

UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

The Computational Microscope of Biomolecules: Molecular Recognition and Protein Functional Dynamics Study Through Molecular Dynamics and Brownian Dynamics Simulations

Permalink

<https://escholarship.org/uc/item/3b83611j>

Author

Wu, Kingsley Yu

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

The Computational Microscope of Biomolecules: Molecular Recognition and Protein
Functional Dynamics Study Through Molecular Dynamics and Brownian Dynamics
Simulations

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

by

Kingsley Yu Wu

March 2023

Dissertation Committee:

Dr. Chia-en A. Chang, Chairperson

Dr. Gregory J.O. Beran

Dr. Ryan Julian

Copyright by
Kingsley Yu Wu
2023

The Dissertation of Kingsley Yu Wu is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

First and foremost, I am extremely grateful to my Ph.D. advisor, Prof. Chia-en A. Chang for her invaluable research advice, continuous support, and patience during my Ph.D. study. Her personal persistent characteristics inspired me throughout my six-year of research and supported my progression as a graduate student. Thank you!

Many thanks also go to my committee members, Prof. Gregory J.O. Beran, Prof. Ryan Julian, and Prof. De-en Jiang (now in Vanderbilt U.) for their suggestive and enlightening questions during my qualify examinations and oral presentations. I would like to thank our collaborators Prof. Jiayu Liao from Bioengineering at UCR, Prof. Leonard Mueller from Chemistry at UCR, Dr. Mohamed Hamed and Dr. Yate-Ching Yuan from City of Hope for their valuable collaboration work and feedback during our project meetings. Thank you!

I would like to thank my supervisor Dr. Kai Zhu (now in Beigene Inc), senior scientists Dr. Huan Rui, Dr. Shu-ching Ou, and director Dr. Yax Sun from Amgen Inc for helping me during my internship in summer 2021. Thank you!

Also, I would like to thank my group members who helped me with my research through the time we shared. I would like to especially thank Dr. Zhiye Tang, Dr. Wanli You, Dr. Yuliana K. Bosken, Dr. Tim Cholko, and Dr. Shivansh Kaushik, post-docs, Dr. Si-Han Chen, Dr. Talant Ruzmetov, and Dr. Rahman Homaidur in my group who inspired me by many of their work and discussions we had. I also want to acknowledge my amazing friends from inside and outside of UCR for being with me since the first day we met. Thank you!

Finally, I want to take the opportunity to extend my greatest appreciation to the University of California, Riverside, Graduate Program in Chemistry, and Graduate Student Association for providing me the opportunity to carry out my research, supporting me for conference travel, and the wonderful time in Riverside. Also, I thank Chinese American Engineers and Scientists Association of Southern California (CESASC) for awarding me 2022 CESASC scholarship.

To everyone I mentioned above and the ones I forgot. Thank you for supporting me, challenging me, and helping me grow on this journey!

The text of this dissertation, in full, is a reprint of the materials used as it appears in the following publications:

Chapter 2: Wu, Kingsley Y., David Doan, Marco Medrano, and Chia-en A. Chang. "Modeling structural interconversion in Alzheimers' amyloid beta peptide with classical and intrinsically disordered protein force fields." *Journal of Biomolecular Structure and Dynamics* (2021): 1-18.

Chapter 5: Cholko, Timothy, Shivansh Kaushik, Kingsley Y. Wu, Ruben Montes, and Chia-en A. Chang. "GeomBD3: Brownian Dynamics Simulation Software for Biological and Engineered Systems." *Journal of Chemical Information and Modeling* (2022).

Chapter 2 includes the article without being rewritten. Chapter 5 includes my contributions (manuscript writing, data analysis, code testing, *etc.*) to the article.

Dedication

I dedicate this dissertation to my family members and my lover Lain Lee. They endowed me with the tools to think and care for myself and taught me to be curious and analytical.

Without those abilities I could not have made it through this.

Thank you for your constant love and support.

ABSTRACT OF THE DISSERTATION

The Computational Microscope of Biomolecules: Molecular Recognition and Protein Functional Dynamics Study Through Molecular Dynamics and Brownian Dynamics Simulations

by

Kingsley Yu Wu

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, March 2023
Dr. Chia-en A. Chang, Chairperson

Molecular recognition and functional dynamics are fundamental in chemical processes, often seen in those chemical and biological systems. The molecular recognition process guides molecular species with preferential binding and unbinding events via physical differentiation and complementarity. Many experimental techniques show promising results, yet limitations in studying such fundamental process at a molecular level. With the advanced and rapid development of computational technology and mathematical algorithms, molecular modeling has become an effective and powerful tool in studying molecular recognition processes, such as investigating the interactions between molecular species and understanding the biological dynamic processes, making it significant and indispensable contributions to modern medicinal chemistry, biology, and drug development. The state-of-the-art techniques of computational chemistry and molecular modeling can be employed to study a wide range of chemical and biological systems of interest, which allow people to understand the structural-functional-relation in detail at a microscopic view, and gain chemical/biological information which deem to be challenging

by experimental measurements. The work in this dissertation focuses on modeling the recognition processes and protein functional dynamics of biomolecules in different biological systems using multiple computational approaches, such as homology modeling, molecular docking including protein-protein and protein-ligand docking, structural-based alignment and superposition, molecular dynamics (MD) and Brownian dynamics (BD) simulations, enhanced samplings, and MM/GB(PB)SA-based energy calculations. We also apply various post-analysis tools to understand the systems of interest's dynamic nature and motions in details in this dissertation.

A strong focus in two projects describe here use MD simulations to demonstrate physical and chemical principles that govern ligand's effectiveness in the complex and understand the ligand binding affinity. One project focuses on applying MD simulations to investigate and demonstrate the protein's functional dynamics, which deem to link to its characteristic features in the disease. In addition, diffusional association and dissociation binding events between receptor and substrate are also studied in this dissertation project. We apply BD simulations to model such events with over microseconds to millisecond timescale. A computational efficient BD simulation software developed in our group is introduced and applied to study different structural features that affects the substrate binding efficiency.

Table of Contents

Chapter 1: Introduction

1.1 Overview of computational chemistry and molecular modeling.....	1
1.2 Molecular Dynamics Simulations.....	2
1.2.1 Molecular Mechanics.....	3
1.2.2 Molecular Dynamics Algorithm.....	6
1.2.3 Solvent and Ion Models.....	9
1.2.4 Molecular Dynamics Simulation Protocols.....	11
1.2.4.1 Preparation of Initial Structure.....	12
1.2.4.2 Minimization and Solvation.....	13
1.2.4.3 Equilibrium and Production run.....	13
1.3 Enhanced Sampling.....	14
1.4 Brownian Dynamics Simulations.....	16
1.4.1 Brownian Dynamics Simulation Protocol.....	16
1.5 Molecular Docking.....	17
1.6 Post-Analysis of Simulation Data.....	18
1.6.1 Root-Mean-Square Deviation.....	19
1.6.2 Root-Mean-Square Fluctuation.....	19
1.6.3 Radius of Gyration.....	20
1.6.4 Configurational Entropy Calculation.....	20
1.6.5 Principal Component Analysis.....	21
1.6.6 MM/PBSA and MM/GBSA Energy Calculations.....	22

1.6.7 Diffusion Coefficient.....	24
1.7 References.....	26

Chapter 2: Modeling Structural Interconversion in Alzheimers' Amyloid Beta

Peptide With Classical and Intrinsically Disordered Protein Force Fields

2.1 Abstract.....	35
2.2 Introduction.....	36
2.3 Methods and Material.....	40
2.4 Results and Discussion.....	44
2.5 Conclusions.....	69
2.6 References.....	72
2.7 Acknowledgements.....	82

Chapter 3: Functional Dynamics in PROTACs-Induced Protein Degradation

Machinery

3.1 Abstract.....	83
3.2 Introduction.....	83
3.3 Methods and Material.....	87
3.4 Results and Discussion.....	96
3.5 Conclusions.....	108
3.6 References	110

Chapter 4: CACNA1C is A Biomarker and Therapeutic Target for Patients With Ovarian Cancer

4.1 Abstract.....	115
4.2 Introduction.....	116
4.3 Methods and Material.....	117
4.4 Results and Discussion.....	125
4.5 Conclusions.....	138
4.6 References.....	140
4.7 Acknowledgements.....	143

Chapter 5: GeomBD3: Brownian Dynamics Simulation Software for Biological and Engineered Systems

5.1 Abstract.....	144
5.2 Introduction.....	144
5.3 Methods.....	147
5.4 Example Applications.....	158
5.5 Conclusions.....	168
5.6 References.....	170
5.7 Acknowledgements.....	174

Chapter 6: Multiscale Modelling of Enzymatic Efficiency in Multienzyme Complex Contributed From Inter-Enzyme Distance, Orientation and Side Reactions: Insights From All-Atom Molecular Dynamics and Brownian Dynamics Simulations

6.1 Abstract.....	175
-------------------	-----

6.2 Introduction.....	176
6.3 Methods and Materials.....	180
6.4 Results and Discussion.....	186
6.5 Conclusions.....	201
6.6 References.....	203
Chapter 7: Conclusions and Future work.....	211
7.1 Conclusion.....	211
7.2 Future work.....	212
7.3 References.....	215

List of Tables

Chapter 2:

Table 2.1 Summary of molecular dynamics simulations

Table 2.2 Root mean squared deviation differences between simulated and experimental NMR observables in A β 42 ^{α -helix} and A β 42 ^{β -strand}

Table 2.3 The average Rg values calculated from all the cMD working systems along with a comparison with other computational studies. The unit of Rg is Å

Chapter 3:

Table 3.1 Accessible lysine residues identified in the CRL4A E3 ligase complexes for each PROTAC

Table 3.2 MD simulations setup for the CRBN-dBETx-BRD4^{BD1} complexes, degradation machinery complex C4 from category A, degradation machinery complex C12 and from category B

Table 3.3 Binding scores predicted by MOE protein protein docking for the CRBN-dBETx-BRD4^{BD1} complexes (*calculated in kcal/mol)

Table 3.4 Summary of the predicted CRBN-dBETx-BRD4^{BD1} ensembles

Chapter 4:

Table 4.1 Binding modes with relative scorings of NIFE

Chapter 5:

Table 5.1 Hybridization time, hybridizing probe site, fraction of 2D hybridization (γ), and DNA-SAM interaction energies on the two different SAMs used for probe cluster simulations

Table 5.2 Average association time $\langle t_{\text{assoc}} \rangle$, probability of direct binding (p_{direc}), and surface dwell time $\langle t_{\text{dwell}} \rangle$ of CdRP in the crowding and diluted models

Chapter 6:

Table 6.1 List of all aMD simulations for substrate 1,2-carboxyphenylamino-1-deoxy-d-ribose 5-phosphate (CdRP) dissociation from PRAI and association to IGPS. Red: The substrate successfully dissociated from the pocket of PRAI. Blue: The substrate dissociated

from the PRAI and associated to the IGPS successfully. Black: The substrate did not successfully dissociate from the PRIA pocket. (*a), different boost potentials

Table 6.2 Calculation of diffusion coefficient for different BD simulations

Table 6.3 Medium Association time (T_{asso}), Binder Percentage in IGPS and in the near Competitors for Different Inter-enzyme Distance Setups in three Simulation Conditions

Table 6.4 Direct Binding Percentage and Dwell time for Different Inter-Enzyme Distance Setups in Three Simulation Conditions

Table 6.5 Medium Association time (T_{asso}) and Dwell time for Different Orientation Setups in Diluted Simulation Condition

List of Figures

Chapter 1:

Figure 1.1 Molecular mechanic force field used for classical MD simulations

Figure 1.2 Molecular dynamics basic algorithm

Figure 1.3 Standard MD simulation protocol

Chapter 2:

Figure 2.1 Sequence of 42 residues amyloid beta peptide (A β 42), NMR-resolved A β 42 ^{α -helix} and A β 42 ^{β -strand}

Figure 2.2 Root mean squared deviation analysis for A β 42 of two monomers

Figure 2.3 Conformers of clustering number with respect to the cumulative time (100 ns window) under two force fields for two monomers

Figure 2.4 Root mean squared fluctuation analysis of the combined 3 μ s in cMD for A β 42 peptide under *ff14SB* and *ff14IDPSFF* of two monomers

Figure 2.5 Comparison of chemical shifts for HN, C α , C β , and N between NMR experimental data and MD prediction

Figure 2.6 Probability distribution on the 2D space of root mean squared deviation and radius of gyration. Data were combined from three MD runs for each model

Figure 2.7 Transformation of secondary structures for A β 42 ^{α -helix}

Figure 2.8 Percentage of individual residue with different secondary structures for the A β 42 ^{α -helix} *ff14SB* and *ff14IDPSFF*

Figure 2.9 Transformation of secondary structures for A β 42 ^{β -strand}

Figure 2.10 Probability of intrapeptide contact map

Figure 2.11 Cluster index versus the percentage of occupied conformers for all evaluated working systems

Figure 2.12 Representative conformers of the top 8 clusters and their occupied percentage derived from simulations with the two force fields *ff14SB* and *ff14IDPSFF*

Figure 2.13 Distributions of structures with given solvent accessible surface areas (SASAs) for the working systems

Figure 2.14 Salt bridge distance measurements as a function of simulation timeline

Chapter 3:

Figure 3.1 PROTAC-mediated degradation machinery complex.

Figure 3.2 Cartoon representations of A. BRD4^{BD1} (PDB ID 3MXF), lysine residues are shown in black licorice; B. CRBN E3 ligase (PDB ID 4CI3); C. BRD4^{BD1} sequence representation

Figure 3.3 Flowchart of CRL4A E3 ligase complexes construction. A

Figure 3.4 CRL4A E3 ligase complexes with 12 DDB1 conformations.

Figure 3.5 Crystal structure alignments of reported PROTACs (dBET1: PDB ID 4ZC9, dBET6: PDB ID 6BOY, dBET23: PDB ID 6BN7, dBET57: PDB ID 6BNB, and dBET70: PDB ID 6BN9)

Figure 3.6 Overlapping diagram representation of common CRBN-dBET_x-BRD4^{BD1} ensembles found in each degradation machinery complex

Figure 3.7 CRBN-dBET_x-BRD4^{BD1} ensembles generated from protein-protein docking. All ensembles were superimposed based on CRBN (cyan, thick, ribbon), BRD4^{BD1} poses were shown as thin, ribbon, rainbow color)

Figure 3.8 CRBN-dBET_x-BRD4^{BD1} ensembles align with CRL4A E3 ligase complex to make degradation machinery complex

Figure 3.9 Superpositions of predicted CRBN-dBET-BRD4^{BD1} with reported crystal structures (Black)

Figure 3.10. Snapshots of three degradation machinery complexes DDB1/CUL4A/NEDD8/Rbx1/E2/Ub/CRBN-dBET_x-BRD4^{BD1} from MD simulations

Figure 3.11. PCA major motion modes of predicted ternary complexes

Chapter 4:

Figure 4.1 Structure of the constructed hCaV1.2 channel

Figure 4.2 Ligand binding sites in the structure of the constructed hCaV1.2 channel

Figure 4.3 CACNA1C mRNA gene expression mature strands was analyzed as reads-per-million-per-reads; data were downloaded from TCGA using TCGA Assembler

Figure 4.4 IHC of hCaV1.2 in paired OC sections

Figure 4.5 Cytotoxic effect of NIFE on human ovarian cancer cell lines (A2780R, OVCAR8 and SKOV3)

Figure 4.6 NIFE binds to the fenestration enclosed by domain III (cyan) and IV (blue) of hCaV1.2

Figure 4.7 Snapshot of the constructed hCaV1.2 ion channel embedded in a POPC bilayer with water molecules, sodium, and chloride ions

Figure 4.8 Comparison of the RMSD over 200 ns MD simulations between the Apoprotein and NIFE bound structures

Figure 4.9 Snapshot of residues around the NIFE binding pocket in apoprotein structure and NIFE bound structure

Figure 4.10 Clustering conformations of residues surrounding the ligand binding pocket

Figure 4.11 The permeation path of hCaV1.2 apoprotein (A) and NIFE bound (B) structures calculated by HOLE program

Chapter 5:

Figure 5.1 Diagram of GeomBD3 workflow

Figure 5.2 HIV protease is placed 20 Å above the SAM parallel to the yz-plane and 2D Structure of xk263 ligand

Figure 5.3 GeomBD3 molecular model of a DNA biosensor with multiple probes (blue strands) and the target DNA (yellow strand) which binds to the probes

Figure 5.4 (A) Model with crowders (magenta). (B) Model with diluted environment. The E1 (PRAI, red) and E2 (IGPS, blue) are biocatalytic cascade enzymes in the tryptophan biosynthesis pathway. Substrate (CdRP) was produced in the PRAI and catalyzed in the IGPS to make IGP. The crowders were placed randomly in the sphere, in which the radius of the sphere was set to 341 Å corresponding to 10.0 μ M concentration

Figure 5.5 Acetylcholinesterase with common ligand association sites found by *ProbdX* indicated by pink dots. Higher dot density indicates higher probability of association

Chapter 6:

Figure 6.1 A. *Escherichia coli* (*E.coli*), mesophile, bifunctional TrpCF fusion protein (PDB 1PII, *trpCF*), phosphoribosyl anthranilate isomerase (PRAI), blue, Indoleglycerol phosphate synthase (IGPS), red. B. *Thermotoga maritima* (*T.mari*), thermophile, monomeric IGPS (PDB 1J5T, *trpC*). C. *Thermotoga maritima* (*T.mari*), thermophile, monomeric PRAI (PDB 1LBM, *trpF*), 1,2-carboxyphenylamino-1-deoxy-d-ribulose 5-phosphate (CdRP) in the catalytic site of PRAI

Figure 6.2 Schematic of simulation setups

Figure 6.3 2D chemical structure of 1,2-carboxyphenylamino-1-deoxy-d-ribulose 5-phosphate (CdRP), and Indoleglycerol phosphate synthase (IGPS) with K55 shown in vdW representation

Figure 6.4 Successful CdRP dissociations from aMD simulations with the final locations on the surface of the PRAI:IGPS fusion complex. One successful CdRP association to the IGPS pocket is shown in blue

Figure 6.5 The association pathway of CdRP in the PRAI:IGPS fusion complex in a 400 ns aMD.

Figure 6.6 Binding time distribution for PRAI and IGPS enzymes at the different inter-enzyme distances

Figure 6.7 Substrate binding map on the surface of the enzymes

Figure 6.8 Binding time distribution for PRAI and IGPS enzymes at the different binding pocket orientation

Figure 6.9 Screenshot of non-specific interactions between substrate, enzymes and crowders.

Chapter 1 Introduction

1.1 Overview of Computational Chemistry and Molecular Modeling

Molecular recognition and functional dynamics in biomolecules underlie majority, if not all, of the biological functions and processes. Since introduced in early 1980s, the phrase “molecular recognition” has driven much modern chemical and biological research, and led the prosperity of many new technologies in enzymology, genetics, cell biology, bioengineering and drug discovery.^{1,2} Timely binding of the suitable biomolecular partners and proper protein functional dynamics are central for many biomolecules in many cellular activities, including signal transduction, protein-protein interactions, enzyme catalysis, and chemical transport.³ In the past few decades, computational studies of biological processes, such as protein dynamics and conformational change, protein-ligand binding, change of association/dissociation rates, in chemical and biological systems have taken the advantage of the development in computational power and resources. Hence, computational modeling and simulation of molecular recognition processes and protein dynamics have taken an integral predictive and explanatory role in material science, enzyme engineering, and pharmaceutical drug design.

This dissertation project aims to use modern computational chemistry theory and various modeling approaches to broaden the fundamental understanding of chemical and biological processes involved in the thermodynamics, kinetics, and dynamics of molecular entities. The following sections in this chapter detail the theories, simulation algorithms, simulation techniques and protocols, and various post analysis methods. According to the simulation

techniques and protocols, the subsequent chapters are organized for further discussion with the following arrangement: Chapter 2, 3, and 4 are mainly MD-based simulations, while chapter 5 and 6 focused on BD-based simulations. The enhanced sampling method is applied, and the details of functional dynamics and conformational changes of A β 42 peptide are investigated in chapter 2. The protein-ligand binding interactions, the residues' flexibility upon ligand binding, and their protein dynamics are investigated in PROTAC and in hCav1.2 calcium channel projects in chapter 3 and 4, respectively. In addition, the association process of enzyme and substrate and the substrate diffusion, which requires large simulation timescale are also discovered. Specifically, a computational efficient Brownian dynamics simulation package is developed and introduced in chapter 5 to determine substrate diffusion and elucidate various factors that influence the successful encounter of ligands to protein targets. Then, an enzyme system TrpCF utilizes this BD-based simulation package to address many questions in chapter 6.

1.2 Molecular Dynamics Simulations

Since first introduced over 40 years ago,⁴ Molecular dynamics (MD) simulations have advanced from a simple method to simulate movements of several thousands of atoms to study biomolecular systems, including peptides,⁵⁻⁸ proteins,⁹⁻¹⁴ nucleotides,¹⁵⁻¹⁷ lipids (membrane included)¹⁸⁻²¹ and viruses.²²⁻²⁵ Nowadays, with the increasing computational power and advanced algorithm architecture, many large biomolecular systems with atomic details can be simulated by MD simulations with timescale less than a few microseconds for system sizes of 10^5 - 10^6 atoms.^{18,26} The theoretical framework of molecular dynamics is

highly derived by the principles of statistical mechanics, with which one can calculate system properties from simulation data.²⁷⁻²⁹ In the statistical mechanics of a classical system, the macroscopic property of a system depends on its microstates, that is, the states of all its smallest constituent parts. In simulations, the generated data are instantaneous snapshots of a system. These snapshots referred as ensembles of system configurations representing energies contributed from atoms, which depend on atoms' positions r and momentum p . They are required to determine ensemble's microscopic information, which in turn, connecting the macroscopic information of the system. Experimentally, the macroscopic information of a system can be obtained from a measurement taken over a period, so the value is an average over the time of the measurement.

Molecular dynamics simulation approximates molecular movements in the context of N-body particles (atoms, for example) governed by the Newton's law. Forces exerting on each atom are computed from the potential energy function derivation, defined by molecular mechanic force fields detailed in the following section. At each time step, usually in femtosecond, atom positions are updated, and integrated stepwise resulting in the trajectory of that system over a certain period of time, usually in nanosecond to microsecond timescale.

1.2.1 Molecular Mechanics

The potential energy function describes the atomic and molecular interactions. It is often called "force field"³⁰ and represented by the equilibrium bond length or angle, partial atomic charge, and van der Waals atomic radii. (Figure 1.1 and Eq. 1.1) It is the heart of

all chemical simulations and is empirically parameterized to accurately describe the structure and energies of a molecular system. As such, their accuracy has immense influence on the ability of simulations to reproduce and predict results of real chemical systems.³¹⁻³³

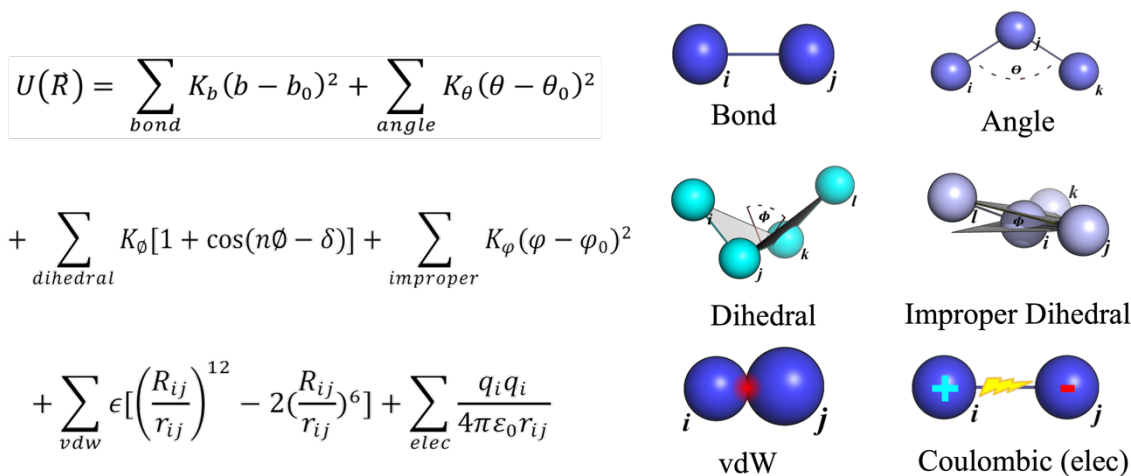


Figure 1.1 Molecular mechanic force field used for classical MD simulations (CHARMM force field is described here). Left, energy functions used to derive forces for atom movements. Right, bonded (bond, angle, torsion) and non-bonded (vdW and electrostatic) atoms illustration.

$$U(R) = (U_{bond} + U_{angle} + U_{dihedral} + U_{improper}) + (U_{elet} + U_{vdw}) \quad \text{Eq. 1.1}$$

Each energy term can be further represented by the following equations:

$$U_{bond} = \sum K_b(b - b_0)^2 \quad \text{Eq. 1.2}$$

where K_b is a force constant and b and b_0 are the current and equilibrium bond lengths.

Similarly, bond angle terms represent the energy as the angle between three sequentially bonded atoms changes,

$$U_{angle} = \sum K_\theta(\theta - \theta_0)^2 \quad \text{Eq. 1.3}$$

where K_θ is a force constant and θ and θ_0 are the current and equilibrium bond angles.

Dihedral angles between a set of three bonded atom pairs are critical for describing the potential energy of differing molecular geometries. A simple example is the difference between a *cis* or *trans* conformation of 2-butanol. The dihedral angle energy for any set of three bonded atom pairs is calculated as

$$U_{dihedral} = \sum K_\phi [1 + \cos(n\phi - \delta)] \quad \text{Eq. 1.4}$$

where K_ϕ is the barrier height, n gives the periodicity of the function in a 360° rotation, ϕ is the measured angle and δ is the phase shift angle.

Another form of dihedral, called improper dihedral, is used to describe the energy related to planarity (and deviations thereof) of a set of four bonded atoms,

$$U_{improper} = \sum K_\phi (\phi - \phi_0)^2 \quad \text{Eq. 1.5}$$

where K_ϕ is the force constant of the improper angle, ϕ is the improper in radians and ϕ_0 is the equilibrium improper.

In addition to these four potential energy terms representing bonded atoms, two non-bonded potential energy terms are considered:

$$U_{vdW} = \sum \epsilon_{ij} \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}}{r_{ij}} \right)^6 \right] \quad \text{Eq. 1.6}$$

where r_{ij} is the distance between two atoms, ϵ_{ij} is the depth of the potential well, and R_{ij} is the equilibrium position of two atoms. The exponents 12 and 6 denote the van der Waals force in the form of a 12-6 Lennard-Jones potential.

$$U_{elec} = \sum \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad \text{Eq. 1.7}$$

where q_i and q_j represent the charges of interacting atoms i and j, r_{ij} is the distance between two atoms, ϵ_0 is the electric permittivity of free space.

Together these six potential energy terms having bonded and non-bonded interactions, make up a full molecular mechanic force field that computes the potential energy of simulated system. Depending on empirical parametrizations of these energy terms, many simulation packages having slightly modifications in the force field are developed. They are AMBER,^{34,35} CHARMM,³⁶ GROMOS,³⁷ NAMD,³⁸ OPLS,³⁹ and many others. It is worth to mention that parameters in these force fields are modified and optimized according to the selected input protein structures, short peptide fragments, and lipids and DNAs. Different simulated performances can be seen from different selected force fields. Hence, chapter 2 investigates and shows that distinct protein dynamics is relied on different force fields that are applied.

1.2.2 Molecular Dynamics Algorithm

Molecular dynamics simulates behaviors as time evolves of a molecular system by solving Newton's equation of motion, given a potential energy function and atoms' coordinates. The algorithm in MD simulations starts with Newton's second law

$$F_i = m_i a_i \quad \text{Eq. 1.8}$$

where F_i is the force acting on atom i, m is the mass of the atom, and a is the acceleration of atom i. The force on atom can be related to gradient of the potential energy of atom i.

$$F_i = -\frac{dU}{dr} \quad \text{Eq. 1.9}$$

and one can relate these equations as follow

$$-\frac{dU}{dr} = m_i a_i \quad \text{Eq. 1.10}$$

Then, the derivatives of the acceleration a and velocity v of an atom with respect to time t and position r can be described as

$$a = \frac{dv}{dt} = \frac{d^2r}{dt^2} \quad \text{Eq. 1.11}$$

$$v = \frac{dr}{dt} \quad \text{Eq. 1.12}$$

Integration gives rise to the following, assuming constant acceleration over the time interval

$$v = \int a dt = at + v \quad \text{Eq. 1.13}$$

$$r_{i+1} = \int v dt = \int (at + v) dt = \frac{at^2}{2} + v_i t + r_i \quad \text{Eq. 1.14}$$

The system with propagating movement forward in time is then finding the acceleration of atoms from the difference in potential energy of the current configuration and the previous one, velocity of atom is updated, and atom moves to its next position $r_{t+\Delta t}$. Repeat the processes over all atoms in the systems for as many steps as needed. Mathematical expression is to combine equations 1.9 and 1.14 to get

$$r_{t+\Delta t} = -\frac{1}{2m_i} \frac{dU}{dr} t^2 + v_t t + r_t \quad \text{Eq. 1.15}$$

Two commonly used mathematical algorithms, Verlet and Leap-frog, are briefly touched without further discussion. They are described as in equations 1.16-1.18 for Verlet and 1.19-1.21 for Leap-frog, respectively,

$$r(t + dt) = r(t) + v(t)dt + \frac{1}{2}a(t)dt^2 \quad \text{Eq. 1.16}$$

$$r(t - dt) = r(t) - v(t)dt + \frac{1}{2}a(t)dt^2 \quad \text{Eq. 1.17}$$

Combining the above two to get

$$r(t + dt) = 2r(t) - r(t - dt) + a(t)dt^2 \quad \text{Eq. 1.18}$$

$$v\left(t + \frac{1}{2}dt\right) = v\left(t - \frac{1}{2}dt\right) + a(t)dt \quad \text{Eq. 1.19}$$

$$r(t + dt) = r(t) + v\left(t + \frac{1}{2}dt\right)dt \quad \text{Eq. 1.20}$$

The velocity at time t can be approximated by the relationship.

$$v(t) = \frac{1}{2}\left[v\left(t - \frac{1}{2}dt\right) + v\left(t + \frac{1}{2}dt\right)\right] \quad \text{Eq. 1.21}$$

Of note, the time step must be determined with care to avoid non-physical dynamics. For example, a common MD simulation time step is set between 1 to 4 femtoseconds because a bond stretch involving hydrogen happens on the order of 10^{-14} . In our cases, 2 femtosecond is commonly used. Figure 1.2 summaries the molecular mechanic basic algorithm.

