, LYS214-GLA14, TRP158-PHE31, LYS165-GLA35, LYS32-SER160, LYS18-GLU216, LYS165-GLA35, LYS214-GLA14, LYS165-GLA35, ARG36-SER163, LEU39-SER205, VAL207-SER43, SER161-ILE30, SER162-ASP33, GLU216-GLA14, SER161-ASP33
FXaGLA-lipid	SER3- POPC113, SER3-POPC87, POPC87-GLA20, POPC87-PHE4, LYS9-POPC88, POPC87-GLA20, POPC87-PHE4,	ARG28-POPC21, GLA32-POPC33, SER3- POPC113, SER3-POPC87, POPC87-GLA20, POPC87-PHE4, LYS9-POPC88,	LYS9-POPS88, ARG28-POPS21, POPC33-GLA32, LYS10-POPS21, LYS9-POPS88, POPC33-GLA32, LYS10-POPS21,

Domain/Residues	Systems		
	TF-FVIIa-FXa Ternary Complex in POPC:POPS (CONTROL)	TF-FVIIaFXa Ternary Complex in POPC:POPS: CHL (CHOL)	PalmitoylatedTF-FVIIa-FXa Ternary Complex in POPC:POPS: CHL (CYSP)
	POPC33-GLA32, LEU5-POPS124, SER3-POPC87, ALA1-POPC87, LYS10-POPC35, SER3-POPC83, POPS47-GLA25, LYS10-POPS21, GLA6-POPC87	POPC87-GLA20, POPC87-PHE4, POPC33-GLA32, LEU5-POPS124, SER3-POPC87, ALA1-POPC87, LYS10-POPC35, SER3-POPC83, POPS47-GLA25, LYS10-POPS21, GLA6-POPC87, ARG28-POPC33, LEU5-POPS124, HSE12-POPS21, LYS10-CHL142, LYS9-CHL186, POPC51-GLA14, LYS10-CHL141, LYS10-CHL119, LYS10-CHL142, LYS9-CHL186, POPC51-GLA14, LYS10-CHL141, LYS10-CHL119	LYS9- POPC88, POPC33-GLA32, POPS124-LEU5, SER3-POPC83, POPS47-GLA25, LYS10-POPS21, GLA6-POPC87, ARG28-POPC33, LEU5-POPS124, HSE12-POPS21, LYS10-CHL142, LYS9-CHL186, POPC51-GLA14, LYS10-CHL141, LYS10-CHL119, LYS10-CHL119, SER3-POPS140, PHE4-POPS140, LYS10-POPC51, GLY11-POPS21 POPS21-VAL30, HSE12-POPC51,
TF-FVIIaLightChain	GLY78-ASP58, ARG79-GLU24, ARG79-GLU56, LYS85-ASP61, SER47-GLU94, SER47-ASN93, LYS165-GLA35, LYS62-GLN110, LYS62-GLU128, LYS20-CYS70, LYS46-GLU94, ARG36-SER163, TRP158-THR37, GLN110-GLN64, LYS48-GLU77, ARG79-GLU56, ARG36-SER160, ARG36-SER160, TRP158-THR37	LYS32-SER160, LYS214-GLA14, ARG79-GLU24, ARG79-GLU56, LYS18-GLU208, LYS85-ASP61, GLY78-ASP58, ARG79-GLU56, SER47-GLU94, LYS165-GLA35, LYS62-GLU128, LYS62-GLU130, GLU94-SER47, LYS20-CYS70, LYS20-GLY78, GLN110-GLN64, TRP158-THR37, LYS48-GLU77, LYS18-GLU208, LYS165-GLA35, ARG36-SER163, LYS48-GLU77, LYS18-GLU216, GLY107-TYR51, LYS214-GLA14,	LYS18-GLU208, LYS18-GLU216, GLY78-ASP58, LYS85-ASP61, ARG79-GLU56, ARG79-GLU24, LYS62-GLU128, GLU216-GLA14, LYS20-GLY78, ARG36-SER162, ARG36-SER163, GLN110-GLN64, LYS48-GLU77, LEU65-GLN110, LYS18-GLU216, LYS32-ASN184, LYS85-ASP61, SER161-ASP33, LYS62-ASP129, VAL207-SER43, LYS46-GLU94, LEU13-GLN212, ASN93-LYS48, ARG79-GLU56, LYS85-ASP61,

Domain/Residues	Systems		
	TF-FVIIa-FXa Ternary Complex in POPC:POPS (CONTROL)	TF-FVIIaFXa Ternary Complex in POPC:POPS: CHL (CHOL)	PalmitoylatedTF-FVIIa-FXa Ternary Complex in POPC:POPS: CHL (CYSP)
		LYS85-ASP61, SER161-ILE30, LYS48-ASP86	SER161-ASP33, ARG79-ASP54, LYS32-GLU183, LYS32-GLU183, SER47-VAL92, LYS62-GLU128, LYS48-ASP86
TF-FXaGLAdomain	LYS201- ASP35, LYS165-ASP33, LYS201-GLU39, LYS165-GLA32, LYS36-ASP204, ARG200-ASP35, ARG200-ASP48, LYS201-ASP35, LYS201-GLU39, LYS166-GLA32, ARG200-GLU39, LYS36-ASP204, ARG200- ASP35	LYS201- ASP35, LYS165-GLA32, LYS165-VAL30, TYR156-GLA32, ARG200-ASP35, LYS165-GLA32, ARG200-ASP35, ARG200-GLU39, LYS166-GLA32, ARG200-GLU39, LYS166-GLA32, ARG200-ASP48, ASN199-CYS50	LYS166-GLA32, LYS201- ASP35, LYS165-GLA32, LYS165-VAL30, TYR156-GLA32, ARG200-ASP35, LYS165-GLA32, ARG200-ASP35, ARG200-GLU39, LYS166-GLA32, ARG200-GLU39, LYS166-GLA32, ARG200-GLU39, LYS166-GLA32, ARG200-ASP48, ASN199-CYS50
TF-lipid	CYS245-POPC224, TYR222-POPC29, ARG246-POPS230, ARG246-POPC250, LYS247-POPC220, LYS244-POPC250, LYS244-POPS230, LYS247-POPC209, LYS244-POPC263, HSD243-POPC258, LYS255-POPC205	ARG246-POPC263, LYS247-POPC263, LYS244-POPC247, LYS255-POPC263, LYS244-POPC312, LYS255-POPC264, SER241-POPC247, CYS245-POPC224, ARG246-POPC263, LYS244-POPS298, ASN261-CHL1327, ALA248-CHL1283, TYR222-CHL16, SER263-POPC343, ARG246-POPC230, ARG246-POPS230, ARG246-POPC247, LYS244-LYS244, GLY251-POPC201	LYS244-POPC226, LYS247-POPS211, LYS244-POPC226, TYR222-CHL16, CHL117-PHE221, ARG246-CHL1236, CYS245-POPC224, SER253-POPC252, HSD243-POPC247, LYS255-POPC252, TRP254-POPC252, POPC342-SER263, LYS247-POPS247, ARG246-POPS211, LYS244-POPS211, GLY251-POPC201, GLY251-POPC201, LYS247-POPS211, ARG246-POPS230, LYS247-POPS211, LYS244-POPC226, LYS244-POPC263

3.4 Contact Probability estimation between Protein and lipid

To decipher protein-lipid interactions, the computation of contact probability can be very helpful [59]. Contact probability plot can provide us with information about the residues which are in contact with lipid and may give us clues about the importance of those residues responsible for lipid-protein interaction [Table 3- (4.1-4.3)]. Therefore, we have generated a contact probability plot using inhouse VMD TCL script for all the three systems: TF-FVIIa-FXa Ternary Complex in POPC: POPS (CONTROL), TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CHOL), and Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CYSP) for 200ns. Contact Probability plots were constructed for residues of Chain B of FXa in contact with lipid (Figure 3-7 A-C), residues of Chain L of FVIIa in contact with lipid (Figure 3-7 D-F), residues of Chain P of TF in contact with lipid (Figure 3-7 G-I). From Contact Probability data between Chain B of FXa with lipid, it is observed that there are notable distinct differences in the contact fraction for residue numbers 6(GLA), 15(ARG), 19(GLA), 20(GLA), 28(ARG), 32(GLA) with Lipid. (Table 3-4.1) (Figure 3-8 A-C) It is noteworthy that Arg in position 15 and GLA in position 6 have 100% contact probability with lipid for both CHOL AND CYSP sets but this phenomenon is absent in the CONTROL set. Also, significant differences are observed in residue numbers 20(GLA), 28(ARG), and 32(GLA) of chain B of FXa, having contact fractions of 1.07%, 1.85%, and 8.06% respectively in the CONTROL system. This indicates the presence of strong and stable lipid-protein interactions in the CHOL and CYSP systems than in CONTROL. Similarly, from Contact Probability data between Chain L of FVIIa with lipid, it is observed that there are distinct differences in the contact fraction for residue numbers 1(ALA), 18(LYS), 26(GLA), and 28(ARG) with lipid. (Table 3-4.2) (Figure 3-8 D-F). Notably, GLA in position 26 has 100% contact probability with lipid for both CHOL AND CYSP sets but only 3.24% in the CONTROL set. A similar trend is observed for residue 8 which has 98.15% contact probability with lipid for both CHOL and 68.42% in the CYSP set but only 3.56% in CONTROL. Likewise, significant differences are also observed in residue numbers 1(ALA), and 28(ARG) of chain B of FXa. Likewise, from Contact Probability data between Chain P of TF with lipid, it is observed that there are distinct differences in the contact fraction for residue numbers 117(GLU), 122(LYS), 159(LYS), 179(VAL), 183(GLU), 212(GLN), 219(GLU), 245(CYS), 255(LYS) with lipid. (Table 3-4.3) (Figure 3-8 G-I) It is important to note that LYS in position 122, CYS in position 245, and LYS in position 159 have 100% contact probability with lipid for CYSP set as compared to the

CONTROL set which has only 48.82%, 62.12%, and 68.10% for the residues 122(LYS), 245(CYS), 159(LYS) respectively. CHOL set has a significant contact fraction of 100%, 82.12%, and 94.54% for the residues 122(LYS), 245(CYS), and 159(LYS) respectively. It is clear from the above data, that Cholesterol and Palmitoylation at CYS²⁴⁵ residue enhance the protein-lipid interactions. This is in line with the experimental studies which show that cholesterol has a strong influence on membrane receptors by unique molecular interactions or by changes of fluidity in the membrane [5].

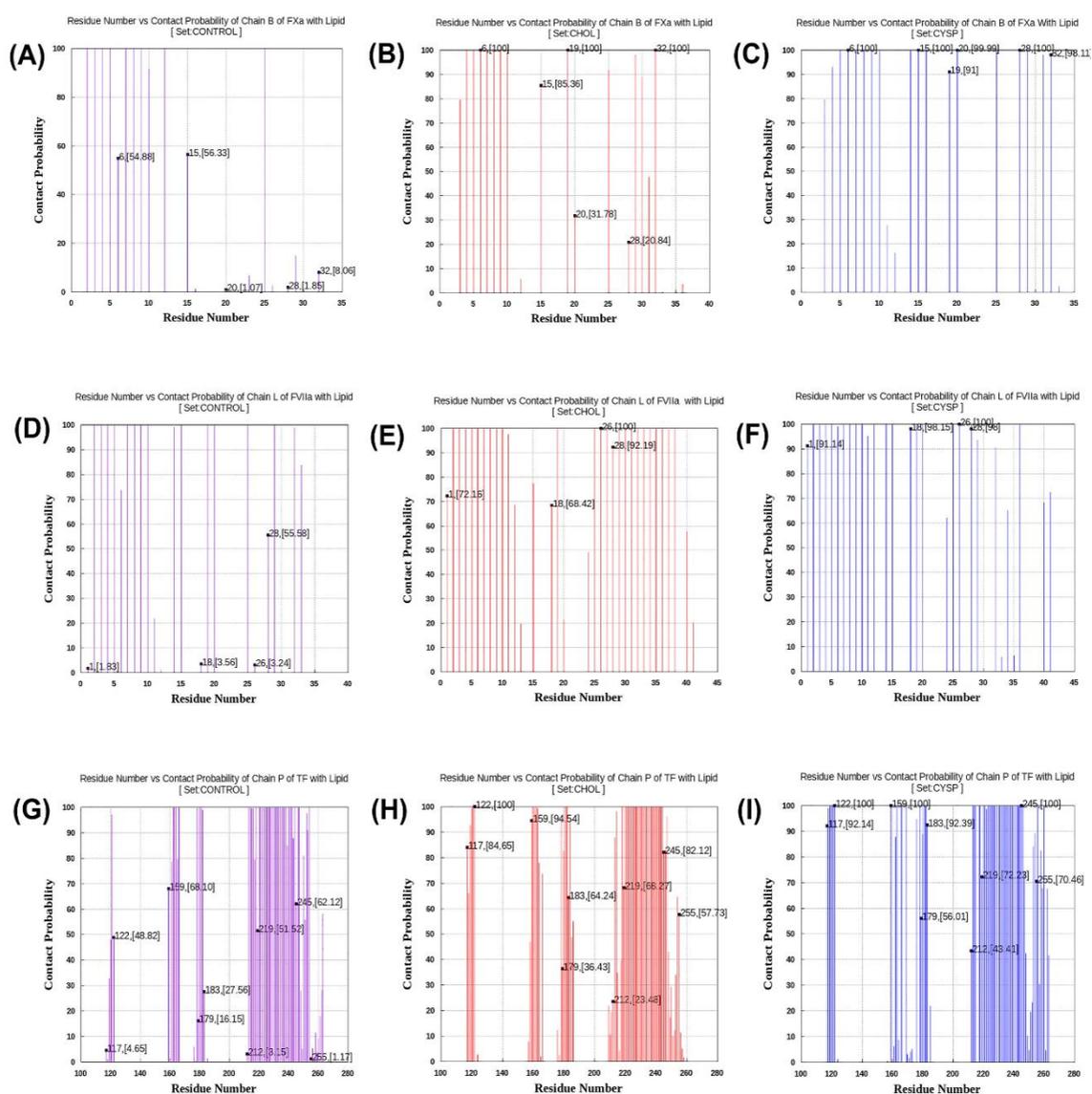


Figure 3-7: Detailed contact probability profile (lipid-protein interaction) in FVIIa-FXa Ternary Complex in POPC: POPS(CONTROL), TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CHOL), and Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CYSP) calculated from the 200ns simulation period of all the three complexes. (A) Contact probability vs residues of chain B of FXa in contact with lipid (CONTROL) (B) Contact

probability vs residues of chain B of FXa in contact with lipid (CHOL) (C) Contact probability vs residues of chain B of FXa in contact with lipid (CYSP) (D) Contact probability vs residues of chain L of FVIIa in contact with lipid (CONTROL) (E) Contact probability vs residues of chain L of FVIIa in contact with lipid (CHOL) (E) Contact probability vs residues of chain L of FVIIa in contact with lipid (CYSP) (G) Contact probability vs residues of chain P of TF in contact with lipid (CONTROL) (H) Contact probability vs residues of chain P of TF in contact with lipid (CHOL) (I) Contact probability vs residues of chain P of TF in contact with lipid (CYSP).

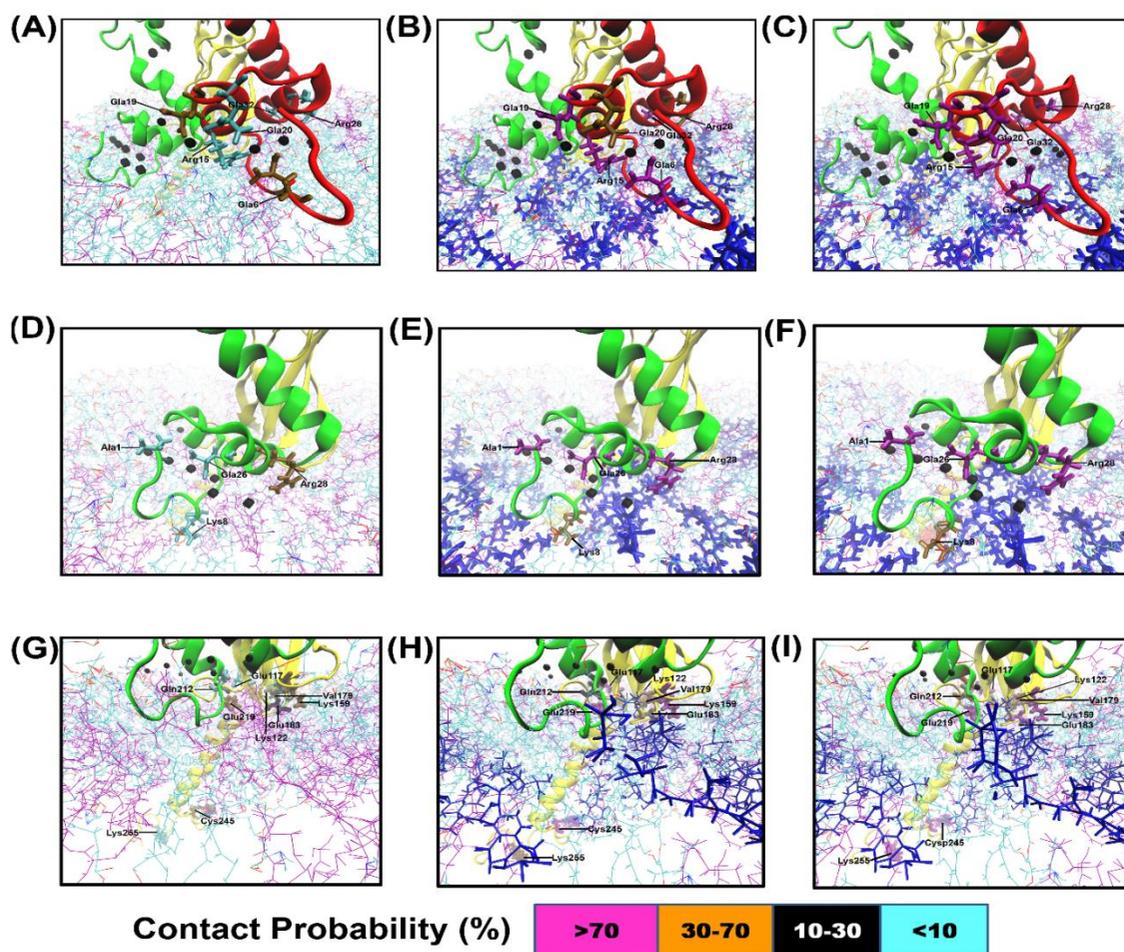


Figure 3-8: Schematic representation of the residues in contact with lipid (contact probability profile) of TF-FVIIa-FXa Ternary Complex in POPC:POPS(CONTROL), TF-FVIIa-FXa Ternary Complex in POPC:POPS:CHL (CHOL) and Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC:POPS:CHL (CYSP) (A) Closer view of the residues of chain B of FXa (CONTROL) in contact with lipid: 6(GLA), 15(ARG), 19(GLA), 20(GLA), 28(ARG), 32(GLA). (B) Closer view of the residues of chain B of FXa (CHOL) in contact with lipid:6(GLA), 15(ARG), 19(GLA), 20(GLA), 28(ARG), 32(GLA). (C) Closer view of the residues of chain B of FXa (CYSP) in contact with lipid : 6(GLA), 15(ARG), 19(GLA), 20(GLA), 28(ARG), 32(GLA) (D) Closer view of the residues of chain L of FVIIa (CONTROL) in contact with lipid: 1(ALA), 18(LYS), 26(GLA) and 28(ARG) (E) Closer view

of the residues of chain L of FVIIa (CHOL) in contact with lipid: 1(ALA), 18(LYS), 26(GLA) and 28(ARG)(F) Closer view of the residues of chain L of FVIIa (CYSP) in contact with lipid: 1(ALA), 18(LYS), 26(GLA) and 28(ARG) (G)Contact probability vs residues of chain P(CONTROL) of TF in contact with lipid 117(GLU), 122(LYS), 159(LYS), 179(VAL), 183(GLU), 212(GLN), 219(GLU), 245(CYS), 255(LYS)(H)Contact probability vs residues of chain P of TF (CHOL) in contact with lipid 117(GLU), 122(LYS), 159(LYS), 179(VAL), 183(GLU), 212(GLN), 219(GLU), 245(CYS), 255(LYS)(I)Contact probability vs residues of chain P of TF (CYSP) in contact with lipid 117(GLU), 122(LYS), 159(LYS), 179(VAL), 183(GLU), 212(GLN), 219(GLU), 245(CYS), 255(LYS). Residues are shown in licorice representation. The chain B of FXa is shown by a red color ribbon representation. The chain L of FVIIa is shown by a green color ribbon representation. The chain P of TF is shown by a yellow color ribbon representation. In the bottom panel, phospholipid bilayer is shown in lines representation wherein POPC is shown as purple color lines, POPS shown as ice blue color lines, and CHOL in blue color lines. Contact Probability (%)colored based on occupancy during the simulations: Purple for >70%, blue for Dark yellow%, black for 10-30%, and sky blue for <10%.

Table 3-4.1: Contact probability between Chain B of FXa with lipid.

Residue No	Residue Name	Contact Probability CONTROL	Contact Probability CHOL	Contact Probability CYSP
6	GLA	54.88	100	100
15	ARG	0	100	100
19	GLA	56.33	85.36	100
20	GLA	1.07	31.78	100
28	ARG	1.85	20.84	100
32	GLA	8.06	100	98.11

Table 3-4.2: Contact Probability between Chain L of FVIIa with lipid.

Residue No	Residue Name	Contact Probability CONTROL	Contact Probability CHOL	Contact Probability CYSP
1	ALA	1.83	72.16	91.14
18	LYS	3.56	68.42	98.15
26	GLA	3.24	100	100
28	ARG	55.58	92.19	98

Table 3-4.3: Contact Probability data between Chain P of TF with lipid.

Residue No	Residue Name	Contact Probability CONTROL	Contact Probability CHOL	Contact Probability CYSP
117	GLU	4.65	84.65	92.14
122	LYS	48.82	100	100
159	LYS	68.10	94.54	100
179	VAL	16.15	36.43	56.01
183	GLU	27.56	64.24	92.39
212	GLN	3.15	23.48	43.41
219	GLU	51.52	68.27	72.23
245	CYS	62.12	82.12	100
255	LYS	1.17	57.73	70.46

3.5 Protein stability predictions by FoldX

Lastly, to check whether the presence of cholesterol and palmitoylation of Tissue factor has any effect on the binding affinity between chain L of FVIIa and TF, we have calculated the FoldX energy function using the FoldX optimization algorithm between FVIIa (light chain) and TF for all the three systems: CONTROL, CHOL, and CYSP (Table 3-5). FoldX is an algorithm that is used for the prediction of free energy changes based on the entropic contributions to protein interactions, which also contributes to the stability of proteins and protein complexes [71]. FoldX force field includes several

terms: Van der Waals, solvation, H-bond, electrostatic and entropic terms for the backbone and side chains of protein and protein complexes.

The free energy profile (ΔG) of the target protein is calculated using the following equation:

$$\Delta G = W_{vdw} \times \Delta G_{vdw} + W_{solvH} \times \Delta G_{solvH} + W_{solvP} \times \Delta G_{solvP} + \Delta G_{wb} + \Delta G_{hbond} + \Delta G_{el} + \Delta G_{Kon} + W_{Nc} \times T \times \Delta S_{Nc} + W_{sc} \times T \times \Delta S_{sc}$$

Where,

ΔG_{vdw} = summation of van der Waals contributions of all atoms.

ΔG_{solvH} and ΔG_{solvP} = solvation energy difference between polar and apolar groups, respectively, from unfolded to the folded state.

ΔG_{wb} = free energy difference provided by water molecule making more than one H-bond to the protein and that interaction cannot be taken into account with non-explicit solvent approximations.

ΔG_{hbond} = free energy difference between intra- and inter-molecular H-bond formation.

ΔG_{el} = electrostatic contribution of charged groups.

ΔG_{Kon} = effect of electrostatic interactions on the association constant k_{on} , when the interaction energies are calculated between protein complexes.

ΔS_{mc} = entropy cost to fix the backbone in the folded state.

ΔS_{sc} = entropy cost to fix a side-chain of protein in a specific conformation.

Coordinates of protein atoms (target) were extracted for all the three systems (CONTROL, CHOL, and CYSP) from the equilibrated MD simulation system and subjected to free energy calculation using the FoldX optimization algorithm. Lower the energy value obtained from the FoldX energy function indicates greater binding affinity between the interacting proteins. We found that the binding affinity between chain L of FVIIa and TF is less (-1.45 kcal/mol) for CONTROL than CHOL (-3.48kcal/mol) and

CYSP (-4.44 kcal/mol). This indicates that the CONTROL system is thermodynamically less stable than CHOL and CYSP and Palmitoylation of Tissue factor and cholesterol may play a role in enhancing the binding affinity between chain L of FVIIa and TF.

Table 3-5: Interaction energy between TF and FVIIa (light chain) using the FoldX algorithm.

FoldX energy profile System	Interaction between TF and FVIIa (L-chain)		
	TF-FVIIa-FXa Ternary Complex in POPC: POPS (CONTROL)	TF-FVIIaFXa Ternary Complex in POPC:POPS: CHLc (CHOL)	PalmitoylatedTF- FVIIa-FXa Ternary Complex in POPC:POPS: CHL (CYSP)
BackHbond	-2.66	-3.39	-3.49
SideHbond	-5.89	-6.44	-5.06
Energy-vdW	-15.32	-22.14	18.14
Electro	-3.44	-4.47	-2.36
Energy-SolvP	23.61	32.83	25.35
Energy-SolvH	-19.61	-28.99	23.62
Energy-vdWclash	3.09	3.74	1.27
Energy-torsion	0.29	0.25	0.41
Backbone-vdWclash	1.32	2.32	1.92
Entropy-sidechain	14.01	17.75	16.19
Entropy-mainchain	4.85	7.87	5.32
Water bonds	0	0	0
Helix dipole	-0.11	-0.02	0.06
Loop-entropy	0	0	0
Cis-bond	0	0	0
Disulfide	0	0	0
kn electrostatic	-0.26	-0.47	-0.38
Partial covalent interactions	0	0	0
Energy-ionisation	0	0	0
Entropy complex	2.38	2.38	2.38
Total interaction energy (kcal mol⁻¹)	-1.45	-3.48	-4.44

4. Conclusions

To gain insight into the effect of cholesterol and Tissue Factor Palmitoylation on TF-FVIIa-FXa ternary complex in POPC-POPS-CHOL lipid environment, we had performed a comparative study among TF-FVIIa-FXa Ternary Complex in POPC: POPS(CONTROL), TF-FVIIa-FXa Ternary Complex in POPC: POPS: CHL (CHOL) and Palmitoylated TF-FVIIa-FXa Ternary Complex in POPC:POPS: CHL (CYSP). In

this study, we subjected CONTROL, CHOL, and CYSP systems to MD simulation in POPC: POPS and POPC:POPS: CHL environment. RMSD and RMSF profiles captured from the simulation suggest that CHOL and CYSP systems are more stable compared to CONTROL. Structural and Conformational analysis indicates lesser structural integrity of the subdomains of TF in TF-FVIIa-FXa Ternary Complex (CONTROL) compared to TF-FVIIa-FXa Ternary Complex (CHOL) and Palmitoylated TF-FVIIa-FXa Ternary Complex (CYSP). Our study also suggests that the presence of cholesterol and palmitoylation may play a critical role in influencing FVIIa activity by maintaining optimal distances among catalytic triad (CT) residues, thereby providing extra stability to the SP domain. From the SASA profile, it can be deduced that there is an enhancement in hydrophobicity of proteins, indicative of a more buried structure with augmented protein-protein interaction in TF-FVIIa-FXa Ternary Complex (CHOL) and Palmitoylated TF-FVIIa-FXa Ternary Complex (CYSP). The radius of Gyration data suggests that the presence of cholesterol and Palmitoylation of the Tissue Factor contributes to structural rigidity and compactness of the GLA domains of both FVIIa and FXa respectively. While evaluating H-bond interactions formed during MD Simulations, we have found many novel interactions between the various domains of CHOL and CYSP systems which may be responsible for making these two systems structurally more stable, compared to CONTROL. From Contact Probability data, it is evident that Cholesterol and Palmitoylation at CYS²⁴⁵ residue enhance the protein-lipid interactions thus having a profound influence on protein activity. Lastly, from the FoldX protein stability study, it can be inferred that the CONTROL system is thermodynamically less stable than CHOL and CYSP.

References:

1. Mackman N. Role of tissue factor in hemostasis, thrombosis, and vascular development. *Arteriosclerosis, thrombosis, and vascular biology*. 2004 Jun 1;24(6):1015-22.
2. Davie EW, Fujikawa K, Kisiel W. The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry*. 1991 Oct 1;30(43):10363-70.
3. Ardissino D, Merlini PA, Arlens R, Coppola R, Bramucci E, Lucreziotti S, Repetto A, Fetiveau R, Mannucci PM. Tissue factor in human coronary atherosclerotic plaques. *Clinica chimica acta*. 2000 Feb 15;291(2):235-40.
4. Mandal SK, Iakhiaev A, Pendurthi UR, Rao LV. Acute cholesterol depletion impairs functional expression of tissue factor in fibroblasts: modulation of tissue factor activity by membrane cholesterol. *Blood*. 2005 Jan 1;105(1):153-60.
5. Gimpl G, Burger K, Fahrenholz F. Cholesterol as modulator of receptor function. *Biochemistry*. 1997 Sep 9;36(36):10959-74.
6. Parpal S, Karlsson M, Thorn H, Strålfors P. Cholesterol depletion disrupts caveolae and insulin receptor signaling for metabolic control via insulin receptor substrate-1, but not for mitogen-activated protein kinase control. *Journal of Biological Chemistry*. 2001 Mar 30;276(13):9670-8.
7. Mulder AB, Smit JW, Bom VJ, Blom NR, Ruiters MH, Halie MR, van der Meer J. Association of smooth muscle cell tissue factor with caveolae. *Blood*. 1996;88:1306-1313.
8. Sevinsky JR, Rao LV, Ruf W. Ligand-induced protease receptor translocation into caveolae: a mechanism for regulating cell surface proteolysis of the tissue factor-dependent coagulation pathway. *The Journal of cell biology*. 1996 Apr 15;133(2):293-304.
9. Schuff-Werner P, Claus G, Armstrong VW, Kostering H, Seidel D. Enhanced procoagulatory activity (PCA) of human monocytes/macrophages after in vitro stimulation with chemically modified LDL. *Atherosclerosis*. 1989 Aug 1;78(2-3):109-12.

