8-2015

Exploring Structure-Dynamics-Function Relationship in Proteins, Protein: Ligand and Protein: Protein Systems through Computational Methods

Karan Pal Kapoor
University of Tennessee - Knoxville, kkapoor@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Karan Pal Kapoor entitled "Exploring Structure-Dynamics-Function Relationship in Proteins, Protein: Ligand and Protein: Protein Systems through Computational Methods." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Life Sciences.

Jerome Baudry, Major Professor

We have read this dissertation and recommend its acceptance:

Jeremy Smith, Elizabeth Howell, Robert Hinde, Hong Guo

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Exploring Structure-Dynamics-Function Relationship in Proteins, Protein: Ligand and Protein: Protein Systems through Computational Methods

**A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville**

**Karan Pal Kapoor
August 2015**

DEDICATION

This dissertation is dedicated to my parents and my dear friend Sam.

ACKNOWLEDGEMENTS

ABSTRACT

The study focuses on understanding the dynamic nature of interactions between molecules and macromolecules. Molecular modeling and simulation technologies are employed to understand how the chemical constitution of the protein, specific interactions and dynamics of its structure provide the basis of its mechanism of function. The structure-dynamics-function relationship is investigated from quantum to macromolecular-assembly level, with applications in the field of rationale drug discovery and in improving efficiency of renewable sources of energy. Results presented include investigating the role of dynamics in the following:

1) In interactions between molecules: analyzing dynamic nature of a specific non-covalent interaction known as "anion-π [pi]" in RmlC protein.

2) In interactions between molecules and macromolecules: defining the structural basis of testosterone activation of GPRC6A.

3) In disrupting the function using specific substrate interactions: incorporating protein dynamics and flexibility in structure-based drug-discovery approach targeting the prothrombinase coagulation complex.

4) In interactions between macromolecules: elucidating the protein-protein binding and dynamics of electron-transport proteins, Ferrodoxin and Cytochrome c6, with Cyanobacterial Photosystem I.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

PDB   Protein Data Bank

QM   Quantum mechanical

BF   Benzene-Formate

KM   Kitaura-Morokuma

MD   Molecular Dynamics

PME   Particle Mesh Eswald

STAAR   STatistical Analysis of Aromatic Rings

PMF   Potential of mean force

RMSD   Root mean square deviation

Ocn   Osteocalcin

L-Arg   L-Arginine

Egr-1   Early growth response protein 1

GPCR   G-protein coupled receptor

GPRC6A   G-protein coupled receptor family C group 6 member A

T   Testosterone

CasR   Calcium sensing receptor

DHT   Dihydrotestosterone

mGluR   Metabotropic glutamate receptor

PTase   Prothrombinase

PT   Prothrombin

FXa   Factor Xa

FVa	Factor Va
SAR	Structure activity relationships
PSI	Photosystem I
PSII	Photosystem II
Fd	Ferredoxin
Cytc6	Cytochrome c6
PC	Plastocyanin
FTT	Fast-fourier transform

INTRODUCTION

Prologue

Proteins are of critical importance for all living organisms. They perform an array of functions, including catalyzing metabolic reactions, maintaining structural integrity of the cell, transporting molecules from one location to another, to name a few. Proteins have a chemical nature defined by its amino acid sequence. This unique sequence of residues fold into unique three-dimensional structures, which defines not only its size and shape but also its function (1). The three-dimensional structures can form complementary sites for binding other molecules, for example, active site in the case of reactions catalyzed by enzymes (2), and protein binding sites on protein surface in formation of oligomers (3). Hence, protein sequence and its structure can be compared to its fingerprint, where different structures represent different personalities and functions of the proteins.

The function of a protein is also modulated by other molecules (4). Recognition and binding of a substrate to the protein involves specific interactions between protein residues and the substrate molecule. Covalent and non-covalent interactions play a critical role in maintaining the structural integrity of the protein itself (5,6). These interactions can be described both at the quantum scale with high accuracy, and by using semi-empirical and classical methods, that allow certain approximations. Hence, in order to understand the behavior of protein, it is important to understand and quantify all possible interactions that control protein structural integrity and drive its function (7).

Biophysical methods like crystallography and NMR provide a static picture of the proteins (8,9). But, it is important to recognize that proteins are 'alive' molecules (10). They are flexible at room temperatures and can undergo changes in different interactions, leading to changes in its three-dimensional structure and function. This flexibility and stochastic fluctuations lead to

2

different 'active' and 'non-active' states of the protein, directing reactions preferentially along functionally important pathways (11). Thus, in order to exploit this quality of biological molecules, for example in medicine and renewable power generation, computational methods that simulate molecular structure and interactions can be employed in exploring the structures (and interactions) that are unresolved through experiments (12).

Scope of Study

Protein dynamics are central to understanding the structure-function correlation in biological systems (13). The study presented here characterizes the dynamic nature of interactions in proteins, protein: ligand and protein: protein systems, and in understanding the structural and dynamical properties of macromolecular systems that furnish their biological function using computational methods (14). In particular, this research describes a new non-covalent interaction in proteins at the quantum and macromolecular level, the structural basis of substrate interactions/activation and function is explored in a membrane protein, structure-based drug discovery methods are extended to include the inherent flexibility of biological molecules in order to make better predictions in protein-ligand binding, and protein-protein interactions and dynamics are studied at the macromolecule-systems level.

Understanding and quantifying different interactions in biological systems opens the avenue for designing new proteins/peptides and molecules with novel properties. Modeling interactions between protein and small molecules forms an important area of research in discovering new drugs for life-threatening diseases (15). Also, understanding the interactions between different macromolecules, especially those involved in plant/bacteria photosystem, can be used to modify and improve the efficiency of these systems for electricity production (16).

Molecular Interactions

Atom and molecules constitute matter and form a complex network of weak and strong interactions conferring specific properties to the matter. The same is true for biological molecules. This interplay between strong and weak interactions in a macromolecule like protein results in a unique structure for that protein. The primary structure of a protein is determined by the bonded interactions between its amino acids. Secondary structure like alpha helix and beta strands are defined by patterns of non-bonded interactions, like hydrogen bonds, between the main-chain atoms. Anion-π are an emerging non-bonded interactions that are present in biomolecules between negatively charged amino acid side chains and the positively charged ring edge of aromatic groups (17,18). These interactions have also been beneficially exploited in field structure-based drug-discovery and in development of effective inhibitors (19).

CHAPTER 1 introduces the theory of anion-π interactions and recent advances made in this field. The study aims at investigating the chemical, structural and thermodynamic characteristics of the anion-π interactions in RmlC protein (20), and characterizes the time- and temperature dependent dynamic stability of these interactions using molecular dynamics (MD) simulations, potential of mean force (PMF) and semi-empirical quantum energy calculations.

Structure and function

The structure of a protein is determined by its primary sequence of amino acids, which is encoded by the sequence of nucleotides in the gene (DNA). This unique structure of the protein and its active site allows the protein to interact with and bind different ligands, conferring it a unique function. So, in order to better understand the biological function of proteins, it is

imperative to correctly model the binding, interactions as well as the structural changes involved when a ligand binds to the protein.

G protein–coupled receptors (GPCRs) are heptahelical membrane receptors, that sense molecules outside the cell, activating inside signal transduction pathways and, ultimately, cellular responses (21). Steroid hormone Testosterone (T) plays a key role in development of reproductive tissues, muscles and bones (22), and has been shown to lead to the downstream signaling in a newly characterized Family C GPCR, GPRC6A (23). **CHAPTER 2** deals with providing the structural evidence for the T binding to GPRC6A, using homology modeling and molecular docking, to gain additional insights into this role of T in the activation of GPRC6A.

'Dynamics' bridges the gap

Proteins are not static entities. At room temperature, they populate an ensembles of conformations. Transitions between these states occur on a variety of time scales (ns to s), linking these ensemble of structures to functionally relevant phenomena such as allosteric signaling and enzyme catalysis (24,25). As proteins have this dynamic nature, changes in its structure also complement the interplay between different interactions that maintain both its structural integrity and function.

The prothrombinase (PTase) enzymatic complex, consisting of the enzyme factor Xa (FXa), and a protein cofactor, factor Va (FVa), catalyzes the cleavage of prothrombin (PT) leading to the formation of thrombin and clot-formation (26,27). The goal in **CHAPTER 3** is to target FXa within the PTase complex, but instead of seeking another active-site directed inhibitor that have the serious liability of excessive bleeding (28), MD simulations and ensemble-based high-throughput virtual screening methods are used to identify compounds that bind to FXa and alter

the interaction between FXa and its cofactor FVa. These compounds have the potential to down-modulate the PTase activity in an allosteric and non-competitive manner.

Macromolecular assemblies

The functions of macromolecules are inter-connected. All the pieces in the puzzle need to come together to make the system work. Some small molecules can interact and bind to a protein, leading to its downstream signaling, activating/deactivating several other members in the same pathway (24). The final function or product formation is the product of this long assembly line. Treating individual elements of the same pathway or system separately only provides an isolated view of the whole process. As the biological system are dynamic and complex, their behavior can be hard to predict from the properties of individual subunits. So, in order to better understand the whole process, the complete system needs to be modeled.

Photosystem is an interesting biological system that converts light energy into chemical energy through a series of redox reactions between two large, multi-subunit proteins Photosystem II (PSII) and Photosystem I (PSI) (29), and electron-transport proteins like Ferredoxin (Fd) and Cytochrome c6 (Cytc6) (30,31). **CHAPTER 4** aims at developing new structural models between the stromal and luminal domains of PSI and Fd and Cytc6, respectively, using rigid-body molecular docking. MD simulations are being utilized to examine the atomic-level intermolecular interactions and dynamics at the protein-protein interface in these proteins to further understand and illustrate the small changes taking place in the transfer of electrons between the separate subunits.

LIST OF REFERENCES

1.   Murray RF, Harper HW, Granner DK, Mayes PA, Rodwell VW (2006). Harper's Illustrated Biochemistry. New York: Lange Medical Books/McGraw-Hill. pp. 30–34

2.   EBI External Services (2010-01-20). "The Catalytic Site Atlas at The European Bioinformatics Institute". Ebi.ac.uk. Retrieved 2011-01-16.

3.   Mayssam H. Ali and Barbara Imperiali. Protein oligomerization: How and why. Bioorganic & Medicinal Chemistry **13** (2005) 5013–5020

4.   Daniel Atkinson. Modulation of Protein Function. - Nature. Volume XIII. Elsevier, Dec 2, 2012

5.   Campbell, Neil A.; Brad Williamson; Robin J. Heyden (2006). *Biology: Exploring Life*. Boston, Massachusetts: Pearson Prentice Hall.

6.   Anslyn, Eric (2004). *Modern Physical Organic Chemistry*. Sausalito, CA: University Science.

7.   Smith AJ, Zhang X, Leach AG, Houk KN (Jan 2009). "Beyond picomolar affinities: quantitative aspects of noncovalent and covalent binding of drugs to proteins". *Journal of Medicinal Chemistry* **52** (2): 225-33

8.   Gregory A. Petsko, Dagmar Ringe. Protein Structure and Function. New Science Press, Jan 1, 2004 - Science.

9.   Ayyalusamy Ramamoorthy. Beyond NMR spectra of antimicrobial peptides: Dynamical images at atomic resolution and functional insights. Solid State Nuclear Magnetic Resonance**. 35**(4), July 2009, Pages 201–207

10.   Katherine Henzler-Wildman and Dorothee Kern. Dynamic personalities of proteins. *Nature* **450**, 964-972 (13 December 2007).

11.   Van Holde KE, Mathews CK (1996). Biochemistry. Menlo Park, California: Benjamin/Cummings Pub. Co., Inc. pp. 368–75.

12.   Stewart A. Adcock and J. Andrew McCammon. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. Chem Rev. 2006 May; **106**(5): 1589–1615.

13.   Cristian Micheletti. Comparing proteins by their internal dynamics: Exploring structure–function relationships beyond static structural alignments. Physics of Life Reviews Volume **10**, Issue 1, March 2013, Pages 1–26

14.   Pramod C Nair and John O Miners. Molecular dynamics simulations: from structure function relationships to drug discovery. In Silico Pharmacol. 2014; **2**: 4.

15.   Sliwoski G, Kothiwale S, Meiler J, Lowe EW Jr. Computational methods in drug discovery. Pharmacol Rev. 2013 Dec 31; **66**(1):334-95.

16.   Nguyen K, Bruce BD. Growing green electricity: progress and strategies for use of photosystem I for sustainable photovoltaic energy conversion. Biochim Biophys Acta. 2014 Sep;**1837**(9):1553-66.

17.   Jackson, M. R., Beahm, R., Duvvuru, S., Narasimhan, C., Wu, J., Wang, H. N., Philip, V. M., Hinde, R. J., and Howell, E. E. (2007) A preference for edgewise interactions between aromatic rings and carboxylate anions: the biological relevance of anion-quadrupole interactions. *The journal of physical chemistry. B* **111**, 8242-8249

18.   Philip, V., Harris, J., Adams, R., Nguyen, D., Spiers, J., Baudry, J., Howell, E. E., and Hinde, R. J. (2011) A survey of aspartate-phenylalanine and glutamate-phenylalanine interactions in the protein data bank: searching for anion-pi pairs. *Biochemistry* **50**, 2939-2950

19. Krieger, I. V., Freundlich, J. S., Gawandi, V. B., Roberts, J. P., Gawandi, V. B., Sun, Q., Owen, J. L., Fraile, M. T., Huss, S. I., Lavandera, J. L., Ioerger, T. R., and Sacchettini, J. C. (2012) Structure-guided discovery of phenyl-diketo acids as potent inhibitors of M. tuberculosis malate synthase. *Chemistry & biology* **19**, 1556-1567

20. Christendat, D., Saridakis, V., Dharamsi, A., Bochkarev, A., Pai, E. F., Arrowsmith, C. H., and Edwards, A. M. (2000) Crystal structure of dTDP-4-keto-6-deoxy-D-hexulose 3,5-epimerase from Methanobacterium thermoautotrophicum complexed with dTDP. *The Journal of biological chemistry* **275**, 24608-24612

21. Trzaskowski, B; Latek, D; Yuan, S; Ghoshdastider, U; Debinski, A; Filipek, S (2012). "Action of molecular switches in GPCRs--theoretical and experimental studies". *Current medicinal chemistry* **19**(8): 1090–109.

22. Mooradian AD, Morley JE, Korenman SG (February 1987). "Biological actions of androgens". *Endocr. Rev.* **8** (1): 1–28.

23. Wellendorph P, Brauner-Osborne H. Molecular cloning, expression, and sequence analysis of GPRC6A, a novel family C G-protein-coupled receptor. Gene 2004; **335**:37-46

24. Bu Z, Callaway DJ (2011). "Proteins move! Protein dynamics and long-range allostery in cell signaling". *Advances in Protein Chemistry and Structural Biology* **83**: 163–221.

25. Fraser JS, Clarkson MW, Degnan SC, Erion R, Kern D, Alber T (Dec 2009). "Hidden alternative structures of proline isomerase essential for catalysis". *Nature* **462** (7273): 669–673.

26. Kane WH, Davie EW. Blood coagulation factors V and VIII: structural and functional similarities and their relationship to hemorrhagic and thrombotic disorders. *Blood*. 1988;**71**(3):539-555.

27. Davie EW, Fujikawa K, Kisiel W. The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry*. 1991;**30**(43):10363-10370.

28. Alexander JH, Lopes RD, James S, et al. Apixaban with antiplatelet therapy after acute coronary syndrome. *N Engl J Med*. 2011;**365**(8):699-708.

29. Jagannathan, B; Golbeck, JH (2009). "Photosynthesis:Microbial". *Encyclopedia of Microbiology, 3rd Ed*: 325–341.

30. Muhlenhoff, U. *et al.* Characterization of a redox-active cross-linked complex between cyanobacterial photosystem I and its physiological acceptor flavodoxin. *The EMBO journal* **15**, 488-497 (1996)

31. Hope A.B. (2000): *Electron transfers amongst cytochrome f, plastocyanin and photosystem I: kinetics and mechanisms*. In: Biochim Biophys Acta. **1456**(1) S. 5-26.

Chapter 1

ANION-Π NETWORKS IN PROTEINS ARE

HIGHLY DYNAMIC

Research described in this chapter is reproduced from a manuscript prepared by Karan Kapoor, Amit Upadhyay, Robert Hinde, Elizabeth Howell and Jerome Baudry.

My contribution in the paper includes (1) Compiling and interpretation of the literature (2) Designing and carrying out all computation research (except solvent analysis) (3) Understanding the literature and interpretation of the results (4) Preparation of tables and figures (5) Writing and editing.

Abstract

The dynamics of anion-π, (or anion-quadrupole) interactions formed between negatively charged (Asp/Glu) and aromatic (Phe/Trp/Tyr) side chains is computationally characterized in RmlC (PDB: 1EP0), a homodimeric epimerase. Empirical force field-based molecular dynamics simulations shows anion-π pairs and triplets (anion-anion-π, anion-$\pi-\pi$) being formed in conformations sampled by the protein, extending the anion-π interactions beyond those observed in the crystal structure. The combined energies from the anion-π interactions sampled during the protein trajectory may provide a significant contribution to the overall stability of the protein, with an average of -1.6 kcal/mol per pair. These results suggest that near-planar anion-π pairs can exist – sometimes transiently - both at the active site and the dimer interface, which may play a role in maintaining the structural stability and function of the protein.

Introduction

Non-covalent interactions play a constitutive role in maintaining the structural integrity of proteins, as well as in protein-protein and protein-ligand interactions. In addition to the well-known and well-described hydrogen bonds, van der Waals interactions and hydrophobic interactions, other non-covalent bond interactions have been suggested to contribute to the stabilization of protein structures and of protein-ligand complexes Our current analysis and understanding of protein structure, protein-protein and protein-ligand interactions is in large part based on our understanding of non-bonded interactions between atoms, functional groups or molecular domains. Among non-bonded interactions, hydrogen bonds, van der Waals interactions, hydrophobic interactions, and ion pairs are well described and well understood. However, other types of non-bonded interactions have been suggested to contribute to the stabilization of protein structures and of protein-ligand interactions (1-7). One of these is the anion-π interactions between an anion and the aromatic ring. These interactions have previously been studied with a focus on electron-deficient π rings, by incorporating strong electron-withdrawing substituents such as fluorobenzene derivatives, fluoro-s-triazine, and tetrafluoroethene, and negatively charged molecules and ions, as well as in the case of metallacycles (8-11). The positive charge on the aromatic ring edge arises from the quadrupole moment of the ring, leading to the anion-quadrupole or anion-π interaction with the negatively charged anion. Wheeler et al. have recently provided an alternative explanation for the origin of anion-π interactions involving substituted benzenes and N-heterocycles and suggested that the positive electrostatic potentials and molecular quadrupole moments characteristic of π-acidic azines, which underlie the ability of these rings to bind anions above their centers, arise from the position of nuclear charges, not changes in the π-electron density distribution. (12,13).

Anion-π interactions have also been found to be present in biomolecules between negatively charged amino acid side chains and the positively charged ring edge of aromatic groups. Small molecule analogs (benzene/phenol/indole interacting with formate) have been used to characterize the pairwise interaction of aromatic amino acids (phenylalanine, tyrosine, tryptophan) with anionic amino acids (aspartic and glutamic acids) at the MP2 level of quantum mechanical (QM) theory (14). The study shows that the overall stabilization associated with an anion-π interaction to be potentially substantial, as large -9.5 kcal/mol for a benzene-formate (BF) planar dimer at van der Waals contact distance where the interaction is expected to be at a maximum.

Studies by Kallenbach et al. using short α-helical peptides with glutamate/ phenylalanine pairs positioned at i and i+4 spacing indicate that this pairwise interaction provides ~0.5 kcal/mol additional stability to the helix (15-17). Breberina et al. find a significant percentage of anion-π interacting residues located as stabilization centers in Sm/LSm proteins, with energies in the range -2 to -9 kcal mol-1 calculated by CHARMM22 (18). In the specific context of intermolecular binding interactions, Schwans et al. find two active site phenylalanines (F54, F116) help hold the general base (D38) in position for catalysis in ketosteroid isomerase (19). They also identified 46 enzymes that appear to use an aromatic group to position Asp or Glu so that it can act as a general base in the mechanism, suggesting an important functional role in biology of anion-π interactions. These interactions have also been beneficially exploited in structure-based drug-discovery, where Sacchettini et al. developed effective inhibitors targeting malate synthase in Mycobacterium tuberculosis, where anion–π interactions play an important role (20).

The energetics and structural properties of anion-π interactions present in the Protein Data Bank (PDB) have been investigated for Phe-Asp and Phe-Glu pairs (21). An average of ~3.4 anion-π pairs per protein structure in the PDB were identified, and Kitaura-Morokuma (KM) QM analysis

indicates that 81% of these pairs exhibit stabilizing energies as low as -8.7 kcal/mol for pairs present at near-planar angles. Nearly planar aromatic- anionic amino acid pairs were also found to occur more often than expected from a random angular distribution, while axial aromatic-anionic pairs occur less often than expected from a random distribution, suggesting a contribution of anion–π interactions to protein stability.

A number of proteins that exhibit clusters of such interactions have been identified (21). Some of these clusters involve relatively long and linear chains of anion-π interactions, or several anions clustering around a single resonant side chain. Pairs occur in α helices as well as in β strands and sometimes form "anion-π networks" involving several amino acids. In crystal structure of one of these proteins, RmlC, a homodimeric deoxythymidine diphosphate epimerase (dTDP-4-keto-6-deoxy-d-hexulose 3,5-epimerase; PDB code 1EP0) a network of five anion-π pairs (in the monomer) was identified involving seven neighboring residues.

There is hence a growing body of results, theoretical and experimental, that points to the structure, thermodynamic and functional importance of anion-π interactions in biomolecules. The previous work has essentially focused on static structures however, and essentially focused on the enthalpy of anion-π interactions. The present work aims at investigating the dynamics characteristics of the anion-π network in RmlC using Molecular dynamics (MD) simulations. In particular we investigate whether the Phe-Asp or Phe-Glu anion-π interactions identified in the crystal structure remain stable on the free energy surface sampled by the protein, or whether these interactions are 'drowned out' by solvent interactions and/or otherwise larger scale protein dynamics. Inversely, we investigate if anion-π interactions that would not been observed in the crystal structure appear in the dynamics, because, for instance, crystal packing forces could (22)

'drown out' otherwise relatively weak anion-π interactions initially present at non-ideal geometries.

Materials and methods

Anion-π cluster in RmlC

Deoxythymidine diphosphate (dTDP)-4-keto-6-deoxy-d-hexulose 3,5-epimerase (RmlC; PDB ID- 1EP0), a homodimer with each monomer consisting of 183 residues, is involved in the biosynthesis of dTDP-l-rhamnose which is an essential component of the bacterial cell wall (23). The dimer interface is formed by an extensive set of hydrophobic and electrostatic contacts. Charged residues (Asp-24, Arg-26, Glu-35, Asp-50, Glu-52, Arg-61, Arg-76, Lys-134) and aromatic residues (Tyr-28, Phe-33, Phe47) are present at this interface. The substrate binding site is also lined with a number of charged residues (Asp-24, Arg-26, Glu-31, Glu-52, Arg-61, His-64, Lys-73, Asp-84, His- 120, Asp-144, Lys-171, Asp-172) and aromatic residues (Phe-29, Phe-122, Tyr-133, Tyr- 139, Trp-175).

The crystal structure was analyzed with the STAAR (STatistical Analysis of Aromatic Rings) program (24). STAAR locates phenylalanine rings and determines their centers of mass (CM). For each aromatic ring, STAAR then calculates the distance r between the ring's center of mass and the center of charge (CC) for the Glu or Asp carboxylate group, as well as the angle θ between the plane of the ring and the vector connecting the ring center of mass with this center of charge. The cutoff criteria of Gallivan and Dougherty (25) was adopted i.e., those pairs possessing a distance r of ≤ 7 Å were chosen for analysis to eliminate cases in which a water molecule could fit between the two residues and diminish the interaction energy. The Phe and Glu or Asp anion-π

pairs identified, both intramolecular and intermolecular, were then parsed from the STAAR output for their respective distance and angles using perl scripts.

MD Simulations

The dynamics of the RmlC dimer was characterized in molecular dynamics (MD) simulations. The asymmetric unit representing the dimer was generated using the biological assembly of the RmlC X-ray crystal structure (PDB code: 1EP0) (23). This PDB entry has a 1.50 Å resolution, and no gaps in the structure. The structure was protonated according to estimated PkAs using the Protonate-3D facility in MOE-2010 (26), with an ion concentration of 0.1 mol/L in the Generalized Born (GB) electrostatics model. Crystallographic water molecules were retained and a periodic solvation cube of water molecules was created using the Solvate facility in MOE-2010. Twenty-six positive sodium ions were added to neutralize the system and the final system consisted of 47892 atoms, with 14,002 water molecules.

In previous studies the correlation between the CHARMM-22 force-field (27) and ab initio interaction energies has shown that the ab initio anion-π interaction energies can be reproduced correctly by the CHARMM empirical force field for basic functional groups, albeit with a magnitude that underestimates the ab initio results by up to $\sim$50% (14,21). The system was energy-minimized to a gradient of 10-5 RMS kcal/mol/Å2 using a 8Å -10Å nonbonded cutoff distance. NAMD2 input files were generated using the Dynamics facility in MOE-2010. A time-step of 2 fs/step was used for all calculations and water molecules were held rigid. A constant temperature of 300K was maintained throughout the simulation using Langevin dynamics. Particle Mesh Eswald (PME) was used for calculation of the electrostatic interactions with the periodic boundary conditions.

MD simulation of the complete system using NVT ensemble was performed on the University of Tennessee High Performance Computing architecture (Newton), using the NAMD2 simulation engine (28). A 100 ps initial equilibration was ran followed by a 150 ns production run with structures saved at every 250 timesteps (0.5ps), giving a total of 300,200 trajectory frames including 200 frames (100ps) for equilibrium of the system. The resulting trajectories were analyzed for convergence of temperature and RMSD using perl scripts to parse data, Prody v1.2 (29) and Gnuplot v4.6.

Anion-π pair analysis in MD simulation

The trajectory files in DCD format from the MD production run were converted to PDB format using Prody v1.2, generating a total of 300,000 structures, i.e., one for each trajectory frame. Perl scripts were used to modify these files to make them compatible with STAAR. The names of the amino acid residues, residue numbers, atom names and chain-ID were parsed from the starting PSF file using perl scripts and added to the generated PDB files to maintain consistency in residue naming and numbering. These snapshots were parsed with the STAAR program for Phe-Glu or Asp pairs with a threshold distance of 11 Å. The anion-π pairs present in the crystal structure were parsed from the STAAR output for their respective distance and angles using perl scripts. Distance and angle plots were generated using Gnuplot-4.6.

The STAAR output was again parsed for Phe-Glu or Asp pairs not seen in the crystal structure but that may form in the monomer during the course of the MD simulation, with a threshold distance of 6 Å (6 Å or below at any point during MD) between the pairs using perl scripts. The output was also parsed for anion-π pairs present between the two monomers (intermolecular) with the same threshold distance.

Solvent Analysis

To quantify the amount of water molecules that can possibly interact and shield the interactions between the partners of the anion-π pairs identified by STAAR, the center of mass for Phe and the center of charge for Asp/Glu were obtained by parsing the STAAR output files. The number of water molecules for a given anion-π pair was calculated according to two different criteria: (i) number of water molecules within 4 Å of the center of mass of Phe and 4 Å of the center of charge of the anion, and (ii) water molecules in the region located between the Phe residue and the anion for a given anion-π pair. The latter was calculated by first determining the coordinates of the midpoint between the center of mass of Phe residue and the center of charge of the anion. The distance threshold was chosen as half of the distance between the center of mass of Phe residue and the center of charge of the anion.

Semi-empirical quantum energy calculations

The STAAR program was used to perform the quantum energy calculation as described in earlier work (24). Briefly, the approach is to reduce the Phe and Glu or Asp pairs identified to benzene-formate (BF) pairs, followed by addition of hydrogens to the BF pairs using ProDrg2 (30). As the pKa values for Asp and Glu are low (i.e., 3.5 - 4.5) (31), Asp and Glu are sassumed to be ionized. The resulting file are converted from PDB coordinates to an xyz format using Babel v2.3.3 (32) and the appropriate input files for PC GAMESS (June 1999 version) (33) generated. PC GAMESS running a Kitaura-Morokuma (KM) energy decomposition analysis (34,35) was used for interaction energy calculation between the anion-π pairs. GAMESS energy calculations were performed on these pairs for every 200th point i.e. at a time interval of 100ps. Hence,

interaction energies were calculated for a total of 1500 snapshots for each pair in the MD trajectory. Perl scripts, Excel spreadsheets and Macros were used to sort through the data.

Potential of Mean force (PMF)

The PMF describes the free energy changes of a system as a function of reaction coordinates of the system (36). Here, the distances 'R' and angles between the anion-π pairs were used to generate angle-distance frequency distribution over the entire trajectory. For normalizing this frequency distribution, 1 million points were randomly distributed on a unit sphere using python scripts, and the distribution of points at different polar angles (0 to 180°) was calculated. Figure S1 and S2 show the resulting Gaussian distribution of these points. The area of sphere (and solid angle) is proportional to R2. Accordingly, the distribution of points with same probability density function and at different distances (different sphere radii) was also calculated. Figure S3 shows the resulting exponential distribution of these points. The random distribution of points based on angles and distances was used to normalize the angle-distance frequency distribution over the entire trajectory.

The normalized frequency distribution was used to generate the corresponding free energy maps for the anion-π pairs investigated here. For Potential of mean force (PMF) calculations the following equation was used:

$$\Delta G = -kT*\ln(\text{distribution ratio})$$

With k = 0.5961 kcal/mol at T=300K.

Results

Anion-π cluster in the RmlC crystal structure

Anion-π interactions in RmlC crystal structure were found to consist of a total of ten anion-π pairs, eight intramolecular and two intermolecular, comprising two clusters of four and five anion-π pairs at the dimer interface and in the monomer respectively, shown in Figure 1A, and their respective distance and angles are given in Table 1. The intramolecular clusters in the monomers each consist of interactions between seven residues (Figure 1B). The cluster shows the presence of several aromatic groups surrounding one anion, Phe122 and Phe112 around Asp84 and Phe38 and Phe4 around Glu111 respectively, forming possible π-π-anion triplets, Anions, Glu-31 and Glu-111, can also be seen to cluster around one aromatic group, Phe-4, forming an anion-π-anion triplet. Out of the residues forming this cluster, two residues- Glu31 and Phe122, are present in the active site of the protein.

The intermolecular cluster at the dimer interface consists of interactions between eight residues (Figure 1C). The cluster also contains possible π-π-anion (Phe33-Phe47-Asp50) and anion-π-anion (Glu35-Phe33-Asp50) triplets. Two symmetrical intermolecular pairs, Phe33-Asp50 and Asp50-Phe33 (Phe and Asp residues present in separate monomers), are present in this cluster.

MD Simulations and Anion-π pair analysis

Figure S4 shows the RMSD plot for the protein backbone over the time course of the trajectory. Based on the RMSD plot, the RmlC homodimer is stable, stabilizing after the first 2 ns of the production run around the value of 3.5 Å. The ten anion-π pairs present in the crystal

structure are found to sample different sidechain conformations in the MD snapshots, as shown in Figure S5.

Distance vs time and angle vs time plots for the eight intramolecular pairs present in the crystal structure are shown in Figures S6A and S6B respectively, and the minimum, maximum and average distances and angles for each pair are given in Table 2. These pairs sample distance between 4 - 8 Å and angles 0 - 60°. The minimum average distances and angles are shown by Phe112- Asp84, whereas the maximum distances and angles are shown by Phe47- Asp50 and Phe118- Asp88 respectively. Noticeably, the pair Phe122- Asp84 samples planar configurations intermittently throughout the trajectory. Distance vs time and angle vs time plots for the two intermolecular pairs present in the crystal structure are shown in Figures S6C and S6D respectively, and the minimum, maximum and average distances and angles are given in Table 2. These two anion-π pairs, Phe33-Asp50 and Asp50-Phe33 (first and second residues present in separate monomers), sample distances between 4 - 6 A and angles between 0 - 40°.

In addition to these pairs, a total of seven intra-molecular Phe-Glu or Asp pairs were seen appearing during the MD simulation and not in the starting crystal structure, showing distances < 6 Å at some point in the trajectory (Figure 2A). Distance vs time and angle vs time plots are shown in Figures 2B and 2C respectively, and the minimum, maximum and average distances and angles sampled by each pair, along with the distances and angles seen in the crystal structure are given in Table 3. The pairs contain two anion-π-anion triplets, with one of these (Glu31-Phe29-Asp24)present at the active site of the protein, and the other (Glu116-Phe113-Glu44) present in both monomers. These seven pairs were also found to sample different sidechain conformations, as shown in Figure S7. The pairs sample distance between 5 - 10 Å and angles 0 - 60°. The minimum average distances and angles are shown by Phe33-Glu111 and Phe113-Glu116, whereas

the maximum distances and angles are shown by Phe113-Glu44 and Phe4-Glu3, respectively. Phe4-Glu3 and Phe83-Asp157 transition between 'close' and 'distant' conformations with the corresponding transitions observed in the angles as well. Out of these, Phe4-Glu3 transitions between conformations that are closer in distance with the corresponding angles becoming more planar, whereas Phe83-Asp157 transitions between conformations that are closer in distance with the corresponding angles becoming more axial. These pairs may form anion-π interactions intermittently whenever the distance/angle falls below a certain threshold. Phe113-Glu44 also shows transition to the 'close' conformation at around 75 ns.

Solvent Analysis

The simulations showed that considerable number of water molecules may be present surrounding the anion-π pairs as well as between the pairs compared to the crystal structure. The number of water molecules surrounding each of the partners in the anion- π pairs is shown in Figure S8 and Table 4. In case of intramolecular pairs that were identified in the crystal structure, the maximum number of water molecules during the course of the simulation was lowest for Phe112-Asp84 and Phe122-Asp84 with three water molecules each (Figure S8A). There were no water molecules surrounding these pairs in the crystal structure even after explicit solvation. The largest number of water molecules was observed for Phe33-Glu35 followed by Phe4-Glu31 and Phe47-Asp50, all of which involve residues located near the surface of the protein. Considering the number of water molecules between the residues of the anion- π pairs, no water molecules were found to be present between Phe118-Asp88 (Figure 3A and Table 4), while the largest number of water molecules were observed for Phe33-Glu35, Phe47-Asp50 and Phe4-Glu31. Pairs comprising of partners near the surface, in general, showed more water molecules surrounding the pairs as

well as between the pairs (Table 4). In case of inter-molecular pairs, the number of water molecules surrounding the residues ranged from 1-9 at different points during the simulation (Figure S8B). The number of water molecules between the partners of the pairs ranged from 0-2 (Figure 3B).

For anion pairs identified by MD simulations and not observed in the crystal structure, the number of water molecules was much higher. The surrounding water ranged from 0-17 (Figure S8C) while water molecules between the pairs ranged from 0-12 (Figure 3C). Phe4-Glu3 shows least number of water molecules corresponding to the times when the pairs transition to conformations that are closer in distance ('close' conformations). Phe33-Glu111 pair was associated with the least number of water molecules throughout the simulation amongst the non-crystal structure pairs. The non-crystal structure pairs involved at least one residue which is close to the surface corresponding with the large number of water molecules associated with these pairs (Table 4).

Semi-empirical quantum energy calculations

The KM energy vs time plots for the ten anion-π pairs present in the crystal structure are shown in Figure 4A and 4B. The minimum, maximum and average interaction energies for these pairs during MD are given in Table 5, along with the interaction energies seen in the crystal structure. These pairs show energies ranging between +3 to -8 kcal/mol, with lower energies observed for pairs sampling shorter distances and near-planar angles, and vice-versa. Except Phe4-Glu111, all pairs sample lower energy conformations at some point during MD than that seen in crystal structure. The lowest interaction energy of -7.53 kcal/mol is shown by Phe38- Glu111, whereas Asp50-Phe33 shows the most prominent change (ΔE of -4.83 kcal/mol) between the minimum energy shown during MD and the crystal structure. Phe122-Asp84 intermittently shows

25

low energy configurations (~ -7 kcal/mol) corresponding to the times it samples planar angles. Phe47- Asp50 shows ~ 0 kcal/mol throughout the trajectory, as the distance and angle between this pair were stable at around 8 A and 45° respectively.

The energy vs time plot for the seven anion-π pairs only seen during the simulation are shown in Figure 4C, and the minimum, maximum and average interaction energies for these pairs along with the interaction energies seen in the crystal structure are given in Table 6. These pairs too sample lower energy conformations during MD than that seen in crystal structure, with Phe33-Glu111 showing both the lowest energy (-4.71 kcal/mol) and the most prominent change (ΔE of -3.21 kcal/mol) between the minimum energy structure in MD and the crystal structure. Phe4-Glu3 and Phe83-Asp157 show negative and positive energies intermittently at the times in the trajectory when the pairs sample planar and axial angles respectively. Phe113-Glu44 shows an interaction energy of -2.37 kcal/mol at 75 ns where it samples the 'close' configuration.

Potential of Mean force

The free energy maps of one intramolecular and one intermolecular anion-π pair present in the crystal structure, as well as one pair appearing during the MD simulation are shown in Figure 5. The free energy maps for all other pairs are shown in Figure S9.

Figure 5A shows an anion-π pair exhibiting average values of distances and angles close to the values seen in the crystal structure. These relatively small dynamics variations exhibit a narrow distance (~4.0 to ~6.3 Å) and angles (0 to ~44°) range sampled by this specific anion-π pair. The corresponding PMF (Figure 5B) correspondingly exhibit a narrow free energy basin of the energy landscape.

Figure 5C shows one of the symmetrical intermolecular anion-π pairs showing dynamic variations corresponding to a slightly wider range of distances (~3.8 to ~6.6 Å) and angles (0 to ~60). The pair shows average values of distances and angles during the MD close to the values in the crystal structure but with larger fluctuations. The corresponding PMF (Figure 5D) indicates a relatively wider associated free energy basin.

Figure 5E shows an anion-π pair only seen during the MD simulation showing dynamic variations with a wide range of distances (~4.8 to ~10.9 Å) and angles (10 to ~76°), sampling 'close' and 'distant' states (relative to the conformation in the crystal structure) intermittently throughout the simulation. The pair samples distances and angles both above and below the values shown in the crystal structure (Table 3). The PMF (Figure 5F) indicates two possible free energy minimas.

Discussion

Proteins are not static structures, but undergo dynamical variations at room temperature and in solution that can lead to changes in the amount and strength of anion-π interactions. Some relatively weak anion-π interactions cease to be present, and some new interactions are formed over time. Out of the twelve negatively charged or aromatic amino acid residues present at the dimer interface and the active site of the protein, nine have been shown to possibly form anion-π interactions (Asp-24, Phe-29, Glu-31, Phe-33, Glu-35, Phe47, Asp-50, Asp-84, Phe-122). Among these, two symmetrical pairs, Phe33-Asp50, are present at the dimer interface, and an anion-π triplet, Glu31-Phe29-Asp24, is present at the active site of RmlC. Together these results suggest that anion-π interactions may contribute to both the structural stability and the function of the protein. Possible anion-π pairs and triplets only observed in other conformations sampled by the

protein during the MD form an extensive network of anion-π interactions extending through most of the protein structure (Figure 6A and 6B).

From the previous analysis of the PDB, edgewise geometries (small θ values between the pairs) are found to be more common than would be expected from a random distribution (21). At the same time, QM calculations of optimized BF (benzene-formate) pairs (14) show that the strongest interaction energies are also associated with edgewise interactions (-9.5 kcal/mol for planar BF dimer). This trend continues to be seen during the MD simulation of RmlC, with sixteen out of the seventeen pairs showing average angle values of less than 45°. As expected, fifteen out of these seventeen pairs show negative (i.e., stabilizing) average interaction energies as calculated by the KM energy decomposition analysis (Table 5 and 6), with the minimum energies shown by the pairs at near planar angles. The calculated interaction energies for the pairs during MD can be more stabilizing (by up to -4.8 kcal/mol for Asp50- Phe33) than the interaction energy calculated in the crystal structure, indicating that the pairs are able to sample lower energy conformations not seen in the crystal structure. While some of these anion-π interactions are relatively weak, they are still significantly above kT (i.e., ~0.6 kcal/mol at 300 K) and can occur frequently. The combined ΔE from the seventeen pairs may provide a significant contribution to the overall stability of the protein, with an average of -1.6 kcal/mol per pair.

The PMF calculated for all the pairs (Figure 5 and Figure S9) shows that some pairs are very stable throughout the molecular dynamics simulations, while some other pairs are only transiently formed, with both narrow and wide free energy basins of the energy landscape, respectively. This indicates the variable amount of stability provided by these interactions to the protein structure, with pairs with narrow free energy basins likely sampling conformations close to the native state for these pairs and hence more stable. The wide fluctuations in the angles and

distances of some of the pairs (Figure 5E and S9) transitioning between 'close' and 'distant' states and corresponding to wide PMF variations, can partly be attributed to the lack of convergence over the simulation time scale (150 ns).

The effects of the solvent molecules on the polarization of surrounding organic moieties are implicitly incorporated in the effective charges and other parameters of the pair-additive force fields (37). As an electrostatic component of the anion-π interaction exists, the burial of the pair would minimize potential disruption of the anion-π geometry by water (14,21). As can be expected, pairs comprising of partners near the surface exhibit more water molecules surrounding the pairs, and show relatively larger fluctuations in distances and angles (Table 3 and 4) than more buried pairs. On the other hand, well buried pairs like Phe112-Asp84 and Phe122-Asp84 are exposed to very few waters throughout the simulations, and show low fluctuations in distances and angles. These variations in angle and distance are not due only to solvent screening, but to otherwise more crowded environments. Nonetheless, some pairs form anion-π only intermittently ('close' and 'distant' conformations), likely because one or both of the anion- π partners prefer at time to interact with the nearby water molecules than with each other, preventing the pairs from staying in the favorable 'close' conformations at planar angles for long periods of time.