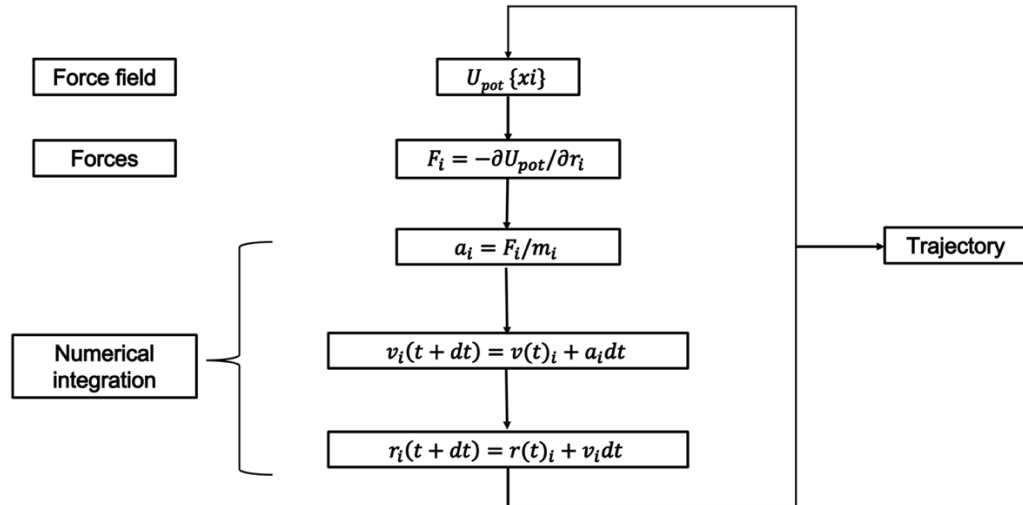


Figure 1.2 Molecular mechanic basic algorithm. Upot, potential energy; t, simulation timestep; dt, time iteration; v, velocity; r, atomic coordinate; F, forces; a, acceleration; m, atomic mass; subscript i, each cartesian coordinate of the n atom being simulated.

1.2.3 Solvent and Ion Models

As many chemical and biological processes occur in solution, it is therefore common and realistic to model the simulations with solvent and counter ions to ensure the modeling accuracy. Water as solvent is typical in MD simulations unless stated otherwise. Appropriate counter ions such as Na⁺ and Cl⁻ are commonly selected to neutralize the whole system complex. The solvation can be either implicit or explicit. Common explicit water models include TIP3P, TIP4P, TIP5P and SPC(/E).⁴⁰⁻⁴² The explicit water model is considered more accurate in molecular mechanical model of water. However, the large number of water molecules that are necessary to solvate a molecular system dramatically slows down the calculation of the potential energy function. An alternative selection in water modeling is implicit solvent model, which the solvent is replaced by a continuous medium. Two popular implicit solvent models are Poisson-Boltzmann (PB) and Generalized Born (GB) models.

The PB method calculates the electrostatic potential in a solvated system by modeling the water as a high dielectric medium.^{43,44} The effects of ions in solution are defined in PB equation

$$\nabla \cdot [\epsilon(r)\nabla\phi(r)] = -4\pi\rho(r) - 4\pi\lambda(r) \sum_i z_i c_i e^{\left(\frac{-z_i\phi(r)}{k_b T}\right)} \quad \text{Eq. 1.22}$$

where $\epsilon(r)$ is the dielectric constant, $\phi(r)$ is the electrostatic potential, $\rho(r)$ is the solute charge, $\lambda(r)$ is the Stern layer masking function, k_b is Boltzmann's constant, T is the temperature, z_i is the charge of ion type i , and c_i is the number density of ion type i in solution far from the solute.⁴⁵ Summation is performed over all ion types (i.e. monovalent

cations, divalent anions). Formulations of the PB equation can also be solved to calculate the solvation free energy of a molecular system. Due to the computational cost of solving the PB equation, it is commonly used to investigate energies of static structures of a molecular model with post-analysis such as MMPBSA calculation detailed in section 1.1.6, rather than as a part of a dynamics simulation.

The simplest type of GB or the typical GB model used in research is approximated as the following equation

$$\Delta G_{solv} = -\frac{1}{2} \sum_{ij} \frac{q_i q_j}{f_{GB}} \left(1 - \frac{e^{(-\kappa f_{GB})}}{\epsilon} \right) \quad \text{Eq. 1.23}$$

where q_i and q_j are the charges on two interaction particles, ϵ is the solvent dielectric, and κ is Debye-Huckel electrostatic salt screening parameter. The choice of function f_{GB} takes the following

$$f_{GB} = \sqrt{l_{ij} + R_i R_j e^{\left(\frac{-l_{ij}^2}{4R_i R_j} \right)}} \quad \text{Eq. 1.24}$$

where l_{ij} is the distance between two interacting particles, R_i and R_j are the effective Born radii of the atoms. The effective Born radii describes the degree to which a particle is buried within its parent molecule. For an isolated ionic particle, this is equal to its van der Waals radius. The calculation of these radii is imperative to the accuracy of the model.^{46,47} GB is a commonly used electrostatic water model in dynamics simulations due to its relative simplicity and computational practicality.

1.2.4 Molecular Dynamics Simulation Protocols

Simulating an all-atom molecular dynamics system requires careful preparation of the structure, such as protonation states of the protein and ligand, because the potential energy function is very sensitive to small changes in the coordinates of the system. Thus, steps must be taken before a molecular dynamics simulation is run to ensure a stable system at equilibrium. The common protocol for performing biosystem of MD simulations is listed in Figure 1.3 and detailed in the following sections.

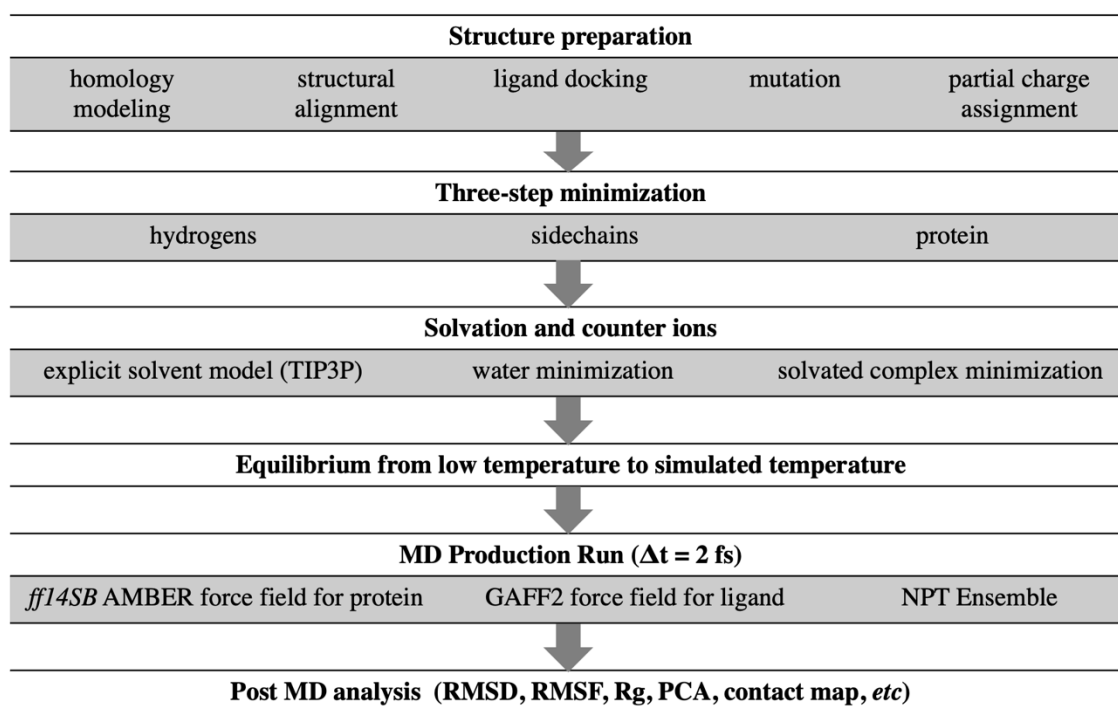


Figure 1.3 Standard MD simulation protocol.

1.2.4.1 Preparation of Initial Structure

To begin with MD simulations of biomolecules, atomistic-resolved structures can be obtained from protein data bank.⁴⁸ X-ray crystallography, NMR, and Cryo-EM techniques are the most common and powerful experimental tools to resolve biomolecular structures. Recently, the advances in machine learning and artificial intelligence have also promoted and predicted structures with high accuracy and precision, such as AlphaFold.⁴⁹ Structures with high resolution and are intact are always preferred, but homology modeling is an alternative and supplemental method to construct a system,⁵⁰ based on the sequence identity and structure similarity with other known structures. If some regions or segments in a system are missed or not well-resolved, one can conventionally “borrow” regions and segments from other structures to construct the protein of interest after both structures are well aligned and superimposed. In case of small molecules or drug compounds, coordinates can be obtained from PubChem⁵¹ and ZINC⁵² databases. If small molecules or compounds are not available from the websites, software and packages, such as Avogadro,⁵³ VEGA ZZ,⁵⁴ and MOE⁵⁵ are options to build, edit, and visualize molecules.

Once the initial coordinates of the protein and ligand are set, the protonation states of the biomolecules can be assigned by manually adding or removing hydrogens. Because the pKa value of each residue is sensitive to hydrogen bonding, desolvation effect, and electrostatic interactions surrounding its local environment. One can calculate the true pKa values of residues using different programs, such as PROPKA3,⁵⁶ DEPTH,⁵⁷ H++,⁵⁸ MCCE2,⁵⁹ and UHBD.⁶⁰

1.2.4.2 Minimization and Solvation

Energy minimization is a necessary step in the MD simulations because error can be introduced during the system preparation step or with inappropriate initial structures. In particular, coordinates of a system downloaded from the PDB website may have small deviations, and energy minimization alleviates by removing steric clashes or close contacts within the atoms. This voids the energy function calculation to “blow up”. Then, restraints on a part of protein atoms while maintain important features of the original structure to gradually bring the system to the target condition. If a segment or a loop region is built during the system preparation step, usually, loop minimization will be taken care of prior to the conventional energy minimization step to “smooth” the connection. The hydrogen atoms of the system are usually minimized by restraining all atoms except hydrogen. The next is to minimize the sidechain, and then, the entire system (complex). The entire system including system complex, water molecules, and ions is minimized finally. Several minimum search algorithms are commonly used to bring the energy to local energy minima, such as steepest descent, conjugate gradient, and Newton-Raphson methods.⁶¹⁻⁶³ The solvation step is to solvate the entire system in a water box, which can be rectangular or octagon. It is suggestive to solvate the entire system having a 10-12 Å away from the box edge to void the periodic boundary issue.⁶⁴

1.2.4.3 Equilibrium and Production Run

Equilibrium after minimization is to ensure the entire system can be simulated with desired temperature in the production run. To do so, velocities are assigned starting with at low

temperature, and equilibrium proceeds to heat up to reach desired temperature at a 25 K temperature segments. The Andersen thermostat, Langevin thermostat and Nose-Hoover thermostat are frequently considered for controlling the temperature.⁶⁵ Equilibration simulation should be run until the energy of the system is stable over time. Once the system has reached equilibrium, a production simulation can be performed. Typical ensemble, *e.g.*, NPT, NVE, or NVT, is chosen according to the nature of the biomolecular system. In the production run step, in which a “real” MD simulation trajectory is generated, and protein’s conformational dynamics, thermodynamics properties, protein ligand binding affinity, and free energy values can be obtained and analyzed. Multiple simulation packages are developed in the community. AMBER,⁴⁵ CHARMM,⁶⁶ and GROMACS⁶⁷ are commonly used. In recent years, the use of graphical processing unit (GPU) cards to accelerate calculations has been a breakthrough in computational simulation field. With the use of GPU cards, which include many arithmetic units working in parallel, MD simulations can be simulated with many magnitudes speed of times. Many major simulation codes have already been rewritten to incorporate GPUs like ACEMD and OpenMM.

1.3 Enhanced Sampling

Unbiased MD (cMD) simulation is common, while bias, accelerated MD (aMD) is also necessary when one wants to simulate dynamics or conformational changes, such as protein ligand dissociation process, which are beyond the capable timescales of cMD. Frequently used enhanced sampling algorithms are aMD,⁶⁸ Gaussian Accelerated MD (GaMD),⁶⁹ and Replica exchange MD (REMD).^{70,71} Only aMD is further expanded in this dissertation.

The principal of aMD is to improve the conformational space sampling by reducing energy barriers separating different states of a system. The method modifies the potential energy landscape by raising energy wells that are below a certain threshold level, while leaving those above this level unaffected. As a result, barriers between adjacent energy basins are reduced, allowing the system to cross the energy barriers much easier than that in a conventional MD simulation. When potential energy of the system falls below a reference energy U , a boost potential energy $\Delta V(r)$ will be added to make the modified potential energy $V^*(r)$ as a sum of the original potential energy and a boost potential energy. The aMD modification of the potential is defined by the following equations:

$$V^*(r) = V(r) + \Delta V(r) \quad \text{Eq. 1.25}$$

$$\Delta V(r) = \begin{cases} 0 & \text{if } V(r) \geq U \\ \frac{(U-V(r))^2}{\alpha+V+V(r)} & \text{if } V(r) < U \end{cases} \quad \text{Eq. 1.26}$$

where $V(r)$ is the original potential, U is the reference energy, and $V^*(r)$ is the modified potential. $\Delta V(r)$ is the boost potential, α is the acceleration factor that determines the shape of the modified potential. To run an aMD, a short and quick cMD (20ns) is performed to obtain dihedral and total potential energies. Four parameters: energy threshold potential (EthreshP), boost energy potential (alphaP), energy threshold dihedral potential (EthreshD), and boost energy dihedral potential (alphaD) are calculated based on the value of acceleration factor α . Once these values are calculated, they are added to MD simulation input file as flags to run in the standard MD simulation protocol.

1.4 Brownian Dynamics Simulations

To simulate substrate diffusional movement in the molecular recognition process in long timescale when structural details is less important in a research objective or it is challenging to observe the entire diffusional process experimentally, Brownian dynamics simulation is considered. It is derived from Langevin dynamics⁷² that uses stochastic differential equations to model the dynamics of a system that has omitted solvent degrees of freedom. Such a simulation typically pairs with an implicit model, such as GBSA. The Brownian dynamics of a particle i can be simulated using the following equation

$$r_i(t + \Delta t) = -\frac{D_i}{k_b T} \frac{dU}{dr_i} \Delta t + \sqrt{2D_i \Delta t} \xi + r_i(t) \quad \text{Eq. 1.27}$$

where D_i is the diffusion coefficient, ξ is the zero mean, stationary Gaussian process satisfying. k_b is Boltzmann constant and T being the temperature. U is a given potential energy function. Δt is the time step, which should be $\gg \frac{m_i D_i}{k_b T}$, but small enough to prevent introducing error.⁷² The diffusion coefficient D_i can be derived from experiment or calculated with respect to the radius of a spherical particle, a_i , and the viscosity of the solvent, η

$$D_i = \frac{k_b T}{6\pi\eta a_i} \quad \text{Eq. 1.28}$$

1.4.1 Brownian Dynamics Simulation Protocol

The protein/receptor in a BD simulation is often a rigid body. To simulate a substrate diffusing to a binding pocket of its receptor, for instance, the pocket in the receptor should be easy to be “accessible” because of the lack of protein conformational rearrangement and

the absence of internal degrees of freedom makes the integration time step much longer than it could. Then, the initial position of the substrate can be determined. Multiple replicas of the substrate can be initiated and let them diffuse freely. Once bound to the destination, the diffusing time is recorded, and a new replica is reproduced till the simulation is terminated. Many BD simulation packages have been developed to study large biological molecules and cellular components.⁷³ Chapter 5 introduces and details one BD simulation package GeomBD3.

1.5 Molecular Docking

In many cases, it is still challenging to know where a chemical compound binds to a pocket of a protein or one protein binds to another protein to affect its function, when x-ray crystallography, nuclear magnetic resonance spectroscopy (NMR) or cryo-EM techniques are not available to resolved ligand-bound structures. Molecular docking has become an indispensable and effective modeling tool in the drug design, antibody design, and drug discovery field.⁷⁴⁻⁷⁷ Molecular docking can be widely decomposed into protein-protein docking and protein-ligand docking. In the protein-protein docking, one protein can be held rigid while another protein is docked to the first protein on the fly to search for possible protein-protein poses. In the protein-ligand docking, however, the protein is held rigid, while the residues near the protein's binding pocket can be flexible or rigid depending on research needs,⁷⁸⁻⁸⁰ and ligand is docked to the binding pocket on the fly to search for possible protein-ligand interaction poses. Molecular docking predicts bound conformational states and its corresponding scoring function, sometimes people refer it as

energetic affinities.^{81,82} Many conformational states are generated and subsequently subjected to an evaluation function to determine the states with the most favorable intermolecular interactions. Then, these states or conformations are scored, ranked, and clustered.

In the context of drug discovery, it is desirable to screen large libraries of small compounds against a molecular receptor to predict drug candidates with high binding affinity for further empirical study. Traditionally, high throughput virtual screening has been used to test libraries of small molecules against a molecular target receptor. Docking methods for *in silico* virtual screening were developed to replace or complement existing high throughput screening methods.^{83,84} Owing to the development of computational power and resource, many free or commercial docking software and packages, such as AutoDock 4.2,⁸⁵ AutoDock Vina 1.2.0,⁸⁶ HADDOCK 2.4,⁸⁷ GOLD,⁸⁸ RosettaDock⁸⁹, and MOE⁵⁵ are available to meet people's research need.

1.6 Post-Analysis of Simulation Data

In general, generating protein and ligand movements using MD simulations or studying subtract diffusional movements using BD simulations is one step, and often straightforward unless stated otherwise by following standard MD or BD simulation protocols mentioned above. Post analyzing these simulations, however, can be tricky, time consuming, and sometimes difficult since many gigabytes or even terabytes of raw data were generated to ensure accuracy/sufficiency of the protein functional dynamics and/or substrate diffusion. A common goal of such simulations is to elucidate protein's functional dynamics and its

conformational changes upon ligand/substrate binding. Depending on the research objectives and applications, several quantitative properties can be further analyzed using common and popular analysis techniques described below.

1.6.1 Root-Mean-Square Deviation

Root-mean-square deviation (RMSD) is an analysis tool often used to calculate the averaged deviation of atom positions in a protein system with respect to the reference positions at a specific timescale. One can specify either the whole system, or just the backbone, the sidechain, or the loop regions for structural equilibrium checking and conformational changes. The equation to calculate RMSD is described as

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - x_0)^2} \quad \text{Eq. 1.29}$$

where x is the measured distance or coordinate, subscript denotes the reference and N is the number of atoms.

1.6.2 Root-Mean-Square Fluctuation

Like RMSD, Root-mean-square fluctuation (RMSF) is also commonly used in MD simulations analysis, except, it is defined to calculate the deviation of one atom i or a group of atoms over a period (time T) from its reference atom positions. For example, RMSF can be used to study loop fluctuation in a protein system or can provide structural flexibility upon ligand binding. The RMSF can be described as

$$RMSF = \sqrt{\frac{1}{T} \sum_{j=1}^T (x_i(t_j) - \bar{x}_i)^2} \quad \text{Eq. 1.30}$$

where x_i is the coordinate or measured distance at time t_i and $\bar{x}_i$ is the reference position of atom i . T is the total time.

1.6.3 Radius of Gyration

Another common structural analysis is the radius of gyration (R_g), which quantifies the degree of compactness of a molecule. The R_g is described as

$$R_g = \sqrt{\frac{\sum_i^n m_i s_i^2}{\sum_i^n m_i}} \quad \text{Eq. 1.31}$$

where m is the mass and s is the distance from the center-of-mass of atom I in a molecule of n atoms. Often it is useful to track the shape change or the compactness a system around a central point. For example, one dissertation project below (chapter 2) used R_g to track the protein's shape change and compactness with respect to the system's RMSD measurement.

1.6.4 Configurational Entropy Calculation

From a statistical mechanics point-of-view, configurational entropy refers to microscopic entropy of conformational entropy in the study of molecules. The conformational dynamics of proteins play critical roles in their functions and regulating ligand binding. Binding affinity of a ligand to protein partner depends on the balance between intermolecular interactions and configurational entropy changes. Both ligand and protein are constantly moving before two encounter each other, but the restriction of motion of the small ligand upon binding to a protein causes changes of translational, rotational, and configurational entropy, and thus a penalty in the binding affinity. One dissertation project herein investigate the residues' sidechain flexibility upon ligand binding using our in-house tool

T-analyst.⁹⁰ Configurational entropy includes both conformational and vibrational terms, which are reflected by the number of energy wells and the average width of the occupied wells, respectively.⁹⁰ Using bond-angle-torsion (BAT) coordinate instead of cartesian coordinates, one can decompose the total entropy into individual terms based on the torsion coordinates, including backbone dihedrals (phi, psi and omega) and sidechain dihedrals.⁹¹ The torsional entropy can be described using the Gibbs entropy formula

$$S_{conf(i)} = -R \sum P_i \ln (P_i) \quad \text{Eq. 1.32}$$

Where P_i is the probability distribution of each torsional angle i and R is the gas constant.

1.6.5 Principal Component Analysis

Principal component analysis (PCA) is commonly used to study random motions of system of interested from long MD trajectories in MD simulations.⁹² Mathematically, PCA is similar to quasi-Harmonic approximation, in which the eigenvectors of the covariance matrix are calculated for intrinsic motion creation. It can be calculated by the following steps: 1, alignment of the calculated covariance matrix from trajectory's atomic coordinates with the study structure (often reference structure); 2, conversion of eigenvectors by diagonalization; 3, creation of motion for each principal component mode from the eigenvectors on average atomic positions. Given the fact that only minimal data is required for convergence, it is common to only considered $C\alpha$ atoms in each residue of the protein for covariance matrix construction. The principal component of each conformation along the PC mode can be computed by:

$$q_i = R^T (x_i - x_a) \quad \text{Eq. 1.33}$$

where the x represents the conformation along the trajectory, i represents the conformation number, a represents the average conformation in the trajectory, q represents the principal component on one PCA mode, R^T represents the transpose of the eigenvector matrix of the covariance matrix.

By reducing it from high dimensionality to low dimensionality with ordered eigenvalues transformation, it has been shown that system's major fluctuation can be observed and described by the first few principal components with alpha carbon data in the backbone movement. Because of the low dimensional representations of system's dynamics, one can compute the free energy landscape by projecting the trajectory onto the space of the first few principal component modes and calculating the free energy from population on these first few dimension spaces, which can represent metastable conformational or transition state of the system.⁹³

1.6.6 MM/PBSA and MM/GBSA Energy Calculations

Molecular Mechanics Poisson Boltzmann Surface Area (MM/PBSA) and its variant Molecular Mechanics Generalized Born Surface Area (MM/GBSA) are examples of end point technique for evaluating specific interactions and binding energies of protein- ligand system from an MD simulation.⁹⁴⁻⁹⁷ The former technique is more accurate but more expensive, whereas the later one is less accurate but less expensive. To compute energy, MM/PBSA replaces the explicit solvent with an implicit PB solvent model, which enables the program to separate the complex system into the protein, the ligand, and the original complex. In, MM/GBSA, the PB solvent model is replaced by more efficient but less

accurate GB solvent model. In addition to reporting the different interaction energies for the ligand-receptor complex, it can also break the interaction energy down to residue-residue interactions. The binding interaction energy ΔG_{bind} , associated with the binding of a protein (P) and ligand (L) to form protein-ligand complex (PL), can be calculated as the following:

$$\Delta G_{bind} = G_{PL} - G_P - G_L \quad \text{Eq. 1.34}$$

It can be decomposed into contributions of different interactions and expressed as

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{sol} - T\Delta S \quad \text{Eq. 1.35}$$

in which,

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{vdW} + \Delta E_{ele} \quad \text{Eq. 1.36}$$

$$\Delta G_{sol} = \Delta G_{PB/GB} + \Delta G_{SA} \quad \text{Eq. 1.37}$$

$$\Delta G_{SA} = \gamma \cdot SASA + b \quad \text{Eq. 1.38}$$

where ΔE_{int} represents the changes in valance energy, ΔE_{vdW} represents the changes in van der Waal interaction energy, and ΔE_{ele} represents the changes in electronic interaction energy. ΔG_{sol} is the sum of the polar contribution of solvation free energy $\Delta G_{PB/GB}$ and non-polar contribution ΔG_{SA} between the solute and the continuum solvent that is estimated using solvent-accessible surface area (SASA).⁹⁸ γ is the surface tension (0.005 kcal/mol·Å²) and b is a correction constant (0.92 kcal/mol). The change in conformational entropy

$-T\Delta S$ is usually calculated by normal-mode analysis on a set of conformational snapshots taken from MD simulations but often neglected due to heavy computational cost.

1.6.7 Diffusion Coefficient

The molecular diffusion during a long timescale molecular recognition process is often influenced by many factors, such as the environment temperature, viscosity of the solvent, ionic concentration, the volume of the cell environment, and the intermolecular potentials. Calculating the diffusion coefficient (D) of a substrate in a BD simulation can be described as

$$D = \lim_{t \rightarrow \infty} \frac{(r_t - r_0)^2}{2nt} \quad \text{Eq. 1.39}$$

where $(r_t - r_0)$ represents the displacement a molecule at time t , and n represents the dimensionality of the diffusion. The diffusion coefficient (D) can be used to calculate molecular association and dissociation rates. One can calculate the diffusion-limited association rate k_{on} of two molecules a and b analytically using

$$k_{on} = 4\pi(D_a + D_b)(r_a + r_b) \quad \text{Eq. 1.40}$$

Where D is the diffusion coefficient and r is the radius of the molecule a and b .⁹⁹ If attractive or repulsive potential exist between molecules a and b , the k_{on} is computed by

$$k_{on} = \frac{4\pi(D_a + D_b)(r_a + r_b)}{\int_{a+b}^{\infty} \frac{1}{r^2} e^{-\beta E(r)} dr} \quad \text{Eq. 1.41}$$

Where β is inverse temperature divided by Boltzmann's constant and $E(r)$ is the intermolecular potential.¹⁰⁰

1.7 References

1. Gellman SH. Introduction: molecular recognition. *Chemical reviews* 1997;97(5):1231-1232.
2. Gibbs AC. Elements and modulation of functional dynamics. *Journal of Medicinal Chemistry* 2014;57(19):7819-7837.
3. McCammon JA. Theory of biomolecular recognition. *Current opinion in structural biology* 1998;8(2):245-249.
4. McCammon J, Gelin B, Karplus M. Dynamics of folded proteins. *nature*, 267. 1977.
5. Amrhein S, Oelmeier SA, Dismar F, Hubbuch Jr. Molecular dynamics simulations approach for the characterization of peptides with respect to hydrophobicity. *The Journal of Physical Chemistry B* 2014;118(7):1707-1714.
6. Damjanovic J, Miao J, Huang H, Lin Y-S. Elucidating solution structures of cyclic peptides using molecular dynamics simulations. *Chemical Reviews* 2021;121(4):2292-2324.
7. Ulmschneider JP, Ulmschneider MB. Molecular dynamics simulations are redefining our view of peptides interacting with biological membranes. *Accounts of chemical research* 2018;51(5):1106-1116.
8. Wu KY, Doan D, Medrano M, Chang C-eA. Modeling structural interconversion in Alzheimers' amyloid beta peptide with classical and intrinsically disordered protein force fields. *Journal of Biomolecular Structure and Dynamics* 2021:1-18.
9. Bosken YK, Cholko T, Lou YC, Wu KP, Chang CEA. Insights Into Dynamics of Inhibitor and Ubiquitin-Like Protein Binding in SARS-CoV-2 Papain-Like Protease. *Frontiers in Molecular Biosciences* 2020;7.
10. Cholko T, Chen W, Tang Z, Chang CeA. A molecular dynamics investigation of CDK8/CycC and ligand binding: conformational flexibility and implication in drug discovery. *Journal of Computer-Aided Molecular Design* 2018;32(6):671-685.
11. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nature structural biology* 2002;9(9):646-652.

12. Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M. Simultaneous determination of protein structure and dynamics. *Nature* 2005;433(7022):128-132.
13. D'Amico RN, Bosken YK, O'Rourke KF, Murray AM, Admasu W, Chang C-EA, Boehr DD. Substitution of a surface-exposed residue involved in an allosteric network enhances tryptophan synthase function in cells. *Frontiers in molecular biosciences* 2021;8:679915.
14. Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. *Neuron* 2018;99(6):1129-1143.
15. Tumbi KM, Nandekar PP, Shaikh N, Kesharwani SS, Sangamwar AT. Molecular dynamics simulation studies for DNA sequence recognition by reactive metabolites of anticancer compounds. *Journal of Molecular Recognition* 2014;27(3):138-150.
16. Jawad B, Poudel L, Podgornik R, Steinmetz NF, Ching W-Y. Molecular mechanism and binding free energy of doxorubicin intercalation in DNA. *Physical Chemistry Chemical Physics* 2019;21(7):3877-3893.
17. Cholko T, Kaushik S, Chia-en AC. Dynamics and molecular interactions of single-stranded DNA in nucleic acid biosensors with varied surface properties. *Physical Chemistry Chemical Physics* 2019;21(29):16367-16380.
18. Flood E, Boiteux C, Lev B, Vorobyov I, Allen TW. Atomistic simulations of membrane ion channel conduction, gating, and modulation. *Chemical Reviews* 2019;119(13):7737-7832.
19. Nguyen THT, Liu Z, Moore PB. Molecular dynamics simulations of homo-oligomeric bundles embedded within a lipid bilayer. *Biophysical Journal* 2013;105(7):1569-1580.
20. Venable RM, Brown FL, Pastor RW. Mechanical properties of lipid bilayers from molecular dynamics simulation. *Chemistry and physics of lipids* 2015;192:60-74.
21. Feller SE. Molecular dynamics simulations of lipid bilayers. *Current opinion in colloid & interface science* 2000;5(3-4):217-223.
22. de Almeida H, Bastos IM, Ribeiro BM, Maigret B, Santana JM. New binding site conformations of the dengue virus NS3 protease accessed by molecular dynamics simulation. *PloS one* 2013;8(8):e72402.

23. Gupta G, Lim L, Song J. NMR and MD studies reveal that the isolated dengue NS3 protease is an intrinsically disordered chymotrypsin fold which absolutely requests NS2B for correct folding and functional dynamics. *PLoS One* 2015;10(8):e0134823.
24. Smondyrev AM, Voth GA. Molecular dynamics simulation of proton transport through the influenza A virus M2 channel. *Biophysical Journal* 2002;83(4):1987-1996.
25. Huber RG, Marzinek JK, Holdbrook DA, Bond PJ. Multiscale molecular dynamics simulation approaches to the structure and dynamics of viruses. *Progress in Biophysics and Molecular Biology* 2017;128:121-132.
26. Şterbuleac D. Molecular dynamics: a powerful tool for studying the medicinal chemistry of ion channel modulators. *RSC Medicinal Chemistry* 2021;12(9):1503-1518.
27. Leach AR. *Molecular modelling: principles and applications*. Pearson education; 2001.
28. Zuckerman DM. *Statistical physics of biomolecules: an introduction*. CRC press; 2010.
29. Jensen F. *Introduction to computational chemistry*. John wiley & sons; 2017.
30. Allen MP, Frenkel D, Talbot J. Molecular dynamics simulation using hard particles. *Computer physics reports* 1989;9(6):301-353.
31. Guvench O, MacKerell AD. Comparison of protein force fields for molecular dynamics simulations. *Molecular modeling of proteins* 2008:63-88.
32. Piana S, Klepeis JL, Shaw DE. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Current opinion in structural biology* 2014;24:98-105.
33. Lopes PE, Guvench O, MacKerell AD. Current status of protein force fields for molecular dynamics simulations. *Molecular modeling of proteins*: Springer; 2015. p 47-71.
34. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Structure, Function, and Bioinformatics* 2006;65(3):712-725.
35. Case DA, Cheatham III TE, Darden T, Gohlke H, Luo R, Merz Jr KM, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. *Journal of computational chemistry* 2005;26(16):1668-1688.

36. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry* 2010;31(4):671-690.
37. Scott WR, Hünenberger PH, Tironi IG, Mark AE, Billeter SR, Fennel J, Torda AE, Huber T, Krüger P, van Gunsteren WF. The GROMOS biomolecular simulation program package. *The Journal of Physical Chemistry A* 1999;103(19):3596-3607.
38. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. *Journal of computational chemistry* 2005;26(16):1781-1802.
39. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society* 1996;118(45):11225-11236.
40. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* 1983;79(2):926-935.
41. Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *The Journal of Physical Chemistry A* 2001;105(43):9954-9960.
42. Zielkiewicz J. Structural properties of water: Comparison of the SPC, SPCE, TIP4P, and TIP5P models of water. *The Journal of chemical physics* 2005;123(10):104501.
43. Lu B, Zhou Y, Holst M, McCammon J. Recent progress in numerical methods for the Poisson-Boltzmann equation in biophysical applications. *Commun Comput Phys* 2008;3(5):973-1009.
44. Lu B, Zhang D, McCammon JA. Computation of electrostatic forces between solvated molecules determined by the Poisson-Boltzmann equation using a boundary element method. *The Journal of chemical physics* 2005;122(21):214102.
45. Case DA, Aktulga HM, Belfon K, Ben-Shalom I, Brozell SR, Cerutti DS, Cheatham III TE, Cruzeiro VWD, Darden TA, Duke RE. *Amber 2021*. University of California, San Francisco; 2021.
46. Onufriev A, Case DA, Bashford D. Effective Born radii in the generalized Born approximation: the importance of being perfect. *Journal of computational chemistry* 2002;23(14):1297-1304.

47. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics* 2004;55(2):383-394.
48. Burley SK, Bhikadiya C, Bi C, Bittrich S, Chen L, Crichlow GV, Christie CH, Dalenberg K, Di Costanzo L, Duarte JM. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic acids research* 2021;49(D1):D437-D451.
49. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596(7873):583-589.
50. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* 2018;46(W1):W296-W303.
51. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B. PubChem in 2021: new data content and improved web interfaces. *Nucleic acids research* 2021;49(D1):D1388-D1395.
52. Sterling T, Irwin JJ. ZINC 15—ligand discovery for everyone. *Journal of chemical information and modeling* 2015;55(11):2324-2337.
53. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of cheminformatics* 2012;4(1):1-17.
54. Pedretti A, Mazzolari A, Gervasoni S, Fumagalli L, Vistoli G. The VEGA suite of programs: An versatile platform for cheminformatics and drug design projects. *Bioinformatics* 2021;37(8):1174-1175.
55. Inc CCG. Molecular operating environment (MOE). Chemical Computing Group Inc. 1010 Sherbooke St. West, Suite# 910, Montreal ...; 2019.
56. Olsson MH, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: consistent treatment of internal and surface residues in empirical p K a predictions. *Journal of chemical theory and computation* 2011;7(2):525-537.
57. Tan KP, Nguyen TB, Patel S, Varadarajan R, Madhusudhan MS. Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. *Nucleic acids research* 2013;41(W1):W314-W321.

58. Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A. H++: a server for estimating pK_as and adding missing hydrogens to macromolecules. *Nucleic acids research* 2005;33(suppl_2):W368-W371.
59. Song Y, Mao J, Gunner MR. MCCE2: improving protein pK_a calculations with extensive side chain rotamer sampling. *Journal of computational chemistry* 2009;30(14):2231-2247.
60. Raquet X, Lounnas V, Lamotte-Brasseur J, Frere J, Wade R. pK_a calculations for class A beta-lactamases: methodological and mechanistic implications. *Biophysical journal* 1997;73(5):2416-2426.
61. Curry HB. The method of steepest descent for non-linear minimization problems. *Quarterly of Applied Mathematics* 1944;2(3):258-261.
62. Štich I, Car R, Parrinello M, Baroni S. Conjugate gradient minimization of the energy functional: A new method for electronic structure calculation. *Physical Review B* 1989;39(8):4997.
63. Ben-Israel A. A Newton-Raphson method for the solution of systems of equations. *Journal of Mathematical analysis and applications* 1966;15(2):243-252.
64. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of computational chemistry* 2003;24(16):1999-2012.
65. Andersen HC. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of chemical physics* 1980;72(4):2384-2393.
66. Brooks BR, Brooks III CL, Mackerell Jr AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S. CHARMM: the biomolecular simulation program. *Journal of computational chemistry* 2009;30(10):1545-1614.
67. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, Van Der Spoel D. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 2013;29(7):845-854.
68. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *The Journal of chemical physics* 2004;120(24):11919-11929.

69. Miao Y, Feher VA, McCammon JA. Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation. *Journal of chemical theory and computation* 2015;11(8):3584-3595.
70. Voter AF. Parallel replica method for dynamics of infrequent events. *Physical Review B* 1998;57(22):R13985.
71. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters* 1999;314(1-2):141-151.
72. Ermak DL, McCammon JA. Brownian dynamics with hydrodynamic interactions. *The Journal of chemical physics* 1978;69(4):1352-1360.
73. Huber GA, McCammon JA. Brownian dynamics simulations of biological molecules. *Trends in chemistry* 2019;1(8):727-738.
74. Meng X-Y, Zhang H-X, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design* 2011;7(2):146-157.
75. Lybrand TP. Ligand—protein docking and rational drug design. *Current opinion in structural biology* 1995;5(2):224-228.
76. Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. *Molecules* 2015;20(7):13384-13421.
77. Saikia S, Bordoloi M. Molecular docking: challenges, advances and its use in drug discovery perspective. *Current drug targets* 2019;20(5):501-521.
78. Sandak B, Wolfson HJ, Nussinov R. Flexible docking allowing induced fit in proteins: insights from an open to closed conformational isomers. *Proteins: Structure, Function, and Bioinformatics* 1998;32(2):159-174.
79. Nabuurs SB, Wagener M, De Vlieg J. A flexible approach to induced fit docking. *Journal of medicinal chemistry* 2007;50(26):6507-6518.
80. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Current opinion in structural biology* 2008;18(2):178-184.
81. Wang R, Lu Y, Wang S. Comparative evaluation of 11 scoring functions for molecular docking. *Journal of medicinal chemistry* 2003;46(12):2287-2303.

82. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature reviews Drug discovery* 2004;3(11):935-949.
83. Shoichet BK. Virtual screening of chemical libraries. *Nature* 2004;432(7019):862-865.
84. Lin X, Li X, Lin X. A review on applications of computational methods in drug screening and design. *Molecules* 2020;25(6):1375.
85. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry* 2009;30(16):2785-2791.
86. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2. 0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling* 2021;61(8):3891-3898.
87. Honorato RV, Koukos PI, Jiménez-García B, Tsaregorodtsev A, Verlato M, Giachetti A, Rosato A, Bonvin AM. Structural biology in the clouds: the WeNMR-EOSC ecosystem. *Frontiers in Molecular Biosciences* 2021:708.
88. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein–ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics* 2003;52(4):609-623.
89. Marze NA, Roy Burman SS, Sheffler W, Gray JJ. Efficient flexible backbone protein–protein docking for challenging targets. *Bioinformatics* 2018;34(20):3461-3469.
90. Ai R, Fatmi MQ, Chang C-e, A. T-Analyst: a program for efficient analysis of protein conformational changes by torsion angles. *Journal of computer-aided molecular design* 2010;24(10):819-827.
91. Pitzer KS. Energy levels and thermodynamic functions for molecules with internal rotation: II. Unsymmetrical tops attached to a rigid frame. *The Journal of Chemical Physics* 1946;14(4):239-243.
92. Bro R, Smilde AK. Principal component analysis. *Analytical methods* 2014;6(9):2812-2831.
93. Zhou R, Berne BJ, Germain R. The free energy landscape for β hairpin folding in explicit water. *Proceedings of the National Academy of Sciences* 2001;98(26):14931-14936.

94. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of chemical research* 2000;33(12):889-897.
95. Massova I, Kollman PA. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspectives in drug discovery and design* 2000;18(1):113-135.
96. Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *Journal of chemical information and modeling* 2011;51(1):69-82.
97. Wang E, Sun H, Wang J, Wang Z, Liu H, Zhang JZ, Hou T. End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chemical reviews* 2019;119(16):9478-9508.
98. Gilson MK, Honig B. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins: Structure, Function, and Bioinformatics* 1988;4(1):7-18.
99. Berg OG, von Hippel PH. Diffusion-controlled macromolecular interactions. *Annual review of biophysics and biophysical chemistry* 1985;14(1):131-158.
100. Jackson MB. *Molecular and cellular biophysics*. Cambridge University Press; 2006.

Chapter 2: Modeling Structural Interconversion in Alzheimer's Amyloid Beta Peptide With Classical and Intrinsically Disordered Protein Force Fields

2.1 Abstract

A comprehensive understanding of the aggregation mechanism in amyloid beta 42 (A β 42) peptide is imperative for developing therapeutic drugs to prevent or treat Alzheimer's disease. Because of the high flexibility and lack of native tertiary structures of A β 42, molecular dynamics (MD) simulations may help elucidate the peptide's dynamics with atomic details and collectively improve ensembles not seen in experiments. We applied microsecond-timescale MD simulations to investigate the dynamics and conformational changes of A β 42 by using a newly developed Amber force field (*ff14IDPSFF*). We compared the *ff14IDPSFF* and the regular *ff14SB* force field by examining the conformational changes of two distinct A β 42 monomers in explicit solvent. Conformational ensembles obtained by simulations depend on the force field and initial structure, A β 42 $^{\alpha\text{-helix}}$, or A β 42 $^{\beta\text{-strand}}$. The *ff14IDPSFF* sampled a high ratio of disordered structures and diverse β -strand secondary structures; in contrast, *ff14SB* favored helicity during the A β 42 $^{\alpha\text{-helix}}$ simulations. The conformations obtained from A β 42 $^{\beta\text{-strand}}$ simulations maintained a balanced content in the disordered and helical structures when simulated by *ff14SB*, but the conformers clearly favored disordered and β -sheet structures simulated by *ff14IDPSFF*. The results obtained with *ff14IDPSFF* qualitatively reproduced the NMR chemical shifts well. In-depth peptide and cluster analysis revealed some characteristic features that may be linked to early onset of the fibril-like structure. The C-terminal region

(mainly M35-V40) featured in-registered anti-parallel β -strand (β -hairpin) conformations with tested systems. Our work should expand the knowledge of force field and structure dependency in MD simulations and reveals the underlying structural mechanism–function relationship in A β 42 peptides.

2.2 Introduction

Intrinsically disordered proteins (IDPs) are associated with neurodegenerative disorders such as Parkinson's disease, Alzheimer's disease (AD), prion disease, Huntington's disease, and type II diabetes.¹⁻⁴ Amyloid beta peptides belong to the class of IDPs associated with AD. At the neurophysiological level, the pathology of AD is characterized by the presence of insoluble, densely packed, and aggregated peptide fibrils in the brain. Before forming the amyloid plaques, the soluble and structured oligomers of the amyloid beta (A β) peptides undoubtedly have neurotoxic effects.⁵⁻⁷ Therefore, investigating the dynamics of the peptide is crucial to understanding the molecular processes that lead to cell toxicity as well as the action of small molecule inhibitors for AD.⁸

As compared with its alloform A β 40, A β 42 (Figure 2.1 A) is more neurotoxic and aggregates faster in amyloid plaques.⁹⁻¹¹ The N-terminal region (residues D1 to K16) is highly flexible in the whole sequence and has metal-binding ability under abnormal physiological conditions.¹²⁻¹⁵ The region mainly with non-polar residues spanning L17 to A21 is known as the central hydrophobic region (CHR). This region has a close relation with the aggregation event.¹⁶ Following the CHR is the central polar region (CPR), which

contains many hydrophilic residues (E22 to G29). The final C-terminal hydrophobic region contains more non-polar residues (A30 to A42). X-ray crystallography, solid-state NMR spectroscopy, cryo-electron microscopy, and other approaches¹⁷⁻²⁶ have been advanced to determine static structures with high resolution, but exploration of A β peptide dynamics and the structural mechanism–function association remains elusive. Molecular modeling methods are powerful tools for sampling more conformations and investigating the biophysical mechanisms of the A β peptide.

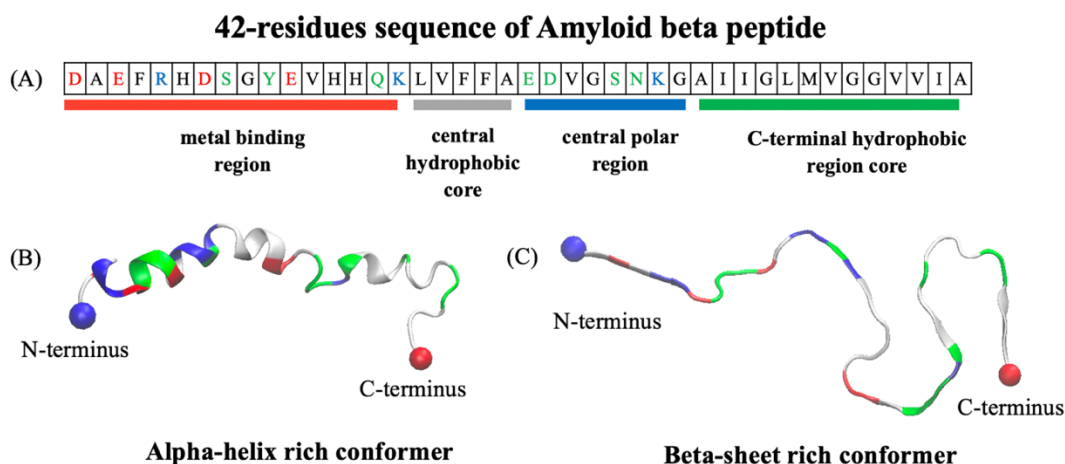


Figure 2.1: (A) Sequence of 42 residues amyloid beta peptide (A β 42), (Red: polar, negative charge residues; Blue: polar, positive charge residues; Black: non-polar residues; Green: polar, neutral residues). (B) NMR-resolved structure of A β 42 $^{\alpha}$ -helix (the 4th model, PDB ID: 1Z0Q). (C) NMR-resolved structure of A β 42 $^{\beta}$ -strand (strand A of the 5th model, PDB ID: 2NAO); blue sphere, N-termini, red sphere, C-termini.

Molecular dynamics (MD) simulations are emerging tools used to sample atomistic details of structural conformations and dynamics of molecules not achieved with experiments. Over the past decades, studies have used different force fields and MD algorithms to understand the dynamics and conformations of A β 42, ranging from monomeric to fibrillar

states.²⁷⁻⁴⁵ However, conventional force fields with parameterization used in MD studies were optimized for structured proteins to agree with experimental data or often involved quantum calculations of small molecules or short peptide fragments. Studies have focused on evaluating the performance of different force fields in the hope of finding suitable force fields for A β simulation.^{43,46-53} For instance, Man *et al.* evaluated the performance of 17 widely used atomistic MM force fields on a seven-residue peptide from AMBER, GROMOS, OPLS and CHARMM families and suggested that five force fields were good candidates for amyloid peptide simulations: two are AMBER99SB-ILDN and AMBER14SB and three are CHARMM C22*, C36 and C36m.⁵² Carballo-Pacheco *et al.* applied various AMBER-based OPLS and CHARMM22* force fields with replica exchange MD (REMD) to investigate the dynamics of A β 42: CHARMM22* reproduced better NMR observables for C α and HN chemical shifts than the other force fields.⁴⁹ Rosenman *et al.* analyzed the ensembles of two isoforms (A β 40 and A β 42) and mutant forms of A β 40 by using REMD with three different force fields to understand the sequence-dependent dynamics.⁴⁷ However, concluding on a universal force field for simulating amyloid peptides (A β , tau and α -synuclein, *etc.*) is challenging because their structural preferences depend on the force field, simulation protocol and conditions used.⁵⁴⁻

⁵⁷ MD simulations of IDPs using conventional force fields may incorrectly characterize the features of IDPs, thereby affecting the quantitative analysis and hiding realistic ensemble descriptions. Recently, IDP-specific force fields such as *ff99IDPs*,⁵⁸ *ff14IDPs*,⁵⁹ *ff14IDPSFF*,⁶⁰ *A99SB-disp*,³¹ *OPLSIDPSFF*,⁶¹ *CHARMM36IDPSFF*⁶² and *ff99SBnmr2*⁶³ were developed to refine and improve the modeling of IDPs. Because these updated IDP-

specific force fields were parameterized based on short coils and turn peptides to improve the backbone dihedral angles or were optimized to achieve agreement with disordered peptides, conformational ensembles obtained with these force fields reproduced the quantitative characteristics of IDPs and intrinsically disordered regions. Specifically, *ff14IDPSFF* showed improved performance in sampling various IDPs, with high corresponding values to the experimental observables.^{53,60,64,65} Krupa *et al.* used conventional MD (cMD) and REMD to study the dynamics of A β 42 by comparing the performance of *ff14IDSPFF* and *CHARMM36m* with their original versions, *ff99SB* and *ff14SB*, and *CHARMM*, respectively, and they have shown that *ff14IDPSFF* yielded similar results in REMD sampling but improved conformational sampling in cMD.³⁸ Recently, Rahman *et al.*, conducted MD simulations using various IDP-specify force fields, namely, *ff99IDPs*, *ff14IDPs* and *ff14IDPSFF*, to study two different initial structures of A β 42: one α -helix rich content (33.33%) conformer and one model conformer (α -helix content 26.19%). The authors concluded similar results for these force fields.⁵³ The question remains as to whether and how to distinguish the initial conformation for such a flexible peptide to see a significant difference on computational study, or is there an interconversion event occurring within the same A β conformer but different results observed with different tested force fields?

This study compared the newly developed but not extensively explored *ff14IDPSFF* force field and the regular *ff14SB* force field by examining the conformational changes of two totally different A β 42 monomers in explicit solvent. The initial structures are shown in Figures 2.1 B and 2.1 C. We sampled the conformer ensembles by using cMD and

accelerated MD (aMD) simulations to attain microsecond-timescale (total 15.6 μ s) MD and performed post-analysis to access the structural stability, flexibility and secondary structure properties. We evaluated the performance of the force fields by comparing the predicted structures with NMR chemical shift values and accessed the radius of gyration (Rg) values from the simulations for comparison with other studies. We also revealed some characteristic features in the sequence that were found in common from our study. We observed clear differences with the two tested force fields in the same initial structures, and similar conformational features within A β 42 were revealed by using distinct initial conformations. Our work broadens the spectrum of force-field and initial structure selection and provides useful insights into the conformational changes in A β 42.

2.3 Methods and Material

MD Simulations

Two NMR structures (PDB ID: 1Z0Q,⁶⁶ monomer and PDB ID: 2NAO,⁶⁷ fibril) were used in the simulations. In 1Z0Q, with an α -helix rich conformer, 30 conformations were deposited, and the 4th conformation with extended structure was selected as our initial structure A β 42 ^{α -helix} for cMD runs. For 2NAO, with β -strand rich fibril, we selected the strand A of the 5th conformation as our initial structure A β 42 ^{β -strand} for cMD runs. In A β 42 ^{α -helix} (Figure 2.1 B), this peptide is mixed, with mostly helices spanning S8 to V24 and 3-10 helix spanning I31 to V36, connected with the coil structure. In A β 42 ^{β -strand} (Figure 2.1 C), the β -strand segment spans D1 to D7, Q15 to F19, and V40 to A42, connected with coils,

a turn and the isolated bridge structures. Table 2.1 lists all working systems and conditions according to the force fields and simulation method used.

Table 2.1. Summary of molecular dynamics (MD) simulations.

Simulation index	Force field	Simulation Method	Length (ns)	Boost potential α
A1 (1,2,3)	<i>ff14SB</i>	cMD	1000	-
A1I (1,2,3)	<i>ff14IDPSFF</i>	cMD	1000	-
A2 (1,2,3)	<i>ff14SB</i>	aMD	100	0.2
A2I (1,2,3)	<i>ff14IDPSFF</i>	aMD	100	0.2
A3 (1,2,3)	<i>ff14SB</i>	aMD	100	0.5
A3I (1,2,3)	<i>ff14IDPSFF</i>	aMD	100	0.5
A4 (1,2,3)	<i>ff14SB</i>	aMD	100	1.0
A4I (1,2,3)	<i>ff14IDPSFF</i>	aMD	100	1.0
B1 (1,2,3)	<i>ff14SB</i>	cMD	1000	-
B1I (1,2,3)	<i>ff14IDPSFF</i>	cMD	1000	-
B2 (1,2,3)	<i>ff14SB</i>	aMD	100	0.2
B2I (1,2,3)	<i>ff14IDPSFF</i>	aMD	100	0.2
B3 (1,2,3)	<i>ff14SB</i>	aMD	100	0.5
B3I (1,2,3)	<i>ff14IDPSFF</i>	aMD	100	0.5
B4 (1,2,3)	<i>ff14SB</i>	aMD	100	1.0
B4I (1,2,3)	<i>ff14IDPSFF</i>	aMD	100	1.0

In aMD, the initial conformation was obtained from the last frame of the cMD. All simulations were performed with 3 random number seeds noted as (1, 2, 3) in the simulation index column. All MD simulations were performed with the Amber 18 simulations package.⁶⁸ Amber *ff14SB* and *ff14IDPSFF*⁶⁰ force fields were used for all working systems. To use *ff14IDPSFF*, the regular topology file with optimized grid-based energies (CMAP) for all 20 amino acids was converted to the new topology file.⁵⁸ A standard Amber protocol was used. Na⁺ ions were added to neutralize the systems. The following minimization steps were performed for the hydrogen atom, side chains, and the entire protein complex. Then, the systems were solvated with a rectangular (orthogonal) TIP3P water box⁶⁹ with the cutting edge of box 12 Å away from the solutes. Minimization for the water molecules followed by the entire system was performed to correct any inconsistency. The water molecules with solutes fixed at 300 K were equilibrated in the NPT ensemble, followed by heating the systems with an increment of 25 K temperature up to the final 300 K. This step was used to fully relax the system before entering the production run. Long-range electrostatic interactions were computed by the particle mesh Ewald method.⁷⁰ The SHAKE algorithm was used to constrain the bonds involving hydrogen atoms.⁷¹ The Langevin thermostat^{72,73} with a damping constant of 2 ps⁻¹ was used to maintain a temperature of 300 K. The production run in NPT ensemble had a 2-fs time step. cMD with a 1-μs production run for each working system was simulated for 3 seeds. We collected the resulting trajectories every 1 ps and re-saved the frames for analysis at intervals of 100 ps.

To further enhance the samplings of the monomers, all working systems underwent aMD simulations.⁷⁴ Briefly, cMD simulations were run at 10 ns to obtain dihedral and total potential energies. To simulate aMD with different boost potentials, we calculated four parameters — energy threshold potential (EthreshP), boost energy potential (alphaP), energy threshold dihedral potential (EthreshD), and boost energy dihedral potential (alphaD) — based on the established protocol,⁷⁵ and different boost potentials ($\alpha=0.2, 0.5$ and 1.0) were applied in the alphaD parameter to each working system. All aMD simulations were performed at 100 ns for 3 seeds (Table 2.1) under the same protocol and conditions run by cMD. We used Define Secondary Structure Protein (DSSP) and conformational clustering, for all aMD runs.

Post-Analysis

We used visual MD (VMD)⁷⁶ for visualization. Cpptraj implemented in AmberTools18⁶⁸ was used to calculate the root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), solvent accessible surface areas (SASA), and intra-peptide contact map. Chemical shifts were downloaded from the BioMasRes Database (BMRB)⁷⁷ (BMRB ID: 6554 for A β 42 $^{\alpha}$ -helix and BMRB ID: 26692 for A β 42 $^{\beta}$ -strand). The chemical shifts were estimated by using the SHIFTX2 program⁷⁸ and results were compared with the NMR data. We used Multiscale Modeling Tools for Structural Biology (MMTSB)⁷⁹ with a K-means clustering program based on an RMSD cutoff of 4.5 Å to proceed with conformational clustering. The secondary structures for A β 42 were assigned

based on the DSSP algorithm⁸⁰ implemented in VMD. The salt bridge distance measurement with a cutoff of 4.0 Å was calculated in VMD.

2.4 Results and Discussion

Assessment of Overall Stability and Flexibility of Peptide Aβ42

To examine the stability and flexibility of all working systems, we first calculated the Cα atomic RMSDs (Figure 2.2). Figure 2.2 B illustrates a larger conformational adjustment with *ff14IDPSFF* simulations initiated from Aβ42^{α-helix} as compared with *ff14SB* simulations (Figure 2.2 A). For Aβ42^{β-strand} simulated by the two force fields, *ff14SB* (Figure 2.2 C) conferred a smaller average conformational fluctuation (RMSD values from 10 to 14 Å) as compared with *ff14IDPSFF* (Figure 2.2 D) (RMSD values from 6 to 14 Å), which is similar to the range 5 to 15 Å observed by Krupa *et al.*,³⁸ who used cMD with the same IDP force field. Structured proteins typically fluctuate with less dynamics (RMSD < 3 Å), whereas in Aβ42, the RMSD values ranged from 6 to 18 Å, which indicates that it unfolded and refolded during the simulations. To further monitor the convergence of all working systems, we calculated the cumulative clusters in a time-dependent window. For each 100-ns time window, the cluster of conformers was calculated (Figure 2.3). For simulations of Aβ42^{α-helix}, the overall number of clusters increased more with *ff14IDPSFF* than *ff14SB*. A similar trend occurred with simulations of Aβ42^{β-strand}, with a higher number of clusters sampled with *ff14IDPSFF* than *ff14SB*. Notably, the cluster numbers of all working systems reached a plateau toward the end of the simulations, which indicates the convergence of the conformation sampling. However, existing free energy landscape

studies suggested that the intrinsically disordered A β 42 has more distinct minima than does a typical globular protein.^{81,82} Our MD runs did not likely sample all energy minima. The conformations may be kinetically trapped in an energy funnel,⁸³ and enhanced sampling techniques, such as REMD simulation with a set of parameters (temperature, number of replicas, exchange rates, *etc.*) need to be applied to overcome the energy barriers. As a result, to sample more conformations, we also used aMD to obtain additional conformations.