10. Lesnik P, Rouis M, Skarlatos S, Kruth HS, Chapman MJ. Uptake of exogenous free cholesterol induces upregulation of tissue factor expression in human monocyte-derived macrophages. *Proceedings of the National Academy of Sciences*. 1992 Nov 1;89(21):10370-4.
11. Lewis JC, Bennett-Cain AL, DeMars CS, Doellgast GJ, Grant KW, Jones NL, Gupta M. Procoagulant activity after exposure of monocyte-derived macrophages to minimally oxidized low density lipoprotein. Co-localization of tissue factor antigen and nascent fibrin fibers at the cell surface. *The American journal of pathology*. 1995 Oct;147(4):1029.
12. Wada H, Kaneko T, Wakita Y, Minamikawa K, Nagaya S, Tamaki S, Deguchi K, Shirakawa S. Effect of lipoproteins on tissue factor activity and PAI-II antigen in human monocytes and macrophages. *International journal of cardiology*. 1994 Dec 1;47(1):S21-5.
13. Drake TA, Hannani K, Fei HH, Lavi S, Berliner JA. Minimally oxidized low-density lipoprotein induces tissue factor expression in cultured human endothelial cells. *The American journal of pathology*. 1991 Mar;138(3):601.
14. Petit L, Lesnik P, Dachet C, Moreau M, Chapman MJ. Tissue factor pathway inhibitor is expressed by human monocyte-derived macrophages: relationship to tissue factor induction by cholesterol and oxidized LDL. *Arteriosclerosis, thrombosis, and vascular biology*. 1999 Feb;19(2):309-15.
15. Brand K, Banka CL, Mackman N, Terkeltaub RA, Fan ST, Curtiss LK. Oxidized LDL enhances lipopolysaccharide-induced tissue factor expression in human adherent monocytes. *Arteriosclerosis and Thrombosis: A Journal of Vascular Biology*. 1994 May;14(5):790-7.
16. Schlichting E, Henriksen T, Lyberg T. Lipoproteins do not modulate the tissue factor activity, plasminogen activator or tumour necrosis factor production induced by lipopolysaccharide stimulation of human monocytes. *Scandinavian Journal of Clinical and Laboratory Investigation*. 1994 Jan 1;54(6):465-73.
17. van den Eijnden MM, van Noort JT, Hollaar L, van der Laarse A, Bertina RM. Cholesterol or triglyceride loading of human monocyte-derived macrophages by incubation with modified lipoproteins does not induce tissue factor expression. *Arteriosclerosis, thrombosis, and vascular biology*. 1999 Feb;19(2):384-92.

18. Penn MS, Patel CV, Cui MZ, DiCorleto PE, Chisolm GM. LDL increases inactive tissue factor on vascular smooth muscle cell surfaces: hydrogen peroxide activates latent cell surface tissue factor. *Circulation*. 1999 Apr 6;99(13):1753-9.
19. Dietzen DJ, Page KL, Tetzloff TA. Lipid rafts are necessary for tonic inhibition of cellular tissue factor procoagulant activity. *Blood*. 2004 Apr 15;103(8):3038-44.
20. Awasthi V, Mandal SK, Papanna V, Rao LV, Pendurthi UR. Modulation of tissue factor–factor VIIa signaling by lipid rafts and caveolae. *Arteriosclerosis, thrombosis, and vascular biology*. 2007 Jun 1;27(6):1447-55.
21. Mandal SK, Iakhiaev A, Pendurthi UR, Rao LV. Acute cholesterol depletion impairs functional expression of tissue factor in fibroblasts: modulation of tissue factor activity by membrane cholesterol. *Blood*. 2005 Jan 1;105(1):153-60.
22. Aikawa M, Voglic SJ, Sugiyama S, Rabkin E, Taubman MB, Fallon JT, Libby P. Dietary lipid lowering reduces tissue factor expression in rabbit atheroma. *Circulation*. 1999 Sep 14;100(11):1215-22.
23. Stansfeld PJ, Sansom MS. Molecular simulation approaches to membrane proteins. *Structure*. 2011 Nov 9;19(11):1562-72.
24. Yeste-Velasco M, Linder ME, Lu YJ. Protein S-palmitoylation and cancer. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*. 2015 Aug 1;1856(1):107-20.
25. Baldwin AC, Green CD, Olson LK, Moxley MA, Corbett JA. A role for aberrant protein palmitoylation in FFA-induced ER stress and β -cell death. *American Journal of Physiology-Endocrinology and Metabolism*. 2012 Jun 1;302(11):E1390-8.
26. Park EC. Palmitoylation in Alzheimer's disease and other. *Mol. Cell Biol*. 2007;8(1):74-84.
27. Milligan G, Parenti M, Magee AI. The dynamic role of palmitoylation in signal transduction. *Trends in biochemical sciences*. 1995 May 1;20(5):181-6.
28. Resh MD. Fatty acylation of proteins: new insights into membrane targeting of myristoylated and palmitoylated proteins. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*. 1999 Aug 12;1451(1):1-6.

29. Mumby SM. Reversible palmitoylation of signaling proteins. *Current opinion in cell biology*. 1997 Apr 1;9(2):148-54.
30. Dorfleutner A, Ruf W. Regulation of tissue factor cytoplasmic domain phosphorylation by palmitoylation. *Blood*. 2003 Dec 1;102(12):3998-4005.
31. Bach RR, Moldow CF. Mechanism of tissue factor activation on HL-60 cells. *Blood, The Journal of the American Society of Hematology*. 1997 May 1;89(9):3270-6.
32. Ettelaie C, Featherby S, Rondon AM, Greenman J, Versteeg HH, Maraveyas A. De-palmitoylation of tissue factor regulates its activity, phosphorylation and cellular functions. *Cancers*. 2021 Jan;13(15):3837.
33. Rao LV, Pendurthi UR. Regulation of tissue factor coagulant activity on cell surfaces. *Journal of Thrombosis and Haemostasis*. 2012 Nov;10(11):2242-53.
34. Jo S, Lim JB, Klauda JB, Im W. CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophysical journal*. 2009 Jul 8;97(1):50-8.
35. de Meyer F, Smit B. Effect of cholesterol on the structure of a phospholipid bilayer. *Proceedings of the National Academy of Sciences*. 2009 Mar 10;106(10):3654-8.
36. Asawakarn T, Cladera J, O'Shea P. Effects of the membrane dipole potential on the interaction of saquinavir with phospholipid membranes and plasma membrane receptors of Caco-2 cells. *Journal of Biological Chemistry*. 2001 Oct 19;276(42):38457-63.
37. Banner DW, D'Arcy A, Chène C, Winkler FK, Guha A, Konigsberg WH, Nemerson Y, Kirchhofer D. The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature*. 1996 Mar;380(6569):41-6.
38. Pike AC, Brzozowski AM, Roberts SM, Olsen OH, Persson E. Structure of human factor VIIa and its implications for the triggering of blood coagulation. *Proceedings of the National Academy of Sciences*. 1999 Aug 3;96(16):8925-30.
39. Murakami MT, Rios-Steiner J, Weaver SE, Tulinsky A, Geiger JH, Arni RK. Intermolecular interactions and characterization of the novel factor Xa exosite

- involved in macromolecular recognition and inhibition: crystal structure of human Gla-domainless factor Xa complexed with the anticoagulant protein NAPc2 from the hematophagous nematode *Ancylostoma caninum*. *Journal of molecular biology*. 2007 Feb 16;366(2):602-10.
40. Harlos K, Martin DM, O'Brien DA, Jones EY, Stuart DI, Polikarpov I, Miller A, Tuddenham EG, Boys CW. Crystal structure of the extracellular region of human tissue factor. *Nature*. 1994 Aug;370(6491):662-6.
41. Muller YA, Ultsch MH, de Vos AM. The crystal structure of the extracellular domain of human tissue factor refined to 1.7 Å resolution. *Journal of molecular biology*. 1996 Feb 16;256(1):144-59.
42. Lee CJ, Wu S, Bartolotti LJ, Pedersen LG. Molecular dynamics simulations of the binary complex of human tissue factor (TF1–242) and factor VIIa (TF1–242/fVIIa) on a 4: 1 POPC/POPS lipid bilayer. *Journal of thrombosis and haemostasis: JTH*. 2012 Nov;10(11):2402.
43. Vasur J, Kawai R, Andersson E, Igarashi K, Sandgren M, Samejima M, Ståhlberg J. X- ray crystal structures of *Phanerochaete chrysosporium* Laminarinase 16A in complex with products from lichenin and laminarin hydrolysis. *The FEBS journal*. 2009 Jul;276(14):3858-69.
44. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular biology*. 1990 Oct 5;215(3):403-10.
45. Sen M, Herzik Jr M, Craft Jr JW, Creath AL, Agrawal S, Ruf W, Legge GB. Spectroscopic characterization of successive phosphorylation of the tissue factor cytoplasmic region. *The open spectroscopy journal*. 2009 Jan 1;3:58.
46. Norledge BV, Petrovan RJ, Ruf W, Olson AJ. The tissue factor/factor VIIa/factor Xa complex: a model built by docking and site- directed mutagenesis. *Proteins: Structure, Function, and Bioinformatics*. 2003 Nov 15;53(3):640-8.
47. Brasher SV, Smith BO, Fogh RH, Nietlispach D, Thiru A, Nielsen PR, Broadhurst RW, Ball LJ, Murzina NV, Laue ED. The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer. *The EMBO journal*. 2000 Apr 3;19(7):1587-97.

48. Shaw G, Gan J, Zhou YN, Zhi H, Subburaman P, Zhang R, Joachimiak A, Jin DJ, Ji X. Structure of RapA, a Swi2/Snf2 protein that recycles RNA polymerase during transcription. *Structure*. 2008 Sep 10;16(9):1417-27.
49. Bach R, Konigsberg WH, Nemerson Y. Human tissue factor contains thioester-linked palmitate and stearate on the cytoplasmic half-cystine. *Biochemistry*. 1988 Jun 1;27(12):4227-31.
50. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. *Journal of computational chemistry*. 2005 Dec;26(16):1781-802.
51. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan SA, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*. 1983 Jun;4(2):187-217.
52. MacKerell Jr AD, Bashford D, Bellott ML, Dunbrack Jr RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry B*. 1998 Apr 30;102(18):3586-616.
53. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *Journal of molecular graphics*. 1996 Feb 1;14(1):33-8.
54. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*. 1983 Jul 15;79(2):926-35.
55. Darden T, York D, Pedersen L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of chemical physics*. 1993 Jun 15;98(12):10089-92.
56. Martyna GJ, Tobias DJ, Klein ML. Constant pressure molecular dynamics algorithms. *The Journal of chemical physics*. 1994 Sep 1;101(5):4177-89.
57. Feller SE, Zhang Y, Pastor RW, Brooks BR. Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of chemical physics*. 1995 Sep 15;103(11):4613-21.
58. Andersen HC. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *Journal of computational Physics*. 1983 Oct 1;52(1):24-34.

59. Wang L, Zhang J, Wang D, Song C. Membrane contact probability: An essential and predictive character for the structural and functional studies of membrane proteins. *PLOS Computational Biology*. 2022 Mar 30;18(3):e1009972.
60. Vadivel K, Bajaj SP. Structural biology of factor VIIa/tissue factor initiated coagulation. *Frontiers in bioscience: a journal and virtual library*. 2012 Jun 1;17:2476.
61. Prasad R, Sen P. Structural modulation of factor VIIa by full-length tissue factor (TF1-263): Implication of novel interactions between EGF2 domain and TF. *Journal of Biomolecular Structure and Dynamics*. 2018 Feb 17;36(3):621-33.
62. Prasad R, Sen P. Molecular determinants involved in differential behaviour between soluble tissue factor and full-length tissue factor towards factor VIIa. *Physical Chemistry Chemical Physics*. 2017;19(33):22230-42.
63. Ruf W, Miles DJ, Rehemtulla A, Edgington TS. Tissue factor residues 157-167 are required for efficient proteolytic activation of factor X and factor VII. *Journal of Biological Chemistry*. 1992 Nov 15;267(31):22206-10.
64. Lins L, Thomas A, Brasseur R. Analysis of accessible surface of residues in proteins. *Protein science*. 2003 Jul;12(7):1406-17.
65. Belting M, Dorrell MI, Sandgren S, Aguilar E, Ahamed J, Dorfleutner A, Carmeliet P, Mueller BM, Friedlander M, Ruf W. Regulation of angiogenesis by tissue factor cytoplasmic domain signaling. *Nature medicine*. 2004 May;10(5):502-9.
66. Mizuno H, Fujimoto Z, Atoda H, Morita T. Crystal structure of an anticoagulant protein in complex with the Gla domain of factor X. *Proceedings of the National Academy of Sciences*. 2001 Jun 19;98(13):7230-4.
67. Madsen JJ, Persson E, Olsen OH. Tissue factor activates allosteric networks in factor VII a through structural and dynamic changes. *Journal of Thrombosis and Haemostasis*. 2015 Feb;13(2):262-7.
68. Prasad R, Sen P. Structural modulation of factor VIIa by full-length tissue factor (TF1-263): Implication of novel interactions between EGF2 domain and TF. *Journal of Biomolecular Structure and Dynamics*. 2018 Feb 17;36(3):621-33.

69. McCallum CD, Su B, Neuenschwander PF, Morrissey JH, Johnson AE. Tissue factor positions and maintains the factor VIIa active site far above the membrane surface even in the absence of the factor viia gla domain: A fluorescence resonance energy transfer study. *Journal of Biological Chemistry*. 1997 Nov 28;272(48):30160-6.
70. McCallum CD, Hapak RC, Neuenschwander PF, Morrissey JH, Johnson AE. The location of the active site of blood coagulation factor VIIa above the membrane surface and its reorientation upon association with tissue factor: a fluorescence energy transfer study. *Journal of Biological Chemistry*. 1996 Nov 8;271(45):28168-75.
71. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *Journal of molecular biology*. 2002 Jul 5;320(2):369-87.

Chapter 4

Identification of potential antifibrinolytic compounds against kringle-1 and serine protease domain of Plasminogen and kringle-2 domain of tissue-type plasminogen activator using combined virtual screening, molecular docking, and molecular dynamics simulation approaches.

1. Introduction

Fibrinolysis is a process of removal (lysing) of clots formed due to the triggering of the hemostatic pathway, in conditions such as pathological thrombosis and vascular occlusion [1]. The principal mediator of fibrinolysis is plasminogen, which is transformed into plasmin by the plasminogen activators (Tissue-type plasminogen activator and Urokinase plasminogen activator) Figure 4-1[2]. Tissue-type plasminogen activator (tPA or PLAT) is a serine protease found on the surface of endothelial cells that plays an important role in degrading polymerized fibrin by hydrolyzing the Arg560-Va1561 peptide bond in plasminogen (Plg) [3]. t-PA consists of six well-defined structural modules: a finger domain; an epidermal growth factor domain; two kringle domains and a serine protease two-domain unit with trypsin-like specificity [4]. Because of the presence of a lysine binding site in the above-mentioned two kringle domains and its stimulatory effect on Plg activation by fibrin, the scientific community has shown significant interest in the kringle domain [5]. The zymogen protease Plg and its active form plasmin (Plm), carry out important functions in blood clot disintegration apart from other cellular processes such as bacterial pathogenesis, tissue remodeling, and cell migration [6]. Plg comprises a serine protease (SP) domain, a Pan-apple (PAP) domain, and 5 kringle (KR 1-5) domains. These domains are ordered in a closed configuration that is sustained through interdomain interactions facilitated by the canonical lysine-binding sites (LBSs) of the KR domains. The binding of Plg to lysine-containing substrates or receptors is also facilitated by the LBSs of the KR domains. Post binding, Plg takes an open configuration that is triggered by tissue plasminogen activators (tPAs) or urokinase plasminogen activators (uPAs) [7]. To regulate this

binding process, α_2 -antiplasmin and α_2 -macroglobulin block the release of activated Plm from the clot or the cell surface [8]. Irregular balance of Plg/Plm system and immoderate fibrinolytic activity leads to fatal hemorrhagic disorders, thrombotic vascular injury, and severe complications during general surgery or major trauma [9].

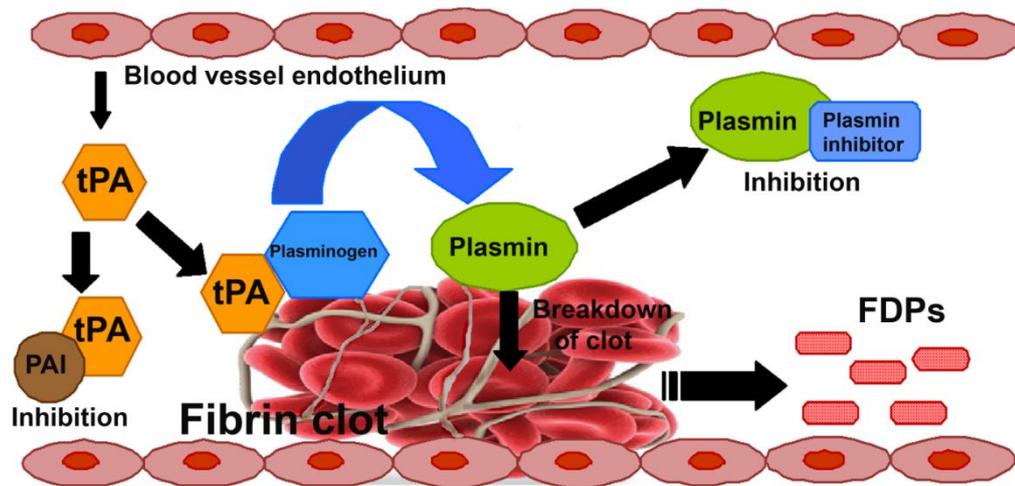


Figure 4-1: The interplay of enzymes in the process of fibrinolysis. Abbreviations used are FDPs, fibrin degradation products; PAI, plasminogen activator inhibitors; tPA, tissue plasminogen activator. (Adapted from Bhattacharjee P, Bhattacharyya D. An insight into the abnormal fibrin clots—its pathophysiological roles. *Fibrinolysis and thrombolysis*. 2014 May 7:1-29.)

Presently, the Plm inhibitors aprotinin and tranexamic acid (TXA) are clinically used to treat severe hemorrhages by blocking new entry to target substrates and are the most extensively used inhibitor for the Plg/Plm system [10]. TXA is a reasonably well-tolerated lysine analog that binds to the LBSs, culminating in the development of open Plg [11]. Post binding, TXA obstructs Plm activation by stopping it from binding to substrates and/or receptors. However, one recent study has shown that TXA may not block the activity of Plm which has already been formed leading to dangerous complications in patients suffering from extreme blood loss [12]. Hence, there is a requirement to synthesize specific inhibitors which can attach to the protease active site of Plg. Cross reaction of Plm inhibitors with other plasma serine proteases also challenges their development because the reactions generate multiple side effects like headaches, nasal symptoms, or back, abdominal, and muscle pain. In some cases of treatment with TXA, seizures were reported as a side effect which in turn were traced to gamma-aminobutyric acid antagonistic activity (GABA_A) [13]. Also, the gastrointestinal side effects associated with a high dose of TXA can be attributed to

GABAa [14]. Based on the findings of these studies, the ideal novel oral fibrinolysis inhibitor is expected to work on a mechanism of action similar to that of TXA, having suitable dosing and higher selectivity over GABAa. Of late, some new encouraging drugs have been discovered like 4-PIOL and PS-112 (TXA-derived active site inhibitors) but, to date, their detailed profiling and testing have not been performed [15-16]. Even though considerable research has been undertaken, clinically safe fibrinolysis inhibitors have not reached the market. Fibrinolysis can be suppressed by targeting each of the three important protein domains: the kringle-2 domain of tissue plasminogen activator, the kringle-1 domain of plasminogen, and the serine protease domain of plasminogen. Previous studies reported that the catalytically active domain is constituted by the serine protease domain while the kringle domains are crucial for protein-protein interactions such as binding to fibrin [17]. Therefore, we have selected the above-mentioned three individual protein domains as our respective targets [18-20]. Herein, we attempt to find potential inhibitors having minimum side effects with maximum specificity and efficacy using *in silico* approaches. In this present study, combined approaches of structure-based virtual screening and molecular docking were employed to discern prospective hits from the ligand database [21]. Thereafter, ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) properties of these identified ligands were evaluated and a molecular dynamics simulation of 200ns was carried out for each protein-ligand complex. Furthermore, principal component analysis (PCA) was performed and binding free energy was calculated using the Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) approach for all the protein-ligand complexes. The identified lead compounds are expected to contribute to the development of potent anti-fibrinolytic agents.

2. Methodology

2.1. Protein Preparation:

The crystal structures of the kringle-2 domain of tissue plasminogen activator, the kringle-1 domain of plasminogen, and the serine protease domain of plasminogen with PDB IDs: 1PK2, 4CIK, 5UGG were retrieved from the protein data bank (PDB) respectively [22-25]. All the three protein structures were analyzed in the pymol and discovery Studio 3.5 package (Accelrys, San Diego, CA, USA), and the missing residues in the structures were added using Modeller [26-28]. The co-crystallized

ligands with IDs: ACA, XO3, 89M were extracted from protein crystal structures 1PK2, 4CIK, and 5UGG respectively using discovery Studio 3.5. The water molecules were eliminated, and protein preparation was done using the default settings of discovery Studio 3.5. Further, the protein structures were optimized using CHARMM (Chemistry at Harvard Macromolecular Mechanics) force field, and energy minimization was done by a smart minimizer algorithm incorporating conjugate gradient energy protocol [29]. Subsequently, these protein structures were selected as targets for virtual screening.

2.2. Structure-based virtual screening

The library of one million ligands was retrieved from the ZINC 15 database [30]. Initially, these one million ligands were screened individually against 1PK2, 4CIK, and 5UGG protein targets respectively using RASPD [31]. This tool rapidly identifies hit molecules against the protein targets based on their physicochemical properties and calculates the binding energy. Subsequently, the hits were further evaluated based on Lipinski parameters such as hydrogen bonds acceptors & donors, molar refractivity, ADMET, Wiener index, and volume of functional groups & protein. The residues which were bound to the co-crystallized ligands were selected as active site pockets of the target proteins for virtual screening. From RASPD, the top 1500 hits (500 for each protein target) were retrieved for further analysis.

2.3 Molecular Docking

The top 500 hits filtered out from virtual screening (1500 total) were subjected to molecular docking against the three protein targets 1PK2, 4CIK, and 5UGG using three docking tools such as Autodock Vina, Glide, and ParDOCK/BAPPL+ respectively [32-35].

2.3.1. Glide (Grid-based ligand docking with energetic)

Firstly, the ligands were prepared using the LigPrep module of Schrodinger Maestro version 2.9, 2019 so that the ligands could be supplied to glide for docking in a state as they would occur in a protein-ligand complex [36]. LigPrep uses Optimized Potential Liquid Simulation (OPLS) 2005 force field and consists of a series of steps to optimize the structures. By default, LigPrep adds hydrogens, removes unwanted molecules, and

minimizes the ligand structure. Thereafter, docking was performed using the software package Glide version 2.9, 2019. Protein structures were prepared in the Protein Preparation Wizard of Schrodinger in which proteins were refined by adding missing hydrogens and assigning proper bond orders. A grid file was generated using the Receptor Grid Generation protocol on these minimized structures, such that the grid points were adjusted to the active site pocket in X, Y, Z dimensions with $25 \times 25 \times 25$ points and the ligands were allowed to move freely. We have used Glide docking modules namely simple precision (SP) and extra precision (XP) for docking calculations [37]. The conformations obtained from SP were used as input for XP to generate multiple binding poses. The binding affinity was computed based on the glide scoring function for all the three target proteins i.e., 1PK2, 4CIK, and 5UGG respectively. Thereafter, the top ten ligands were ranked according to their binding affinity using GScore (GlideScore) scoring function which is a combination of Coulombic energies, Van der Waals energies, and the internal strain of the ligand [38].

$$\text{GScore} = 0.065 * \text{vdW} + 0.130 * \text{Coul} + \text{Lipo} + \text{Hbond} + \text{Metal} + \text{BuryP} + \text{RotB} + \text{Site.}$$
where,

Glide evdW - Van der Waal energy

Glide Ecoul - Coulomb energy

Glide Lipo - Lipophilic contact term

Glide Hbond - Hydrogen-bonding term

Glide Metal- Metal-binding term

Glide BuryP- Penalty for buried polar groups

Glide RotB- Penalty for freezing rotatable bonds

Glide Site- Polar interactions in the active site

2.3.2. Auto Dock Vina

As mentioned earlier, 1500 hits (500 hits per protein) were processed and subjected to docking analysis in AutoDock Vina. ADT tool was used to load the target protein structure and convert proteins to PDBQT format. Subsequently, the ligand hits were also converted into PDBQT format in Open Babel [39]. The hydrogen atoms and water molecules from the protein structure were removed and polar hydrogen atoms were added. Thereafter, Kollman charges were incorporated into the target protein structures [40]. The auto-grid box was set across X-Y-Z directions, i.e., 23x20x18 points for

tissue-type plasminogen activator kringle-2 domain(1PK2), 18x20x18 points for plasminogen kringle-1 domain (4CIK), and 29x28x28 points for serine protease domain of plasminogen (5UGG) respectively. The grid was set up to encompass the active site pocket, with a spacing of 1 Å. The ligand hits were ranked according to their binding energies and finally, the best performing top ten ligands for each protein target were selected.

2.3.3. ParDOCK & Bappl+

The same set of 1500 hits (500 hits per protein) was screened by the Monte Carlo method involving an all-atom energy-based system using default parameters of ParDOCK software, hosted in the Sanjeevani server [41]. The ligands were positioned optimally around the active site of the protein target and ranked based on their interacting energies. Subsequently, the scoring function of all the aforementioned ligands was derived using the Bappl+ tool which utilizes the Random Forest algorithm scoring function [42]. In Bappl+ the default parameters were set and formal charges were assigned for each ligand. Based on these settings, the binding affinities of ligands were computed and the top ten compounds were ranked accordingly.

2.4. ADMET and TOPKAT

All the top ten identified ligands obtained by screening from Glide, Autodock Vina, ParDOCK, and Bappl+ tools were visually inspected in Pymol and Schrodinger maestro. Based on the docking scores, out of these ten identified ligands, the top three were selected for ADMET and TOPKAT (Toxicity Prediction by Computer Assisted Technology) analysis for each protein target. ADMET properties such as CYP2D6 & plasma protein inhibition, intestinal adsorption, aqueous solubility, hepatotoxicity, blood-brain barrier level were evaluated using the Discovery studio 3.5 ADMET tool kit. Furthermore, Lipinski's rule of five was also checked for these ligands [43]. Thereafter, the toxicity of each ligand was estimated using the TOPKAT tool, which contains a robust Quantitative Structure-Toxicity Relationship (QSTR) modeling system to predict the accurate toxicity endpoints [44]. The toxicity properties such as rodent carcinogenicity, Ames mutagenicity, skin irritation, and developmental toxicity potential were also investigated.

2.5. Molecular Dynamics Simulation (MD)

The top three ranked ligands for each protein target were subjected to MD simulation using GROMACS (Groingen Machine for Chemical Simulations) 2019 suite implementing GROMOS (54a7) force field [45,46]. The protein-ligand complex was solvated in a cubic box of 0.8 nm using periodic boundary conditions and a Simple point-charge water model (SPCE) [47]. The Na⁺ and Cl⁻ ions were added to neutralize the system and maintain the concentration of 0.15 mol/L. The PRODRG server was used to generate the ligand parameters and topology [48]. After that, internal constraints of the protein-ligand complex were relaxed by 2000 steps of steepest descent energy minimization with a max force constraint of 1000 KJ/mol, leading to restraining positions of all heavy atoms. Before MD simulations, the systems were heated using a V-rescale thermostat to attain the temperature of 310 K with 0.1ps as the constant of coupling and achieved equilibration in NVT (Number of atoms, Volume of the system, and Temperature of the system). Then solvent density was sustained using a Parrinello-Rahman barostat with the pressure of 1 bar, coupling constant of 0.1 ps, and temperature of 310 K to obtain equilibration in NPT (Number of atoms in the system, the pressure of the system, and temperature of the system) by gradually discharging the restraint on heavy atoms step by step [49]. Finally, an MD simulation was performed for the equilibrated structures for 5 ns with an integration time step of 2fs. The electrostatic interactions of long-range were implemented using particle-mesh Ewald sum with a cutoff of 1.0 nm [50]. During simulation, the LINCS algorithm was used to constrain all bond lengths, and the SETTLE algorithm was used for restraining water molecules [51-52]. The resultant structure from the NPT equilibration phase was employed for the final production run in the NPT ensemble for 200 ns simulation time. Finally, the trajectory analysis such as RMSD, RMSF, RoG, SASA, H-bonds, and PCA of protein-ligand complexes was performed using gromacs utilities and plotted in xmgrace [53].

2.6. Principal Component Analysis

PCA is a statistical method that is applied to reduce data complexity and is useful for analyzing the large-scale motion of proteins [54]. Generally, internal motions of every protein molecule are ingrained and are required for their proper biological functioning like conformational changes to various biological environments, binding to substrate,

etc. Due to problems in interpreting these internal motions of the protein, PCA is utilized to minimize the large dimensions of the data set to identify the prominent principal components (PCs). These PCs represent the major contributors responsible for elucidating crucial information about the dynamic changes of the protein. In PCA, a covariance matrix was constructed from trajectory data after eliminating unwanted motions (translational and rotational). This was followed by diagonalization of the covariance matrix using the *gmx_covar* tool which is a part of the Gromacs module. By diagonalization of covariance matrices containing backbone C alpha atoms data, the eigenvector and eigenvalues were obtained which corresponds to the change in protein trajectory throughout the simulation time. The *gmx_anaeig* tool of Gromacs was used to analyze and plot the trajectories of the backbone C alpha atoms of all the systems. In our study, the first two projections (PC 1 and PC 2) with the highest eigenvalues were taken into consideration as they are responsible for 80–90% of the collective motions of the C alpha backbone atoms.