Previous comparisons of ab-initio and empirical force fields calculations of anion- π interactions indicate that the agreement between these two approaches is qualitatively correct, but that empirical force fields can underestimate the ab-initio calculated interaction energies for anion-π interactions by about 50% (14,21), especially close to the low distance / low angle values of the anion-π pairs geometries. This means that the strength of interactions may be underestimated in the present MD simulations. The of observations made here, based on using an empirical forcefield, will need to be further validated in future experiments looking at the stability of the

proteins by mutagenesis studies, and by using new polarizable force fields like Amoeba (38) and Drude (39) that explicitly include polarization in their parameterization and in calculation of the forces. This will characterize the interplay between the strength of anion-π interactions and the folding and stability of the protein structure.

LIST OF REFERENCES

1.   Shi, G., Ding, Y., and Fang, H. (2012) Unexpectedly strong anion-pi interactions on the graphene flakes. *Journal of computational chemistry* **33**, 1328-1337

2.   Rosokha, Y. S., Lindeman, S. V., Rosokha, S. V., and Kochi, J. K. (2004) Halide recognition through diagnostic "anion-pi" interactions: molecular complexes of Cl-, Br-, and I- with olefinic and aromatic pi receptors. *Angewandte Chemie* **43**, 4650-4652

3.   de Hoog, P., Gamez, P., Mutikainen, I., Turpeinen, U., and Reedijk, J. (2004) An aromatic anion receptor: anion-pi interactions do exist. *Angewandte Chemie* **43**, 5815-5817

4.   Lucas, X., Quinonero, D., Frontera, A., and Deya, P. M. (2009) Counterintuitive substituent effect of the ethynyl group in ion-pi interactions. *The journal of physical chemistry. A* **113**, 10367-10375

5.   Frontera, A., Saczewski, F., Gdaniec, M., Dziemidowicz-Borys, E., Kurland, A., Deya, P. M., Quinonero, D., and Garau, C. (2005) Anion-pi interactions in cyanuric acids: a combined crystallographic and computational study. *Chemistry* **11**, 6560-6567

6.   Berryman, O. B., Hof, F., Hynes, M. J., and Johnson, D. W. (2006) Anion-pi interaction augments halide binding in solution. *Chemical communications*, 506-508

7.   Frontera, A., Gamez, P., Mascal, M., Mooibroek, T. J., and Reedijk, J. (2011) Putting anion-pi interactions into perspective. *Angewandte Chemie* **50**, 9564-9583

8.   Lucas, X., Frontera, A., Quinonero, D., and Deya, P. M. (2010) Substituent effects in ion-pi interactions: fine-tuning via the ethynyl group. *The journal of physical chemistry. A* **114**, 1926-1930

9.   Chen, Y., and Wang, F. (2015) Theoretical study of interactions between electron-deficient arenes and coinage metal anions. *Journal of molecular modeling* **21**, 2584

10.   Zhang, J., Zhou, B., Sun, Z. R., and Wang, X. B. (2015) Photoelectron spectroscopy and theoretical studies of anion-pi interactions: binding strength and anion specificity. *Physical chemistry chemical physics : PCCP* **17**, 3131-3141

11.   Chifotides, H. T., and Dunbar, K. R. (2013) Anion-pi interactions in supramolecular architectures. *Accounts of chemical research* **46**, 894-906

12.   Wheeler, S. E., and Bloom, J. W. (2014) Toward a more complete understanding of noncovalent interactions involving aromatic rings. *The journal of physical chemistry. A* **118**, 6133-6147

13.   Wheeler, S. E., and Bloom, J. W. (2014) Anion-pi interactions and positive electrostatic potentials of N-heterocycles arise from the positions of the nuclei, not changes in the pi-electron distribution. *Chemical communications* **50**, 11118-11121

14.   Jackson, M. R., Beahm, R., Duvvuru, S., Narasimhan, C., Wu, J., Wang, H. N., Philip, V. M., Hinde, R. J., and Howell, E. E. (2007) A preference for edgewise interactions between aromatic rings and carboxylate anions: the biological relevance of anion-quadrupole interactions. *The journal of physical chemistry. B* **111**, 8242-8249

15.   Shi, Z., Olson, C. A., and Kallenbach, N. R. (2002) Cation-pi interaction in model alpha-helical peptides. *Journal of the American Chemical Society* **124**, 3284-3291

16.   Olson, C. A., Shi, Z., and Kallenbach, N. R. (2001) Polar interactions with aromatic side chains in alpha-helical peptides: Ch...O H-bonding and cation-pi interactions. *Journal of the American Chemical Society* **123**, 6451-6452

17.   Shi, Z., Olson, C. A., Bell, A. J., Jr., and Kallenbach, N. R. (2002) Non-classical helix-stabilizing interactions: C-H...O H-bonding between Phe and Glu side chains in alpha-helical peptides. *Biophysical chemistry* **101-102**, 267-279

18.   Breberina, L. M., Milcic, M. K., Nikolic, M. R., and Stojanovic, S. D. (2014) Contribution of anion-pi interactions to the stability of Sm/LSm proteins. *Journal of biological inorganic chemistry : JBIC : a publication of the Society of Biological Inorganic Chemistry*

19.   Schwans, J. P., Sunden, F., Lassila, J. K., Gonzalez, A., Tsai, Y., and Herschlag, D. (2013) Use of anion-aromatic interactions to position the general base in the ketosteroid isomerase active site. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 11308-11313

20.   Krieger, I. V., Freundlich, J. S., Gawandi, V. B., Roberts, J. P., Gawandi, V. B., Sun, Q., Owen, J. L., Fraile, M. T., Huss, S. I., Lavandera, J. L., Ioerger, T. R., and Sacchettini, J. C. (2012) Structure-guided discovery of phenyl-diketo acids as potent inhibitors of M. tuberculosis malate synthase. *Chemistry & biology* **19**, 1556-1567

21.   Philip, V., Harris, J., Adams, R., Nguyen, D., Spiers, J., Baudry, J., Howell, E. E., and Hinde, R. J. (2011) A survey of aspartate-phenylalanine and glutamate-phenylalanine interactions in the protein data bank: searching for anion-pi pairs. *Biochemistry* **50**, 2939-2950

22.   Hugh P. G. Thompsona, G. M. D. (2014) Which conformations make stable crystal structures? Mapping crystalline molecular geometries to the conformational energy landscape. *Chemical science* **5**, 3173-3182

23.   Christendat, D., Saridakis, V., Dharamsi, A., Bochkarev, A., Pai, E. F., Arrowsmith, C. H., and Edwards, A. M. (2000) Crystal structure of dTDP-4-keto-6-deoxy-D-hexulose 3,5-epimerase from Methanobacterium thermoautotrophicum complexed with dTDP. *The Journal of biological chemistry* **275**, 24608-24612

24.   Jenkins, D. D., Harris, J. B., Howell, E. E., Hinde, R. J., and Baudry, J. (2013) STAAR: statistical analysis of aromatic rings. *Journal of computational chemistry* **34**, 518-522

25.   Gallivan, J. P., and Dougherty, D. A. (1999) Cation-pi interactions in structural biology. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 9459-9464

26.   (2014) Molecular Operating Environment (MOE), 2013.08. in *Chemical Computing Group Inc.*

27.   MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry. B* **102**, 3586-3616

28.   Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *Journal of computational chemistry* **26**, 1781-1802

29.   Bakan, A., Meireles, L. M., and Bahar, I. (2011) ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* **27**, 1575-1577

30. Schuttelkopf, A. W., and van Aalten, D. M. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta crystallographica. Section D, Biological crystallography* **60**, 1355-1363

31. Pace, C. N., Grimsley, G. R., and Scholtz, J. M. (2009) Protein ionizable groups: pK values and their contribution to protein stability and solubility. *The Journal of biological chemistry* **284**, 13285-13289

32. O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011) Open Babel: An open chemical toolbox. *Journal of cheminformatics* **3**, 33

33. M.W.Schmidt, K. K. B., J.A.Boatz, S.T.Elbert, M.S.Gordon, J.H.Jensen, S.Koseki, N.Matsunaga, K.A.Nguyen, S.Su, T.L.Windus, M.Dupuis, J.A.Montgomery J. . (1993) General Atomic and Molecular Electronic Structure System. *Comput. Chem.*, 1347-1363

34. Kitaura, K., and Morokuma, K. (1976) New energy decomposition scheme for molecular interactions within Hartree-Fock approximation. *Int. J. Quantum Chem.* **10**, 325–340

35. Ishida, K., Morokuma, K., and Komornicki, A. (1977) The intrinsic reaction coordinate. An ab initio calculation for HNC - HCN and H^-+CH_4 - CH_4+H^-. *J. Chem. Phys.* **66**, 2153–2156

36. Roux, B. (1995) The calculation of the potential of mean force using computer simulations. *Computer Physics Communications* **91**, 275–282

37. Leontyev, I., and Stuchebrukhov, A. (2011) Accounting for electronic polarization in non-polarizable force fields. *Physical chemistry chemical physics : PCCP* **13**, 2613-2626

38. Shi, Y., Xia, Z., Zhang, J., Best, R., Wu, C., Ponder, J. W., and Ren, P. (2013) The Polarizable Atomic Multipole-based AMOEBA Force Field for Proteins. *Journal of chemical theory and computation* **9**, 4046-4063

39.   Lopes, P. E., Huang, J., Shim, J., Luo, Y., Li, H., Roux, B., and Mackerell, A. D., Jr. (2013) Force Field for Peptides and Proteins based on the Classical Drude Oscillator. *Journal of chemical theory and computation* **9**, 5430-5449

APPENDIX

Table 1.1. Anion-π pairs in crystal structure. The Phe and Glu or Asp anion-π pairs identified, both intramolecular and intermolecular, and their respective distance and angles in the crystal structure.

Intra-/Inter-molecular pairs	Residue-1	Rsidue-2	Distance (A)	Angle (deg)
Intra	Phe112	Asp84	4.59	0.62
Intra	Phe38	Glu111	5.49	0.74
Intra	Phe4	Glu31	5.03	0.10
Intra	Phe122	Asp84	5.11	23.98
Intra	Phe33	Glu35	6.09	4.98
Intra	Phe4	Glu111	6.17	12.33
Intra	Phe47	Asp50	6.88	36.01
Intra	Phe118	Asp88	4.31	73.86
Inter	Phe33	Asp50	5.02	33.28
Inter	Asp50	Phe33	5.02	33.28

Table 1.2. Distances and angles of anion-π pairs in crystal structure. The minimum, maximum and average distances and angles for the ten anion-π pairs during MD simulation.

Anion-π pairs	Min Distance (A)	Max Distance (A)	Avg Distance (A)	Min Angle (deg)	Max Angle (deg)	Avg Angle (deg)
Phe112- Asp84	3.94	6.32	4.75	0.00	43.83	6.56
Phe38- Glu111	3.90	6.76	4.89	0.00	44.74	12.92
Phe4- Glu31	3.73	6.74	5.07	0.00	61.77	17.91
Phe122- Asp84	3.91	6.39	4.89	0.01	68.62	35.76
Phe33- Glu35	4.73	8.18	6.29	0.00	36.50	6.63
Phe4- Glu111	5.44	9.71	7.41	0.01	85.34	27.23
Phe47- Asp50	6.08	9.26	7.94	13.69	75.10	44.87
Phe118- Asp88	3.02	5.73	4.44	17.99	89.95	69.88
Phe33- Asp50	3.75	6.61	4.83	0.03	60.44	28.10
Asp50- Phe33	3.85	6.87	4.88	0.00	55.30	20.40

Table 1.3. Distances and angles of anion-π pairs seen during MD simulation. The minimum, maximum and average distances and angles for the seven anion-π pairs only seen during MD simulation. The respective distances and angles values in the crystal structure are shown in brackets.

Anion-π pairs (distance, angle in Xtal)	Min Distance (A)	Max Distance (A)	Avg Distance (A)	Min Angle (deg)	Max Angle (deg)	Avg Angle (deg)
Phe33-Glu111 (7.13, 13.60)	5.12	8.43	6.45	0.00	44.68	10.99
Phe29-Glu31 (7.23, 12.95)	4.85	9.51	7.41	0.00	68.39	21.58
Phe4-Glu3 (8.84, 54.98)	4.75	10.86	8.94	0.06	75.79	38.05
Phe83-Asp157 (8.39, 26.80)	5.89	10.94	9.09	1.91	73.89	32.22
Phe29-Asp24 (8.99, 34.49)	5.83	10.01	7.74	0.76	58.07	28.59
Phe113-Glu116 (8.42, 10.11)	5.70	10.99	8.70	0.00	40.84	10.17
Phe113-Glu44 (8.11, 41.71)	5.31	10.99	9.40	0.02	58.17	24.38

Table 1.4. Solvent analysis of anion-π pairs. The minimum and maximum number of water molecules for the ten anion-π pairs (present in crystal structure) and seven anion-π pairs only seen during MD simulation.

Anion-π pairs	Depth (A.u.)		Surrounding				Between			
	Phe Depth	Anion Depth	Xtal	Solvated Xtal	Min	Max	Xtal	Solvated Xtal	Min	Max
Phe112-Asp84	9.85	9.58	0	0	0	3	0	0	0	1
Phe38-Glu111	8.08	9.86	1	1	0	5	0	0	0	1
Phe4-Glu31	5.5	4.29	1	4	1	12	0	0	0	4
Phe122-Asp84	10.59	9.58	0	0	0	4	0	0	0	2
Phe33-Glu35	4.59	4.2	2	4	1	14	1	1	0	5
Phe4-Glu111	5.5	9.86	2	2	1	7	0	0	0	2
Phe47-Asp50	6.41	3.85	2	1	1	8	0	1	0	5
Phe118-Asp88	8.48	8.02	0	0	0	3	0	0	0	0
Asp50-Phe33	4.59	3.85	2	1	1	8	0	0	0	2
Phe33-Asp50	4.59	3.85	1	2	1	9	0	0	0	2
Phe33-Glu111	4.59	9.86	1	1	1	6	1	1	0	3
Phe29-Glu31	4.04	4.29	0	3	1	15	0	2	0	12
Phe4-Glu3	5.5	3.86	1	4	0	18	1	1	0	17
Phe83-Asp157	6.67	3.95	1	6	3	15	1	3	0	13
Phe29-Asp24	4.04	4.09	2	1	2	16	2	6	1	14
Phe113-Glu116	9.24	4.66	1	4	2	14	1	0	0	11
Phe113-Glu44	9.24	3.77	0	6	0	16	1	3	0	15

Table 1.5. Interaction energies of anion-π pairs in crystal structure. The minimum, maximum and average interaction energies for ten anion-π pairs during MD, along with the interaction energies seen in the crystal structure.

Anion-π pairs	Min Interaction energy (kcal/mol)	Max Interaction energy (kcal/mol)	Avg Interaction energy (kcal/mol)	Xtal Interaction energy (kcal/mol)
Phe112- Asp84	-7.37	+0.2	-5.09	-6.44
Phe38- Glu111	-7.53	-0.99	-4.65	-4.83
Phe4- Glu31	-6.86	+0.13	-3.18	-3.62
Phe122- Asp84	-7.29	+1.3	-1.58	-2.87
Phe33- Glu35	-4.62	-0.64	-1.98	-2.21
Phe4- Glu111	-0.94	+0.24	-0.71	-2.09
Phe47- Asp50	-0.97	+0.56	-0.21	-0.32
Phe118- Asp88	-2.87	+2.98	+1.13	+ 1.45
Phe33- Asp50	-6.02	+0.61	-2.53	-2.48
Asp50- Phe33	-7.31	+0.25	-3.59	-2.48

Table 1.6. Interaction energies of anion-π pairs seen during MD simulation. The minimum, maximum and average interaction energies for seven anion-π pairs only seen during MD simulation, along with their interaction energies in the crystal structure.

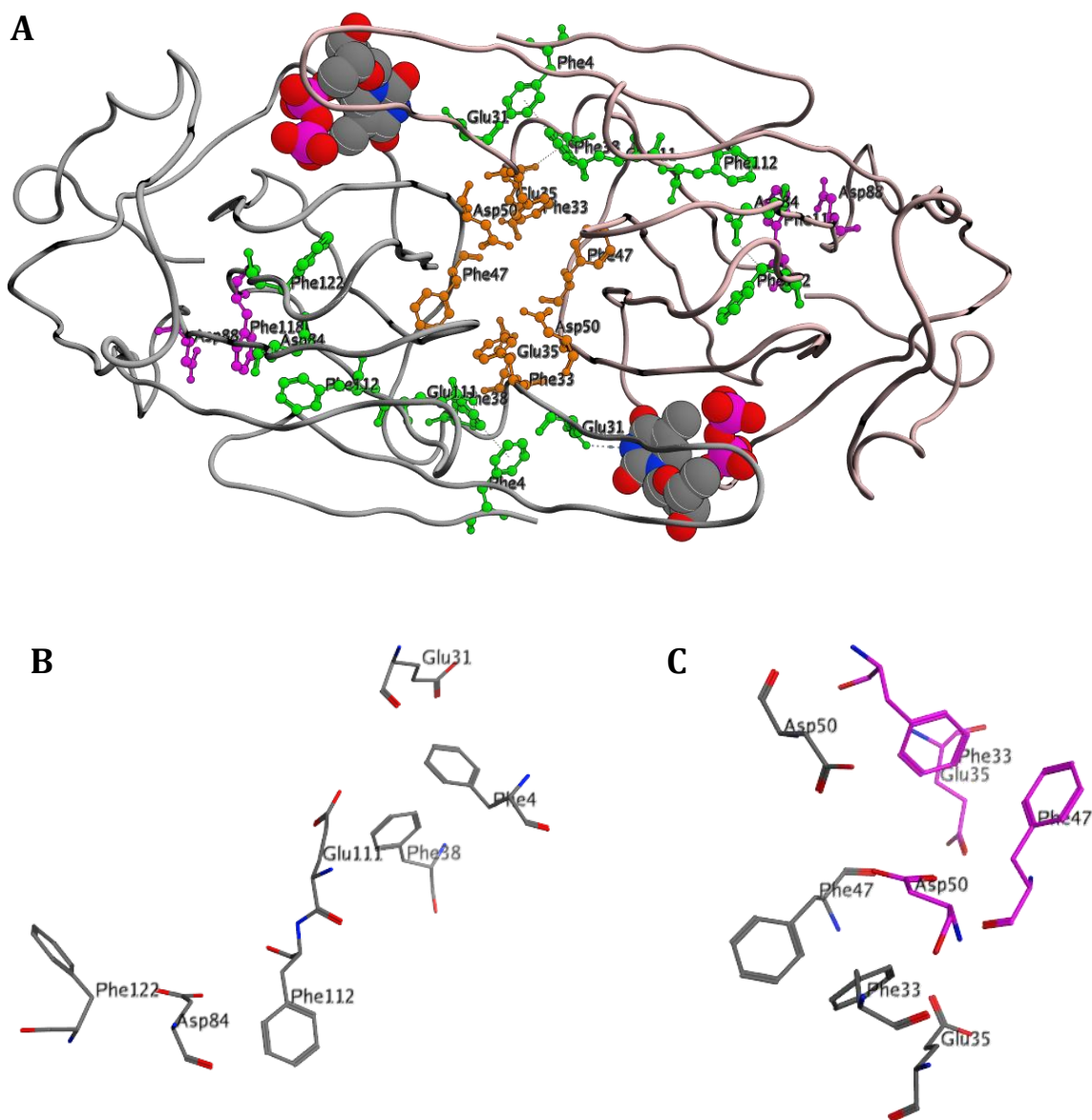| Anion-π pairs	Min Interaction energy (kcal/mol)	Max Interaction energy (kcal/mol)	Avg Interaction energy (kcal/mol)	Xtal Interaction energy (kcal/mol)
Phe33-Glu111	-4.71	-0.7	-2.10	-1.50
Phe29-Glu31	-2.74	+0.18	-0.89	-1.13
Phe4-Glu3	-2.41	+0.37	-0.14	+0.20
Phe83-Asp157	-0.5	+0.68	-0.15	-0.35
Phe29-Asp24	-1.42	+0.22	-0.55	-0.31
Phe113-Glu116	-1.9	-0.02	-0.65	-0.72
Phe113-Glu44	-2.37	+0.09	-0.31	-0.05

Figure 1.1. Anion-π pairs in crystal structure. The two monomers are shown in grey and pink ribbon conformation. (A) Intrarmolecular cluster in two monomers shown in green, intermolecular cluster at dimer interface shown in orange, and a single anion-π pair shown in pink. Substrate dTDP bound in the active site shown in space-filling configuration. (B) Intramolecular and (C) Intermolecular cluster (residues in separate monomers colored grey and pink).

Figure 1.2. Distances and angles of anion-π pairs seen during MD simulation. The two monomers are shown in grey and pink ribbon conformation. (A) Anion-π triplet at active site shown in green and another triplet in red. Single anion-π pairs shown in blue and pink. Substrate dTDP bound in the active site shown in space-filling configuration. (B) Distance vs Time and (C) Angle vs Time plots for the seven intramolecular anion-π pairs only seen during the MD simulation.
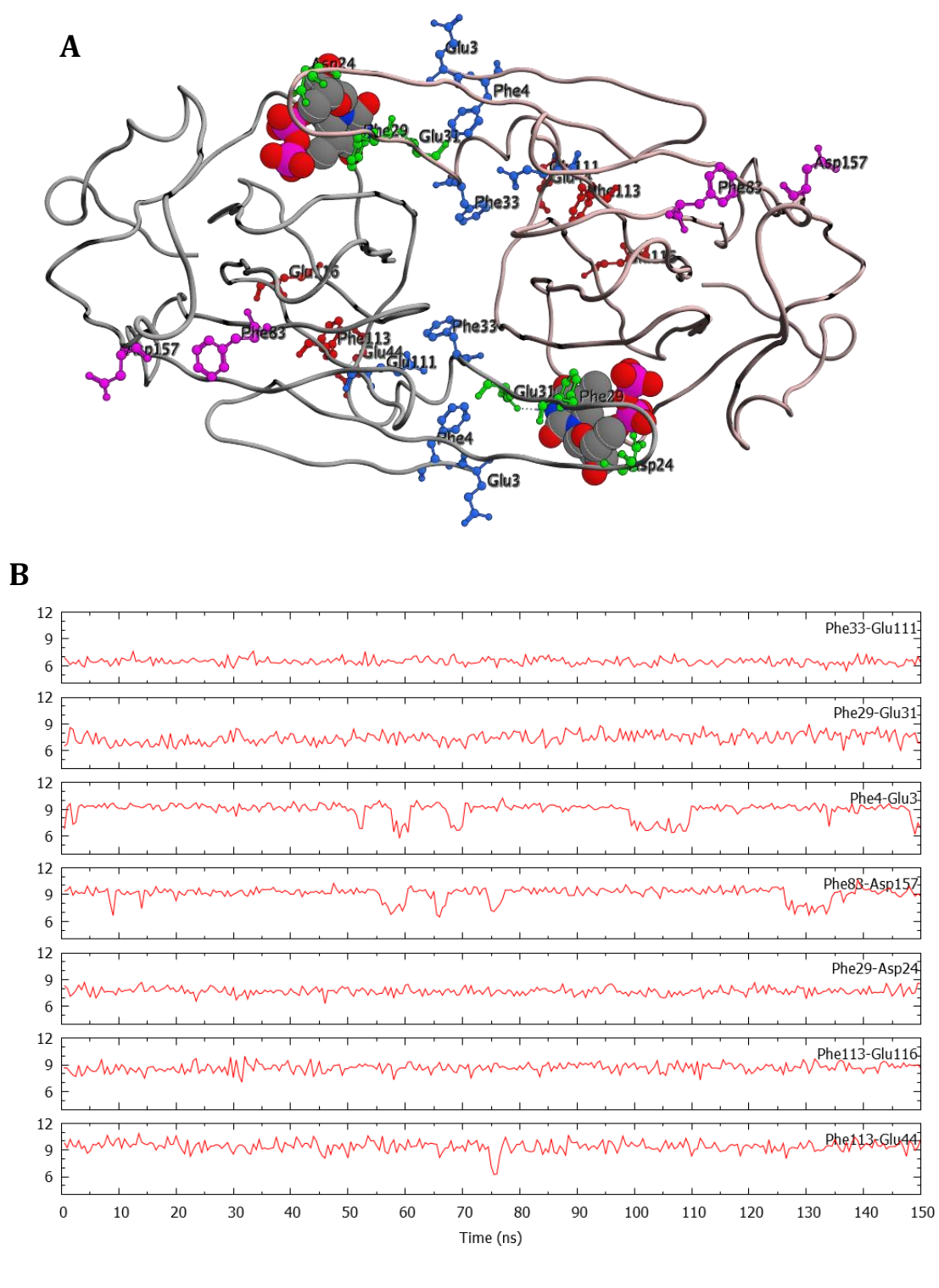
Figure 1.2 continued

C

Figure 1.2 continued

47

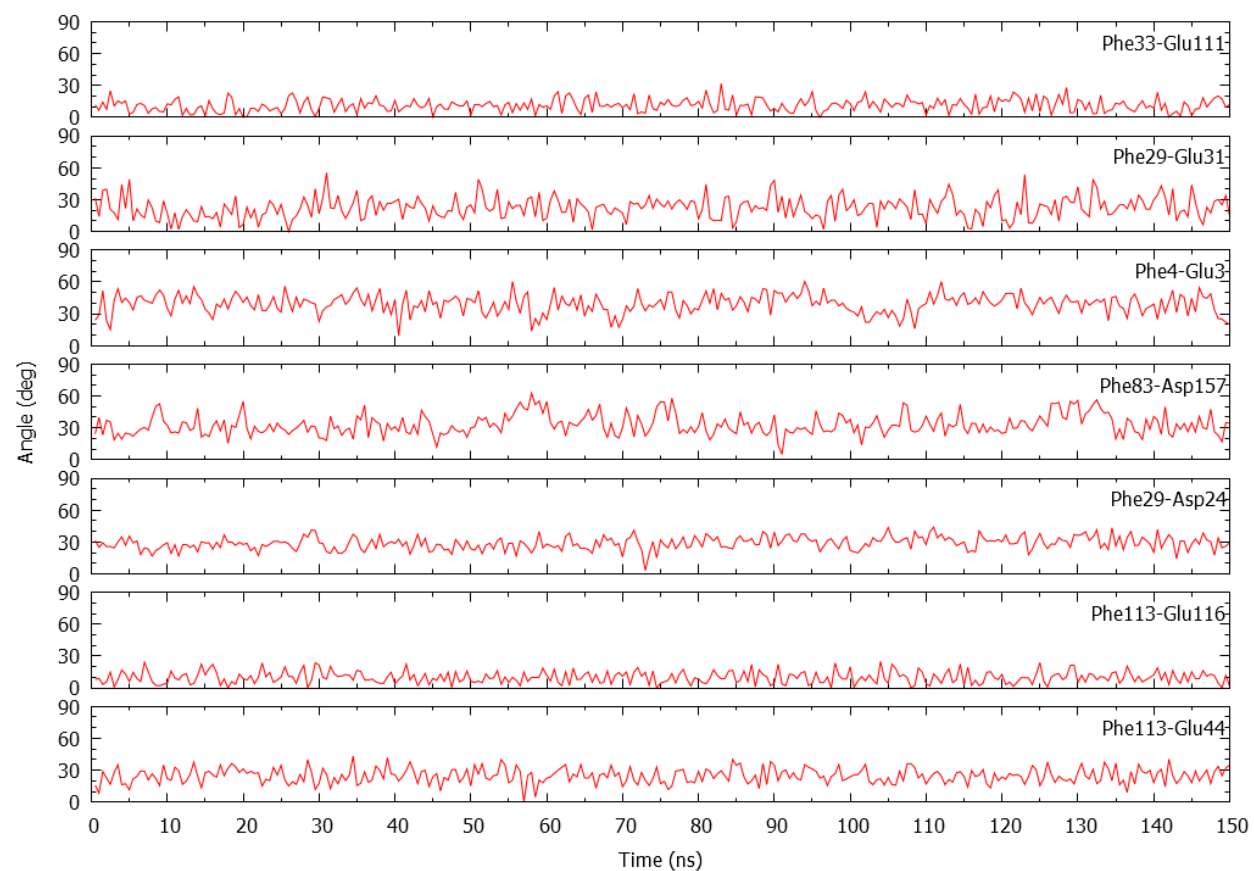**Figure 1.3. Solvent analysis of anion-π pairs.** Total number of water molecules in the region between the Phe and anion residues for (A) eight intramolecular anion-π pairs in the crystal structure, (B) the two symmetrical pairs present at the dimer interface, and (C) seven intramolecular anion-π pairs identified during MD simulations.

A
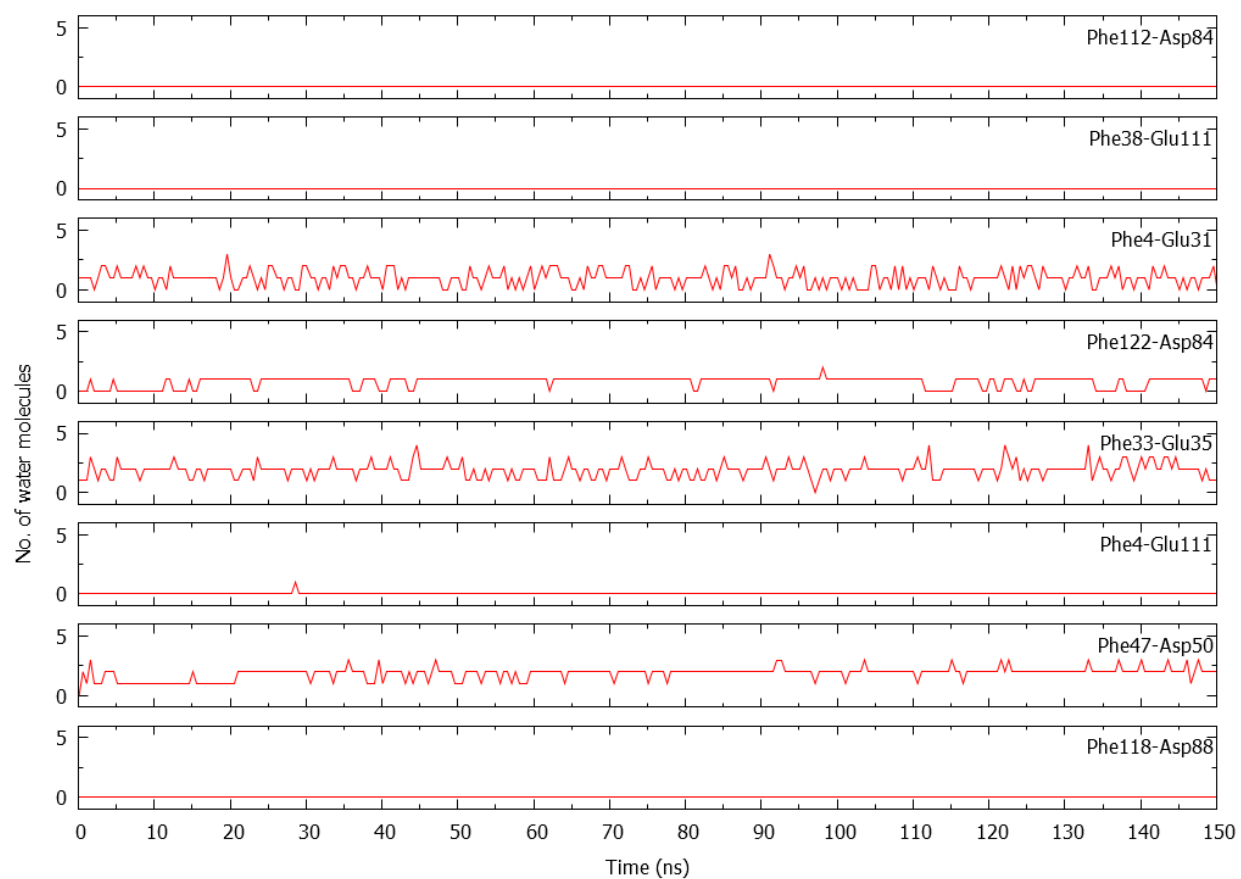
B

Figure 1.3 continued

49
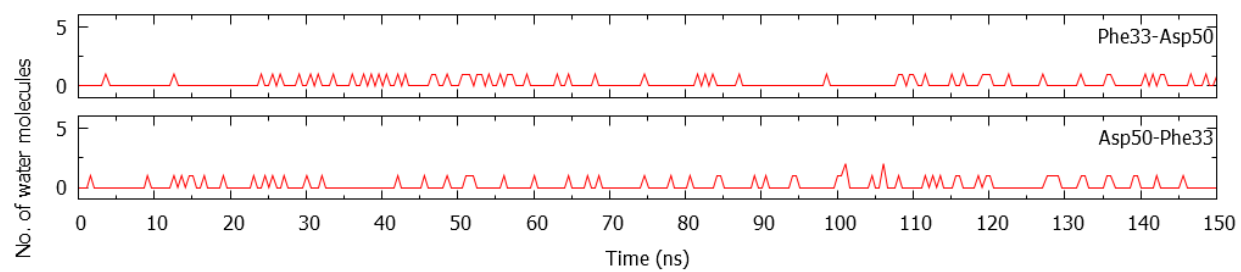
C
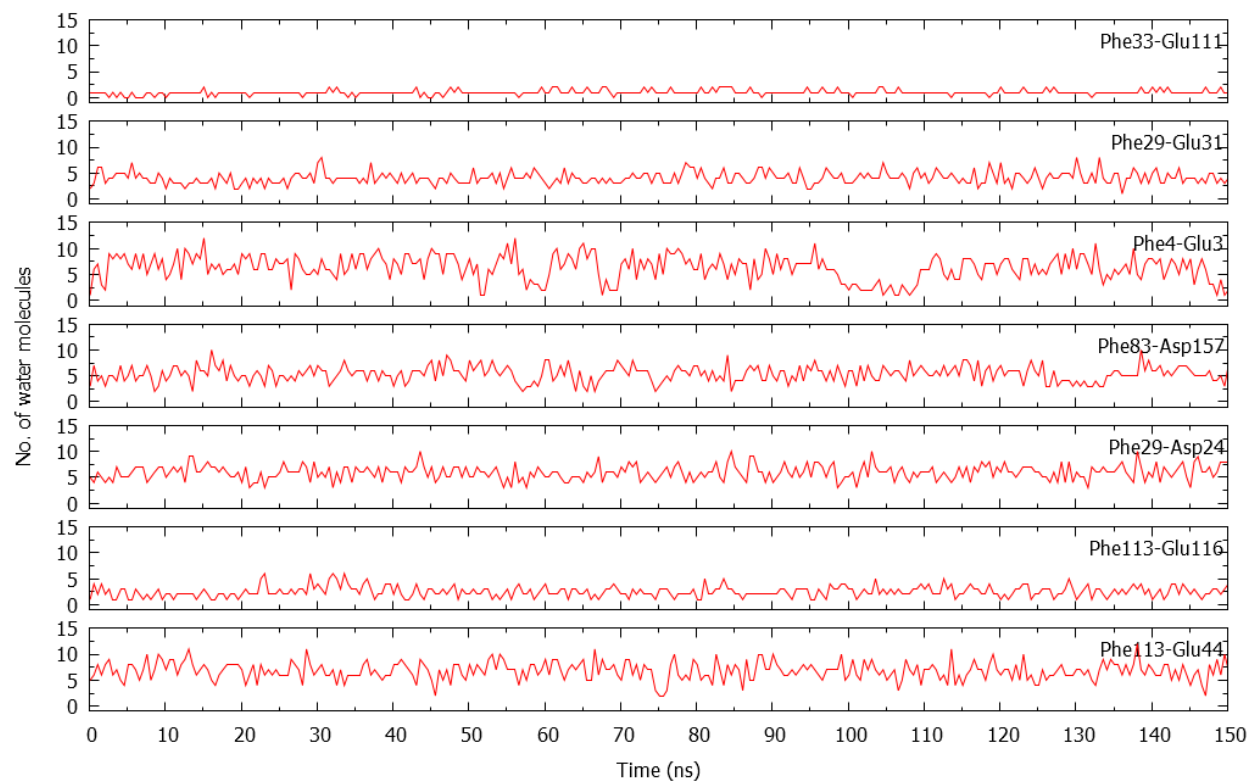
Figure 1.3 continued

Figure 1.4. Interaction energies of anion-π pairs. Interaction Energy vs Time plots for (A) eight intramolecular anion-π pairs in the crystal structure, (B) two symmetrical pairs present at dimer interface and (C) seven intramolecular anion-π pairs only seen during the MD simulation.

A

Phe112-Asp84

Phe38-Glu111

Phe4-Glu31

Phe122-Asp84

Phe33-Glu35

Phe4-Glu111

Phe47-Asp50

Phe118-Asp88

Interaction Energy (kcal/mol)

Time (ns)

B

Phe33-Asp50

Asp50-Phe33

Int. Energy (kcal/mol)

Time (ns)
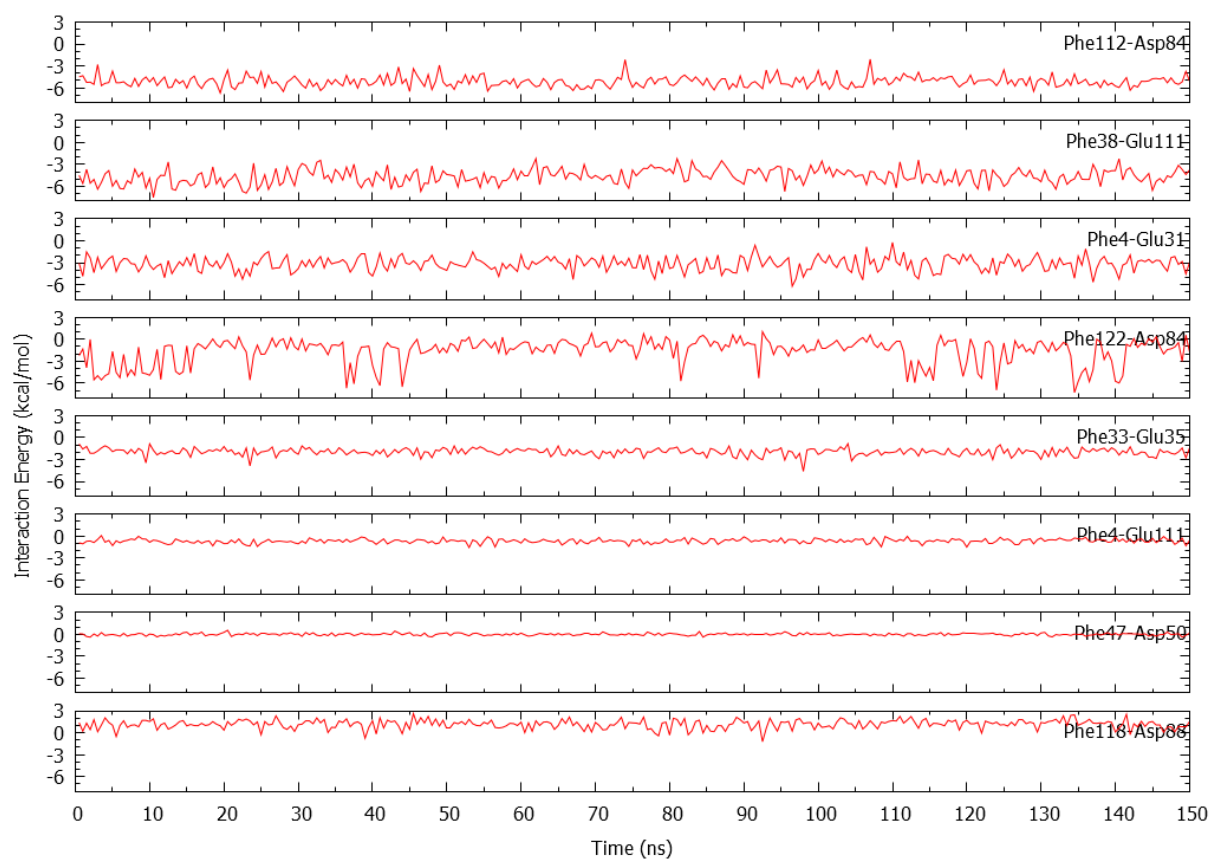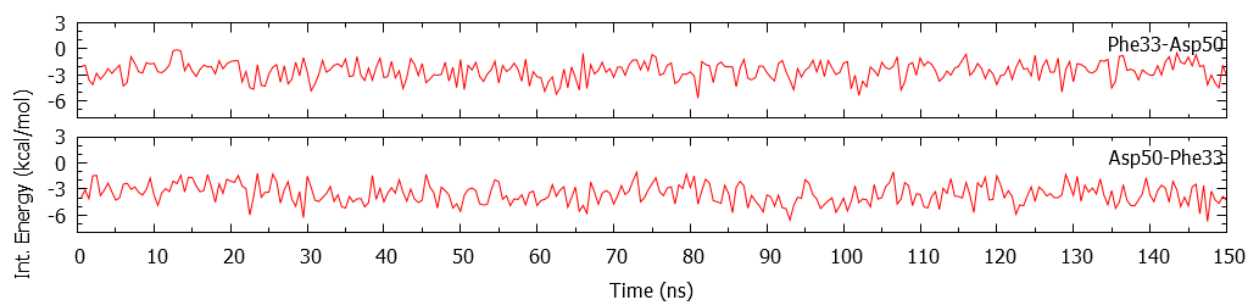
Figure 1.4 continued

C
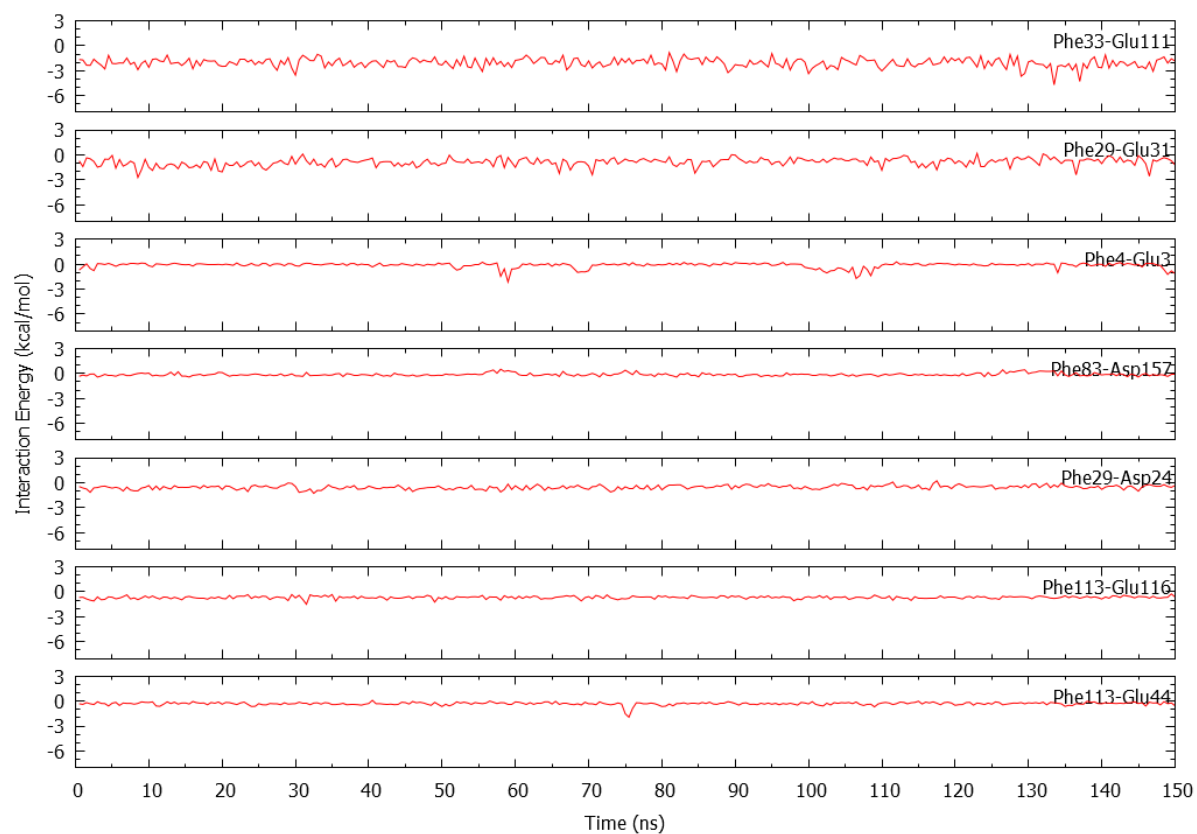
Figure 1.4 continued

Figure 1.5. Angle/distance conformation distributions and free energy (PMF) maps. Angle/distance conformation distribution maps and free energy (PMF) maps for (A,B) Phe112-Asp84, (C,D) Phe33-Glu35 and (E,F) Phe4-Glu3. The angle-distance values in the crystal structure are depicted as white dots on the plots.