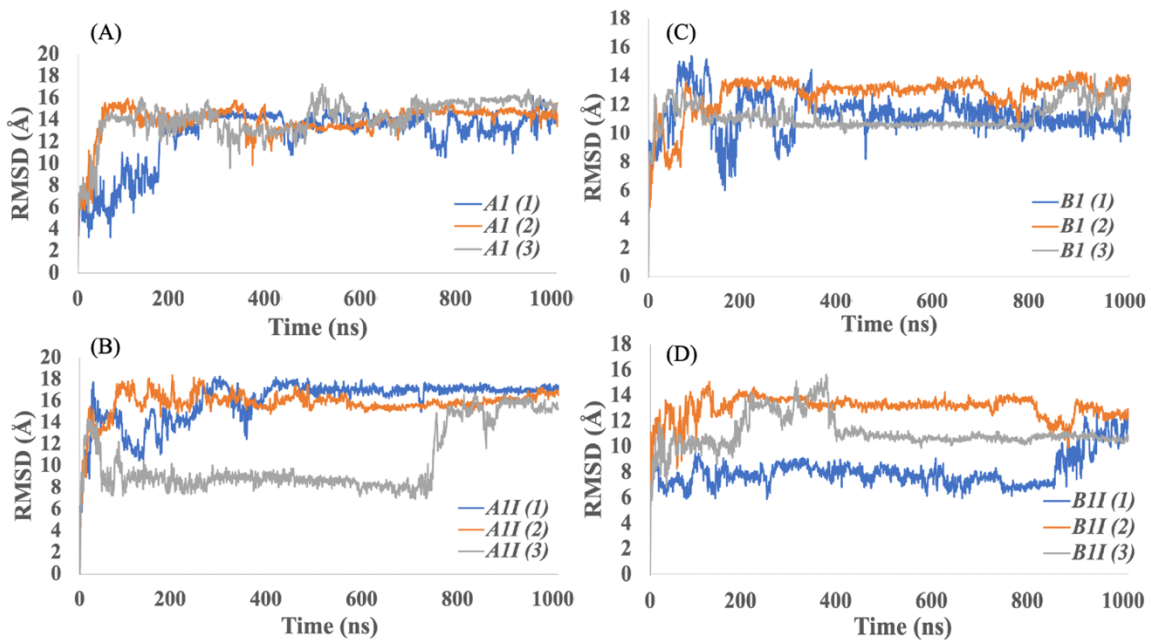


Figure 2.2: RMSD for A β 42 of two monomers using (A) RMSD for A β 42 $^{\alpha}$ -helix with *ff14SB* [Simulation index: A1 (1,2,3)]. (B) RMSD for A β 42 $^{\alpha}$ -helix with *ff14IDPSFF* [Simulation index: A1I (1,2,3)]. (C) RMSD for A β 42 $^{\beta}$ -strand with *ff14SB* [Simulation index: B1 (1,2,3)]. (D) RMSD for A β 42 $^{\beta}$ -strand with *ff14IDPSFF* [Simulation index: B1I (1,2,3)]. Each initial conformation has 3 independent runs using 3 random-number seeds (1, 2, 3).

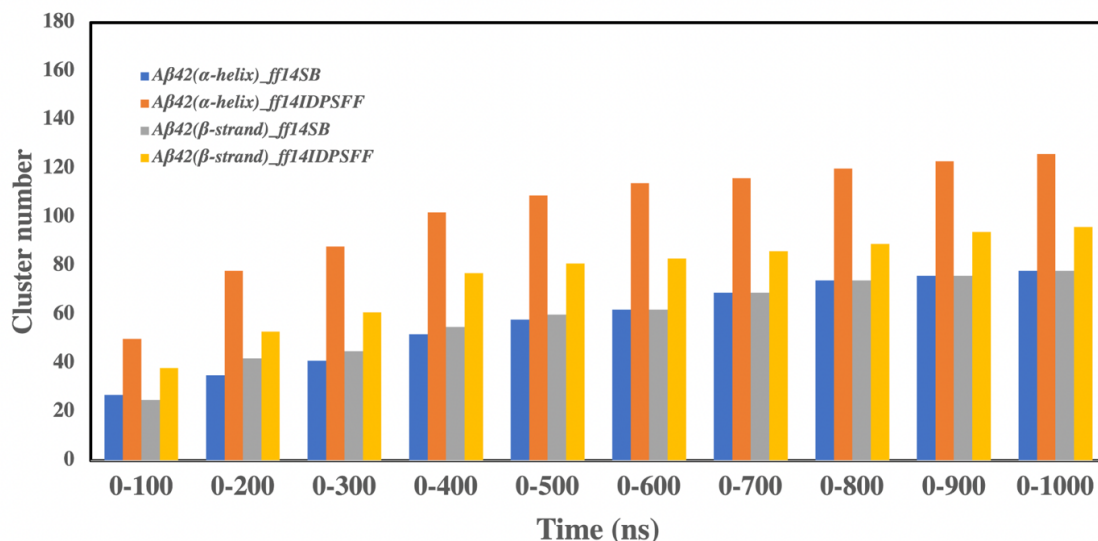


Figure 2.3: Conformers of clustering number with respect to the cumulative time (100 ns window) under two force fields for two monomers.

To illustrate the overall flexibility of the evaluated Aβ42 monomers, we computed the Cα RMSF (Figure 2.4). The central hydrophobic core region (gray bar) was more stable than the N- and C-termini. The first 10 residues (or the metal binding region) are sometimes missing in other structures, but unexpectedly, we did not observe that this region was more flexible than the C-terminal region. Fluctuation in simulations was high with *ff14IDPSFF* for both monomers as compared with *ff14SB*. This observation can explain the relatively high ratio of disordered conformations simulated with *ff14IDPSFF*.

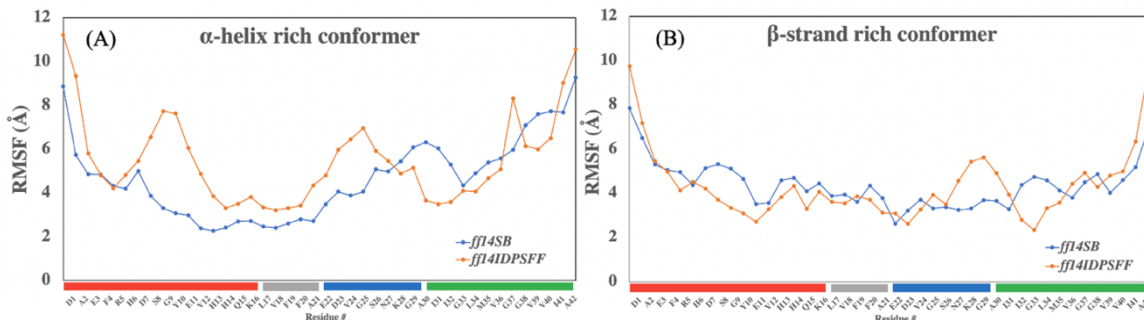
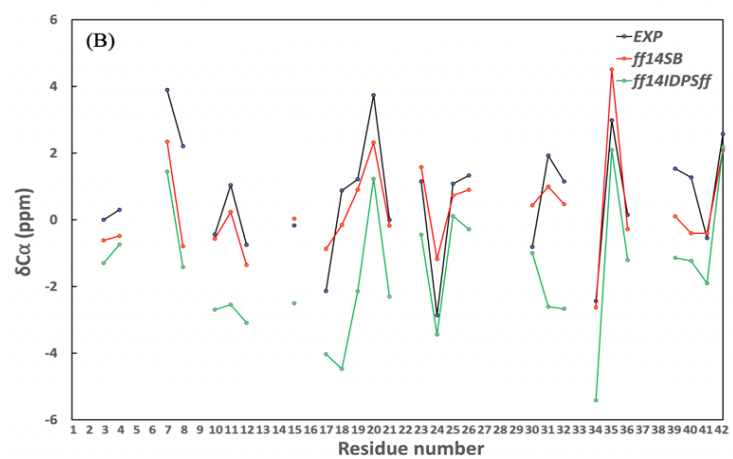
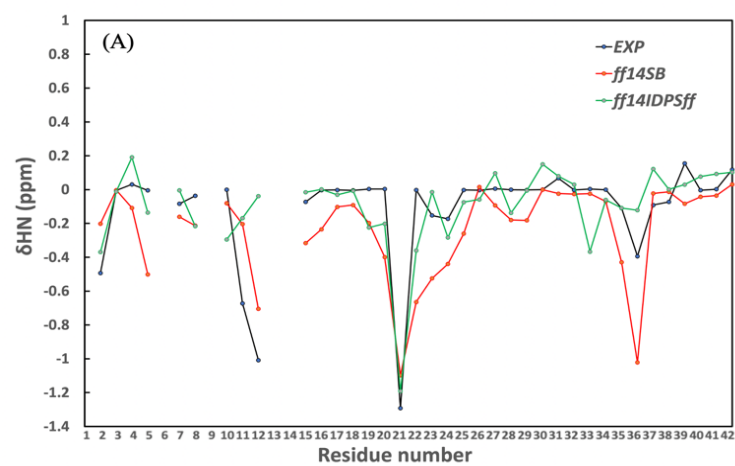


Figure 2.4: RMSF analysis of the combined 3 μ s in cMD for A β 42peptide under *ff14SB* and *ff14IDPSFF* of two monomers. (A) Simulation index A1 (blue) and A1I (orange) for A β 42 $^{\alpha}$ -helix; (B) Simulation index B1 (blue) and B1I (orange) for A β 42 $^{\beta}$ -strand.

Comparison of Force Field Performance With Experimental Data

Comparing the chemical shifts ($C\alpha$, $C\beta$, HN, and N) between the experimental NMR observables^{84,85} and the predicted structures is a robust and straightforward method to evaluate the performance of force fields. With the available chemical shift residue dataset, we compared the HN and $C\alpha$ using simulations from A β 42 $^{\alpha}$ -helix (Figure 2.5 A and B, respectively). The $C\alpha$, $C\beta$ and N from simulations of A β 42 $^{\beta}$ -strand were calculated and compared in Figure 2.5 C, D and E, respectively. The difference in mean RMSD values for the observables between the experimental data and the predicted structures are in Table 2.2. For both A β 42 $^{\alpha}$ -helix and A β 42 $^{\beta}$ -strand monomers, the simulation results and experimental data showed the same trend with the two tested force fields,^{57,66,67} with *ff14IDPSFF* providing better results for NMR observables except for $C\alpha$ in A β 42 $^{\alpha}$ -helix. Sampling was improved with the regular *ff14SB* for $C\alpha$ in A β 42 $^{\alpha}$ -helix. Similar results were observed in other computational simulations.^{38,53} For the $C\alpha$ chemical shift in A β 42 $^{\alpha}$ -helix, the calculated value deviated from the experimental data with *ff14IDPSFF* simulations (Figure 2.5 B). The difference in RMSD was calculated as 0.928 ppm with *ff14IDPSFF* versus 0.194 ppm

with *ff14SB*. Such a deviation in *ff14IDPSFF* corresponds to the parameter optimized by fitting with small peptides and coils, whereas for *ff14SB*, the favorable helicity is obvious when simulating a helical rich conformer by this force field. This was not the case for the C α chemical shift in A β 42 ^{β -strand} (Figure 2.5 C). We observed better results in agreement with the experimental data with *ff14IDPSFF*, with a difference in RMSD of 0.518 ppm versus 0.941 ppm with *ff14SB*. The difference in RMSD value for the C β chemical shift indicated that the calculated values under *ff14IDPSFF* simulation are closer to the experimental data (0.364 ppm with *ff14IDPSFF* vs 0.505 ppm with *ff14SB*). Figure 2.5 E shows that the predicted data for the N chemical shifts for both tested force fields are quite large in simulating A β 42 ^{β -strand}. The difference in RMSD was higher with *ff14IDPSFF* than *ff14SB* (0.666 vs 0.564 ppm).



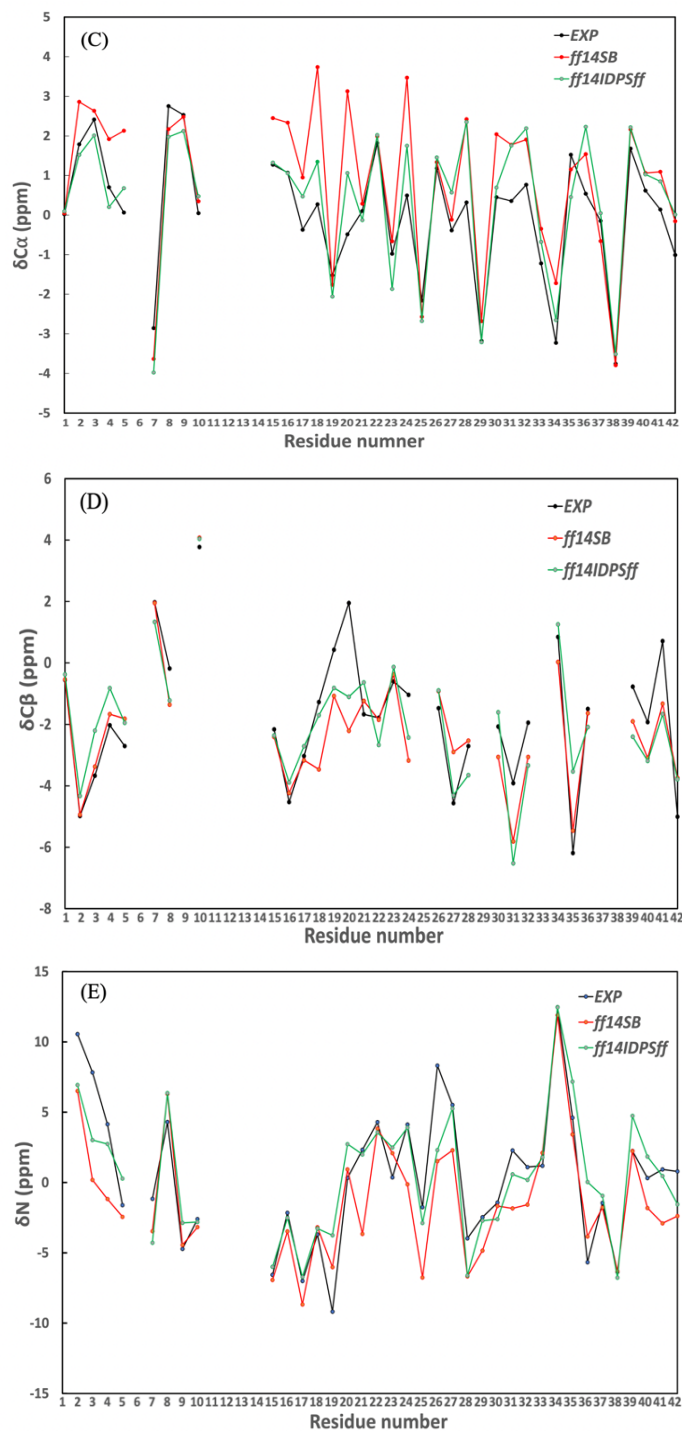


Figure 2.5: Comparison of chemical shifts for HN, $C\alpha$, $C\beta$, and N between NMR experimental data and MD prediction. (A) HN chemical shift for $A\beta_{42}^{\alpha\text{-helix}}$ simulations; (B) $C\alpha$ chemical shift for $A\beta_{42}^{\alpha\text{-helix}}$ simulations; (C) $C\alpha$ chemical shift for $A\beta_{42}^{\beta\text{-strand}}$ simulations; (D) $C\beta$ chemical shift for $A\beta_{42}^{\beta\text{-strand}}$ simulations; (E) N chemical shift for $A\beta_{42}^{\beta\text{-strand}}$ simulation.

Table 2.2 Root mean square deviation differences between simulated and experimental NMR observables in A β 42 ^{α -helix} and A β 42 ^{β -strand}

Chemical Shift (ppm)					
	Force fields	HN	C α	C β	N
A β ₄₂ ^{α-helix}	<i>ff14SB</i>	0.378	0.194	-	-
	<i>ff14IDPSFF</i>	0.338	0.928	-	-
A β ₄₂ ^{β-strand}	<i>ff14SB</i>	-	0.941	0.505	0.564
	<i>ff14IDPSFF</i>	-	0.518	0.364	0.666

To further elucidate the compactness and measure the sizes of our simulated A β 42 structures, we plotted the distribution with coordinates of Rg and RMSD (Figure 2.6) for the whole trajectories. For A β 42 ^{α -helix} simulations, the distribution of structural compactness with *ff14IDPSFF* (Figure 2.6 B) ranged from 8.2 to 23.8 Å for Rg, whereas the distribution with *ff14SB* (Figure 2.6 A) ranged from 8.2 to 22.1 Å for Rg. Also, another distribution of high probability represented a different conformation ranging from 18.0 to 20.0 Å for Rg. This observation indicates a larger populated structural distribution and diverse conformations with *ff14IDPSFF*. However, simulations starting from A β 42 ^{β -strand} showed wider dispersed regions, with mean Rg values ranging from 8.8 to 17.2 Å with *ff14SB* and 8.2 to 19.8 Å with *ff14IDPSFF* (Figure 2.6 C and 2.6 D, respectively). A β 42 ^{α -helix} simulated by *ff14IDPSFF* was able to create more disordered and small structural compactness and flexibility. The calculated Rg values for both the A β 42 ^{α -helix} and A β 42 ^{β -strand} are in Table 2.3, along with the comparison with other studies. Our simulated Rg distribution showed similar with the results from Chakraborty *et al.*⁴⁰ and Krupa *et al.*,³⁸ but our Rg values were ~4.0 to 6.0 Å lower than the value reported by Chakraborty *et al.*

(16.9 ± 0.13 Å)⁴⁰ and ~ 2.0 Å lower than the value reported by Meng *et al.* (15.9 ± 0.1 Å),²⁶ whose results were shown in comparison with fluorescence resonance energy transfer results. The estimated average Rg values for the four systems in our study (11.3 ± 2.0 Å, 13.4 ± 4.0 Å, 10.9 ± 1.0 Å, and 11.2 ± 1.4 Å) agree well with the values reported by Huy *et. al.*,⁸⁶ Carballo-Pacheco *et al.*⁴⁹ and Rahman *et al.*⁵³

Table 2.3: The Average Rg values calculated from all the cMD working systems along with a comparison with other computaitonal studies. The unit of Rg is Å

Current study	Carballo-Pacheco <i>et al.</i> , ⁴⁹	Caliskan <i>et al.</i> , ⁴³	Chakraborty <i>et al.</i> , ⁴⁰	Meng, <i>et al.</i> , ²⁶
11.3 ± 2.0 (A β 42 $^{\alpha}$ -helix, <i>ff14SB</i>)	10.9 ± 1.2 (A β 42, <i>AMBER99SB</i>)	11.9 ± 3.1 (A β 40, <i>a03ws</i>)	16.5 ± 0.12 (A β 40, <i>CG-SOP-IDP</i>)	15.6 ± 0.1 (A β 40, ff99SBws)
13.4 ± 4.0 (A β 42 $^{\alpha}$ -helix, <i>ff14IDPSFF</i>)	10.5 ± 0.8 (A β 42, <i>AMBER99SB*ILD</i> <i>N</i>)	13.6 ± 2.7 (A β 40, <i>a99sb-disp</i>)	16.9 ± 0.13 (A β 42, <i>CG-SOP-IDP</i>)	15.9 ± 0.1 (A β 42, ff99SBws)
10.9 ± 1.0 (A β 42 $^{\beta}$ -strand <i>ff14SB</i>)	10.7 ± 0.7 (A β 42, <i>AMBER99SBILDN</i> <i>-NMR</i>)	16.8 ± 3.7 (A β 40, <i>a99sb-ildn</i>)		
11.2 ± 1.4 (A β 42 $^{\beta}$ -strand, <i>ff14IDPSFF</i>)	10.1 ± 0.8 (A β 42, <i>OPLS</i>)	14.1 ± 3.4 (A β 40, <i>c36m</i>)		
	11.1 ± 1.2 (A β 42, <i>CHARMM22*</i>)	11.0 ± 0.8 (A β 40, <i>a99sb*-ildn</i>)		
		17.6 ± 3.9 (A β 40, <i>a99sb-ucb</i>)		
		12.6 ± 2.2 (A β 40, <i>c22*</i>)		

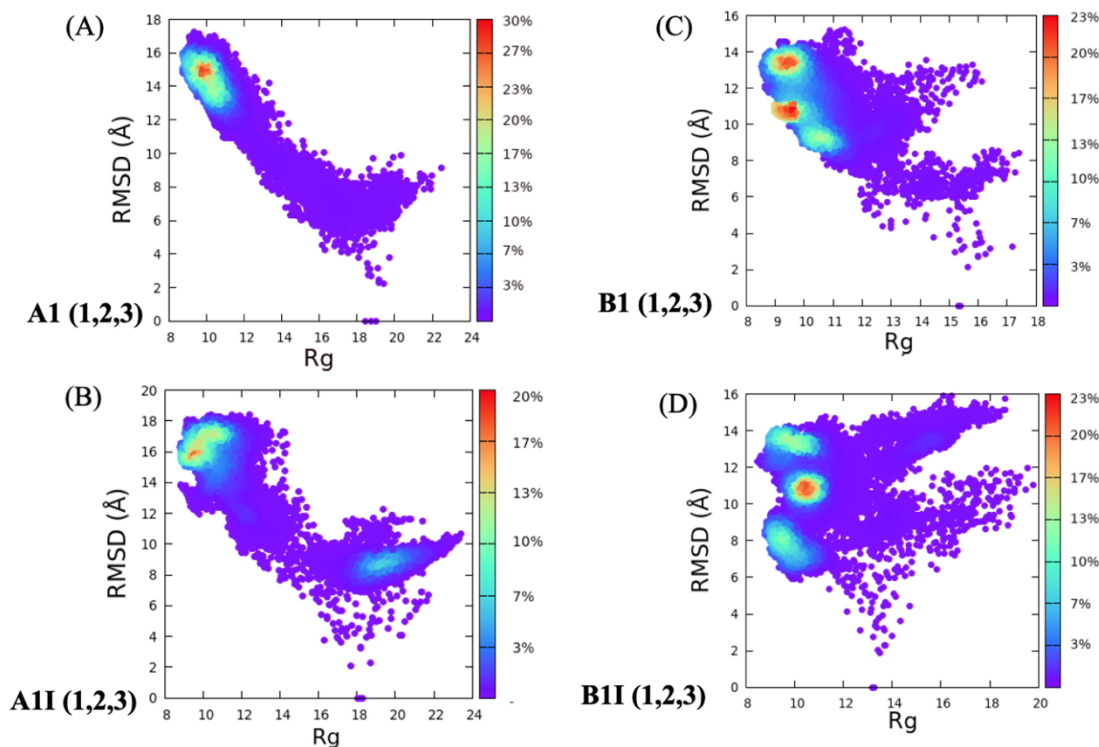


Figure 2.6: Probability distribution on the 2D space of root mean-square deviation (RMSD) and radius of gyration (Rg). Data were combined from three molecular dynamics (MD) runs for each model.

To structurally characterize the conformational interconversion, we examined the secondary structure changes during our MD runs. The time-dependent contents of each secondary structure are in Figures 2.7 and 2.9. The secondary structure content is reported in Figure 2.8.

Investigation of the Transformation of the Secondary Structure in A β 42 to Reveal the Key Features of the Residues

A β 42 $^{\alpha}$ -helix: We analyzed changes in secondary structures during cMD with the two force fields. In general, *ff14IDPSFF* yielded more frequent interconversion between secondary structures than did *ff14SB*. For MD run A1 (1), on the C-terminal, the transition event was

from α -helix to the coil and to the β -sheet within the residues M35, V36, V39, and V40, over ~ 0 to 300 ns (Figure 2.7 A). V36 transitioned from the turn (green) to the β -bridge (brown)/ β -sheet (yellow), which lasted for ~ 150 ns, and transformed back and forth from the turn to 3-10 helix (blue), lasting for ~ 150 ns, before switching back to the β -sheet for another ~ 200 ns. However, V39 and V40 transitioned from the turn/coil to the β -sheet over ~ 0 to 300 ns and retained mostly β -sheets after ~ 470 ns, lasting for ~ 200 ns. G37 and G38 did not form a β -sheet but were more prone to alternate between the turn and 3-10 helix because of lack of side chains. Notably, G37 and G38 were in the coil conformation because they connected the hydrophobic residues in our simulation. Previous research indicated that G37A mutation can disrupt the β -sheet formation.²⁷ These secondary structure transitions of the residues in the C-terminal and its longevity to be in that specific secondary structure may suggest that the region of the conformer is more susceptible to aggregation. In other words, the time it takes for the residue to morph from the α -helix to the β -sheet and how long the β -sheet conformation lasts within the MD run could imply how prone the residue is to aggregation.

In the MD run A1I (1) (Figure 2.7 B), on the C-terminal, although M35-V40 is unique as compared with Figure 4A, residues V36-G38 did not form the β -sheet conformation. Also, V39, V40 and I41 formed the β -sheet conformation later in the MD run, over ~ 300 ns. The V39 and V40 longevity of the β -sheet in Figure 4B being similar to that in Figure 4A could imply that those residues are prone to aggregation.

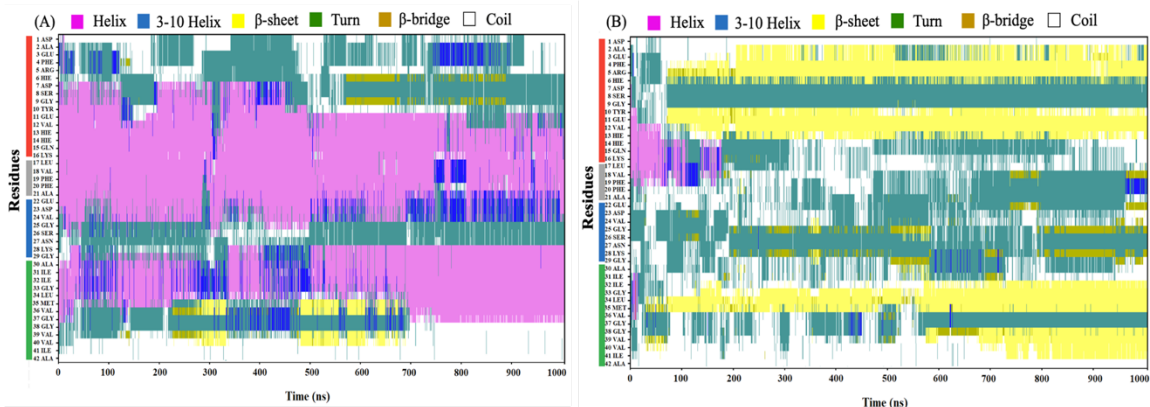


Figure 2.7: Transformation of secondary structures for A β 42 α -helix. (A) (B), Representative trajectory for the time evolution of secondary structures for A β 42 α -helix with *ff14SB* [Simulation index: A1 (1)] and *ff14IDPSFF* [Simulation index: A1I (1)]. (pink: helix; blue: 3-10 short helix; yellow: β -sheet; green: turn, brown: β -bridge, white: coil).

To quantify the content of secondary structures that A β 42 produced in these simulations and gain more in-depth insight into the conformers, we calculated the percentage of helix (red, orange, gray), β -sheet (yellow)/ β -bridge (green), turn (blue), and coil (dark blue) elements as a function of residues within the A β 42 sequence. In Figure 2.8 A and B, the MD run A1 (1) showed that helical elements were more prominent in simulations with *ff14SB* (red). Two segments with more than 50% helicity persisted in the region E11 to V24 and extended from residues K28 to M35, with a small break in the G25, S26 and N27 residues having turn structures. This relatively high content of helical structure was shown in experimental studies.^{87,88} The prominent helical content was also observed at the same peptide regions in recent simulation studies^{43,49,89} The coil elements generally persisted toward the terminus of A β 42. The estimated percentage of helical/turn/coil content in this MD simulation was > 80%, which is similar to the result reported by Krupa *et al.* (~85%) using the same force field in cMD.³⁸ Figure 2.8 B shows that MD run A1I (1) produced a short α -helix and 3-10 helix segments that were approximately $\leq 10\%$ content in the region

H13 to F20 (red and gray) but produced β -sheet elements toward the N- and C-termini. The overall appearance of the β -sheet was $> 80\%$ in the residues spanning I32 to M35, with a long strand appearing at regions Y10-H13 and A2-H6, possibly because of the bulky, alkyl side chain within these residues that therefore could correspond to the increased aggregation rate.³⁵ The overall β -sheet content was estimated at $\sim 30\%$ in the whole sequence, which was also observed in a recent study using the same force field³⁸ and in experiment findings.^{87,90} The central polar region appeared to be mostly coil and turn structures, with a few β -bridge structures at G25 and K28.

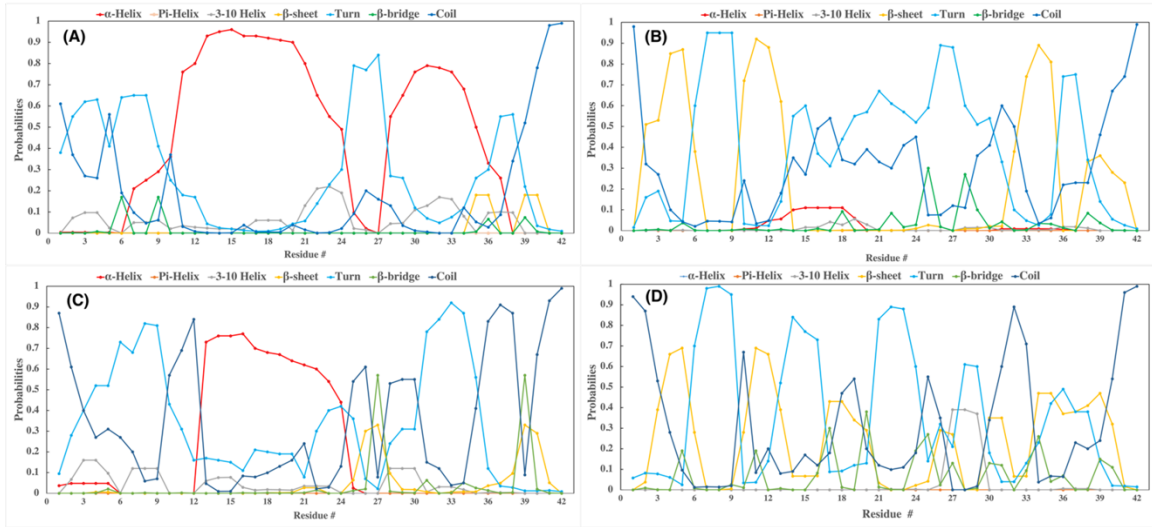


Figure 2.8 (A) (B), Percentage of individual residue with different secondary structures for the $A\beta_{42}^{\alpha\text{-helix}}$ *ff14SB* (simulation index A1 (1)) and *ff14IDPSFF* (simulation index A1I (1)) and, respectively. (C) (D), Percentage of individual residue with different secondary structures for the $A\beta_{42}^{\beta\text{-strand}}$ *ff14SB* (Simulation index B1 (1)) and *ff14IDPSFF* (Simulation index B1I (1)), respectively.