2.7. Free energy calculation

To evaluate the binding free energy (ΔG) of protein-ligand complexes, the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) method was employed [55]. We computed ΔG for the last 50ns of the production run using the *g_mmpbsa* tool of the Gromacs module [56]. Binding free energy was calculated for the identified ligands by estimating the bound and unbound state differences with the protein targets.

$$\Delta G_{\text{binding}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}}) \longrightarrow (1)$$

In equation (1), the free energies of protein and ligands are indicated by G_{protein} and G_{ligand} . The free energy of the protein-ligand complex is represented by the G_{complex} .

$$G_x = (E_{\text{MM}}) - TS + (G_{\text{solv}}) \longrightarrow (2)$$

Similarly, equation (2) is used to calculate the bound or unbound state's free energy. The x indicates the unbound states or protein-ligand complex. E_{MM} is used to calculate the average molecular mechanics energy and TS indicates the entropic contribution. The solvation free energy is indicated by G_{solv} .

$$E_{\text{MM}} = E_{\text{bonded}} + E_{\text{nonbonded}} \longrightarrow (3)$$

Equation (3) shows that bonded and non-bonded (i.e. vander Waal's & electrostatic) interactions were considered for calculating the molecular mechanics energy (E_{MM}) for protein-ligand interaction.

$$G_{solv} = G_{nonpolar} + G_{polar} \longrightarrow (4)$$

The linearized Poisson Boltzmann equation is represented by G_{solv} , where polar and non-polar hydrophobic contribution (G_{polar} & $G_{nonpolar}$) is taken into consideration for estimating the solvent accessible surface area. It was assumed that as all ligands are binding to one distinct protein target, therefore these ligands will contribute similar entropic energy. Hence, in our analysis complicated entropic contribution was not taken into account and was removed at the time of calculation.

3. Results and discussion

3.1 Virtual screening (VS)

VS is the most widely used technique in the drug discovery process. The complete flow chart of our current study is represented in Figure 4-2. As mentioned earlier, we have selected three protein targets which known to play a crucial role in the fibrinolysis process i.e., tissue-type plasminogen activator kringle-2 domain (1PK2), plasminogen kringle-1 domain (4CIK), and serine protease domain of plasminogen (5UGG). From the crystallographic structures of protein i.e., 1PK2, 4CIK, and 5UGG, the co-crystallized ligands with ligand IDs ACA, XO3, and 89M were removed from the protein complex, and the residues found interacting with these ligands were selected as active site pockets. We have used the structure-based drug design approach to identify the lead compounds which could bind to the specific pocket region accordingly [57]. As mentioned earlier, one million molecules were retrieved from the ZINC database and post-screening of these compounds in the RASPD tool we obtained 500 hits for each protein target (1500 hits total). These 1500 compounds (500 per protein) were optimized in the Discovery studio and molecular docking was performed by three different tools accordingly.

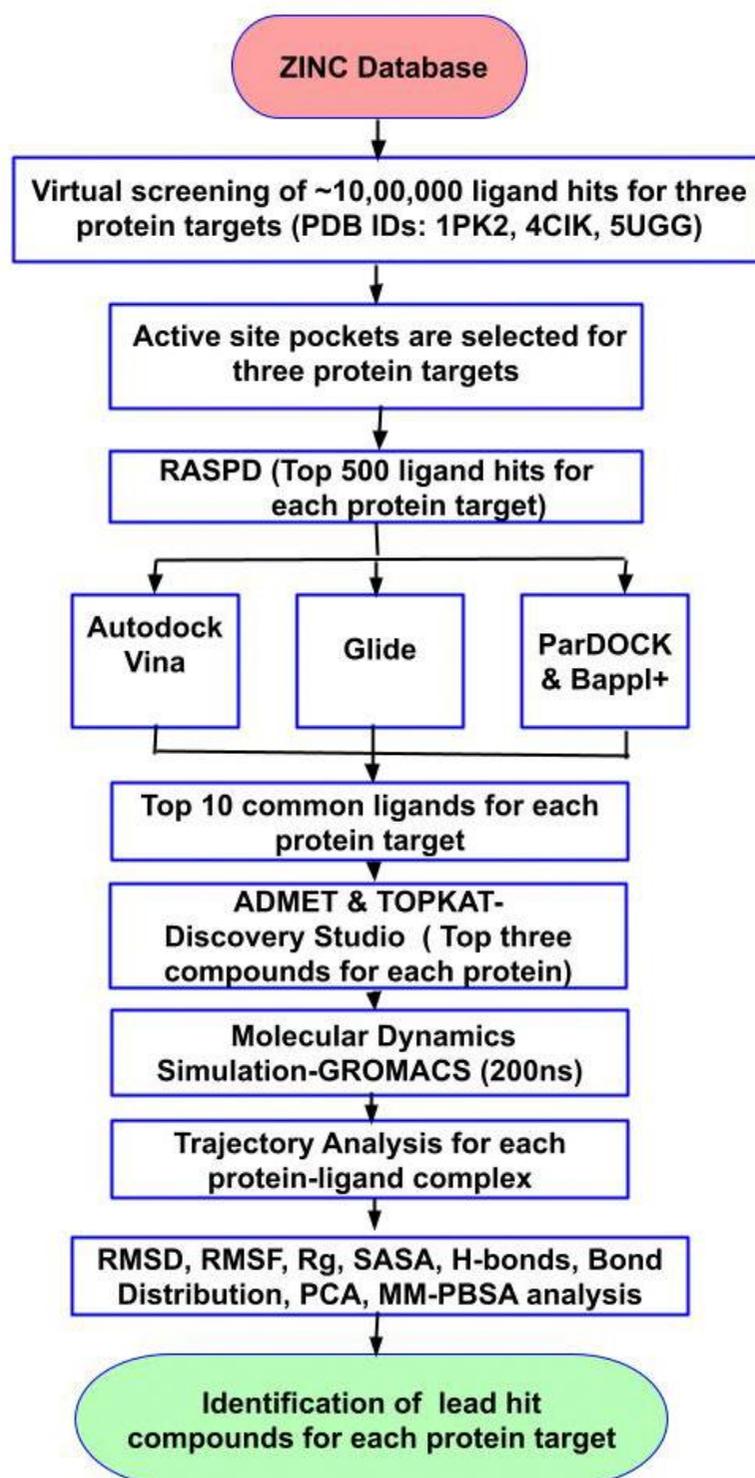


Figure 4-2: Flow chart of the virtual screening process using combined molecular docking and molecular dynamics simulation approaches.

3.2 Comparative Molecular Docking

To understand the interaction profile of 1500 ligands (500 per protein) with the protein targets, we conducted a comparative molecular docking analysis using three different

tools, i.e., Auto dock Vina, Glide, and ParDOCK/BAPPL+. The 500 hits were individually docked to the active site pocket of 1PK2, 4CIK, and 5UGG respectively. The key active site residues of the 1PK2 protein target include TRP80, THR79, LEU78, ARG71, TRP69, ASP63, and VAL40. Similarly, the binding site of the 4CIK protein target includes ARG35, ASP55, ASP57, TRP62, TYR64, ARG71, and TYR72 active site residues. Likewise, for the 5UGG protein target, the residues selected for docking include PHE587, CYX588, HID603, LYS607, and SER736, CYX737, GLN738, GLY739, SER741, THR759, TRP761, and GLY762. Based on the binding energies, the top 10 hits for each target protein were ranked based on the scores obtained from the three docking tools: Autodock vina, ParDOCK/BAPPL+, and Glide (Table 4-1, Table 4-2, and Table 4-3). Detailed Glide SP and XP scores of the top three ligands hit against each protein target are also provided in (Table 4-4). Among these 10 hits, we narrowed it down to the top three ligands by selecting the common hits found in all three docking tools. The docking results obtained from docking all the nine identified ligands (P27, P67, P76, C19, C90, C97 U19, U94, U97) and the reference ligands (ACA, XO3, and 89M) with their respective protein targets (1PK2, 4CIK, and 5UGG) are provided in Table 4-5. The interacting residues which include both the active site and novel (other than the active site) residues of the proteins (1PK2, 4CIK, and 5UGG) with their respective ligands (P27, P67, P76, C19, C90, C97 U19, U94, U97) 1PK2 are provided in Table 4-1. The 2D structures of the nine identified compounds (three for each protein target) are provided in Figure 4-3 [(A)-(I)]. As mentioned earlier, for comparative docking studies of the 1PK2 protein target, the ligand ACA was selected as the reference ligand. Its binding energy was found to be -3.9 kcal/mol and -4.9 kcal/mol in Autodock Vina and ParDOCK/BAPPL+ respectively. Similarly, the Glide score obtained for the reference ligand was -7.25 kcal/mol. Simultaneously, we conducted docking studies for each of the top three ligands i.e., ZINC12211330 (P27), ZINC00718625 (P67), ZINC09970930 (P76) against the 1PK2 protein target. The identified ligands P27, P67, P76 exhibited binding energy of -7.6, -8.3, -8.9 kcal/mol in Autodock Vina and -8.84, -9.60, -9.63 kcal/mol. in ParDOCK/ BAPPL+ respectively. Similarly, the Glide score obtained for P27, P67 & P76 was -8.6, -9.31, -9.42 kcal/mol respectively. As shown, the interacting residues of 1PK2 with P27 are LYS39, VAL40,

Table 4-1: AUTODOCK-VINA scores of the top 10 identified ligands docked with their respective protein targets.

AUTODOCK-VINA SCORES (TOP 10)					
1PK2		4CIK		5UGG	
Ligands	Binding energy (kcal/mol)	Ligands	Binding energy (kcal/mol)	Ligands	Binding energy (kcal/mol)
P76	-8.9	C97	-8.8	U97	-11.4
P27	-8.3	C38	-8.6	U94	-11
P95	-7.8	C19	-8.5	U19	-10.9
P67	-7.6	C90	-8.2	U90	-10.63
P03	-7.58	C63	-8.1	U26	-10.6
P83	-7.57	C20	-8	U08	-10.5
P09	-7.54	C07	-7.9	U51	-10.3
P01	-7.51	C10	-7.8	U80	-10.2
P71	-7.4	C27	-7.7	U18	-10
P78	-7.32	C66	-7.6	U87	-9.88
ACA-std	-3.9	XO3-std	-7.4	89M-std	-9

Table 4-2: PARDOCK/BAPPL+ scores of the top 10 identified ligands docked with their respective protein targets.

PARDOCK/BAPPL+ SCORES (TOP 10)					
1PK2		4CIK		5UGG	
Ligands	Binding energy (kcal/mol)	Ligands	Binding energy (kcal/mol)	Ligands	Binding energy (kcal/mol)
P66	-11.85	C86	-11.03	U97	-8.33
P74	-11.09	C62	-10.76	U7	-8.08
P79	-10.41	C83	-10.71	U94	-8.02
P78	-10.18	C22	-10.68	U59	-8.01
P44	-9.98	C38	-10.6	U19	-7.9
P48	-9.85	C12	-10.57	U55	-7.86
P22	-9.73	C97	-9.47	U15	-7.74
P76	-9.63	C18	-9.2	U81	-7.7
P27	-9.6	C19	-9	U35	-7.69
P67	-8.84	C90	-8.9	U40	-7.6
ACA-std	-4.9	XO3-std	-6.66	89M-std	-5.9

Table 4-3: GLIDE scores of the top 10 identified ligands docked with their respective protein targets.

GLIDE SCORES (TOP 10)					
1PK2		4CIK		5UGG	
Ligands	Binding energy (kcal/mol)	Ligands	Binding energy (kcal/mol)	Ligands	Binding energy (kcal/mol)
P82	-10.11	C32	-9.9	U88	-9.78
P86	-9.98	C62	-9.5	U97	-9.6
P67	-9.7	C97	-9.3	U96	-9.53
P48	-9.56	C19	-8.92	U94	-9.41
P76	-9.42	C74	-8.81	U49	-9.33
P27	-9.31	C90	-8.75	U22	-9.21
P67	-8.6	C36	-8.56	U19	-8.87
P22	-8.4	C49	-8.51	U11	-8.76
P64	-8.37	C40	-8.42	U09	-8.54
P93	-8.1	C15	-8.13	U08	-8.39
ACA-std	-7.25	XO3-std	-7.4	89M-std	-7.5

Table 4-4: GLIDE scores of the top 3 identified ligands docked with their respective protein targets.

GLIDE DOCKING SCORES (TOP 3 identified and reference ligands)					
Protein target	Ligand	Docking Score	XP Gscore	Glide Gscore	Glide energy
1PK2	P27	-9.12	-9.42	-9.42	-39.63
	P67	-8.52	-8.6	-8.6	-37.27
	P76	-9.31	-9.31	-9.31	-35.56
	ACA	-7.11	-7.25	-7.25	-22.85
4CIK	C19	-8.34	-8.92	-8.92	-43.23
	C90	-7.82	-8.75	-8.75	-39.23
	C97	-8.9	-9.3	-9.3	-40.01
	XO3	-6.75	-7.15	-7.15	-31.62
5UGG	U19	-9.21	-9.6	-9.6	-88.73
	U94	-8.67	-9.41	-9.41	-85.92
	U97	-8.33	-8.87	-8.87	-79.37
	89M	-7.45	-7.5	-7.5	-70.75

Table 4-5: Molecular docking results of the top nine identified ligands (P27, P67, P76, C19, C90, C97 U19, U94, U97) and the reference ligands (ACA, XO3, and 89M) bound with their respective protein targets (1PK2, 4CIK and 5UGG) using three docking tools: Autodock vina, ParDOCK/BAPPL+, and Glide respectively.

Protein target	Ligands	Autodock Vina-Binding Energy (Kcal/mol)	ParDOCK and BappL-Binding energy (Kcal/mol)	Glide Score	Interacting residues
1PK2	P27	-8.3	-9.6	-9.31	LYS39, VAL40, TYR41, ASP63, TRP69, HIS71, ARG77, LEU78, THR79, TRP80, GLU81, TYR82
	P67	-7.6	-8.84	-8.6	GLU23, LYS39, VAL40, TYR41, HIS71, ASP63, TRP69, ARG77, LEU78, TRP80, GLU81, TYR82
	P76	-8.9	-9.63	-9.42	LYS39, VAL40, TYR41, TRP69, HIS71, ARG77, LEU78, THR79, TRP80, GLU81, TYR82
	ACA	-3.9	-4.9	-7.25	LYS39, VAL40, TYR41, ASP63, ASP65, LEU78, THR79, TRP80, TRP69
4CIK	C19	-8.5	-9	-8.92	ARG35, ASP55, ASP57, PRO58, GLN59, TRP62, TYR64, ARG71, TYR72, TYR74
	C90	-8.2	-8.9	-8.75	ARG35, ASP55, ASP57, GLN59, TRP62, GLU69, LYS70, ARG71, TYR72, ASP73, TYR74
	C97	-8.8	-9.47	-9.3	ARG35, PHE36, ASP55, ASP57, PRO58, GLN59, TRP62, TYR64, ARG71, TYR72, TYR74

Protein target	Ligands	Autodock Vina-Binding Energy (Kcal/mol)	ParDOCK and BappL-Binding energy (Kcal/mol)	Glide Score	Interacting residues
	XO3	-7.4	-6.66	-7.15	ARG35, PHE36, ARG71, TYR72, TYR74, TYR64, TRP62, ASP55, ASP57
5UGG	U19	-10.9	-7.9	-8.87	HID63, CYS64, LEU65, GLU66, LYS67, SER68, ARG70, SER72, SER73, TYR74, GLN198, SER201
	U94	-11	-8.02	-9.41	SER68, LYS67, TYR74, GLU66, LEU65, CYS64, HID63, CYS197, SER196, ASP195, VAL233, GLY232, CYS225, GLY224, LEU223, GLY222, TRP221, SER220, SER201
	U97	-11.4	-8.33	-9.6	ALA62, HID63, CYS64, LEU65, GLU66, LYS67, SER68, TYR74, SER73, CYS197, SER196, ASP195, GLY232, VAL233, SER220, TRP221, GLY222, LEU223, GLY224, SER201
	89M	-9	-5.9	-7.5	SER68, LYS67, GLU66, LEU65, TYR74, PHE47, CYS64, HID63, GLY146, GLU147, THR148, GLN198, GLY199, SER201, SER220, TRP221, GLY222, LEU223, GLY224

TYR41, ASP63, TRP69, HIS71, ARG77, LEU78, THR79, TRP80, GLU81, TYR82. Likewise, the interacting residues of 1PK2 with P67 are GLU23, LYS39, VAL40, TYR41, HIS71, ASP63, TRP69, ARG77, LEU78, TRP80, GLU81, TYR82. Similarly,

the interacting residues of 1PK2 with P76 are LYS39, VAL40, TYR41, TRP69, HIS71, ARG77, LEU78, THR79, TRP80, GLU81, TYR82. Docking studies for each of the top three ligands i.e., ZINC02060288 (C19), ZINC12455413 (C90), ZINC14888376 (C97) were performed against the 4CIK protein target. The identified ligands C19, C90, C97 exhibited binding energy of -8.2, -8.5, -8.8 kcal/mol in Autodock Vina and -8.90, -9.0, -9.47 kcal/mol in ParDOCK/ BAPPL+ respectively. Similarly, the Glide score obtained for C19, C90, C97 was -8.75, -8.92, -9.3 kcal/mol respectively. The reference ligand XO3 exhibited binding energy of -7.4 kcal/mol in Autodock Vina and -6.66 kcal/mol in ParDOCK/BAPPL+ and -7.25 in Glide respectively. As shown, the interacting residues of 4CIK with C19 are ARG35, ASP55, ASP57, PRO58, GLN59, TRP62, TYR64, ARG71, TYR72, TYR74. Likewise, the interacting residues of 4CIK with C90 are ARG35, ASP55, ASP57, GLN59, TRP62, GLU69, LYS70, ARG71, TYR72, ASP73, TYR74. Similarly, the interacting residues of 4CIK with C97 are ARG35, PHE36, ASP55, ASP57, PRO58, GLN59, TRP62, TYR64, ARG71, TYR72, TYR74. In the same way, docking studies for each of the top three ligands i.e., ZINC12576410 (U19), ZINC04557820 (U94), ZINC11839443 (U97) were conducted against the 5UGG protein target. The identified ligands U19, U94 and U97 exhibited binding energy of -10.9, -11.0, -11.4 kcal/mol in Autodock Vina and -7.9, -8.02, -8.33 kcal/mol in ParDOCK/ BAPPL+ respectively. Similarly, the Glide score obtained for U19, U94, and U97 was -8.87, -9.41, and -9.6 kcal/mol respectively. The reference ligand 89M exhibited binding energy of -9 kcal/mol in Autodock Vina and -5.9 kcal/mol in ParDOCK/BAPPL+ and -7.5 in Glide respectively. As shown, the interacting residues of 5UGG with U19 are HID63, CYS64, LEU65, GLU66, LYS67, SER68, ARG70, SER72, SER73, TYR74, GLN198, SER201. Likewise, the interacting residues of 5UGG with U94 are SER68, LYS67, TYR74, GLU66, LEU65, CYS64, HID63, CYS197, SER196, ASP195, VAL233, GLY232, CYS225, GLY224, LEU223, GLY222, TRP221, SER220, SER201.

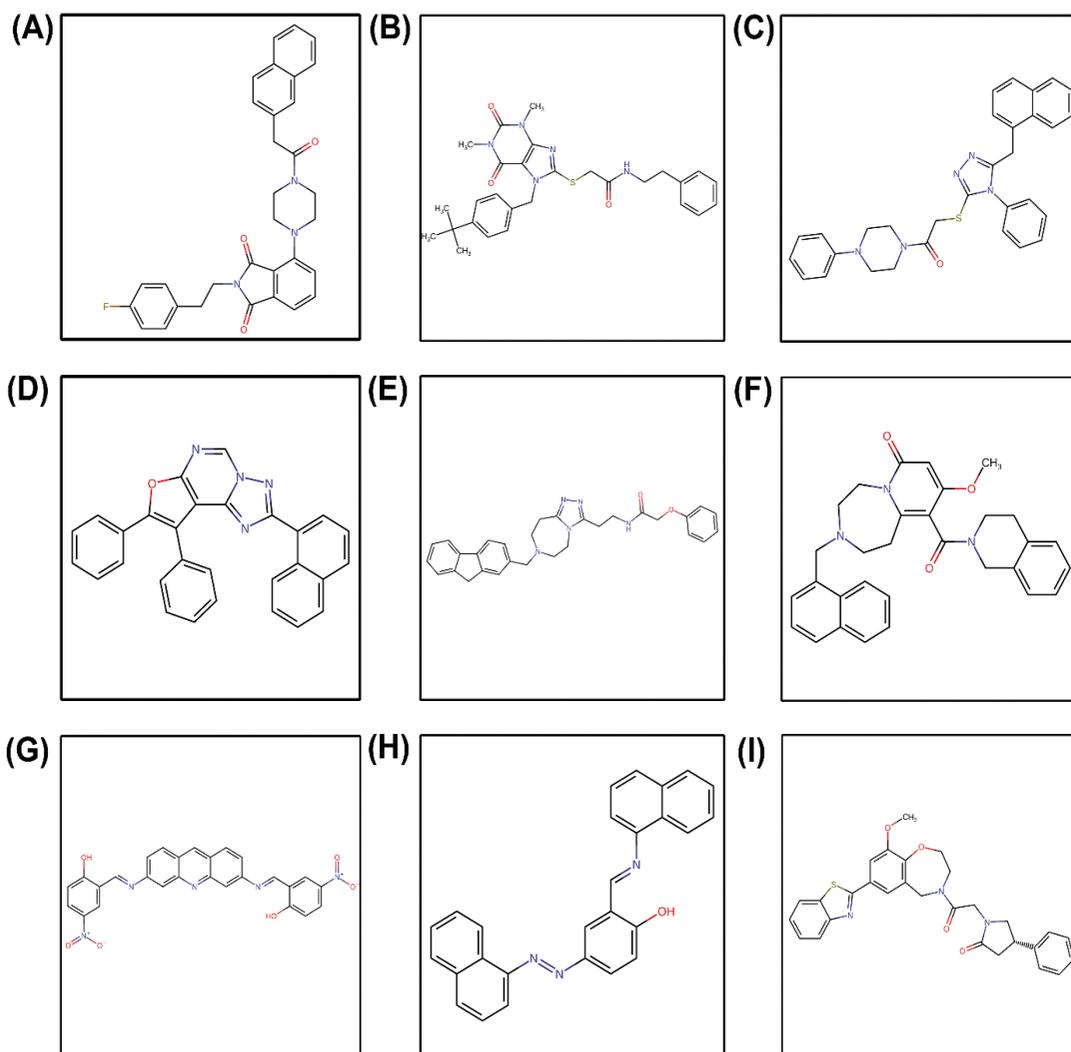


Figure 4-3: The 2D structures of the lead identified compounds (A) ZINC12211330 (P27), (B) ZINC00718625 (P67), (C) ZINC09970930 (P76), (D) ZINC02060288 (C19), (E) ZINC12455413 (C90), (F) ZINC14888376 (C97), (G) ZINC12576410 (U19), (H) ZINC04557820 (U94), (I) ZINC11839443 (U97)

Similarly, the interacting residues of 5UGG with U97 are ALA62, HID63, CYS64, LEU65, GLU66, LYS67, SER68, TYR74, SER73, CYS197, SER196, ASP195, GLY232, VAL233, SER220, TRP221, GLY222, LEU223, GLY224, SER201. Overall, from the docking results, it can be construed that in comparison to reference ligands ACA, XO3, and 89M, the identified ligands P27, P67, P76, C19, C90, C97 U19, U94, U97 exhibited higher binding energy and stronger affinity towards their respective protein targets 1PK2, 4CIK, and 5UGG. Apart from the active site residues of the target proteins, no other residues are found to interact with the reference ligand ACA, XO3, and 89M. Therefore, it can be ascertained that the identified ligands form more favorable interactions with their respective protein targets, thus enhancing stronger and stable binding. The 2D interaction images of all the top nine ligands P27, P67, P76,

C19, C90, C97 U19, U94, U97 complexed with their respective protein targets 1PK2, 4CIK, and 5UGG are also provided in Figure 4-4 [(A)-(I)]. It is noteworthy to mention that out of the nine identified ligands, P76, C97, and U97 are the top three ligands that bind with greater affinity to their respective targets 1PK2, 4CIK, and 5UGG as implied from the docking scores obtained from Autodock vina, ParDOCK/BAPPL+ and Glide (Table 4-5). Post docking, ADMET & TOPKAT properties such as solubility, intestinal adsorption, etc. of the identified ligands and the reference ligands were investigated, As provided in Table 4-6 and Table 4-7, these ligands were found non-hepatotoxic, have the least development toxicity and they don't inhibit the cytochrome 4502D6 (CYP2D6) enzyme. The identified ligands were also found not to bind to plasma protein (PP) and don't exhibit any rodent carcinogenicity, Ames mutagenicity, and skin irritation.

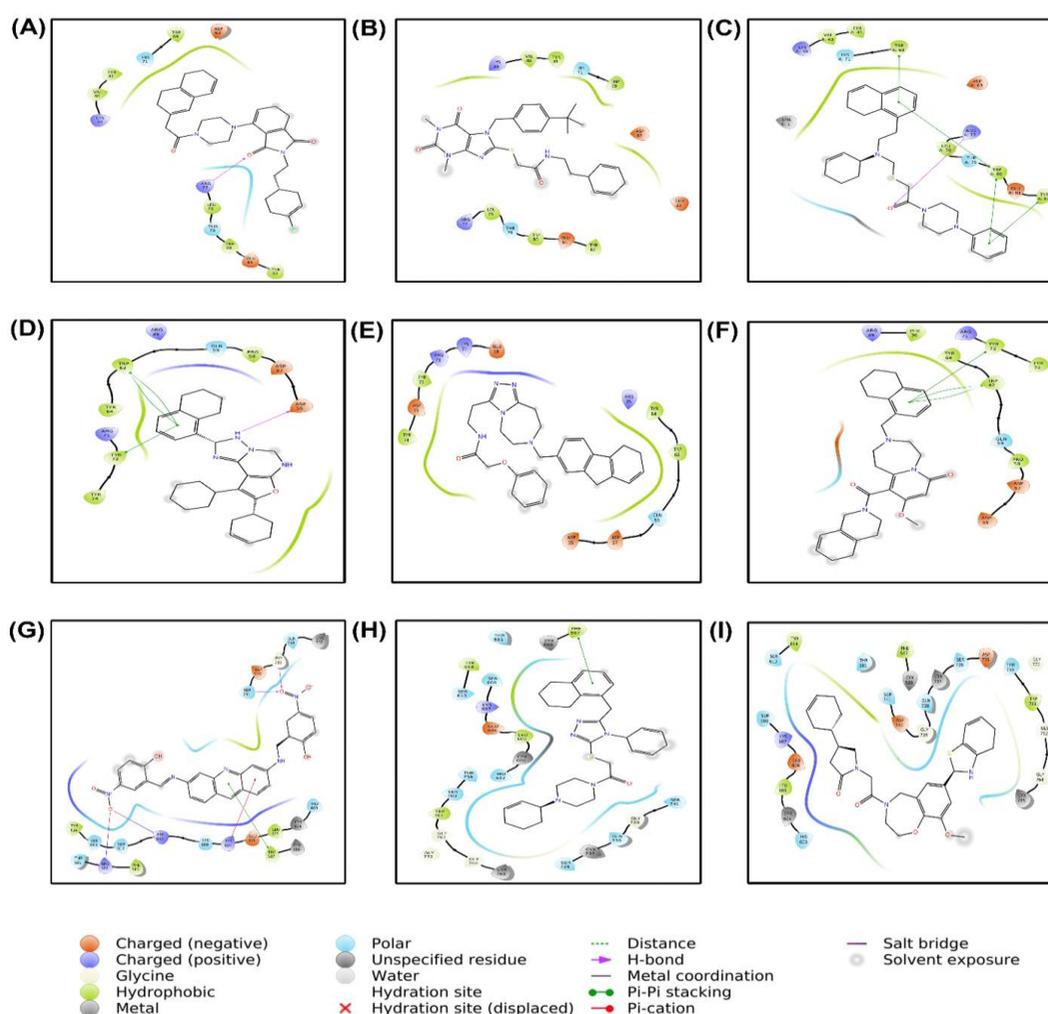


Figure 4-4: The 2D interaction profile of lead molecules into the binding site of their respective protein targets 1PK2, 4CIK and 5UGG (A) ZINC12211330 (P27), (B) ZINC00718625 (P67), (C) ZINC09970930 (P76), (D) ZINC02060288 (C19), (E) ZINC12455413 (C90), (F) ZINC14888376 (C97), (G) ZINC12576410 (U19), (H) ZINC04557820 (U94), (I) ZINC11839443 (U97)

Table 4-6: ADMET properties of top 3 identified ligands for each protein target. ADMET properties: Extension hepatotoxicity: <1 is nontoxic. CYP2D6: -ve is noninhibitors and +ve is inhibition, human intestinal absorption level: 0 (good); 1 (moderate); 2 (low); 3 (very low), (aqueous solubility): 0 (extremely low); 1 (low); 2 (good); 3 (optimal); 4 (too soluble), BBB (blood brain barrier): 0(very high); 1(high); 2 (mediums); 3 (low); 4 (undefined).