A

Phe112-Glu84 (Xtal Distance= 4.59, Angle= 0.62)

B

Phe112-Glu84 (Xtal Distance= 4.59, Angle= 0.62)

Figure 1.5 continued

C

Phe33-Asp50 (Xtal Distance= 5.02, Angle= 33.28)

D

Phe33-Asp50 (Xtal Distance= 5.02, Angle= 33.28)

Figure 1.5 continued

E

Phe4-Glu3 (Xtal Distance= 8.84, Angle= 54.98)

F

Phe4-Glu3 (Xtal Distance= 8.84, Angle= 54.98)
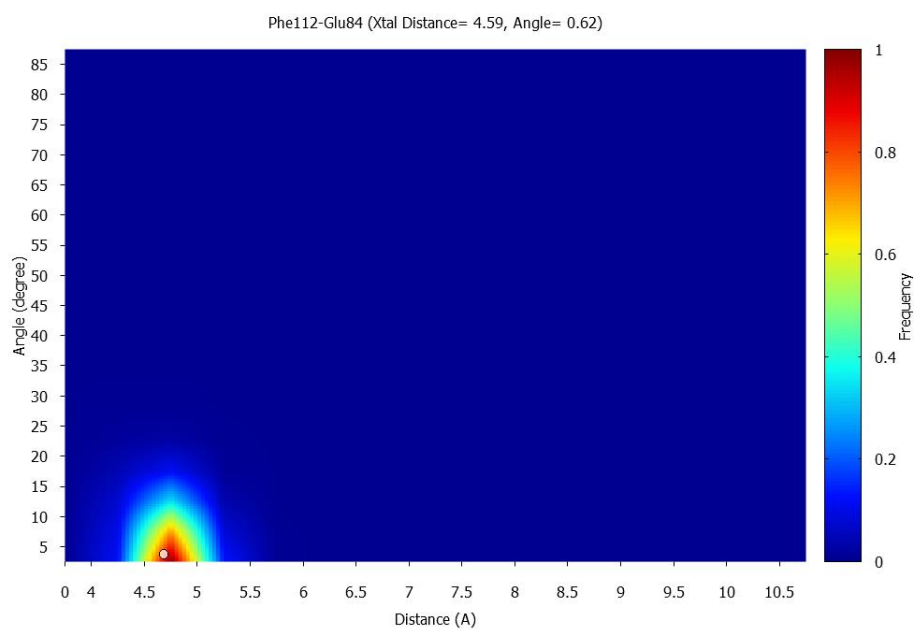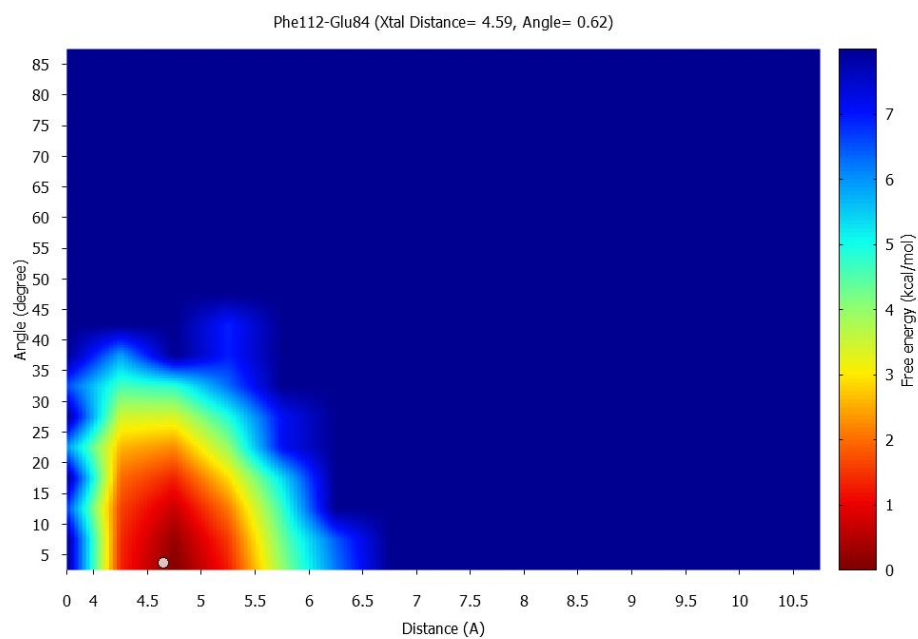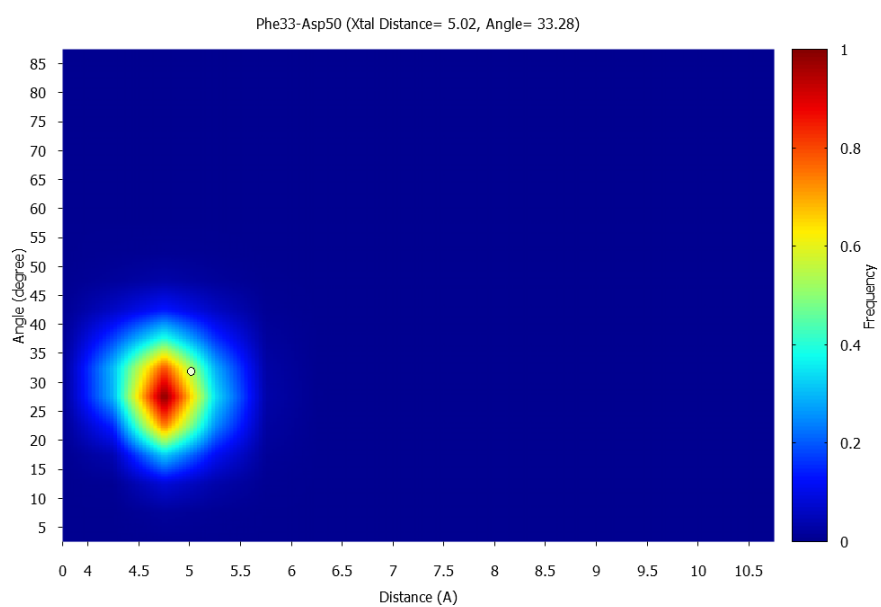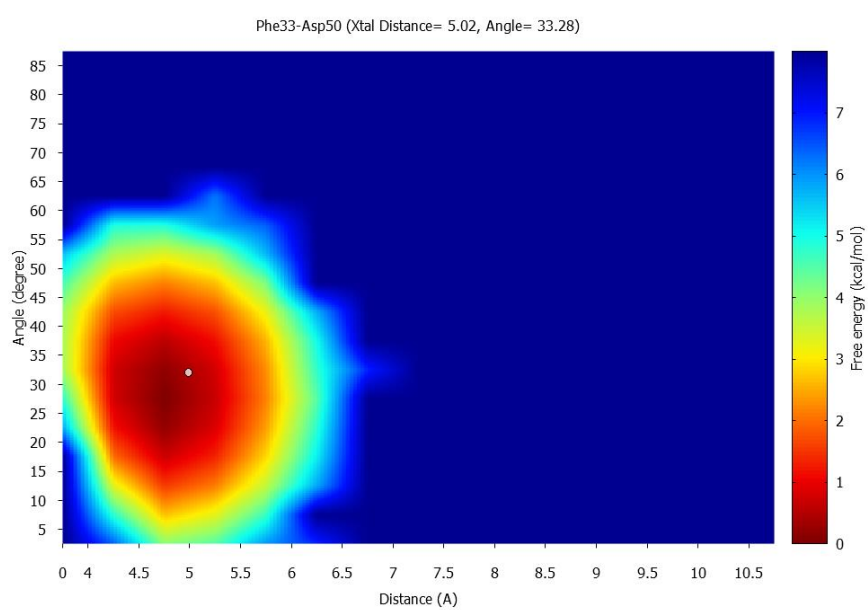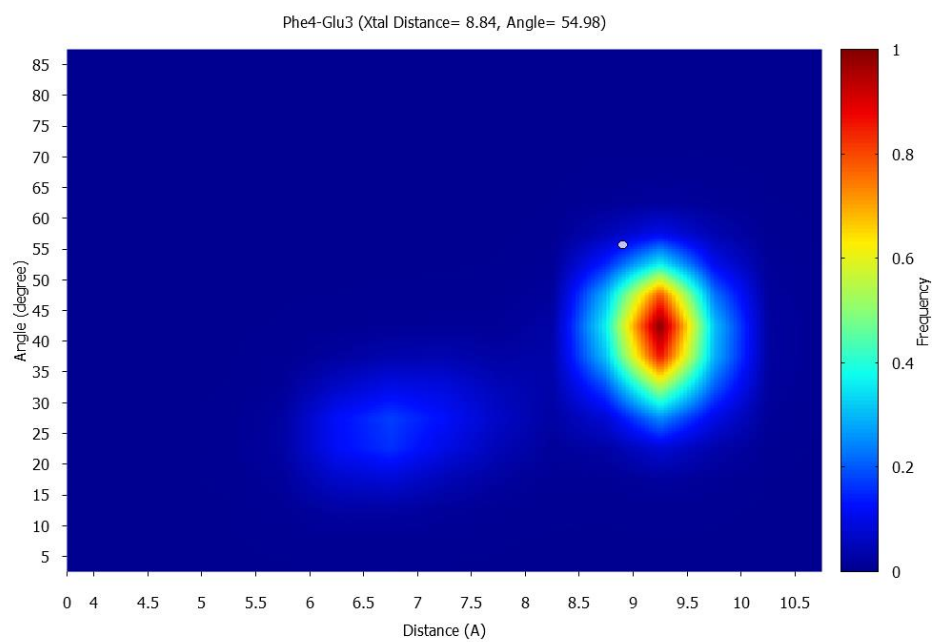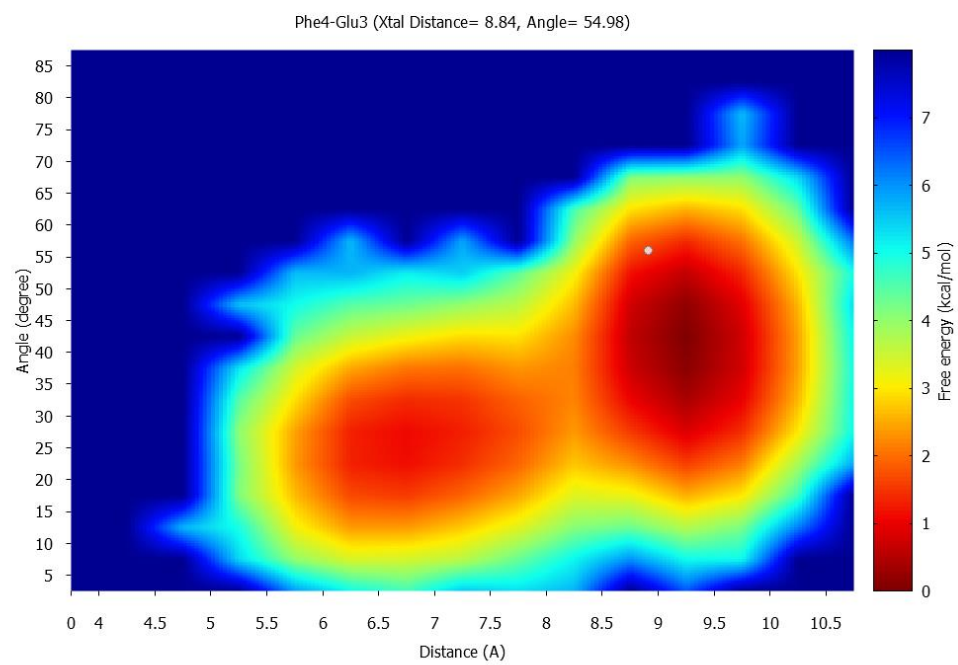
Figure 1.5 continued

Figure 1.6. Anion-π network. The two monomers are shown in grey and pink ribbon conformation. (A) Intrarmolecular cluster in two monomers in green and intermolecular cluster at dimer interface in orange, connected together by single anion-π pairs shown in blue that are only seen during the MD simulation. Substrate dTDP bound in the active site shown in space-filling configuration. (B) The complete anion-π network over the two monomers.

Figure 1.S1. Spherical coordinates in space. Spherical coordinates in space (r, θ, φ): radial

distance r, polar angle θ (theta), and azimuthal angle φ (phi).
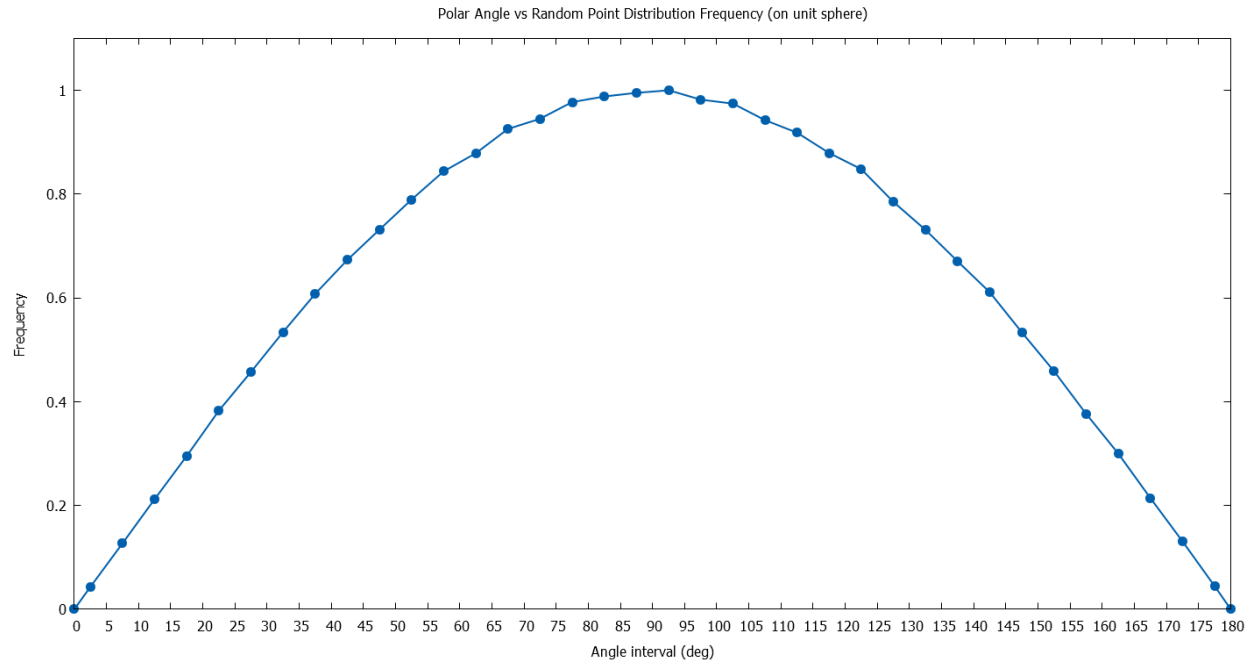
Figure 1.S2. Randomly distributed points on unit sphere. Distribution of 1 million randomly distributed points on a unit sphere at polar angles: 0 to 180°. Each point on the plot represents the distribution in a 5 degree interval. The resulting distribution is Gaussian.

Radius vs Random Point Distribution Frequency (on spheres of different radii)
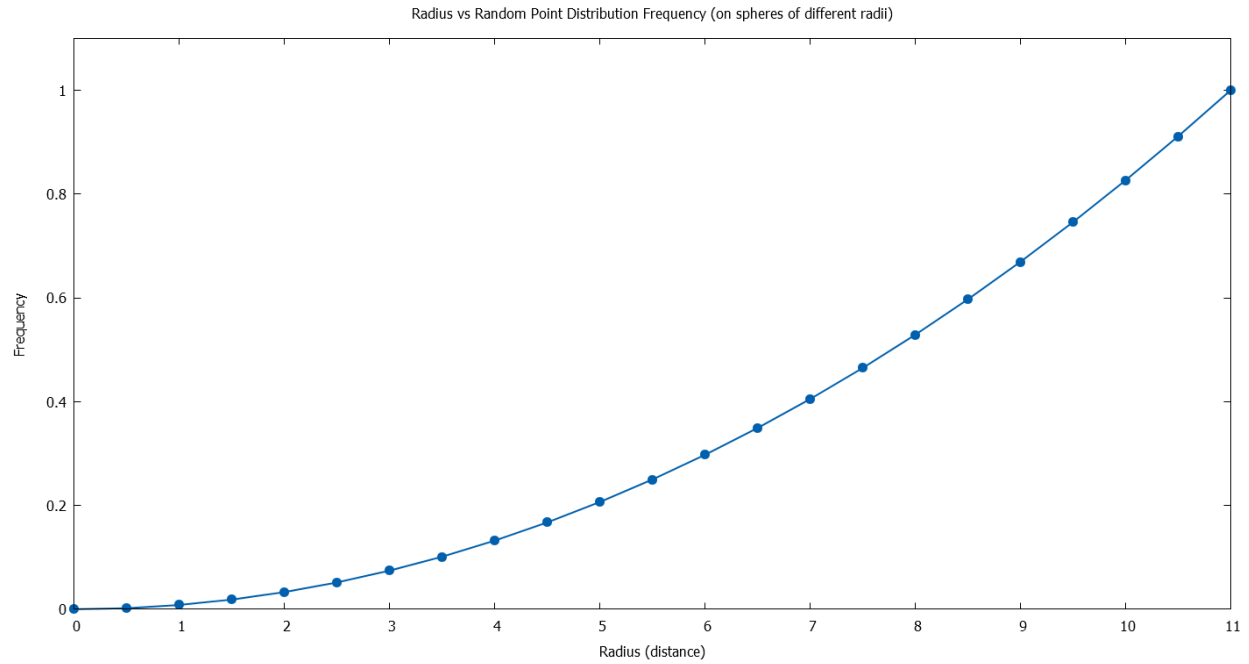
Figure 1.S3. Randomly distributed points at different distances. Distribution of points with same probability density function and at different distances (different sphere radii). The area of sphere (and solid angle) $\propto R^2$. Accordingly, the resulting distribution is exponential.
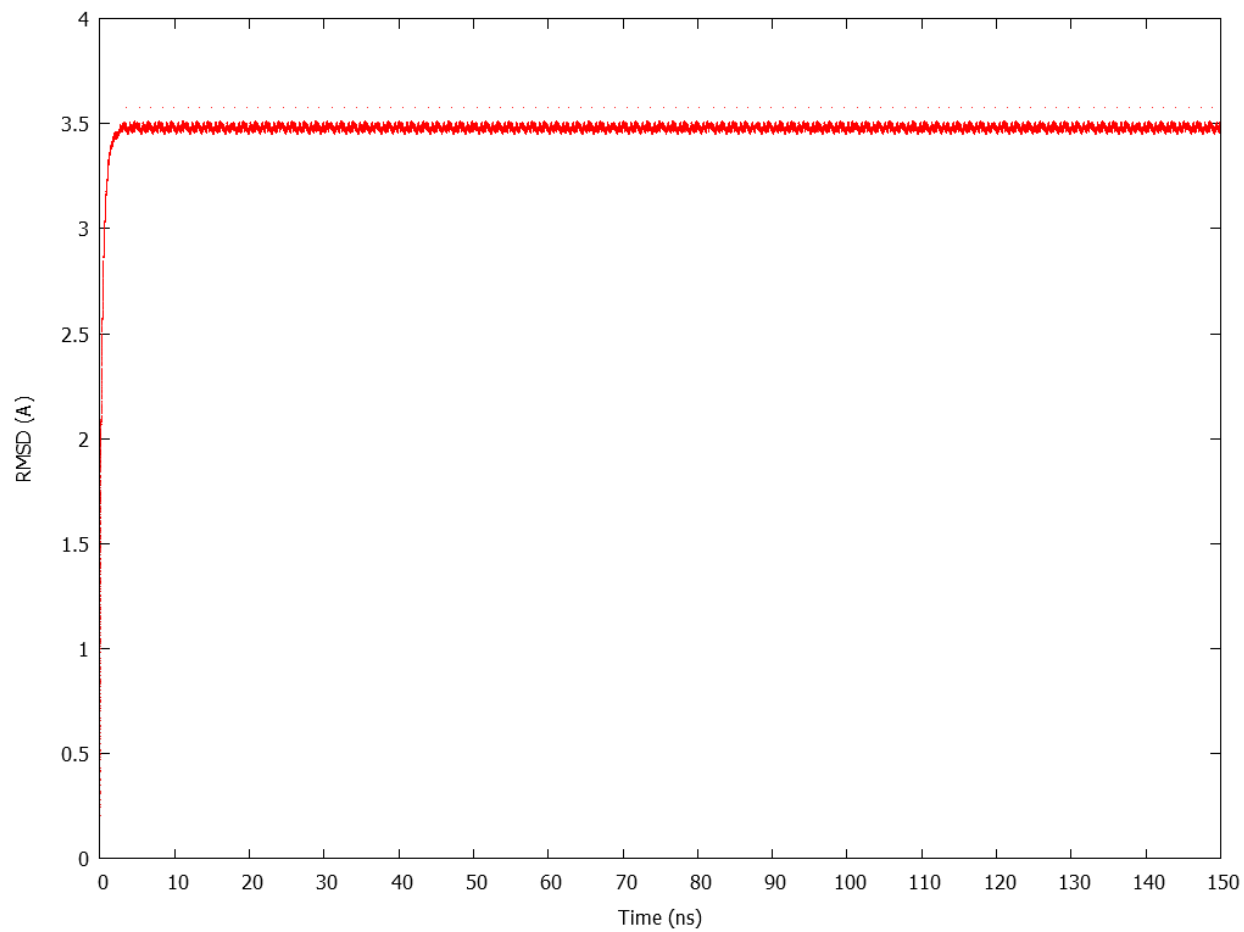
Figure 1.S4. RMSD for protein backbone. RMSD plot for the protein backbone over the time course of the trajectory**.**
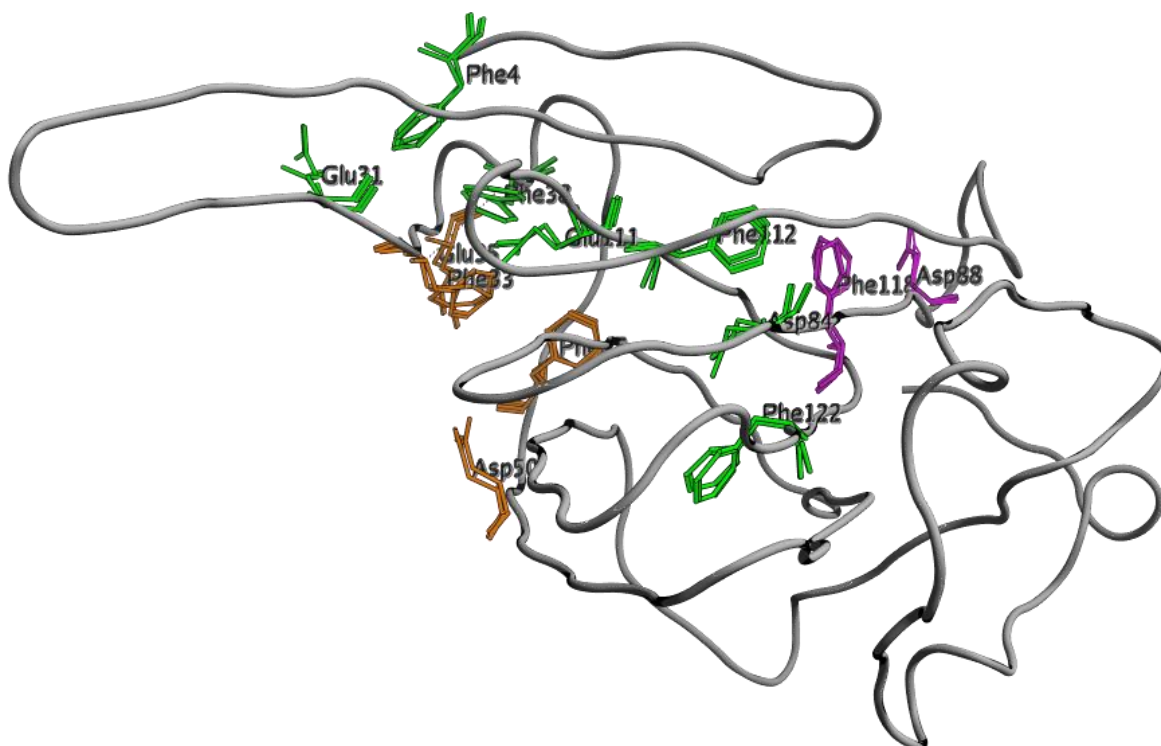
Figure 1.S5. Sidechain conformations sampled by anion-π pairs present in crystal structure.
Sidechain conformations sampled by the ten anion-π pairs present in the crystal structure, in the
MD snapshots showing minimum and maximum angles for each pair (Table 1.2).

Figure 1.S6. Distances and angles of anion-π pairs in crystal structure. (A) Distance vs Time and (B) Angle vs Time plots for the eight intramolecular anion-π pairs in the crystal structure. (C) Distance vs Time and (D) Angle vs Time plots for the two symmetrical pairs present at the dimer interface.

A
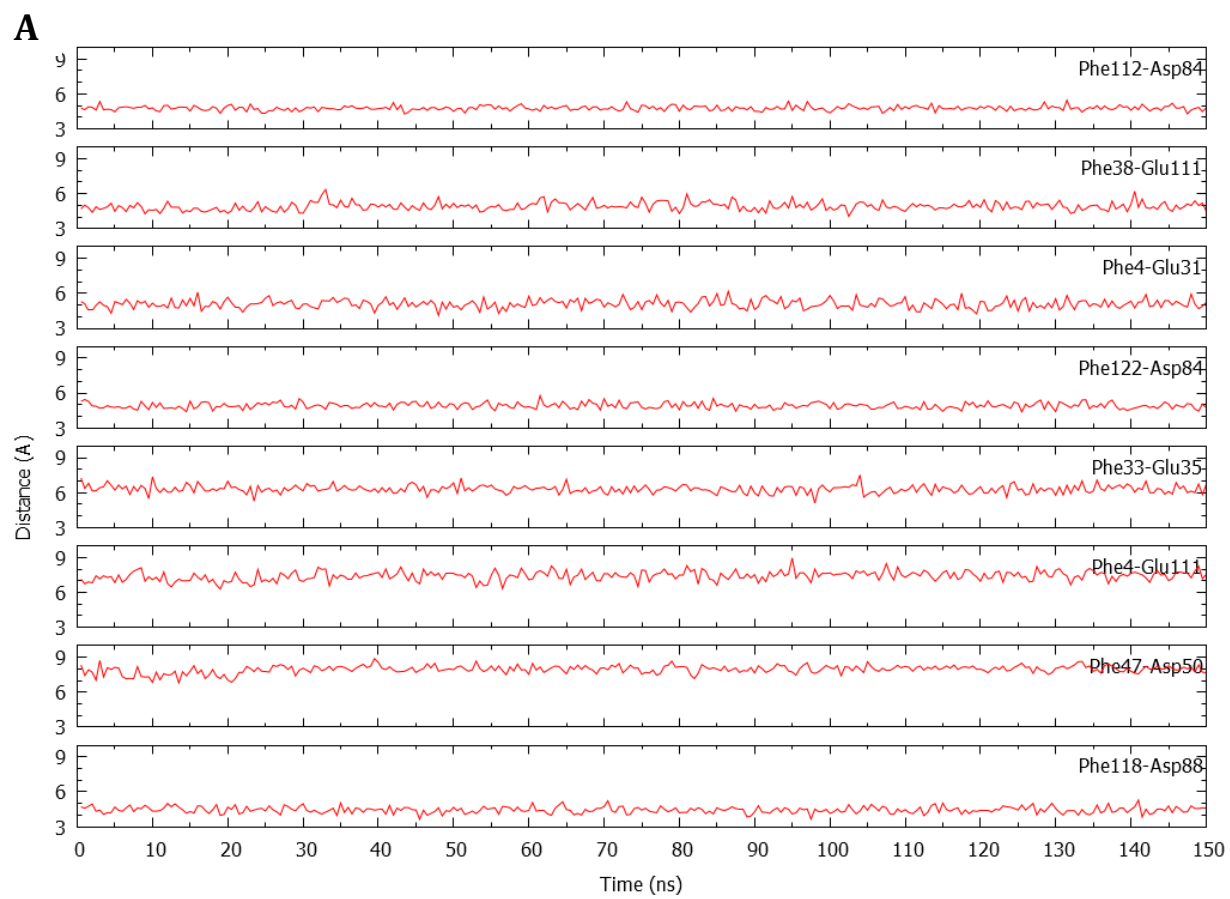
Figure 1.S6 continued

B
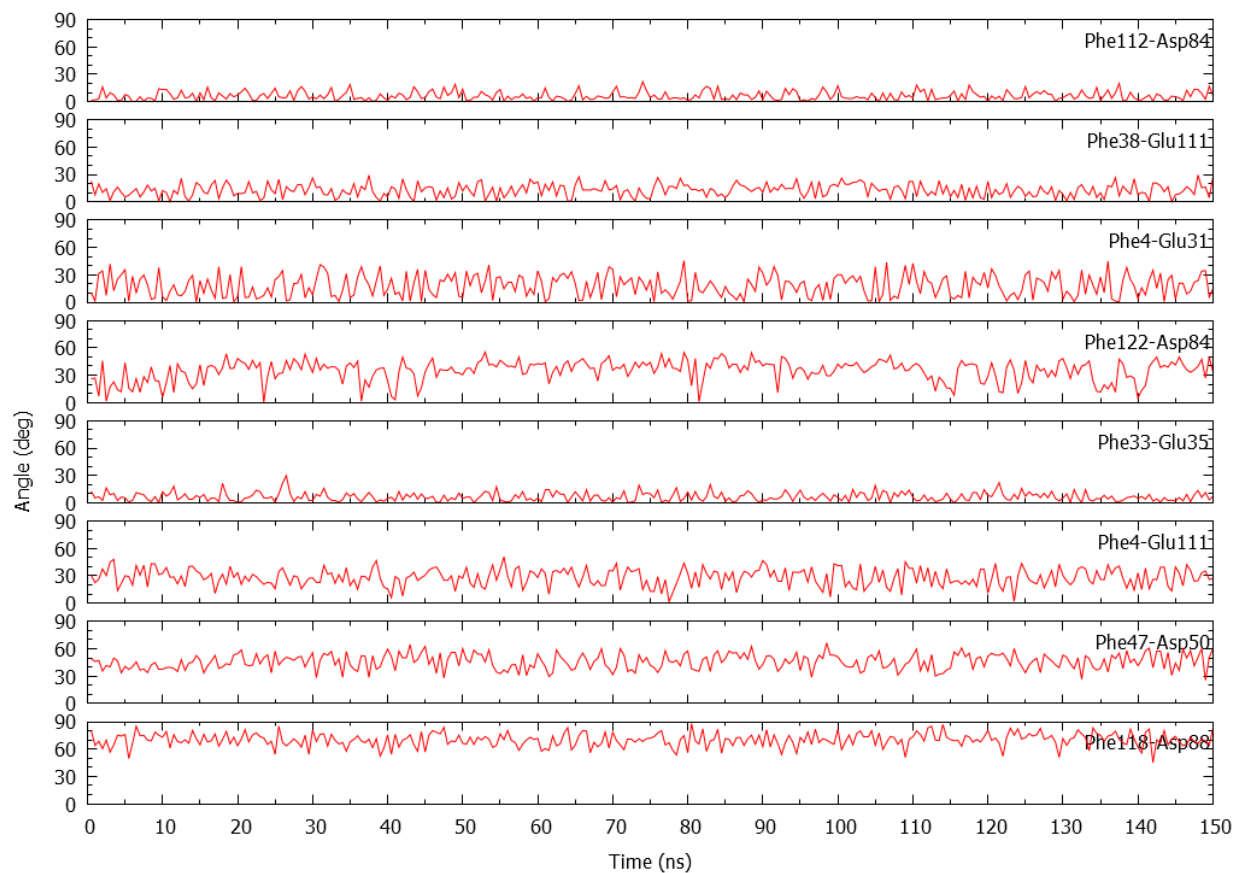
C
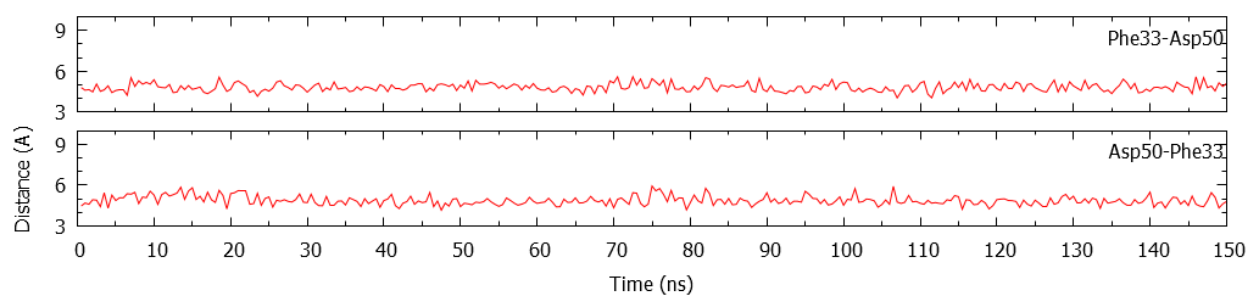
D
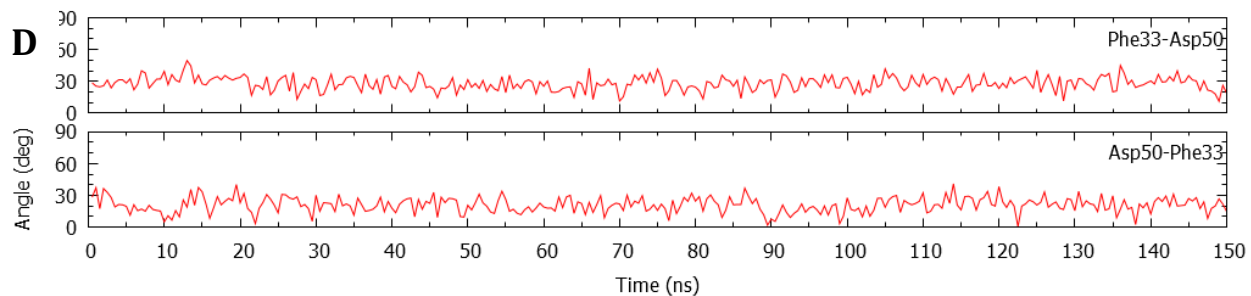
Figure 1.S6 continued
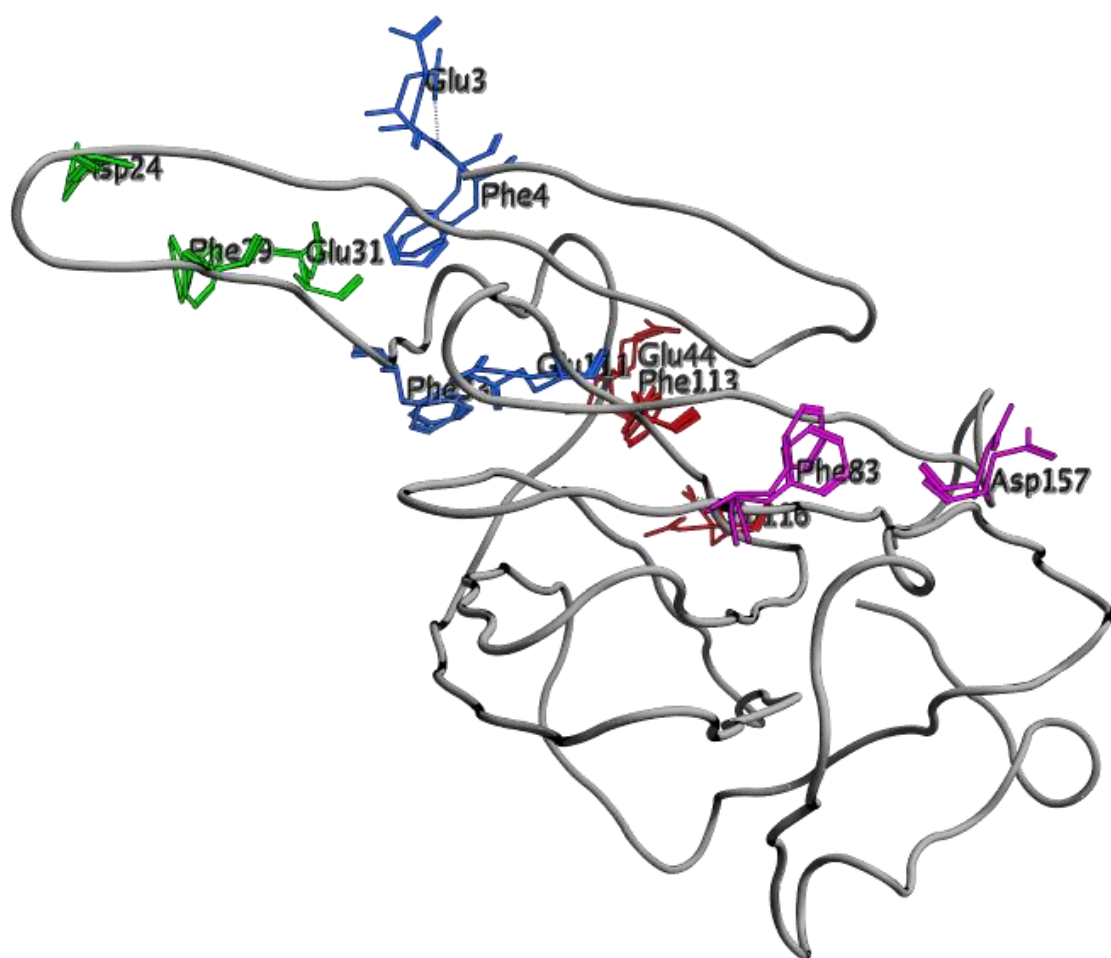
Figure 1.S7. Sidechain conformations sampled by anion-π pairs seen during MD simulation. Sidechain conformations sampled by the seven anion-π pairs only seen during the simulation, in the MD snapshots showing minimum and maximum angles for each pair (Table 1.3).

Figure 1.S8. Solvent molecules around anion-π pairs. Total number of water molecules within 4 Å of center of mass of Phe and center of charge of the anion for (A) eight intramolecular anion-π pairs in the crystal structure, (B) the two symmetrical pairs present at the dimer interface and (C) seven intramolecular anion-π pairs identified during MD simulations.