$A\beta_{42}^{\beta\text{-strand}}$: Figure 2.9 A shows that MD B1(1) transitioned from turns to coils to β -strands.

Taking a closer look at the M35-V40 region, M35 started as a turn conformation, then morphed into a coil within certain timeframes, then changed back to turns. That residue

retained the coil secondary structure within the last ~300 ns. V36 started as a turn conformation as well but quickly transformed into the coil secondary structure after ~100 ns and mostly stayed as the coil structure for the rest of the MD run. G37 and G38 started as a turn secondary structure but morphed to a coil throughout the MD run. V39 and V40 were mostly a mixed form with β -bridges/ β -sheets and coils throughout the MD run. These two residues transformed back and forth between β -sheets and β -bridges until the latter part of the MD run, when V40 transformed to coils for the last ~300 ns. V39 and V40 being hydrophobic and more prone to forming and maintaining the β -sheets and β -bridge secondary structures may imply that these residues are prone to aggregation.

In Figure 2.9 B, M35-V40 showed a trend of transitioning from turns to β -sheets. M35 started as a turn conformation, then morphed into a β -sheet, which lasted for ~100 ns. Then, M35 transformed into coils and turns for a short time until it transformed and sustained the β -sheet structure for ~300 ns. Finally, it morphed back to turns and transitioned to β -sheets, which lasted about ~150 ns. Likewise, V36 showed a pattern similar to that for M35 but with more small transitions to β -sheets at the beginning of the MD run. G37 and G38 started as a coil but transformed back and forth from turns to β -sheets during the MD run. V39 and V40 mainly started as β -bridges and coils, respectively, but had a similar pattern (Figure 2.9 A). These residues sustained a ~300-ns β -sheet structure (timeframe 450-730 ns), which suggests that the region was prone to aggregation. M35, in particular, was more prone to forming β -sheets later in the MD run, so it may be prone to aggregation.

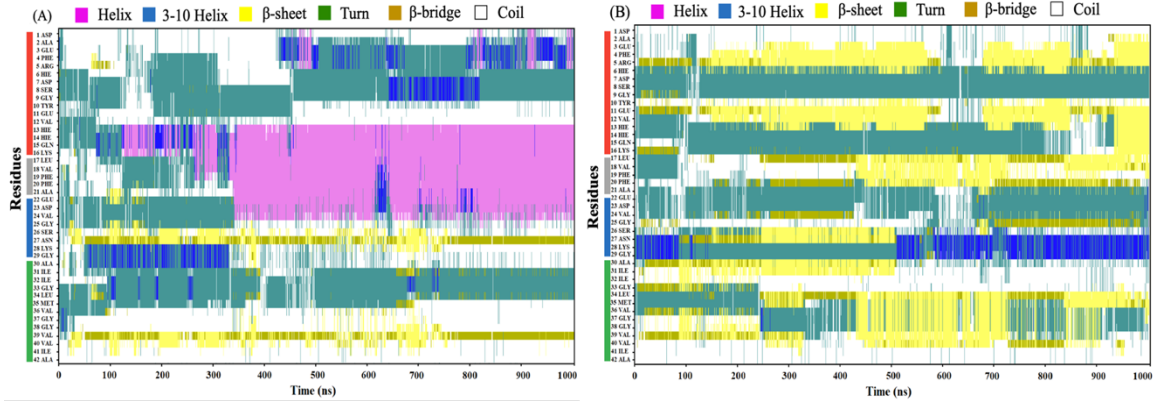


Figure 2.9: Transformation of secondary structures for A β 42 β -strand. (A) (B), Representative trajectory for the time evolution of secondary structures for A β 42 β -strand with *ff14SB* [Simulation index: B1 (1)] and *ff14IDPSFF* [Simulation index: B1I (1)]. (pink: helix; blue: 3-10 short helix; yellow: β -sheet; green: turn, brown: β -bridge, white: coil).

Again, we further quantified the percentage of secondary structures by plotting the propensity of each residue. Residues V36 to I41 of the C-terminal showed less β -sheet content (Figure 2.8 C) and sparse content in the polar region spanning S25 to K28 in MD B1(1). The remaining residues in the sequence were dominated by helix, coil and turn structures. In Figure 2.8 D, MD run B1I (1) showed a high ratio of the β -sheet appearing mainly in the regions located at the N- and C-termini and central hydrophobic region. The high β -sheet content observed in our MD study was in good agreement with Rosenman *et al.*⁴⁷ and Lincoff *et al.*⁸⁹ but not with another study reporting helix and β -sheet contents of 11.93% and 10.6% with *ff14IDPSFF*, thus indicating a quantitative balance in helix and β -sheet content for their α -helix-rich conformer.⁵³ Previously, *ff14SB* was believed to overestimate helix content,^{38,50,91} so this force field may simulate more helix conformers in A β 42 monomers, although we selected two distinct monomeric states. Force field

favorability in secondary structure determination remains a substantial problem because the A β 42 simulation strongly depends on the force field selection and simulation protocol.

Mapping Residue Contacts to Characterize the Self-Assembly Regions

To characterize the intramolecular interactions within the monomers, we created intrapeptide contact maps to show the probabilities of the backbone contacts. These maps reveal the characteristic monomeric tertiary structures of A β 42 defined by four different regions: N-terminal region (N) (D1–K16), hydrophobic central region (CH) (L17–A21), central polar region (P) (E22–I31) and C-terminal region (C) (I32–A42).

In our analysis, two residues were in contact on this map if their C α –C α distance was ≤ 7.5 Å. Figures 2.10 A and B compare the two force fields for A β 42 $^{\alpha}$ -helix. With *ff14IDPSFF* simulation [Figure 2.10 B, MD run A1I (1,2,3)], there was high-probability contact between residues A2-S8 and E11-Q15, residues G9-H14 and the CH region, the CH region and V31-M35, and the N and C regions. In contrast, with *ff14SB* simulation, there was sparse contact in these regions [Figure 2.10 A, MD run A1 (1,2,3)], so the backbone contact may not be obvious in our simulation with the cutoff criteria. Also, the polar region P had the least contact with the two force fields, so this region could be very dynamic. Our findings agree with Rosenman *et al.*,⁴⁷ who calculated the tertiary structure with OPLS and AMBER99sb-ILDN for A β 42 and CHARMM22* for A β 40. Figures 2.10 C and D compare the two force fields for A β 42 $^{\beta}$ -strand. Similar to A1I, as shown in Figure 2.10 D, the MD run B1I (1,2,3) shows a high probability of residue contact between residues A2-S8 and E11-Q15, residues G9-H14 and the CH region, N and P regions, CH and P regions,

CH and C regions, and residues E11-Q15 and the C region. The high contact ratio between residues could lead to the increased breadth of the β -sheets after in-register stacking within this monomer. With *ff14SB* simulation, in the β -strand shown in Figure 2.10 C, MD run B1 (1,2,3), we did not observe close contact between N and P regions; however, we observed contacts between N and CH regions, CH and P regions, and P and C regions. These regions may be prone to forming short β -hairpins and discrete segments of parallel β -strands rather than stacking in-register β -sheet conformations. These short β -hairpins and discrete segments of parallel β -strands were found linked to early-onset aggregation.⁹²⁻

94

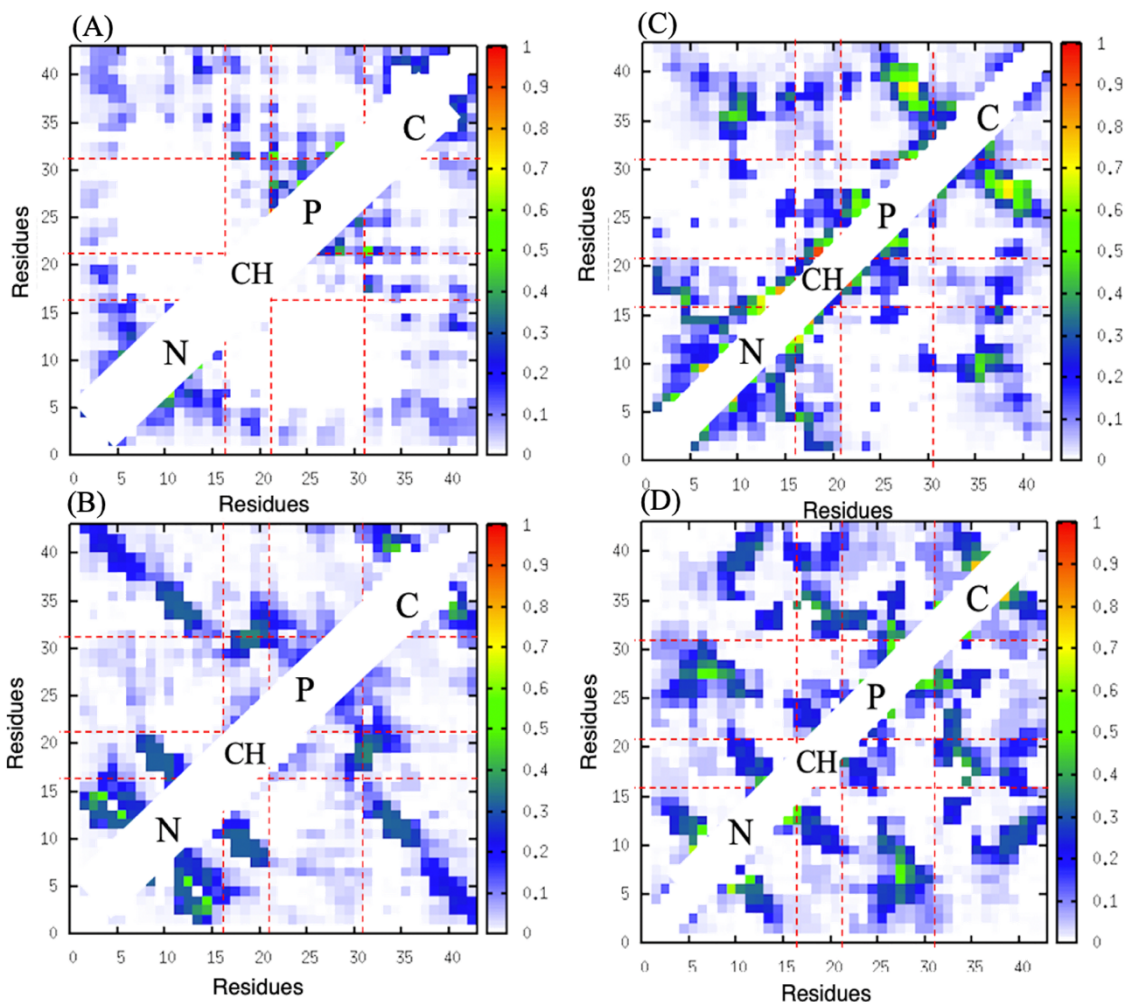


Figure 2.10: Probability of intrapeptide contact map. (A) and (B) A β 42 $^{\alpha}$ -helix [Simulation index: A1 (1,2,3) and A1I (1,2,3), respectively]; (C) and (D) A β 42 $^{\beta}$ -strand [Simulation index: B1 (1,2,3) and simulation index B1I (1,2,3), respectively]. N-terminal region (N, residues 1–16), hydrophobic central region (CH, residues 17–21), central polar region (P, residues 22–31), and C-terminal region (C, residues 32–42).

Comparison of Conformation Clusters Sampled by the Two Force Fields in A β 42 to Reveal the Fingerprints of the Fibril Structure

To compare the difference in conformer distribution with the two force fields and to glean further insights into the heterogeneity of A β 42 structures, we used conformer clustering. We clustered conformations sampled from cMD, and each analysis combined MD runs with 3 random number seeds to obtain a total of 30000 snapshots. The MD runs A1 (1, 2,

3) and A1I (1, 2, 3) generated 104 and 208 clusters, respectively. When MD runs were initiated with the $A\beta_{42}^{\beta\text{-strand}}$ conformation, we obtained fewer clusters: 95 and 106 clusters from simulations B1 (1, 2, 3) and B1I (1, 2, 3), respectively (Figure 2.11). The population of monomeric states in $A\beta_{1-42}$ highly depended on the force field used. Hence, *ff14IDPSFF* strongly favored the disordered states and therefore a high ratio of clusters. The inset in Figure 2.11 shows the population percentage of the top 20 clusters. Initiated by a β -strand conformation, *ff14SB* sampled a few conformations with a large population, whereas *ff14IDPSFF* produced more heterogenous conformers with a similar population. Similar results can be seen in MD runs A1I (1, 2, 3) and A1 (1, 2, 3). *ff14SB* sampled a few highly popular conformations as compared with the clusters generated by *ff14IDPSFF*, although the differences were not as pronounced as runs initiated by a β -strand conformation.

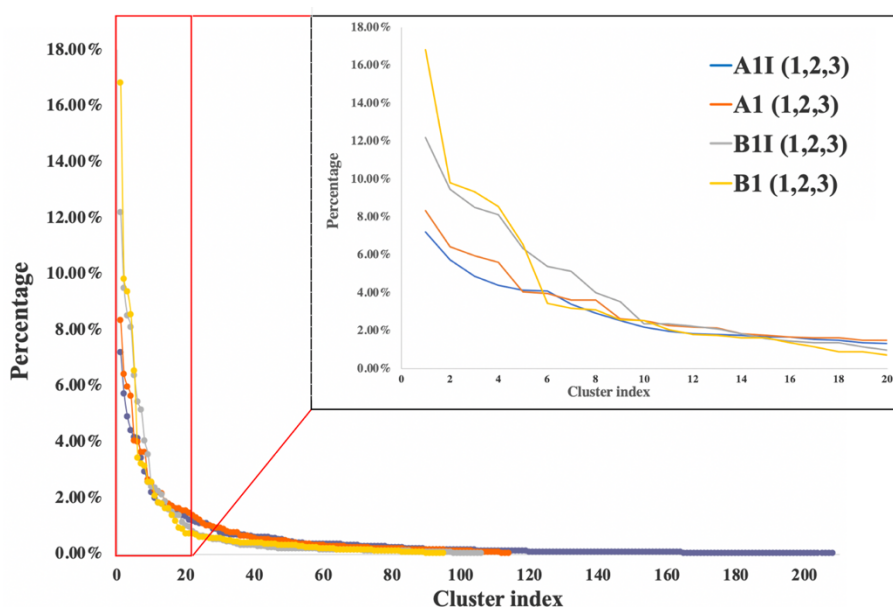


Figure 2.11: Cluster index versus the percentage of occupied conformers for all evaluated working systems. (Inset: slope with the top 20 clusters).

The difference in the monomeric states of A β 42 can be further visualized by checking the most popular conformations obtained with the two force fields. Figures 2.12 A and B present the eight largest clusters: the top eight clusters occupied 38.2% and 33.3% of the total conformer population from A1 (1,2,3) and A1I (1,2,3), respectively. Starting with A β 42 ^{α -helix}, β -sheets and β -strands were generated with *ff14IDPSFF* [A1I (1,2,3)], mainly located at the N- and C-termini as well as the CH region. The central polar region was dominated by coil and turn conformations. The preference for β -sheets and anti-parallel β -hairpins is not uncommon: they were also observed and reported in studies using GROMOS ffG53a6,²⁹ AMBER99SB and OPLS,⁴⁹ and *ff14IDPSFF* force fields.³⁸

With *ff14SB*, A β 42 ^{α -helix} produced a large number of helical structures, thus indicating little change throughout the simulations. Such rich helix conformation was also observed in the studies of Gerben *et al.*, using AMBER 03 and CHARMM22+CMAP,⁴⁶ and Carballo-Pacheco *et al.*, using the AMBER99SBILDN-NMR force field.⁴⁹ The exception is in the C-terminal, which showed a short β -hairpin and short 3-10 helix with about the same percentage, which became the dominant conformation when using GROMOS9653a6⁴⁶ and GROMOSffG43a2.²⁹ In the cluster conformation obtained in Rahman *et al.*, using the same initial structure and force field, among the top three conformations reported with their six tested force fields, only *ff03w* conferred an anti-parallel β -hairpin conformation. The remaining conformations were disordered or had helix conformations,⁵³ which could be due to simulation protocol or analysis criteria differences.

For A β ₄₂ ^{β -strand}, the top eight clusters occupied 60.9% of the total conformer population (Figure 2.12 C) and 53.0% of the total conformer population (Figure 2.12 D) for B1 (1,2,3) and B1I (1,2,3), respectively. With *ff14IDPSFF*, B1I (1,2,3), A β ₄₂ ^{β -strand} maintained a large percentage of β -sheet and coil structures in the top eight clusters, so *ff14IDPSFF* produced a reasonably heterogeneous ensemble of structures dominated by β -hairpin and anti-parallel β -sheets. With *ff14SB*, MD run B1 (1,2,3), A β ₄₂ ^{β -strand} maintained much higher percentages within each cluster, so the force field sampled a small number of structures. The distinct structures were attributed mostly to the formation of short helix, coil and turn conformations that encompass most of the A β ₁₋₄₂ sequence. Surprisingly, a high percentage of β -hairpins and β -sheets in anti-paralleled orientations were also observed in the C-terminus and extended to residue G25 (9.8% in cluster #2), which agrees well with the NMR study.⁹⁵

Our cluster conformation analysis revealed some fingerprints linked to the onset of early fibril-like structures, which were also seen in other studies.^{40,42} This observation may suggest that self-assembly could be initiated in the C-terminal because of mostly non-polar residues.

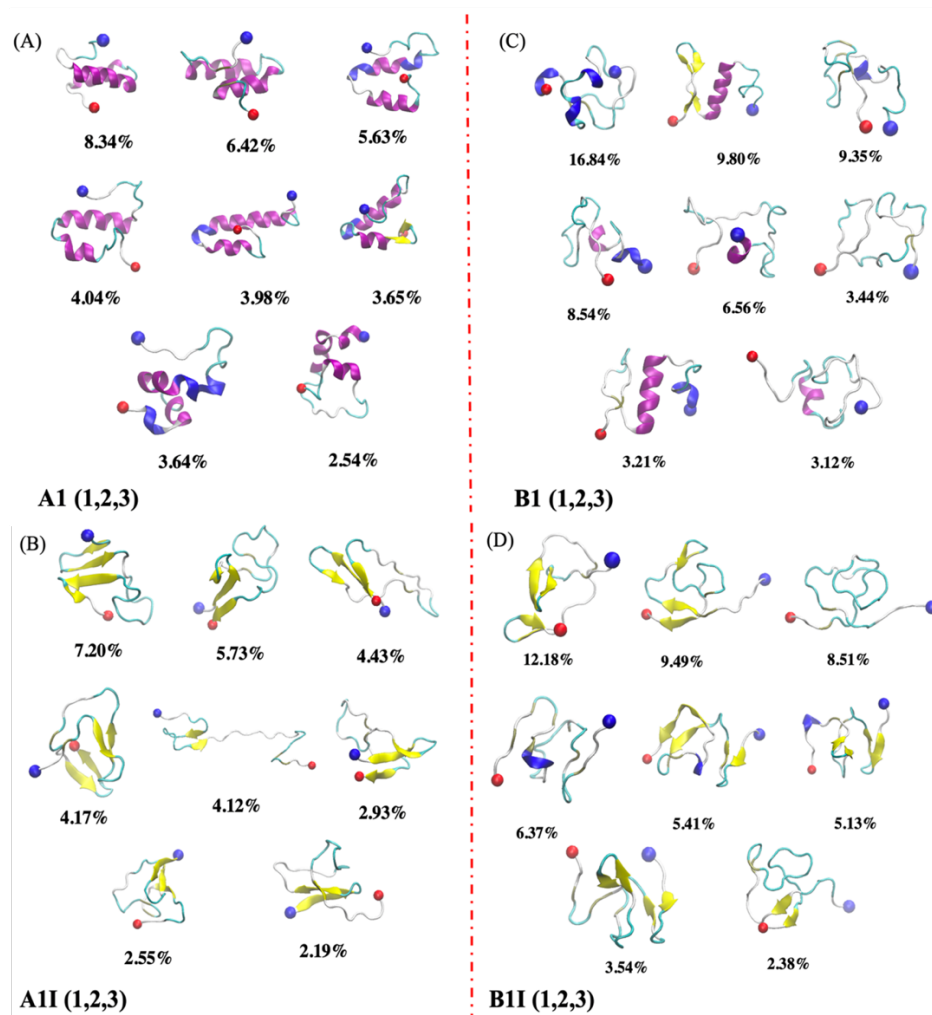


Figure 2.12: Representative conformers of the top 8 clusters and their occupied percentage derived from simulations with the two force fields *ff14SB* and *ff14IDPSFF*. (A) and (B) Aβ42 α -helix with *ff14SB* and *ff14IDPSFF*, respectively. (C) and (D) Aβ42 β -strand with *ff14SB* and *ff14IDPSFF*, respectively.

Conformers can be trapped easily in a local energy minimum when using cMD. To overcome this issue, we extended the sampling by performing aMD simulations, with which we obtained some conformations that may not be seen in cMD. Adding aMD in simulation studies can reproduce structures and results in close agreement with experiments in other studies.⁹⁶⁻⁹⁸ We simulated 100 ns for each system but acknowledged

that our systems did not reach convergence in such a short time because convergence depends on the property of interest. Among three different boost potentials, 0.2, 0.5 and 1.0, we further characterized the conformers of clustering by selecting the simulation performed under the 0.5 boost potential. The rationale is based on the trajectories of the timeline for the secondary structures. The 0.2 boost potential in aMD overestimated the structured conformers, such as the helix and sheet structures, whereas the 1.0 boost potential in aMD overestimated the disordered conformers, such as the turn and coil structures. By justifying the suitability of the 0.5 boost potential in the aMD for A β 42, we applied the K-means clustering method for the trajectories of all aMDs and obtained conformational diversity. With *ff14IDPSFF* used for both MD runs A3I (1,2,3) and B3I (1,2,3), structurally, with a few exceptions, the clusters exhibited an extended length of β -strand contents in their top 5 clusters but low helical content for the α -helix conformer [A3 (1,2,3)] and a high degree of unstructured coils and short segments of helical content for A β 42 ^{β -strand} [B3 (1,2,3)]. As a result, the conformers obtained with the aMD simulations may be supplemental to the conformational ensembles, which could not be revealed by cMD simulation.

Investigation of Solvent Accessible Surface Area (SASA) and Two Key Salt Bridges, E22/K28 and D23/K28, to Reveal Solvent Effect in Structure Stability and Conformation

We conducted SASA analysis to investigate the solvent effect in structure conformation. The distribution of SASA values was relatively broader and right-shifted for both monomers by using *ff14IDPSFF* (Figure 2.13). Particularly, the result for A β 42 ^{α -helix} simulations had another significant peak near 41 nm², which indicates a fraction of

extended conformations exposed to the solvent. This finding might suggest that the *ff14IDPSFF* force field was able to simulate some particular unfolded states, which may not be feasible with *ff14SB*. The estimated SASA values of 32.8, 33, 32.2 and 33.3 nm² for all four working systems were in good agreement with other studies^{38,86}; a value close to 41 nm² was reported by Duan *et al.*,⁹⁹ using *AMBERff03*, and Watts *et al.*,¹⁰⁰ using the A β 40 peptide with CHARMM36. Finally, we measured the salt bridge distance formed by two pairs of residues, E22/K28 and D23/K28, which were established by NMR studies. The salt bridge formed between these residues stabilized the A β 42 peptide, which then in the later stage, formed A β fibrils.^{101,102} We set a cutoff of 4.0 Å between the positive charge atom N involving the sidechain K28 and the negative charge atom O involving the sidechain E22 or D23. We obtained a total of 3000-ns time lengths by combining three MD runs for each system. To find the frequency of the salt bridge formation throughout all MD simulations, we calculated the average estimated time for salt bridge formation over the total run time of the MD simulation, as exemplified in equation 1 below:

$$\frac{\Sigma(\text{estimated time length of salt bridge formation})}{(\text{total MD simulation runtime})} * 100\% \quad \text{Eq. 2.1}$$

Figure 2.14 showed the distance of salt bridges E22/K28 and D23/K28, respectively, over 3000 ns. Using equation 1 with *ff14SB*, the estimated total time of salt bridge formation was ~1100 ns over the total run time (3000 ns), for a ~37% frequency of E22/K28 salt bridge formation. In contrast, with *ff14IDPSFF*, the frequency of E22/K28 salt bridge formation was ~9% of run time (orange line [A1I (1,2,3)]). D23/K28 (Figure 2.14B) formed a salt bridge infrequently, at ~1.3% (blue) with *ff14SB* and ~15% (orange) with *ff14IDPSFF*. Starting from A β 42 ^{β -strand}, the E22/K28 salt bridge was unfavorable with the

two force fields (yellow and gray lines), but the frequency of D23/K28 with *ff14IDPSFF* was increased, ~10% [B1I (1,2,3)]. Thus, salt bridge interactions between E22/K28 and D23/K28 should exist alternately to dominate the stabilization of the central polar region, thus resulting in large conformational changes and accessing the solvent dynamically.

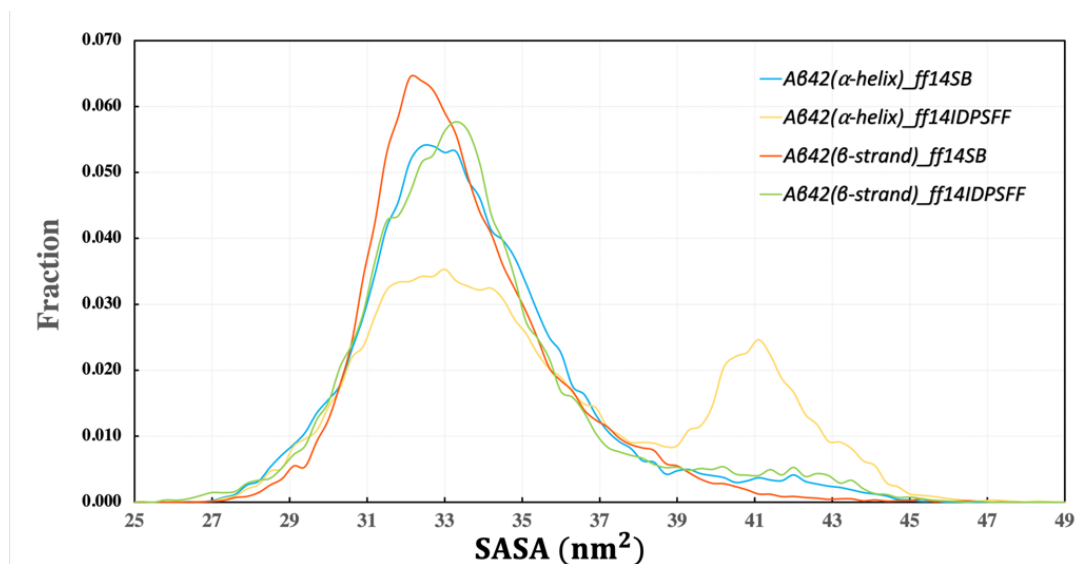


Figure 2.13: Distributions of structures with given solvent accessible surface areas (SASAs) for the working systems.

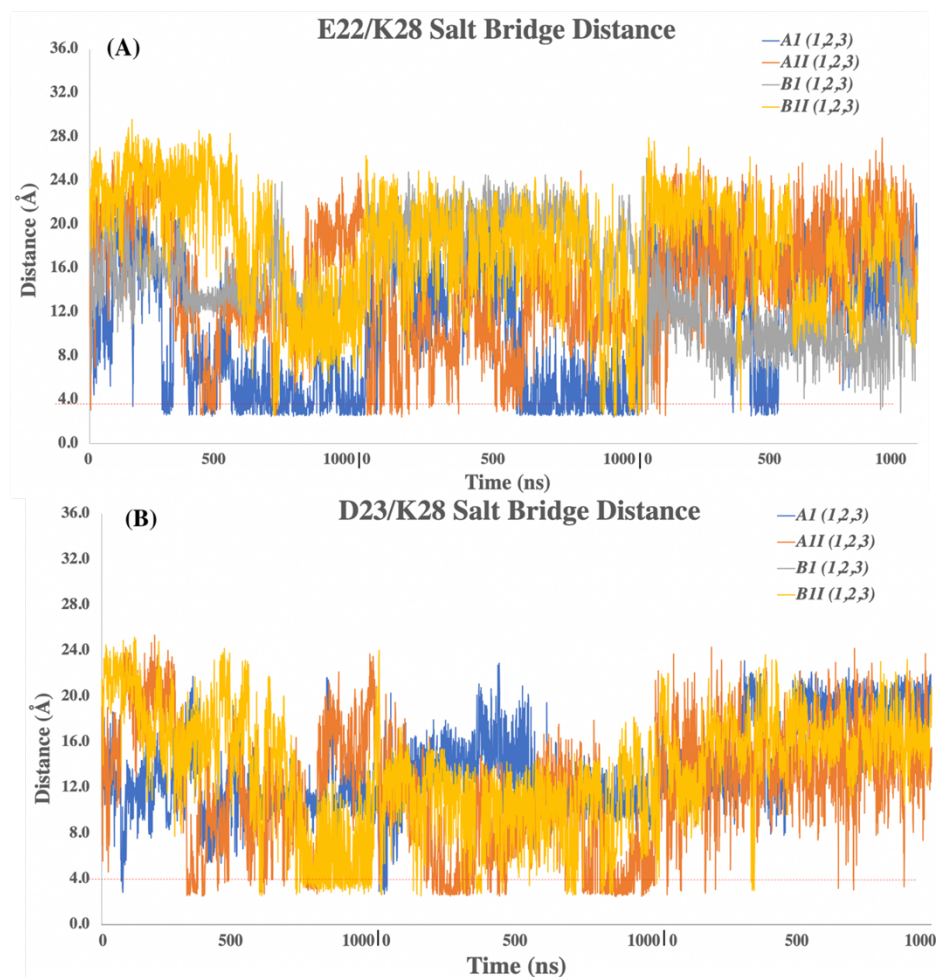


Figure 2.14: Salt bridge distance measurements as a function of simulation timeline. (A) E22/K28; (B) D23/K28. Red dash line represents the cutoff distance of 4.0 Å.