Protein	Ligand name	Solubility level	BBB Penetration level	Absorption level	Extension CYP2D6	Extension hepatotoxicity
1PK2	P27	1	3	0	-1.52 (false)	-11.08 (false)
	P67	2	3	1	-9.95 (false)	-18.01 (false)
	P76	2	3	1	-7.24 (false)	-6.26 (false)
	ACA	5	3	0	-4.03 (false)	-11.83 (false)
4CIK	C19	2	3	1	-0.13 (false)	-14.09 (false)
	C90	2	3	0	-4.45 (false)	-7.81 (false)
	C97	3	3	0	-1.8 (false)	-4.85 (false)
	XO3	4	3	0	-4.11 (false)	-6.21 (false)
5UGG	U19	2	4	2	-3.21 (false)	-7.68 (false)
	U94	1	3	1	-7.24 (false)	-5.11 (false)
	U97	2	4	0	-7.28 (false)	-6.51 (false)
	89M	2	4	0	-9.51 (false)	-5.81 (false)

Table 4-7: TOPKAT properties of top 3 identified ligands for each protein target. TOPKAT properties: NC -Non-carcinogenic, C-Carcinogenic, NT-Non-Toxic, T-Toxic, NI-Non-Irritant, I-Irritant, NM-Non-Mutagenic, M-Mutagenic

Protein	Ligand name	NTP carcinogenicity call (male rat) (v3.2)	NTP carcinogenicity call (female rat) (v3.2)	Developmental toxicity potential (DTP) (v3.1)	Skin irritation (v6.1)	Ames mutagenicity (v3.1)
1PK2	P27	NC	NC	NT	NI	NM
	P67	NC	NC	NT	NI	NM
	P76	NC	NC	NT	NI	NM
	ACA	NC	NC	T	NI	NM
4CIK	C19	NC	NC	NT	NI	NM
	C90	NC	NC	NT	NI	NM
	C97	NC	NC	NT	NI	NM
	XO3	C	C	T	NI	NM
5UGG	U19	NC	NC	NT	NI	NM
	U94	NC	NC	NT	NI	NM
	U97	NC	NC	NT	NI	NM
	89M	C	C	NT	I	M

3.3 Trajectory Analysis

Molecular Dynamics Simulation (MDS) plays a significant role in studying conformational changes of protein-ligand complexes and provides valuable insights into the prediction and identification of protein-ligand interactions [58]. The top nine ligands P27, P67, P76, C19, C90, C97, U19, U94, U97 docked with their respective protein targets i.e., plasminogen activator kringle-2 domain (1PK2), plasminogen kringle-1 domain (4CIK), and serine protease domain of plasminogen (5UGG) were subjected to molecular dynamics simulation in GROMACS for 200ns duration. Also, MD simulation for 200ns was performed for reference ligands i.e., ACA, XO3, and 89M docked with 1PK2, 4CIK, and 5UGG respectively. Additionally, we performed a simulation run of 200ns for all the three proteins 1PK2, 4CIK, and 5UGG respectively without ligand (APO). Thereafter, we performed trajectory analysis such as Root-mean-square-deviation (RMSD), Root-mean-square fluctuation (RMSF), Radius of gyration (Rg), number of Hydrogen bonds, bond distribution, Principal component analysis (PCA), and binding free-energy calculation (MMPBSA) for all the systems (standalone proteins, protein-reference ligand complex, and protein-identified ligand complex).

3.3.1 Root-mean-square deviation analysis

To elucidate the influence of ligands on the conformational stability of the protein, we have analyzed the root-mean-square deviation (RMSD) of backbone atoms of all the systems using the standard *g_rms* function of GROMACS for an overall time of 200 ns simulation run (Figure 4-5 A-C). As provided in Table 4-8, 1PK2 protein docked with the reference ligand ACA and the identified legends P27, P67, P76 shows an average deviation of 0.32 ± 0.04 nm, 0.45 ± 0.05 nm, 0.36 ± 0.04 nm, 0.4 ± 0.03 nm respectively. The 1PK2 APO protein exhibits an average deviation of 0.44 ± 0.05 nm. From the RMSD graphs, it can be inferred that all the systems are equilibrated and have converged within a range of 0.2-0.5nm (Figure 4-5A). Similarly, 4CIK protein docked with ligands X03 (reference ligand), C19, C90, C97 shows an average deviation of 0.29 ± 0.03 nm, 0.29 ± 0.03 nm, 0.3 ± 0.03 nm, 0.34 ± 0.04 nm respectively. The 4CIK APO protein exhibits an average deviation of 0.34 ± 0.06 nm. This suggests that all the

systems are equilibrated and have converged within a range of 0.2 - 0.4 nm (Figure 4-5B, Table 4-8). Likewise, 5UGG protein docked with ligands 89M (reference ligand), U19, U94, U97 exhibited an average deviation of 0.24 ± 0.02 nm, 0.29 ± 0.02 nm, 0.3 ± 0.02 nm, 0.31 ± 0.02 nm respectively. The 5UGG APO protein shows deviations of 0.24 ± 0.02 nm. This implies that all the systems that have converged within a range of 0.2-0.35 nm have attained equilibration (Figure 4-5C, Table 4-8). Although the RMSD profile shows an insignificant difference in the average value and minor fluctuations throughout, convergence is observed which indicates that our simulations have achieved stable trajectories. Hence, other conformational dynamics analysis is required to deduce further conclusions.

Table 4-8: RMSD values of all the protein-ligand complexes

RMSD		
Protein-Ligand Complex (30-200ns)	Average(nm)	SD -/+
1PK2-APO	0.44	0.05
1PK2-D27	0.45	0.05
1PK2-D67	0.36	0.04
1PK2-D76	0.4	0.03
1PK2-ACA	0.32	0.05
4CIK-APO	0.34	0.06
4CIK-D19	0.29	0.03
4CIK-D90	0.3	0.03
4CIK-D97	0.34	0.04
4CIK-XO3	0.31	0.03
5UGG-APO	0.27	0.02
5UGG-D19	0.29	0.02
5UGG-D94	0.3	0.02
5UGG-D97	0.31	0.02
5UGG-89M	0.24	0.02

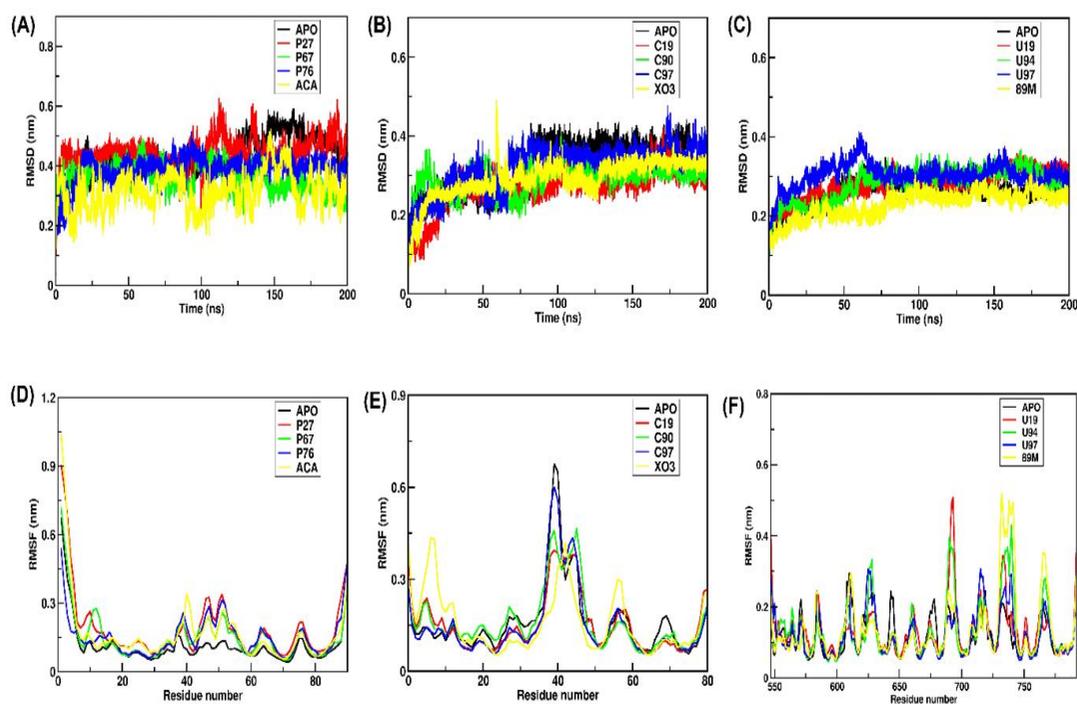


Figure 4-5: RMSD profile of all the systems: 1PK2, 4CIK, and 5UGG. (A) RMSD profile of 1PK2 protein-ligand complexes (top three identified ligands P27, P67, P76, and reference ligand ACA) along with APO. (B) RMSD profile of 4CIK protein-ligand complexes (top three identified ligands C19, C90, C97 and reference ligand XO3) along with APO. (C) RMSD profile of 5UGG protein-ligand complexes (top three identified ligands U19, U94, U97, and reference ligand 89m) along with APO. (D) RMSF profile of 1PK2 protein-ligand complexes (top three identified ligands P27, P67, P76, and reference ligand ACA) along with APO. (E) RMSF profile of 4CIK protein-ligand complexes (top three identified ligands C19, C90, C97 and reference ligand XO3) along with APO. (F) RMSF profile of 5UGG protein-ligand complexes (top three identified ligands U19, U94, U97, and reference ligand 89m) along with APO.

3.3.2 Residue flexibility analysis

To understand the overall residue-wise protein dynamics in the systems, the evaluation of root mean square fluctuations (RMSF) plays a very important role. Therefore, to check the stability and flexibility of the residues, we have calculated the RMSFs for an entire simulation run of 200ns using the *gmx_rmsf* module of GROMACS (Figure 4-5 D-F). From the RMSF profile, it can be understood that 1PK2 protein docked with ligands P27, P67, and P76 exhibited a similar pattern of fluctuation like the reference ligand ACA at active site residues TRP80, THR79, LEU78, ARG 71, TRP69, ASP63 respectively. 1PK2 docked with ligands P27, P67 and P76 show a fluctuation peak of

values 0.15nm, 0.16nm, 0.18nm at residue VAL40 respectively but a higher fluctuation peak of 0.36 nm is observed at residue VAL40 for the reference ligand. This indicates that the ligands P27, P67, and P76 are contributing to the greater rigidity of the binding pocket of 1PK2 (Figure 4-5D). Similarly, 4CIK protein docked with ligands C19, C90, and C97 exhibited a similar pattern of fluctuation like the reference ligand XO3 at active site residues ARG35, TRP62, TYR64, ARG71, TYR72 respectively. Likewise, 4CIK docked with ligands C19, C90, and C97 exhibited fluctuation of 0.15 nm at the residues ASP55 and ASP57 but in the case of reference ligand XO3, a higher fluctuation of 0.3 nm is observed. Also, XO3 shows a higher peak of 0.42nm at residues ASN9 and LYS8. This suggests that the identified ligands C19, C90, and C97 have made the binding pocket more rigid and intact (Figure 4-5E). In the same manner, 5UGG protein docked with ligands U19, U94, and U97 displayed the fluctuation pattern resembling the reference ligand 89M at active site residues LYS67, PHE68, CYX83, HID87, GLN239 respectively. Likewise, 5UGG docked with ligands U19, U94, and U97 exhibited fluctuation of 0.2nm, 0.2nm, and 0.28nm at the residues GLY762, SER765, CYS763 respectively, but in the case of reference ligand 89M, a higher fluctuation of 0.32 nm is observed. Also, 89M shows a higher peak of 0.5nm at residues TRP736, THR737, and GLY738. This propounds that the identified ligands U19, U94, and U97 have made the binding pocket more inflexible (Figure 4-5F). From the RMSF profile of all the systems, it can be inferred that the ligands P27, P67, P76, C19, C90, C97, U19, U94, U97 are responsible for the restricted dynamics at the active site residue pocket of their respective protein targets and for imparting structural rigidity.

3.3.3 Compactness analysis

To study the compactness of all the protein-ligand complexes, the radius of gyration (Rg) is computed using the *gmx_gyrate* function of GROMACS for a simulation time of 200ns. Rg is explained as the distance measured between the terminal end of protein and its center of mass during simulation. When a ligand docks to a protein, there is a conformational shift that changes the radius of gyration [59]. The compact protein structure tends to maintain a low Rg average deviation thus showing dynamic stability.

The combined Rg plot of all ligands is represented in Figure 4-6 [A-C]. As provided in Figure 4-6A and Table 4-9, 1PK2 protein docked with ligands ACA (reference ligand), P27, P67, P76 shows an average Rg deviation of 1.31 ± 0.02 nm, 1.29 ± 0.01 nm, 1.27 ± 0.01 nm, 1.28 ± 0.02 nm respectively from 30-200 ns simulation time. Similarly, 4CIK protein docked with ligands X03 (reference ligand), C19, C90, C97 shows an average Rg deviation of 1.26 ± 0.01 nm, 1.23 ± 0.01 nm, 1.22 ± 0.01 nm, 1.21 ± 0.01 nm respectively from 30-200 ns simulation time (Figure 4-6B and Table 4-9). Likewise, 5UGG protein docked with ligands 89M (reference ligand), U19, U94, U97 exhibited an average Rg deviation of 1.77 ± 0.01 nm, 1.74 ± 0.01 nm, 1.75 ± 0.01 nm, 1.73 ± 0.01 nm respectively from 30-200 ns simulation time (Figure 4-6C and Table 4-9). Collectively, from the radius of gyration profile, it can be concluded that the identified ligands P27, P67, P76, C19, C90, C97, U19, U94, U97 are responsible for making the structure more compact compared to the reference ligands ACA, X03, and 89M respectively.

Table 4-9: Rg values of all the protein-ligand complexes.

Radius of gyration		
Protein-Ligand Complex	Average(nm)	SD -/+
(30-200ns)		
1PK2-APO	1.26	0.02
1PK2-D27	1.29	0.01
1PK2-D67	1.27	0.01
1PK2-D76	1.28	0.02
1PK2-ACA	1.31	0.02
4CIK-APO	1.2	0.01
4CIK-D19	1.23	0.01
4CIK-D90	1.22	0.01
4CIK-D97	1.21	0.01
4CIK-XO3	1.26	0.01
5UGG-APO	1.74	0.01
5UGG-D19	1.74	0.01
5UGG-D94	1.75	0.01
5UGG-D97	1.73	0.01
5UGG-89M	1.77	0.01

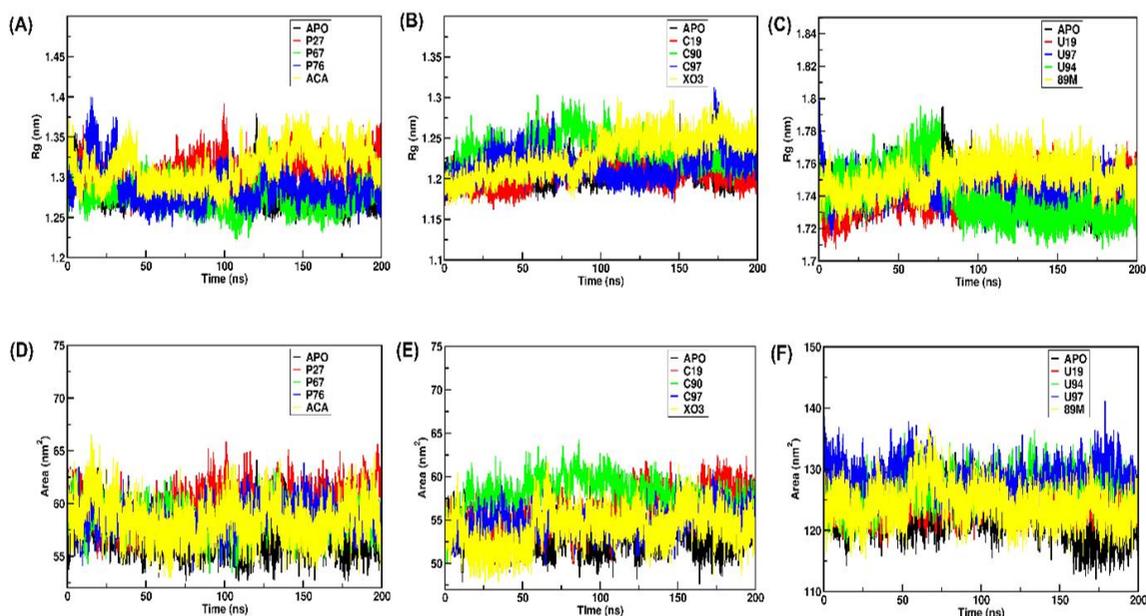


Figure 4-6: Radius of gyration profile of all the systems: 1PK2, 4CIK, and 5UGG (A) Radius of gyration profile of 1PK2 protein-ligand complexes (top three identified ligands P27, P67, P76, and reference ligand ACA) along with APO. (B) The radius of gyration profile of 4CIK protein-ligand complexes (top three identified ligands C19, C90, C97 and reference ligand XO3) along with APO. (C) The radius of gyration profile of 5UGG protein-ligand complexes (top three identified ligands U19, U94, U97, and reference ligand 89m) along with APO. (D) SASA profile of 1PK2 protein-ligand complexes (top three identified ligands P27, P67, P76, and reference ligand ACA) along with APO. (E) SASA profile of 4CIK protein-ligand complexes (top three identified ligands C19, C90, C97 and reference ligand XO3) along with APO. (F) SASA profile of 5UGG protein-ligand complexes top three identified ligands U19, U94, U97 and reference ligand 89m) along with APO.

3.3.4 Solvent Accessible Surface Area analysis

To calculate the solvent-accessible surface area (SASA) of a molecular surface, a method is employed which involves the *in-silico* rolling of a spherical probe that approximates a water molecule, encompassing a full-atom protein model [60]. This procedure takes into account the expansion of the van der Waals radius for each atom by 1.4 Å (the radius of a polar solvent probe) followed by the calculation of the surface area of these atoms with expanded radius. Therefore, the SASA values of all the complexes were computed using the *gmx_sasa* function of GROMACS for a simulation time of 200ns (Figure 4-6 D-F).

As provided in Figure 4-6D and Table 4-10, 1PK2 protein docked with ligands ACA (reference ligand), P27, P67, P76 shows an average SASA value of $58.94 \pm 1.7 \text{ nm}^2$, $58.67 \pm 1.83 \text{ nm}^2$, $58.34 \pm 1.54 \text{ nm}^2$, $58.76 \pm 1.53 \text{ nm}^2$ respectively from 30-200 ns simulation timeframe. It is observed that the SASA curves followed the same trend as the reference ligand ACA. Similarly, 4CIK protein docked with ligands X03 (reference ligand), C19, C90, C97 shows an average SASA value of $54.46 \pm 2.0 \text{ nm}^2$, $54.18 \pm 2.3 \text{ nm}^2$, $54.28 \pm 1.72 \text{ nm}^2$, $53.21 \pm 1.7 \text{ nm}^2$ respectively from 30-200 ns simulation time frame (Figure 4-6E and Table 4-10). The C19, C90, and C97 show a similar pattern of average SASA values when compared to X03. Likewise, 5UGG protein docked with ligands 89M (reference ligand), U19, U94, U97 exhibited an average SASA value of $124.75 \pm 2.85 \text{ nm}^2$, $123.61 \pm 2.36 \text{ nm}^2$, $124.32 \pm 2.62 \text{ nm}^2$, $124.24 \pm 2.84 \text{ nm}^2$ respectively from 30-200 ns simulation time frame (Figure 4-6F and Table 4-10). The same pattern of SASA values was obtained for all the identified ligands.

Table 4-10: SASA values of all the protein-ligand complexes.

SASA		
Protein-Ligand Complex	Average(nm)	SD +/-
(30-200ns)		
1PK2-APO	56.9	1.86
1PK2-D27	58.67	1.83
1PK2-D67	58.34	1.54
1PK2-D76	58.76	1.53
1PK2-ACA	58.94	1.7
4CIK-APO	53.01	1.56
4CIK-D19	54.18	2.3
4CIK-D90	54.28	1.72
4CIK-D97	53.21	1.7
4CIK-XO3	54.46	2
5UGG-APO	122.63	3.2
5UGG-D19	123.61	2.36
5UGG-D94	124.32	2.62
5UGG-D97	124.24	2.84
5UGG-89M	124.75	2.85

It is observed that similar to reference ligands, the identified ligands make the active site residues of the protein targets readily accessible to the solvent surface, thus exposing the proteins to the hydration shell. Also, the SASA values obtained from the analysis of proteins docked with our identified ligands are found to be significantly similar to the reference protein-ligand complexes.

3.3.5 Hydrogen Bond and Bond distribution analysis

The hydrogen bonds (H-bonds) are known to play a key role in forming stable contacts between the ligands and protein target [61]. In this study, H-bond analysis was performed on all the protein-ligand systems for a total simulation run of 200ns. The number of H-bonds and their bond frequencies are plotted using the *gmx_hbond* tool of GROMACS and are represented in Figure 4-7 A-F. It is observed that the ligand P67 formed one stable H-bond of length 0.27 nm with the protein target 1PK2, and its frequency to form this bond is 20 % (Figure 4-7A, 4-7D). Likewise, the P76 formed 2-3 H-bonds during the simulation period and its frequency to form a bond of length of 0.28 nm is 17.5% (Figure 4-7A, 4-7D). Similarly, P27 had formed a bond of length 0.29 nm with a frequency of 17% (Figure 4-7A, 4-7D). The reference ligand ACA did not form stable H-bonds and the frequency to form a bond of length of 0.28 nm is only 10% which is very less compared to identified ligands (Figure 4-7A, 4-7D). In the case of the 4CIK protein target, the ligand C97 established stable 1-2 H-bonds during the simulation period and had a 23% frequency to form a bond of length 0.28nm (Figure 4-7B, 4-7E). In the same way, the C90 exhibited 1-2 H-bonds during the simulation period, and its ability to form a bond of length 0.27 is 18% (Figure 4-7B, 4-7E). Similarly, the ligand C19 formed 1 H-bond, and its frequency to form a bond of length of 0.28 nm is 18% (Figure 4-7B, 4-7E). Lastly, we observed that the reference ligand XO3 formed unstable H-bonds during the simulation, and its frequency to form a bond of length 0.26 nm is only 10%, which is less compared to identified ligands (Figure 4-7B, 4-7E).

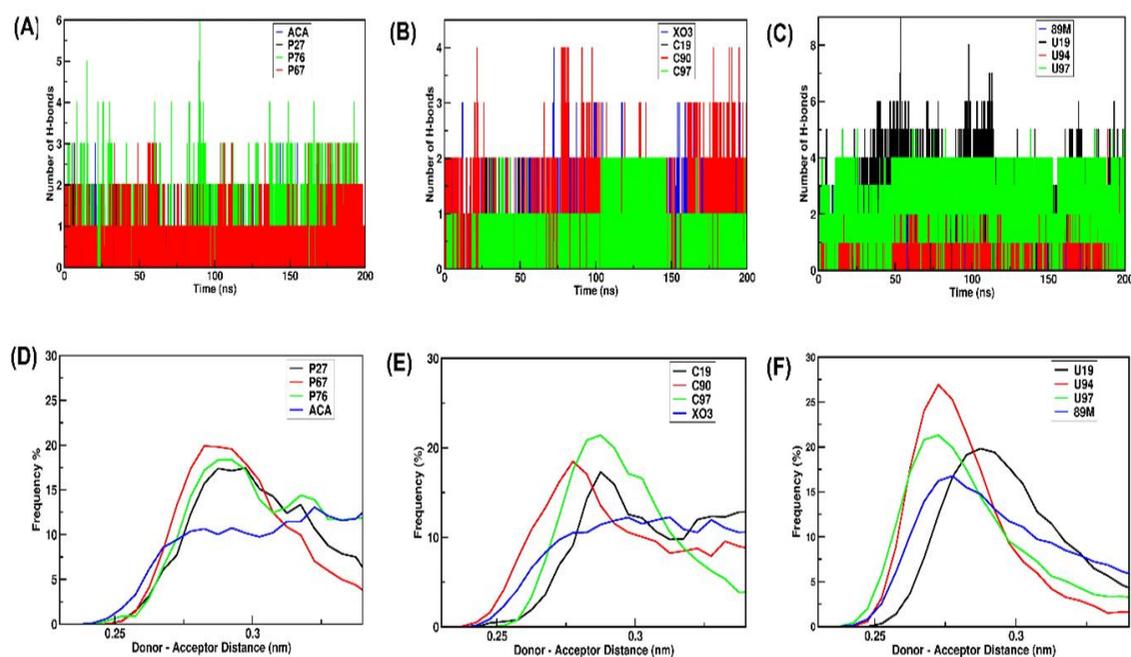


Figure 4-7: Hydrogen bonding profile of all the systems: 1PK2, 4CIK, and 5UGG. (A) Hydrogen bonding pattern of 1PK2 protein-ligand complexes (top three identified ligands P27, P67, P76, and reference ligand ACA) along with APO. (B) Hydrogen bonding pattern of 4CIK protein-ligand complexes (top three identified ligands C19, C90, C97 and reference ligand XO3) along with APO. (C) Hydrogen bonding pattern of 5UGG protein-ligand complexes (top three identified ligands U19, U94, U97, and reference ligand 89m) along with APO. (D) H-Bond frequencies of 1PK2 protein-ligand complexes (top three identified ligands P27, P67, P76, and reference ligand ACA) along with APO. (E) H-Bond frequencies of 4CIK protein-ligand complexes (top three identified ligands C19, C90, C97, and reference ligand XO3) along with APO. (F) H-Bond frequencies of 5UGG protein-ligand complexes top three identified ligands U19, U94, U97 and reference ligand 89m) along with APO.

Likewise, in the case of the 5UGG protein target, the ligand U97 formed 1-2 stable H-bonds during the simulation time, and its frequency to form a bond of length 0.27 nm is 22% (Figure 4-7C, 4-7F). The U94 ligand formed 2-4 stable H-bonds and the frequency to form a bond of length 0.27 nm is 27% (Figure 4-7C, 4-7F). Also, the ligand U19 formed 2-4 H-bonds, and its frequency to form a bond of length of 0.28 nm is 20% (Figure 4-7C, 4-7F). Lastly, the reference ligand 89M formed 1-2 unstable H-bonds, and the frequency to form a bond of length 0.27 nm is 15% (Figure 4-7C, 4-7F). Overall, from the H-bond analysis, it can be concluded that our identified ligands form stable and stronger H-bonds of length ≥ 0.28 nm with their target proteins compared to reference ligands.

3.3.6 Principal component analysis

In general, a protein-ligand complex that occupies a smaller phase space with a stable cluster denotes a highly stable complex [62]. Therefore, to explore this phenomenon, we have performed Principal Component Analysis (PCA) on the trajectories obtained from the 200ns simulation period for all the protein-ligand complexes as shown in the 2D plots of PCA (4-8 A-C). From the PCA graph of 1PK2, it is observed that the complexes formed by P27, P67, and P76 with 1PK2 are clustered very closely and the area coverage is less compared to the reference ligand ACA (Figure 4-8A). In the same way, from the PCA graph of 4CIK, it is observed that the complexes formed by C19, C90, and C97 with 4CIK are closely clustered and occupy less area compared to reference ligand XO3 (Figure 4-8B). Similarly, from the PCA graph of 5UGG, it is observed that the complexes formed by U19, U94, and U97 with 5UGG are clustered and occupy small phase space compared to reference ligand 89M (Figure 4-8C). Thus, from the PCA analysis, it is implied that our identified ligands impart rigidity to their respective protein-ligand complexes by occupying smaller phase space with stable cluster formation compared to reference ligands.

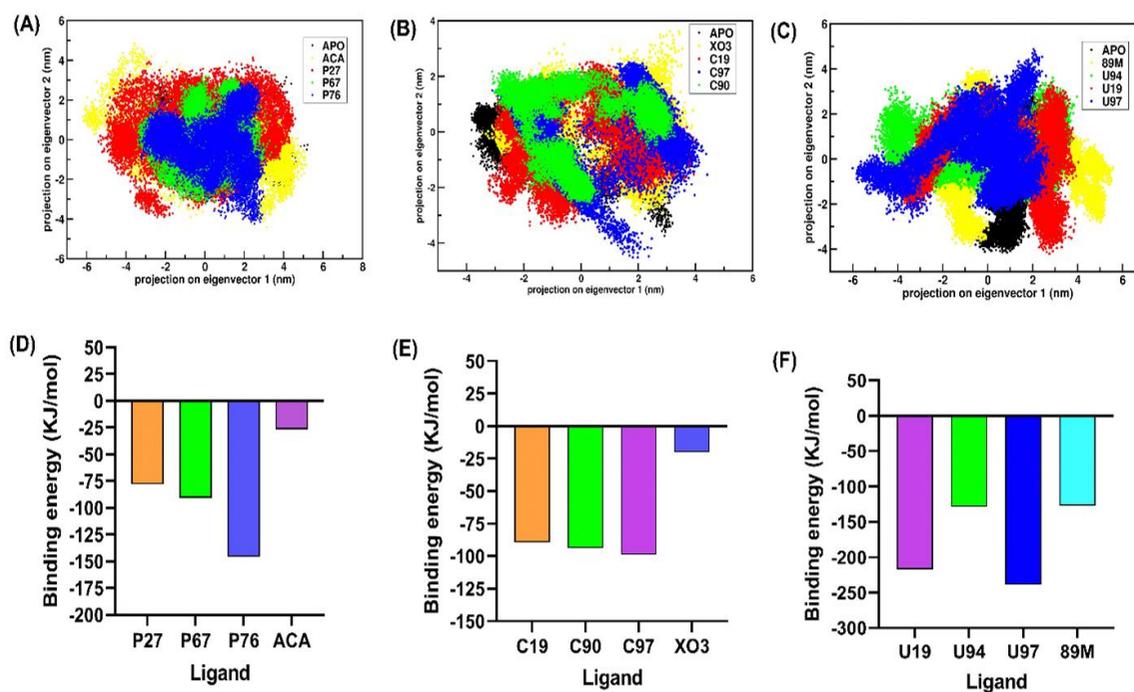


Figure 4-8: Principal component analysis and MMPBSA analysis of all the systems: 1PK2, 4CIK, and 5UGG. (A) Projection of motion of Apo-1PK2 in black; 1PK2-ACA complex in yellow; 1PK2-P27 complex in red; 1PK2-P67 in green; 1PK2-P76 in blue. (B) Projection of motion of Apo-4CIK in black; 4CIK-XO3 complex in yellow; 4CIK -C19 complex in red; 4CIK -C90 in

green; 4CIK-C97 in blue. (C) Projection of motion of Apo-5UGG in black; 5UGG-89M complex in yellow; 5UGG -U19 complex in red; 5UGG -U94 in green; 5UGG -U97 in blue. (D) Binding energy profile (ΔG) of 1PK2 complexed with the identified ligands P27, P67, P76, and reference ligand ACA. (E) Binding energy profile (ΔG) of 4CIK complexed with the identified ligands C19, C90, C97, and reference ligand XO3. (F) Binding energy profile (ΔG) of 5UGG complexed with the identified ligands U19, U94, U97, and reference ligand 89M.