A
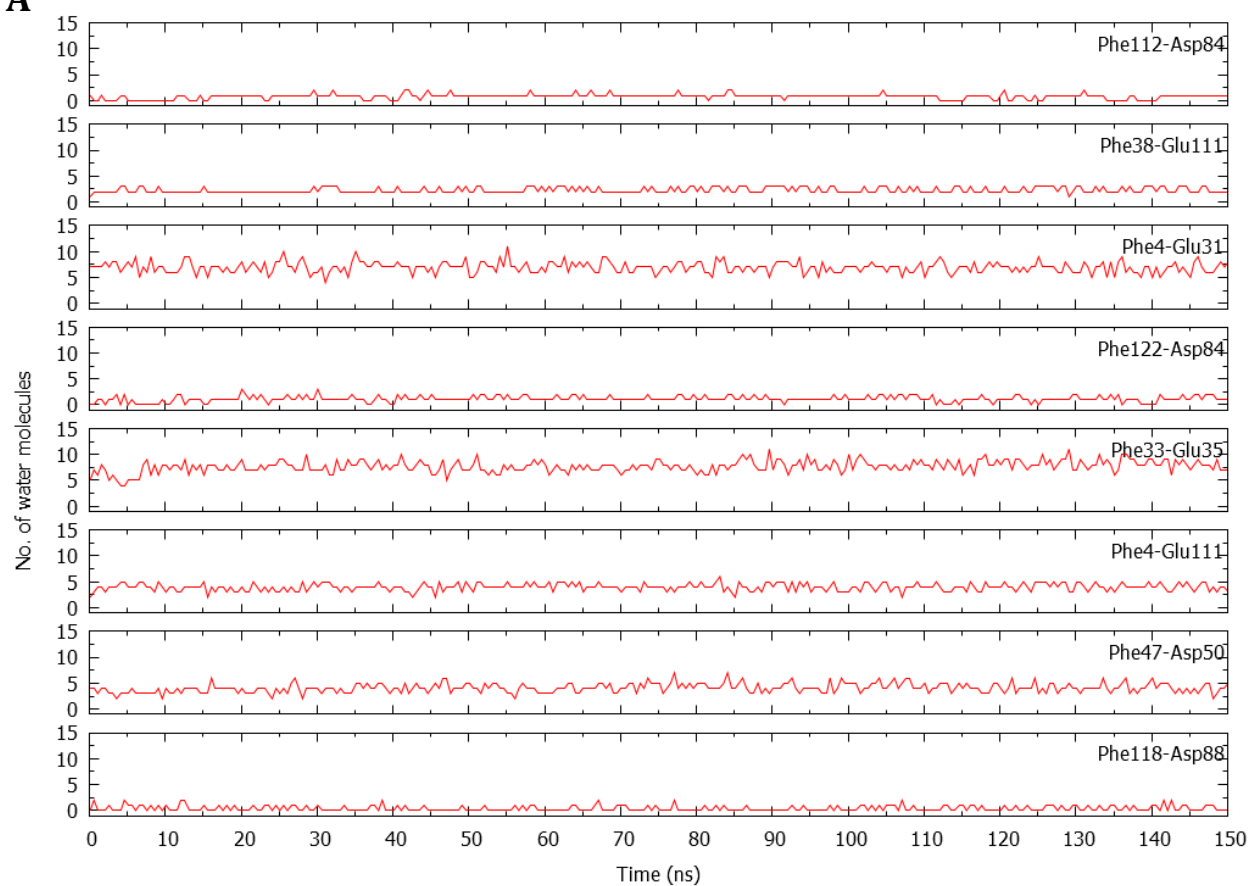
B
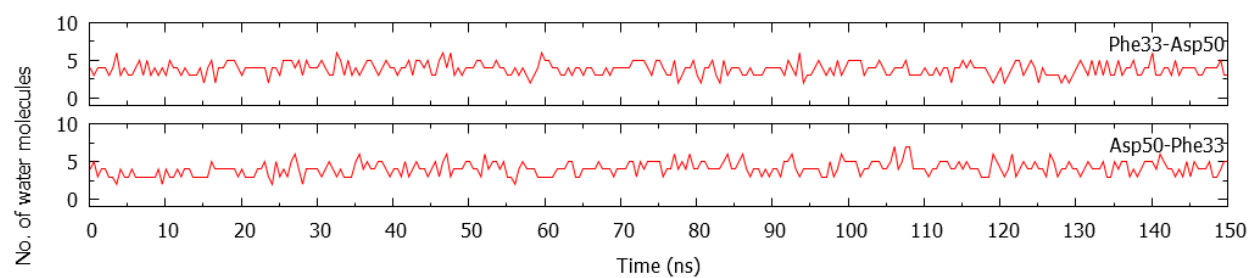
Figure 1.S8 continued

C
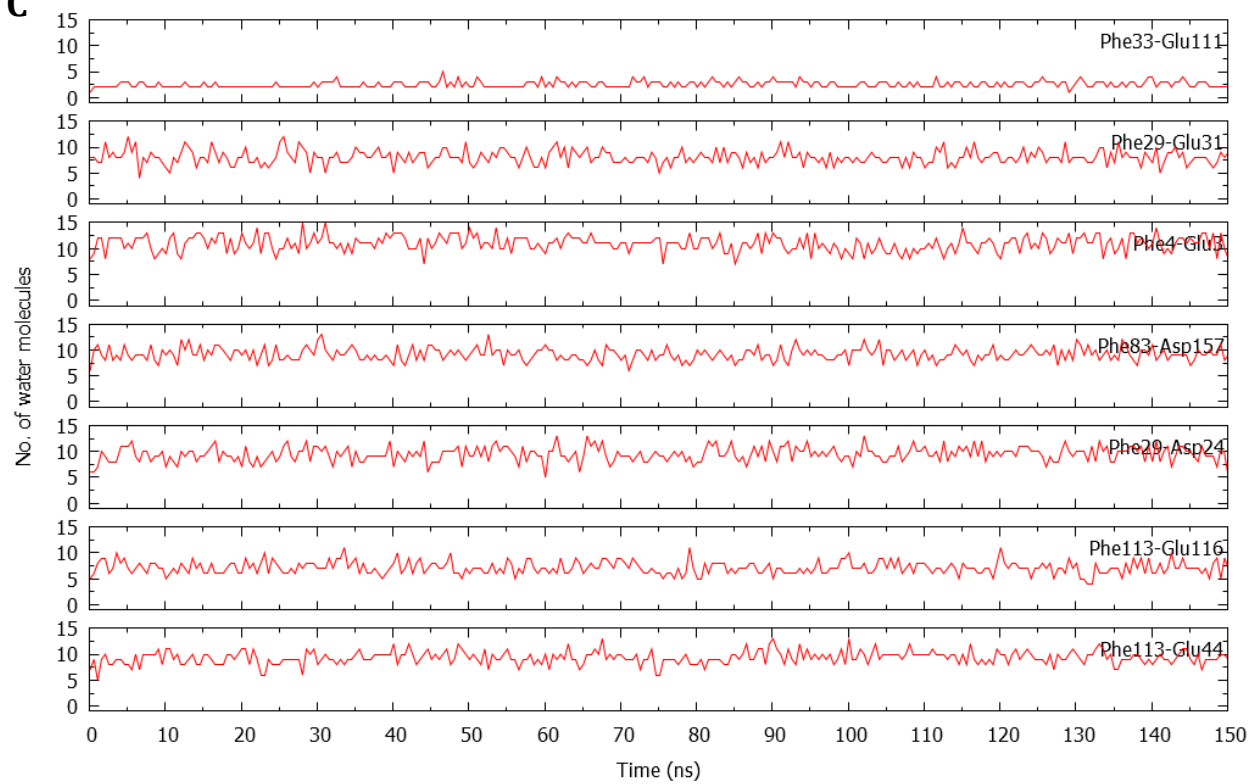
Figure 1.S8 continued

Figure 1.S9. Angle/distance conformation distribution and free energy (PMF) maps. Angle/distance conformation distribution maps and free energy (PMF) maps for the remaining anion-π pairs. The angle-distance values in the crystal structure are depicted as white dots on the plots.

Phe38-Glu111 (Xtal Distance= 5.49 A, Angle= 0.74 deg)

Phe38-Glu111 (Xtal Distance= 5.49 A, Angle= 0.74 deg)

Figure 1.S9 continued

Phe4-Glu31 (Xtal Distance= 5.03 A, Angle= 0.10 deg)

Phe4-Glu31 (Xtal Distance= 5.03 A, Angle= 0.10 deg)

Figure 1.S9 continued

Phe122-Asp84 (Xtal Distance= 5.11 A, Angle= 23.98 deg)

Phe122-Asp84 (Xtal Distance= 5.11 A, Angle= 23.98 deg)

Figure 1.S9 continued

Phe33-Glu35 (Xtal Distance= 6.09, Angle= 4.98)

Phe33-Glu35 (Xtal Distance= 6.09, Angle= 4.98)

Figure 1.S9 continued

Phe4-Glu111 (Xtal Distance= 6.17 A, Angle= 12.33 deg)

Phe4-Glu111 (Xtal Distance= 6.17 A, Angle= 12.33 deg)

Figure 1.S9 continued

Phe47-Asp50 (Xtal Distance= 6.88 A, Angle= 36.01 deg)

Phe47-Asp50 (Xtal Distance= 6.88 A, Angle= 36.01 deg)

Figure 1.S9 continued

Phe118-Asp88 (Xtal Distance= 4.31 A, Angle= 73.86 deg)

Phe118-Asp88 (Xtal Distance= 4.31 A, Angle= 73.86 deg)

Figure 1.S9 continued

Asp50-Phe33 (Xtal Distance= 5.02, Angle= 33.28)

Asp50-Phe33 (Xtal Distance= 5.02, Angle= 33.28)

Figure 1.S9 continued

Phe33-Glu111 (Xtal Distance= 7.13 A, Angle= 13.60 deg)

Phe33-Glu111 (Xtal Distance= 7.13 A, Angle= 13.60 deg)

Figure 1.S9 continued

Phe29-Glu31 (Xtal Distance= 7.23 A, Angle= 12.95 deg)

Phe29-Glu31 (Xtal Distance= 7.23 A, Angle= 12.95 deg)

Figure 1.S9 continued

Phe83-Asp157 (Xtal Distance= 8.39 A, Angle= 26.80 deg)

Phe83-Asp157 (Xtal Distance= 8.39 A, Angle= 26.80 deg)

Figure 1.S9 continued

Phe29-Asp24 (Xtal Distance= 8.99 A, Angle= 34.49 deg)

Phe29-Asp24 (Xtal Distance= 8.99 A, Angle= 34.49 deg)

Figure 1.S9 continued

Figure 1.S9 continued

Phe113-Glu44 (Xtal Distance= 8.11 A, Angle= 41.71 deg)

Phe113-Glu44 (Xtal Distance= 8.11 A, Angle= 41.71 deg)
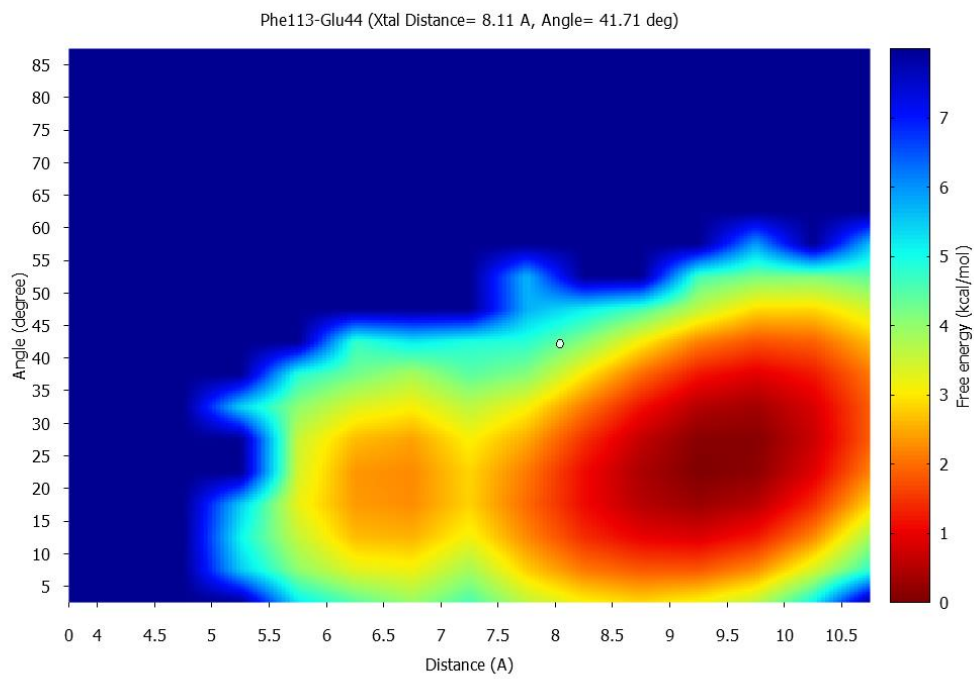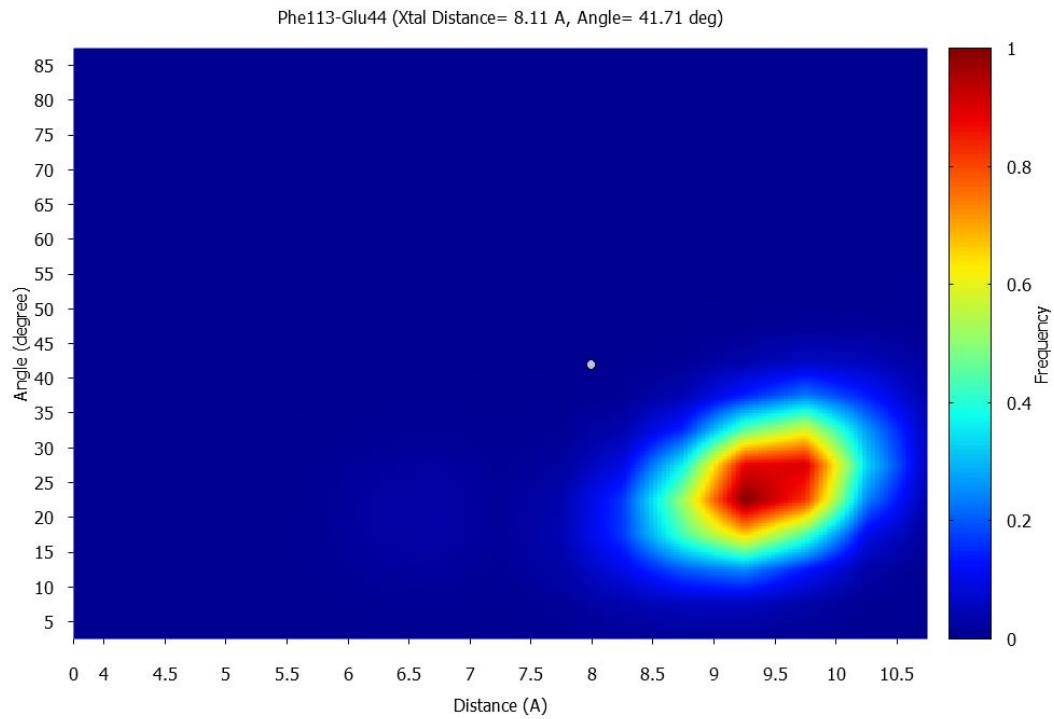
Figure 1.S9 continued

Chapter 2

STRUCTURAL AND FUNCTIONAL EVIDENCE

FOR TESTOSTERONE ACTIVATION OF

GPRC6A

Research described in this chapter is reproduced from an article submitted to the Journal of Molecular Endocrinology by Min Pi, Karan Kapoor, Yunpeng Wu, Susan E. Senogles, Satoru K. Nishimoto, Jeremy C. Smith, Jerome Baudry and L. Darryl Quarles.

My contribution in the paper includes (1) Designing and carrying out all computation research (2) Understanding the literature and interpretation of the results (3) Preparation of tables and figures (4) Writing and editing.

Abstract

GPRC6A is a multi-ligand G-protein coupled receptor (GPCR) that is activated by cations, L-amino acids, and osteocalcin. GPRC6A plays an important role in the regulation of testosterone (T) production and energy metabolism in mice. T has rapid, transcription-independent (non-genomic) effects that are mediated by a putative GPCR. We previously found that T can activate GPRC6A *in vitro*, but the possibility that T is a ligand for GPRC6A remains controversial. Here, we demonstrate direct T binding to GPRC6A by competitive binding assays and construct computational structural models of GPRC6A to identify potential binding poses of T. The modeled binding was confirmed by mutagenesis. Using *Gpr6ca^{-/-}* mice, we define several biological functions regulated by GPRC6A-dependent T rapid signaling, including T stimulation of insulin secretion in pancreatic islets and enzyme expression involved in the biosynthesis of T in Leydig cells. Together our data show that GPRC6A directly mediates the rapid signaling response to T and uncovers previously unrecognized endocrine networks.

Introduction

Classically, testosterone (T) stimulates target gene expression through the nuclear androgen receptor (AR). T is converted by 5α-reductase to dihydrotestosterone (DHT). DHT binds with high affinity to AR to form a transcription factor complex that translocates to the nucleus to activate androgen-responsive genes. Emerging clinical observations indicate that conversion of T to DHT is not obligatory and that T has clinically important functions distinct from activation of AR-mediated gene transcription in some tissues. T also elicits rapid, transcription-independent signaling responses, such as activation of ERK1/2, through a putative membrane G-protein coupled receptor (GPCR). The identity of this T-sensing GPRCR is not known and its physiological significance remains uncertain (1).

GPRC6A is a member of the class-C GPCR family (2-4), which includes the calcium sensing receptor (CasR). Like CasR, GPRC6A is activated by cations and L-amino acids; however, the physiological ligand for GPRC6A is thought to be the bone-derived hormone osteocalcin (Ocn) (5-7). Both GPRC6A and Ocn are involved in the regulation of energy metabolism and T production in mice (8-10), and Gprc6a$^{-/-}$ and Ocn$^{-/-}$ mice are phenocopies (6,11). GPRC6A is widely expressed, including β-cells, bone marrow stromal cells, monocytes, prostate cancer cells, skeletal muscle cells, keratinocytes, and Leydig cells, as well as other tissues known to be targeted by T (11-16). GPRC6A is also a candidate for this T-sensing GPCR (12,17). In this regard, overexpression of GPRC6A imparts the ability of extracellular T to illicit a rapid signaling response in HEK-293 cells, which lack the AR. Moreover, GPRC6A loss-of-function attenuates T stimulated rapid signaling in multiple tissues. Indeed, T rapid signaling is inhibited in bone marrow stromal cells derived from *Gprc6a$^{-/-}$* mice and in 22Rv1 prostate cancer cells after siRNA-mediated knockdown of *GPRC6A*. There is also indirect evidence that T regulates

89

luteinizing hormone secretion (18), stimulates insulin secretion in β-cells, and induces NADPH oxidase isozymes in keratenocytes (17) through GPRC6A (9).

The role of GPRC6A in mediating the effects of T remain controversial, because the binding sites for T in GPRC6A have not been identified, some studies have failed to confirm T-activation of GPRC6A *in vitro* (19), and the physiological roles of T activation of GPRC6A have not been directly demonstrated. Indeed, the global Gprc6a$^{-/-}$ mouse phenotype is complex and has multiple hormonal abnormalities that might indirectly account for the effects attributed to T-activation of GPRC6A (11). To examine direct T binding to GPRC6A we have performed competitive binding assays. Furthermore, we have constructed computational structural models of GPRC6A to identify potential binding poses of T, which enabled confirmatory mutagenesis experiments to be performed. Finally, we directly demonstrate the role of GPRC6A in mediating T effects in regulating β-cell and Leydig functions in cells derived from *Gprc6a$^{-/-}$* mice.

Materials and Methods

Animals

Generation of global *Gprc6a$^{-/-}$* mice has been previously reported (11). Mice were maintained and used in accordance with recommendations as described (National Research Council. 1985; Guide for the Care and Use of Laboratory Animals DHHS Publication NIH 86-23, Institute on Laboratory Animal Resources, Rockville, MD) and following guidelines established by the University of Tennessee Health Science Center Institutional Animal Care and Use Committee. The animal study protocol was approved by the institutional review board at University of Tennessee Health Science Center Institutional Animal Care and Use Committee.

Reagents and antibodies

Insulin (Mouse) Ultrasensitive ELISA kit and mouse C-peptide ELISA kit were obtained from ALPCO Diagnostics. T, glucose and insulin were purchased from Sigma.

Cell culture

All culture reagents were from Invitrogen. Human embryonic kidney HEK-293 cells were obtained from American Type Culture Collection. HEK-293 cells stably transfected with pcDNA3.mGPRC6A were created as previously described (12,20).

Mouse islets isolation and ligand stimulation

Primary islets were isolated using modified method as described (21,22). Briefly, after dissection and mincing, pancreata were digested with 3 ml/pancreas of a collagenase P (1 mg/ml; Roche) solution in complete HBSS (HBSS 1× supplemented with 20 mM Hepes, pH 7.4, and 2 mM CaCl$_2$) for 15 minutes in a 37 ^{o}C shaking water bath. Islets were subsequently purified through a Histopaque 1083 density centrifugation (Sigma). After centrifugation, the islet layer was transferred into petri dishes with wash buffer (HBSS with 10 mM Hepes and 1% FBS; Invitrogen), then handpicked, and cultured in low-glucose medium (RPMI 1640 with 5.6 mM glucose; Invitrogen) for 1 h before being treated for 1 h with T (80 nM). The insulin stimulation index (SI) was calculated as the ratio of media insulin concentrations in T or Ocn divided by the insulin concentration in low glucose conditions.

Real time RT-PCR

For quantitative real-time RT-PCR assessment of *insulin* and *glucagon* gene expression, we isolated total RNA from the islets or pancreas or other tissues of control and *Gprc6a*$^{β-cells\ cko}$ mice by standard TRIzol method (Invitrogen) and reverse transcribed 2.0 μg of total RNAs using cDNA synthesis kit (Bio-Rad). PCR reactions contained 100ng of template (cDNA or RNA), 300nM each of forward and reverse primer, and 1× iQ SYBR Green Supermix (Bio-Rad) in 50μL. Samples were amplified for 40 cycles in an iCycler iQ Real-Time PCR Detection System (Bio-Rad) with an initial melt at 95°C for 10 minutes, followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. PCR product accumulation was monitored at multiple points during each cycle by measuring the increase in fluorescence caused by the binding of SybrGreen I to dsDNA. The threshold cycle (C_t) of tested-gene product from the indicated genotype was normalized to the C_t for cyclophilin A. The primers for mouse cytochrome P450, family 17, subfamily a, polypeptide 1 (cyp17a1) consisted of mCyp17a.F56: agtcaaagacacctaatgccaag and mCyp17.R138: acgtctggggagaaacggt; for cytochrome P450, family 11, subfamily a, polypeptide 1 (cyp11a1) consisted of mCyp11a.F127: aggtccttcaatgagatccctt and mCyp11a.R263: tccctgtaaatggggccatac; and for the *cyclophilin A* consisted of cyclophilin A.For: ctgcactgccaagactgaat and cyclophilin A.Rev: ccacaatgttcatgccttct. Dissociation analysis was used to confirm the presence of a single transcript and the lack of primer-dimer amplification in all PCR reactions.

Saturation analysis of [3H]T binding in HEK-293 cells expressing GPRC6A

Cells were dislodged from tissue culture plates by using a rubber policeman and crude membranes were prepared by Dounce homogenization in 50 mM Tris-HCl pH 7.4 at 25°C with 120 mM NaCl, 1 mM EDTA, 10 mM MgCl$_2$ and a protease inhibitor cocktail for mammalian

tissue culture containing 104 mM AEBSF, 80 μM aprotinin, 2 mM leupeptin, 4 mM bestatin, 1.5 mM pepstatin A and 1.4 mM E-64. The homogenate was subjected to centrifugation at 14,000 x g for 10 minutes and the resulting crude total cellular pellet was resuspended in the buffer above. The membranes were either used immediately for binding or flash frozen at -70° C. The membrane protein concentration was quantified by the method of Bradford, using BSA as the protein standard (23).

For saturation analysis of [3H]T, a range from 0-20 nM [3H] T was used. Routinely, 50-75 ug of membrane protein was used per tube, and the analysis was performed in triplicate. The buffer consisting of 120 mM NaCl, 1 mM EDTA, 10 mM $MgCl_2$ was used as the binding and wash buffer and varying concentrations of [3H]T was added and allowed to incubate for 2 hours at ambient room temperature with agitation. Non-specific ligand binding was determined by the inclusion of 200 nM cold T to a parallel set of tubes. The reactions were terminated by dilution with ice cold buffer and filtration through GF/C glass fiber membranes to retain bound ligand. The retained radioactivity was quantified by liquid scintillation counting.

For saturation analysis of [3H]T binding in the presence of Ca^{2+}, 3mM $CaCl_2$ was added to binding and wash buffer and the assays performed as described.

The saturation data was analyzed using Graphpad Prism 6.0. For analysis of best fit, an extra sum of squares F test was performed to determine if a one site with hill slope or two site fit was superior for the saturation data.

For competition analysis, a concentration of 1nM [3H]T was used, with a concentration of 50-75 μg of cell membrane per tube. The analysis was performed in triplicate. L-Arginine (0-50 mM), or osteocalcin (0-50 μg/ml) were added to the assays and the incubations continued for 2 hours at ambient temperature. The assays were terminated by dilution with ice cold buffer and

filtration through GF/C glass fiber membranes and the retained radioactivity quantified by liquid scintillation counting. Non-specific binding was determined by the addition of 200 nM cold T. The data was analyzed by Graphpad Prism 6.0.

GPRC6A homology modeling

Multiple sequence alignment (MSA) of the GPRC6A sequence with eight family-C sequences was performed using MAFFT (Multiple Alignment using Fast Fourier Transform) sequence alignment and the EINSI alignment method (24). The family-C sequences used are: 1) human calcium-sensing receptor, 2) human probable G-protein coupled receptor-158, 3) human retinoic acid-induced protein-3, 4) human taste receptor type-1 member-1, 5) human metabotropic glutamate receptor-1 (PDB code 4OR2), 6) human metabotropic glutamate receptor-5 (PDB code 4OO9), 7) mouse metabotropic glutamate receptor-3 (PDB code 2E4U), and 8) human gamma-aminobutyric acid B receptor-1 (PDB code 4MQE). Sequences 1 through 4 have no available crystal structure, whereas sequences 5 and 6 have crystal structures for their transmembrane domains, and sequences 7 and 8 have crystal structures for their extracellular domains. In addition, the MAFFT multiple sequence alignment also included seven family-A sequences, which are closest in sequence identity with family-C GPCR's, with existing crystal structures of the short extracellular domain, transmembrane domain and short cytoplasmic domain. These sequences were: bovine rhodopsin receptor (PDB code 3CAP), turkey beta-1 adrenergic receptor 2VT4), human beta-2 adrenergic receptor (PDB code 2RH1), human adenosine A2A receptor (PDB code 2YDV), human 5-hydroxytryptamine receptor-2B (PDB code 4IB4), human 5-hydroxytryptamine receptor-1B (PDB code 4IAR) and rat neurotensin receptor type-1 (PDB code 4GRV).

From the MSA results, the structures of the mGlu-1 and mGlu-5 receptors (with crystal structures of transmembrane domains) were selected as templates for the transmembrane domain modelling. Missing regions in the cytoplasmic loop-2 (seq 688-691) and C-terminus (seq 844-845) of the mGluR-1 template structure were modelled using mGluR-5, and missing regions in the cytoplasmic loop-2 (seq 683-688) and the extracellular loop-2 (seq 721-728) of mGluR-5 template structure were modelled using mGluR-1. Ten main chain models with ten sidechain conformers per mainchain model were generated for each template using the MOE-2012 (Molecular Operating Environment, 2013.08; Chemical Computing Group Inc.) homology modeling facility with the CHARMM27 force-field (25). The GPRC6A homology models were validated using PolyPhobius (26). The best-scoring homology models; one from using mGluR-1 as a template and one from using mGluR-5 as a template, were selected for docking studies based on their predicted GB/VI scores (27), that rank the models based on Coulomb and Generalized Born interaction energies.

Identification of ligand binding sites in GPRC6A models

The recently published crystal structures of mGluR-1 and mGluR-5 contain negative allosteric modulators bound (28,29). The allosteric binding-site residues for these two receptors and the respective CasR and GPRC6A residues in the MSA are listed in Tables 1 and 2. In this site, Gly-667, Ser-669, Trp-795, Phe-798 and Tyr-802 in GPRC6A are found to be conserved between all the four receptors, with conserved or chemically similar residues also present for residues Phe-650, Phe-666, Cys-673, Phe-752, Met-755, Leu-756, Ala-763, Ala-794, Ile-818 and Ile-825.

A number of agonists/antagonists have been reported for the closest family-C member, CasR, that bind to the transmembrane domain of the receptor (38). All residues involved in binding

95

in CasR identified by mutagenesis experiments (Ser-657, Phe-668, Arg-680, Phe-684, Phe-688, Leu-773, Leu-776, Trp-818, Phe-821, Glu-837, Ile-841) are conserved in GPCR6A except Ser-657, Leu-773 and Leu-776, which are replaced by Cys-639, Phe-752 and Met-755, respectively. CasR calcimimetic calindol and the calcilytic NPS2143 have been found to antagonize mouse-GPRC6A (28). The binding site residues for these antagonists in mouse-GPRC6A found by mutagenesis experiments (Phe-666, Phe-670, Trp-797, Glu-816) are conserved in both human-GPRC6A and CasR. Several 2-phenyl-indole derived allosteric antagonists have also been reported in mouse-GPRC6A that are selective only for this receptor among the family-C family receptors (7). The possible binding site residue reported for these antagonists in mouse-GPRC6A (Ile-759) is conserved in human-GPRC6A and not conserved in CasR (Thr-780), mGluR-1 (Asn-760) or mGluR-5 (Asn-747).

The above binding site information from both conservation of residues in the MSA and the mutagenesis experiments was used for identifying binding sites in GPRC6A used in the present docking studies.

Docking of T

Docking of T to the transmembrane domain of the selected GPRC6A homology models was carried out using MOE-2012's Docking facility with the CHARMM27 force-field. CHARMM parameters for the ligands were generated by MOE from a fragment-based approach. The binding site was defined in MOE from the binding-site residues described above using MOE's Site Finder facility. Binding site residue side-chains were allowed to be flexible during the docking using a tethering weight of 0.1. London dG free energy scores were used to rank poses of the docked ligand (36). Potential hydration sites were calculated in the binding sites of both homology models

using the MOE's Solvate facility. Water molecules are predicted to be potentially present in both homology models' binding sites, which were only implicitly included in the docking calculations.

Measurement of cAMP accumulation

HEK-293 and HEK-293 transfected with mouse GPRC6A cDNA cells (10^5 cells/well) (12) were cultured in triplicate in 12-well plates in DMEM supplemented with 10% fetal bovine serum and 1% P/S (100 units/ml of penicillin and 100 ug/ml of streptomycin) for 48 hours followed by overnight incubation in DMEM/F12 containing 0.1% BSA to achieve quiescence. Quiescent cells were treated with vehicle control, 40 nM T or 60 ng/ml bovine osteocalcin (Ocn) or 40 nM T and 60 ng/ml Ocn together for 30 minutes at 37 °C. Then the reaction was stopped and the cells lysed with 0.5 ml 0.1 N HCl. cAMP levels were measured by using Cyclic AMP EIA kit (Cayman Chemical, Ann Arbor, MI) following the manufacture's protocol.

Leydig cells isolation

To isolate Leydig cells from wild type and $Gprc6a^{-/-}$ mice, the testes from 10~12 week-old mice were obtained and enzymatically dispersed with the method as previously described (30). Briefly, the testes from wild type and $Gprc6a^{-/-}$ mice were decapsulated and dispersed with 0.25 mg/ml collagenase (Invitrogen) in medium 199 for 10 min at 4°C with 70~90 rpm shaking. The separated cells were filtered through 100 nm filter, centrifuged at 250 X g and resuspended in 55% isotonic Percoll (Sigma). Following density gradient centrifugation at 25 000 X g for 45 min at 4°C, the PLC fraction was collected between densities of 1.064 and 1.070 g/ml. The cells were washed with HBSS (5 time of volume; Invitrogen), centrifuged at 250 X g and The cells were washed with HBSS (5 time of volume), centrifuged at 250 X g and resuspended in phenol red-free

medium (DMEM/Ham's F-12, D-2906, Sigma) supplemented with 1 mg/ml bovine serum albumin (BSA).

Statistics

We evaluated differences between groups by one-way analysis of variance, followed by *a post-hoc* Tukey's test. Significance was set at $p < 0.05$. All values are expressed as means ± SEM. All computations were performed using the Statgraphic statistical graphics system (STSC Inc.).

Results

Binding of T to GPRC6A

To examine whether T binds to GPRC6A, we performed saturation analysis of [3H]T binding in HEK293 cells overexpressing GPRC6A (Figure 1A). We found that T binds to GPRC6A with a K_d for [3H]T of 9.5 ± 2.41 *nM* (n=5) in the absence of added calcium (control) (Figure 1A), with the data best fitting a model with a Hill slope of 1.9 ± 0.25. GPCRC6A is known to be activated by calcium (7,31). Adding 3*mM* calcium resulted in a shift in the K_d of [3H]T to 7.7 ± 1.7 *nM* with a Hill slope of 1.17 ± 0.13 (n=4) (Fig 1A). Ocn also activates GPRC6A (10,14). Interestingly, Ocn (40 ng/ml) shifted the K_d for [3H]T to 3.3 ± 0.5 *nM*, with a Hill slope of 1.5 ± 0.2 compared to 9.8 ± 1.7 *nM* with a hill slope of 2.0 ± 0.2 in the controls (n=3) (Figure 1B). These data indicate that T binds to GPRC6A and this interaction is modulated by calcium and Ocn.

Computational modeling of T binding to GPRC6A

To investigate the molecular basis for T binding to GPRC6A (9,32), structural models of GPRC6A were constructed and used to identify potential T binding poses. The GPRC6A

98

homology models developed are shown in Figure 2. Helix regions in multiple sequence alignments are highly conserved among GPCRs, whereas the loop regions exhibit low sequence similarity, as shown in Figure 2A. The GPRC6A sequence exhibits a similarity of 43.2% and 44.7% to the mGluR-1 and mGluR-5 receptors respectively, the only sequences for which crystal structures of the transmembrane domain exist (sequence-2 and sequence-3 in Figure 2B). These receptor structures were selected as templates for the homology modeling calculations and the corresponding structural models selected for docking studies.

The highest scoring models are shown in Figure 2C and 2D. These models exhibit very similar structures, with a root mean square deviation (RMSD) of ~4 Å^2 between the backbone atoms. These models were validated against Hidden Markov secondary structure predictions generated by PolyPhobius, and found to be consistent (Table 3 and 4), with only one loop region and one transmembrane helix in the homology models exhibiting deviation from the PolyPhobius-predicted structure for five or more residues.

Docking of Testosterone

Next, we docked T to the above homology models as described in the Methods. The binding sites found in the mGluR-1 and mGluR-5 based models of GPRC6A are shown in Figure 3A and B respectively. The docking scores and the binding pocket residues in these sites are listed in Table 5 for the top-ranking binding poses of T. Thirteen residues in the mGluR-1 model, and twelve in mGluR-5 were identified as possibly interacting with T. In both homology models, the hydrophobic core of T is surrounded by aliphatic side-chains and/or aromatic Phe (F), Trp (W) or Tyr (Y) residues. The hydroxyl and carbonyl moieties of T are located in hydrogen bond acceptor-rich regions, either on side chains or on the backbones of residues in the binding pocket. Six of

99

these predicted binding residues are found in both models, including F-666, F-670, F-752, M-755, L-756 and L-759 (Fig 2 and Table 5).

Validation of T-binding and computational model by mutagenesis

To validate the binding model, we compared the T activation of wild-type and mutant GPRC6A transfected into HEK-293 cells. We observed a dose-dependent effect of T to stimulate cAMP accumulation (Figure 4A) and ERK phosphorylation (Figure 4B) in HEK-293 cells transfected with wild-type GPRC6A. A significant response was observed at 5 nM and a maximal response at a T concentration of 60 nM, using cAMP accumulation as the read out (Figure 4A). No response to T is observed in non-transfected HEK-293 cells in ERK phosphorylation (Figure 4B low panel). Consistent with prior reports (13,14), L-Arginine (L-Arg) (20 mM) stimulated cAMP to a similar magnitude as T in HEK-293 cells expressing GPRC6A but not in untransfected HEK-293 cells (Figure 4C). We also found that the protein kinase A (PKA) inhibitor, H89, blocked GPRC6A-mediated T stimulated ERK phosphorylation (Figure 4D), indicated that PKA is involved in the GPRC6A signaling pathway.

To validate the computational modeling of T binding to GPRC6A, we mutated Phe-666 (F-666) (Fig 2 and Table 5) into alanine by site-directed mutagenesis to investigate its possible interaction with T (33). The Phe-666-Ala (F666A) mutant and the wild type mGPRC6A were transiently transfected into HEK293 cells. Expression of the wild-type and F666A mutant GPRC6A receptors was assessed by Western blotting using a Myc antibody, which recognized the Myc epitope located at the amino-terminal tail of the WT and mutant receptors (33). Expression of the F666A mutant receptor was comparable to that of the WT receptor (Figure 4E, right panel). However, the F666A mutant showed a significant and reproducible decrease in the ERK

100

phosphorylation compared to that of WT receptor in transfected HEK293 cells (Figure 4E, left panel), thus validating the computational model.

Loss of T responses in *Gprc6a^-/-* mice

To evaluate the role of GPRC6A in sensing T *in vivo*, we examined *Egr-1* expression in the pancreas and testis of *Gprc6a^-/-* mice and wild-type littermates after the administration of 200 mg/kg T by intraperitoneal injection. We found that T administration significantly increased *Egr-1* expression in the pancreas and testis of wild-type mice, whereas *Gprc6a^-/-* mice showed no response to T (Figure 5A).

To test this biological response *ex vivo*, we examined the effect of GPRC6A activation in isolated islets and Leydig cells from wild-type and *Gprc6a^-/-* mice. We found that T increased the insulin stimulation index in islets isolated from wild-type but not *Gprc6a^-/-* mice (Figure 5B). Next, we confirmed that GPRC6A is expressed in β-cells. We found that *Gprc6a* transcripts were present in INS-1 rat β-cell line (Figure 5C) (34). In addition, we observed a dose-dependent effect of T to stimulate insulin secretion in INS-1 β-cells (Figure 5D). Finally, we found that 100 nM T significantly stimulated T biosynthesis enzymes, cholesterol side-chain cleavage enzyme (*Cyp11a*) and steroid 17-alpha-monooxygenase (*Cyp17a*) expression in Leydig cells from wild-type mice but not in *Gprc6a^-/-* mice (Figure 5E).

Discussion

The diverse ligand specificity of GPRC6A and the lack of understanding of the structural basis for ligand/receptor interactions have resulted in controversy regarding the possibility that GPRC6A functions as a T-sensing GPCR (12,32). In the current study, we showed that T

101

specifically binds to GPRC6A using [3H]T radioligand binding assays and provided a structural model in which T binds to the heptahelical domain of GPRC6A. Mutagenesis of a critical T binding site identified in this model inhibited T activation of GPRC6A *in vitro*, confirming the model. Finally, we showed that T activates GPRC6A signaling *in vitro* and *in vivo*, leading to regulation of insulin secretion by β-cells and enzymes controlling T production by Leydig cells. Collectively, these data indicate that GPRC6A is a T-sensing receptor that mediates the non-classical, rapid membrane signaling responses to T.

Our computational model of GPRC6A suggests that the binding pocket for T corresponds to the common allosteric site for Class C GPCRs. Residues in GPRC6A predicted to be involved in docking T, include Phe-666, Phe-670, Phe-752, Met-755, Leu-756, Ile-759, Ala-794, Trp-795, Phe-798, Glu-814 and Ile-818. Previous mutagenesis-based studies of Class C GPCR allosteric sites showed that positions of the T binding-pocket residues have also been implicated in the binding of mGluR-1, mGluR-5, and CasR allosteric modulators (Table 1 and 2). Here, we have confirmed the importance of the Phe-666 in T activation of GPRC6A by mutagenesis studies; substitution of Phe-666 with Ala resulted in marked attenuation of T activation of the mutant GPRC6A compared to the wild-type receptor. Phe-666 also overlaps the site for the calcimimetic and calcilytic binding to the closely related calcium sensing receptor, CasR (33). Calcimimetics and calcilytic also respectively stimulate and antagonize mouse-GPRC6A (7,33). Other residues identified in the model are candidates for additional directed mutagenesis studies to confirm their role in T-binding and activation of GPRC6A. Antagonists may also interact with this binding site in GPRC6A (35). Of note, the 2-phenyl-indole derived allosteric antagonist, which is specific for only GPRC6A among the class-C GPCR's, is found to interact with Ile-759. This residue is not conserved in other class-C GPCR's, and may be important in conferring specificity to this receptor.

102

Thus, this site could potentially control the binding of a number of different allosteric modulators (antagonists and agonists) of GPRC6A. Future molecular dynamics computer simulations of the models, followed by selection of representative protein conformers and ensemble-docking of ligands in the snapshots, while computationally expensive, will likely be able to identify the most stable binding modes of ligands in GPRC6A and allow a detailed structure-based description of binding poses (36), as well as identification of novel small molecules that selectively modulate GPRC6A.

The identification of the GPCR mediating the rapid response to T now permits us for the first time to identify the physiological processes regulated by this T-sensing GPRC and to distinguish these from the functions of the androgen receptor (AR). In the current studies, we show that GPRC6A is an important G-protein coupled receptor for T signaling in β-cells. In this regard, T dose-dependently stimulated insulin secretion in INS-1 cells and activated cAMP-dependent pathways in HEK-293 cells expressing GPRC6A. More importantly, T administration stimulated rapid Egr-1 signaling in the pancreas of wild-type mice, but this response was lost in the pancreas of $Gprc6a^{-/-}$ mice. Pancreatic islets from $Gprc6a^{-/-}$ mice had a diminished insulin secretion index. The direct regulation of insulin secretion and β-cell proliferation by GPRC6A raises the possibility of a link between T, male puberty and β-cell mass. Although increased β-cell mass with an accelerated rate of somatic growth of adolescence has not been identified (37), T has been shown to protect against glucotoxicity-induced apoptosis of pancreatic β-cells (38) and T deficiency can contribute to the development of metabolic syndrome (39). Our data suggest that some of the protective effect of T on pancreatic β-cell mass and insulin secretion could be mediated by GPRC6A, in addition to androgen receptor-mediated mechanisms (39).