2.5 Conclusions

Because of their highly flexible properties, use of IDP-specific force fields is essential to accurately model conformations and transitions of IDPs.^{56,57,103} Hence, studies have examined the conformations and protein dynamics sampled by using experiments and molecular simulations of amyloid peptide.^{31,34,38,47,49,50,53,81,82,104-106} Here we used a recently developed but not extensively examined IDP-specific force field, *ff14IDPSFF*, to sample conformations of A β 42 monomers and compared the conformers obtained by a classical

force field, *ff14SB*, by both cMD and aMD. Our study clearly showed that the heterogeneity, dynamics and conformations of A β 42 heavily depended on the initial structure and force field used. Of note, system convergence depends on the property of interest, and our purpose was to ensure a fair comparison in the simulation protocols for two force fields with two starting structures. We calculated the predicted structural chemical shifts and compared the CS observables with the NMR experiments. The force field *ff14IDPSFF* performed qualitatively well in all tested CSs except for C α in A β 42 $^{\alpha\text{-helix}}$, whereas *ff14SB* reproduced the C α CS observable qualitatively well in A β 42 $^{\alpha\text{-helix}}$. We analyzed the probability distribution of the Rg, secondary structural contents, and cluster conformers. Our Rg and secondary structural content values modeled with both *ff14IDPSFF* and *ff14SB* force fields are in qualitative agreement with findings from other studies.^{26,38,40} One key finding noted in our work is that starting with A β 42 $^{\alpha\text{-helix}}$, the *ff14IDPSFF* force field sampled a high ratio of disordered structures and diverse β -sheet secondary structures, whereas *ff14SB* simulated high helical content in the MD simulations. This co-occurrence was also observed in other similar studies.^{38,53} The conformation obtained from the A β 42 $^{\beta\text{-strand}}$ MD simulations maintained a balanced content in the disordered and helical structures with *ff14SB*, but the conformers clearly favored disordered and β -sheet structures with *ff14IDPSFF*. Detailed analysis in the contact map and cluster conformation showed that the C-terminal region (mainly M35-V40) contained in-registered anti-parallel β -sheet (β -hairpin) conformations in both monomers with both force fields. The force field *ff14IDPSFF* sampled more diverse conformations, which could be associated with self-assembly of onset of early fibril-like structures. SASA and salt

bridge analysis (E22/K28 and D23/K28) revealed the simulated unfolded structure, which fell into the central polar region in the peptide.

This study compared two distinct initial structures simulated by two force fields, and the IDP force field *ff14IDPSFF* is suitable to simulate flexible A β 42 monomers. However, further tests for force field improvement are still needed to precisely describe the IDP aggregation, for instance, A β 42 dimerization and polymerization. Considering other factors such as solvent models and simulation protocols (simulation time lengths, temperature condition, *etc.*) is also required to improve the accuracy of MD simulation for better modeling the A β 42 peptide and other IDPs.

2.6 References

1. Caughey B, Lansbury PT. Protofibrils, pores, fibrils, and neurodegenerations: separating the responsible protein aggregates from the innocent bystanders. *Annual review of neuroscience* 2003;26(1):267-298.
2. Chiti F, Dobson CM. Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. *Annual review of biochemistry* 2017;86:27-68.
3. Han X, He G. Toward a rational design to regulate β -amyloid fibrillation for alzheimer's disease treatment. *ACS chemical neuroscience* 2018;9(2):198-210.
4. Panza F, Lozupone M, Logroscino G, Imbimbo BP. A critical appraisal of amyloid- β -targeting therapies for Alzheimer disease. *Nature Reviews Neurology* 2019;15(2):73-88.
5. Cline EN, Bicca MA, Viola KL, Klein WL. The amyloid- β oligomer hypothesis: Beginning of the third decade. *Journal of Alzheimer's Disease* 2018;64(s1):S567-S610.
6. Lee SJC, Nam E, Lee HJ, Savelieff MG, Lim MH. Towards an understanding of amyloid- β oligomers: characterization, toxicity mechanisms, and inhibitors. *Chemical Society Reviews* 2017;46(2):310-323.
7. Qiang W, Yau W-M, Lu J-X, Collinge J, Tycko R. Structural variation in amyloid- β fibrils from Alzheimer's disease clinical subtypes. *Nature* 2017;541(7636):217-221.
8. Zhang M, Ren B, Chen H, Sun Y, Ma J, Jiang B, Zheng J. Molecular Simulations of Amyloid Structures, Toxicity, and Inhibition. *Israel Journal of Chemistry* 2017;57(7):586-601.
9. Finder VH, Vodopivec I, Nitsch RM, Glockshuber R, Forel-strasse A. The recombinant amyloid- β peptide A β 1-42 aggregates faster and is more neurotoxic than synthetic A β 1-42. *Journal of Molecular Biology* 2010;396(1):9-18.
10. Meisl G, Yang X, Hellstrand E, Frohm B, Kirkegaard JB, Cohen SIA. Differences in nucleation behavior underlie the contrasting aggregation kinetics of the A β 40 and A β 42 peptides. *Proceedings of the National Academy of Science* 2014;111(26):9384-9389.
11. Cohen SIA, Arosio P, Presto J, Kurudenkandy FR, Biverstål H, Dolfe L, Dunning C, Yang X, Frohm B, Vendruscolo M and others. A molecular chaperone breaks the

catalytic cycle that generates toxic A β oligomers. *Nature structural & molecular biology* 2015;22(3):207-213.

12. Maynard CJ, Bush AI, Masters CL, Cappai R, Li QX. Metals and amyloid- β in Alzheimer's disease. *International journal of experimental pathology* 2005;86(3):147-159.

13. Adlard PA, Bush AI. Metals and Alzheimer's disease. *Journal of Alzheimer's disease* 2006;10(2-3):145-163.

14. Liao Q, Owen MC, Olubiyi OO, Barz B, Strodel B. Conformational Transitions of the Amyloid- β Peptide Upon Copper (II) Binding and pH Changes. *Israel Journal of Chemistry* 2017;57(7-8):771-784.

15. Roberts BR, Ryan TM, Bush AI, Masters CL, Duce JA. The role of metallobiology and amyloid- β peptides in Alzheimer's disease. *Journal of neurochemistry* 2012;120:149-166.

16. Liu R, McAllister C, Lyubchenko Y, Sierks MR. Residues 17–20 and 30–35 of beta-amyloid play critical roles in aggregation. *Journal of neuroscience research* 2004;75(2):162-171.

17. Petkova AT, Ishii Y, Balbach JJ, Antzutkin ON, Leapman RD, Delaglio F, Tycko R. A structural model for Alzheimer's β -amyloid fibrils based on experimental constraints from solid state NMR. *Proceedings of the National Academy of Sciences* 2002;99(26):16742-16747.

18. Petkova AT, Yau W-m, Tycko R. Experimental constraints on quaternary structure in Alzheimer's β -Amyloid fibrils. *Biochemistry* 2006;45(2):498-512.

19. Luhers T, Ritter C, Adrian M, Riek-loher D, Bohrmann B, Do H, Schubert D, Riek R. 3D structure of Alzheimer's amyloid- β (1–42) fibrils. *Proceedings of the National Academy of Sciences* 2005;102(48):17342-17347.

20. Olofsson A, Lindhagen-persson M, Sauer-eriksson AE, Ohman A. Amide solvent protection analysis demonstrates that amyloid- β (1–40) and amyloid- β (1–42) form different fibrillar structures under identical conditions. *Biochemical Journal* 2007;404(1):63-70.

21. Kirschner DA, Abraham C, Selkoe D. X-ray diffraction from intraneuronal paired helical filaments and extraneuronal amyloid fibers in Alzheimer disease indicates cross-beta conformation. *Proceedings of the National Academy of Sciences* 1986;83(2):503-507.

22. Schmidt M, Rohou A, Lasker K, Yadav JK, Schiene-fischer C, Fändrich M, Gringorieff N. Peptide dimer structure in an A β (1-42) fibril visualized with cryo-EM. *Proceedings of the National Academy of Sciences* 2015;112(38):11858-11863.
23. Antzutkin ON, Leapman RD, Balbach JJ, Tycko R. Supramolecular structural constraints on Alzheimer's β -Amyloid fibrils from electron microscopy and solid-state nuclear magnetic resonance. *Biochemistry* 2002;41(51):15436-15450.
24. Kheterpal I, Zhou S, Cook KD, Wetzel R. A β amyloid fibrils possess a core structure highly resistant to hydrogen exchange. *Proceedings of the National Academy of Sciences* 2000;97(25):13597-13601.
25. Whittemore NA, Mishra R, Kheterpal I, Williams AD, Wetzel R, Serpersu EH. Hydrogen– Deuterium (H/D) Exchange Mapping of A β 1-40 Amyloid Fibril Secondary Structure Using Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* 2005;44(11):4434-4441.
26. Meng F, Bellaiche MM, Kim J-Y, Zerze GH, Best RB, Chung HS. Highly disordered amyloid- β monomer probed by single-molecule FRET and MD simulation. *Biophysical journal* 2018;114(4):870-884.
27. Xu Y, Shen J, Luo X, Zhu W, Chen K, Ma J, Jiang H. Conformational transition of amyloid β -peptide. *Proceedings of the National Academy of Sciences* 2005;102(15):5403-5407.
28. Sgourakis NG, Merced-Serrano M, Boutsidis C, Drineas P, Du Z, Wang C, Garcia AE. Atomic-level characterization of the ensemble of the A β (1–42) monomer in water using unbiased molecular dynamics simulations and spectral algorithms. *Journal of molecular biology* 2011;405(2):570-583.
29. Olubiyi OO, Strodel B. Structures of the amyloid β -peptides A β 1–40 and A β 1–42 as influenced by pH and a D-peptide. *The Journal of Physical Chemistry B* 2012;116(10):3280-3291.
30. Ilie IM, Caflisch A. Disorder at the Tips of a Disease-Relevant A β 42 Amyloid Fibril: A Molecular Dynamics Study. *Journal of Physical Chemistry B* 2018;122(49):11072-11082.
31. Robustelli P, Piana S, Shaw DE. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences* 2018;115(21):E4758-E4766.
32. Zhang Y, Hashemi M, Lv Z, Lyubchenko YL. Self-assembly of the full-length amyloid A β 42 protein in dimers. *Nanoscale* 2016;8(45):18928-18937.

33. Grasso G, Rebella M, Morbiducci U, Tuszynski JA, Danani A, Deriu MA. The Role of Structural Polymorphism in Driving the Mechanical Performance of the Alzheimer's Beta Amyloid Fibrils. *Frontiers in Bioengineering and Biotechnology* 2019;7(April):1-11.
34. Nguyen HL, Krupa P, Hai NM, Linh HQ, Li MS. Structure and Physicochemical Properties of the A β 42 Tetramer: Multiscale Molecular Dynamics Simulations. *The Journal of Physical Chemistry B* 2019;123(34):7253-7269.
35. Thu TTM, Co NT, Tu LA, Li MS. Aggregation rate of amyloid beta peptide is controlled by beta-content in monomeric state. *The Journal of chemical physics* 2019;150(22):225101.
36. Alves NA, Frigori RB. Structural Interconversion in Alzheimer's Amyloid- β (16–35) Peptide in an Aqueous Solution. *The Journal of Physical Chemistry B* 2018;122(6):1869-1875.
37. Masutani K, Yamamori Y, Kim K, Matubayasi N. Free-energy analysis of the hydration and cosolvent effects on the β -sheet aggregation through all-atom molecular dynamics simulation. *The Journal of chemical physics* 2019;150(14):145101.
38. Krupa P, Quoc Huy PD, Li MS. Properties of monomeric A β 42 probed by different sampling methods and force fields: Role of energy components. *The Journal of Chemical Physics* 2019;151(5):055101.
39. Lin Y, Im H, Diem LT, Ham S. Characterizing the structural and thermodynamic properties of A β 42 and A β 40. *Biochemical and Biophysical Research Communications* 2019;510(3):442-448.
40. Chakraborty D, Straub JE, Thirumalai D. Differences in the free energies between the excited states of A β 40 and A β 42 monomers encode their aggregation propensities. *Proceedings of the National Academy of Sciences* 2020;117(33):19926-19937.
41. Nguyen PH, Derreumaux P. Structures of the intrinsically disordered A β , tau and α -synuclein proteins in aqueous solution from computer simulations. *Biophysical Chemistry* 2020:106421.
42. Barz B, Buell AK, Nath S. Compact fibril-like structure of amyloid β -peptide (1–42) monomers. *Chemical Communications* 2021;57(7):947-950.
43. Caliskan M, Mandaci SY, Uversky VN, Coskuner-Weber O. Secondary Structure Dependence of Amyloid- β (1-40) on Simulation Techniques and Force Field Parameters. *Chemical Biology & Drug Design* 2021.

44. Bhattacharya S, Xu L, Thompson D. Long-range Regulation of partially folded Amyloidogenic peptides. *Scientific reports* 2020;10(1):1-17.
45. Hashemi M, Zhang Y, Lv Z, Lyubchenko YL. Spontaneous self-assembly of amyloid β (1–40) into dimers. *Nanoscale Advances* 2019;1(10):3892-3899.
46. Gerben SR, Lemkul JA, Brown AM, Bevan DR. Comparing atomistic molecular mechanics force fields for a difficult target: a case study on the Alzheimer's amyloid β -peptide. *Journal of Biomolecular Structure and Dynamics* 2014;32(11):1817-1832.
47. Rosenman DJ, Wang C, García AE. Characterization of A β monomers through the convergence of ensemble properties among simulations with multiple force fields. *The Journal of Physical Chemistry B* 2016;120(2):259-277.
48. Smith MD, Rao JS, Segelken E, Cruz L. Force-field induced bias in the structure of A β 21–30: A comparison of OPLS, AMBER, CHARMM, and GROMOS force fields. *Journal of chemical information and modeling* 2015;55(12):2587-2595.
49. Carballo-Pacheco M, Strodel B. Comparison of force fields for Alzheimer's A: A case study for intrinsically disordered proteins. *Protein science* 2017;26(2):174-185.
50. Man VH, Nguyen PH, Derreumaux P. High-resolution structures of the amyloid- β 1–42 dimers from the comparison of four atomistic force fields. *The Journal of Physical Chemistry B* 2017;121(24):5977-5987.
51. Siwy CM, Lockhart C, Klimov DK. Is the Conformational Ensemble of Alzheimer's A β 10-40 Peptide Force Field Dependent? *PLoS computational biology* 2017;13(1):e1005314.
52. Man VH, He X, Derreumaux P, Ji B, Xie X-Q, Nguyen PH, Wang J. Effects of all-atom molecular mechanics force fields on amyloid peptide assembly: the case of a β 16–22 dimer. *Journal of chemical theory and computation* 2019;15(2):1440-1452.
53. Rahman MU, Rehman AU, Liu H, Chen H-F. Comparison and evaluation of force fields for intrinsically disordered proteins. *Journal of Chemical Information and Modeling* 2020;60(10):4912-4923.
54. Nasica-Labouze J, Nguyen PH, Sterpone F, Berthoumieu O, Buchete N-V, Coté S, De Simone A, Doig AJ, Faller P, Garcia A and others. Amyloid β Protein and Alzheimer's Disease: When Computer Simulations Complement Experimental Studies. *Chemical Reviews* 2015;115(9):3518-3563.
55. Ilie IM, Caflisch A. Simulation studies of amyloidogenic polypeptides and their aggregates. *Chemical reviews* 2019;119(12):6956-6993.

56. Strodel B. Amyloid aggregation simulations: challenges, advances and perspectives. *Current opinion in structural biology* 2021;67:145-152.
57. Nguyen PH, Ramamoorthy A, Sahoo BR, Zheng J, Faller P, Straub JE, Dominguez L, Shea J-E, Dokholyan NV, De Simone A. Amyloid Oligomers: A Joint Experimental/Computational Perspective on Alzheimer's Disease, Parkinson's Disease, Type II Diabetes, and Amyotrophic Lateral Sclerosis. *Chemical Reviews* 2021.
58. Wang W, Ye W, Jiang C, Luo R, Chen HF. New force field on modeling intrinsically disordered proteins. *Chemical Biology and Drug Design* 2014;84(3):253-269.
59. Song D, Wang W, Ye W, Ji D, Luo R, Chen HF. ff14IDPs force field improving the conformation sampling of intrinsically disordered proteins. *Chemical Biology and Drug Design* 2017;89(1):5-15.
60. Song D, Luo R, Chen HF. The IDP-Specific Force Field ff14IDPSFF Improves the Conformer Sampling of Intrinsically Disordered Proteins. *Journal of Chemical Information and Modeling* 2017;57(5):1166-1178.
61. Yang S, Liu H, Zhang Y, Lu H, Chen H. Residue-specific force field improving the sample of intrinsically disordered proteins and folded proteins. *Journal of chemical information and modeling* 2019;59(11):4793-4805.
62. Liu H, Song D, Lu H, Luo R, Chen HF. Intrinsically disordered protein-specific force field CHARMM36IDPSFF. *Chemical biology & drug design* 2018;92(4):1722-1735.
63. Yu L, Li D-W, Bruschweiler R. Balanced amino-acid specific molecular dynamics force field for the realistic simulation of both folded and disordered proteins. *Journal of Chemical Theory and Computation* 2019.
64. Duong VT, Chen Z, Thapa MT, Luo R. Computational Studies of Intrinsically Disordered Proteins. *Journal of Physical Chemistry B* 2018;122(46):10455-10469.
65. Dan A, Chen HF. Secondary structures transition of tau protein with intrinsically disordered proteins specific force field. *Chemical Biology and Drug Design* 2019;93(3):242-253.
66. Tomaselli S, Esposito V, Vangone P, Nuland NAJV, Bonvin AMJJ, Guerrini R, Tancredi T, Temussi PA, Picone D. The α -to- β conformational transition of Alzheimer's A β -(1-42) peptide in aqueous media is reversible: A step by step conformational analysis suggests the location of β c onformation seeding. *ChembioChem* 2006;7(2):257-267.

67. Wälti MA, Ravotti F, Arai H, Glabe CG, Wall JS, Böckmann A, Güntert P, Meier BH, Riek R. Atomic-resolution structure of a disease-relevant A β (1-42) amyloid fibril. *Proceedings of the National Academy of Sciences* 2016;113(34):E4976-84.
68. Case D, Ben-Shalom I, Brozell S, Cerutti D, Cheatham III T, Cruzeiro V, Darden T, Duke R, Ghoreishi D, Gilson M. AMBER 2018; 2018. University of California, San Francisco.
69. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* 1983;79(2):926-935.
70. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *The Journal of chemical physics* 1995;103(19):8577-8593.
71. Ryckaert J-P, Ciccotti G, Berendsen HJ. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics* 1977;23(3):327-341.
72. Doll J, Dion D. Generalized Langevin equation approach for atom/solid–surface scattering: Numerical techniques for Gaussian generalized Langevin dynamics. *The Journal of Chemical Physics* 1976;65(9):3762-3766.
73. Adelman S. Generalized Langevin theory for many-body problems in chemical dynamics: General formulation and the equivalent harmonic chain representation. *The Journal of Chemical Physics* 1979;71(11):4471-4486.
74. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *The Journal of chemical physics* 2004;120(24):11919-11929.
75. You W, Chang CEA. Role of Molecular Interactions and Protein Rearrangement in the Dissociation Kinetics of p38 α MAP Kinase Type-I/II/III Inhibitors. *Journal of Chemical Information and Modeling* 2018;58(5):968-981.
76. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *Journal of molecular graphics* 1996;14(1):33-38.
77. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z. BioMagResBank. *Nucleic acids research* 2007;36(suppl_1):D402-D408.
78. Han B, Liu Y, Ginzinger SW, Wishart DS. SHIFTX2: significantly improved protein chemical shift prediction. *Journal of biomolecular NMR* 2011;50(1):43.

79. Shao J, Tanner SW, Thompson N, Cheatham TE. Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *Journal of chemical theory and computation* 2007;3(6):2312-2334.
80. Kabsch W, Sander C. Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. 1983;22:2577-2637.
81. Zheng W, Tsai M-Y, Chen M, Wolynes PG. Exploring the aggregation free energy landscape of the amyloid- β protein (1–40). *Proceedings of the National Academy of Sciences* 2016;113(42):11835-11840.
82. Zheng W, Tsai M-Y, Wolynes PG. Comparing the aggregation free energy landscapes of amyloid beta (1–42) and amyloid beta (1–40). *Journal of the American Chemical Society* 2017;139(46):16666-16676.
83. Jia Z, Schmit JD, Chen J. Amyloid assembly is dominated by misregistered kinetic traps on an unbiased energy landscape. *Proceedings of the National Academy of Sciences* 2020;117(19):10322-10328.
84. Roche J, Shen Y, Lee JH, Ying J, Bax A. Monomeric A β 1–40 and A β 1–42 peptides in solution adopt very similar Ramachandran map distributions that closely resemble random coil. *Biochemistry* 2016;55(5):762-775.
85. Hou L, Shao H, Zhang Y, Li H, Menon NK, Neuhaus EB, Brewer JM, Byeon I-JL, Ray DG, Vitek MP. Solution NMR studies of the A β (1–40) and A β (1–42) peptides establish that the Met35 oxidation state affects the mechanism of amyloid formation. *Journal of the American Chemical Society* 2004;126(7):1992-2005.
86. Huy PDQ, Vuong QV, La Penna G, Faller P, Li MS. Impact of Cu (II) binding on structures and dynamics of A β 42 monomer and dimer: Molecular dynamics study. *ACS chemical neuroscience* 2016;7(10):1348-1363.
87. Kirkitadze MD, Condrón MM, Teplow DB. Identification and characterization of key kinetic intermediates in amyloid β -protein fibrillogenesis. *Journal of molecular biology* 2001;312(5):1103-1119.
88. Vivekanandan S, Brender JR, Lee SY, Ramamoorthy A. A partially folded structure of amyloid-beta (1–40) in an aqueous environment. *Biochemical and biophysical research communications* 2011;411(2):312-316.
89. Lincoff J, Sasmal S, Head-Gordon T. The combined force field-sampling problem in simulations of disordered amyloid- β peptides. *The Journal of chemical physics* 2019;150(10):104108.

90. Ono K, Condrón MM, Teplov DB. Structure–neurotoxicity relationships of amyloid β -protein oligomers. *Proceedings of the National Academy of Sciences* 2009;106(35):14745-14750.
91. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Structure, Function, and Bioinformatics* 2006;65(3):712-725.
92. Masuda Y, Uemura S, Nakanishi A, Ohashi R, Takegoshi K, Shimizu T, Shirasawa T, Irie K. Verification of the C-terminal intramolecular β -sheet in A β 42 aggregates using solid-state NMR: implications for potent neurotoxicity through the formation of radicals. *Bioorganic & medicinal chemistry letters* 2008;18(11):3206-3210.
93. Yan Y, Wang C. A β 42 is more rigid than A β 40 at the C terminus: implications for A β aggregation and toxicity. *Journal of molecular biology* 2006;364(5):853-862.
94. Wu C, Murray MM, Bernstein SL, Condrón MM, Bitan G, Shea J-E, Bowers MT. The structure of A β 42 C-terminal fragments probed by a combined experimental and theoretical study. *Journal of molecular biology* 2009;387(2):492-501.
95. Ball KA, Phillips AH, Wemmer DE, Head-Gordon T. Differences in β -strand populations of monomeric A β 40 and A β 42. *Biophysical journal* 2013;104(12):2714-2724.
96. Markwick PR, Bouvignies G, Blackledge M. Exploring multiple timescale motions in protein GB3 using accelerated molecular dynamics and NMR spectroscopy. *Journal of the American Chemical Society* 2007;129(15):4724-4730.
97. Doshi U, Hamelberg D. Towards fast, rigorous and efficient conformational sampling of biomolecules: Advances in accelerated molecular dynamics. *Biochimica et Biophysica Acta (BBA)-General Subjects* 2015;1850(5):878-888.
98. Rodríguez-Bussey IG, Doshi U, Hamelberg D. Enhanced molecular dynamics sampling of drug target conformations. *Biopolymers* 2016;105(1):35-42.
99. Duan S, Guan X, Lin R, Liu X, Yan Y, Lin R, Zhang T, Chen X, Huang J, Sun X. Silibinin inhibits acetylcholinesterase activity and amyloid β peptide aggregation: a dual-target drug for the treatment of Alzheimer's disease. *Neurobiology of aging* 2015;36(5):1792-1807.
100. Watts CR, Gregory AJ, Frisbie CP, Lovas S. Structural properties of amyloid β (1-40) dimer explored by replica exchange molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics* 2017;85(6):1024-1045.

101. Reddy G, Straub JE, Thirumalai D. Influence of preformed asp23– lys28 salt bridge on the conformational fluctuations of monomers and dimers of A β peptides with implications for rates of fibril formation. *The journal of physical chemistry B* 2009;113(4):1162-1172.
102. Sciarretta KL, Gordon DJ, Petkova AT, Tycko R, Meredith SC. A β 40-Lactam (D23/K28) models a conformation highly favorable for nucleation of amyloid. *Biochemistry* 2005;44(16):6003-6014.
103. Nguyen PH, Sterpone F, Derreumaux P. Aggregation of disease-related peptides. *Progress in molecular biology and translational science* 2020;170:435-460.
104. Weber OC, Uversky VN. How accurate are your simulations? Effects of confined aqueous volume and AMBER FF99SB and CHARMM22/CMAP force field parameters on structural ensembles of intrinsically disordered proteins: Amyloid- β 42 in water. *Intrinsically disordered proteins* 2017;5(1):e1377813.
105. Bhattacharya S, Xu L, Thompson D. Molecular simulations reveal terminal group mediated stabilization of helical conformers in both amyloid- β 42 and α -synuclein. *ACS chemical neuroscience* 2019;10(6):2830-2842.
106. Nguyen PH, Campanera JM, Ngo ST, Loquet A, Derreumaux P. Tetrameric A β 40 and A β 42 β -barrel structures by extensive atomistic simulations. i. in a bilayer mimicking a neuronal membrane. *The Journal of Physical Chemistry B* 2019;123(17):3643-3648.

2.7 Acknowledgements

We thank Dr. Hai-Feng Chen for providing the scripts for *ffl4IDPSFF* and Dr. Ray Luo for providing the force fields and helpful discussion. This study is supported by the US National Institutes of Health (GM-109045)

Chapter 3 Functional Dynamics in PROTACs-Induced Protein Degradation Machinery

3.1 Abstract

PROteolysis TArgeting Chimeras (PROTACs) have emerged as an innovative class of drug to induce disease-related targets for degradation. Many determinants governing PROTAC's functionality, activity, and most importantly, the dynamics nature of the ubiquitin-mediated degradation complex remains to be addressed. In this work, we model both the CRBN-dBET-BRD4^{BD1} ternary complex and the DDB1/CUL4A/NEDD8/Rbx1/E2/Ub/CRBN-dBET-BRD4^{BD1} degradation machinery complex and study the degradation machinery complex's functional dynamics using molecular dynamics (MD) simulations. We reveal a critical catalytic chemistry step for the target protein in the context of the full in the CRL4A E3 ligase. Our multi-level modeling approaches including protein protein docking, structures alignment and superposition, and MD simulations captures the dynamic nature of the degradation machinery complex and provides explanation of differences in degradability among different BRD4^{BD1} PROTACs engaging different proximity of lysine residues on BRD4^{BD1}. Our work provides insights into the protein functional dynamics of the overall PROTACs-induced degradation machinery and potentially serves as a conceptual guidance in "rational PROTAC designs".

3.2 Introduction

With the rapid progressive efforts from modern drug discovery and development, many new potential paradigms for disease treatment are emerging from the wealth of medicinal

chemistry and biology generated. Detailed understanding of the molecular basis for the function of the targeted biological system expedites drug development processes. PROteolysis TArgeting Chimera (PROTAC)^{1,2} is such a new paradigm of drug that promotes selective degradation for a variety of prominent disease targets.³⁻⁹ The ternary complex formed by the PROTAC binding to both proteins induces ubiquitination of the target protein by the E3-associated protein complex, which marks it for proteasomal degradation and removal.¹⁰ The fundamental mechanism of PROTACs involves engaging the cell's native protein degradation machinery, such as proteosomes and lysosomes.^{11,12} However, due to its complex degradation mechanism, PROTACs have numerous unknowns left at their many stages of molecular interactions. Several recent experimental studies have indicated that protein-protein interactions play an equal, if not greater, role in ternary complex stability than the individual affinities of the binders towards their targets, and provided important biophysical and structural information.¹³⁻¹⁹ Furthermore, Cell-based experimental studies showed that changing the linker may also alter the effectiveness of the final degradation.^{8,13,20-22}

In addition, It is well-established that the final target protein degradation involved in a three-step (activation-E1, conjugation-E2, and ligation-E3) ubiquitination is a consequence of binding ternary complex brought by a PROTAC, yet its binding affinity and complex formation are not correlated with the target protein degradation, especially given the functional dynamics of the degradation complex, as shown in some recent studies.²³⁻²⁶ For example, a class of bromodomain (BRD)-containing proteins (BRD2, BRD3, BRD4 and BRDT) recognizing the acetyl-lysine in histone is considered an important drug target for

a variety of cancers.^{27,28} A study of several PROTACs and derivative degraders have been designed for the degradation of BRD4^{BD1} based on BRD inhibitors and CRBN-targeting pomalidomide (or thalidomide).^{15,21,29} Although the ternary complex crystal structures are found identical with the similar protein protein contacts and reported binding affinity,^{15,21,30} the degradation capabilities ($DC_{50/5h}$) are significantly different, with dBET70 (~5nM) and dBET6 (~10nM) being the best, followed by dBET23 (~50nM). The poorest degradation capacities are seen in dBET1 (~500nM) and dBET57 (~500nM). (Figure 3.1) Binding affinity and degradation discrepancy has also been reported for von Hippel Lindau (VHL)-based PROTACs.^{23,31} Questions were raised to whether static structural of the ternary complex is sufficient to explain degradation capability. One possible explanation of this discrepancy can be attributed to the proximity and accessibility of target protein's lysine(s) to the E2/Ub cascade in the ubiquitination site.³¹ To fill the knowledge gap between the ternary complex formation and ubiquitin degradation machinery complex, some studies applied docking and structure-based modeling approaches to investigate and predict CDKs-PROTACs and BCL-xl-PROTACs inducing ternary complex formation and target protein degradation in CRBN-based Culling-Ring Ligase 4A (CRL4A) complex and in VHL-based complex, respectively.³²⁻³⁴ Their work revealed the structural relationship that governs the PROTAC-mediated ternary complex formation and ubiquitin degradation machinery using static structures. However, proteins are not rigid, lack of considering proteins' dynamic nature and environment can fail to identify conformations of protein complexes important to spatially organize the target protein's lysine(s) and the catalytic site on E2 for ubiquitination, which is critical in degrader design.