3.3.7 Binding free energy calculation

The results obtained by calculation of the free energy of ligands give insights into ligands' binding potential with bonded and non-bonded entities [63]. In this study, the *g_mmpbsa* tool was used to calculate binding free energy (ΔG) for the last 50ns of the simulation time (Figure 4-8 D-F). From the MMPBSA graphs, it is observed that the reference ligand ACA, and the identified ligands P27, P67, P76 bound to the 1PK2 target protein exhibit binding energy of -26.460 ± 9.222 kJ/mol and -77.975 ± 19.181 kJ/mol, -90.798 ± 37.674 kJ/mol, -145.653 ± 20.078 kJ/mol respectively (Figure 4-8D). Thus, the binding free energies (ΔG) of 1PK2 bound with our identified ligands were found to be significantly better than that of 1PK2 bound with the reference ligand. Additionally, the four individual energy components such as van der Waal energy (ΔE_{vdw}), electrostatic energy (ΔE_{elec}), polar solvation energy ($\Delta G_{polar-solv}$), and SASA energy ($\Delta G_{non-polar}$) were compared for all the four protein-ligand complexes i.e., 1PK2-ACA, 1PK2-P27, 1PK2-67 and 1PK2-P76. The details of the MMPBSA energies of all the protein-ligand complexes are provided in Table 4-11. The individual energy components van der Waal energy (ΔE_{vdw}), electrostatic energy (ΔE_{elec}), polar solvation energy ($\Delta G_{polar-solv}$) and SASA energy ($\Delta G_{non-polar}$) was found to be -40.822 ± 9.960 kJ/mol, -5.951 ± 6.304 kJ/mol, 27.127 ± 13.044 kJ/mol, -6.815 ± 1.117 kJ/mol in 1PK2-ACA complex, -118.529 ± 17.130 kJ/mol, -5.881 ± 6.590 kJ/mol, 61.449 ± 21.931 kJ/mol, -15.014 ± 1.943 kJ/mol in 1PK2-27 complex, -117.567 ± 47.692 kJ/mol, -12.887 ± 9.387 kJ/mol, 51.918 ± 28.232 kJ/mol, -12.262 ± 5.059 kJ/mol in 1PK2-67 complex and -171.612 ± 24.161 kJ/mol, -0.102 ± 3.913 kJ/mol, 42.307 ± 10.310 kJ/mol, -16.246 ± 2.225 kJ/mol kJ/mol in 1PK2-76 complex respectively. Similarly, from the MMPBSA graphs, it is observed that the reference ligand XO3 and

the identified ligands C19, C90, C97 bound to the 4CIK protein target exhibit binding energy of -20.066 +/- 30.373 kJ/mol and -89.316 +/- 38.412 kJ/mol, -93.881 +/- 39.414 kJ/mol, -98.949 +/- 32.293 kJ/mol respectively (Figure 4-8E). The details of the energy components are provided in Table 4-11 for all the four protein-ligand complexes i.e.,

Table 4-11: MMPBSA detailed energies (Binding energy and the individual energy components van der Waal energy (ΔE_{vdw}), electrostatic energy (ΔE_{elec}), polar solvation energy ($\Delta G_{polar-solv}$), SASA energy ($\Delta G_{non-polar}$) of all the protein-ligand complexes.

MMPBSA DETAILED ENERGIES						
Protein target	Ligands	van der Waal energy	Electrostatic energy	Polar solvation energy	SASA energy	Binding energy
1PK2	P27	-118.529 +/- 17.130 kJ/mol	-5.881 +/- 6.590 kJ/mol	61.449 +/- 21.931 kJ/mol	-15.014 +/- 1.943 kJ/mol	-77.975 +/- 19.181 kJ/mol
	P67	-117.567 +/- 47.692 kJ/mol	-12.887 +/- 9.387 kJ/mol	51.918 +/- 28.232 kJ/mol	-12.262 +/- 5.059 kJ/mol	-90.798 +/- 37.674 kJ/mol
	P76	-171.612 +/- 24.161 kJ/mol	-0.102 +/- 3.913 kJ/mol	42.307 +/- 10.310 kJ/mol	-16.246 +/- 2.225 kJ/mol	-145.653 +/- 20.078 kJ/mol
	ACA	-40.822 +/- 9.960 kJ/mol	-5.951 +/- 6.304 kJ/mol	27.127 +/- 13.044 kJ/mol	-6.815 +/- 1.117 kJ/mol	-26.460 +/- 9.222 kJ/mol
4CIK	C19	-178.135 +/- 49.112 kJ/mol	-19.139 +/- 11.123 kJ/mol	126.667 +/- 27.807 kJ/mol	-18.703 +/- 4.933 kJ/mol	-89.31 +/- 39.414 kJ/mol
	C90	-178.135 +/- 49.112 kJ/mol	-8.939 +/- 11.123 kJ/mol	111.896 +/- 27.807 kJ/mol	-18.703 +/- 4.933 kJ/mol	-93.881 +/- 39.414 kJ/mol
	C97	-128.666 +/- 35.350 kJ/mol	0.790 +/- 3.810 kJ/mol	41.817 +/- 25.156 kJ/mol	-12.890 +/- 3.743 kJ/mol	-98.949 +/- 32.293 kJ/mol

MMPBSA DETAILED ENERGIES						
Protein target	Ligands	van der Waal energy	Electrostatic energy	Polar solvation energy	SASA energy	Binding energy
	XO3	-45.470 +/- 7.831 kJ/mol	-6.123 +/- 7.523 kJ/mol	38.916 +/- 27.368 kJ/mol	-7.388 +/- 1.775 kJ/mol	-20.066 +/- 30.373 kJ/mol
5UGG	U19	-326.449 +/- 12.469 kJ/mol	-28.751 +/- 14.338 kJ/mol	164.799 +/- 11.051 kJ/mol	-26.749 +/- 0.969 kJ/mol	-217.150 +/- 16.376 kJ/mol
	U94	-193.346 +/- 18.772 kJ/mol	-24.924 +/- 13.795 kJ/mol	106.994 +/- 23.339 kJ/mol	-17.398 +/- 1.411 kJ/mol	-128.674 +/- 20.963 kJ/mol
	U97	-332.423 +/- 12.146 kJ/mol	-28.614 +/- 7.827 kJ/mol	148.128 +/- 10.778 kJ/mol	-25.937 +/- 1.160 kJ/mol	-238.846 +/- 17.788 kJ/mol
	89M	-199.690 +/- 9.099 kJ/mol	-30.462 +/- 10.217 kJ/mol	126.314 +/- 23.798 kJ/mol	-23.143 +/- 1.662 kJ/mol	-126.981 +/- 19.836 kJ/mol

4CIK-XO3, 4CIK-C19, 4CIK-C90, and 4CIK-97. The ΔE_{vdW} component, the ΔE_{elec} component, the $\Delta G_{\text{polar-solv}}$ component and the $\Delta G_{\text{non-polar}}$ component was found to be -45.470 +/- 7.831 kJ/mol, -6.123 +/- 7.523 kJ/mol, 38.916 +/- 27.368 kJ/mol, -7.388 +/- 1.775 kJ/mol in 4CIK-XO3 complex, -178.135 +/- 49.112 kJ/mol, -19.139 +/- 11.123 kJ/mol, 126.667 +/- 27.807 kJ/mol, -18.703 +/- 4.933 kJ/mol in 4CIK-C19 complex, -178.135 +/- 49.112 kJ/mol, -8.939 +/- 11.123 kJ/mol, 111.896 +/- 27.807 kJ/mol, -18.703 +/- 4.933 kJ/mol in 4CIK-C90 complex, -128.666 +/- 35.350 kJ/mol, 0.790 +/- 3.810 kJ/mol, 41.817 +/- 25.156 kJ/mol, -12.890 +/- 3.743 kJ/mol in 4CIK-97 complex respectively. Likewise, from the MMPBSA graphs, it is observed that the reference ligand 89M and the identified ligands U19, U94, U97 bound to the 5UGG protein target, exhibits binding energy of -126.981 +/- 19.836 kJ/mol, and -217.150 +/- 16.376 kJ/mol, -168.674 +/- 20.963 kJ/mol, -238.846 +/- 17.788 kJ/mol respectively (Figure 4-8F). The ΔE_{vdW} component, the ΔE_{elec} component, the $\Delta G_{\text{polar-solv}}$ component and the $\Delta G_{\text{non-polar}}$ component was found to be -199.690 +/- 9.099 kJ/mol, -30.462 +/-

10.217 kJ/mol, 126.314 +/- 23.798 kJ/mol, -23.143 +/- 1.662 kJ/mol in 5UGG-89M complex, -326.449 +/- 12.469 kJ/mol, -28.751 +/- 14.338 kJ/mol, 164.799 +/- 11.051 kJ/mol, -26.749 +/- 0.969 kJ/mol in 5UGG-U19 complex, -193.346 +/- 18.772 kJ/mol. -24.924 +/- 13.795 kJ/mol, 106.994 +/- 23.339 kJ/mol, -17.398 +/- 1.411 kJ/mol in 5UGG-U94 complex, -332.423 +/- 12.146 kJ/mol, -28.614 +/- 7.827 kJ/mol, 148.128 +/- 10.778 kJ/mol, -25.937 +/- 1.160 kJ/mol in 5UGG-U97 complex respectively. The 3D images of the top nine identified ligands P27, P67, P76, C19, C90, C97, U19, U94, U97 bound to the active site residues of their respective protein targets 1PK2, 4CIK, and 5UGG are shown in Figure 4-9 [(A)-(I)]. Furthermore, from the binding energy calculation, it can also be concluded that out of the nine identified ligands, the top three ligands having the best binding energy are P76, C97, and U97 against their protein targets 1PK2, 4CIK, and 5UGG respectively.

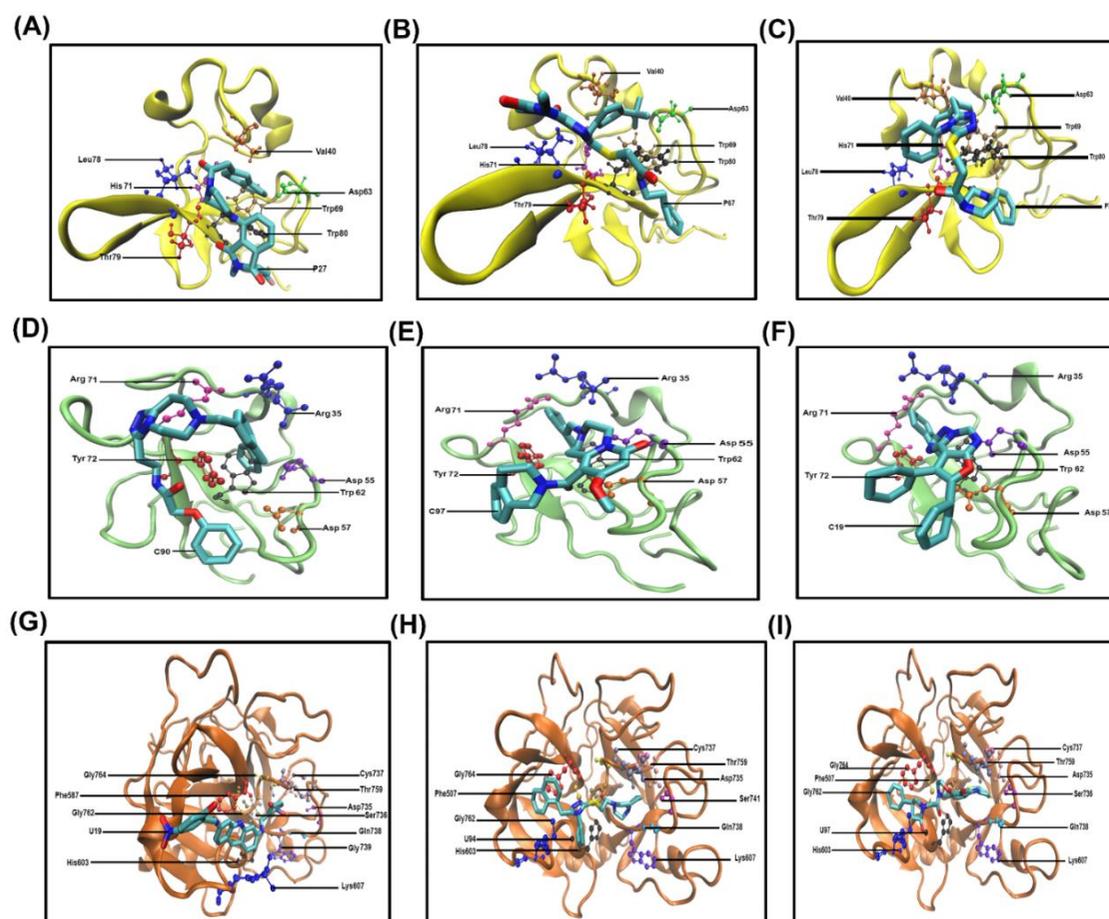


Figure 4-9: Schematic representation of the top three identified ligands docked to the active site residues of their respective protein targets using VMD. (A) P76 bound to 1PK2 (B) C97 bound to 4CIK (C) U97 bound to 5UGG. Proteins are shown in the New Cartoon representation. Active site residues of the proteins are shown in licorice representation. Identified ligands are shown in CPK representation.

From the MMPBSA profile, it can be inferred that the identified ligands docked with their respective target protein have the highest electrostatic energy, SASA energy, and Vander Waal energy in comparison with reference ligands. This also indicates that despite intermittent hydrogen bond formations as observed during h-bond analysis, the ligands were conserved in the binding pocket by nonbonded interactions (electrostatic, polar, and nonpolar). Based on overall MD analysis and MMPBSA energies, we have found the ligands P76, C97, and U97 to be the best performing amongst the top nine ligands.

4. Conclusions

In this study, we have identified compounds that can inhibit fibrinolysis by targeting each of the three important protein domains: the kringle-2 domain of tissue-type plasminogen activator, the kringle-1 domain of plasminogen, and the serine protease domain of plasminogen respectively. The top nine ligands (P27, P67, P76, C19, C90, C97, U19, U94, U97) were obtained by screening one million molecules from the Zinc database. These ligands were docked to their respective protein targets (1PK2, 4CIK, and 5UGG) using Autodock-Vina, Schrodinger-Glide, and ParDOCK/BAPPL+. Docking scores of these top nine ligands were evaluated and their ADMET profiling was also performed keeping the reported ligands (ACA, XO3, and 89M) as reference. All the top nine ligands, with the best scores and good ADMET results, were individually subjected to MD simulation for a period of 200 ns followed by detailed trajectory analysis. RMSD profile suggests that all the protein-ligand complexes have attained stable trajectories and are well equilibrated. From the RMSF analysis of all the systems, it can be inferred that the identified ligands are responsible for the restricted dynamics at the active site residue pocket of their respective protein targets, thereby imparting structural rigidity to the protein compared to the reference ligands. Rg data of all the systems reveals that in comparison to reference ligands, our identified ligands contribute greater compactness to the structure of their respective protein targets. From the SASA analysis, it is evident that similar to reference ligands, the identified ligands make the active site residues of the protein targets well exposed to the solvent, thereby making them readily accessible. From the h-bond analysis and bond distribution, it was observed that the identified ligands tend to form stronger H-bonds of length ≥ 0.28 nm with their target proteins compared to reference ligands. PCA analysis helps us to

understand that in comparison to reference ligands our identified ligands occupy smaller phase space and form stable clusters, thus conferring more stability to the protein-ligand complexes. The results obtained by analyzing molecular docking and MD simulation data are well-validated by the binding free energy calculation. From the MMPBSA profile, it is observed that, in the identified protein-ligand complexes, there exist strong interactions in terms of energy components other than hydrogen bonding such as electrostatic energy, SASA energy, and Vander Waal energy compared to reference protein-ligand complexes. Based on overall MD analysis and MMPBSA energies, we have found the ligands P76, C97, and U97 to be the best performing amongst the top nine ligands. Hence, our study strongly substantiates the point that our identified ligands can act as strong inhibitors and further experimental studies can be conducted to prove the therapeutic potential of these promising antifibrinolytic agents.

References:

1. Chapin JC, Hajjar KA. Fibrinolysis and the control of blood coagulation. *Blood reviews*. 2015 Jan 1;29(1):17-24.
2. Harpel PC, Chang TS, Verderber E. Tissue plasminogen activator and urokinase mediate the binding of Glu-plasminogen to plasma fibrin I. Evidence for new binding sites in plasmin-degraded fibrin I. *Journal of Biological Chemistry*. 1985 Apr 10;260(7):4432-40.
3. Fears R. Binding of plasminogen activators to fibrin: characterization and pharmacological consequences. *Biochemical Journal*. 1989 Jul 15;261(2):313.
4. Novokhatny VV, Ingham KC, Medved LV. Domain structure and domain-domain interactions of recombinant tissue plasminogen activator. *Journal of Biological Chemistry*. 1991 Jul 15;266(20):12994-3002.
5. Verheijen JH, Caspers MP, Chang GT, De Munk GA, Pouwels PH, Enger- Valk BE. Involvement of finger domain and kringle-2 domain of tissue-type plasminogen activator in fibrin binding and stimulation of activity by fibrin. *The EMBO journal*. 1986 Dec;5(13):3525-30.
6. Korhonen TK, Haiko J, Laakkonen L, Järvinen HM, Westerlund-Wikström B. Fibrinolytic and coagulative activities of *Yersinia pestis*. *Frontiers in cellular and infection microbiology*. 2013 Jul 26;3:35.
7. Lijnen, Henri R., Marc Hoylaerts, and Desire Collen. "Isolation and characterization of a human plasma protein with affinity for the lysine binding sites in plasminogen. Role in the regulation of fibrinolysis and identification as histidine-rich glycoprotein." *Journal of biological chemistry* 255.21 (1980): 10214-10222.
8. Ayón-Núñez DA, Fragoso G, Bobes RJ, Laclette JP. Plasminogen-binding proteins as an evasion mechanism of the host's innate immunity in infectious diseases. *Bioscience Reports*. 2018 Oct 31;38(5).
9. Colomina MJ, Méndez E, Sabate A. Altered fibrinolysis during and after surgery. *In Seminars in Thrombosis and Hemostasis* 2021 Apr 20. Thieme Medical Publishers, Inc.

10. McCormack PL. Tranexamic acid. *Drugs*. 2012 Mar;72(5):585-617.
11. Nishida T, Kinoshita T, Yamakawa K. Tranexamic acid and trauma-induced coagulopathy. *Journal of Intensive Care*. 2017 Dec;5(1):1-7.
12. Roberts I, Shakur H, Coats T, Hunt B, Balogun E, Barnetson L, Cook L, Kawahara T, Perel P, Prieto-Merino D, Ramos M. The CRASH-2 trial: a randomised controlled trial and economic evaluation of the effects of tranexamic acid on death, vascular occlusive events and transfusion requirement in bleeding trauma patients. *Health Technol Assess*. 2013 Mar 1;17(10):1-79.
13. Hunter GR, Young GB. Seizures after cardiac surgery. *Journal of Cardiothoracic and Vascular Anesthesia*. 2011 Apr 1;25(2):299-305.
14. Ockerman A, Vanassche T, Garip M, Vandenbrielle C, Engelen MM, Martens J, Politis C, Jacobs R, Verhamme P. Tranexamic acid for the prevention and treatment of bleeding in surgery, trauma and bleeding disorders: a narrative review. *Thrombosis Journal*. 2021 Dec;19(1):1-6.
15. Mortensen M, Krall J, Kongstad KT, Brygger BM, Lenzi O, Francotte P, Sørensen TE, Nielsen B, Jensen AA, Smart TG, Frølund B. Developing New 4-PIOL and 4-PHP Analogues for Photoinactivation of γ -Aminobutyric Acid Type A Receptors. *ACS Chemical Neuroscience*. 2019 Oct 7;10(11):4669-84.
16. Law RH, Wu G, Leung EW, Hidaka K, Quek AJ, Caradoc-Davies TT, Jeevarajah D, Conroy PJ, Kirby NM, Norton RS, Tsuda Y. X-ray crystal structure of plasmin with tranexamic acid-derived active site inhibitors. *Blood advances*. 2017 May 9;1(12):766-71.
17. Wiman B, Collen D. Molecular mechanism of physiological fibrinolysis. *Nature*. 1978 Apr;272(5653):549-50.
18. Van Zonneveld AJ, Veerman H, Pannekoek H. On the interaction of the finger and the kringle-2 domain of tissue-type plasminogen activator with fibrin. Inhibition of kringle-2 binding to fibrin by epsilon-amino caproic acid. *Journal of Biological Chemistry*. 1986 Oct 25;261(30):14214-8.
19. Castellino FJ, McCance SG. The kringle domains of human plasminogen. From Plasminogen-related growth factors. 1997 Jan 1.

20. Keragala CB, Medcalf RL. Plasminogen: an enigmatic zymogen. *Blood, The Journal of the American Society of Hematology*. 2021 May 27;137(21):2881-9.
21. Lionta E, Spyrou G, K Vassilatis D, Cournia Z. Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Current topics in medicinal chemistry*. 2014 Aug 1;14(16):1923-38.
22. Byeon IJ, Llinás M. Solution structure of the tissue-type plasminogen activator kringle-2 domain complexed to 6-aminohexanoic acid an antifibrinolytic drug. *Journal of molecular biology*. 1991 Dec 20;222(4):1035-51.
23. Cheng L, Pettersen D, Ohlsson B, Schell P, Karle M, Evertsson E, Pahlén S, Jonforsen M, Plowright AT, Boström J, Fex T. Discovery of the fibrinolysis inhibitor AZD6564, acting via interference of a protein–protein interaction. *ACS medicinal chemistry letters*. 2014 May 8;5(5):538-43.
24. Law RH, Wu G, Leung EW, Hidaka K, Quek AJ, Caradoc-Davies TT, Jeevarajah D, Conroy PJ, Kirby NM, Norton RS, Tsuda Y. X-ray crystal structure of plasmin with tranexamic acid–derived active site inhibitors. *Blood advances*. 2017 May 9;1(12):766-71.
25. Bank PD. Protein data bank. *Nature New Biol*. 1971; 233:223.
26. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
27. Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Current protocols in bioinformatics*. 2016 Jun;54(1):5-6.
28. Systemes, D. BIOVIA. "Discovery Studio Modeling Environment." *Dassault Systèmes, San Diego* (2016).
29. Brooks BR, Brooks III CL, Mackerell Jr AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A. CHARMM: the biomolecular simulation program. *Journal of computational chemistry*. 2009 Jul 30;30(10):1545-614.
30. Sterling T, Irwin JJ. ZINC 15–ligand discovery for everyone. *Journal of chemical information and modeling*. 2015 Nov 23;55(11):2324-37.

31. Mukherjee G, Jayaram B. A rapid identification of hit molecules for target proteins via physico-chemical descriptors. *Physical Chemistry Chemical Physics*. 2013;15(23):9107-16.
32. Trott, O., & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2), 455-461. (2010)
33. Glide version 6.2. (2019). New York, NY: Schrodinger, LLC.
34. Gupta, A., et al. "ParDOCK: an all atom energy based Monte Carlo docking protocol for protein-ligand complexes." *Protein and peptide letters* 14.7 (2007): 632-646.
35. Soni A, Bhat R, Jayaram B. Improving the binding affinity estimations of protein–ligand complexes using machine-learning facilitated force field method. *Journal of Computer-Aided Molecular Design*. 2020 Aug;34(8):817-30.
36. LigPrep, version 2.9. (2013). New York, NY: Schrodinger, LLC.
37. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrín PC, Mainz DT. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein– ligand complexes. *Journal of medicinal chemistry*. 2006 Oct 19;49(21):6177-96.
38. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of medicinal chemistry*. 2004 Mar 25;47(7):1750-9.
39. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. *Journal of cheminformatics*. 2011 Dec;3(1):1-4.
40. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of chemical research*. 2000 Dec 19;33(12):889-97.

41. Jayaram B, Singh T, Mukherjee G, Mathur A, Shekhar S, Shekhar V. Sanjeevini: a freely accessible web-server for target directed lead molecule discovery. In *BMC bioinformatics* 2012 Dec (Vol. 13, No. 17, pp. 1-13). BioMed Central.
42. Li H, Leung KS, Wong MH, Ballester PJ. Improving AutoDock Vina using random forest: the growing accuracy of binding affinity prediction by the effective exploitation of larger data sets.. 2015 Feb;34(2- 3):115-26.
43. Lipinski CA. Lead-and drug-like compounds: the rule-of-five revolution. *Drug discovery today: Technologies*. 2004 Dec 1;1(4):337-41.
44. Prival MJ. Evaluation of the TOPKAT system for predicting the carcinogenicity of chemicals. *Environmental and molecular mutagenesis*. 2001;37(1):55-69.
45. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, Van Der Spoel D, Hess B. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*. 2013 Apr 1;29(7):845-54.
46. Huang W, Lin Z, van Gunsteren WF. Validation of the GROMOS 54A7 force field with respect to β -peptide folding. *Journal of chemical theory and computation*. 2011 May 10;7(5):1237-43.
47. Lombardero M, Martin C, Jorge S, Lado F, Lomba E. An integral equation study of a simple point charge model of water. *The Journal of chemical physics*. 1999 Jan 8;110(2):1148-53.
48. Schüttelkopf AW, Van Aalten DM. PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallographica Section D: Biological Crystallography*. 2004 Aug 1;60(8):1355-63.
49. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*. 1981 Dec;52(12):7182-90.
50. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *The Journal of chemical physics*. 1995 Nov 15;103(19):8577-93.

51. Hess B, Bekker H, Berendsen HJ, Fraaije JG. LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry*. 1997 Sep;18(12):1463-72.
52. Miyamoto S, Kollman PA. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of computational chemistry*. 1992 Oct;13(8):952-62.
53. Turner PJ. XMGRACE, Version 5.1. 19. Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR. 2005;2.
54. David CC, Jacobs DJ. Principal component analysis: a method for determining the essential dynamics of proteins. *Protein dynamics* Humana Press, Totowa, NJ 2014 :193-226.
55. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert opinion on drug discovery*. 2015 May 4;10(5):449-61.
56. Kumari R, Kumar R, Open Source Drug Discovery Consortium, Lynn A. g_mmpbsa A GROMACS tool for high-throughput MM-PBSA calculations. *Journal of chemical information and modeling*. 2014 Jul 28;54(7):1951-62.
57. Anderson AC. The process of structure-based drug design. *Chemistry & biology*. 2003 Sep 1;10(9):787-97.
58. Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. *BMC biology*. 2011 Dec;9(1):1-9.
59. Seeliger D, De Groot BL. Conformational transitions upon ligand binding: holo-structure prediction from apo conformations. *PLoS computational biology*. 2010 Jan 8;6(1):e1000634.
60. Durham E, Dorr B, Woetzel N, Staritzbichler R, Meiler J. Solvent accessible surface area approximations for rapid and accurate protein structure prediction. *Journal of molecular modeling*. 2009 Sep;15(9):1093-108.
61. Chen D, Oezguen N, Urvil P, Ferguson C, Dann SM, Savidge TC. Regulation of protein-ligand binding affinity by hydrogen bond pairing. *Science advances*. 2016 Mar 25;2(3):e1501240.

62. Papaleo E, Mereghetti P, Fantucci P, Grandori R, De Gioia L. 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*. 2009 Jun 1;27(8):889-99.
63. Wang C, Greene DA, Xiao L, Qi R, Luo R. Recent developments and applications of the MMPBSA method. *Frontiers in molecular biosciences*. 2018 Jan 10;4:87.

Chapter 5

Evaluation of compounds having potential anti-cancer activity against programmed death-ligand 1 (PD-L1) using combined molecular docking, and molecular dynamics simulation approaches.

1. Introduction

The T cells (T lymphocytes) of the immune system have the potential to selectively recognize and destroy pathogens or abnormal cells, including cancer cells, by orchestrating a coordinated immune response including innate and adaptive responses [1]. Immune checkpoints are several costimulatory and co-inhibitory pathways that regulate the antigen recognition of the T-cell receptor in the process of the immune response [2-3]. In normal conditions, checkpoints ensure that the immune system cells do not mistakenly kill healthy cells during an immune response (known as an autoimmune reaction) [4]. Cancer cells can exploit these immune checkpoints as a way to evade immune detection and elimination by distorting its co-stimulatory and coinhibitory molecules [5-6]. Immune checkpoint therapy is currently considered to be one of the most promising strategies in the field of cancer immunotherapy [7-8]. Immune checkpoint therapy predominantly relies on restoring the unregulated co-signaling axis to release immune cells for removing malignant cells by administration of co-stimulatory signals agonists or co-inhibitory signals antagonists [9]. Cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) and programmed cell death protein 1 (PD-1), as well as its ligand PD-L1, are the most well-recognized co-inhibitory immune checkpoint molecules [10]. Accumulating evidence delineates that the PD-1/ PD-L1 signaling pathway plays a pivotal role in the maintenance of peripheral tolerance in normal tissues and in preventing autoimmunity [11]. Thus, it is not surprising that tumor cells can exploit the PD-1/PD-L1 axis to evade antitumor immune responses and eventually progress, propagate, and metastasize [12-14]. Previous studies suggest that the interaction of PD-L1/PD-1 in the tumor microenvironment promotes cancer development and progression by enhancing tumor cell proliferation and survival [15]. An encouraging number of studies report that targeting the programmed death 1 (PD-

1)/PD-1 ligand (PD-L1) immune checkpoint promotes an effective immune response against cancer cells and may be of significant therapeutic value (Figure 5-1) [16].