We also showed here that T activates GPRC6A in Leydig cells. Isolated Leydig cells from *Gprc6a*$^{-/-}$ mice exhibited diminished expression of the T-stimulated *Cyp11a* and *Cyp17a*, which regulate T biosynthesis. In addition, T stimulated rapid Egr-1 signaling in the testes of wild-type mice, but not in the testes of *Gprc6a*$^{-/-}$ mice, indicating that loss-of-GPRC6A disrupts a putative autocrine positive feedback loop whereby T regulates its own production. Our observation could account for the heretofore unexplained finding that testicular interstitial fluid has effects on Leydig cell testosterone secretion that are independent of LH and hCG (40), as well as the finding that Leydig cell production of T and Cyp17a expression are unaffected by the loss of AR (41). T activation of T production through GPRC6A is also consistent with the actions of another GPRC6A ligand, Ocn, to stimulate T production (12,14).

Other studies have shown that GPRC6A is important in prostate cancer. GPRC6A is increased in prostate cancer cells and is linked to prostate cancer progression in genome wide associative studies (13,42). Ablation of GPRC6A attenuates prostate cancer progression in a mouse model (13), suggesting that the AR-independent effects of T in resistant prostate cancer might be mediated by GPRC6A. GPRC6A has also been shown to mediate the functional response to T in skin keratinocytes (17).

The fact that GPRC6A is also expressed in liver, adipocytes, muscle, skin, and bone, as well as other tissues, and is activated by T, as well as multiple structurally distinct ligands, predict additional metabolic functions and integration of endocrine networks involving multiple organs. Indeed, global *Gprc6a*$^{-/-}$ mice, in addition to abnormalities of β-cells, Leydig cells, and prostate, also have abnormalities in hepatic, bone, muscle and adipocyte function. Since GPRC6A and AR are also co-expressed in many of these tissues, there may be other organs where differences in rapid and classical T effects may be physiologically important. Interestingly, consistent with AR-

independent effects of T that are mediated through GPRC6A, recent clinical studies show that graded doses of testosterone administration to patients treated with 5α-reductase inhibitor has effects on lean body and muscle mass, sexual functions, hematocrit, cholesterol and other biological effects that are consistent with the predicted functions of GPRC6A derived from the phenotype of *Gprc6a*$^{-/-}$ mice (43).

In conclusion, an understanding of the tissue specific functions of GPRC6A and the structural basis for its binding to distinct ligands is revealing new connections between endocrine networks that heretofore were not thought to be related. Future studies that conditionally delete GPRC6A and AR in these and other tissues are now possible to differentiate between GPRC6A and AR in mediating T functions. The structural modeling presented here also represents a first step towards developing agonists and antagonists for this novel druggable target.

LIST OF REFERENCES

1. Benten WP, Lieberherr M, Stamm O, Wrehlke C, Guo Z, Wunderlich F. Testosterone signaling through internalizable surface receptors in androgen receptor-free macrophages. Mol Biol Cell 1999; 10:3113-3123

2. Karsenty G, Oury F. Regulation of male fertility by the bone-derived hormone osteocalcin. Mol Cell Endocrinol 2014; 382:521-526

3. Oury F, Khrimian L, Denny CA, Gardin A, Chamouni A, Goeden N, Huang YY, Lee H, Srinivas P, Gao XB, Suyama S, Langer T, Mann JJ, Horvath TL, Bonnin A, Karsenty G. Maternal and offspring pools of osteocalcin influence brain development and functions. Cell 2013; 155:228-241

4. Wellendorph P, Brauner-Osborne H. Molecular cloning, expression, and sequence analysis of GPRC6A, a novel family C G-protein-coupled receptor. Gene 2004; 335:37-46

5. Wei J, Hanna T, Suda N, Karsenty G, Ducy P. Osteocalcin promotes beta-cell proliferation during development and adulthood through Gprc6a. Diabetes 2014; 63:1021-1031

6. Oury F, Ferron M, Huizhen W, Confavreux C, Xu L, Lacombe J, Srinivas P, Chamouni A, Lugani F, Lejeune H, Kumar TR, Plotton I, Karsenty G. Osteocalcin regulates murine and human fertility through a pancreas-bone-testis axis. The Journal of clinical investigation 2013; 123:2421-2433

7. Pi M, Faber P, Ekema G, Jackson PD, Ting A, Wang N, Fontilla-Poole M, Mays RW, Brunden KR, Harrington JJ, Quarles LD. Identification of a novel extracellular cation-sensing G-protein-coupled receptor. J Biol Chem 2005; 280:40201-40209

8.   Fulzele K, Riddle RC, DiGirolamo DJ, Cao X, Wan C, Chen D, Faugere MC, Aja S, Hussain MA, Bruning JC, Clemens TL. Insulin receptor signaling in osteoblasts regulates postnatal bone acquisition and body composition. Cell 2010; 142:309-319

9.   Ferron M, Wei J, Yoshizawa T, Del Fattore A, DePinho RA, Teti A, Ducy P, Karsenty G. Insulin signaling in osteoblasts integrates bone remodeling and energy metabolism. Cell 2010; 142:296-308

10.   Oury F, Sumara G, Sumara O, Ferron M, Chang H, Smith CE, Hermo L, Suarez S, Roth BL, Ducy P, Karsenty G. Endocrine regulation of male fertility by the skeleton. Cell 2011; 144:796-809

11.   Pi M, Chen L, Huang MZ, Zhu W, Ringhofer B, Luo J, Christenson L, Li B, Zhang J, Jackson PD, Faber P, Brunden KR, Harrington JJ, Quarles LD. GPRC6A null mice exhibit osteopenia, feminization and metabolic syndrome. PLoS One 2008; 3:e3858

12.   Pi M, Parrill AL, Quarles LD. GPRC6A mediates the non-genomic effects of steroids. J Biol Chem 2010; 285:39953-39964

13.   Pi M, Quarles LD. Multiligand specificity and wide tissue expression of GPRC6A reveals new endocrine networks. Endocrinology 2012; 153:2062-2069

14.   Pi M, Wu Y, Quarles LD. GPRC6A mediates responses to osteocalcin in beta-cells in vitro and pancreas in vivo. J Bone Miner Res 2011; 26:1680-1683

15.   Pi M, Zhang L, Lei SF, Huang MZ, Zhu W, Zhang J, Shen H, Deng HW, Quarles LD. Impaired osteoblast function in GPRC6A null mice. J Bone Miner Res 2010; 25:1092-1102

16.   Tsuka S, Aonuma F, Higashi S, Ohsumi T, Nagano K, Mizokami A, Kawakubo-Yasukochi T, Masaki C, Hosokawa R, Hirata M, Takeuchi H. Promotion of insulin-induced glucose

uptake in C2C12 myotubes by osteocalcin. Biochemical and biophysical research communications 2015; 459:437-442

17. Ko E, Choi H, Kim B, Kim M, Park KN, Bae IH, Sung YK, Lee TR, Shin DW, Bae YS. Testosterone stimulates Duox1 activity through GPRC6A in skin keratinocytes. J Biol Chem 2014;

18. Wei J, Hanna T, Suda N, Karsenty G, Ducy P. Osteocalcin promotes beta-cell proliferation during development and adulthood through Gprc6a. Diabetes 2013;

19. Jacobsen SE, Norskov-Lauritsen L, Thomsen AR, Smajilovic S, Wellendorph P, Larsson NH, Lehmann A, Bhatia VK, Brauner-Osborne H. Delineation of the GPRC6A receptor signaling pathways using a mammalian cell line stably expressing the receptor. The Journal of pharmacology and experimental therapeutics 2013; 347:298-309

20. Kuang D, Yao Y, Lam J, Tsushima RG, Hampson DR. Cloning and characterization of a family C orphan G-protein coupled receptor. J Neurochem 2005; 93:383-391

21. Liu M, Shapiro ME. A new method for isolation of murine islets with markedly improved yields. Transplant Proc 1995; 27:3208-3210

22. Gerling IC, Serreze DV, Christianson SW, Leiter EH. Intrathymic islet cell transplantation reduces beta-cell autoimmunity and prevents diabetes in NOD/Lt mice. Diabetes 1992; 41:1672-1676

23. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 1976; 72:248-254

24. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013; 30:772-780

25. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FT, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. The journal of physical chemistry B 1998; 102:3586-3616

26. Kall L, Krogh A, Sonnhammer EL. An HMM posterior decoder for sequence feature prediction that includes homology information. Bioinformatics 2005; 21 Suppl 1:i251-257

27. Labute P. The generalized Born/volume integral implicit solvent model: estimation of the free energy of hydration using London dispersion instead of atomic surface area. J Comput Chem 2008; 29:1693-1698

28. Wu H, Wang C, Gregory KJ, Han GW, Cho HP, Xia Y, Niswender CM, Katritch V, Meiler J, Cherezov V, Conn PJ, Stevens RC. Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator. Science 2014; 344:58-64

29. Dore AS, Okrasa K, Patel JC, Serrano-Vega M, Bennett K, Cooke RM, Errey JC, Jazayeri A, Khan S, Tehan B, Weir M, Wiggin GR, Marshall FH. Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. Nature 2014; 511:557-562

30. Xiao YC, Hardy DO, Sottas CM, Li XK, Ge RS. Inhibition of LH-stimulated androgen production in rat immature Leydig cells: Effects on nuclear receptor steroidogenic factor 1 by FGF2. Growth Factors 2010; 28:1-9

31. Christiansen B, Hansen KB, Wellendorph P, Brauner-Osborne H. Pharmacological characterization of mouse GPRC6A, an L-alpha-amino-acid receptor modulated by divalent cations. Br J Pharmacol 2007; 150:798-807

32. Clemmensen C, Smajilovic S, Wellendorph P, Brauner-Osborne H. The G protein-coupled receptor, class C, group 6, subtype A (GPRC6A) receptor: From cloning to physiological function. Br J Pharmacol 2013;

33. Faure H, Gorojankina T, Rice N, Dauban P, Dodd RH, Brauner-Osborne H, Rognan D, Ruat M. Molecular determinants of non-competitive antagonist binding to the mouse GPRC6A receptor. Cell Calcium 2009; 46:323-332

34. Hohmeier HE, Mulder H, Chen G, Henkel-Rieger R, Prentki M, Newgard CB. Isolation of INS-1-derived cell lines with robust ATP-sensitive K+ channel-dependent and -independent glucose-stimulated insulin secretion. Diabetes 2000; 49:424-430

35. Gloriam DE, Wellendorph P, Johansen LD, Thomsen AR, Phonekeo K, Pedersen DS, Brauner-Osborne H. Chemogenomic discovery of allosteric antagonists at the GPRC6A receptor. Chemistry & biology 2011; 18:1489-1498

36. Ellingson SR, Miao Y, Baudry JY, Smith JC. Multi-Conformer Ensemble Docking to Difficult Protein Targets. The journal of physical chemistry B 2014;

37. Meier JJ, Butler AE, Saisho Y, Monchamp T, Galasso R, Bhushan A, Rizza RA, Butler PC. Beta-cell replication is the primary mechanism subserving the postnatal expansion of beta-cell mass in humans. Diabetes 2008; 57:1584-1594

38. Hanchang W, Semprasert N, Limjindaporn T, Yenchitsomanus PT, Kooptiwut S. Testosterone protects against glucotoxicity-induced apoptosis of pancreatic beta-cells (INS-1) and male mouse pancreatic islets. Endocrinology 2013; 154:4058-4067

39. Zitzmann M. Testosterone deficiency, insulin resistance and the metabolic syndrome. Nat Rev Endocrinol 2009; 5:673-681

40. Sharpe RM. Intratesticular regulation of testosterone secretion: comparison of the effects and interactions of hCG, an LHRH agonist and testicular interstitial fluid on Leydig cell testosterone secretion in vitro. Mol Cell Endocrinol 1985; 41:247-255

41. O'Hara L, McInnes K, Simitsidellis I, Morgan S, Atanassova N, Slowikowska-Hilczer J, Kula K, Szarras-Czapnik M, Milne L, Mitchell RT, Smith LB. Autocrine androgen action is essential for Leydig cell maturation and function, and protects against late-onset Leydig cell apoptosis in both mice and men. FASEB journal : official publication of the Federation of American Societies for Experimental Biology 2015; 29:894-910

42. Takata R, Akamatsu S, Kubo M, Takahashi A, Hosono N, Kawaguchi T, Tsunoda T, Inazawa J, Kamatani N, Ogawa O, Fujioka T, Nakamura Y, Nakagawa H. Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. Nat Genet 2010; 42:751-754

43. Bhasin S, Travison TG, Storer TW, Lakshman K, Kaushik M, Mazer NA, Ngyuen AH, Davda MN, Jara H, Aakil A, Anderson S, Knapp PE, Hanka S, Mohammed N, Daou P, Miciek R, Ulloor J, Zhang A, Brooks B, Orwoll K, Hede-Brierley L, Eder R, Elmi A, Bhasin G, Collins L, Singh R, Basaria S. Effect of testosterone supplementation with and without a dual 5alpha-reductase inhibitor on fat-free mass in men with suppressed testosterone production: a randomized controlled trial. Jama 2012; 307:931-939

APPENDIX

Table 2.1. mGluR-1 allosteric binding site residues for negative allosteric modulator (antagonist) and alignment with other receptors (28,29). mGlu1 and CasR residues with '*' shown to be important for binding allosteric modulators by mutagenesis experiments. GPRC6A residues with '*' show reduced binding affinity for allosteric modulators by mutagenesis experiments in mouse-GPRC6A and are conserved in human-GPRC6A.

mGlu1 Numbering	mGlu1	mGlu5	CasR	GPRC6A	Consensus (% similarity)	Conservation (% identical)
648	L	L-635	F-668*	F-650	100 (nonpolar)	50 (L,F)
660	Q	Q-647	R-680*	R-662	50 (polar/charged)	50 (Q,R)
661	R	R-648	Q-681	Q-663	50 (charged/polar)	50 (R,Q)
664	V	I-651	F-684*	F-666*	100 (nonpolar)	50 (F)
668	S*	P-655	F-688*	F-670*	75 (nonpolar)	50 (F)
748	T	T-735	E-767	E-746	50 (polar/charged)	50 (T,E)
753	V	V-740	L-773*	F-752	100 (nonpolar)	50 (V)
756	P*	P-743	L-776*	M-755	100 (nonpolar)	50 (P)
757	L*	L-744	I-777	L-756	100 (nonpolar)	75 (L)
760	N*	N-747	T-780	I-759*	75 (polar)	50 (N)
761	G	G-748	C-781	A-760	75 (nonpolar)	50 (G)
794	T	T-781	F-814	Y-791	75 (polar)	50 (T)
797	I	I-784	V-817	A-794	100 (nonpolar)	50 (I)
798	W*	W-785	W-818*	W-795*	100 (nonpolar)	100 (W)
801	F*	F-788	F-821*	F-798	100 (nonpolar)	100 (F)
805	Y*	Y-792	Y-825	Y-802	100 (polar)	100 (Y)
811	K	K-798	V-833	V-810	50 (charged/nonpolar)	50 (K,V)
812	I	I-799	S-834	P-811	75 (nonpolar)	50 (I)
815	T*	M-802	E-837*	E-814*	50 (charged)	50 (E)
818	A*	S-805	A-840	V-817	75 (nonpolar)	50 (A)
822	S	S-809	A-844	S-821	75 (polar)	75 (S)

Table 2.2. mGluR-5 allosteric binding site residues for negative allosteric modulator (antagonist) and alignment with other receptors (28,29). mGlu5 and CasR residues with '*' shown to be important for binding allosteric modulators by mutagenesis experiments. GPRC6A residues with '*' show reduced binding affinity for allosteric modulators by mutagenesis experiments in mouse-GPRC6A and are conserved in human-GPRC6A.

mGlu5 Numbering	mGlu 5	mGlu 1	CasR	GPRC6 A	Consensus (% similarity)	Conservation (% identical)
624	G	G-637	S-657*	C-639	50 (nonpolar/polar)	50 (G)
625	I	I-638	L-658	H-640	75 (nonpolar)	50 (I)
628	G	G-641	C-661	N-643	50 (nonpolar/polar)	50 (G)
651	I*	V-664	F-684*	F-666*	100 (nonpolar)	50 (F)
652	G	G-665	G-685	G-667	100 (nonpolar)	100 (G)
654	S	S-667	S-687	S-669	100 (polar)	100 (S)
655	P*	S-668	F-688*	F-670*	75 (nonpolar)	50 (F)
658	S*	C-671	C-691	C-673	100 (polar)	75 (C)
659	Y*	Y-672	I-692	I-674	50 (polar/nonpolar)	50 (Y,I)
740	V	V-753	L-773*	F-752	100 (nonpolar)	50 (V)
743	P	P-756	L-776*	M-755	100 (nonpolar)	50 (P)
744	L	L-757	I-777	L-756	100 (nonpolar)	75 (L)
747	N*	N-760	T-780	I-759*	75 (polar)	50 (N)
751	I	I-764	A-784	A-763	100 (nonpolar)	50 (I,A)
781	T	T-794	F-814	Y-791	75 (polar)	50 (T)
784	I	I-797	V-817	A-794	100 (nonpolar)	50 (I)
785	W*	W-798	W-818*	W-795*	100 (nonpolar)	100 (W)
788	F*	F-801	F-821*	F-798	100 (nonpolar)	100 (F)
802	M*	T-815	E-837*	E-814*	50 (charged)	50 (E)
805	S*	A-818	A-840	V-817	75 (nonpolar)	50 (A)
806	V	V-819	I-841*	I-818	100 (nonpolar)	50 (V,I)
809	S*	S-822	A-844	S-821	75 (polar)	75 (S)
810	A*	V-823	S-845	N-822	50 (nonpolar/polar)	25 (A)
813	A	A-826	L-848	I-825	100 (nonpolar)	50 (A)

Table 2.3. PolyPhobius secondary structure prediction for GPRC6A sequence.

Residue Number	Structure Prediction
1-18	SIGNAL PEPTIDE
19-592	NON CYTOPLASMIC
593-616	TRANSMEM
617-630	CYTOPLASMIC
631-651	TRANSMEM
652-663	NON CYTOPLASMIC
664-683	TRANSMEM
684-702	CYTOPLASMIC
703-727	TRANSMEM
728-750	NON CYTOPLASMIC
751-773	TRANSMEM
774-782	CYTOPLASMIC
783-804	TRANSMEM
805-811	NON CYTOPLASMIC
812-833	TRANSMEM
834-926	CYTOPLASMIC

Table 2.4. Secondary structure features of mGluR-1/mGluR-5 models generated in MOE-2012. *

indicates the regions that show deviation from predicted structure for 5 or more residues.

Residue Number	Structure
1-18	SIGNAL PEPTIDE
19-591	NON CYTOPLASMIC
592-617	TRANSMEM
618-633	CYTOPLASMIC
634-651	TRANSMEM
652-657*	NON CYTOPLASMIC
658-686*	TRANSMEM
687-703	CYTOPLASMIC
704-725	TRANSMEM
726-750	NON CYTOPLASMIC
751-772	TRANSMEM
773-778	CYTOPLASMIC
779-804	TRANSMEM
805-811	NON CYTOPLASMIC
812-837	TRANSMEM
838-926	CYTOPLASMIC

Table 2.5. Docking scores and binding pocket residues for testosterone binding in GPRC6A models generated in MOE-2012.

Model	Docking Score	Binding Site Residues
mGluR-1	-11.58	C-659, R-662, Q-663, F-666, G-667, F-670, E-746, A-751, F-752, M-755, L-756, I-759, E-814
mGluR-5	-12.31	F-666, F-670, F-752, M-755, L-756, L-759, Y-791, A-794, W-795, F-798, V-817, S-821

C=Cys; R=Arg; Q=Gln; F=Phe; G=Gly; E-Glu; A=Ala; M=Met; L=Leu; I= Ile; Y=Tyr; W=Trp; V=Val; S=Ser
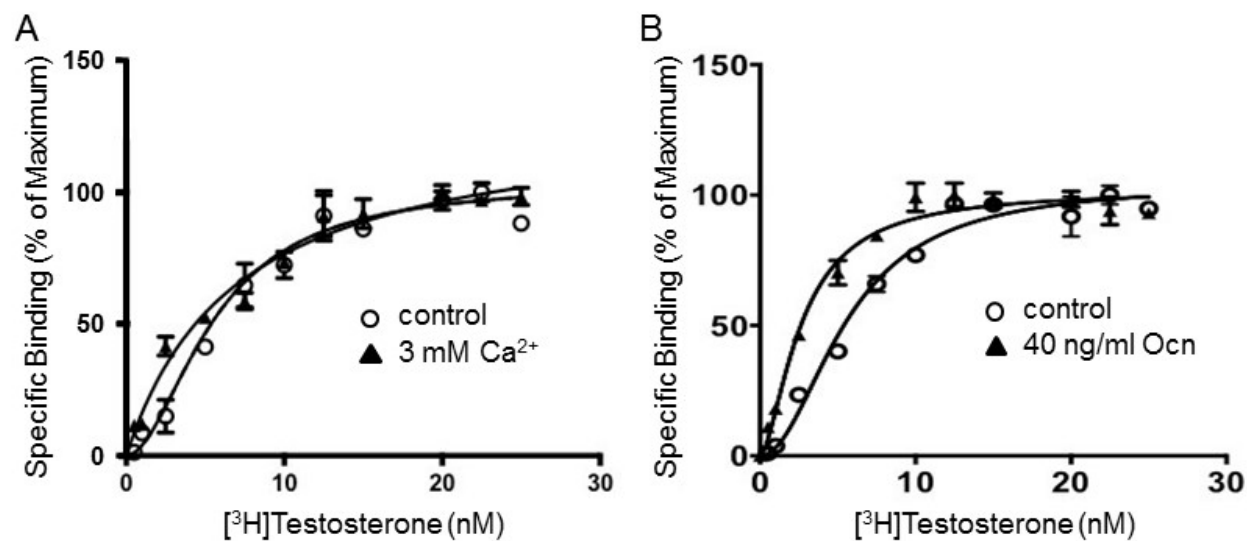
Figure 2.1. Evidence for T-binding and activation of GPRC6A. (A) Saturation analysis of [3H] T binding in HEK-293 cells expressing GPRC6A. The figure shows the saturation isotherm of [3H] T with or without Ca^{2+} (A) or with or without Ocn (B).
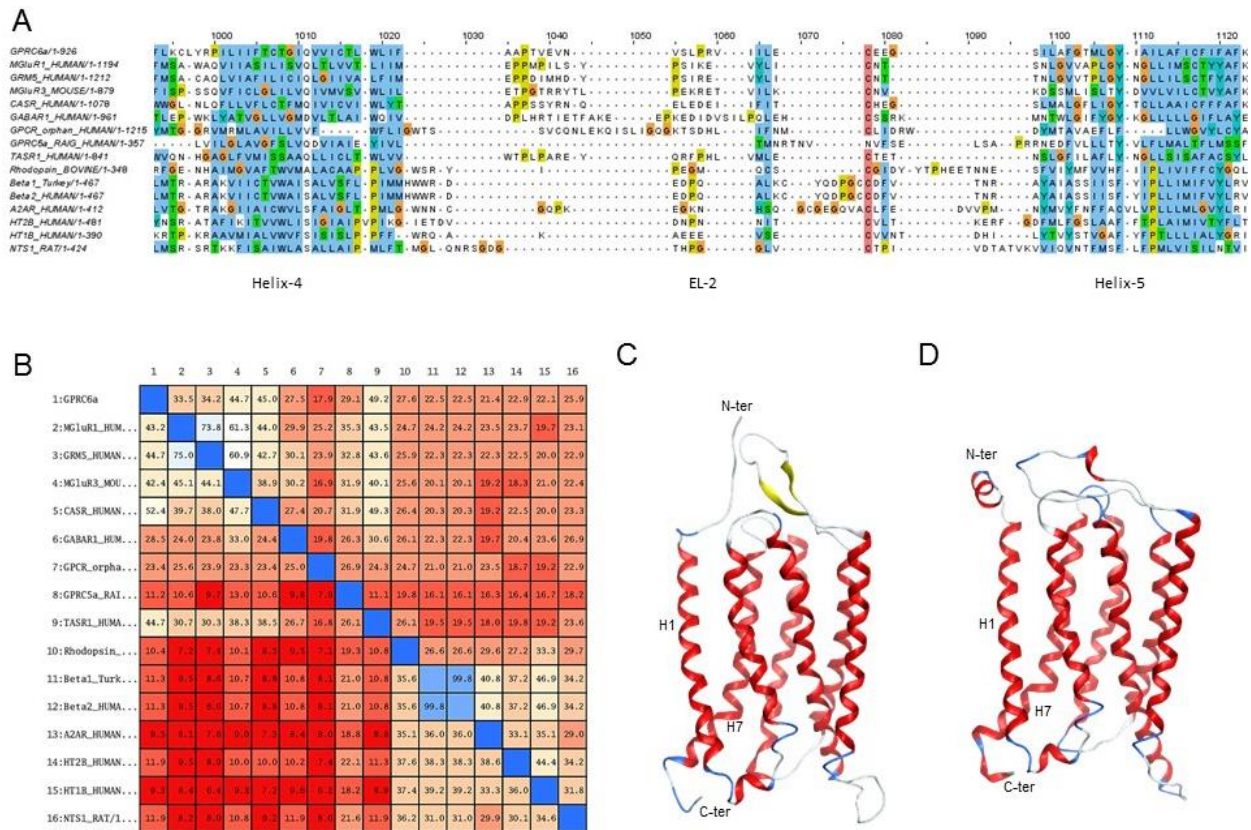
Figure 2.2. GPRC6A Homology Modelling. (A) MSA between 16 sequences shown using clustralx color code. Alignment is shown for helix-4, extracellular loop-2 and helix-5 regions. (B) Sequence similarity scores between 16 sequences after MSA. Sequence 1: GPRC6A, Sequences 2-9: class-C GPCR's, Sequences 10-16: class-A GPCR's. mGluR-1 (sequence-2) and mGluR-5 (sequence-3) taken as main templates for transmembrane domain modelling (C, D) GPRC6A transmembrane homology models based on the mGlu-1 and mGlu-5 receptor structures.
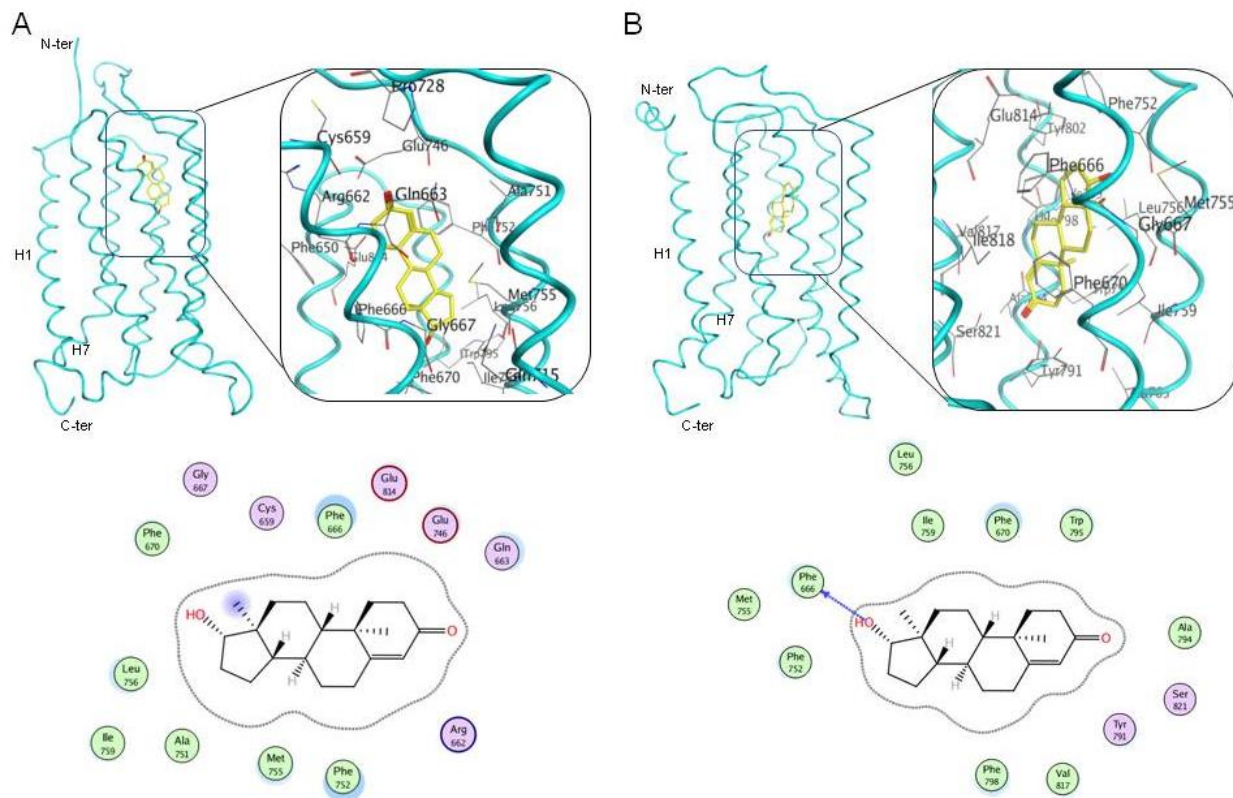
Figure 2.3. Docking of Testosterone to GPRC6A. (A) (top): Allosteric site in the mGluR-1-based template model, T is shown in yellow stick representation. (Bottom): Residues surrounding T in the binding pocket. (B) (top): Allosteric site in the mGlur-5-based model, T is shown in yellow stick representation. (Bottom): Residues surrounding T in the binding pocket.

Figure 2.4. Evidence for T activation of GPRC6A. Dose dependent effects of T on cAMP response (A) and ERK phosphorylation (B) in HEK293 cells expressing GPRC6A. (C) Comparison of cAMP accumulation stimulated by T with other GPRC6A ligands, L-Arginine (Arg) in HEK293 cells transfected with and without GPRC6A. * and ** Significant difference from control group and stimulated group at $p<0.05$ and $p<0.01$ (n≥4). (D) PKA inhibitor H89 blocked GPRC6A-mediated ERK phosphorylation stimulation. (F) Functional effect of mutations affecting the F666A mGPRC6A residue. HEK293 cells were transfected with the mock plasmid or with a plasmid encoding the WT or F666A mGPRC6A mutants. Analysis of mGPRC6A expression by Western blot was shown in right panel. Immunoblot analysis of whole cell lysates (5 µg proteins) from HEK293 cells transiently transfected with an empty vector (MOCK), or a vector containing the WT or the indicated mGPRC6A mutant, was performed following 3~8% SDS polyacrylamide gel electrophoresis. Mouse GPRC6A (mGPRC6A) proteins were detected using the anti-Myc antibody (Cell signaling). The position of the molecular mass markers is shown on the left (kDa). Arrow heads on the right indicate the molecular weight of two major bands corresponding to mGPRC6A. Black arrows on the right indicate major bands corresponding to the WT and mutant receptors, and white arrow head indicates the dimer of the receptors, respectively.
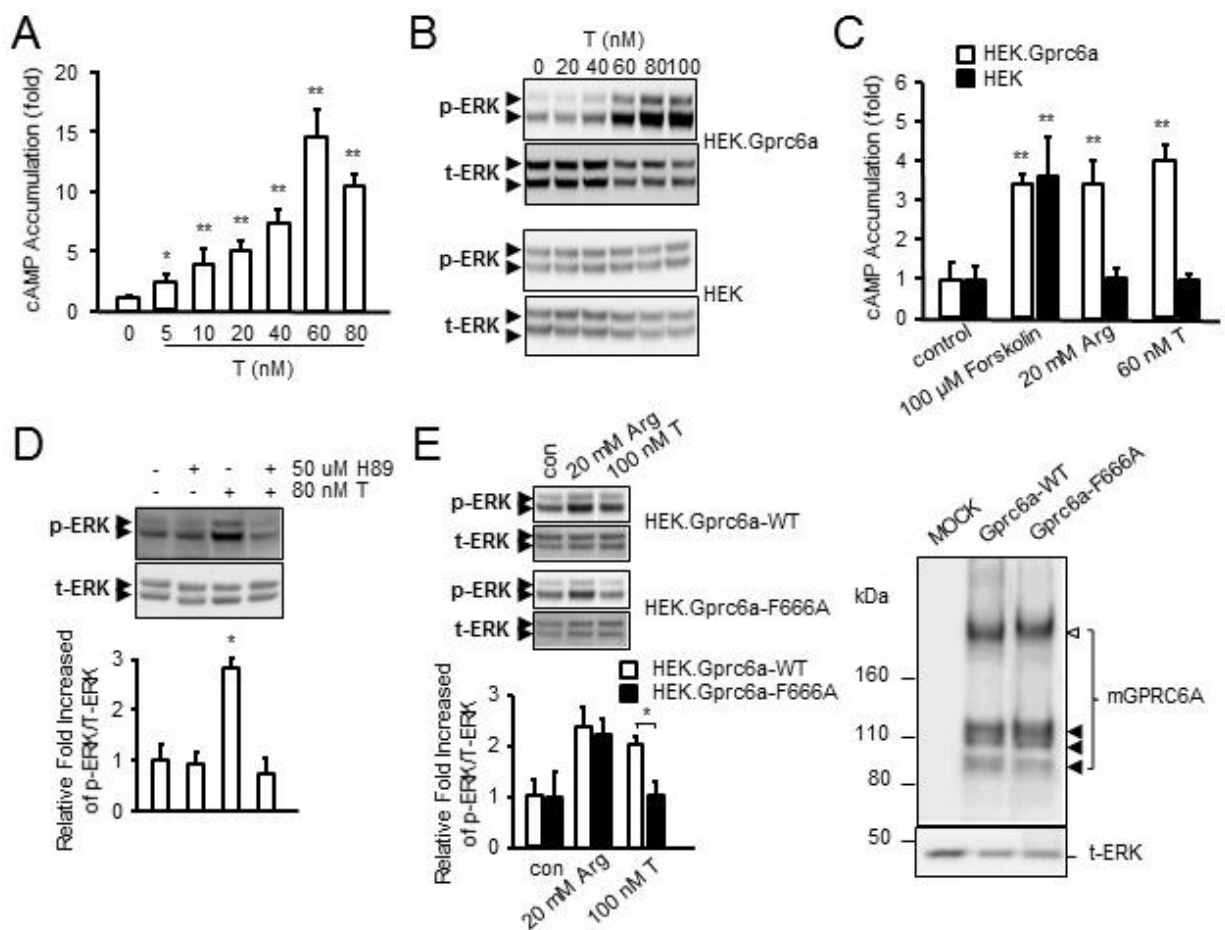
Figure 2.5. β-cell and Leydig cells dysfunction in *Gprc6a⁻/⁻* mice. (A) Effects of systemic administration of T on *Egr-1* expression in pancreas and testis. T (200 mg/kg) or vehicle was injected into mice intraperitoneally, *Egr-1* mRNA abundance in various mouse tissue 1 hour after injection was determined by real time PCR. Values represent the mean ± SEM. * significant difference from control group and stimulated group at $p<0.05$ (n ≥ 3). (B) Effects of T on insulin stimulation index in islets isolated from wild-type and *Gprc6a⁻/⁻* mice. Values represent the mean ± SEM. * significant difference from control group and stimulated group at $p<0.05$ (n ≥ 3). # indicates significant different between T treated control group and *Gprc6a⁻/⁻* mice ($P< 0.05$; n ≥ 3). (C) RT-PCR analysis showing that *Gprc6a* message is expressed in INS-1 cells. A similar band was observed in rat kidney and pancreas which are known to express GPRC6A. The primers sequence are rGPRC6A.For535: aaaatccgctttccttcgttr; and GPRC6A.Rev1400: tgggcatcaaaatgaaatgar. (D) Effects of T to stimulate insulin secretion in rat β-cells INS-1. Values represent the mean ± SEM.* and ** significant difference from control group and stimulated group at $p<0.05$ and $p<0.01$ (n=4). (E) Effects of T on message expression of enzymes regulating T biosynthesis in isolated Leydig cells. Values represent the mean ± SEM. * significant difference from control group and stimulated group at $p<0.05$ (n ≥ 3).
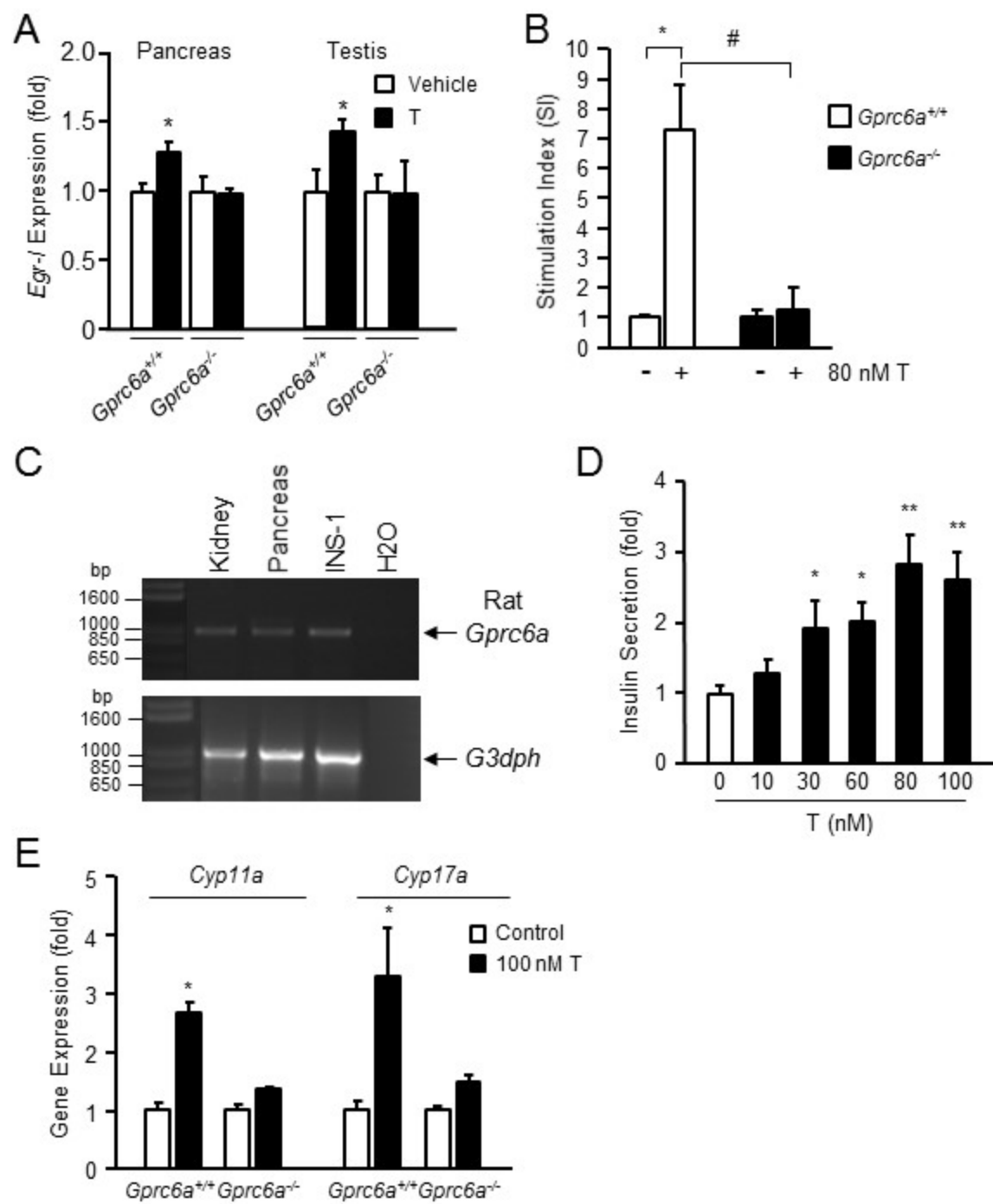
Chapter 3

NOVEL ALLOSTERIC INHIBITORS OF THE

PROTHROMBINASE ENZYME COMPLEX

Research described in this chapter is reproduced from an article submitted to the Blood Journal by Karan Kapoor, Nicole McGill, Cynthia B. Peterson, Harold V. Meyers, Michael N. Blackburn and Jerome Baudry.

My contribution in the paper includes (1) Compiling and interpretation of the literature (2) Designing and carrying out all computation research (3) Understanding the literature and interpretation of the results (4) Preparation of tables and figures (5) Writing and editing.

Abstract

The risk of serious bleeding, particularly at high dosage, is a major liability of anticoagulant drugs that are active-site competitive inhibitors targeting the Factor Xa (FXa) prothrombin (PT) binding site. The present work identifies a new class of anticoagulants that can act as allosteric inhibitors of the prothrombinase (PTase) complex composed of FXa and Factor Va (FVa). This new class of anticoagulants was identified though an integrated computational/experimental approach. A 300 ns Molecular Dynamic (MD) simulation of the FXa heavy chain was used to identify conformational snapshots representing a structural ensemble sampled by the protein. These snapshots were used in ensemble docking calculations of a 281,128 compound drug-like subset of the ZINC database against FXa, targeting potential protein/protein interaction sites in the FXa-FVa binding interface. From about three million docking calculations, five hundred and thirty-five compounds and their structural analogs were selected for experimental validation through a series of binding and inhibition assays, based on their predicted binding affinities to FXa. From these, ten compounds represented by three families of inhibitors were experimentally identified that achieve dose-independent partial inhibition of PTase activity in a non-active site dependent and self-limiting mechanism.

Introduction

The blood coagulation system represents the first line of defense against blood loss following injury. This system consists of a cascade of circulating inactive serine protease zymogens as well as regulatory cofactors and inhibitors. Each active enzyme, once generated from its zymogen, specifically cleaves the next zymogen in the cascade pathway to produce an active protease, a process that is repeated until finally thrombin cleaves the fibrinopeptides from fibrinogen to produce fibrin which polymerizes to make a blood clot. The amplification provided by this system leads to the potential for explosive clot formation.(1) Although efficient clotting limits the loss of blood at a site of trauma, it also poses the risk of systemic coagulation resulting in massive thrombosis. This can result in myocardial infarction, unstable angina, atrial fibrillation, stroke, pulmonary embolism and deep vein thrombosis.