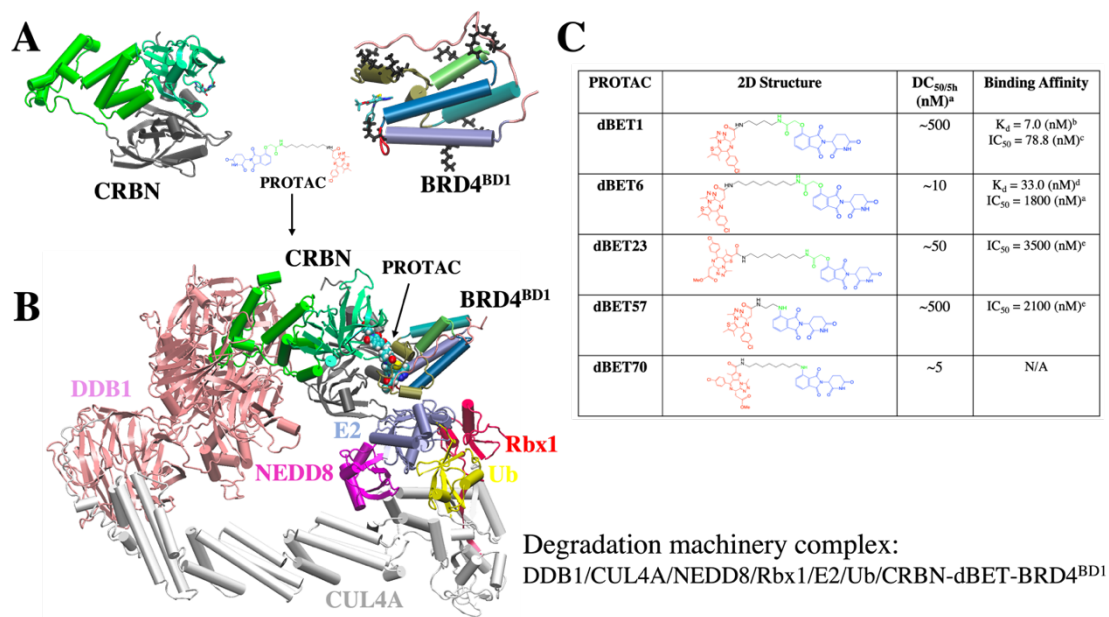


Figure 3.1: PROTAC-mediated degradation machinery complex. A, protein of interest (POI) BRD4^{BD1} (cyan: α Z, purple: α C, blue: α B, light green: α A, tan: ZA-loop, red: BC-loop, black licorice: lysine residues) and CRBN E3 ligase (gray: NTD, green: HBD, mint: CTD). B, degradation machinery complex: DDB1/CUL4A/NEDD8/Rbx1/E2/Ub/CRBN-dBET-BRD4^{BD1}. C, 2D molecular structure (red: JQ1, BRD4^{BD1} ligand, blue: pomalidomide, CRBN ligand, black: linker), degradation profile DC_{50/5h}, and binding affinity; a, data were obtained from ref #¹⁵, IC₅₀: Inhibition of CRBN by competition binding assay. b, data were obtained from ref #³⁰. c, data were obtained ref #²¹, inhibit growth of RS4;11 cells. d, data were obtained from ref #³⁵, K_d was tested by AlphaLISA assay. e, IC₅₀ was calculated in the presence of 40nM CRBN and 1uM BRD4^{BD1} competition binding assay.

Thus, we aim to understand the structural and protein functional dynamics of the PROTAC- induced ubiquitin degradation complex, which ultimately guide “rational PROTAC designs”. In this work, we model and identify structural features of PROTACs that lead to stable ternary complex formation and provide explanation of the binding affinity and degradability reported for the BRD4-based PROTACs. Our modeling approaches extend to simulate the CRBN-dBET-BRD4^{BD1} ternary complex and its dynamics and model the CUL4A E3 ligase complex (DDB1/CUL4A/NEDD8/Rbx1/E2/Ub). Then, we further model the full CUL4A E3 ligase-

based ubiquitin degradation machinery complex and carry out MD simulations to investigate the degradation machinery complex's functional dynamics. Our work demonstrates mechanistic knowledge of the BRD4^{BD1} PROTAC-mediated ternary complex formation and addresses structural functional relationship in the ubiquitin degradation machinery complex engaged by the ternary complex. Our work shows importance of the protein dynamic nature in the PROTAC-mediated BRD4^{BD1} degradation, which is beyond what is provided by static crystal structures, hence allowing efficient PROTAC design.

3.3 Methods and Material

Model CRBN-dBETx-BRD4^{BD1} Ternary Complex

Inhibitor (JQ1)-bound BRD4^{BD1} crystal structure (PDB ID: 3MXF)³⁶ and pomalidomide-bound CRBN crystal structure (PDB ID: 4CI3)³⁷ were downloaded from the PDB website (Figure 3.2). Five BRD4^{BD1} PROTACs (dBET1, dBET6, dBET23, dBET57 and dBET70) were prepared using builder program in MOE. Ternary complex ensembles were generated using method 4B protocol implemented in MOE program (Chemical Computing Group).³⁸ CRBN-dBETx-BRD4^{BD1} was modelled in four steps: 1, CRBN-BRD4^{BD1} docking; 2, dBET conformational sampling; 3, docked CRBN-BRD4^{BD1} poses and dBETs conformations alignment; 4, scoring and clustering of modeled CRBN-dBETx-BRD4^{BD1}. Before performing protein-protein docking, two crystal structures were prepared by adding missing hydrogens and assigning appropriate protonation state for each protein. In the protein-protein docking step, two binary complexes interacted without linker connecting two warheads. The protein-protein poses were generated, and conformations were stored for the next step. In the PROTAC conformational sampling step, five dBETs were provided

for robust conformational lowModeMD searching. Each iteration was set to 10000. Ligand core RMSD was set as default. In the dBETx conformational searching process, each binding warhead in the dBETs was held rigid to retain its bound conformation, while the linker region was simulated on-the-fly. Maximum common substructure approach was applied in the program to maximize dBET warheads alignment with the binding warheads. The simulated dBET conformations were then stored for the next step. In the protein-protein pose and dBETs conformation alignment step, protein-protein poses having two warheads in each binding pocket were aligned with the dBETs' two binding moieties. The generated ternary complex CRBN-dBETx-BRD4^{BD1} ensembles were minimized, scored, clustered. Selected represented conformations with docking scores were stored as pdb format for further visualization analysis.

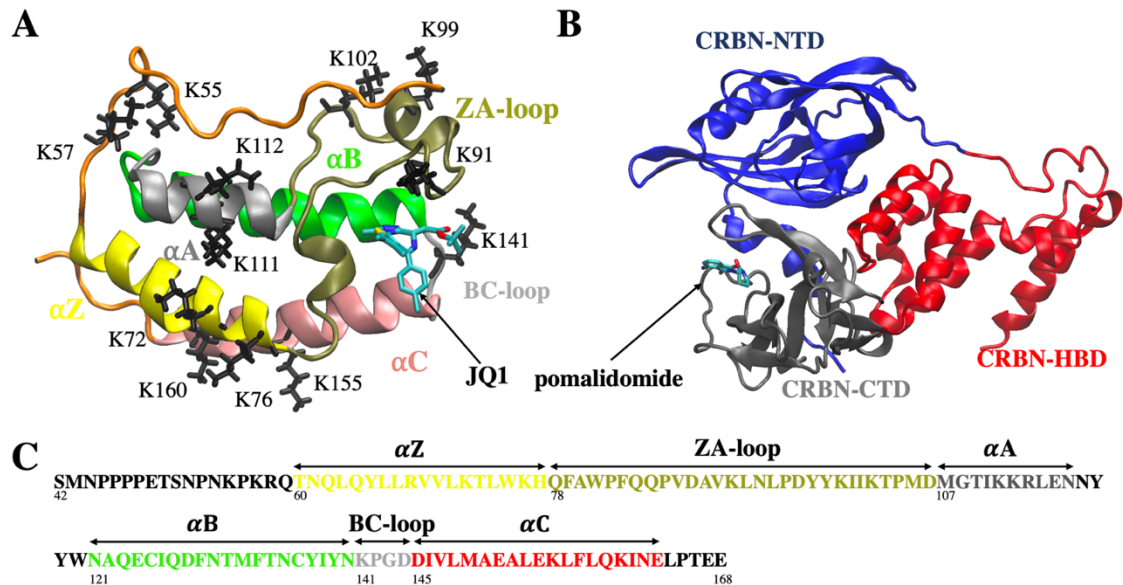


Figure 3.2: Cartoon representations of A. BRD4^{BD1} (PDB ID 3MXF), lysine residues are shown in black licorice; B. CRBN E3 ligase (PDB ID 4CI3); C. BRD4^{BD1} sequence representation.

Model CRL4A E3 Ligase Complex: DDB1/CUL4A/NEDD8/Rbx1/E2/Ub/CRBN

We obtained 12 DDB1 PDB structures from PDB website and named them with index ID: A1_DDB1 (PDB ID: 4E54), A2_DDB1 (PDB ID: 3E0C), A3_DDB1 (PDB ID: 3I8E), A4_DDB1 (PDB ID: 4TZ4), A5_DDB1 (PDB ID: 4A0L), B1_DDB1 (PDB ID: 2B5L), B2_DDB1 (PDB ID: 3EI4), B3_DDB1 (PDB ID: 3EI3), B4_DDB1 (PDB ID: 6PAI), C1_DDB1 (PDB ID: 6FCV), C2_DDB1 (PDB ID: 4A08), C3_DDB1 (PDB ID: 4A0B). CUL4A with DDB1 crystal structure was downloaded (PDB ID: 2HYE) and used as a template. 12 DDB1 crystal structures were superimposed to the CUL4A/DDB1 template via BPB domain to make 12 DDB1/CUL4A structures. CRBN/DDB1 from A4_DDB1 (PDB ID: 4TZ4) was used as a template, and the rest of the DDB1/CUL4A structures were superimposed via BPA/BPC domain to make 12 DDB1/CUL4A/CRBN complexes. Next, E2D1 (PDB ID 5FER chain B) was aligned to (PDB ID 6TTU) to replace E2D2. We then aligned the 12 CRBN/DDB1/CUL4A complexes via residues in the C-terminal domain (residues from 416 to 672) in the CUL4A. Components having NEDD8, Rbx1, E2 and Ub with CUL1 (PDB ID 6TTU) were added to the above 12 CRBN/DDB1/CUL4A structures. The final 12 DDB1/CUL4A/Rbx1/NEDD8/E2/Ub/CRBN complexes were ready for the next step (Figure 3.3 for flowchart). Next, we analyzed and categorized the 12 CRL4A E3 ligase complexes DDB1/CUL4A/NEDD8/Rbx1/E2/Ub/CRBN, based on the interface distance between the CRBN E3 ligase and the E2. Specifically, we measured residue distances from CRBN E3 to the E2 interface plane among these 12 CRL4A E3 ligase complexes. If the CRBN E3 ligase residue distances to E2 plane is less than 10Å and no overlapping or clashing, we grouped those CRL4A E3 ligase complexes into category A;

If the CRBN E3 ligase residue distances to E2 plane is greater than 10Å (the exact gap distance ranges from ~14Å to ~34Å), we grouped those complexes into category B. If CRBN E3 and E2 overlaps, we grouped those complexes in category C, considering them clashing and abandoned (Figure 3.4)

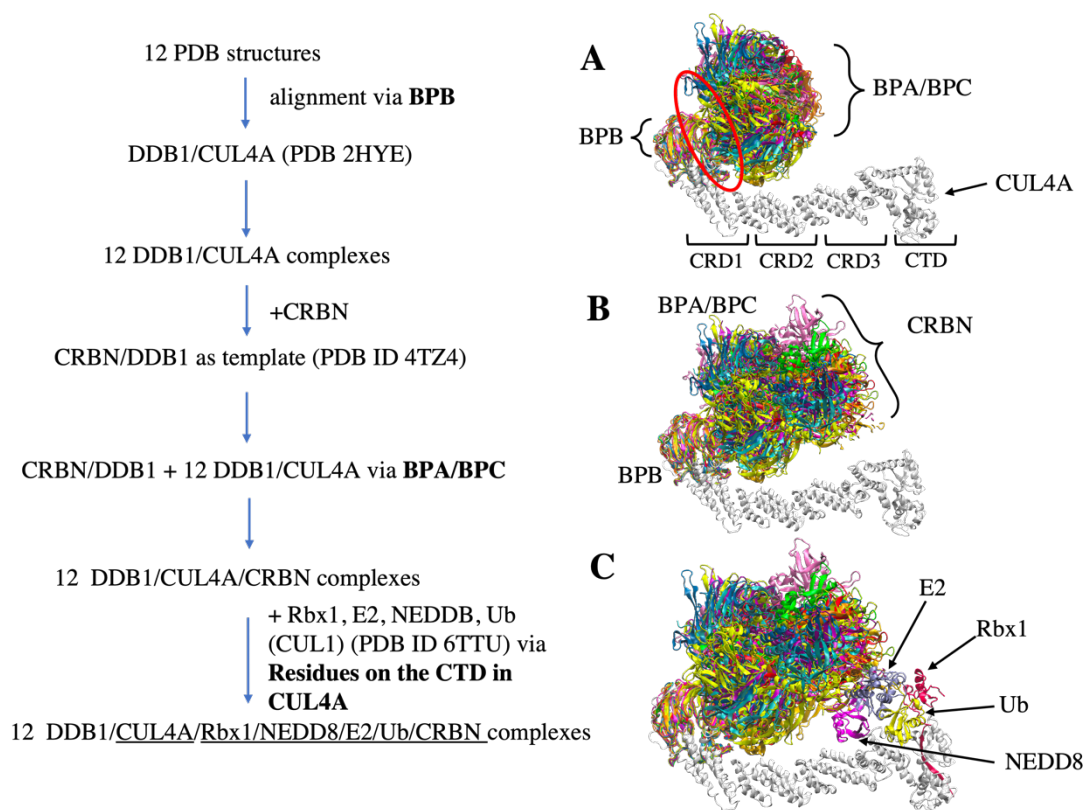


Figure 3.3: Flowchart of CUL4A E3 ligase complexes construction. A. structure alignments of 12 DDB1 PDB structures to the CUL4A. B. structure alignment of CRBN E3 ligase to the DDB1/CUL4A. C. E2/Rbx1/Ubiquitin/NEDD8 components are added to the CRBN/DDB1/CUL4A complex.

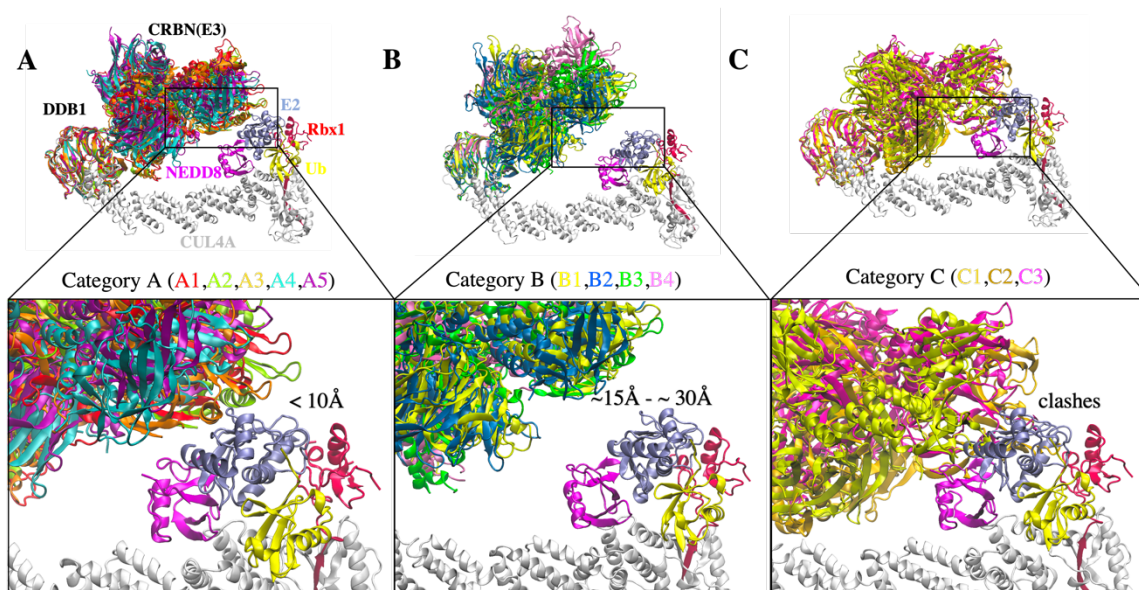


Figure 3.4: CRL4A E3 ligase complexes with 12 DDB1 conformations. A. category A with the surface distance between the E2 and E3 is less than 10 Å. Complexes A1, A2, A3, A4, and A5 are in category A. B. category B with the surface distance between the E2 and E3 is between ~15 Å to ~30 Å. Complexes B1, B2, B3, and B4 are in category B. category C with the E2 and E3 clashes. Complexes C1, C2, and C3 are in category C.

Model Degradation Machinery Complex: DDB1/CUL4A/NEDD8/Rbx1/E2/Ub/CRBN-dBETx-BRD4^{BD1}

Next, CRBN-dBETx-BRD4^{BD1} conformations generated from protein-protein docking modeling step were aligned with the CRL4A E3 ligase complexes taking both category A and B (total of nine CRL4A E3 ligase complexes) via common CRBN structure. Because there are 12 lysine residues on the surface of BRD4^{BD1} (Figure 3.2 A), we considered all the lysine residues that potentially lead to ubiquitination by measuring the distance between the oxygen from C-terminal glycine of the ubiquitin and the nitrogen atom of surface exposed lysine residue(s) of the BRD4^{BD1} using VMD³⁹ (Table 3.1). Among the modelled CRBN-dBETx-BRD4^{BD1} conformations with each CRL4A E3 ligase complexes (A1, A2, A3, A4, A5, B1, B2, B3, B4), only the measured distance is less than or close to 16 Å without clashing were kept for further study, as suggested by the previous study.³³

Table 3.1: Accessible lysine residues identified in the CRL4A E3 ligase complexes for each PROTAC.

PROTAC	A1	A2	A3	A4	A5
dBET1	55,57,102,112	55,57,112	112,155	112	55,57,72,91,111,112,155,160
dBET6	55,102,112,160	55,57,72,112,160	111,112,155	112,111	72,91,111,112,155,160
dBET23	55,102,141,155,160	55,99,102,112,141,155,160	72,76,111,155	N/A	155,160
dBET57	55,112	55,112	N/A	111,112	111,112
dBET70	55,141,160	55,99,102,112,141,155,160	76,155,160	N/A	155,160

PROTAC	B1	B2	B3	B4
dBET1	57,55,72,102,111,112	55,57,72,102,111,112,160	N/A	N/A
dBET6	55,57,72,99,102,111,112,160	55,57,72,102,111,112,160	N/A	N/A
dBET23	55,57,99,102,112	55,57,99,102,112	N/A	N/A
dBET57	57,55,111,112	55,57,72,111,112	N/A	N/A
dBET70	55,57,99,102,112	55,57,99,102	N/A	N/A

Set Up Molecular Dynamic Simulations of Modelled CRBN-dBETx-BRD4^{BD1} and Degradation Machine Complexes: DDB1/CUL4A/NEDD8/Rbx1/E2/Ub/CRBN-dBETx-BRD4^{BD1}

The top selected CRBN-dBETx-BRD4^{BD1} conformations from combining protein-protein docking and structural based steps were used for molecular dynamics (MD) simulations to capture the dynamics of CRBN-dBETx-BRD4^{BD1} and degradation machinery complexes (Table 3.2). For each dBET, we selected one CRBN-dBETx-BRD4^{BD1} conformation, but selected two CRBN-dBET6-BRD4^{BD1} conformations for comparison. For reference, we also investigated and performed MD simulations for resolved crystal structures. Among five studied dBETs in this work, only CRBN-dBET6-BRD4^{BD1} (PDB ID 6BOY) and CRBN-dBET23-BRD4^{BD1} (PDB ID 6BN7), have pdb structures for starting the MD simulations. Other ternary complexes either miss CRBN E3 component (PDB ID 4ZC9, no dBET1³⁰) or miss dBET ligands (PDB ID 6BNB, no dBET57, PDB ID 6BN9 no

dBET70) (Figure 3.5) were not considered. Before proceeding to MD simulations, missing part in each protein component of each CRL4A E3 ligase complex was built either by using structural alignment or homology modeling method.⁴⁰ In brief, we used (PDB ID 4A0L) as template for A1_DDB1, (PDB ID 3L7K) as template for A3_DDB1, and (PDB ID 5HTB) as template for A5_DDB1 in homology modeling, while the rest CRL4A E3 ligase complexes (A2, A4, B1, B2, B3, B4) built the missing part directly using homology modeling method. We picked one CRL4A E3 ligase complex from each category: A1 in category A and B1 in category B based on the high populations in that complex. Then two criteria were considered to select the top CRBN-dBETx-BRD4^{BD1} for the ternary complex MD simulations and for degradation machinery complex MD simulations: 1, the common CRBN-dBETx-BRD4^{BD1} shared in the category (Figure 3.6); 2, the best protein-protein docking score (energy based, Table 3.3).

Table 3.2: MD simulations setup for the CRBN-dBETx-BRD4^{BD1} complexes, degradation machinery complex C4 from category A, degradation machinery complex C12 and from category B

MD index	Run	MD index	Run	MD index	Run
CRBN-dBET1_35-BRD4 ^{BD1}	3	A_md_c4_dBET1_35	1	B_md_c12_dBET1_md3*	1
CRBN-dBET6_77-BRD4 ^{BD1}	3	A_md_c4_dBET6_77	1	B_md_c12_dBET6_116	1
CRBN-dBET6_116-BRD4 ^{BD1}	3	A_md_c4_dBET6_116	1	B_md_c12_dBET6_77	1
CRBN-dBET23_14-BRD4 ^{BD1}	3	A_md_c4_dBET23_14	1	B_md_c12_dBET23_14	1
CRBN-dBET57_9-BRD4 ^{BD1}	3	A_md_c4_dBET57_9	1	B_md_c12_dBET57_9	1
CRBN-dBET70_91-BRD4 ^{BD1}	3	A_md_c4_dBET70_91	1	B_md_c12_dBET70_91	1
CRBN-xray_dBET6-BRD4 ^{BD1}	3	A_xray_c4_dBET6	1	B_xray_c12_dBET6	1
CRBN-xray_dBET23-BRD4 ^{BD1}	3	A_xray_c4_dBET23	1	B_xray_c12_dBET23	1

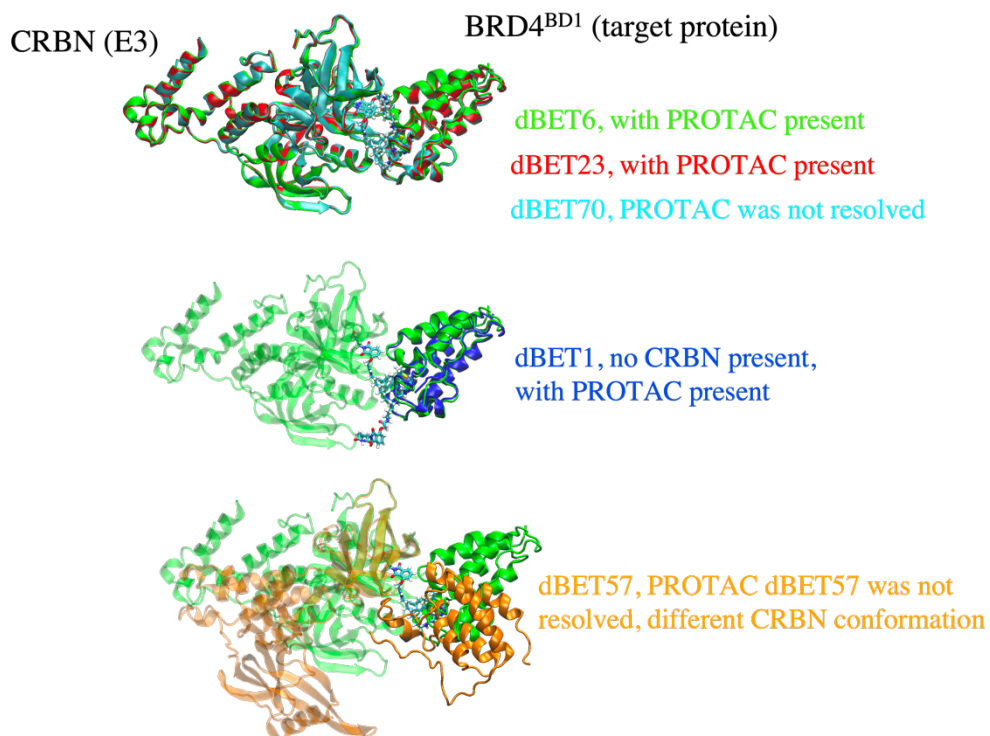


Figure 3.5: Crystal structure alignments of reported PROTACs (dBET1: PDB ID 4ZC9, dBET6: PDB ID 6BOY, dBET23: PDB ID 6BN7, dBET57: PDB ID 6BNB, and dBET70: PDB ID 6BN9). Top, superimpose of dBET6, dBET23, and dBET70 crystal structures. Middle, superimpose of BRD4^{BD1} from dBET1 with the BRD4^{BD1} from dBET6. Bottom, superimpose of CTD of CRBN from dBET57 with CTD of CRBN from dBET6.

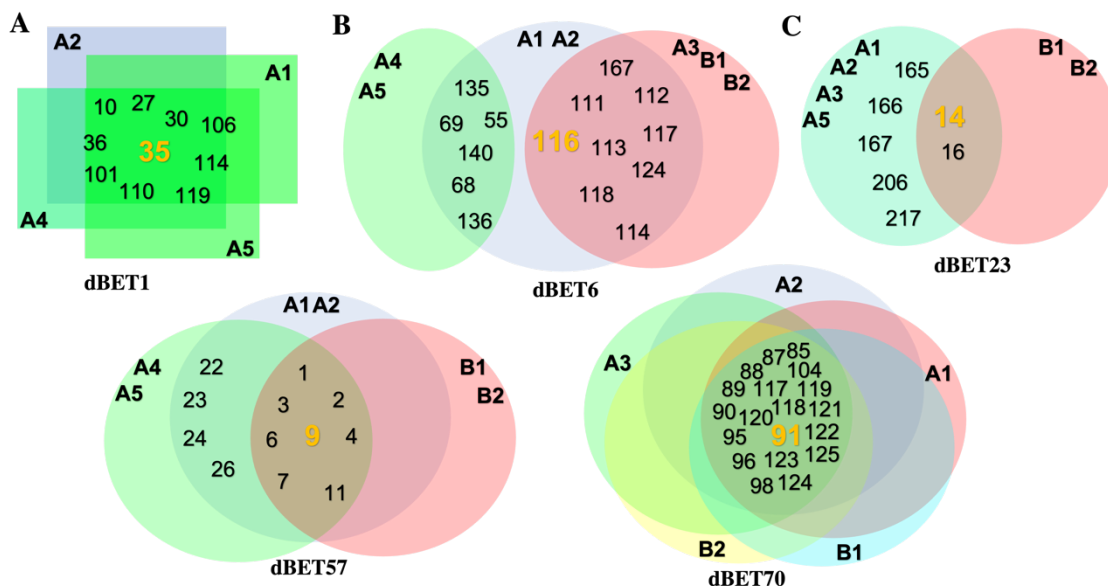


Figure 3.6: Overlapping diagram representation of common CRBN-dBETx-BRD4^{BD1} ensembles found in each degradation machinery complex. A, dBET1. B, dBET6. C, dBET23. D, dBET57. E, dBET70.