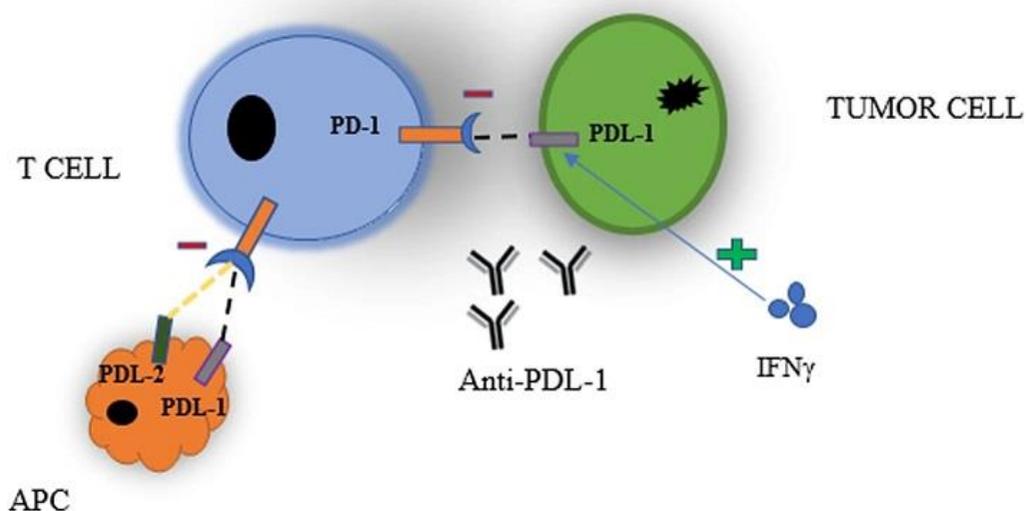


Figure 5-1: PD-1 and PD-L1 interaction of tumor cells and antigen-presenting cells (APC) with T cells that inhibit immune response. IFN γ help induces or maintains the expression of PD-L1. Anti-PD-L1 inhibits the interaction between PD-1 and PD-L1. [Adapted from Akinleye, A., Rasool, Z. Immune checkpoint inhibitors of PD-L1 as cancer therapeutics. *J Hematol Oncol* 12, 92 (2019)].

In this regard, the U.S. Food and Drug Administration (FDA) has approved targeted therapies like monoclonal antibodies (mAbs) namely nivolumab, pembrolizumab, and cemiplimab as well as PD-L1 inhibitors, atezolizumab, durvalumab, and avelumab [17-18]. These inhibitors have demonstrated their unique advantages which include well-tolerance, longer survival time, and higher potency against tumors [19-20]. Although mAbs exhibit high specificity and are ideal for blocking protein–protein interaction, some shortcomings exist like small half-life, less oral bioavailability, low stability, high manufacturing cost, immune-related adverse effects, and incomplete response to cancer patients [21-22]. In comparison to mAbs, small molecules have manageable pharmacodynamic and pharmacokinetic profiles, which provide them some definite advantages such as easy usage, higher membrane permeability, oral bioavailability, easy combination with other treatments, negligible immunogenicity, and rapid binding to the target protein [23-24]. Therefore, designing and synthesizing small molecules is a promising approach to the development of novel and inexpensive immune checkpoint inhibitors for PD-1/PD-L1 [25]. Furthermore, in combination with the mAbs, novel

small-molecule inhibitors are contemplated to attain synergistically anticancer effects [26]. Recently, novel small-molecule inhibitors bearing triazine scaffolds were successfully synthesized and are found to have superior antitumor activity [27]. Triazine is found to be a promising scaffold for designing inhibitors for the PD-1/PD-L1 pathway. It has a wide range of biological applications including antimalarial, antibacterial, antiviral, anti-inflammatory, anticancer, antileukemia, and anti-HIV activities [28]. Besides that triazine exhibit almost all types of chemical interactions like hydrogen, coordination bond (through aromatic N-atom), aromatic stacking interactions, electrostatic and charge-transfer interactions. Owing to these unique properties of triazine and the ease of performing chemical reactions, it is feasible to synthesize derivatives of triazine. Of note, biphenyl scaffold is also known to be a potent chemical moiety for blocking the PD-1/PD-L1 signaling pathway due to their desirable molecular properties such as manufacturability and high stability [29]. Previous literature propounded that the biphenyl moiety is essential for interaction with PD-L1 [30]. Synthesis of biphenyl derivative have been extensively carried out by amination, halogenation, sulphonation, alkylation, hydroxylation, metal complexation, etc. and have yielded a broad range of compounds with a wide range of activities such as antifungal, antimicrobial, anti-diabetic, antiproliferative, immunosuppressant, anti-inflammatory, analgesic, anti-cancer, etc. [31]. Hence, a novel class of novel small-molecule inhibitors was designed and synthesized by conjugating these two moieties (biphenyl and triazine) to prevent the binding of PD-L1 to its receptor PD-1 and ultimately blocking the downstream signaling. The 2D structures of these ten in-house synthesized novel small-molecule inhibitors are provided in Figure 5-2 [(a)-(j)]. In this present study, evaluation of these synthesized small molecules was done using molecular docking followed by ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) analysis. Thereafter, a molecular dynamics simulation of 100ns was carried out for each protein-inhibitor complex. Furthermore, principal component analysis (PCA) was performed and binding free energy was calculated using the Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) approach for all the protein-inhibitor complexes. The novel small-molecule inhibitors are expected to contribute to the development of potent anti-cancer agents.

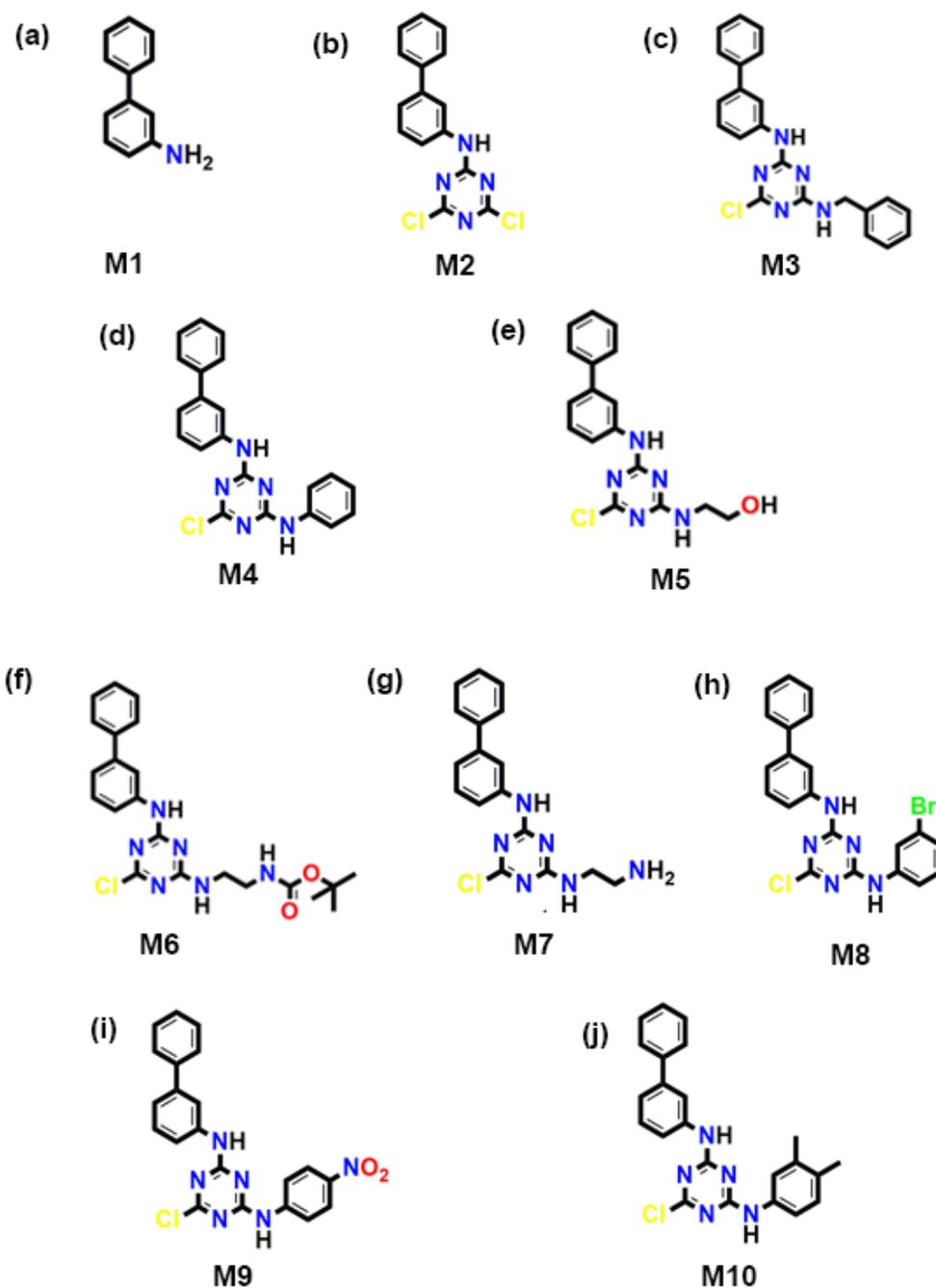


Figure 5-2: The 2D structures of synthesized novel small-molecule inhibitors. (a) - [1,1'-biphenyl]-3-amine (M1), (b) N-([1,1'-biphenyl]-3-yl)-4,6-dichloro-1,3,5-triazin-2-amine (M2), (c) N2-([1,1'-biphenyl]-3-yl)-N4-benzyl-6-chloro-1,3,5-triazine-2,4-diamine (M3), (d) N2-([1,1'-biphenyl]-3-yl)-6-chloro-N4-phenyl-1,3,5-triazine-2,4-diamine (M4), (e) 2-((4-([1,1'-biphenyl]-3-ylamino)-6-chloro-1,3,5-triazin-2-yl)amino)ethanol (M5), (f) tert-butyl (2-((4-([1,1'-biphenyl]-3-ylamino)-6-chloro-1,3,5-triazin-2-yl)amino)ethyl)carbamate (M6), (g) N2-([1,1'-biphenyl]-3-yl)-N4-(2-aminoethyl)-6-chloro-1,3,5-triazine-2,4-diamine (M7), (h) N2-([1,1'-biphenyl]-3-yl)-N4-(3-bromophenyl)-6-chloro-1,3,5-triazine-2,4-diamine (M8), (i) N2-([1,1'-biphenyl]-3-yl)-6-chloro-N4-(4-nitrophenyl)-1,3,5-triazine-2,4-diamine (M9) (j) N2-([1,1'-biphenyl]-3-yl)-6-chloro-N4-(3,4-dimethylphenyl)-1,3,5-triazine-2,4-diamine (M10)

2. Methodology

2.1. Selection of the PD-L1 protein structure and a reference ligand for docking purposes:

The development of novel small-molecule inhibitors that interfere with the PD-1/PD-L1 pathway has been slow in comparison to the development of mAbs [32-33]. This may be due to insufficient structural information which is hindering rational drug design and development [34-35]. In a previous report, a few molecules were disclosed by Bristol-Myers Squibb which are claimed to be “useful as inhibitors of the PD-1/PD-L1 protein/protein interaction” however no comprehensive information was provided [36]. In a recent study, the interaction of Bristol-Myers Squibb (BMS) compounds with the target protein was characterized and it was shown that the compounds directly and specifically bind to PD-L1, not PD-1, dissociating a preformed PD-1/PD-L1 complex *in vitro* effectively [37]. Additionally, this report provided the first crystal structure of novel small-molecule inhibitors bound to their target protein PD-L1 (in this case PDL1-dimer). Furthermore, this study contributed valuable insights into the protein-inhibitor interaction and pinpointed the druggable “hotspots” at the surface of PD-L1. Therefore, we have selected the above-mentioned crystal structure i.e., 5J89 (chain C and chain D) for our study and retrieved it from the protein data bank (PDB) [38]. The co-crystallized ligand with ID: 6GX (reference ligand) was extracted from protein crystal structure 5J89 using the discovery Studio 3.5 package (Accelrys, San Diego, CA, USA) and viewed in pymol [39-40]. The water molecules were eliminated, and protein preparation was done using the default settings of discovery Studio 3.5. Further, the protein structure was optimized using CHARMM (Chemistry at Harvard Macromolecular Mechanics) force field, and energy minimization was done by a smart minimizer algorithm incorporating conjugate gradient energy protocol [41].

2.2 Molecular Docking

The novel synthesized novel small-molecule inhibitors MK1-MK10 along with the reference ligand 6GX were subjected to molecular docking against the protein target 5J89 using the docking tools such as Autodock Vina and ParDOCK/BAPPL+ respectively [42-44].

2.2.1. Auto Dock Vina

All the ten synthesized novel small-molecule inhibitors along with the reference ligand were processed and subjected to docking analysis in AutoDock Vina. ADT tool was used to load the target protein structure and convert proteins to PDBQT format. Subsequently, the ligands were converted into PDBQT format in Open Babel [45]. The hydrogen atoms and water molecules from the protein structure were removed and polar hydrogen atoms were added. Thereafter, Kollman charges were incorporated into the target protein structures [46]. The auto-grid box was set across X-Y-Z directions, i.e., 13.1x32.9x184.3 points for PD-L1 (5J89) and a size of 23 points respectively. The grid was set up to encompass the active site pocket, with a spacing of 1 Å.

2.2.2. ParDOCK & Bappl+

Similarly, all the ten novel small-molecule inhibitors along with the reference ligand were subjected to evaluation by the Monte Carlo method involving an all-atom energy-based system using default parameters of ParDOCK software, hosted in the Sanjeevani server [47]. The ligands were positioned optimally around the active site of the protein target and ranked based on their interacting energies. Subsequently, the scoring function of all the aforementioned ligands was derived using the Bappl+ tool which utilizes the Random Forest algorithm scoring function [48]. In Bappl+ the default parameters were set and formal charges were assigned for each ligand. Based on these settings, the binding affinities of ligands were computed and the compounds were ranked accordingly.

2.3. ADMET and TOPKAT

Post docking by the Autodock Vina, and ParDOCK/Bappl+ tools, all the ten novel compounds were visually inspected in Pymol and Schrodinger maestro. ADMET and TOPKAT (Toxicity Prediction by Computer Assisted Technology) analysis was performed for each ligand. ADMET properties such as CYP2D6 & plasma protein inhibition, intestinal adsorption, aqueous solubility, hepatotoxicity, blood-brain barrier level were evaluated using the Discovery studio 3.5 ADMET tool kit. Furthermore, the novel synthesized novel small-molecule inhibitors along with the reference ligand were evaluated based on Lipinski parameters such as hydrogen bonds acceptors & donors, molar refractivity, ADMET, Wiener index, and volume of functional groups & protein

[49]. Thereafter, the toxicity of each ligand was estimated using the TOPKAT tool, which contains a robust Quantitative Structure-Toxicity Relationship (QSTR) modeling system to predict the accurate toxicity endpoints [50]. The toxicity properties such as rodent carcinogenicity, Ames mutagenicity, skin irritation, and developmental toxicity potential were also investigated.

2.4. Molecular Dynamics Simulation (MD)

All the ten novel compounds were subjected to MD simulation using GROMACS (Groingen Machine for Chemical Simulations) 2019 suite implementing GROMOS (54a7) force field [51-52]. The protein-inhibitor complex was solvated in a cubic box of 0.8 nm using periodic boundary conditions and a Simple point-charge water model (SPCE) [53]. The Na⁺ and Cl⁻ ions were added to neutralize the system and maintain the concentration of 0.15 mol/L. The PRODRG server was used to generate the ligand parameters and topology [54]. Thereafter, internal constraints of the protein-inhibitor complex were relaxed by 2000 steps of steepest descent energy minimization with a max force constraint of 1000 KJ/mol, leading to restraining positions of all heavy atoms. Before MD simulations, the systems were heated using a V-rescale thermostat to attain the temperature of 310 K with 0.1ps as constant of coupling and achieved equilibration in NVT (Number of atoms, Volume of the system, and Temperature of the system). Then solvent density was sustained using a Parrinello-Rahman barostat with the pressure of 1 bar, coupling constant of 0.1 ps, and temperature of 310 K to obtain equilibration in NPT (Number of atoms in the system, the pressure of the system, and temperature of the system) by gradually discharging the restraint on heavy atoms step by step [55]. Finally, an MD simulation was performed for the equilibrated structures for 5 ns with an integration time step of 2fs. The electrostatic interactions of long-range were implemented using particle-mesh Ewald sum with a cutoff of 1.0 nm [56]. During simulation, the LINCS algorithm was used to constrain all bond lengths, and the SETTLE algorithm was used for restraining water molecules [57-58]. The resultant structure from the NPT equilibration phase was employed for the final production run in the NPT ensemble for 200 ns simulation time. Finally, the trajectory analysis such as RMSD, RMSF, RoG, SASA, H-bonds, PCA of protein-inhibitor complexes was performed using gromacs utilities and plotted in xmgrace [59].

2.5. Principal Component Analysis

PCA is a statistical method that is applied to reduce data complexity and is useful for analyzing the large-scale motion of proteins [60]. Generally, internal motions of every protein molecule are ingrained and are required for their proper biological functioning like conformational changes to various biological environments, binding to substrate, etc. Due to problems in interpreting these internal motions of the protein, PCA is utilized to minimize the large dimensions of the data set to identify the prominent principal components (PCs). These PCs represent the major contributors responsible for elucidating crucial information about the dynamic changes of the protein. In PCA, a covariance matrix was constructed from trajectory data after eliminating unwanted motions (translational and rotational). This was followed by diagonalization of the covariance matrix using the *gmx_covar* tool which is a part of the Gromacs module. By diagonalization of covariance matrices containing backbone C alpha atoms data, the eigenvector and eigenvalues were obtained which corresponds to the change in protein trajectory throughout the simulation time. The *gmx ana eig* tool of Gromacs was used to analyze and plot the trajectories of the backbone C alpha atoms of all the systems. In our study, the first two projections (PC 1 and PC 2) with the highest eigenvalues were taken into consideration as they are responsible for 80–90% of the collective motions of the C alpha backbone atoms.

2.6. Free energy calculation

To evaluate the binding free energy (ΔG) of protein-inhibitor complexes, the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) method was employed [6]. We computed ΔG for the last 50ns of the production run using the *g_mmpbsa* tool of the Gromacs module [62]. Binding free energy was calculated for the novel small-molecule inhibitors by estimating the bound and unbound state differences with the protein targets.

$$\Delta G_{\text{binding}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}}) \longrightarrow (1)$$

In equation (1), the free energies of protein and ligands are indicated by G_{protein} and G_{ligand} . The free energy of the protein-inhibitor complex is represented by the G_{complex}

$$G_x = (E_{\text{MM}}) - TS + (G_{\text{solvent}}) \longrightarrow (2)$$

Similarly, equation (2) is used to calculate the bound or unbound state's free energy. The x indicates the unbound states or protein-inhibitor complex. E_{MM} is used to calculate the average molecular mechanics energy and TS indicates the entropic contribution. The solvation free energy is indicated by G_{solv}

$$E_{MM} = E_{bonded} + E_{nonbonded} \longrightarrow (3)$$

Equation (3) shows that bonded and non-bonded (i.e. vander Waal's & electrostatic) interactions were considered for calculating the molecular mechanics energy (E_{MM}) for protein-inhibitor interaction.

$$G_{solv} = G_{nonpolar} + G_{polar} \longrightarrow (4)$$

The linearized Poisson Boltzmann equation is represented by G_{solv} , where polar and non-polar hydrophobic contribution (G_{polar} & $G_{nonpolar}$) is taken into consideration for estimating the solvent accessible surface area. It was assumed that as all ligands are binding to one distinct protein target, therefore these ligands will contribute similar entropic energy. Hence, in our analysis complicated entropic contribution was not taken into account and was removed at the time of calculation.

3. Results and discussion

3.1 Comparative Molecular Docking

To understand the interaction profile of the novel compounds with the protein target 5J89, a comparative molecular docking analysis was conducted using two different tools, i.e., Auto dock Vina and ParDOCK/BAPPL+ respectively. All the novel ligands were individually docked to the active site cleft between chain C and chain D of the 5J89 protein target respectively. The key active site residues of the 5J89 protein target include THR20, TYR56, MET115, ILE116, SER117, ALA121, ASP122, TYR123, LYS124 of chain C and ILE54, TYR56, GLN66, MET115, ILE116, SER117 of chain D respectively. The results were obtained from docking all the ten novel small-molecule inhibitors (M1, M2, M3, M4, M5, M6 M7, M8, M9, M10) and the reference ligand (6GX) with their respective protein target (5J89) as provided in Table 5-1. The interacting residues which include both the active site and novel (other than the active site) residues of the protein (5J89) with the respective ligands (M1, M2, M3, M4, M5,

M6 M7, M8, M9, M10) are also provided in Table 5-1. As mentioned earlier, for comparative docking studies of the 5J89 protein target, the ligand 6GX was selected as the reference ligand. Its binding energy was found to be -9.3 kcal/mol and -9.89 kcal/mol in Autodock Vina and ParDOCK/BAPPL+ respectively. Simultaneously, we conducted docking studies for all the ten ligands i.e., M1, M2, M3, M4, M5, M6 M7, M8, M9, and M10 against the 5J89 protein target. The novel small-molecule inhibitors M1, M2, and M3 exhibited binding energy of -9.9, -10.6, -10.7 kcal/mol in Autodock Vina and -10.76, -11.4, -11.78 kcal/mol. in ParDOCK/ BAPPL+ respectively. Docking studies for the novel small-molecule inhibitors M4, M5, M6 reveal binding energy of -11.3, -10.98, -10.3 kcal/mol in Autodock Vina, and -12.01, -11.87, -11.45 kcal/mol in ParDOCK/ BAPPL+ respectively. Likewise, the novel small-molecule inhibitors M7, M8, M9, M10 exhibited binding energy of -9.82, -11.2, -11.02, -11.1 kcal/mol in Autodock Vina and -10.34, -11.97, -11.88, -12.26 kcal/mol in ParDOCK/ BAPPL+ respectively. Overall, from the docking results, it can be construed that in comparison to reference ligand 6GX, the novel small-molecule inhibitors M1, M2, M3, M4, M5, M6 M7, M8, M9, and M10 exhibited higher binding energy and stronger affinity towards their respective protein target 5J89. As indicated in Table 5-1, apart from the active site residues of the target proteins, no other residues are found to interact with the reference ligand 6GX. Therefore, it can be ascertained that the novel small-molecule inhibitors form more favorable interactions with their respective protein targets, thus enhancing stronger and stable binding. The 2D interaction images of all the top ten ligands M1, M2, M3, M4, M5, M6 M7, M8, M9, and M10 complexed with their respective protein targets 5J89 are also provided in Figure 5-3 [(a)-(j)]. It is noteworthy to mention that out of the ten novel small-molecule inhibitors, M3, M4, and M10 are the top three ligands that bind with greater affinity to their respective targets 5J89 as implied from the docking scores obtained from Autodock vina and ParDOCK/BAPPL. Post docking, ADMET & TOPKAT properties such as solubility, intestinal adsorption, etc. of the novel small-molecule inhibitors and the reference ligand were investigated, As provided in Table 5-2 and Table 5-3, these ligands were found non-hepatotoxic and they don't inhibit the cytochrome 4502D6 (CYP2D6) enzyme. The novel small-molecule inhibitors were also found not binding to Plasma Protein (PP) and don't exhibit any rodent carcinogenicity, Ames mutagenicity, skin irritation, with the least development toxicity.

Table 5-1: Molecular docking results of ten novel small-molecule inhibitors (M1-M10) and the reference ligand (6GX) bound with their respective protein target (5J89) using two docking tools: Autodock vina, ParDOCK/BAPPL+ respectively.

PROTEIN TARGET 5J89			
Ligands	Autodock Vina-Binding Energy (Kcal/mol)	ParDock/BappL+ Binding energy (Kcal/mol)	Interacting residues
M1	-9.9	-10.76	Chain C: ILE54, VAL55, TRY56, SER117, ILE116, MET115, ALA121, ASP122, TYR123
			Chain D : SER117, ILE116, MET115, ILE54, VAL55, TYR56, ALA121, ASP122, TYR123
M2	-10.6	-11.4	Chain C: ILE54, VAL55, TRY56, SER117, ILE116, MET115, ALA121, ASP122, TYR123
			Chain D: SER117, ILE116, MET115, GLN66, VAL68, ILE54, VAL55, TYR56, ALA121, ASP122, TYR123
M3	-10.7	-11.78	Chain C: ILE54, VAL55, TRY56, SER117, ILE116, MET115, ALA121, ASP122, TYR123, ARG125
			Chain D : ILE54, VAL55, TYR56, ASN63, GLN66, VAL68, SER117, ILE116, MET115, ALA121, ASP122, TYR123
M4	-11.3	-12.01	Chain C: ILE54, VAL55, TRY56, MET115, ILE116, SER117, GLY120, ALA121, ASP122, TYR123, PHE19, ALA18
			Chain D: MET115, ILE116, SER117, GLN66, VAL68,

PROTEIN TARGET 5J89			
Ligands	Autodock Vina- Binding Energy (Kcal/mol)	ParDock/BappL+ Binding energy (Kcal/mol)	Interacting residues
			ILE54, VAL55, TYR56, ASN63, ALA121, ASP122, TYR123
M5	-10.98	-11.87	Chain C: ILE54, VAL55, TRY56, MET115, ILE116, SER117, GLY120, ALA121, ASP122, TYR123, LYS124, ARG125
			Chain D: MET115, ILE116, SER117, GLN66, VAL68, ILE54, VAL55, TYR56, ASN63, ALA121, ASP122, TYR123
M6	-10.3	-11.45	Chain C: ILE54, VAL55, TRY56, MET115, ILE116, SER117, ALA121, ASP122, TYR123, LYS124, ARG125
			Chain D: MET115, ILE116, SER117, GLN66, VAL68, ILE54, VAL55, TYR56, GLU58, ASP61, LYS 62, ASN63
M7	-9.82	-10.34	Chain C: ILE54, VAL55, TRY56, MET115, ILE116, SER117, ALA121, ASP122, TYR123, LYS124, ARG125
			Chain D: MET115, ILE116, SER117, GLN66, VAL68, ILE54, VAL55, TYR56, ASN63, ALA121, ASP122, TYR123
M8	-11.2	-11.97	Chain C: ILE54, VAL55, TRY56, MET115, ILE116, SER117, GLY120, ALA121, ASP122, TYR123, PHE19, ALA18
			Chain D: MET115, ILE116, SER117, GLN66, VAL68,

PROTEIN TARGET 5J89			
Ligands	Autodock Vina-Binding Energy (Kcal/mol)	ParDock/BappL+ Binding energy (Kcal/mol)	Interacting residues
			ILE54, VAL55, TYR56, ALA121, ASP122, TYR123
M9	-11.02	-11.88	Chain C: ILE54, TYR56, MET115, ILE116, SER117, ALA121, ASP122, TYR123, LYS124, ARG125
			Chain D: MET115, ILE116, SER117, ASN63, GLN66, VAL68, ILE54, TYR56, GLU58, ALA121, ASP122, TYR123
M10	-11.1	-12.26	Chain C : ILE54, VAL55, MET115, ILE116, SER117, ALA121, ASP122, TYR123, LYS124, ARG125
			Chain D: MET115, ILE116, SER117, GLN66, VAL68, ILE54, VAL55, TYR56, ASN63, ALA121, ASP122, TYR123, VAL76
6GX-std	-9.3	-9.89	Chain C: ILE54, TYR56, ALA18, PHE19, THR20, MET115, ILE116, SER117, GLY120, ALA121, ASP122, TYR123
			Chain D: MET115, ILE116, SER117, ASN63, GLN66, VAL68, ILE54, VAL55, TYR56, ALA121, ASP122, TYR123

Table 5-2: ADMET properties of novel identified ligands (M1-M10) and the reference ligand (6GX). ADMET properties: Extension hepatotoxicity: <1 is nontoxic. CYP2D6: -ve is noninhibitors and +ve is inhibition, human intestinal absorption level: 0 (good); 1 (moderate); 2 (low); 3 (very low), (aqueous solubility): 0 (extremely low); 1 (low); 2 (good); 3 (optimal); 4 (too soluble), BBB (blood brain barrier): 0(very high); 1(high); 2 (mediums); 3 (low); 4 (undefined).