Several new oral Factor Xa and thrombin inhibitors have received approval for clinical use in anticoagulant therapy. These new oral anticoagulants or NOACs include Xarelto, Eliquis and Savaysa,(2) which are FXa active site inhibitors, and Pradaxa(3) which is a thrombin inhibitor. These drugs are a significant improvement over Coumadin but as use of these drugs expands, bleeding remains a concern,(4) particularly in the treatment of arterial thrombosis.

Each of these drugs binds in the active site of a coagulation protease and is a competitive inhibitor of its target enzyme. Following classical Michaelis-Menten kinetics, as competitive inhibitor concentration increases, enzyme activity decreases until at saturating concentrations, activity approaches zero. Although prescribed drug levels are designed to achieve a safe and effective concentration, increased drug plasma levels can result from impaired renal clearance, drug-drug interactions, improper time between dosing and accidental over-dosing. These increases in drug plasma concentrations give rise to a potential bleeding diathesis. Alternative approaches

to anticoagulant therapy with truly novel starting points are needed to circumvent the bleeding liability of these active-site targeted agents.

The prothrombinase (PTase) enzymatic complex, consisting of the enzyme factor Xa (FXa), a protein cofactor, factor Va (FVa), associated on the surface of negatively charged phospholipid membranes in the presence of divalent metal ions,(5,6) forms a critical junction of the blood coagulation cascade pathways.(7) This complex catalyzes the cleavage of prothrombin (PT) leading to the formation of thrombin and subsequent clot-formation.(8) FXa alone can slowly activate PT, but the rate of thrombin formation is enhanced by >105 by the presence of the cofactor, FVa and Ca2+ on phospholipid membranes in the PTase complex.(9,10)

FXa consists of two chains– the light chain of 139 residues, and the heavy chain of 305 residues that contains the catalytic active site and is connected via a disulfide (Cys132-Cys302) bridge to the light chain.(11) A wealth of structural data is available for this protein; over 150 crystal structures of the protein or protein-ligand complexes have been deposited in the Brookhaven Protein Data Base (PDB).(12) Activated human FVa consists of a non-covalent calcium-dependent complex between the heavy chain (Hc: A1 (1–303)-A2 (317–656) domains) and the light chain (Lc: A3 (1546–1877)-C1 (1878–2036)-C2 (2037–2196) domains) in which the cofactor binding site is exposed for interaction with FXa.(13,14) Several studies have shown that both the Hc and Lc of FVa are involved in the binding of FXa, and also give some insights about FVa residues interacting with FXa.(15-19) In addition, some regions in the Hc of FXa have been suggested to be important for the interaction with FVa.(20-26) High-resolution crystal structures of FXa are available,(27) and homology models of FVa using the crystal structure of inactivated bovine FVa (FVai)(28) have also been built. Using these two structures, models of the FVa–FXa complex have been generated that are in agreement with available experimental data.(29)

The goal of this study is to target FXa within the PTase complex, but instead of seeking another active-site directed inhibitor, high-throughput virtual screening methods are used to identify compounds that are likely to bind to FXa and alter the interaction between FXa and FVa, and hence have the potential to down-modulate the PTase activity in an allosteric and non-competitive manner.

Materials and Methods

FXa-FVa binding site

The active site-inhibited human FXa structure, solved by X-ray crystallography (PDB code: 1XKA), was used as the starting structure.(27) It has a 2.3 Å resolution and no gaps in the structure. MD simulations in previous studies showed that the light chain, which is connected to a negatively charged phospholipid membrane in the PTase complex, and does not form any interactions with FVa, is highly flexible in aqueous solution and undergoes angstroms-wide fluctuations within a few nanoseconds of the MD.(30) To focus on the serine protease domain that interacts with FVa, this flexible part was removed and only the globular heavy chain was used in the MD calculations. Homology models of FVa using the crystal structure of inactivated bovine FVa (FVai) have been built by other groups.(28) Using these models and the crystal structure of FXa, models of the FVa–FXa complex have been generated that are in agreement with available experimental data.(29) These models and the available experimental data were used to identify the regions of FXa that have been suggested to be important for the interaction with FVa.

Figure 1A shows the FXa binding surface that is proposed to interact with FVa, based on the FVa–FXa complex models, consisting of twenty-eight residues. They form part of 3 helices-H1, H2 and H3 and one loop, L1, shown in Figure 1B.

Recently, the crystal structure of the FX:FV complex from the venom of Pseudonaja textilis has been solved.(31) The complex shows specific interactions between the structured A2 and A3 domains of FV and FX. The highly flexible terminal region following the A2 domain forms non-directional/non-specific ionic interactions and lacks a fixed binding mode, and only a small hydrophobic portion at the end of this region (IFADIFI) was well resolved.

The FXa binding interface selected in the current study can be compared to the crystal structure of the FX:FV complex from P. textilis. Human FX and FV sequences show an identity of 50.5% and 53.5 % with the P. textilis sequences, respectively. Figures 1C shows the superposition of the human FXa crystal structure with the FX-FV complex from P. textilis (PDB code: 4BXS). The selected human FXa:FVa binding site consisting of 3 helices and 1 loop also forms the interface between FX:FV complex from P. textilis (Figure 1C). The FXa/FVa contact region mapped on human FXa shown on Figure 1A is essentially reproducing the P. textitlis's protein:protein interface.

MD Simulations and Clustering

The dynamics (MD) of the FXa heavy chain was characterized using molecular dynamics (MD) simulations. The active site ligand and the light chain were removed. One calcium ion known to be structurally important(32,33) was kept in the structure. The structure was protonated according to estimated pKa calculations using the Protonate-3D facility in MOE, version 2011,(34) with an ionic concentration of 0.1 mol/L in the Generalized Born (GB) electrostatics model. A periodic solvation cube of water molecules was created using the Solvate facility in MOE-2011. Seven negative chloride ions were added to neutralize the system, and the final system consists of 22,516 atoms, with 6279 water molecules. The system was energy minimized to a gradient of 10-

5 RMS kcal/mol/Å2 using the CHARMM-22 force field(35) and a 8Å-10Å nonbonded cutoff. NAMD2 input files were generated using the Dynamics facility in MOE-2011. MD simulations using the NAMD2 simulation engine(36) were performed using an integration time-step of 2 fs/step, holding covalent bonds in water molecules rigid. A constant temperature of 300K was maintained throughout the simulation using Langevin dynamics. Particle Mesh Eswald (PME) was used for calculation of the electrostatic interactions with the periodic boundary conditions. Simulations were performed in the NVT ensemble. The MD simulation was run on the University of Tennessee High Performance Computing architecture (Newton). A 100 ps initial equilibration was performed followed by a 300 ns production trajectory. The resulting trajectories were analyzed for convergence of temperature and RMSD using perl scripts for parsing data, Prody v1.2(37) and Gnuplot v4.6.

Structures of the protein along the MD trajectory were generated every 50ps using Prody v1.2, i.e., generating a total of 6,000 structures. Nearest-neighbor RMSD clustering was used to cluster these structures with Maxcluster(38) using RMSD distance between the data point represented by each structure. Two structures were considered part of the same cluster if they were closer to each other than the cut-off threshold distance, i.e. if they were nearest-neighbors. The entire ensemble docking process flowchart is shown in Figure 2.

Ligand and Receptor Preparation for Docking

Virtual screening was performed using the program Autodock Vina(39) on 12 structures: 11 structures obtained from the MD clustering described above, and the FXa crystal structure. The 'clean drug-like' subset of the ZINC database(40) was used, consisting of ~11.1 million compounds without reactive functional groups, and that do not violate any of the Lipinski's criteria. The subset

132

of compounds used for virtual screening contained 244,493 compounds (referred as "Subset A" from here on) and covered 90% of the 'clean drug-like' compounds of ZINC database's chemical diversity based on Tanimoto's similarity coefficient.(41) These compounds were converted from the smile format to the PDB format using Babel v2.3.3(42) with the generate-3D option. The geometries of the ligands were optimized using the MMFF94 force field(43) in Babel. The compounds' coordinates in PDB format were converted to PDBQT format using the Prepare Ligand scripts provided in ADT/MGL tools v1.5.4.(44) The FXa structures in PDB format from the clustered MD trajectory were also converted to PDBQT format directly in ADT tools. The configuration files for each FXa structure, including the docking grid box dimensions and the box center, used as input for docking in Autodock Vina, were created using ADT tools.

The Subset A compounds were docked in the crystal structure and in the first 11 representative structures from MD trajectory clustering, as indicated in Figure 2, allowing additional conformations of the ligand by sampling the rotatable bonds, and keeping the protein's sidechains rigid. The docking was performed "agnostically," i.e., binding could happen anywhere on the protein without imposing a particular binding site using Autodock Vina and an exhaustiveness value of 100 on the Newton cluster.

Docking Analysis and Secondary Compound Selection

Perl scripts were used to select the best-scoring binding mode reported in the AutoDock Vina output for each ligand and to parse these binding modes in order to identify ligands that bind in the regions of FXa involved in FVa interactions (steps III in Figure 2).

The top 80 ranked compounds in the crystal structure and top 40 ranked compounds in each of the other 11 MD snapshots (representing a total of 520 compounds) were identified from

133

the docking calculations. ZincPharmer,(45) a pharmacophore search tool for screening the Zinc database, was used to identify an additional 42,309 compounds in the ZINC database that are chemically similar to the docking hits but with larger molecular weights (step IV in Figure 2). These compounds were parsed to keep only the drug-like compounds with properties as defined in the Zinc database, i.e., $150 \leq$ mol wt ≤ 500, xlogp ≤ 5, $-5 \leq$ net charge ≤ 5, rotatable bonds ≤ 8, polar surface area < 150 Å2, no. of H donors ≤ 5, no. of H acceptors ≤ 10, $-400 \leq$ polar desolvation energy ≤ 1 kcal/mol, $-100 \leq$ apolar desolvation energy ≤ 40 kcal/mol, and also to remove any duplicates, giving a total of 36,635 compounds (referred as 'Subset B' from here on), corresponding to step V in Figure 2. These compounds were again docked in the 12 structures using Autodock Vina on the UTK's Newton High Performance cluster with the same configuration settings as described above. Perl scripts were again used to parse the docking results for compounds that bind in the regions of FXa involved in FVa interactions.

Analysis of Compound Lists

The results from docking of Subset A (step II in Figure 2) and Subset B (step VI in Figure 2), indicated that compounds can bind in different sub-binding sites at the FXa-FVa interface, in different protein MD snapshots. Based on the clustering of the binding loci for these compounds, the FXa-FVa interface was divided into ten sub-binding sites.

The computational hits were analyzed using two main criteria (in addition to their predicted binding energies): the number of MD snapshots to which they were predicted to bind, and the number of sites in the protein where a compound is predicted to bind. This allows the identification of compounds that bind to a large number of target structural variations and binding site(s). A value named "snapshot count" was calculated using Perl scripts that gives the total number of

134

snapshots in which a given compound binds to the FXa-FVa interface. Compounds found to bind to most snapshots were checked for the different sub-binding sites (on FXa-FVa interface), yielding another value named "binding-site score" that gives the number of sub-binding sites that a given compound is docked.

Two lists of compounds were generated for experimental validations: first, a list that contained 288 compounds consisting of the strongest binders in the crystal structure and the 11 snapshots, as predicted by Vina docking scores (shown in step VIIIa of Figure 2). Second, a list was generated that contains 247 compounds with "snapshot count" scores of three and above (i.e., including compounds that successfully bind to three or more different protein conformations) and "binding-site scores" of one or two (i.e. removing promiscuous compounds that bind in many sites in different snapshots), shown in step VIIIb of Figure 2. The first list thus contains compounds that are predicted to bind with the highest binding affinities at the FXa-FVa interface, and the second list contains the compounds that can bind to different protein conformations but at the same time are selective for particular binding sites at the interface.

Visual inspection was performed on the compounds in the two lists to insure their commercial availability and affordable cost, and to identify and remove compounds with reactive functional groups (e.g. aldehydes and alkyl bromides), excessive hydrophobicity (LogP >5), a lack of functionalization (e.g. polyaromatics with few or no functional groups) or an excess of functionalization, the latter two of which would render synthetic analoging difficult for any follow up structure-activity relationship (SAR) studies. The compounds ordered for experimental testing were selected to sample different sub-binding sites at the interface as predicted by the docking, be of affordable cost and presenting desirable chemical features.

Experimental Validation

Initial screening of these compounds for potential FXa inhibitors was performed using a prothrombinase assay as described by Nesheim et al.(9) FXa, FVa, prothrombin and thrombin were purchased from Hematologic Technologies. The thrombin and FXa substrates S-2238 and S-2222, respectively, were obtained from diaPharma. Apixaban was purchased from Selleckchem. PTase activity measurements were initiated by the addition of prothrombin to the reaction mixture, and the rate of thrombin generation was determined by measuring thrombin activity towards the chromogenic substrate S-2238. Replicate samples were analyzed and the rates were corrected for the very low rate of hydrolysis of S-2238 by FXa. FXa activity was measured using the substrate S-2222. Assay data was analyzed using GraphPad Prism 5.0f to obtain kinetic parameters.

Structural analogs of the identified hits were searched from the list of all the compounds used for docking, including Subsets A and B, using Tanimoto coefficient and the MACCS structural keys fingerprints(46) in MOE-2011. Similar compounds not present in this list were also identified directly from the ZINC database using substructure searching. Further filters described above were also applied on these compounds to order a second set of compounds. All compounds were dissolved in anhydrous spectroscopic grade DMSO at 10 mM and were stored in sealed vials at -20°C. Compounds were diluted in the assay to final concentrations of 5 to 100 μM. The final DMSO concentration in the assay was kept constant at 1%.

Binding Sites Analysis

The initial FXa inhibitors identified by PTase activity assays (and given in results) were re-docked in the binding sites identified with Vina using MOE-2011 with the CHARMM27 Force-field, allowing protein sidechain flexibility. Additional conformations of the ligand were allowed

by sampling the rotatable bonds. Binding site residues side-chains were allowed to move away from their original location using a tethering weight of 0.1. The GBVI/WSA dG scoring function(47) was used for scoring different poses of the docked ligand. The top ranked binding poses were energy-minimized and the compounds' predicted binding free energies were calculated using a MM/GBVI-adapted protocol.(48)

Results

MD Simulations and Clustering

The FXa heavy chain structure including the FXa-FVa binding interface residues were found to sample different conformations in the MD snapshots as shown in Figure 3A. Figure 3B shows the RMSD plot for the protein backbone over the time course of the trajectory. Generally, MD simulations are considered stable when the backbone RMSD is in the low angstrom area.(30) However, in the case of FXa, higher RMSD values ranging between 2 and 7 Å have been reported.(49-51) Based on the RMSD plot in Figure 3B, the FXa structure converges after the first 100 ns of the trajectory.

Nearest-neighbor RMSD clustering was used to divide the trajectory into 15 clusters as described in Methods. Figure 3C shows the superposition of these structures, also represented on the RMSD plot (Figure 3B) over the time course of the trajectory.

Docking in AutoDock Vina

The docking of compound Subset A in 12 FXa structures (crystal structure and 11 representative structures from 15 clusters) represented a total of ~2.9 million docking calculations (step II in Figure 2). After selecting the best scoring poses of docked ligands as predicted by

AutoDock Vina and parsing these for the selected FXa-FVa binding sites, it was found that in the crystal structure most of the compounds (99.8%) bind to sites outside of the FXa-FVa interface, mostly in the protein active site, and only around 0.2% of compounds screened (532 compounds in the case of crystal structure) bind to FXa within the FXa-FVa interface. As the MD snapshots sample larger conformational space, more compounds bind at this interface compared to the number that bind to the crystal structure. Docking of Subset B compounds in the 12 structures represented another ~0.45 million docking calculations (step VI in Figure 2). In the case of crystal structure, only 177 of these compounds (0.5%) bind to FXa within the FXa-FVa interface, and the higher hit rate compared to Subset A could be expected due to the larger MW's of Subset B compounds. Figure 4A shows the comparison between a compound from Subset A and one from Subset B binding in the FXa-FVa interface in the crystal structure (both compounds were later identified as inhibitors). As one would expect, larger compounds are found to bind to larger regions on the FXa surface, and also often (although not in all cases) exhibit better binding (docking) scores than smaller compounds.

Docking Analysis and Compound Lists

In the crystal structure the compounds that bind in the FXa-FVa interface were found to bind essentially in four regions represented by four grooves, shown in Figure 4B. Taking the docking results from all snapshots together, the interface was divided into ten sub-binding sites, as shown in Figure 4C and Table 1, with sub-binding sites 1, 4, 5, 6 and 9 corresponding to groove-A, groove-B, groove-C, groove-D and groove-E in the crystal structure, respectively. Other binding sites were not observed in the crystal structure but only in MD snapshots. The ten binding subsites are thus defined by the localization of compounds observed in the docking studies as

138

opposed to the topological features of the protein surface as observed in the crystal structure. It was found that compounds with top snapshot count scores were concentrated in sub-binding sites 9 (52%), 7 (27%) and 10 (4%), respectively. Compounds with low binding-site scores of 1 and 2 mostly bind to binding site 9 as well. This site is relatively large and present at the center of the binding interface.

The strongest binders as predicted by Vina docking scores were also concentrated in sub-binding site 9 (43%), 7 (38%) and 10 (22%). These compounds are weighted heavily from Subset B (~91%), compounds that arose from the pharmacophore based search around the best-ranked compounds from Subset A, and represent an overall higher molecular weight.

Experimental Validation

After employing further filters to the compounds in the two lists based on compounds' availability, cost, chemical similarity, and presence of reactive functional groups, compounds were ordered for experimental validation, shown in step IX of Figure 2. The majority of these selected compounds were predicted to bind in sites 9 and 7, with some representation of the remaining binding sites. Compounds were screened for solubility in the assay buffer (about half were soluble at 100 μM) and for their effect on PTase and FXa enzyme activity. Initial screening of these compounds at 100 μM concentration was performed in a PTase activity assay measuring inhibition of the rate of generation of active thrombin using the thrombin specific chromogenic substrate, S-2238. Further screening of these compounds with inhibitory activity was performed as a function of inhibitor concentration from about 5 μM to 100 μM. PTase activity assays identified 16 compounds as inhibitors for the PTase activity, out of the initial compounds suggested from the in silico calculations. Eight of these 16 compounds were not computationally predicted to bind well

139

in the crystal structure but were predicted to bind well in one or more of the MD snapshots. These inhibitors would not have been identified if the docking calculations had been made using only the crystal structure. Ten more inhibitors were identified from a screen of structural analogs (step Xa and Xb in Figure 2) of those compounds that were obtained for initial screening. A total of 26 inhibitors were identified.

These compounds inhibit PTase activation of thrombin from prothrombin in the presence of FVa, but at a concentration of 100 μM, these compounds do not inhibit the activity of FXa towards the chromogenic substrate S-2222 indicating that they do not bind to the active site of the FXa molecule. Similarly, the compounds do not inhibit thrombin activity towards its chromogenic substrate, S-2238. The compounds thus appear to be specific inhibitors of the PTase complex, and inhibition is dependent upon the interaction between FXa and FVa.

When these 26 compounds were assayed as a function of inhibitor concentration, 10 compounds exhibited incomplete or partial inhibition of PTase activity even at high concentrations of inhibitor. Figure 5A, which shows percent inhibition as a function of [I]/Ki, contrasts the effects of five of these partial inhibitors on PTase activity with the anticoagulant apixaban, which is an active site competitive inhibitor of FXa. The characteristic feature of these novel inhibitors is that they achieve or approach plateau levels of inhibition as the concentration of inhibitor is increased. Maximal inhibition, calculated from the inhibition curves, is 100% for apixaban and 25%, 40%, 80% and 90% for SBC-160,029, SBC-160,119, SBC-160,042 and SBC-160,109/SBC-160,012, respectively.

SBC-160,029 was identified by the computational screen as a member of a substituted tetrazole series and binds to FXa in four snapshots, including the crystal structure. SBC-160,029 docked exclusively in Site 9. SBC-160,029 inhibits PTase activity to about 25% with a Ki of 15

µM. Two analogs (Table 2), SBC-160,124 and SBC-160,125, exhibit similar inhibitory activity and comparable Ki values.

Compounds SBC-160,012 and SBC-160,042 were also identified as site 9 binders; both compounds also dock into site 7 in other MD snapshots. SBC-160,012 docks in site 9 in the crystal structure and two other snapshots, but docks to site 7 in snapshot 10. SBC-160,042 does not dock in the crystal structure, but docks in site 9 in one snapshot and in site 7 in another (Table 3). With the exception of SBC-160,042 (Ki = 30 uM) these inhibitors have similar affinity for FXa with Ki values of 10 ± 5 µM, but the maximal inhibition for each compound varies from about 25% (029) to about 95% (012). Comparing the series of site 9 inhibitors indicated that maximal inhibition increased with increasing size of the inhibitor. Figure 5B shows the change in maximum PTase inhibition with the molecular weights of the compounds. It can be seen that larger compounds inhibit PTase activity to a greater extent. SBC-160,109, which is a tight binding analog of SBC-160,042 with higher molecular weight, shows greater inhibition, contrasted with SBC-160,119, which has lower molecular weight and lower inhibition than SBC-160,042. Interestingly, SBC-160,012, which has a relatively small molecular weight, is a clear outlier from this trend suggesting a complex inhibition mechanism.

These hits comprise three families of inhibitors as represented by SBC-160,029, SBC-160,012 and SBC-160,042, respectively. SBC-160,099, SBC-160,109, SBC-160,112 and SBC-160,119 were selected from a panel of structural analogs of SBC-160,042.

In summary, from around 3.4 million docking calculations, the computational results yielded 535 compounds that could potentially bind at the FXa-FVa interaction region. Twenty six hits were identified, with 10 of them represented by three families of inhibitors exhibiting partial inhibition of PTase activity in which the extent of inhibition reaches a maximal value so that further

increases in inhibitor concentration do not further inhibit PTase enzyme activity. We refer to this as self-limiting inhibition. Four of these self-limiting inhibitors originated directly from the docking calculations.

Sub-binding Sites Analysis

Table 3 shows three families of inhibitors, represented by parent compounds SBC-160,012, SBC-160,042 and SBC-160,029. These compounds were identified as non-active site or allosteric inhibitors by PTase and FXa activity assays. Compound SBC-160,029 binds in only site 9, whereas SBC-160,012 and SBC-160,042 are able to potentially bind in either site 9 or site 7 in separate snapshots. Table 4 lists the difference in predicted binding energies calculated using the MM/GBVI protocol in MOE for the two compounds binding in site 7 and in site 9 in snapshot 10. The results indicate that compound SBC-160,012 is predicted to bind more strongly in binding site 7 than in binding site 9. Compound SBC-160,042, however, is predicted to bind more strongly in binding site 9.

Table 5 gives the cluster population from which the representative snapshots (centroid structures) were selected. Among the snapshots to which these compounds bind, site 9 was identified in snapshots 2 and 4, and these show relatively high cluster populations. Binding site 7, however, was identified only in snapshot 10 that represents a cluster of structures that was found only 0.5% of the time in the molecular dynamics trajectory; i.e. snapshot 10 represents structures rarely sampled. This makes binding in site 7 for compound SBC-160,012 a rare but possible binding pose, which is nonetheless quite significant in terms of the strength of the interaction ($\Delta\Delta G$ -3.72).

Site 9 was observed in the crystal structure, but its shape and structure change significantly during the MD simulation. Figure 6A shows a compound identified initially as an inhibitor that binds to site 9 only in the later snapshots and not in the crystal structure. Sub-binding site 10 was not seen in the crystal structure and became apparent in 3 of the 11 snapshots from the MD trajectory. Figure 6B shows one of the compounds initially found as a non-active site binder docked in snapshot 7, and how it is not possible for it to bind to the crystal structure.

Discussion

Results from clustering the MD trajectory indicate that the regions of FXa that interact with FVa sample a relatively large accessible conformational space, as seen in Figures 3A and C, and that ligands are predicted to interact with FXa in a way that would not necessarily be observed from an analysis that focused solely on the starting crystal structure. These structural variations shown in the MD translated into a diverse set of compounds that potentially bind to FXa's conformations.

Larger compounds are found to bind to larger sites on the FXa surface, with binding (docking) scores that predict tighter binding compared to smaller compounds in the Subset A, which will likely prove advantageous in modulating the FXa-FVa protein-protein interaction. Figure 5B supports the hypothesis that larger compounds are better able to alter the FXa-FVa interaction. Compound 12, the outlier in Figure 5B, which is predicted to bind preferentially to binding site 7 according to free-energy calculations, may follow a more complex inhibition mechanism.

Compounds with low binding site scores of 1 and 2 mostly bind to sub-binding site 9, a relatively large area at the center of the FVa binding interface. This site is particularly important

for inhibiting the FXa-FVa interaction, as compounds that bind at this site interact with residues from all three helices of interest (H1, H2 and H3). This site was observed in the crystal structure, but its shape and structure change significantly during the MD simulation, as shown Figure 6A. Eleven of the 26 non-active site inhibitors identified by PTase activity assays (including 10 structural analogs which were not docked to generate binding site data) are predicted to bind to site 9. This includes 3 out of the 10 compounds (comprising 6 of the 10 analogs) that achieve self-limiting inhibition of PTase activity.

Among the 10 self-limiting inhibitors, only two compounds were observed to bind to the crystal structure. Sub-binding site 10 was not seen in the crystal structure and became apparent in 3 of the 11 snapshots from the MD trajectory, where two loops that are close to each other in the crystal structure open up, as shown in Figure 6B. SBC-160,064 is a non-active site-inhibitor, inhibiting PTase activity by about 20% at 100 μM, and is predicted to bind in this site but would not have been identified from the crystal structure searches alone.

Tables 3 and 5 indicate that, while compounds appear to bind in site 9 (snapshots 2, 4 5 and 9) an order of magnitude more often than in site 7 (snapshot 10), the predicted binding affinity of compound SBC-160,012 is stronger in site 7 by 3.72 kcal/mol. Hence, this compound's predicted binding site is difficult to assess: site 9 is more often accessible but binding in site 7 is more stable. It is theoretically possible that the compound may bind at both sites and possibly simultaneously. Inversely, binding of compound SBC-160,042 is predicted to happen most likely in binding site 9 since the predicted binding free energy of this compound is better in site 9 than in site 7 by 2.38 kcal/mol.

The comparison of the human structure and the crystal structure of the prothrombinase complex from the venom of P. textilis(31) suggests that the binding sites investigated here are

relevant. Figures 7A and B show the superposition of human FXa (crystal structure) bound with a site-9 inhibitor in its predicted binding mode, with the FX-FV complex from P. textilis. Site-9 inhibitors discovered from the computational approach bind directly at the interface of FXa and the A2/A3 domains of FV, possibly interfering with the specific interactions between these proteins. Figures 7C and D show the superposition of human FXa (snapshot 10 from MD simulation) containing an inhibitor predicted to bind in site-7, with the P. textilis FX-FV complex. In this snapshot the flexible loop L1 folds to form site-7, which is not seen in the crystal structure of human or P. textilis FXa. As seen in this MD snapshot, binding in site-7 can interfere with the interactions between human FXa and the small hydrophobic portion of the terminal region in FV.

Compounds binding to other sites on the human FXa-FVa interface did not inhibit the prothrombinase activity possibly because of the non-specific nature of ionic interactions between FX and the flexible terminal region following the A2 domain of FV in the P. textilis crystal structure. This flexible region of FV may sample alternate binding modes that can overcome the interference of the inhibitors binding to FXa at these sites.

The initial aim of this study to identify compounds that reduce the PTase activity has been met as shown in Figure 5A. In future virtual screening studies, additional compounds can be identified for experimental validation. In addition, lead optimization can be aided by the chemical structures and binding poses of positive hits predicted to bind at both site 7 and site 9. Library design can be based on compound modifications guided by pharmacophore features associated with active compounds, while at the same time avoiding those features associated with inactive compounds (e.g., volume exclusion). It should also be possible to differentiate between binding in site 7 and site 9 from the experimental testing of compounds originating from the pharmacophore features of the active compounds binding at these two sites, as well as directed mutagenesis studies

145

that would affect the predicted potential binding modes identified computationally. These compounds will be used to identify new hits and to develop structure-activity relationships (SAR) in order to identify more potent and selective lead compounds against PTase.

Conclusion

This study used an ensemble docking approach to identify structural variations of FXa using MD simulations, and targeted these conformations with compounds that specifically bind to the FXa-FVa interface to modulate the PTase activity in an allosteric and non-active site dependent manner. In contrast to competitive active site inhibitors, which completely inhibit enzyme activity at saturating inhibitor concentrations, these allosteric PTase inhibitors exhibit a plateau in inhibition referred to as 'self-limiting inhibition'. PTase activity assays identified a total of 26 compounds as inhibitors. Only 8 of these compounds were predicted to dock in the crystal structure. From these non-active site inhibitors, 10 compounds, 4 of which were identified by initial screening and 6 that were structural analogs, achieved self-limiting inhibition of PTase. Only two of these compounds bind to the crystal structure, and two compounds and their analogs would not have been identified if the virtual screen had used only this one structure.

This study exemplifies a successful MD simulation and high-throughput virtual screening approach to drug discovery targeting a protein-protein complex. Amaro et al. showed the importance of employing molecular dynamics to account for protein flexibility in the discovery of new molecules for African sleeping sickness.(52) Here we apply a similar ensemble docking approach, but use a larger compound database covering a larger chemical space, with around 3.4 million docking calculations completed in total. In future studies the same approach can be extended to even larger compound databases covering millions of compounds and targeting

multiple protein targets, using massive high-throughput screenings on supercomputers like TITAN.(53,54) This can increase dramatically the number and diversity of compounds that can be identified, while at the same time taking into account the inherent flexibility and the dynamic nature of macromolecules and their complexes under physiological conditions.

LIST OF REFERENCES

1.   Davie, E. W., and Ratnoff, O. D. (1964) Waterfall Sequence for Intrinsic Blood Clotting. *Science* **145**, 1310-1312

2.   Weitz, J. I. (2015) Anticoagulation therapy in 2015: where we are and where we are going. *Journal of thrombosis and thrombolysis*

3.   Bauer, K. A. (2013) Pros and cons of new oral anticoagulants. *Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program* **2013**, 464-470

4.   Alexander, J. H., Lopes, R. D., James, S., Kilaru, R., He, Y., Mohan, P., Bhatt, D. L., Goodman, S., Verheugt, F. W., Flather, M., Huber, K., Liaw, D., Husted, S. E., Lopez-Sendon, J., De Caterina, R., Jansky, P., Darius, H., Vinereanu, D., Cornel, J. H., Cools, F., Atar, D., Leiva-Pons, J. L., Keltai, M., Ogawa, H., Pais, P., Parkhomenko, A., Ruzyllo, W., Diaz, R., White, H., Ruda, M., Geraldes, M., Lawrence, J., Harrington, R. A., Wallentin, L., and Investigators, A.-. (2011) Apixaban with antiplatelet therapy after acute coronary syndrome. *The New England journal of medicine* **365**, 699-708

5.   Kane, W. H., and Davie, E. W. (1988) Blood coagulation factors V and VIII: structural and functional similarities and their relationship to hemorrhagic and thrombotic disorders. *Blood* **71**, 539-555

6.   Mann, K. G., Nesheim, M. E., Church, W. R., Haley, P., and Krishnaswamy, S. (1990) Surface-dependent reactions of the vitamin K-dependent enzyme complexes. *Blood* **76**, 1-16

7.   Davie, E. W., Fujikawa, K., and Kisiel, W. (1991) The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry* **30**, 10363-10370

148

8. Bianchini, E. P., Orcutt, S. J., Panizzi, P., Bock, P. E., and Krishnaswamy, S. (2005) Ratcheting of the substrate from the zymogen to proteinase conformations directs the sequential cleavage of prothrombin by prothrombinase. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 10099-10104

9. Nesheim, M. E., Taswell, J. B., and Mann, K. G. (1979) The contribution of bovine Factor V and Factor Va to the activity of prothrombinase. *The Journal of biological chemistry* **254**, 10952-10962

10. Rosing, J., Tans, G., Govers-Riemslag, J. W., Zwaal, R. F., and Hemker, H. C. (1980) The role of phospholipids and factor Va in the prothrombinase complex. *The Journal of biological chemistry* **255**, 274-283

11. Borensztajn, K., Peppelenbosch, M. P., and Spek, C. A. (2008) Factor Xa: at the crossroads between coagulation and signaling in physiology and disease. *Trends in molecular medicine* **14**, 429-440

12. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic acids research* **28**, 235-242

13. Steen, M., and Dahlback, B. (2002) Thrombin-mediated proteolysis of factor V resulting in gradual B-domain release and exposure of the factor Xa-binding site. *The Journal of biological chemistry* **277**, 38424-38430

14. Toso, R., and Camire, R. M. (2004) Removal of B-domain sequences from factor V rather than specific proteolysis underlies the mechanism by which cofactor function is realized. *The Journal of biological chemistry* **279**, 21643-21650

15. Heeb, M. J., Kojima, Y., Hackeng, T. M., and Griffin, J. H. (1996) Binding sites for blood coagulation factor Xa and protein S involving residues 493-506 in factor Va. *Protein science : a publication of the Protein Society* **5**, 1883-1889

16. Kalafatis, M., and Beck, D. O. (2002) Identification of a binding site for blood coagulation factor Xa on the heavy chain of factor Va. Amino acid residues 323-331 of factor V represent an interactive site for activated factor X. *Biochemistry* **41**, 12715-12728

17. Kalafatis, M., Xue, J., Lawler, C. M., and Mann, K. G. (1994) Contribution of the heavy and light chains of factor Va to the interaction with factor Xa. *Biochemistry* **33**, 6538-6545

18. Kojima, Y., Heeb, M. J., Gale, A. J., Hackeng, T. M., and Griffin, J. H. (1998) Binding site for blood coagulation factor Xa involving residues 311-325 in factor Va. *The Journal of biological chemistry* **273**, 14900-14905

19. Steen, M., Villoutreix, B. O., Norstrom, E. A., Yamazaki, T., and Dahlback, B. (2002) Defining the factor Xa-binding site on factor Va by site-directed glycosylation. *The Journal of biological chemistry* **277**, 50022-50029

20. Kittur, F. S., Manithody, C., and Rezaie, A. R. (2004) Role of the N-terminal epidermal growth factor-like domain of factor X/Xa. *The Journal of biological chemistry* **279**, 24189-24196

21. Rezaie, A. R. (2000) Identification of basic residues in the heparin-binding exosite of factor Xa critical for heparin and factor Va binding. *The Journal of biological chemistry* **275**, 3320-3327

22. Rezaie, A. R., and Kittur, F. S. (2004) The critical role of the 185-189-loop in the factor Xa interaction with Na+ and factor Va in the prothrombinase complex. *The Journal of biological chemistry* **279**, 48262-48269

23. Rudolph, A. E., Porche-Sorbet, R., and Miletich, J. P. (2000) Substitution of asparagine for arginine 347 of recombinant factor Xa markedly reduces factor Va binding. *Biochemistry* **39**, 2861-2867

24. Rudolph, A. E., Porche-Sorbet, R., and Miletich, J. P. (2001) Definition of a factor Va binding site in factor Xa. *The Journal of biological chemistry* **276**, 5123-5128

25. Thiec, F., Cherel, G., and Christophe, O. D. (2003) Role of the Gla and first epidermal growth factor-like domains of factor X in the prothrombinase and tissue factor-factor VIIa complexes. *The Journal of biological chemistry* **278**, 10393-10399

26. Yegneswaran, S., Mesters, R. M., and Griffin, J. H. (2003) Identification of distinct sequences in human blood coagulation factor Xa and prothrombin essential for substrate and cofactor recognition in the prothrombinase complex. *The Journal of biological chemistry* **278**, 33312-33318

27. Kamata, K., Kawamoto, H., Honma, T., Iwama, T., and Kim, S. H. (1998) Structural basis for chemical inhibition of human blood coagulation factor Xa. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 6630-6635

28. Adams, T. E., Hockin, M. F., Mann, K. G., and Everse, S. J. (2004) The crystal structure of activated protein C-inactivated bovine factor Va: Implications for cofactor function. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 8918-8923

29. Autin, L., Steen, M., Dahlback, B., and Villoutreix, B. O. (2006) Proposed structural models of the prothrombinase (FXa-FVa) complex. *Proteins* **63**, 440-450

30. Wallnoefer, H. G., Handschuh, S., Liedl, K. R., and Fox, T. (2010) Stabilizing of a globular protein by a highly complex water network: a molecular dynamics simulation study on factor Xa. *The journal of physical chemistry. B* **114**, 7405-7412

31. Lechtenberg, B. C., Murray-Rust, T. A., Johnson, D. J., Adams, T. E., Krishnaswamy, S., Camire, R. M., and Huntington, J. A. (2013) Crystal structure of the prothrombinase complex from the venom of Pseudonaja textilis. *Blood* **122**, 2777-2783

32. Griffon, N., and Di Stasio, E. (2001) Thermodynamics of Na+ binding to coagulation serine proteases. *Biophysical chemistry* **90**, 89-96

33. Underwood, M. C., Zhong, D., Mathur, A., Heyduk, T., and Bajaj, S. P. (2000) Thermodynamic linkage between the S1 site, the Na+ site, and the Ca2+ site in the protease domain of human coagulation factor xa. Studies on catalytic efficiency and inhibitor binding. *The Journal of biological chemistry* **275**, 36876-36884

34. (2014) Molecular Operating Environment (MOE), 2013.08. in *Chemical Computing Group Inc.*

35. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry. B* **102**, 3586-3616

36.   Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *Journal of computational chemistry* **26**, 1781-1802

37.   Bakan, A., Meireles, L. M., and Bahar, I. (2011) ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* **27**, 1575-1577

38.   Shortle, D., Simons, K. T., and Baker, D. (1998) Clustering of low-energy conformations near the native structures of small proteins. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 11158-11162

39.   Trott, O., and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* **31**, 455-461

40.   Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., and Coleman, R. G. (2012) ZINC: a free tool to discover chemistry for biology. *Journal of chemical information and modeling* **52**, 1757-1768

41.   Voigt, J. H., Bienfait, B., Wang, S., and Nicklaus, M. C. (2001) Comparison of the NCI open database with seven large chemical structural databases. *Journal of chemical information and computer sciences* **41**, 702-712

42.   O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011) Open Babel: An open chemical toolbox. *Journal of cheminformatics* **3**, 33

43.   Halgren, T. A. (1998) Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **17**, 490-519

44.   Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry* **30**, 2785-2791

45.   Camacho, D. R. K. a. C. J. (2012) ZINCPharmer: pharmacophore search of the ZINC database. *Nucleic acids research* **40**, W409-W414

46.   Martin, R. D. B. a. Y. C. (1996) Use of Structure−Activity Data To Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. *J. Chem. Inf. Comput. Sci* **36**, 572–584

47.   Corbeil, C. R., Williams, C. I., and Labute, P. (2012) Variability in docking success rates due to dataset preparation. *Journal of computer-aided molecular design* **26**, 775-786

48.   Labute, P. (2008) The generalized Born/volume integral implicit solvent model: estimation of the free energy of hydration using London dispersion instead of atomic surface area. *Journal of computational chemistry* **29**, 1693-1698

49.   Daura, X., Haaksma, E., and van Gunsteren, W. F. (2000) Factor Xa: simulation studies with an eye to inhibitor design. *Journal of computer-aided molecular design* **14**, 507-529

50.   Singh, N., and Briggs, J. M. (2008) Molecular dynamics simulations of Factor Xa: insight into conformational transition of its binding subsites. *Biopolymers* **89**, 1104-1113

51.   Venkateswarlu, D., Perera, L., Darden, T., and Pedersen, L. G. (2002) Structure and dynamics of zymogen human blood coagulation factor X. *Biophysical journal* **82**, 1190-1206

52.   Amaro, R. E., Schnaufer, A., Interthal, H., Hol, W., Stuart, K. D., and McCammon, J. A. (2008) Discovery of drug-like inhibitors of an essential RNA-editing ligase in

Trypanosoma brucei. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 17278-17283

53.   Ellingson, S. R., Dakshanamurthy, S., Brown, M., Smith, J. C., and Baudry, J. (2014) Accelerating Virtual High-Throughput Ligand Docking: current technology and case study on a petascale supercomputer. *Concurrency and computation : practice & experience* **26**, 1268-1277

54.   S. Ellingson, J. C. S., and J. Baudry. (2014) Polypharmacology and Supercomputer-Based Docking: Opportunities and Challenges. *Molecular Simulation*

APPENDIX

Table 3.1. Ten sub-binding sites identified at the FXa:FVa interface and the corresponding FXa residues in these sites that are proposed to interact with FVa (Figure 3.1A).

Sub-binding site	Residues
1	Arg-125, Asp-126, Trp-127
2	Val-231, Leu-235
3	Val-231, Ile-238
4	Arg-125, Asp-126, Val-231, Thr-232, Ala-233, Phe-234, Leu-235, Lys-236, Trp-237, Ile-238, Asp-239, Arg-240
5	Glu-129, Thr-232, Ala-233, Phe-234, Leu-235, Lys-236, Trp-237, Ile-238
6	Ser-241, Met-242, Lys-243, Arg-93, Phe-94
7	Arg-93, Phe-94, Thr-95, Lys-96
8	Arg-93, Phe-94, Thr-95, Lys-96
9	Arg-125, Asp-126, Trp-127, Ala-128, Glu-129, Asp-164, Arg-165, Asn-166, Ser-167, Cys-168, Lys-169, Leu-170, Val-231, Thr-232, Ala-233
10	Asp-164, Arg-165, Asn-166, Ser-167, Cys-168, Lys-169, Leu-170

Table 3.2. Nine self-limiting inhibitors identified by PTase activity assays. The compounds without "Binding Sites" values are the structural analogs of compounds obtained for initial screening. Max % inhibition and Ki, were obtained by fitting activity inhibition as a function of inhibitor concentration.

SBC ID	Inhibition @ 100 uM	Max % inhibition	Ki, µM	Binding Sites	Binds Xtal Structure
SBC-160,010	65-70	80±10	25	7/8*	No
SBC-160,012	85-90	95±5	10	9,9,9,7	Yes
SBC-160,029	15-30	25±5	15	9,9,9,9	Yes
SBC-160,042	60-70	85±10	30	9,7	No
SBC-160,109	85-95	90±5	5	-	-
SBC-160,114	20-25	25±5	10	-	-
SBC-160,119	35-40	40±5	10	-	-
SBC-160,124	15-20	23±5	5	-	-
SBC-160,125	20-25	25±5	20	-	-

*Binding across both sites 7 and 8

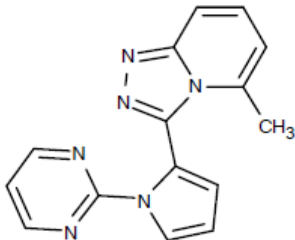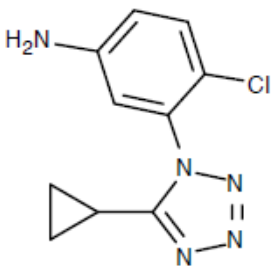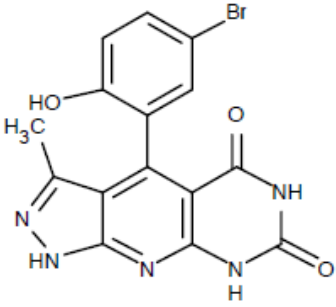**Table 3.3.** Compounds experimentally tested to be active. Binding snapshot 1 represents the crystal structure.

Compound	Structure	Molecular Weight	Binding Snapshots	Binding sites
SBC-160,012		276	1,2,4,10	9,9,9,7
SBC-160,029		236	1,2,5,9	9,9,9,9
SBC-160,042		388	4,10	9,7

Table 3.4. Differences in binding free energy for selected compounds in binding site 9 and binding site 7 in snapshot 10.

Compound	$\Delta\Delta G$ (ΔG site 7 – ΔG site 9) Kcal/mol
SBC-160,012	-3.72
SBC-160,042	+2.38

Table 3.5. Cluster population for the snapshots that compounds 012, 029 and 042 bind to.

Snapshot number	Cluster population (in % of MD structures)
2	10.7
4	4
5	1.3
9	2.3
10	0.5

Figure 3.1. FXa-FVa binding site. (A) Human FXa residues proposed to interact with human FVa, based on the FVa–FXa complex models and the available experimental data. Binding site residues 93-96, 125-129, 164-170, 231-244 shown in green. (B) Residues (green) form part of 3 helices- H1, H2 and H3 and a loop, L1. (C) Superposition of human FXa with the FX-FV complex from *P. textilis*. Only superposed human FXa (red) and *P. textilis* FV (blue) are shown and *P. textilis* FX structure is not shown for clarity. Binding interface (green) in human FXa proposed to interact with human FVa.
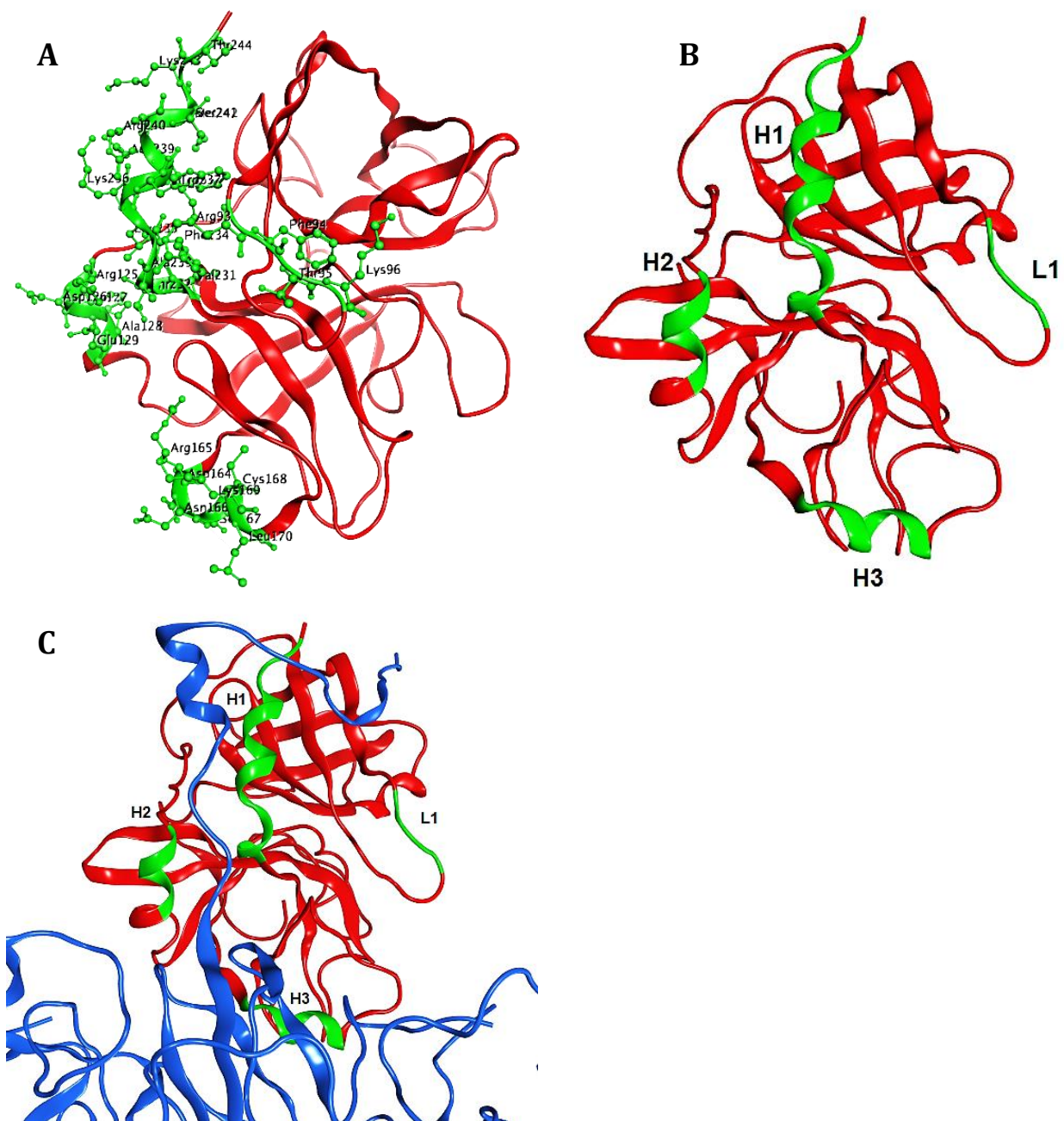
Figure 3.2. Flowchart showing the Ensemble docking process. I. (a) 300 ns MD simulation carried out followed by (b) Nearest-neighbor RMSD clustering, generating 15 snapshots. II. 244K compounds docked, covering 90% of 11.1 million "clean drug-like" subset from the Zinc database. III. Parsed for compounds that bind to FXa:FVa interface in the crystal structure or at least one of the 11 snapshots. Only 532 out of 244,493 docked compounds bind at this interface in the crystal structure. IV. (a) Top 80 compounds in the crystal structure + top 40 compounds in each of 11 snapshots selected (520 total) for (b) pharmacophore search with ZincPharmer for larger compounds (MW 280-600 amu's), against purchasable Zinc database compounds. V. Compounds parsed for drug-like properties and redundant compounds removed. VI. 37,000 larger compounds docked in 11 snapshots + crystal structure. VII. Parsed again for compounds that bind to FXa:FVa interface in the crystal structure or at least one of the 11 snapshots. Only 177 out of 36,635 docked larger compounds bind at this interface in the crystal structure. VIII. (a) Top 50 compounds in crystal structure + top 30 compounds from each of 11 snapshots based on docking scores selected: 288 compounds (after removing redundancy), from which 44 bind to the crystal structure and others bind to MD snapshots. (b) 247 compounds selected that bind to most number of snapshots and only 1 or 2 different binding sites on FXa-FVa interface. Out of these 79 bind to the crystal structure and others bind to MD snapshots. IX. Compound selection based on compounds' availability, cost, chemical diversity, and lack of reactive functional groups. X. (a) Analog generation based on initial positive hits. (b) Experimental validation of these analogs.
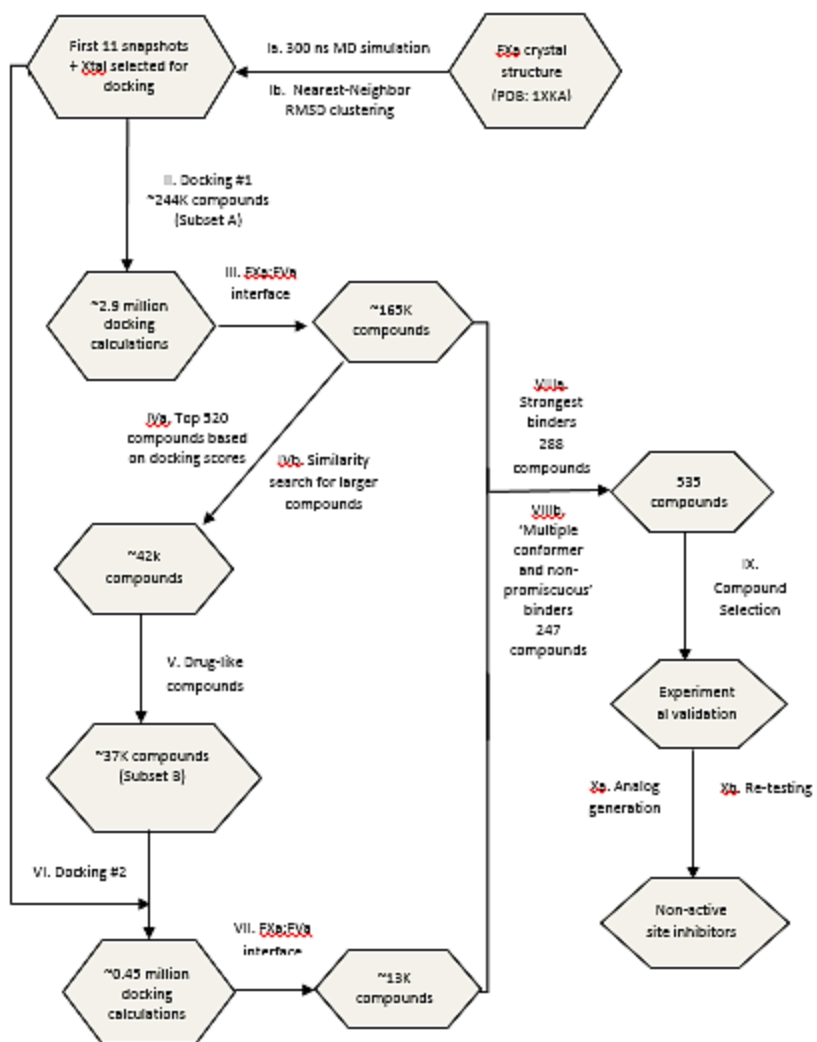
164

Figure 3.3. MD Simulations and Clustering. (A) Superposition of 3 FXa structures at different times in the MD trajectory. Starting crystal structure shown in green, structure at 150 ns in blue and structure at 300 ns in orange, with binding site residues shown as stick-ball representation. (B) RMSD plot over the time course of the trajectory. (C) Superposition of representative non-redundant structures from 15 clusters with binding site residues shown in white. Structures in (C) represented on the RMSD plot (B) as black dots.
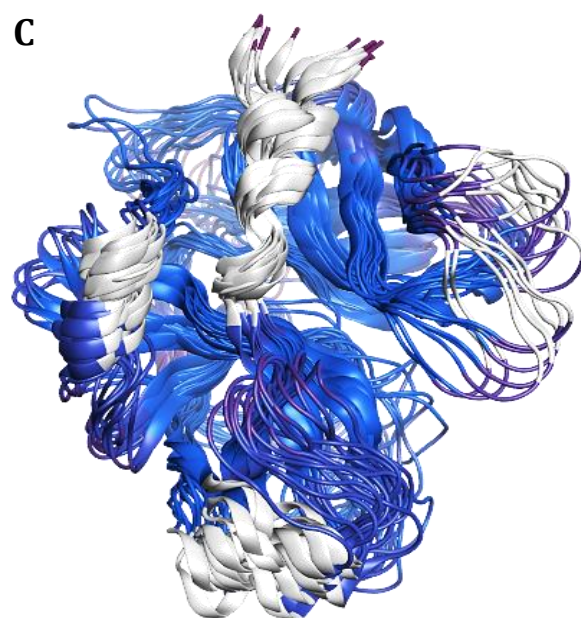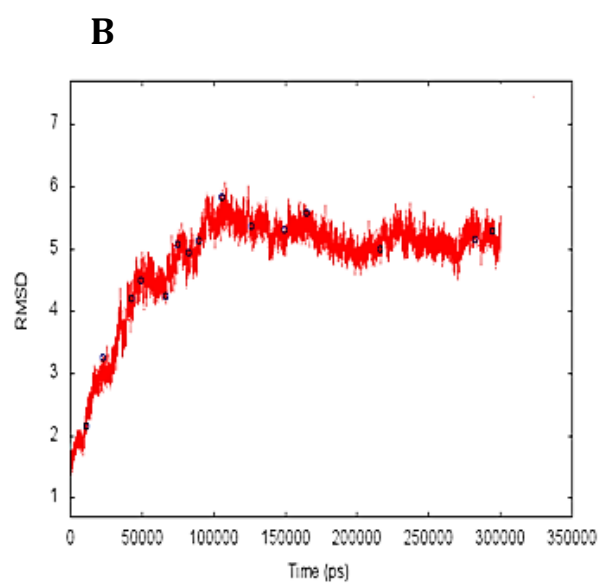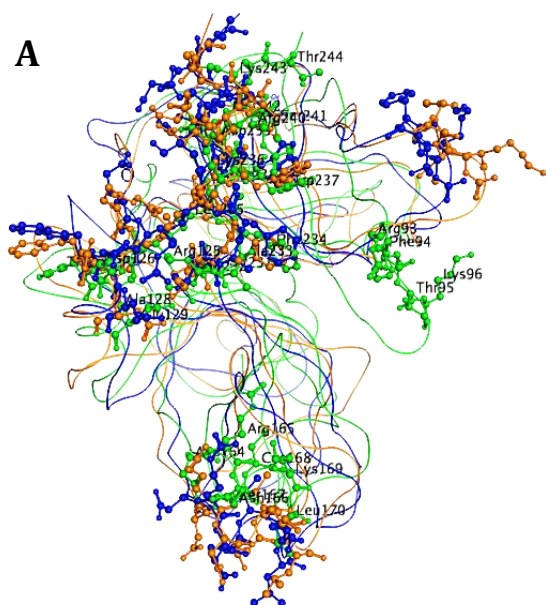
A

B

C

Figure 3.4. Docking analysis. (A) Smaller original compound from the Subset A in blue and larger similar compound from the Subset B in brown binding in the FXa-Va interface in the crystal structure. The exposed regions are shown as red on the molecular surface, hydrophobic regions in green and polar regions in magenta. Larger compounds cover more binding surface. (B) In crystal structure compounds bind in four regions: grooves A, B, C, D and E. (C) Ten binding sites considered taking all 12 structures into account.
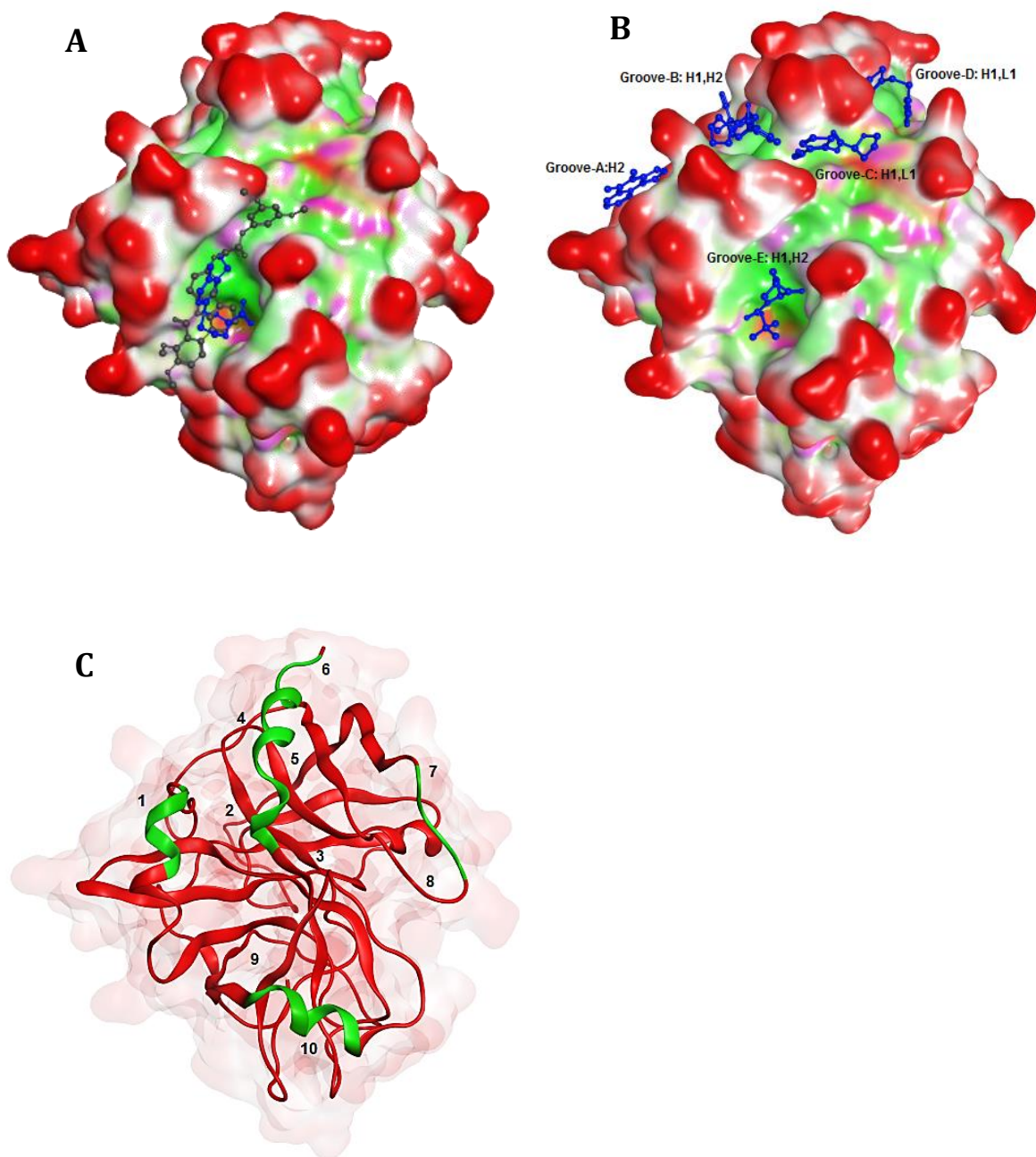
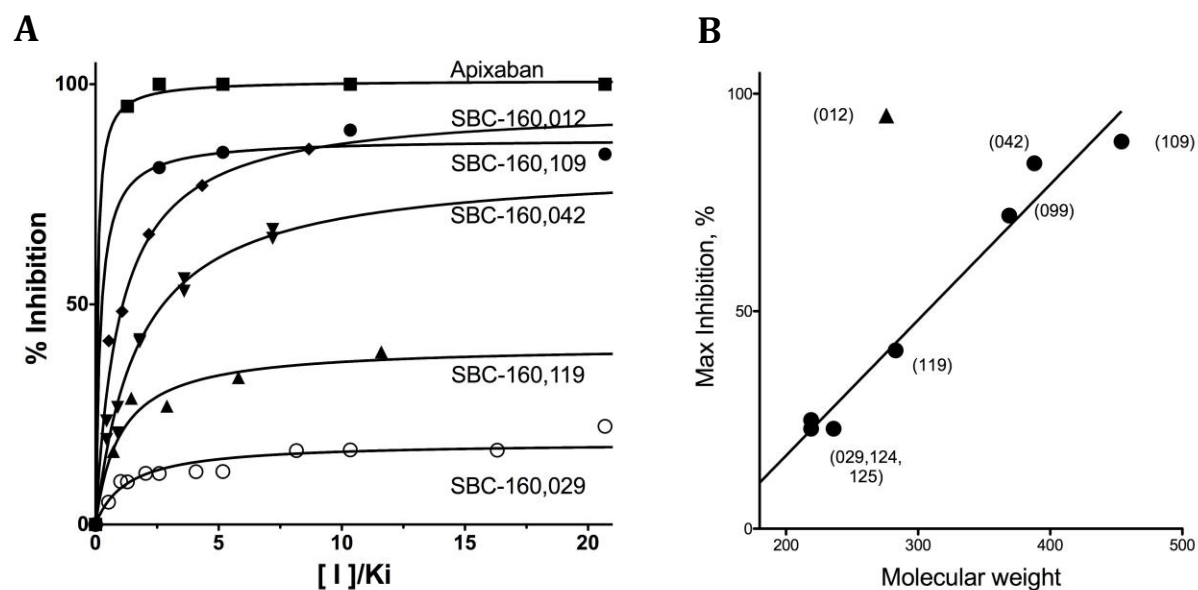**Figure 3.5. PT inhibition.** (A) Effect of the identified inhibitors on PT activity based on activity measurements, presented as % inhibition vs. [I]/Ki. (B) Plot of the maximum inhibition vs. the molecular weights of identified PTase inhibitors and corresponding analogs.
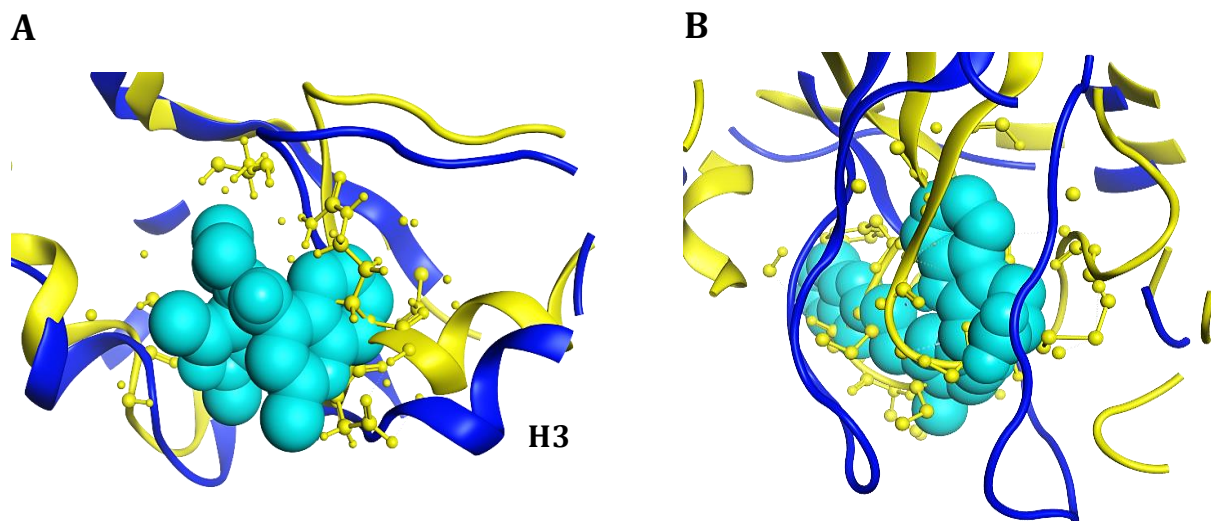
Figure 3.6. Structural variations with MD. Superposition of loop backbones in FXa-FVa interacting region. Yellow: FXa crystal structure. Blue: snapshots from the MD trajectory. (A) Inhibitor (cyan) binds to sub-binding site 9 in snapshot 3 where the Helix 3 (H3) is pushed out, and not in the crystal structure. (B) Non-active site binder (cyan) binds to sub-binding site 10 in snapshot 7 where two loops move apart, opening the binding site. These loops occlude the binding site in the crystal structure.

Figure 3.7. Site 9 and 7 in FX-FV complex. (A) Superposition of human FXa (crystal structure) bound with a site-9 inhibitor, with the FX-FV complex from *P. textilis*. Only superposed human FXa and *P. textilis* FV are shown and *P. textilis* FX structure is not shown for clarity. Red: human FXa structure. Cyan: A2 domain from *P. textilis* FV. Green: A3 domain from *P. textilis* FV. Purple: Flexible terminal region in FV. Site-9 inhibitor binds directly at the interface of FXa and A2/A3 domains of FV. (B) 90 degree rotation of (A). (C) Superposition of human FXa (snapshot 10 from MD simulation) bound with site-7 inhibitor, with the FX-FV complex. Hydrophobic portion of terminal region shown in stick conformation. (D) 90 degree rotation of (C).
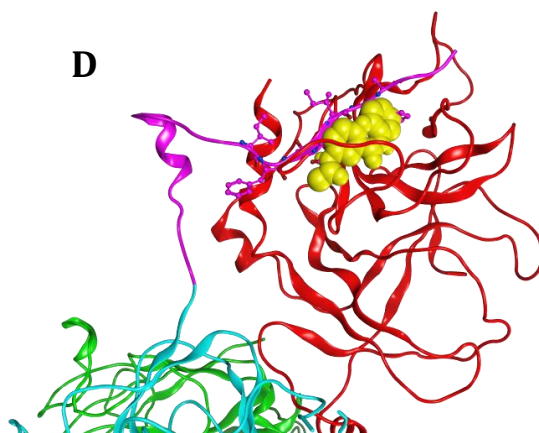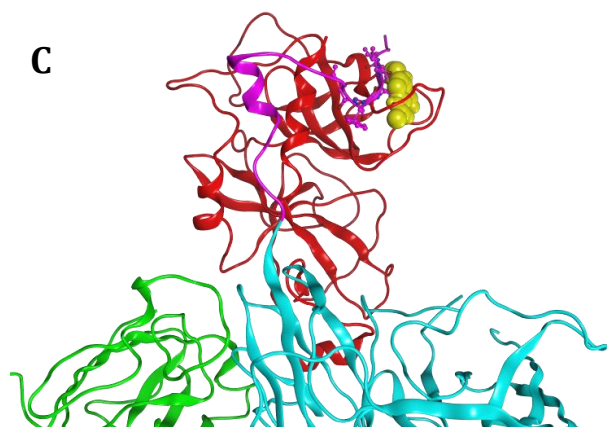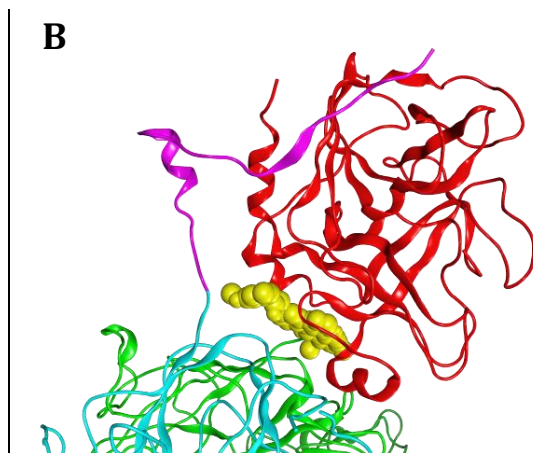
Chapter 4

ELUCIDATION OF THE PROTEIN-PROTEIN

BINDING/DYNAMICS OF ELECTRON

TRANSPORT PROTEINS WITH

CYANOBACTERIAL PHOTOSYSTEM I

Research described in this chapter is reproduced from a manuscript prepared by Karan Kapoor, Derek J. Cashman, Barry D. Bruce and Jerome Baudry.

My contribution in the paper includes (1) Designing and carrying out all computation research (2) Understanding the literature and interpretation of the results (3) Preparation of tables and figures (4) Writing and editing.

Abstract

The Luminal domain (PsaA, PsaB, and PsaF) of photosystem I (PSI) in Cyanobacteria accepts electrons from Cytochrome c6. After a series of electron transfer steps, the stromal domain (PsaC, PsaD, and PsaE) transfers these electrons and reduces the transiently bound ferredoxin (Fd) or flavodoxin. Experimental structures exist for all of these protein partners individually, but no experimental structure of the PSI/Cytc6 and PSI/Fd complexes is presently available. Molecular models of Cytc6 docked on the lumen side and Fd docked onto the stromal domain of the cyanobacterial PSI site are constructed here utilizing X-ray and NMR structures of PSI, Cytc6 and Fd. Predictions of potential protein-protein interaction regions are based on previous experimental site-directed mutagenesis and cross-linking studies to guide rigid body docking calculations of these proteins. Agnostic docking without any constraints are also carried out, validating the PSI/Fd results, and providing an alternate binding mode for the PSI/Cytc6 complex. The results are complemented by the calculations of coevolving residues between the interacting partners on the two domains of PSI. These results further validate the complex models and provide a comprehensive picture of the binding and interactions between the PSI and the electron-transport proteins.

Introduction

The process of photosynthesis converts light energy into chemical energy through two large, multi-subunit proteins bound to the thylakoid membranes in the chloroplasts of cyanobacteria, algae and plants. These large, multi-subunit proteins are referred to as Photosystem II (PSII) and Photosystem I (PSI). Upon photo-excitation, PSII extracts electrons from water, and transfers them to the b6/f complex and ultimately to PSI using soluble electron transfer proteins such as Cytc6 or plastocyanin (PC). PSI then undergoes a second photo-excitation, which induces a cascade of electron transfer steps through several redox cofactors within the protein, ending up at one of two terminal electron acceptors, the 4Fe-4S centers in the PsaC subunit. Electrons are then transferred from these 4Fe-4S centers to the 2Fe-2S center of ferredoxin (Fd) (1,2) or the flavin mononucleotide (FMN) cofactor of flavodoxin under iron-deficient conditions (3,4). The reduced Fd transfers two electrons to the ferredoxin-NADP oxidoreductase (FNR) to produce NADPH, which functions as a reducing agent in Calvin Cycle reactions as well as several other metabolic processes (5). Understanding the binding and affinity of Fd and Cytc6 with PSI is a key element of the structure/function relationship of the electron transfer mechanisms during the process of photosynthesis.

The purpose of this study is to examine the atomic-level intermolecular interactions at the protein-protein interface in cyanobacterial photosynthetic proteins using advanced computational methods. Using the NMR structures of cyanobacterial Fd (6,7) and Cytc6 (8), and X-ray crystal structures of cyanobacterial PSI (9), molecular models of Fd docked onto the stromal domain and Cytc6 docked on the luminal domain of the cyanobacterial PSI are constructed. Docking models of PSI-Fd have recently been published by Cashman et al (10). The newly generated models are compared with these previously published models. Docking studies were guided by predictions of

177

protein-protein interaction regions from experimental site-directed mutagenesis (11,12) and cross-linking (13,14) studies. Several plausible docked configurations are ranked based on the predictions of the coevolving residues between the interacting partners, forming the basis of the computational model prediction presented here.

Materials and Methods

Structure preparation

Protein structures from T. elongatus, a thermophilic cyanobacterium, were used to model the interaction sites between the X-ray crystal structure of PSI and the NMR structures of Fd and Cytc6 (6-9). The structural files were downloaded from the PDB (15) using the PDB accession coordinates 1JB0 for photosystem I, 2CJN for Fd and 1C6S for Cytc6. In the case of PSI-Fd models, using MOE, v. 2012 (16) the PsaC, PsaD, and PsaE chains were extracted from the PDB file, and hydrogen atoms were added to the structure according to standard residue protonation. Similarly in the case of PSI-Cytc6 models, the PsaA, PsaB, and PsaF chains were extracted from the PDB file and hydrogen atoms were added. The remainder of the system was omitted, as it is not involved in interacting with Fd or Cytc6.

PSI-Fd and PSI-Cytc6 binding sites

To identify potential protein–protein interaction sites between the stromal domain of PSI (PsaC, PsaD, and PsaE) and Fd, the residues that have been experimentally determined on the basis of site-directed mutagenesis, cross-linking and other studies, to be involved in interactions with both systems (11,12,17,18) were used (Ile-11, Thr-14, Gln-15, Lys-34, Gly-36, Val-48, Lys-51,

Arg-52 from PsaC, His-95, Asp-98, Glu-103, Lys-104, Lys-107, Arg-109 from PsaD and Arg-39 from PsaE).

In the case of interactions between the Luminal domain of PSI (PsaA, PsaB, and PsaF) and Cytc6, the stretch of charged residues in Cytc6 shown to be involved in interactions with the Lumen side of PSI through mutagenesis studies (data unpublished) were used (Gly-63, Arg-64, Leu-65, Thr-66, Asp-67, Glu-68 and Gln-69). The stretch of residues from the multiple sequence alignment of PsaF domain in different species and containing the basic residues possibly involved in interaction with the corresponding acidic residues in Cytc6 (data unpublished) were also used (Ser-11, Pro-12, Ala-13, Phe-14, Gln-15, Lys-16, Arg-17, Ala-18, Ala-19, Ala-20, Ala-21, Val-22, Asn-23, Thr-24, Thr-25). In addition, explicit repulsions were added between Cytc6 and the residues in PsaA, PsaB and PsaF present on the stromal side or interacting with the surrounding membrane regions.

Agnostic docking, without using any constraints in the docking protocol, was done for comparison and to further validate the results generated using the residue filters for both PSI-Fd and PSI-Cytc6 models. In the case of PSI-Cytc6 models, explicit repulsions were retained to obtain a realistic set of results.

Rigid-body docking

Fast-fourier transform (FTT) based rigid-body docking program, Cluspro (19-21), was used for generating docking models between PSI and Fd and between PSI and Cytc6. The working of Cluspro is described as follow- Cluspro treats one of the proteins as the receptor (PSI in this case) and the other as the ligand (Fd or Cytc6). The translation and rotation of ligand with respect to the receptor is carried out on a grid, sampling around 109 positions of the ligand relative to the

receptor. Translation with the best score from each rotation is selected, using pairwise interaction potential scoring function based on attractive and repulsive contributions to the van der Waals interaction energy, electrostatic energy term, and the desolvation contributions. Of all the generated structures, 1000 rotation/translation combinations that have the lowest score are chosen for further clustering. Clustering of these 1000 ligand positions is done using pairwise C-α RMSD (of ligand residues) as the distance measure, within 10 Å of any receptor atom in these docked models. Ligand positions with the most 'neighbors' in 9 Å becomes a cluster center, and its neighbors the members of the cluster. These are removed from the set and the same procedure repeated till all structures are divided into clusters. The biophysical meaning of clustering is isolating highly populated low energy basins of the energy landscape, as large clusters are more likely to include native structures. This means that the cluster centers representing the largest clusters are found to be closest to the native state and are shown to give better results than ranking the structures only based on their energies. After clustering with this hierarchical approach, the ranked complexes are subjected to a (300 step and fixed backbone) van der Waals minimization using the CHARMM (22) potential to remove potential side chain clashes. ClusPro then outputs the centers of the largest clusters.

Calculation of Fe-Fe and Chlorophyll-Heme distances

The models generated using Cluspro were subjected to a further energy minimization of side-chains (fixed backbone) to a 0.01 RMS kcal/mol/Å2 gradient in MOE 2012, using the CHARMM27 forcefield (22). In the case of PSI- Fd models, the original PSI structure (1JB0) was superimposed on these models to obtain the Fe4S4 center in PsaC. Because the 2CJN structure does not include the Fe2S2 center of Fd, these atoms were obtained from the 1ROE Fd NMR

structure. To insert the FeS center into 2CJN, residues 37 through 48 of 1ROE were superimposed in MOE with the same residues of 1ROE, and the FeS cluster was copied from the 1ROE structure into the 2CJN structure.

Similarly, in the case of PSI- Cytc6 models, the original PSI and Cytc6 (1C6S) structures were superimposed on these models to obtain the Chlorophyll molecules in PsaA-PsaB and Heme group in Cytc6, respectively. The distances were calculated between the iron atoms in corresponding Fe-S centers in PsaC and Fd and between the Heme and Chlorophyll A in PsaA/B and Cytc6, respectively.

Results

PSI-Fd Modeling

PSI-Fd models generated using Cluspro consisted of a total of 12 separate clusters represented by their cluster centers. Table 1 shows these 12 clusters, the number of members belonging to each cluster, and the weighted energy scores. Figure 1A shows the RMSD deviation between these twelve clusters. These 12 structures are separated by relatively larger RMSD distances, with the closest centers (New_7 and New_10) having a RMSD deviation of 3.71 Å. These 12 structures can be considered as separate PSI-Fd states that are possibly sampled, weighted against the cluster population represented by each cluster center, i.e. cluster 1 center represents the structure most sampled and hence closest to the native state of the complex.

Table 2 shows the number of inter-residue protein contacts and the Fe-Fe distances between the FeS centers in PsaC and Fd in each of these 12 cluster centers. The top-3 models show more inter-residue contacts compared to previously published models of PSI-Fd (10). The Fe-Fe distance is under 13.5 A for most of these models, with the lowest distance shown by 3rd model

(8.5 Å). These distances closely compares with the Fe–Fe distances of 8.9–13.1 Å between the two FeS centers of PsaC observed in six available crystal structures of PSI from T. elongatus, Pisum sativum, and Arabidopsis thaliana (9,23-26), suggesting that the distances between the generated models may provide an adequate proximity for rapid electron transfer (27). The Fe-Fe distances on average are also smaller than the previously published models. Superposition of these 12 cluster centers shows Fd sampling multiple rotational modes around the same binding site near the PSI FeS center, shown in Figure 1B.

The agnostic docking protocol in Cluspro generated a total of 18 cluster centers, shown in Table 3. Figure 1C shows the superposition of the two top models (cluster centers with largest cluster populations) generated by Cluspro using the residues filter and the agnostic docking protocol. These two models are very similar with a RMSD of only 1.70 Å. The model shows direct interactions between Arg-41, Ser-63, Glu-23 and Glu-93 in Fd and the residues in PsaC, PsaD and PsaE that have been shown to be important for binding Fd through mutagenesis studies (Gln-15 in PsaC, Lys-34 in PsaC, Lys-104 in PsaD and Arg-39 in PsaE, respectively), given in Table 4. This complex structure, the distance between the FeS centers, and two salt bridges between Arg-39 in PsaE and Lys-104 in PsaD and Glu-93 and Glu-23 in Fd, is shown in Figure 1D.

The previously published models of PSI-Fd consisted of a set of 12 models represented by only three clusters shown in Figure 2A: cluster I represented by structures 2, 4, 5 and 7; cluster II represented by structures 1, 3, 6, 10, 11, 12; cluster III represented by structures 8, 9. These 12 structures are separated by relatively shorter RMSD distances. Figure 2B shows the comparison of RMSD of the newly generated models to the previous models. Structures in cluster I are found to be similar to New_12 model; cluster II similar to New_8 model; and cluster III similar to New_6 model. The previous three clusters (consisting of 12 models) likely represent a subset of the

conformational ensemble sampled by PSI-Fd, as seen from the much larger ensemble of structures generated by Cluspro, both by residues filter and agnostic approach.

PSI-Cytc6 Modeling

PSI-Cytc6 models generated using Cluspro consisted of a total of 15 separate clusters represented by their cluster centers. Table 5 shows these 15 clusters, the number of members belonging to each cluster, and the weighted energy scores. Figure 3A and B show the superposition and the RMSD deviation between these 15 cluster centers, respectively. These structures are also separated by relatively large RMSD distances, with the closest centers (2 and 9) having a RMSD deviation of 8.23 Å. The center of cluster 1 represents the center of the most populated cluster, and hence most sampled and closest to the native state of the complex. This structure shows direct interactions between the charged residues in Cytc6, PsaF and PsaB, given in Table 6. The complex structure, the distance between the Heme and Chlorophyll groups, two salt bridges between Glu-617 in PsaB and Lys-8 in Cytc6, and between Lys-16 in PsaF and Asp-67 in Cytc6, are shown in Figure 3C. The distance between the Heme group and Chlorophyll molecule is 14.03 Å, which is high, compared to the value predicted by the Moser-Dutton rule (< 7 Å) for distances suitable for electron transfer between these groups (28).

The agnostic docking protocol in Cluspro generated a total of 29 cluster centers, shown in Table 7. The top cluster center shows interactions between the hydrophobic surfaces of Cytc6, consisting of residues Val-25 and Met-26, and PsaA and PsaB, consisting of residues Trp-655 and Phe-644, shown in Figure 4A and Table 8. The distance between the Heme group and Chlorophyll molecule is 9.66 Å, closer to the value predicted by the Moser-Dutton rule. Trp-655 in PsaA and Trp-631 in PsaB that are well conserved and have been shown to be important in binding Cytc6 in

other species, *Spinacia oleracea* and *Chlamydomonas reinhardtii* (29,30), are present at the interface between these two interacting proteins, shown in Figure 4B.

Figure 4C shows the superposition of the top two PSI-Cytc6 models generated by Cluspro using the residue filter and the agnostic docking protocol. These models show a RMSD of 16.39 Å. The first model shows contacts that are rich in electrostatic interactions, whereas as the agnostic model shows more hydrophobic contacts.

Discussion

The rigid body directed-docking calculations yield several observations with regard to possible interactions between the electron-transport protein, Fd and Cytc6, and PSI. Docking calculations focused on the stromal (PsaC, PsaD, PsaE) and luminal (PsaA, PsaB, PsaF) domains of PSI where these electron-transport proteins bind. Previously published PSI-Fd models (10) had used a 'top-down' approach by using the regions of high Frustration to direct rigid-body docking calculations, with distance constraints applied between the possibly interacting residues/regions. In this study, only the residues in PSI that have been shown to interact with the electron-transport proteins have been used to direct the docking calculations. This allows the receptor and the ligand in the rigid-body docking protocol to sample well all the six degrees of freedom, generating ~ 109 positions of the ligand relative to the receptor. These are then filtered to only keep the structures that are thermodynamically most stable, giving preference to the complexes that show interactions between the input residues. This provides a 'bottoms-up' approach, where the best models are predicted from the possible ~ 109 generated models, and hence decreasing the possibility of sampling only certain fixed rotational/translational modes of the ligand with respect to the receptor.