No.	Ligand name	Solubility level	BBB Penetration level	Absorption level	Extension CYP2D6
1	M1	3	1	0	-4.25 (false)
2	M2	2	1	0	-0.73 (false)
3	M3	1	3	0	-0.11 (false)
4	M4	2	1	0	-0.47 (false)
5	M5	2	2	0	-2.6 (false)
6	M6	1	3	2	-5.6 (false)
7	M7	2	3	0	-2.5 (false)
8	M8	1	3	2	-1.72 (false)
9	M9	2	3	2	-3.31 (false)
10	M10	2	3	1	-0.59 (false)
11	6GX-std	2	3	2	-3.76 (false)

Table 5-3: TOPKAT properties of novel identified ligands (M1-M10) and the reference ligand (6GX) TOPKAT properties: NC -Non-carcinogenic, C-Carcinogenic, NT-Non-Toxic, T-Toxic, NI-Non-Irritant, I-Irritant, NM-Non-Mutagenic, M-Mutagenic

No.	Ligand name	NTP carcinogenicity call (male rat) (v3.2)	NTP carcinogenicity call (female rat) (v3.2)	Developmental toxicity potential (DTP) (v3.1)	Skin irritation (v6.1)	Ames mutagenicity (v3.1)
1	M1	NC	NC	NT	NI	NM
2	M2	NC	NC	NT	NI	NM
3	M3	NC	NC	T	NI	NM
4	M4	NC	NC	NT	NI	NM
5	M5	NC	NC	NT	NI	NM
6	M6	NC	NC	NT	NI	NM
7	M7	NC	NC	NT	NI	NM
8	M8	NC	NC	NT	NI	NM
9	M9	NC	NC	NT	NI	NM
10	M10	NC	NC	NT	NI	NM
11	6GX-std	NC	NC	NT	I	M

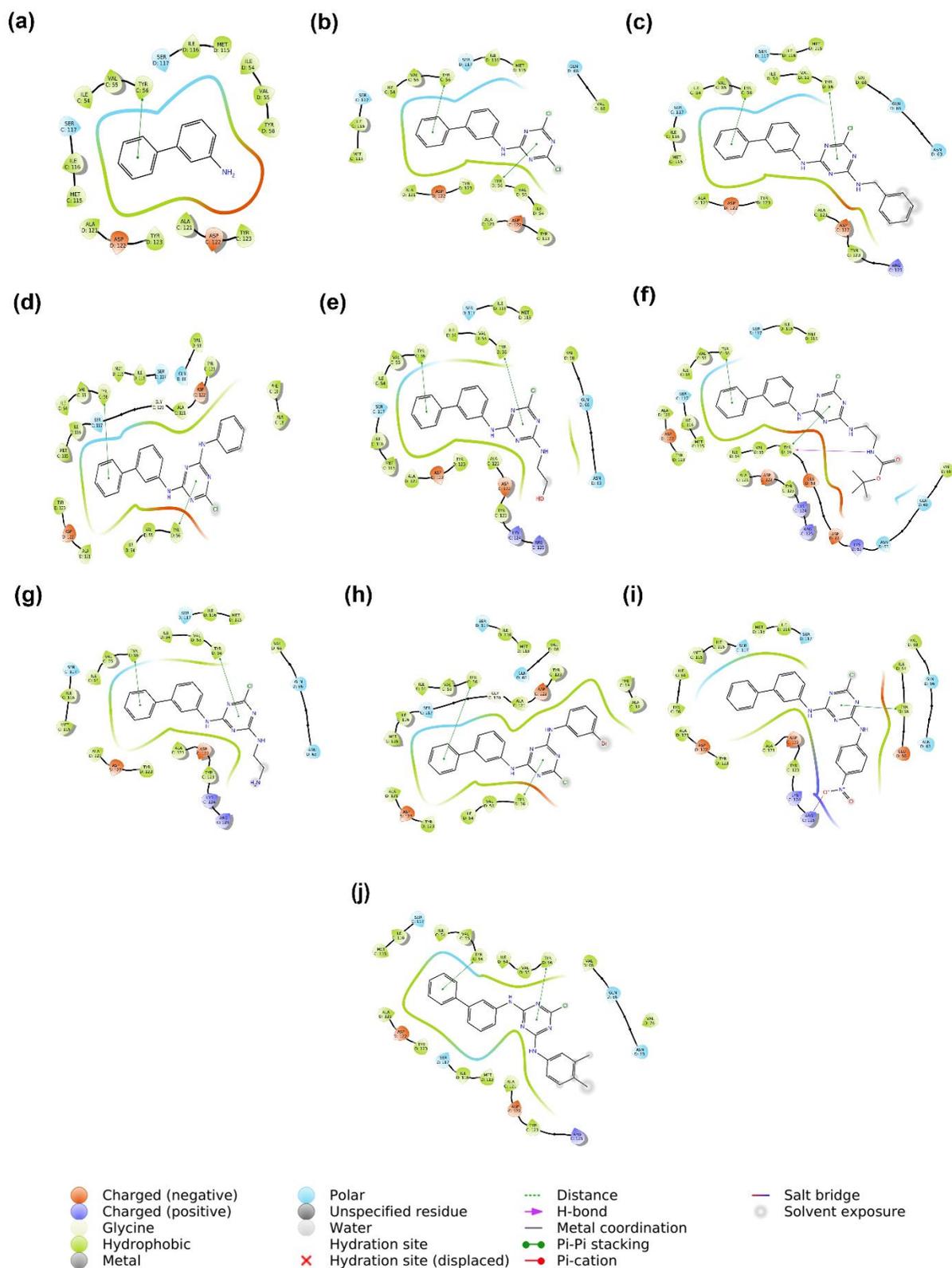


Figure 5-3: The 2D interaction profile of novel identified ligands (M1-M10) docked into the binding site of their respective protein target 5J89.

3.2 Trajectory Analysis

Molecular Dynamics Simulation (MDS) plays a significant role in studying conformational changes of protein-inhibitor complexes and provides valuable insights into the prediction and identification of protein-inhibitor interactions [63]. The ten novel ligands M1, M2, M3, M4, M5, M6, M7, M8, M9, and M10 docked with their respective protein target i.e., 5J89 were subjected to molecular dynamics simulation in GROMACS for 100ns duration. Also, MD simulation for 100ns was performed for reference ligand i.e., 6GX docked with 5J89 respectively. Additionally, we performed a simulation run of 100ns for the protein 5J89 without ligand (APO). Thereafter, we performed trajectory analysis such as Root-mean-square-deviation (RMSD), Root-mean-square fluctuation (RMSF), Radius of gyration (Rg), number of Hydrogen bonds, bond distribution, Principal component analysis (PCA), and binding free-energy calculation (MMPBSA) for all the systems (standalone proteins, protein-reference ligand complex, and protein-identified ligand complex).

3.2.1 Root-mean-square deviation analysis

To elucidate the influence of ligands on the conformational stability of the protein, we have analyzed the root-mean-square deviation (RMSD) of backbone atoms of all the systems using the standard *g_rms* function of GROMACS for an overall time of 100 ns simulation run (Figure 5-4). As provided in Table 5-4, the 5J89 protein docked with the reference ligand 6GX exhibits an average deviation of 0.63 ± 0.07 nm. 5J89 APO protein and the novel inhibitors M1, M2, M3 show an average deviation of 0.33 ± 0.06 nm, 0.35 ± 0.05 nm, 0.29 ± 0.06 , 0.31 ± 0.04 nm respectively. Similarly, 5J89 protein docked with M4, M5, M6, M7 shows an average deviation of 0.42 ± 0.07 nm, 0.42 ± 0.09 nm, 0.27 ± 0.04 nm, 0.32 ± 0.05 nm respectively. Likewise, 5J89 protein docked with M8, M9, M10 exhibited an average deviation of 0.30 ± 0.04 nm, 0.43 ± 0.05 nm, 0.29 ± 0.04 nm respectively. From the RMSD graphs, it can be inferred that all the systems are equilibrated and have converged within a range of 0.2-0.4nm (Figure 5-4). Although the RMSD profile shows an insignificant difference in the average value and minor fluctuations throughout, convergence is observed which indicates that our simulations have achieved stable trajectories. Hence, other conformational dynamics analysis is required to deduce further conclusions.

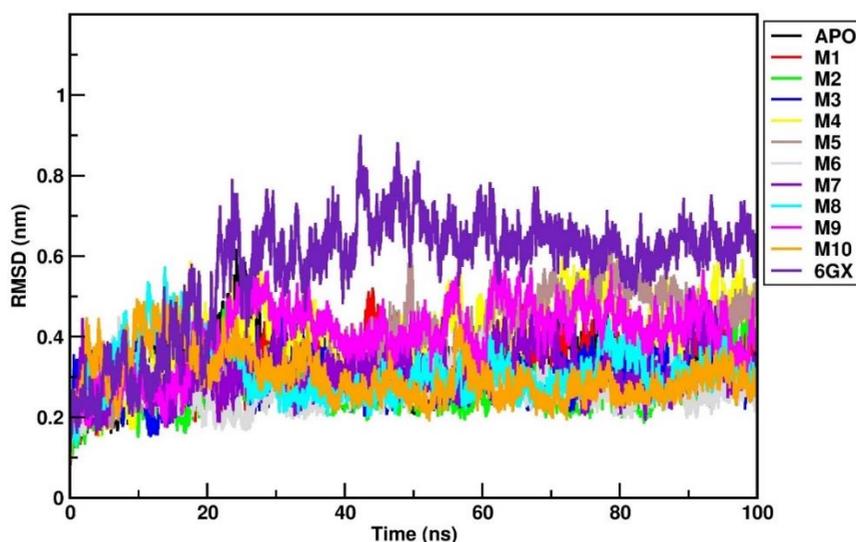


Figure 5-4: RMSD profile of 5J89 protein-inhibitor complexes [novel identified ligands (M1-M10) and reference ligand 6GX] along with APO.

Table 5-4: RMSD profile of 5J89 protein-inhibitor complexes [novel small-molecule inhibitors (M1-M10) and reference ligand 6GX] along with APO.

RMSD		
Protein-Ligand Complex (20-100ns)	Average(nm)	SD +-
5J89-APO	0.33	0.06
5J89-M1	0.35	0.05
5J89-M2	0.29	0.06
5J89-M3	0.31	0.04
5J89-M4	0.42	0.07
5J89-M5	0.42	0.09
5J89-M6	0.27	0.04
5J89-M7	0.32	0.05
5J89-M8	0.30	0.04
5J89-M9	0.43	0.05
5J89-M10	0.29	0.04
5J89-6GX	0.63	0.07

3.2.2 Residue flexibility analysis

To understand the overall residue-wise protein dynamics in the systems, the evaluation of root mean square fluctuations (RMSF) plays a very important role. Therefore, to check the stability and flexibility of the residues, we have calculated the RMSFs of both chain C and chain D of PDL1(5J89) for an entire simulation run of 100ns using the `gmx_rmsf` module of GROMACS (Figure 5-5 a-b). From the RMSF profile of chain C of 5J89, it can be inferred that the ligands M1-M10 exhibited less fluctuation at active site residues respectively. But a higher fluctuation peak of 0.38 nm, 0.2 nm, 0.23 nm, 0.25 nm, 0.29 nm, 0.4 nm, 0.37 nm, 0.35 nm, 0.33 nm is observed at THR20, TYR56, MET115, ILE116, SER117, ALA121, ASP122, TYR123, LYS124 residues of chain C for the reference ligand 6GX. This indicates that the ligands M1-M10 are contributing to greater rigidity to the binding pocket of 5J89. Similarly, from the RMSF profile of chain D of 5J89, it can be construed that the ligands M1-M10 exhibited a similar less fluctuation pattern at active site residues respectively. But the reference ligand 6GXa shows a higher fluctuation peak of 0.18, 0.15, 0.23, 0.13, 0.14, 0.15 nm observed at ILE54, TYR56, GLN66, MET115, ILE116, SER117 residues. This propounds that the novel inhibitors M1-M10 have made the binding pocket more inflexible. From the RMSF profile of all the systems, it can be inferred that the ligands M1-M10 are responsible for the restricted dynamics at the active site residue pocket and for imparting structural rigidity to the protein target 5J89.

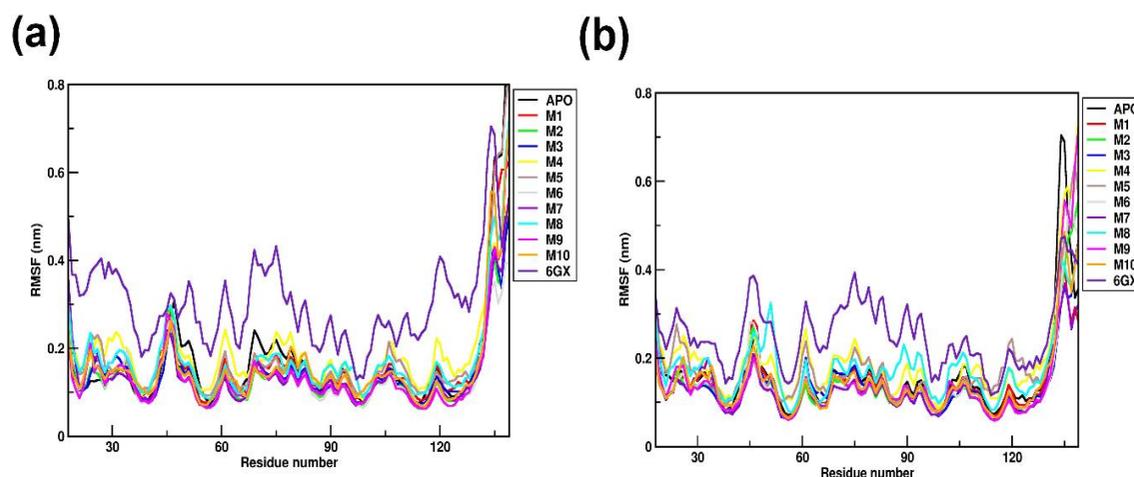


Figure 5-5: RMSF profile of 5J89 protein-inhibitor complexes [novel small-molecule inhibitors (M1-M10) and reference ligand 6GX] along with APO.

3.2.3 Compactness analysis

To study the compactness of all the protein-inhibitor complexes, the radius of gyration (R_g) is computed using the *gmx_gyrate* function of GROMACS for a simulation time of 100ns. R_g is explained as the distance measured between the terminal end of protein and its center of mass during simulation. When a ligand docks to a protein, there is a conformational shift that changes the radius of gyration [64]. The compact protein structure tends to maintain a low R_g average deviation thus showing dynamic stability. The combined R_g plot of all ligands is represented in Figure 5-6. As shown in Table 5-5, 5J89 APO protein and 5J89 docked with novel inhibitors M1, M2, M3 show an average R_g deviation of 2.01 ± 0.04 nm, 1.99 ± 0.02 nm, 2.00 ± 0.03 nm, 2.06 ± 0.02 nm respectively from 30-100 ns simulation time. Similarly, 5J89 protein docked with M4, M5, M6, M7 shows an average R_g deviation of 2.06 ± 0.02 nm, 2.05 ± 0.02 nm, 2.01 ± 0.02 nm, 1.98 ± 0.02 nm respectively from 30-100 ns simulation time. Likewise, 5J89 protein docked with ligands M8, M9, M10 exhibited an average R_g deviation of 2.00 ± 0.03 nm, 2.03 ± 0.02 nm, 1.98 ± 0.02 nm respectively from 30-200 ns simulation time. As provided in Table 5-5 the 5J89 protein docked with the reference ligand 6GX exhibits an average R_g deviation of 2.53 ± 0.03 nm. Collectively, from the radius of gyration profile, it can be concluded that the novel small-molecule inhibitors M1, M2, M3, M4, M5, M6, M7, M8, M9, and M10 are responsible for making the structure more compact compared to the reference ligand 6GX.

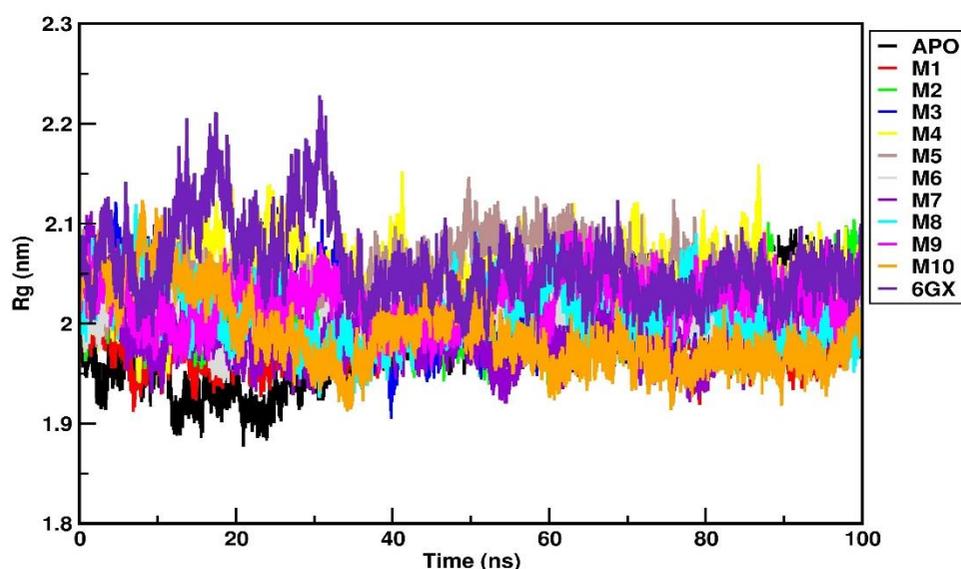


Figure 5-6: R_g profile of 5J89 protein-inhibitor complexes [novel small-molecule inhibitors (M1-M10) and reference ligand 6GX] along with APO.

Table 5-5: Rg values of all the protein-inhibitor complexes [novel small-molecule inhibitors (M1-M10) and reference ligand 6GX] along with APO.

Rg		
Protein-Ligand Complex (20-100ns)	Average(nm)	SD +-
5J89-APO	2.01	0.04
5J89-M1	1.99	0.02
5J89-M2	2.00	0.02
5J89-M3	2.00	0.03
5J89-M4	2.06	0.02
5J89-M5	2.05	0.02
5J89-M6	2.01	0.02
5J89-M7	1.98	0.02
5J89-M8	2.00	0.03
5J89-M9	2.03	0.02
5J89-M10	1.98	0.02
5J89-6GX	2.53	0.03

3.2.4 Solvent Accessible Surface Area analysis

To calculate the solvent-accessible surface area (SASA) of a molecular surface, a method is employed, which involves the *in-silico* rolling of a spherical probe that approximates a water molecule, encompassing a full-atom protein model [65]. This procedure takes into account the expansion of the van der Waals radius for each atom by 1.4 Å (the radius of a polar solvent probe) followed by the calculation of the surface area of these atoms with expanded radius. Therefore, the SASA values of all the complexes were computed using the *gmx_sasa* function of GROMACS for a simulation time of 100 ns.

As provided in Figure 5-7 and Table 5-6, 5J89 docked with novel inhibitors M1, M2, M3, and 6GX (reference ligand) show an average SASA value of $139.42 \pm 3.06 \text{ nm}^2$, $141.66 \pm 2.64 \text{ nm}^2$, $141.23 \pm 2.92 \text{ nm}^2$, $144.35 \pm 2.86 \text{ nm}^2$ respectively from 30-100 ns simulation timeframe. It is observed that the SASA curves followed the same trend as the reference ligand 6GX. Similarly, 5J89 APO protein and 5J89 protein docked with ligands M4, M5, M6 show an average SASA value of $139.20 \pm 2.79 \text{ nm}^2$, 142.99

$\pm 2.71 \text{ nm}^2$, $140.72 \pm 2.86 \text{ nm}^2$, $141.35 \pm 2.61 \text{ nm}^2$ respectively from 30-100 ns simulation time frame (Figure 5-7 and Table 5-6). Likewise, 5J89 protein docked with ligands M7, M8, M9, and M10 exhibited an average SASA value of $141.21 \pm 2.31 \text{ nm}^2$, $143.72 \pm 3.03 \text{ nm}^2$, $142.57 \pm 2.45 \text{ nm}^2$, $140.63 \pm 2.68 \text{ nm}^2$ respectively from 30-100 ns simulation time frame (Figure 5-7 and Table 5-6). The values obtained from the SASA analysis of protein structures docked with our novel small-molecule inhibitors are found to be significantly similar to that of the reference protein-ligand complex (5J89-6GX). Hence, it can be interpreted that similar to reference ligands, the novel small-molecule inhibitors make the active site residues of the protein targets readily accessible to the solvent surface, thus exposing the proteins to the hydration shell.

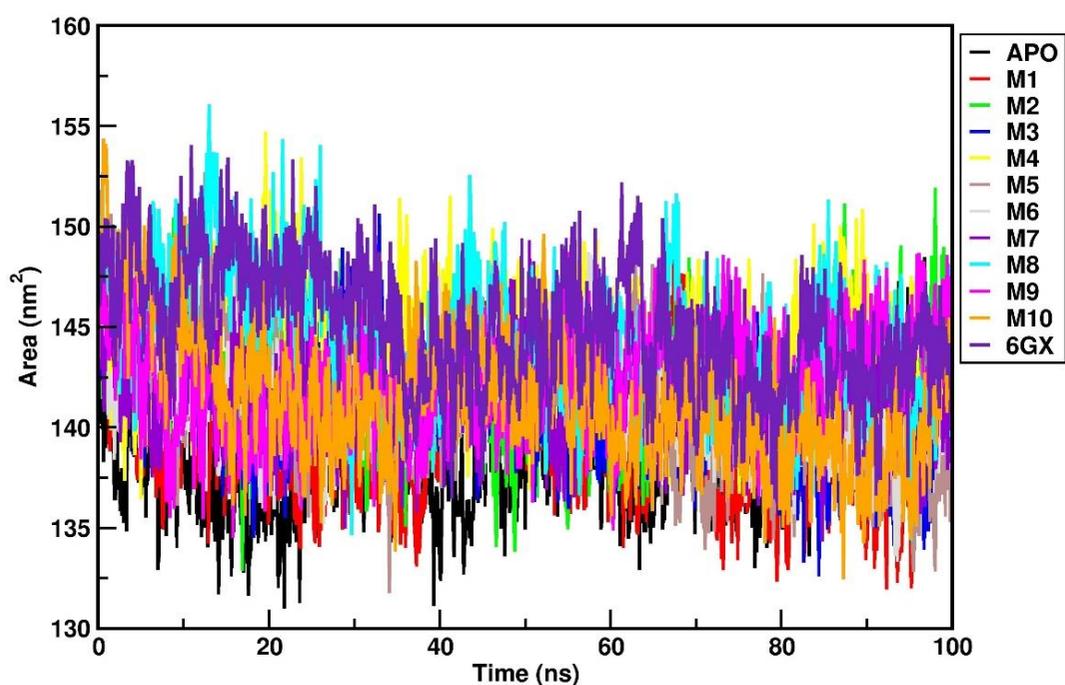


Figure 5-7: SASA profile of 5J89 protein-inhibitor complexes [novel small-molecule inhibitors (M1-M10) and reference ligand 6GX] along with APO.

Table 5-6: SASA values of all the protein-inhibitor complexes [novel small-molecule inhibitors (M1-M10) and reference ligand 6GX] along with APO.

SASA		
Protein-Ligand Complex (20-100ns)	Average(nm)	SD +-
5J89-APO	139.20	2.79
5J89-M1	139.42	3.06
5J89-M2	141.66	2.64
5J89-M3	141.23	2.92
5J89-M4	142.99	2.71
5J89-M5	140.72	2.86
5J89-M6	141.35	2.61
5J89-M7	141.21	2.31
5J89-M8	143.72	3.03
5J89-M9	142.57	2.45
5J89-M10	140.63	2.68
5J89-6GX	144.35	2.86

3.2.5 Hydrogen Bond and Bond distribution analysis

The hydrogen bonds (H-bonds) are known to play a key role in forming stable contacts between the ligands and protein target [66]. In this study, H-bond analysis was performed on all the protein-inhibitor systems for a total simulation run of 100ns. The number of H-bonds and their bond frequencies are plotted using the `gmx_hbond` tool of GROMACS and are represented in Figure 5-8. It is observed that the ligand M1 formed an unstable H-bond of length 0.34 nm with the protein target 5J89, and its frequency to form this bond is 5%. Likewise, the M2 formed one stable H-bond during the simulation period, and its frequency to form a bond of the length of 0.31 nm is 16%. Similarly, M3 had formed stable 1-2 H-bonds of length 0.27 nm with a frequency of 17.5%. In the case of the 5J89 protein target, the ligand M4 established stable 1-2 H-bonds during the simulation period and had a 21.3% frequency to form a bond of the length of 0.26nm. In the same way, the M5 exhibited one stable H-bond during the

simulation period, and its ability to form a bond of length 0.32 is 15%. Similarly, the ligand M6 formed 1 H-bond, and its frequency to form a bond of length 0.28 nm is 15%. Likewise, in the case of the 5J89 protein target, the ligand M7 formed 1-2 stable H-bonds during the simulation time, and its frequency to form a bond of length 0.34 nm is 16%. Also, the ligand M8 formed one H-bond, and its frequency to form a bond of length 0.34 nm is 16.1%. Similarly, The M9 ligand formed 1-3 stable H-bonds, and the frequency to form a bond of length 0.28 nm is 15%. Likewise, the ligand M10 formed 1-2 stable H-bonds during the simulation time, and its frequency to form a bond of length 0.27 nm is 16.2%. Lastly, the reference ligand 6GX formed 1-3 unstable H-bonds, and the frequency to form a bond of length 0.34 nm is 18%. Overall, from the h-bond analysis, it can be concluded that our identified ligands- M3, M4, M10 form stable and stronger H-bonds of length < 0.27 nm with their target protein compared to the reference ligands.

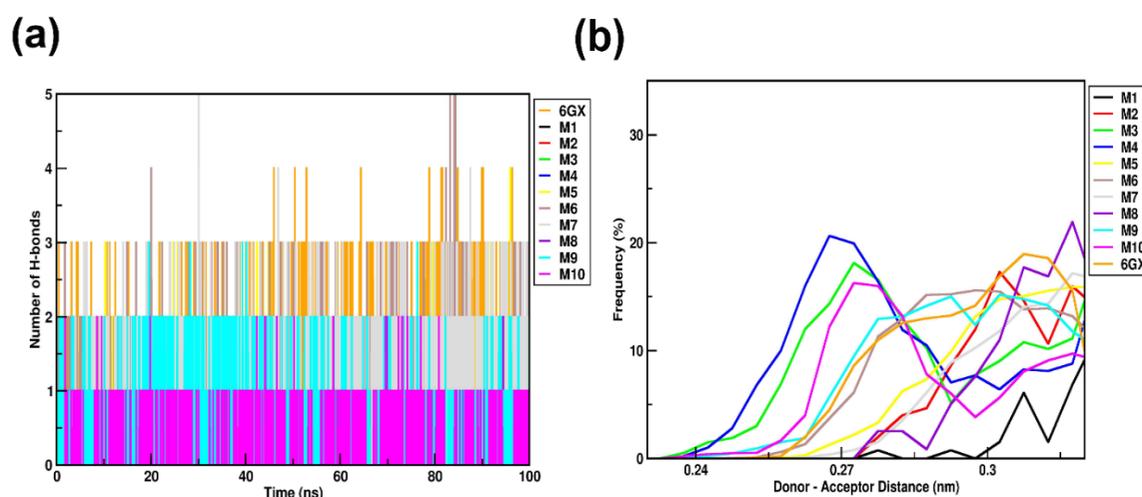


Figure 5-8: Hydrogen bonding pattern of 5J89 protein-inhibitor complexes [novel identified ligands (M1-M10) and reference ligand 6GX] along with APO.

3.2.6 Principal component analysis

In general, a protein-inhibitor complex that occupies a smaller phase space with a stable cluster denotes a highly stable complex [67]. Therefore, to explore this phenomenon, we have performed Principal Component Analysis (PCA) on the trajectories obtained from the 100ns simulation period for all the protein-inhibitor complexes as shown in the 2D plots of PCA (Figure 5-9). From the PCA graph, it is observed that the complexes formed by M1-M10 with 5J89 are clustered very closely, occupy small

phase space and the area coverage is less compared to reference ligand 6GX. Thus, from the PCA analysis, it is implied that our novel small-molecule inhibitors restrict the protein motion, thus imparting rigidity to their respective protein-inhibitor complexes compared to reference ligands.

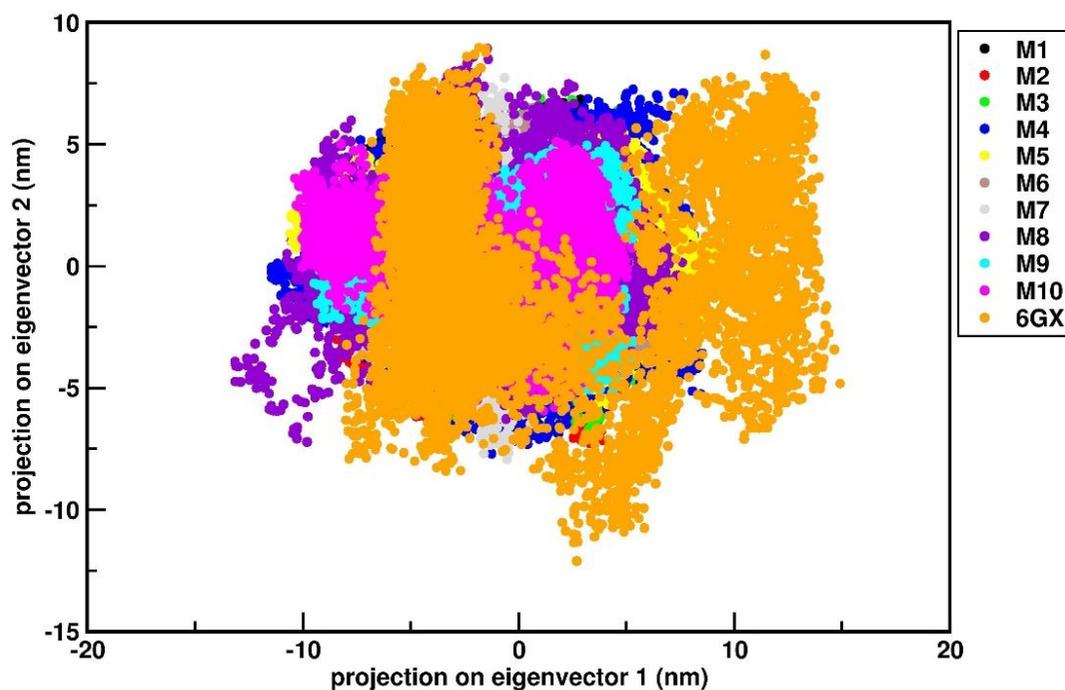


Figure 5-9: Principal component analysis: Projection of motion of 5J89-M1 in black; 5J89-M2 complex in red; 5J89-M3 in green; 5J89-M4 in blue, 5J89-M5 in yellow; 5J89-M6 complex in pink; 5J89-M7 complex in grey; 5J89-M8 in violet; 5J89-M9 in cyan, 5J89-M10 in purple; 5J89-6GX complex in orange.