The top cluster center for the PSI-Fd model shows direct interactions between the residues experimentally predicted to be important for binding Fd, Gln-15 in PsaC, Lys-34 in PsaC, Lys-104 in PsaD and Arg-39 in PsaE, and the Arg-41, Ser-63, Glu-23 and Glu-93 residues of Fd. There are two salt bridges present between Arg-39 in PsaE and Glu-93 in Fd, and Lys-104 in PsaD and Glu-23 in Fd. Low RMSD between the top PSI-Fd models (Figure 1C) generated using the residues filter and the agnostic docking protocol highlights the preference for the top binding mode generated by using two separate approaches.

Previous PSI-Fd models can be represented by a subset of the newly generated models; hence the new results provide a more comprehensive picture of the possible binding modes. Fd is found to sample multiple rotational modes around the same binding site near the PSI FeS center. It is possible that Fd may bind in two or more different orientations with PSI. In future, further refinement using molecular dynamics simulation on high-performance supercomputers of the generated docked complexes, will enable a thorough analysis of the protein dynamics and conformational changes that occur upon binding. This will help to determine if the system converges to a single binding mode, or if the complex samples multiple modes throughout.

Two separate models are generated for PSI-Cytc6 using the residues filter and the agnostic docking protocol, showing preference for electrostatic and hydrophobic interactions respectively. In the first model, Asp-67 in Cytc6, that has been shown experimentally to be important for binding with PSI (unpublished data), forms a salt-bridge with Lys-16 in PsaF. In the agnostic model, Trp-655 in PsaA, that is well conserved and has been shown to be important in binding Cytc6 in other species (29,30), forms hydrophobic interactions with Val-25 in Cytc6. Trp-631 in PsaB, also well-conserved, is also present in close proximity to Cytc6. It is possible that there is an equilibrium between these two modes, with the Cytc6 first interacting with PSI through long-range electrostatic

interactions, and then the system equilibrating/relaxing to the second mode where the hydrophobic surfaces in PSI and Cytc6 come together by minimizing the area exposed to the solvent. Between the two modes, the distance between Heme group and Chlorophyll decreases from 14.03 A to 9.66 Å, possibly facilitating in the transfer of electrons from Cytc6 to PSI.

In case of the two PSI-Cytc6 models, future MD simulation in explicit solvent of the two binding modes can show if there is an overlap between the models, and if the first model with preference for electrostatic interactions finally converges to the model showing preferentially hydrophobic interactions in the presence of the surrounding solvent molecules. The distance between Heme and Chlorophyll groups can also be expected to decrease further in the presence of solvent molecules, as the exposed hydrophobic surfaces will try to minimize the solvent accessible area by moving closer. These studies will provide an improved model that describes the atomic interactions and the possible binding mechanism at the luminal domain of photosystem I.

Conclusion

Overall, this study provides a new, detailed insight into visualizing possible binding mechanisms between PSI, Fd and Cytc6. These complexes are involved in the electron transfer reactions of photosynthesis, and a thorough understanding of their atomic-level interactions and dynamics will help in suggesting future mutagenesis experiments for validation purposes. This will also enable in the design more efficient mutants of these proteins, to optimize rate of electron transfer and hence the rate of energy generation from these systems, leading to the development of biologically-based solar energy cells for electricity generation.

LIST OF REFERENCES

1. Zanetti, G., and Merati, G. (1987) Interaction between photosystem I and ferredoxin. Identification by chemical cross-linking of the polypeptide which binds ferredoxin. *European journal of biochemistry / FEBS* **169**, 143-146

2. Zilber, A. L., and Malkin, R. (1988) Ferredoxin Cross-Links to a 22 kD Subunit of Photosystem I. *Plant physiology* **88**, 810-814

3. Muhlenhoff, U., Kruip, J., Bryant, D. A., Rogner, M., Setif, P., and Boekema, E. (1996) Characterization of a redox-active cross-linked complex between cyanobacterial photosystem I and its physiological acceptor flavodoxin. *The EMBO journal* **15**, 488-497

4. Setif, P. (2001) Ferredoxin and flavodoxin reduction by photosystem I. *Biochimica et biophysica acta* **1507**, 161-179

5. Berg, J. M., Tymoczko, J. L. & Stryer, L. (2007) *Biochemistry (6th Edition)*, W.H. Freeman & Company, New York

6. Hatanaka, H., Tanimura, R., Katoh, S., and Inagaki, F. (1997) Solution structure of ferredoxin from the thermophilic cyanobacterium Synechococcus elongatus and its thermostability. *Journal of molecular biology* **268**, 922-933

7. Baumann, B., Sticht, H., Scharpf, M., Sutter, M., Haehnel, W., and Rosch, P. (1996) Structure of Synechococcus elongatus [Fe2S2] ferredoxin in solution. *Biochemistry* **35**, 12831-12841

8. Beissinger, M., Sticht, H., Sutter, M., Ejchart, A., Haehnel, W., and Rosch, P. (1998) Solution structure of cytochrome c6 from the thermophilic cyanobacterium Synechococcus elongatus. *The EMBO journal* **17**, 27-36

9.   Jordan, P., Fromme, P., Witt, H. T., Klukas, O., Saenger, W., and Krauss, N. (2001) Three-dimensional structure of cyanobacterial photosystem I at 2.5 A resolution. *Nature* **411**, 909-917

10.   Cashman, D. J., Zhu, T., Simmerman, R. F., Scott, C., Bruce, B. D., and Baudry, J. (2014) Molecular interactions between photosystem I and ferredoxin: an integrated energy frustration and experimental model. *Journal of molecular recognition : JMR* **27**, 597-608

11.   Setif, P., Harris, N., Lagoutte, B., Dotson, S., and Weinberger, S. R. (2010) Detection of the photosystem I:ferredoxin complex by backscattering interferometry. *Journal of the American Chemical Society* **132**, 10620-10622

12.   Bottin, H., Hanley, J., and Lagoutte, B. (2001) Role of acidic amino acid residues of PsaD subunit on limiting the affinity of photosystem I for ferredoxin. *Biochemical and biophysical research communications* **287**, 833-836

13.   Lelong, C., Boekema, E. J., Kruip, J., Bottin, H., Rogner, M., and Setif, P. (1996) Characterization of a redox active cross-linked complex between cyanobacterial photosystem I and soluble ferredoxin. *The EMBO journal* **15**, 2160-2168

14.   Lelong, C., Setif, P., Lagoutte, B., and Bottin, H. (1994) Identification of the amino acids involved in the functional interaction between photosystem I and ferredoxin from Synechocystis sp. PCC 6803 by chemical cross-linking. *The Journal of biological chemistry* **269**, 10034-10039

15.   Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) The Protein Data Bank: a

computer-based archival file for macromolecular structures. *Journal of molecular biology* **112**, 535-542

16.   (2014) Molecular Operating Environment (MOE), 2013.08. in *Chemical Computing Group Inc.*

17.   Fischer, N., Setif, P., and Rochaix, J. D. (1999) Site-directed mutagenesis of the PsaC subunit of photosystem I. F(b) is the cluster interacting with soluble ferredoxin. *The Journal of biological chemistry* **274**, 23333-23340

18.   Setif, P., Fischer, N., Lagoutte, B., Bottin, H., and Rochaix, J. D. (2002) The ferredoxin docking site of photosystem I. *Biochimica et biophysica acta* **1555**, 204-209

19.   Kozakov, D., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R., and Vajda, S. (2013) How good is automated protein docking? *Proteins* **81**, 2159-2166

20.   Kozakov, D., Brenke, R., Comeau, S. R., and Vajda, S. (2006) PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins* **65**, 392-406

21.   Comeau, S. R., Gatchell, D. W., Vajda, S., and Camacho, C. J. (2004) ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* **20**, 45-50

22.   MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry. B* **102**, 3586-3616

23. Amunts, A., Drory, O., and Nelson, N. (2007) The structure of a plant photosystem I supercomplex at 3.4 A resolution. *Nature* **447**, 58-63

24. Amunts, A., Toporik, H., Borovikova, A., and Nelson, N. (2010) Structure determination and improved model of plant photosystem I. *The Journal of biological chemistry* **285**, 3478-3486

25. Chapman, H. N., Fromme, P., Barty, A., White, T. A., Kirian, R. A., Aquila, A., Hunter, M. S., Schulz, J., DePonte, D. P., Weierstall, U., Doak, R. B., Maia, F. R., Martin, A. V., Schlichting, I., Lomb, L., Coppola, N., Shoeman, R. L., Epp, S. W., Hartmann, R., Rolles, D., Rudenko, A., Foucar, L., Kimmel, N., Weidenspointner, G., Holl, P., Liang, M., Barthelmess, M., Caleman, C., Boutet, S., Bogan, M. J., Krzywinski, J., Bostedt, C., Bajt, S., Gumprecht, L., Rudek, B., Erk, B., Schmidt, C., Homke, A., Reich, C., Pietschner, D., Struder, L., Hauser, G., Gorke, H., Ullrich, J., Herrmann, S., Schaller, G., Schopper, F., Soltau, H., Kuhnel, K. U., Messerschmidt, M., Bozek, J. D., Hau-Riege, S. P., Frank, M., Hampton, C. Y., Sierra, R. G., Starodub, D., Williams, G. J., Hajdu, J., Timneanu, N., Seibert, M. M., Andreasson, J., Rocker, A., Jonsson, O., Svenda, M., Stern, S., Nass, K., Andritschke, R., Schroter, C. D., Krasniqi, F., Bott, M., Schmidt, K. E., Wang, X., Grotjohann, I., Holton, J. M., Barends, T. R., Neutze, R., Marchesini, S., Fromme, R., Schorb, S., Rupp, D., Adolph, M., Gorkhover, T., Andersson, I., Hirsemann, H., Potdevin, G., Graafsma, H., Nilsson, B., and Spence, J. C. (2011) Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73-77

26. Brunger, A. T., Adams, P. D., Fromme, P., Fromme, R., Levitt, M., and Schroder, G. F. (2012) Improving the accuracy of macromolecular structure refinement at 7 A resolution. *Structure* **20**, 957-966

27.   Page, C. C., Moser, C. C., Chen, X., and Dutton, P. L. (1999) Natural engineering principles of electron tunnelling in biological oxidation-reduction. *Nature* **402**, 47-52

28.   Moser, C. C., Page, C. C., Farid, R., and Dutton, P. L. (1995) Biological electron transfer. *Journal of bioenergetics and biomembranes* **27**, 263-274

29.   Ueda, T., Nomoto, N., Koga, M., Ogasa, H., Ogawa, Y., Matsumoto, M., Stampoulis, P., Sode, K., Terasawa, H., and Shimada, I. (2012) Structural basis of efficient electron transport between photosynthetic membrane proteins and plastocyanin in spinach revealed using nuclear magnetic resonance. *The Plant cell* **24**, 4173-4186

30.   Sommer, F., Drepper, F., Haehnel, W., and Hippler, M. (2004) The hydrophobic recognition site formed by residues PsaA-Trp651 and PsaB-Trp627 of photosystem I in Chlamydomonas reinhardtii confers distinct selectivity for binding of plastocyanin and cytochrome c6. *The Journal of biological chemistry* **279**, 20009-20017

APPENDIX

Table 4.1. PSI-Fd models generated using Cluspro. The 12 clusters, the number of members belonging to each cluster, and the weighted energy scores.

Cluster	Members	Representative	Weighted Score
1	247	Center	-744
1	247	Lowest Energy	-852
2	163	Center	-712.6
2	163	Lowest Energy	-850.2
3	140	Center	-714.4
3	140	Lowest Energy	-815.2
4	119	Center	-723.2
4	119	Lowest Energy	-882.8
5	106	Center	-712.4
5	106	Lowest Energy	-870.7
6	53	Center	-729.7
6	53	Lowest Energy	-784.2
7	44	Center	-740.2
7	44	Lowest Energy	-808.2
8	38	Center	-723
8	38	Lowest Energy	-769.5
9	22	Center	-747.9
9	22	Lowest Energy	-805.5
10	18	Center	-720.5
10	18	Lowest Energy	-758
11	13	Center	-754.3
11	13	Lowest Energy	-754.3
12	6	Center	-706.3
12	6	Lowest Energy	-753.4

Table 4.2. PSI-Fd models contacts and Fe-Fe distances. The number of inter-residue protein contacts and the Fe-Fe distances between the FeS centers in PsaC and Fd in each of the 12 cluster centers.

Models	Inter-residue protein contacts	Fe-Fe distance (Å)
1	29	11.7
2	24	9.9
3	25	8.5
4	26	9.7
5	23	13.5
6	27	9.95
7	27	12.4
8	22	10.8
9	33	15.7
10	25	12.6
11	25	21.6
12	29	12.4

Table 4.3. PSI-Fd models generated using Cluspro using agnostic docking. The 18 clusters, the number of members belonging to each cluster, and the weighted energy scores.

Cluster	Members	Representative	Weighted Score
0	248	Center	-615.7
0	248	Lowest Energy	-698.8
1	99	Center	-597.5
1	99	Lowest Energy	-667.4
2	85	Center	-590.8
2	85	Lowest Energy	-685.6
3	79	Center	-592.9
3	79	Lowest Energy	-648.6
4	77	Center	-590.6
4	77	Lowest Energy	-728.1
5	64	Center	-595.3
5	64	Lowest Energy	-711.5
6	57	Center	-607.1
6	57	Lowest Energy	-690.5
7	51	Center	-622
7	51	Lowest Energy	-650.9
8	43	Center	-685.4
8	43	Lowest Energy	-685.4
9	39	Center	-589
9	39	Lowest Energy	-646
10	32	Center	-626.9
10	32	Lowest Energy	-676.7
11	28	Center	-603.9
11	28	Lowest Energy	-636
12	17	Center	-583.1
12	17	Lowest Energy	-632.8
13	16	Center	-585.1
13	16	Lowest Energy	-647.8
14	14	Center	-592.6
14	14	Lowest Energy	-614.9
15	14	Center	-595.4
15	14	Lowest Energy	-684.1
16	4	Center	-585.8
16	4	Lowest Energy	-628.2
17	4	Center	-596.8
17	4	Lowest Energy	-604.7

Table 4.4. PSI-Fd best model interactions. Interactions between PSI and Fd in the top model (cluster center with largest cluster populations).

Type	Chain	Position	Residue	Chain	Position	Residue
HB	PsaC	13	CYS13.O	Fd	41	ARG41.NE
HB	PsaC	15	GLN15.OE1	Fd	41	ARG41.NE
HB	PsaC	18	ARG18.NH1	Fd	23	GLU23.OE1
HB	PsaC	18	ARG18.NE	Fd	30	GLU30.OE2
HB	PsaC	18	ARG18.NH2	Fd	31	GLU31.OE1
HB	PsaC	34	LYS34.NZ	Fd	63	SER63.OG
HB	PsaD	73	ARG73.NH1	Fd	31	GLU31.OE1
HB	PsaD	76	LYS76.NZ	Fd	22	ASP22.OD2
HB	PsaD	104	LYS104.NZ	Fd	23	GLU23.OE2
HB	PsaE	3	ARG3.NH1	Fd	66	ASP66.OD1
HB	PsaE	3	ARG3.NH2	Fd	68	ASP68.OD1
HB	PsaE	25	SER25.OG	Fd	44	ALA44.O
HB	PsaE	39	ARG39.NH1	Fd	46	SER46.O
HB	PsaE	39	ARG39.NH2	Fd	93	GLU93.OE2
HB	PsaE	48	TYR48.N	Fd	35	ASP35.OD2
HB	PsaE	51	SER51.OG	Fd	35	ASP35.OD2
HB	PsaE	53	SER53.OG	Fd	11	ASP11.OD1
HB	PsaE	57	THR57.O	Fd	39	SER39.OG
HB	PsaE	57	THR57.OG1	Fd	47	THR47.OG1
ION	PsaC	18	ARG18.NH1	Fd	23	GLU23.OE1
ION	PsaC	18	ARG18.NH1	Fd	27	ASP27.OD2
ION	PsaC	18	ARG18.NE	Fd	30	GLU30.OE2
ION	PsaC	18	ARG18.NH2	Fd	31	GLU31.OE1
ION	PsaD	73	ARG73.NH1	Fd	31	GLU31.OE1
ION	PsaD	76	LYS76.NZ	Fd	22	ASP22.OD2
ION	PsaD	104	LYS104.NZ	Fd	23	GLU23.OE2
ION	PsaE	3	ARG3.NH1	Fd	66	ASP66.OD1
ION	PsaE	3	ARG3.NH2	Fd	68	ASP68.OD1
ION	PsaE	39	ARG39.NH2	Fd	93	GLU93.OE2

Table 4.5. PSI-Cytc6 models generated using Cluspro. The 15 clusters, the number of

members belonging to each cluster, and the weighted energy scores.

| Cluster	Members	Representative	Weighted Score
0	234	Center	-998.5
0	234	Lowest Energy	-1129.2
1	149	Center	-1036.9
1	149	Lowest Energy	-1132.2
2	106	Center	-942
2	106	Lowest Energy	-1136.3
3	77	Center	-1009.1
3	77	Lowest Energy	-1130.4
4	70	Center	-938.7
4	70	Lowest Energy	-1082
5	68	Center	-980
5	68	Lowest Energy	-1168.5
6	57	Center	-946.3
6	57	Lowest Energy	-1013.4
7	52	Center	-1122.3
7	52	Lowest Energy	-1122.3
8	46	Center	-952.4
8	46	Lowest Energy	-1068.2
9	36	Center	-963.1
9	36	Lowest Energy	-1069.4
10	26	Center	-935.4
10	26	Lowest Energy	-977
11	25	Center	-1015.8
11	25	Lowest Energy	-1058.2
12	19	Center	-927.1
12	19	Lowest Energy	-1017.9
13	8	Center	-934
13	8	Lowest Energy	-1014.3
14	5	Center	-928.4
14	5	Lowest Energy	-977

Table 4.6. PSI-Cytc6 best model interactions. Interactions between PSI and Cytc6 in the top model.

Type	Chain	Position	Residue	Chain	Position	Residue
HB	PsaA	117	ASP129.OD2	Cytc6	56	ASN56.ND2
HB	PsaA	644	SER659.OG	Cytc6	16	ALA16.O
HB	PsaA	645	GLN660.NE2	Cytc6	24	VAL24.O
HB	PsaA	738	SER753.OG	Cytc6	25	VAL25.O
HB	PsaB	617	GLU617.OE1	Cytc6	8	LYS8.NZ
HB	PsaB	617	GLU617.OE2	Cytc6	66	THR66.OG1
HB	PsaB	618	SER618.OG	Cytc6	13	ASN13.ND2
HB	PsaB	621	TYR621.OH	Cytc6	13	ASN13.OD1
HB	PsaB	621	TYR621.OH	Cytc6	64	ARG64.NH2
HB	PsaB	628	ASP628.OD1	Cytc6	13	ASN13.N
HB	PsaF	16	LYS16.NZ	Cytc6	67	ASP67.O
HB	PsaF	17	ARG17.NH2	Cytc6	64	ARG64.O
HB	PsaF	17	ARG17.NH1	Cytc6	65	LEU65.O
HYD	PsaA	654	LEU669.CD1	Cytc6	25	VAL25.CG1
ION	PsaB	617	GLU617.OE1	Cytc6	8	LYS8.NZ
ION	PsaF	16	LYS16.NZ	Cytc6	67	ASP67.OD2

Table 4.7. PSI-Cytc6 models generated using Cluspro using agnostic docking. The 29 clusters, the number of members belonging to each cluster, and the weighted energy scores.

Cluster	Members	Representative	Weighted Score
0	140	Center	-663.4
0	140	Lowest Energy	-782.3
1	118	Center	-716
1	118	Lowest Energy	-716
2	81	Center	-618.6
2	81	Lowest Energy	-735.5
3	65	Center	-650.9
3	65	Lowest Energy	-778.3
4	47	Center	-708.1
4	47	Lowest Energy	-725.3
5	43	Center	-689.4
5	43	Lowest Energy	-689.4
6	42	Center	-648.7
6	42	Lowest Energy	-707.7
7	40	Center	-654.3
7	40	Lowest Energy	-721.9
8	39	Center	-724.8
8	39	Lowest Energy	-724.8
9	38	Center	-605.8
9	38	Lowest Energy	-677.8
10	38	Center	-663.4
10	38	Lowest Energy	-670.8
11	37	Center	-620
11	37	Lowest Energy	-721
12	37	Center	-673.5
12	37	Lowest Energy	-687.9
13	30	Center	-624.9
13	30	Lowest Energy	-665.6
14	27	Center	-618.4
14	27	Lowest Energy	-709.6
15	24	Center	-617.4
15	24	Lowest Energy	-654.2
16	18	Center	-662.2
16	18	Lowest Energy	-662.2

Table 4.7 continued

Cluster	Members	Representative	Weighted Score
17	14	Center	-644.7
17	14	Lowest Energy	-644.7
18	13	Center	-610.6
18	13	Lowest Energy	-635.6
19	13	Center	-694.2
19	13	Lowest Energy	-694.2
20	13	Center	-691.9
20	13	Lowest Energy	-691.9
21	12	Center	-636.4
21	12	Lowest Energy	-636.4
22	11	Center	-613.6
22	11	Lowest Energy	-662.3
23	11	Center	-677.8
23	11	Lowest Energy	-677.8
24	11	Center	-641.5
24	11	Lowest Energy	-641.5
25	11	Center	-614.8
25	11	Lowest Energy	-646.2
26	6	Center	-607.2
26	6	Lowest Energy	-652.1
27	2	Center	-608.1
27	2	Lowest Energy	-609.1
28	1	Center	-613.8
28	1	Lowest Energy	-613.8

Table 4.8. PSI-Cytc6 best model interactions (agnostic docking). Interactions between PSI and Cytc6 in the top model.

Type	Chain	Position	Residue	Chain	Position	Residue
HB	PsaA	609	THR624.OG1	Cytc6	49	TYR49.OH
HB	PsaA	617	SER632.OG	Cytc6	55	LYS55.NZ
HB	PsaA	618	HIS633.O	Cytc6	49	TYR49.OH
HB	PsaA	623	ASN638.ND2	Cytc6	56	ASN56.O
HB	PsaA	644	SER659.OG	Cytc6	60	ALA60.O
HB	PsaA	644	SER659.OG	Cytc6	64	ARG64.NE
HB	PsaA	645	GLN660.NE2	Cytc6	60	ALA60.O
HB	PsaA	645	GLN660.NE2	Cytc6	61	PHE61.O
HB	PsaA	735	ARG750.NH2	Cytc6	53	HIS53.O
HB	PsaA	738	SER753.OG	Cytc6	63	GLY63.N
HB	PsaB	636	GLN636.NE2	Cytc6	16	ALA16.O
HYD	PsaA	640	TRP655.CZ3	Cytc6	25	VAL25.CG1
HYD	PsaB	644	PHE644.CE1	Cytc6	26	MET26.CE
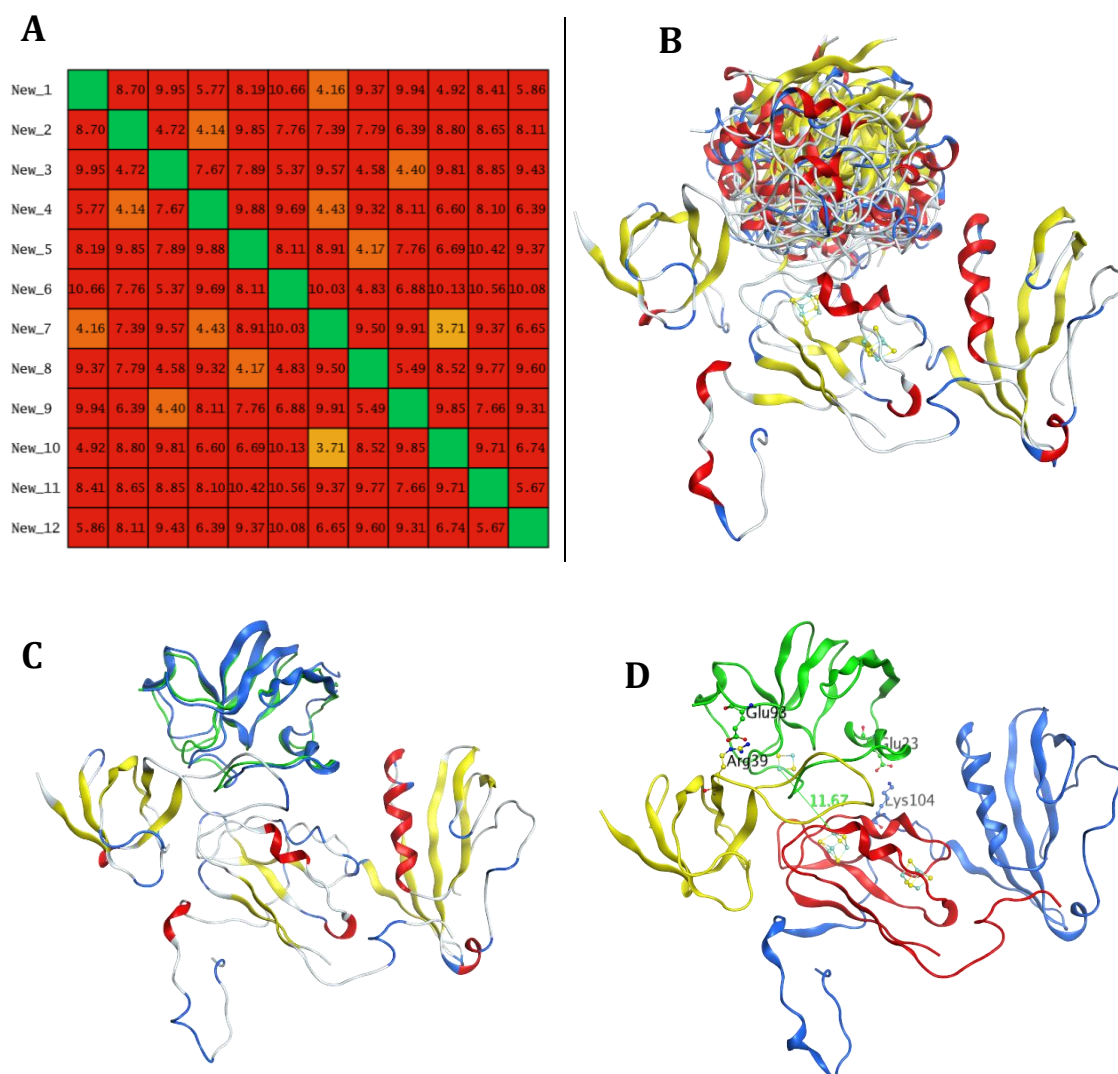
Figures 4.1. PSI-Fd models generated using Cluspro. (A) RMSD deviation between 12 cluster center. (B) Superposition of 12 cluster centers. (C) Superposition of two top models (cluster centers with largest cluster populations) generated by Cluspro using residues filter, Fd in green, and agnostic docking protocol, Fd in blue. (D) Distance between FeS centers and salt bridges between PSI and Fd in top model.

Figures 4.2. Comparison with previously published PSI-Fd models. (A) RMSD deviation between previously published models of PSI-Fd consisting of 12 models represented by 3 clusters. (B) Comparison of RMSD of the newly generated models to the previous models.

A

B

Figures 4.3. PSI-Cytc6 models generated using Cluspro. (A) Superposition of 15 cluster centers. (B) RMSD deviation between 15 cluster center. (C) Distance between Heme and Chlorophyll groups and salt bridges between PSI and Cytc6 in top model. PsaA in purple, PsaB in orange, PsaF in cyan and Cytc6 in red.

A
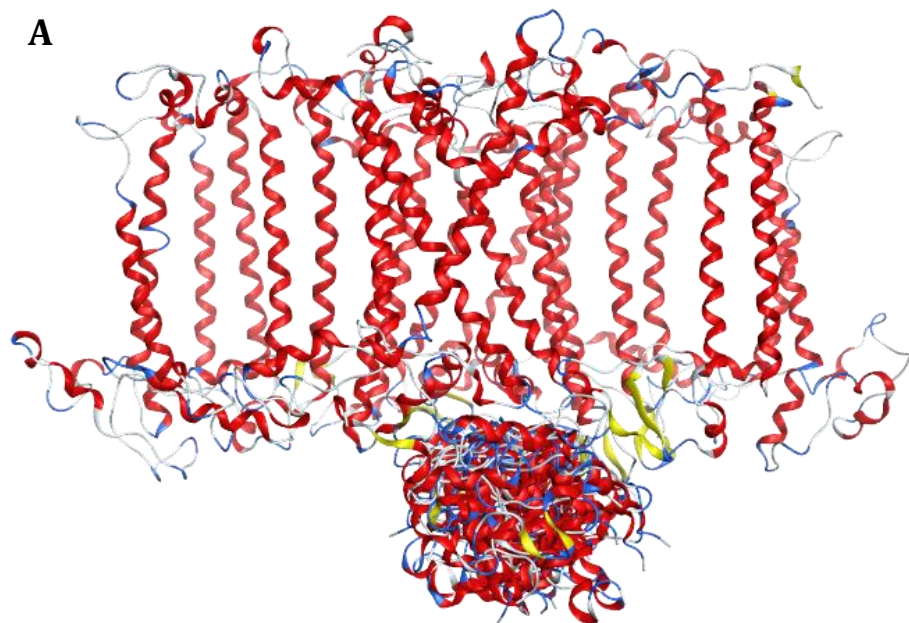
B

	4	8	12	16	20	24	28	32	36	40	44	48	52	56	60
4:model.000....		19.24	8.64	10.30	10.59	18.57	9.49	19.69	17.04	16.51	18.24	22.78	8.92	14.51	18.65
8:model.000....	19.24		18.81	17.50	20.17	10.11	18.57	12.22	8.23	12.91	23.19	11.37	18.90	15.26	27.25
12:model.000....	8.64	18.81		15.70	16.56	18.01	10.00	19.56	18.73	14.27	19.32	24.17	15.93	13.47	19.28
16:model.000....	10.30	17.50	15.70		12.09	18.33	14.99	18.47	11.84	18.16	14.26	20.05	9.19	17.93	18.75
20:model.000....	10.59	20.17	16.56	12.09		20.94	12.45	18.94	17.91	19.48	21.50	19.32	9.87	19.05	22.32
24:model.000....	18.57	10.11	18.01	18.33	20.94		17.75	14.05	13.24	9.47	22.12	17.52	19.13	11.45	25.18
28:model.000....	9.49	18.57	10.00	14.99	12.45	17.75		16.97	19.11	12.45	21.90	20.25	13.60	10.63	22.78
32:model.000....	19.69	12.22	19.56	18.47	18.94	14.05	16.97		14.04	14.34	22.73	11.08	19.58	16.61	26.73
36:model.000....	17.04	8.23	18.73	11.84	17.91	13.24	19.11	14.04		15.95	17.96	13.93	15.25	17.98	23.04
40:model.000....	16.51	12.91	14.27	18.16	19.48	9.47	12.45	14.34	15.95		22.95	17.92	18.11	9.27	25.14
44:model.000....	18.24	23.19	19.32	14.26	21.50	22.12	21.90	22.73	17.96	22.95		28.37	19.70	23.70	9.27
48:model.000....	22.78	11.37	24.17	20.05	19.32	17.52	20.25	11.08	13.93	17.92	28.37		20.62	20.11	32.82
52:model.000....	8.92	18.90	15.93	9.19	9.87	19.13	13.60	19.58	15.25	18.11	19.70	20.62		17.33	20.94
56:model.000....	14.51	15.26	13.47	17.93	19.05	11.45	10.63	16.61	17.98	9.27	23.70	20.11	17.33		25.85
60:model.000....	18.65	27.25	19.28	18.75	22.32	25.18	22.78	26.73	23.04	25.14	9.27	32.82	20.94	25.85	

Figures 4.3 continued

C
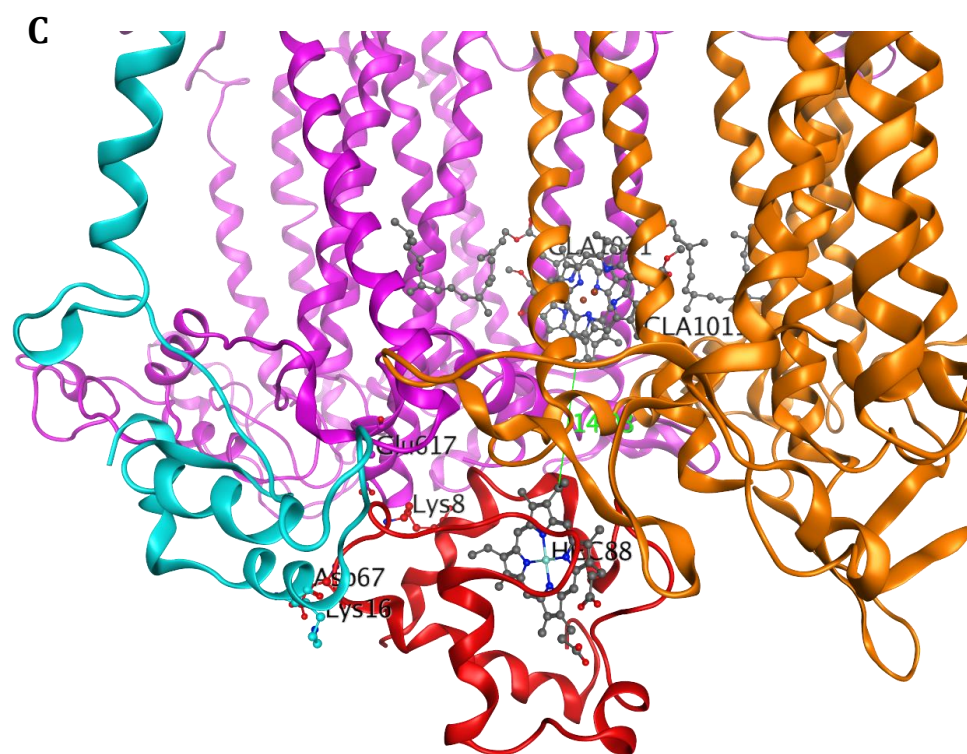
Figures 4.3 continued

Figures 4.4. PSI-Cytc6 models generated using Cluspro using agnostic docking. (A) Distance between Heme and Chlorophyll groups and salt bridges between PSI and Cytc6 in top model. PsaA in purple, PsaB in orange, PsaF in cyan and Cytc6 in red. (B) Trp-655 and Trp-631 present at the interface between Heme (pink) and Chlorophyll (green). (C) Superposition of the top two PSI-Cytc6 models generated by Cluspro using the residue filter (red) and the agnostic docking protocol (green).

A
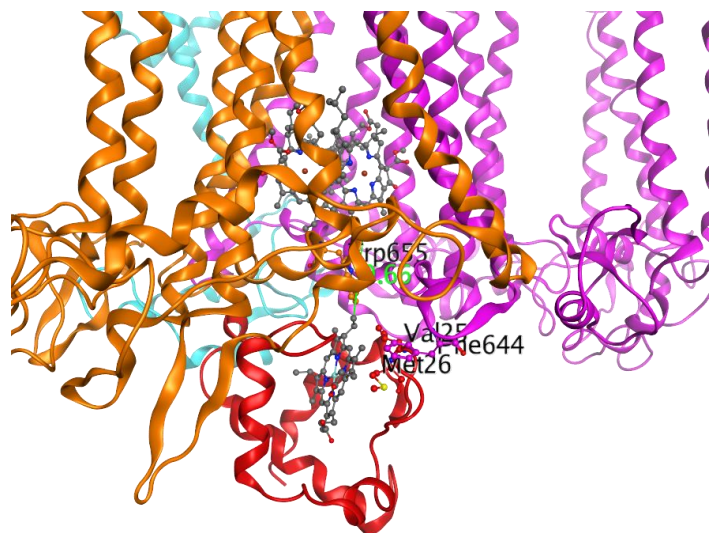
B
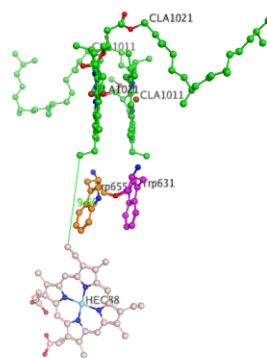
C
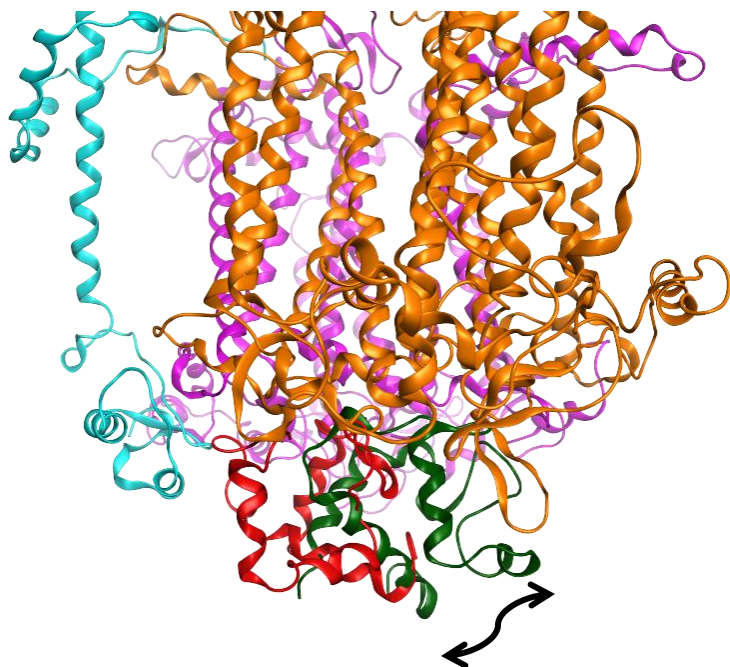
CONCLUSION

The biological levels of organization in living beings form a highly structured hierarchy. This range from a single atom, all the way up to the biosphere. The different subunits- atoms, molecules, amino acid sequences, proteins, cells, tissues, organs, etc., together (as a system) perform the functions necessary for survival and reproduction. Thus, to understand the function at the macroscopic scale that we see around us, it is important to understand the function at the microscopic level first.

There is a direct relationship between the structure of the proteins and the diverse functions that they can perform. The flexible nature and dynamics of macromolecules forms an indispensable link between their structure and function. The interactions that drive most of the biological processes are also influenced by this dynamic nature of biological molecules. The study presented here explores this structure-dynamics-function relationship. Important insights were gained about the dynamics of interactions in proteins, between protein: ligands and protein: protein systems. Throughout this work, different molecular modeling and simulation technologies were employed to tackle specific problems in the field of molecular and structural biophysics.

In **CHAPTER 1**, an emerging non-covalent interaction is identified to potentially play important roles in protein systems in: i) maintaining stability of the folded structure of protein monomers, ii) dimeric interface interactions i.e., forming quaternary assemblies, and iii) active-site interactions in maintaining function. Experimental work like mutagenesis studies can help to validate these results in terms of physical effects of these interactions on the stability of the protein. These interactions have been exploited in field of drug-discovery in developing active-site inhibitors. Hence, in future it is imperative to include these interactions in the description of new forcefields and docking scoring-functions in order to make better predictions both in the fields of molecular modelling and structure-based drug discovery.

CHAPTER 2 focuses on understanding the tissue specific functions of GPRC6A and the structural basis for its binding to distinct ligands, elucidating new systems biology involving coordinated metabolic processes. Future molecular dynamics simulations of the T:GPRC6A models, will likely be able to identify the most stable binding modes of ligands in GPRC6A and allow a detailed structure-based description of binding poses. The study also takes the first step for the development of agonists for GPRC6A that may stimulate insulin production, β-cell proliferation and peripheral tissue sensitivity, and thereby provide a new treatment for Type II diabetes. Conversely, antagonists to GPRC6A might represent new treatments to retard the progression of prostate cancer. .

CHAPTER 3 exemplifies a successful MD simulation and high-throughput virtual screening approach to drug discovery, targeting a protein-protein complex, and shows the importance of explicitly including the flexibility of proteins in making better predictions in drug-discovery. New protein targets for diseases like cancer and diabetes are being discovered rapidly. The same approach can be extended to target these proteins, using even larger compound databases covering millions of compounds. This can increase dramatically the number and diversity of compounds that can be identified, while at the same time providing a more realistic approach that accounts for the dynamic nature of macromolecules and their complexes under physiological conditions.

The work in **CHAPTER 4** allows the atomic-level description of protein-protein interactions and dynamics between electron transport proteins in PS1. Information obtained from the docking models and ongoing MD simulations can be used to guide future experimental studies,

enabling the design of high affinity variants of these proteins to optimize electron transfer and leading to the development of biologically-based solar energy cells for electricity generation.

Overall, the work presented here explores the importance of dynamics in describing both the structure and the function of macromolecules. It becomes indispensable to spend more time and resources in the study of this field, as it forms an important link between the thousands of protein structures solved till now, and their functions. An unwanted change in the structure/interactions/dynamics can lead to a change in the function, leading to diseases. At the same time, this dynamic nature of biological molecules can be exploited for making better drugs and improving the efficiency of renewable sources of energy.

VITA

Karan Kapoor was born on October 11, 1985, in Amritsar, India. He obtained his Bachelor of Engineering degree in Biotechnology from Manipal Institute of Technology, India in 2009. He worked as a Technical Assistant and Systems Engineer at Dell India Inc. and Tata Consultancy Services respectively, from 2009 to 2010. He was accepted to the Graduate School of Genome Science and Technology at the University of Tennessee in 2010, and subsequently joined the lab of Dr. Jerome Baudry in 2011. He received his PhD in Life Sciences with an interdisciplinary minor in Computer Science in 2015.