3.2.7 Binding free energy calculation

The results obtained by calculation of the free energy of ligands give insights into ligands' binding potential with bonded and non-bonded entities [68]. In this study, the *g_mmpbsa* tool was used to calculate binding free energy (ΔG) for the last 50ns of the simulation time. From the MMPBSA graphs, it is observed that the reference ligand 6GX, and the novel small-molecule inhibitors M1, M2, M3 bound to the 5J89 target protein exhibit binding energy of -121.422 ± 10.843 kJ/mol and -150.165 ± 7.395 kJ/mol, \pm kJ/mol, -200.113 ± 11.477 kJ/mol, -230.282 ± 13.028 kJ/mol respectively (Figure 5-10). Similarly, from the MMPBSA graphs, it is observed that the novel small-molecule inhibitors M4, M5, M6 bound to the 5J89 protein target exhibits

binding energy of -233.623 ± 18.600 kJ/mol, -193.142 ± 15.756 kJ/mol, -211.679 ± 15.524 kJ/mol respectively (Table 5-7). Likewise, from the MMPBSA graphs, it is observed that the novel small-molecule inhibitors M7, M8, M9, and M10 bound to the 5J89 protein target, exhibit binding energy of -202.559 ± 13.036 kJ/mol, -222.366 ± 23.224 kJ/mol, -217.321 ± 16.770 kJ/mol, -226.747 ± 13.020 kJ/mol respectively (Table 5-7). Furthermore, from the binding energy calculation, it can also be concluded that out of the ten novel small-molecule inhibitors, the top three ligands having the best binding energy are M3, M4, and M10 against their protein target 5J89. The 3D images of these top three ligands M3, M4, and M10 bound to the active site residues of their respective protein targets 5J89 are shown in Figure 5-11 (a-c). For a better visual representation, the 3D images of both the chains C and D of 5J89 (PD-L1) docked with an inhibitor are provided in Figure 5-12 (a-b). The binding free energies (ΔG) of 5J89 bound with our novel small-molecule inhibitors were found to be significantly better than that of 5J89 bound with the reference ligand. Additionally, the four individual energy components such as van der Waal energy (ΔE_{vdW}), electrostatic energy (ΔE_{elec}), polar solvation energy ($\Delta g_{polar-solv}$), and SASA energy ($\Delta G_{non-polar}$) were compared for all the protein-inhibitor complexes. The details of the MMPBSA energies of all the protein-inhibitor complexes are provided in Table 5-7. It can be inferred that the novel small-molecule inhibitors docked with their respective target protein have the highest electrostatic energy, SASA energy, and Vander Waal energy in comparison with reference ligands. This also indicates that despite intermittent hydrogen bond formations as observed during h-bond analysis, the ligands were conserved in the binding pocket by nonbonded interactions (electrostatic, polar, and nonpolar). Based on overall MD analysis and MMPBSA energies, we have found the ligands M3, M4, and M10 to be the best performing amongst the top ten ligands.

Table 5-7: MMPBSA detailed energies (Binding energy and the individual energy components van der Waal energy (ΔE_{vdw}), electrostatic energy (ΔE_{elec}), polar solvation energy ($\Delta G_{polar-solv}$), SASA energy ($\Delta G_{non-polar}$) of all the protein-ligand complexes [5J89 complexed with the novel small-molecule inhibitors (M1-M10) and reference ligand 6GX].

PDL1- 5J89	MMPBSA DETAILED ENERGIES				
	Ligands	van der Waal energy	Electrostatic energy	Polar solvation energy	SASA energy
6GX (Standard)	-161.862 +/- 12.304 kJ/mol	-41.152 +/- 6.109 kJ/mol	100.883 +/- 8.469 kJ/mol	-19.292 +/- 0.921 kJ/mol	-121.422 +/- 10.843 kJ/mol
M1	-162.952 +/- 6.831 kJ/mol	-0.487 +/- 0.359 kJ/mol	26.122 +/- 5.257 kJ/mol	-12.848 +/- 0.791 kJ/mol	-150.165 +/- 7.395 kJ/mol
M2	-231.755 +/- 7.786 kJ/mol	1.067 +/- 3.420 kJ/mol	48.312 +/- 10.211 kJ/mol	-17.737 +/- 0.744 kJ/mol	-200.113 +/- 11.477 kJ/mol
M3	-280.654 +/- 8.920 kJ/mol	-0.896 +/- 3.762 kJ/mol	73.147 +/- 14.218 kJ/mol	-21.880 +/- 0.977 kJ/mol	-230.282 +/- 13.028 kJ/mol
M4	-276.443 +/- 12.087 kJ/mol	0.237 +/- 5.335 kJ/mol	64.718 +/- 15.913 kJ/mol	-22.135 +/- 1.102 kJ/mol	-233.623 +/- 18.600 kJ/mol
M5	-239.256 +/- 14.084 kJ/mol	-1.300 +/- 4.738 kJ/mol	67.252 +/- 13.490 kJ/mol	-19.838 +/- 1.022 kJ/mol	-193.142 +/- 15.756 kJ/mol
M6	-279.624 +/- 15.584 kJ/mol	-8.908 +/- 13.887 kJ/mol	99.517 +/- 22.634 kJ/mol	-22.663 +/- 1.082 kJ/mol	-211.679 +/- 15.524 kJ/mol
M7	-250.280 +/- 9.823 kJ/mol	3.909 +/- 4.118 kJ/mol	63.635 +/- 10.826 kJ/mol	-19.823 +/- 0.789 kJ/mol	-202.559 +/- 13.036 kJ/mol
M8	-265.642 +/- 15.054 kJ/mol	2.995 +/- 5.021 kJ/mol	61.319 +/- 12.949 kJ/mol	-21.039 +/- 1.014 kJ/mol	-222.366 +/- 23.224 kJ/mol
M9	-292.077 +/- 12.898 kJ/mol	-22.994 +/- 5.571 kJ/mol	119.556 +/- 14.166 kJ/mol	-21.805 +/- 1.096 kJ/mol	-217.321 +/- 16.770 kJ/mol
M10	-279.196 +/- 10.697 kJ/mol	2.510 +/- 5.117 kJ/mol	71.275 +/- 12.145 kJ/mol	-21.335 +/- 1.095 kJ/mol	-226.747 +/- 13.020 kJ/mol

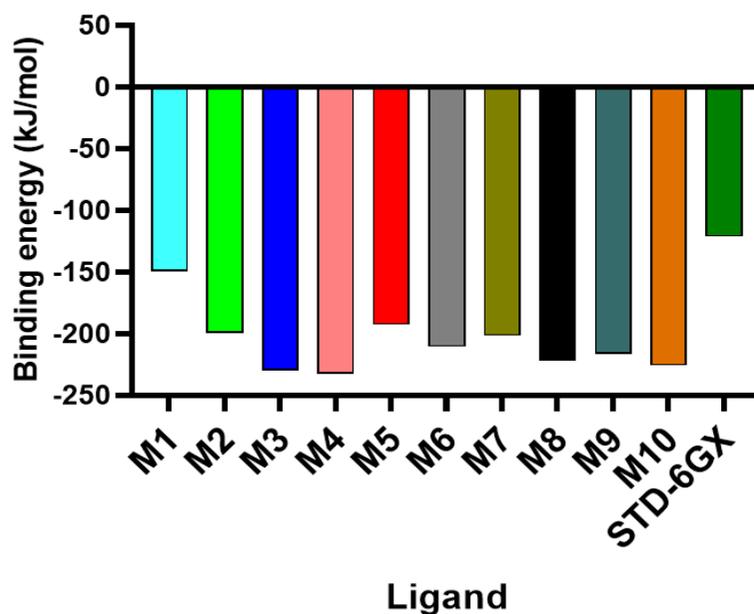


Figure 5-10: Binding energy profile (ΔG) of 5J89 complexed with the novel small-molecule inhibitors (M1-M10) and reference ligand 6GX.

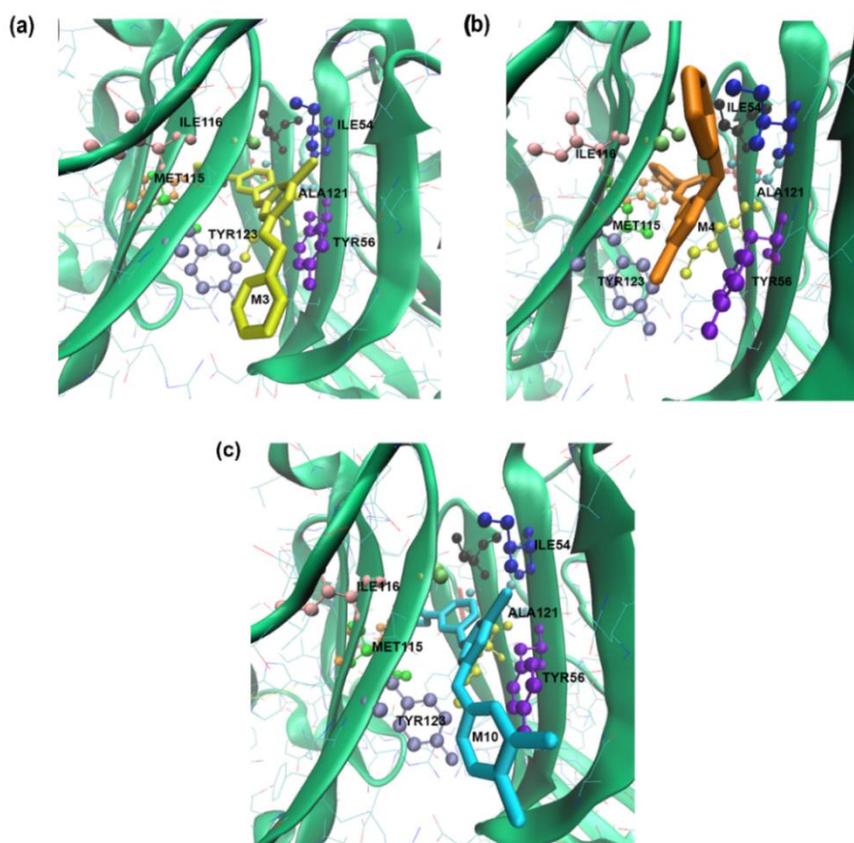


Figure 5-11: Schematic representation of the top three novel small-molecule inhibitors bound to the active site residues of their respective protein target 5J89 using VMD. (a) M3 (b) M4 (c) M10. Proteins are shown in the New Cartoon representation. Active site residues of the protein are shown in licorice representation. Novel ligands M3, M4, and M10 are shown in CPK representation.

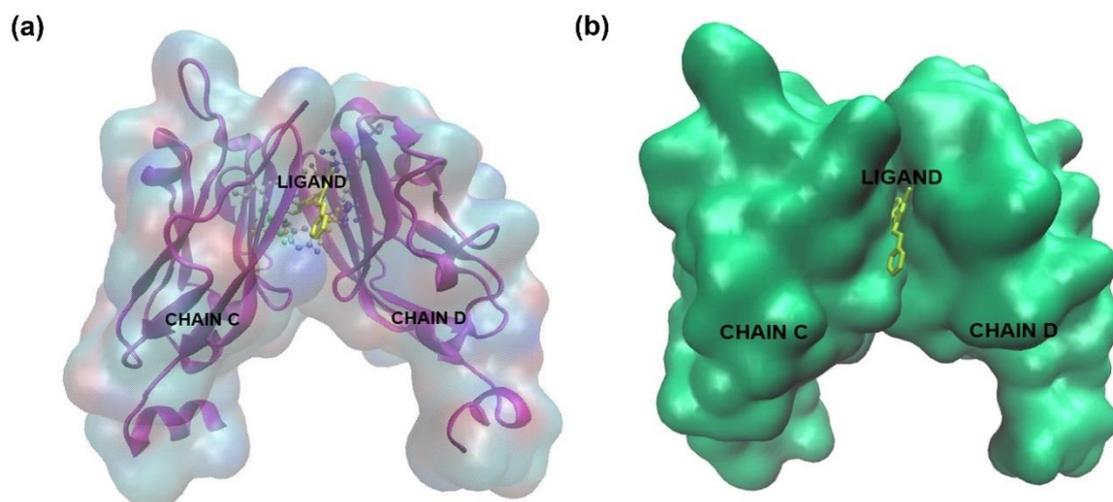


Figure 5-12: 3D representation of PDL1 protein dimer (chain C and chain D) (a) mixed surface and new cartoon representation of both chain C and D in purple color using VMD (b) CPK representation of ligand docked to the binding groove of PDL1 protein dimer shown in surface form using VMD.

4. Conclusions

Previous literature revealed the crystal structure of PDL1 docked to BMS compounds and highlighted the druggable hotspots of the protein. In this study, evaluation of in-house synthesized novel small-molecule inhibitors docked to the binding site cleft was performed using combined docking and molecular dynamics simulation approaches. The novel compounds were docked to the binding cleft of the protein target (5J89) using Autodock-Vina, and ParDOCK/BAPPL+. Docking scores of these top ten ligands were evaluated and their ADMET profiling was also performed keeping the reported ligand 6GX as reference. All the ten novel inhibitors were individually subjected to MD simulation for a period of 100 ns followed by a detailed trajectory analysis. RMSD profile suggests that all the protein-inhibitor complexes have attained stable trajectories and are well equilibrated. From the RMSF analysis of all the systems, it can be inferred that the novel small-molecule inhibitors are responsible for the restricted dynamics at the active site residue pocket of their respective protein targets, thereby imparting structural rigidity to the protein compared to the reference ligands. Rg data of all the systems reveals that in comparison to reference ligands, our novel small-molecule inhibitors contribute greater compactness to the structure of their respective protein targets. From the SASA analysis, it is evident that similar to reference ligands, the novel small-molecule inhibitors make the active site residues of the protein targets well exposed to the solvent, thereby making them readily accessible. From the h-bond

analysis and bond distribution, it was observed that the novel small-molecule inhibitors tend to form stronger H-bonds of length ≥ 0.28 nm with their target proteins compared to reference ligands. PCA analysis helps us to understand that in comparison to reference ligands our novel small-molecule inhibitors occupy smaller phase space and form stable clusters, thus conferring more stability to the protein-inhibitor complexes. The results obtained by analyzing molecular docking and MD simulation data are well-validated by the binding free energy calculation. From the MMPBSA profile, it is observed that, in the identified protein-inhibitor complexes, there exist strong interactions in terms of energy components other than hydrogen bonding such as electrostatic energy, SASA energy, and Vander Waal energy compared to reference protein-inhibitor complexes. Based on overall MD analysis and MMPBSA energies, we have found the ligands M3, M4, and M10 to be the best performing amongst the top ten ligands. Hence, our study strongly substantiates the point that our novel small-molecule inhibitors can act as strong inhibitors and further experimental studies can be conducted to prove the therapeutic potential of these promising anticancer agents.

References:

1. Chaplin DD. Overview of the immune response. *Journal of allergy and clinical immunology*. 2010 Feb 1;125(2):S3-23.
2. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nature Reviews Cancer*. 2012 Apr;12(4):252-64.
3. Fife BT, Bluestone JA. Control of peripheral T- cell tolerance and autoimmunity via the CTLA- 4 and PD- 1 pathways. *Immunological reviews*. 2008 Aug;224(1):166-82.
4. Chen DS, Mellman I. Oncology meets immunology: the cancer-immunity cycle. *immunity*. 2013 Jul 25;39(1):1-0.
5. Beatty GL, Gladney WL. Immune escape mechanisms as a guide for cancer immunotherapy. *Clinical cancer research*. 2015 Feb 15;21(4):687-92.
6. Zappasodi R, Merghoub T, Wolchok JD. Emerging concepts for immune checkpoint blockade-based combination therapies. *Cancer cell*. 2018 Apr 9;33(4):581-98.
7. Marin-Acevedo JA, Dholaria B, Soyano AE, Knutson KL, Chumsri S, Lou Y. Next generation of immune checkpoint therapy in cancer: new developments and challenges. *Journal of hematology & oncology*. 2018 Dec;11(1):1-20.
8. Robert C. A decade of immune-checkpoint inhibitors in cancer therapy. *Nature Communications*. 2020 Jul 30;11(1):1-3.
9. O'Neill RE, Cao X. Co-stimulatory and co-inhibitory pathways in cancer immunotherapy. *Advances in cancer research*. 2019 Jan 1;143:145-94.
10. Francisco LM, Sage PT, Sharpe AH. The PD- 1 pathway in tolerance and autoimmunity. *Immunological reviews*. 2010 Jul;236(1):219-42.
11. Riella LV, Paterson AM, Sharpe AH, Chandraker A. Role of the PD- 1 pathway in the immune response. *American Journal of Transplantation*. 2012 Oct;12(10):2575-87.
12. Keir ME, Butte MJ, Freeman GJ, Sharpe AH. PD-1 and its ligands in tolerance and immunity. *Annu. Rev. Immunol.* 2008 Apr 23;26:677-704.

13. Keir ME, Francisco LM, Sharpe AH. PD-1 and its ligands in T-cell immunity. *Current opinion in immunology*. 2007 Jun 1;19(3):309-14.
14. Keir ME, Liang SC, Guleria I, Latchman YE, Qipo A, Albacker LA, Koulmanda M, Freeman GJ, Sayegh MH, Sharpe AH. Tissue expression of PD-L1 mediates peripheral T cell tolerance. *The Journal of experimental medicine*. 2006 Apr 17;203(4):883-95.
15. Han Y, Liu D, Li L. PD-1/PD-L1 pathway: current researches in cancer. *American journal of cancer research*. 2020;10(3):727.
16. Zhang JY, Yan YY, Li JJ, Adhikari R, Fu LW. PD-1/PD-L1 based combinational cancer therapy: icing on the cake. *Frontiers in pharmacology*. 2020 May 15;11:722.
17. Lin X, Lu X, Luo G, Xiang H. Progress in PD-1/PD-L1 pathway inhibitors: From biomacromolecules to small molecules. *European Journal of Medicinal Chemistry*. 2020 Jan 15;186:111876.
18. Akinleye A, Rasool Z. Immune checkpoint inhibitors of PD-L1 as cancer therapeutics. *Journal of hematology & oncology*. 2019 Dec;12(1):1-3.
19. Chames P, Van Regenmortel M, Weiss E, Baty D. Therapeutic antibodies: successes, limitations and hopes for the future. *British journal of pharmacology*. 2009 May;157(2):220-33.
20. Shukla AA, Thömmes J. Recent advances in large-scale production of monoclonal antibodies and related proteins. *Trends in biotechnology*. 2010 May 1;28(5):253-61.
21. Konstantinidou M, Zarganes-Tzitzikas T, Magiera-Mularz K, Holak TA, Dömling A. Immune Checkpoint PD-1/PD-L1: Is There Life Beyond Antibodies?. *Angewandte Chemie International Edition*. 2018 Apr 23;57(18):4840-8.
22. Wang T, Wu X, Guo C, Zhang K, Xu J, Li Z, Jiang S. Development of inhibitors of the programmed cell death-1/programmed cell death-ligand 1 signaling pathway. *Journal of Medicinal Chemistry*. 2018 Sep 24;62(4):1715-30.
23. Huck BR, Kötzner L, Urbahns K. Small molecules drive big improvements in immuno-oncology therapies. *Angewandte Chemie International Edition*. 2018 Apr 9;57(16):4412-28.

24. Bojadzic D, Buchwald P. Toward small-molecule inhibition of protein–protein interactions: General aspects and recent progress in targeting costimulatory and coinhibitory (immune checkpoint) interactions. *Current topics in medicinal chemistry*. 2018 Mar 1;18(8):674-99.
25. Geng Q, Jiao P, Jin P, Su G, Dong J, Yan B. PD-1/PD-L1 inhibitors for immunoncology: from antibodies to small molecules. *Current Pharmaceutical Design*. 2017 Nov 1;23(39):6033-41.
26. Cheng B, Xiao Y, Xue M, Cao H, Chen J. Recent Advances in the Development of PD-L1 Modulators: Degradable, Downregulators, and Covalent Inhibitors: Miniperspective. *Journal of Medicinal Chemistry*. 2020 Dec 4;63(24):15389-98.
27. Shanmugam M, Narayanan K, Prasad KH, Karthikeyan D, Chandrasekaran L, Atchudan R, Chidambaranathan V. Synthesis, characterization, and antiproliferative and apoptosis inducing effects of novel s-triazine derivatives. *New Journal of Chemistry*. 2018;42(3):1698-714.
28. Manohar S, Pepe A, Gerena CE, Zayas B, Malhotra SV, Rawat DS. Anticancer activity of 4-aminoquinoline-triazine based molecular hybrids. *RSC Advances*. 2014;4(14):7062-7.
29. Kim EH, Ning B, Kawamoto M, Miyatake H, Kobatake E, Ito Y, Akimoto J. Conjugation of biphenyl groups with poly (ethylene glycol) to enhance inhibitory effects on the PD-1/PD-L1 immune checkpoint interaction. *Journal of Materials Chemistry B*. 2020;8(44):10162-71.
30. Skalniak L, Zak KM, Guzik K, Magiera K, Musielak B, Pachota M, Szelazek B, Kocik J, Grudnik P, Tomala M, Krzanik S. Small-molecule inhibitors of PD-1/PD-L1 immune checkpoint alleviate the PD-L1-induced exhaustion of T-cells. *Oncotarget*. 2017 Sep 22;8(42):72167.
31. Jain ZJ, Gide PS, Kankate RS. Biphenyls and their derivatives as synthetically and pharmacologically important aromatic structural moieties. *Arabian Journal of Chemistry*. 2017 May 1;10:S2051-66.
32. Chen T, Li Q, Liu Z, Chen Y, Feng F, Sun H. Peptide-based and small synthetic molecule inhibitors on PD-1/PD-L1 pathway: A new choice for immunotherapy?. *European Journal of Medicinal Chemistry*. 2019 Jan 1;161:378-98.

33. Li X, Shao C, Shi Y, Han W. Lessons learned from the blockade of immune checkpoints in cancer immunotherapy. *Journal of hematology & oncology*. 2018 Dec;11(1):1-26
34. Lin DY, Tanaka Y, Iwasaki M, Gittis AG, Su HP, Mikami B, Okazaki T, Honjo T, Minato N, Garboczi DN. The PD-1/PD-L1 complex resembles the antigen-binding Fv domains of antibodies and T cell receptors. *Proceedings of the National Academy of Sciences*. 2008 Feb 26;105(8):3011-6.
35. Chupak LS, Zheng X. Compounds useful as immunomodulators. Bristol-Myers Squibb Company. 2015 Oct 22;12.
36. Zak KM, Kitel R, Przetocka S, Golik P, Guzik K, Musielak B, Dömling A, Dubin G, Holak TA. Structure of the complex of human programmed death 1, PD-1, and its ligand PD-L1. *Structure*. 2015 Dec 1;23(12):2341-8.
37. Zak KM, Grudnik P, Guzik K, Zieba BJ, Musielak B, Dömling A, Dubin G, Holak TA. Structural basis for small molecule targeting of the programmed death ligand 1 (PD-L1). *Oncotarget*. 2016 May 24;7(21):30323.
38. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic acids research*. 2000 Jan 1;28(1):235-42.
39. Systemes, D. BIOVIA. "Discovery Studio Modeling Environment." *Dassault Systèmes, San Diego* (2016).
40. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
41. B.R. Brooks, C.L. Brooks, A.D. Mackerell, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, M. Karplus, CHARMM: The Biomolecular Simulation Program, *J. Comput. Chem.* 30 (2009) 1545.
42. O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J. Comput. Chem.* 31 (2010) 455–461.

43. A. Gupta, P. Sharma, B. Jayaram, ParDOCK: An All Atom Energy Based Monte Carlo Docking Protocol for Protein-Ligand Complexes, *Protein Pept. Lett.* 14 (2007) 632–646.
44. A. Soni, R. Bhat, B. Jayaram, Improving the binding affinity estimations of protein–ligand complexes using machine-learning facilitated force field method, *J. Comput. Aided. Mol. Des.* 34 (2020) 817–830.
45. N.M. O’Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison, Open Babel: An Open chemical toolbox, *J. Cheminform.* 3 (2011) 1–14.
46. P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, T.E. Cheatham, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models, *Accounts of chemical research.* ACS Publ. 33 (2000) 889–897.
47. B. Jayaram, T. Singh, G. Mukherjee, A. Mathur, S. Shekhar, V. Shekhar, Sanjeevini: a freely accessible web-server for target directed lead molecule discovery., *BMC Bioinformatics.* 13 Suppl 1 (2012) 1–13.
48. H. Li, K. Leung, MH. Wong, PJ Ballester, Improving AutoDock Vina using random forest: the growing accuracy of binding affinity prediction by the effective exploitation of larger data sets, *Molecular informatics Wiley Online Libr.* 34 (2015) 115–126.
49. CA. Lipinski Lead-and drug-like compounds: the rule-of-five revolution. *Drug discoverytoday:Technologies.*1(2004):33741.
50. Prival MJ. Evaluation of the TOPKAT system for predicting the carcinogenicity of chemicals. *Environmental and molecular mutagenesis.* 2001;37(1):55-69.
51. S. Pronk, S. Páll, R. Schulz, P. Larsson, R. Apostolov, MR. Shirts, JC Smith, PM Kasson, D. Van Der Spoel, B. Hess, GROMACS 4.5: a high-throughput and highly parallel open-source molecular simulation toolkit. *Bioinformatics.*29(2013):845-54.
52. W. Huang, Z. Lin, W.F. Van Gunsteren, Validation of the GROMOS 54A7 force field with respect to β -peptide folding, *J. Chem. Theory Comput.* 7 (2011) 1237–1243.

53. J. Zielkiewicz, Structural properties of water: Comparison of the SPC, SPCE, TIP4P, and TIP5P models of water, *J. Chem. Phys.* 123 (2005).
54. AW Schuttelkopf, DM Van Aalten, PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr. D* 60(2004):1355-63.
55. M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, *J. Applied phy.* 52 (1981) 7182.
56. U. Essmann, L. Perera, ML. Berkowitz, T. Darden, H. Lee, LG Pedersen, A smooth particle mesh Ewald method. *J. chem. phy.* 103(1995):8577-93.
57. B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: A Linear Constraint Solver for Molecular Simulations, *J Comput Chem.* 18 (1997) 14631472.
58. S. Miyamoto, P.A. Kollman, Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models, *J. Comput. Chem.* 13 (1992) 952–962.
59. Turner PJ. XMGRACE, Version 5.1. 19. Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR. 2005;2.
60. C.C. David, D.J. Jacobs, Principal component analysis: A method for determining the essential dynamics of proteins, *Methods Mol. Biol.* 1084 (2014) 193–226.
61. S. Genheden, U. Ryde. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities, *Expert opinion on drug discovery* 10 (2015) 449–461..
62. R. Kumari, R.Kumar, g_mmpbsa A GROMACS Tool for High-Throughput MM-PBSA Calculations, *J. Chem. Inf. Model ACS Publ.* 54 (2014) 1951–1962.
63. Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. *BMC biology.* 2011 Dec;9(1):1-9.
64. Seeliger D, De Groot BL. Conformational transitions upon ligand binding: holo-structure prediction from apo conformations. *PLoS computational biology.* 2010 Jan 8;6(1):e1000634.

65. E. Durham, B. Dorr, N. Woetzel, R. Staritzbichler, J. Meiler, Solvent accessible surface area approximations for rapid and accurate protein structure prediction, *J. Mol. Model.* 15 (2009) 1093–1108.
66. Chen D, Oezguen N, Urvil P, Ferguson C, Dann SM, Savidge TC. Regulation of protein-ligand binding affinity by hydrogen bond pairing. *Science advances.* 2016 Mar 25;2(3):e1501240.
67. E. Papaleo, P. Mereghetti, P. Fantucci, R. Grandori, L. De Gioia, Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: the myoglobin case. *J. Mol. Graph. Model.* (2009):889-99.
68. Wang C, Greene DA, Xiao L, Qi R, Luo R. Recent developments and applications of the MMPBSA method. *Frontiers in molecular biosciences.* 2018 Jan 10;4:87.

List of Publications

- **Paperwork included in the thesis:**

1. **Banerjee S**, Prasad R, Sen P. In Silico Mutation and Binding Studies of Human FVIIa γ -Carboxyglutamic Acid-Domain to Endothelial Protein C Receptor: A Molecular Dynamics Simulation Approach. ACS Omega. 2019 Jan 8;4(1):496-508.
2. **Banerjee S**, Sen P. A molecular dynamics simulation study to elucidate the effect of Cholesterol and Tissue Factor Palmitoylation on TF-FVIIa-FXa ternary complex in POPC-POPS-CHOL lipid environment. (Manuscript under revision)
3. **Banerjee S**, M Yeshwanth, Prabhu D, Sekar K, Sen P. Virtual Screening of novel class of inhibitors against Kringle 2 domain of Tissue Plasminogen Activator and SP domain and Kringle 1 domain of Plasminogen having antifibrinolytic property. (Manuscript under review)
4. **Banerjee S**, Mandal S, Yeshwanth M, Paul S., Sen, P. Design, synthesis, and evaluation of compounds having potential anti-cancer activity against programmed death-ligand 1 (PD-L1) using combined molecular docking, molecular dynamics simulation, and biological approaches. (Manuscript under preparation)

- **Paperwork not included in the thesis:**

1. Prasad R, **Banerjee S**, Sen P. Contribution of allosteric disulfide in the structural regulation of membrane-bound tissuefactor-factorVIIa binary complex. Journal of Biomolecular Structure and Dynamics. 2018 Nov 13.
2. Chatterjee A, Paul S, **Banerjee S**, Yeshwanth M, Sen P. Sunshinamide increases cellular ROS levels by targeting TrxR1 and Gpx4 to ER stress and ferroptosis mediated cell death in cancer cells. (Manuscript under preparation)

• **Book chapters not included in the thesis:**

1. Datta D, **Banerjee S**, Ghosh A, Banerjee Mustafi S, Sen P, Raha S. Involvement of Heat Shock Protein 70 (Hsp70) in Gastrointestinal Cancers. In *HSP70 in Human Diseases and Disorders 2018* (pp. 71-91). Springer, Cham.
 2. Mohan A, **Banerjee S**, Sekar K. Role of Advanced Computing in the Drug Discovery Process. *Innovations and Implementations of Computer Aided Drug Discovery Strategies in Rational Drug Design 2021* (pp. 59-90). Springer, Singapore.
-