Table 3.3: Binding scores predicted by MOE protein protein docking for the CRBN-dBETx-BRD4^{BD1} complexes (*calculated in kcal/mol)

dBET1 Conf #	*E _{total}	dBET6 Conf #	*E _{total}	dBET23 Conf #	*E _{total}	dBET57 Conf #	*E _{total}	dBET70 Conf #	*E _{total}
35	-7353.1	68	-7379.1	206	-7272.9	22	-7265.3	91	-7330.9
10	-7352.0	116	-7334.2	14	-7221.1	24	-7265.2	85	-7326.2
27	-7349.2	167	-7320.6	165	-7193.7	26	-7264.9	90	-7323.8
30	-7345.6	111	-7317.8	168	-7183.9	9	-7237.8	117	-7319.3
101	-7304.9	135	-7315.2	166	-7177.8	23	-7225.3	88	-7314.0
110	-7279.3	114	-7309.9	217	-7141.8	1	-7216.9	118	-7308.2
36	-7255.5	113	-7305.5	167	-7141.4	7	-7201.4	120	-7303.2
106	-7236.1	69	-7304.7	16	-7045.1	4	-7153.0	89	-7297.7
114	-7260.9	112	-7289.4			2	-7145.9	119	-7295.4
119	-7142.3	140	-7278.5			3	-7130.2	122	-7288.8
		136	-7271.3			6	-7125.6	125	-7256.5
		124	-7256.3			11	-7065.6	123	-7254.4
		117	-7236.2					98	-7244.7
		77	-7200.3					121	-7237.9
		118	-7177.6					87	-7219.4
								95	-7214.7
								96	-7195.2
								124	-7104.8
								126	-7040.4
								104	-5301.3

MD simulations were performed using AMBER20 software package⁴¹ with AMBER ff14SB force field⁴² for proteins and GAFF2⁴³ for ligand parameterization, respectively. A

standard AMBER simulation protocol was used. The sequential minimization steps were performed for the hydrogen atoms, side chains and the entire protein complex to clean any clashes. The system was solvated in a TIP3P water box⁴⁴ with a distance 12 Å from the edge of the box to the whole system. To neutralize the CRL4A E3 ligase complex system, 40 and 46 Na⁺ ions were added to the water box for C4 cluster and for C12, respectively, while no counter ion was added to the ternary complex. Subsequently, the whole complex system with water was minimized by only focusing on water molecules, and then followed by the entire complex system. Heating of the complex system from 50 K to 300 K with an increment of 25 K was achieved in 200 ps NVT simulation (time step of 2 fs) for each temperature. The system was subsequently equilibrated at 300 K for 500 ps NPT simulation with all atoms relaxed. Long-range electrostatic interactions were computed by the particle mesh Ewald method.⁴⁵ The SHAKE constraint was applied to all atoms including hydrogens.⁴⁶ The Langevin thermostat with a damping constant of 2 ps⁻¹ was applied to maintain a temperature of 300 K. The NPT ensemble at 300 K was used for MD production run for 400 ns, and each frame was collected at 1 ps time interval. We generated eight CRBN-dBETx-BRD4^{BD1} complex structures and performed 3 MD runs with different initial velocities for each structure. Sixteen degradation machinery complex structures were constructed, and we ran one MD for each complex (Table 3.2).

3.4 Results and Discussion

To understand why PROTACs with similar binding ability to form a stable E3-ligand-POI ternary complex may have different induced degradation of a target protein, we need to understand how to achieve effective degradation of POI by the degradation machinery

complex. We therefore combined molecular docking, structural alignment and atomistic MD simulations in explicit solvent to construct the degradation machinery complexes, DDB1/CUL4A/NEDD8/Rbx/E2/Ub/CRBN-dBETx-BRD4^{BD1}, and examine the structure-dynamics-function relationship to reveal how the dBETx induced multi-protein complex may utilize a lysine on BRD4^{BD1} to accomplish BRD4^{BD1} degradation. Because not all CRBN-dBETx-BRD4^{BD1} have crystal structures and a handful conformations are not sufficient to understand protein function here, we performed protein-protein docking to obtain many ternary complexes with five dBETx and carried out MD runs to sample conformational arrangements. Selected conformations of CRBN-dBETx-BRD4^{BD1} were incorporated into the CRL4A E3 ligase complex to construct degradation machinery complex for investigating functional dynamics using MD simulations.

Ternary Complex Ensemble CRBN-dBETx-BRD4^{BD1} Prediction and Analysis

For the sake of modeling feasibility, we selected MOE method 4B protocol to model five BRD4^{BD1} dBETx series PROTACs. Of note, this ternary complex modeling approach does not take into accounts other properties, such as binding kinetics, PROTAC permeability, cell activity, ubiquitin degradability, oral bioavailability, pharmacokinetic/pharmacodynamic (PK/PD), but allows us to sample as many CRBN-dBETx_BRD4^{BD1} ensembles as possible *a priori* with thoroughly searching algorithms. It is conceivable in theory that, longer and more flexible PROTAC linkers create more degrees of freedom, and hence generate more ternary complexes. In the predicted CRBN-dBET57_BRD4^{BD1} conformations, dBET57 with the shortest linker has the least number (40) of ternary complexes, while CRBN-dBET1_BRD4^{BD1}, CRBN-dBET6_BRD4^{BD1}, and

CRBN-dBET23_BRD4^{BD1} have the most ternary complex formations (297, 245 and 250, respectively). (Table 3.4) After carefully checking each of the generated CRBN-dBETx_BRD4^{BD1} for all five PROTACs as suggested, we examined and culled some CRBN-dBETx_BRD4^{BD1} conformations that are not considered reasonable. Figure 3.7 shows the ensembles of each CRBN-dBETx_BRD4^{BD1} complex with percentages after superposition and culling. Although the predicted ternary complex percentages seem to agree with the degradation profile in our findings, it is not conclusive, as other study already showed that the prediction score does not always agree with the degradation.⁴⁷ However, our findings indicated that culling “unreasonable” conformations carried out by modeling work is the first step in PROTAC ternary complex prediction. As we mentioned above, ternary complex prediction does not include other factors, such as ubiquitin degradation and binding kinetics, *etc.*, we advanced in the workflow by considering the ubiquitin degradation machinery complex.

Table 3.4: Summary of the predicted CRBN-dBETx-BRD4^{BD1} ensembles.

PROTAC	DC_{50/5h} (nM)^a	PROTAC Conf.	Total Ternary Conf. Predicted	Selected Reasonable Ternary Conf.	Reasonable Ternary Conf. %
dBET1	~500	125	297	186	62.3
dBET6	~10	152	245	179	73.1
dBET23	~50	212	250	168	67.2
dBET57	~500	25	40	24	60.0
dBET70	~5	143	217	183	84.4

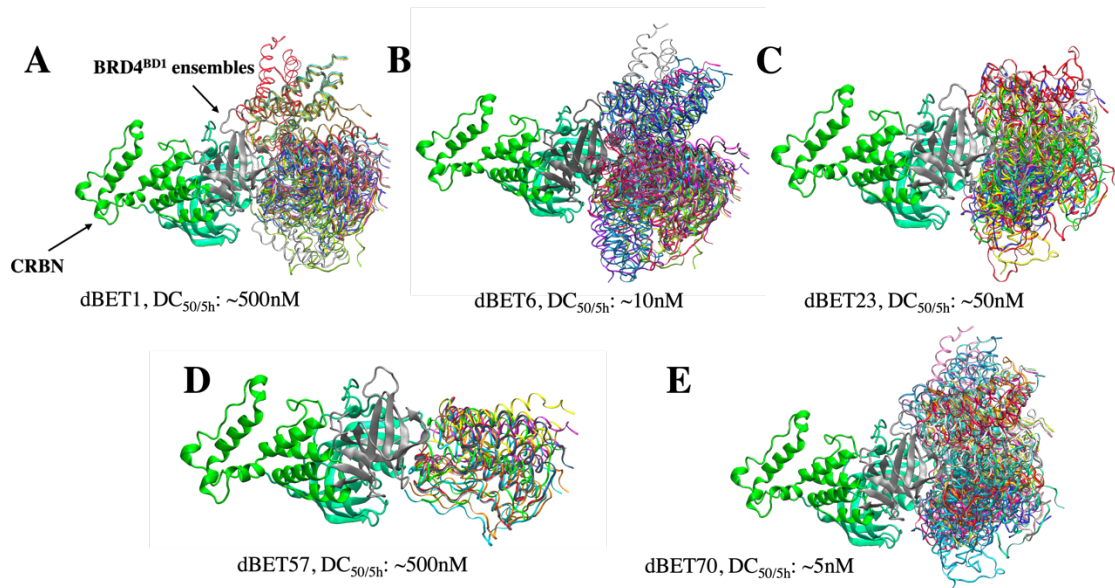


Figure 3.7: CRBN-dBET_x-BRD4^{BD1} ensembles generated from protein-protein docking . All ensembles were superimposed based on CRBN (cyan, thick, ribbon), BRD4^{BD1} poses were shown as thin, ribbon, rainbow color). A. CRBN-dBET1-BRD4^{BD1} ensembles; B. CRBN-dBET6-BRD4^{BD1} ensembles; C. CRBN-dBET23-BRD4^{BD1} ensembles; D. CRBN-dBET57-BRD4^{BD1} ensemble; E. CRBN-dBET70-BRD4^{BD1} ensemble.

Creation of CRL4A E3 Ligase Complex and Structural Alignment of Generated CRBN-dBET_x_BRD4^{BD1} Ensembles to Identify Potential Ubiquitinated Lysine Residue(s)

One of the three components CRBN E3 ligase in a CRBN-dBET-BRD4^{BD1} is also a substrate receptor of the adaptor protein DDB1. There are three WD40 domains (BPA, BPB, and BPC) in a DDB1, and CRBN E3 ligase is sandwiched between the BPA and BPC domains. Being part of the CRL4A E3 ligase complex, the BPB domain in DDB1 is attached non-covalently to the CRD1 of CUL4A.^{48,49} (Figure 3.3) The flexible hinge region connecting the BPB and the interface of BPA and BPC allows a diverse CRL4A E3 ligase complex conformational change. To include our predicted CRBN-dBET_x_BRD4^{BD1} ensembles into the ubiquitin degradation machine, we constructed 12 CRL4A E3 ligase complexes, and each predicted CRBN-dBET_x_BRD4^{BD1} complex was aligned and superimposed via the common structure unit CRBN E3. (See method for details) The

resulting degradation machinery complex includes DDB1 (BPA, BPB and BPC), CUL4A (CRD1, CRD2, CRD3, and CTD), Rbx1, NEDD8, E2, Ubiquitin, CRBN E3, PROTAC and BRD4^{BD1} components. As suggested, the distance between the exposed lysine residues in the BRD4^{BD1} and the glycine residue from the Ubiquitin can be used as a criterion (< 16 Å) to differentiate “degradable” and “non-degradable” ternary complex conformations in this structural-based modeling step.³³ For example, in category A, at least one lysine residue (K112) was found within the designed measure criterion in the predicted CRBN-dBET1_BRD4^{BD1}. (Figure 3.8 and Table 3.1 category A) In addition, we hypothesize that, though the surface distance gap between CRBN E3 and E2 is relatively far (~15 Å to ~30 Å), category B still can be used to differentiate “degradable” and “non-degradable” CRBN-dBETx_BRD4^{BD1} conformations in this structural-based modeling step because proteins constantly move. With this in mind, we observed that at least one lysine residue (K99 and K102) was found within the designed measure criterion for different PROTACs in different degradation machinery complexes, as shown in Figure 3.8 B and Table 3.1 category B. It is worth to point out we did not observe K141 in our CRBN-dBET6_BRD4^{BD1} prediction in both the categories, as reported in the previous work,³³ but K141 was predicted in both degradation machinery complexes A1 and A2 in both CRBN-dBET23_BRD4^{BD1} and CRBN-dBET70_BRD4^{BD1} in our study. (Highlighted in red in Table 3.1). Since three PROTAC crystal structures (PDB ID 6BOY, 6BN7, and 6BN9) have the same BRD4^{BD1} and CRBN E3 protein contacts and orientations, except for PROTAC positions. (Figure 3.5) It is suggestive that our structural-based modeling work is valid for degradation machinery complex construction to identify lysine residues that are potentially

ubiquitinated. After structural-based modeling filtration, the number of reasonable conformations were further lowered.

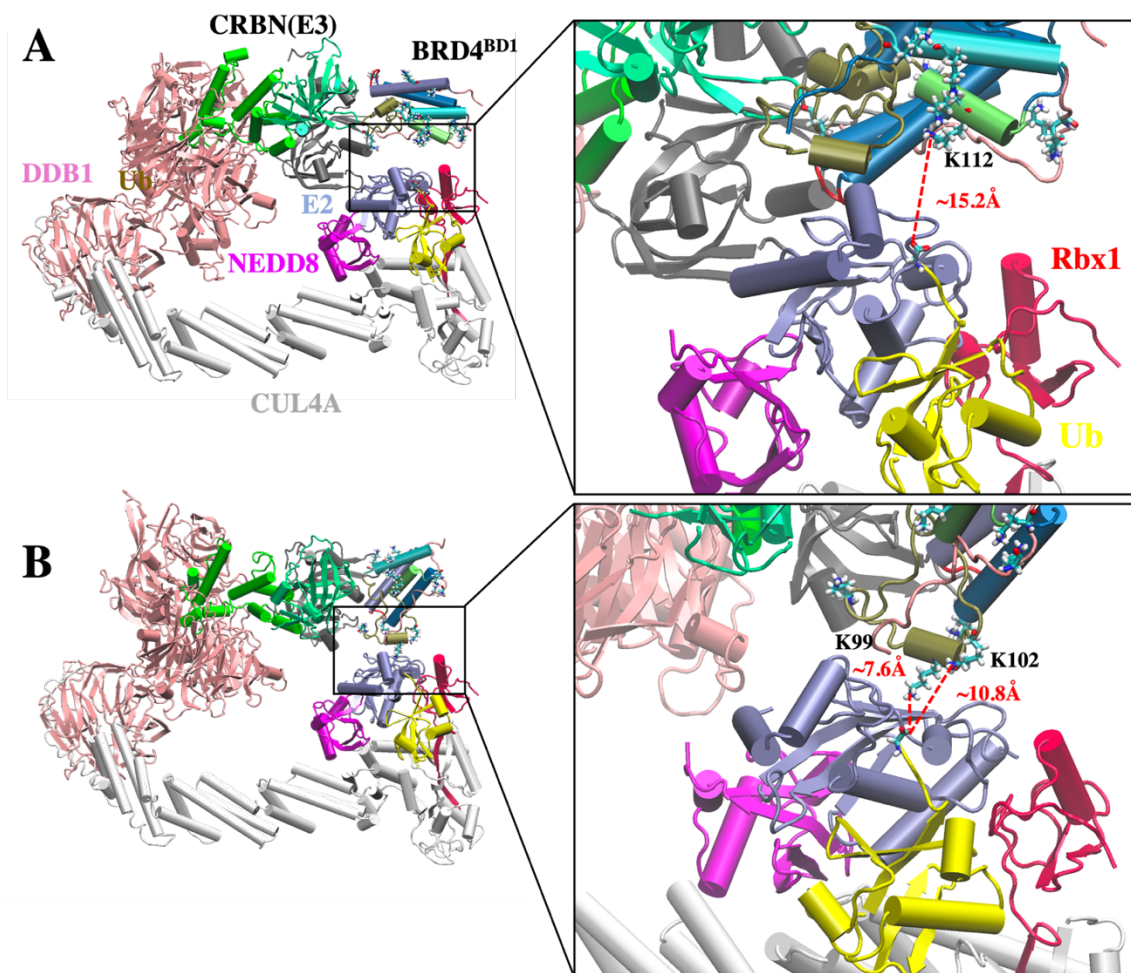


Figure 3.8. CRBN-dBETx-BRD4^{BD1} ensembles align with CUL4A E3 ligase complex to make degradation machinery complex. A, degradation machinery complex A1-DDB1/CUL4A/NEDD8/Rbx1/E2/Ub/CRBN-dBET1-BRD4^{BD1}, distance between the nitrogen of residue K112 in BRD4^{BD1} to oxygen of C-terminal glycine of Ubiquitin is ~15.2 Å. B, degradation machinery complex B1-DDB1/CUL4A/NEDD8/Rbx1/E2/Ub/CRBN-dBET23-BRD4^{BD1}, distance between the nitrogen of residues K99 and K102 in BRD4^{BD1} to oxygen of C-terminal glycine in Ubiquitin are ~ 7.6 Å and ~ 10.8 Å, respectively.

Validation of Filtered CRBN-dBETx_BRD4^{BD1} With Reported Crystal Structures

To perform MD simulations for investigating the CRBN-dBETx_BRD4^{BD1} complex and the degradation machinery complex's functional dynamics, selecting initial structure is

important. Instead of using all the predicted ternary complexes from the previous step, next, we combined the CRBN-dBETx_BRD4^{BD1} complexes from two categories using overlapping diagram representation to identify common CRBN-dBETx_BRD4^{BD1} complexes. (Figure 3.6) Along with that, we also examined the protein protein docking scores of each CRBN-dBETx_BRD4^{BD1} complex. (Table 3.3) Again, with the existing five crystal structures, there is no CRBN E3 reported in the dBET1 ternary complex, and a different set of residues in CTD of CRBN E3 was observed to contact with BRD4^{BD1} in the dBET57 ternary complex. Hence, according to the overlapping diagram representation and MOE docking scores, we came out the conformation #35 for CRBN-dBET1_BRD4^{BD1} complex and the conformation # 9 for CRBN-dBET57_BRD4^{BD1}. For CRBN-dBET6_BRD4^{BD1}, CRBN-dBET23_BRD4^{BD1}, and CRBN-dBET70_BRD4^{BD1}, we superimposed the final ternary complexes with the reported crystal structures to measure the RMSD, validating our predictions. It is interesting to observe our predicted CRBN-dBET23_BRD4^{BD1} and CRBN-dBET70_BRD4^{BD1} are quite close to the crystal structures, with the RMSD of ~1.989 Å in #14 and 2.602 Å in #91, respectively. In the predicted #116 conformation, CRBN-dBET6_BRD4^{BD1}, however, the RMSD is 12.479 Å, having a completely different BRD4^{BD1} and CRBN E3 protein contacts. We then revisited our predicted CRBN-dBET6_BRD4^{BD1} complexes and found that #77 has a similar interprotein contexts with a smaller RMSD value 4.862 Å. (Figure 3.9) We hypothesized that #116 may not be a thermodynamically stable ternary complex (compared to crystal structure), therefore, we wonder how the protein functional dynamics behaves when incorporating these CRBN-dBETx_BRD4^{BD1} complexes into the CRL4A E3 ligase complex to make the

degradation machinery complex. So, we included #116 in CRBN-dBET6_BRD4^{BD1}, along with other CRBN-dBET_x_BRD4^{BD1} complexes for in the following MD-based analysis.

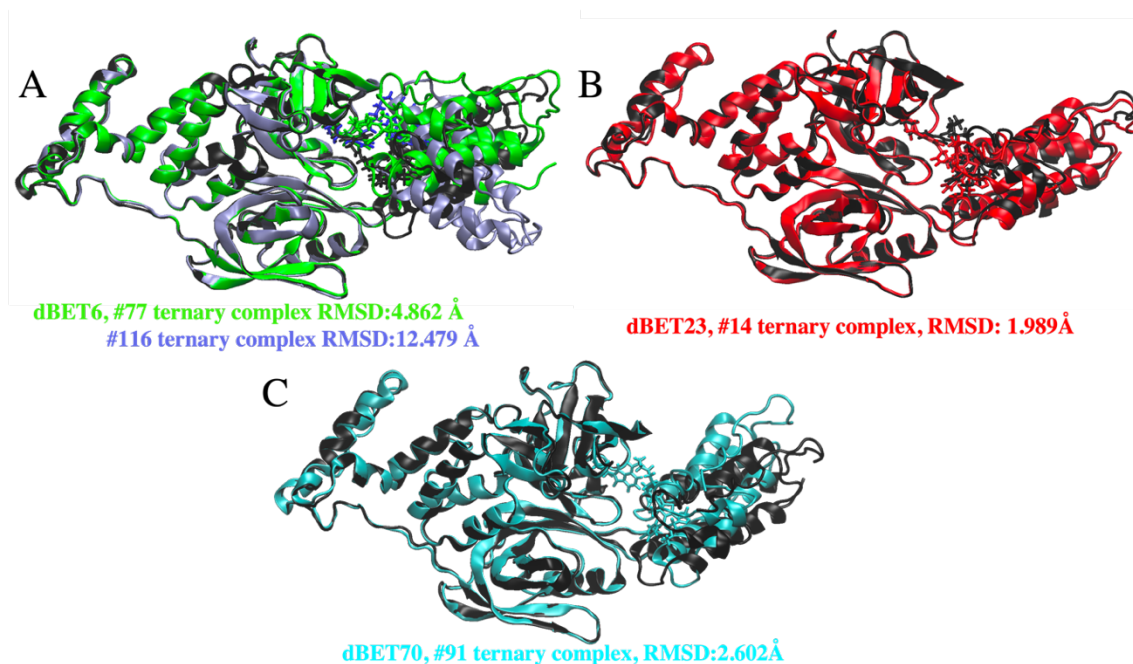


Figure 3.9. Superpositions of predicted CRBN-dBET-BRD4^{BD1} with reported crystal structures (Black). A, predicted CRBN-dBET6-BRD4^{BD1} #77 and #116 have RMSDs of 4.862 Å and 12.479 Å, respectively, with respect to the crystal structure. B, predicted CRBN-dBET23-BRD4^{BD1} #14 has RMSD of 1.989 Å with respect to the crystal structure. C, predicted CRBN-dBET70-BRD4^{BD1} #91 has RMSD of 2.602 Å with respect to the crystal structure.

Degradation Machinery Complex Functional Dynamics Reveals Mechanistic Studies of Target Protein BRD4^{BD1} With E2/Ubiquitin Conjugated Cascade (Catalytic Site)

To answer the questions regarding why dBET6, dBET23, and dBET70 are significantly more efficient in protein degradation than dBET1 and dBET57, we brought atomistic details by studying the functional dynamics of the degradation machinery complex. It was shown that not every lysine residue on the surface of the target protein can be ubiquitinated and shows degradation, due to its inaccessibility to the E2/Ubiquitin catalytic site.^{23,50,51} To test whether the target protein BRD4^{BD1} can bridge the gap in the “ring-forming” and “ring-

open” CRL4A E3 ligase complexes, allowing surface lysine residues accessing E2/Ubiquitin catalytic site easily, a series of MD simulations were performed, included are CRL4A E3 ligase complexes with predicted ternary complexes and crystal structures in both the category A and B. In both the MD simulations of predicted dBET23_14 (Figure 3.10 A and B) and dBET70_90 (Figure 3.10 C and D) CRL4A E3 ligase complexes in category B, we observed that K99 of BRD4^{BD1} in both complexes reaching a negatively charged rich environment consisting of three aspartates and one serine (shown in red jackets in Figure 3.10 B and D), and forms a stable H-bond between the K99 and G75 of ubiquitin (Figure 3.10 A for dBET23_14 and Figure 3.10 C for dBET70_91, respectively). The negatively charged rich environment has been proven to be a critical conserved E2 serine/aspartate (CES/D) site to attract and align positively charged target protein lysine via electrostatic interactions toward this catalytic cleft on E2.^{34,52} Detailed trajectory analysis showed that water molecules enable bridging the ubiquitin G75 and target protein K99, once K99 entering the CES/D catalytic cleft. Our MD simulations in both the predicted ternary complexes of dBET23_14 and dBET70_91 CRL4A E3 ligase complexes, in part, captured the key step contributing to the preference of lysine ubiquitination degradation. In the dBET57_9 CRL4A E3 ligase complex, the accessible target protein lysine residues were observed differently from dBET23_14 and dBET70_91. K55, K57, K111, and K112, instead of K99, were found within the 16 Å distance. (Figure 3.10 E and F) However, none of these lysine residues was observed to reach G75 of ubiquitin within the CES/D catalytic site in our 400 ns MD simulations easily. (Figure 3.10 E) Close attention was paid to the CES/D catalytic region near BRD4^{BD1} where five proline residues

in the loop region (42-59) were observed, which was considered restricting the BRD4^{BD1}'s dynamics movement when approaching to CES/D catalytic site.

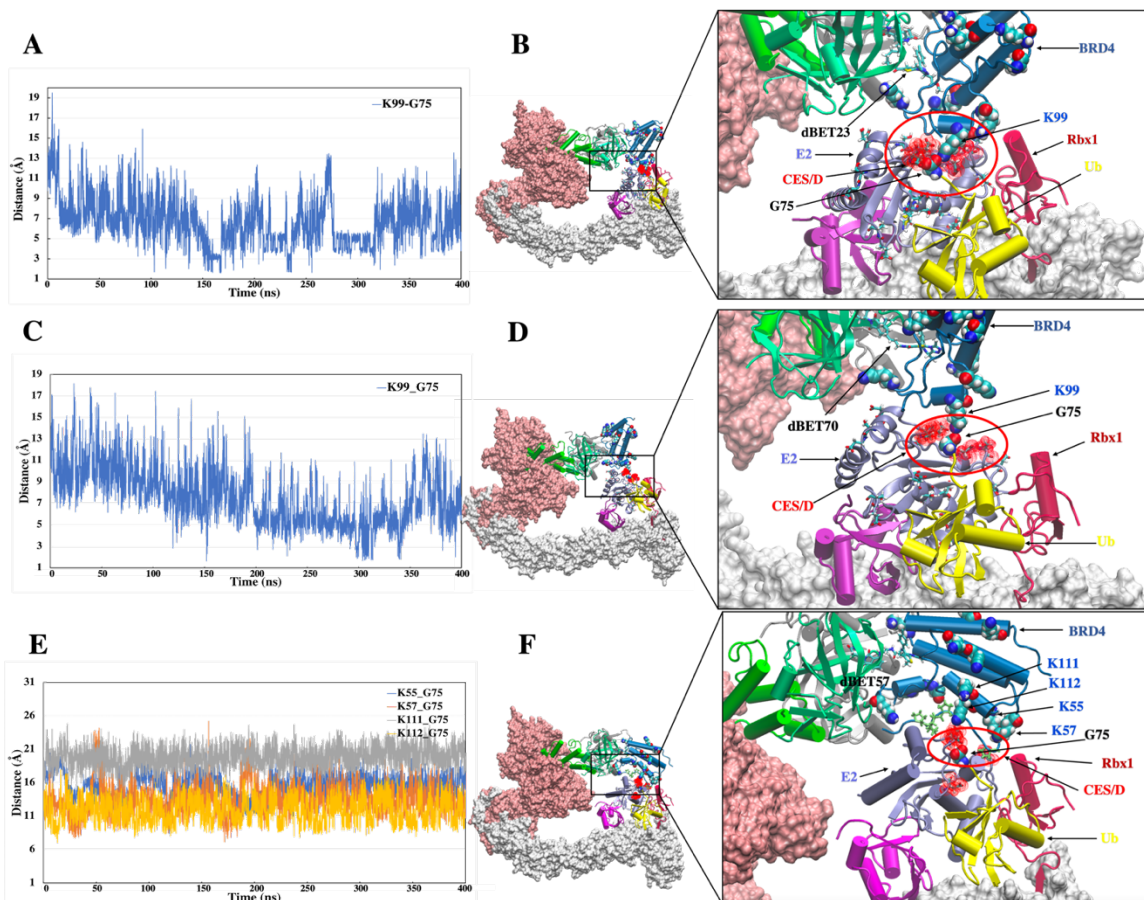


Figure 3.10. Snapshots of three degradation machinery complexes DDB1/CUL4A/NEDD8/Rbx1/E2/Ub/CRBN-dBET_x-BRD4^{BD1} from MD simulations. A, distance measurement of BRD4^{BD1} K99 and ubiquitin G75 for dBET23. B, K99 is shown to reach the CES/D catalytic site of E2 and forms a H-bond between K99 and G75 in the predicted ternary complex (dBET23) MD simulations. C, distance measurement of BRD4^{BD1} K99 and ubiquitin G75 for dBET70. D, K99 is shown to reach the CES/D catalytic site of E2 and forms a H-bond between K99 and G75 in the predicted ternary complex (dBET70) MD simulations. E, distance measurements of BRD4^{BD1} K55, K57, K111, and K112 and ubiquitin G75 for dBET 57. F, K55 and K57 are shown to proximate the CES/D catalytic site of E2 in the predicted ternary complex (dBET57) MD simulations.

Next, we applied PCA to investigate major motion mode of the CRL4A E3 ligase complexes among five PROTACs. In the constructed CRL4A E3 ligase complexes,

E2/Ubiquitin/Rbx1/NEDD8 components were extracted from PDB ID 6TTU and were found attaching to the C-terminal of CUL1 (see method section for details). It is noticeable that there is a hinge loop region locating in the Rbx1, which was proposed to make E2/Ubiquitin/NEDD8 component highly flexible with different loop motions.³⁴ Figure 3.11 shows the first motion mode of dBET6, dBET23, dBET57, and dBET70 in CRL4A E3 ligase complexes in category B) shows the first motion mode of dBET1, dBET6, dBET23, dBET57, and dBET70 in CRL4A E3 ligase complexes in category A. According to the PCA of the entire CRL4A E3 ligase complex in dBET6_77 (Figure 3.11 A), we observed the main motion represents the “breath” motion, in which the hinge loop region of the Rbx1 twists and brings the E2/Ubiquitin/NEDD8 closer to BRD4^{BD1}. Meanwhile, twisting motions were observed in the BPA, BPB, and BPC of DDB1, which permits BRD4^{BD1} swinging to the E2/Ubiquitin cascade. In the CRL4A E3 ligase complex of dBET70_91, (Figure 3.11 B) the main motion mode represents the “open-close” motion, in which two hinge twisting motions were observed in the interface of BPB and BPA/C of DDB1 and the hinge loop region of Rbx1. The synchronizing while twisting movements allow BRD4^{BD1} reaching the CES/D catalytic site more easily. Similarly, two hinge twisting motions were observed in the interface of BPB and BPA/C of DDB1 and E2/Ubiquitin/Rbx1 components in dBET23_14 CRL4A E3 ligase complex, (Figure 3.11 C) which permits the downward movement of BRD4^{BD1} closer to the CES/D catalytic site. In the CRL4A E3 ligase complex of dBET57_9, (Figure 3.11 D) on the contrary, the major motions were observed in the CUL1 region and the DDB1, in which the CUL1 and E2/Ubiquitin/NEDD8/Rbx1 components have twisting motions, while DDB1 and CRBN

E3 have twisting motion in the opposite direction. The BRD4^{BD1} was found to “dock” on the E2 and swinging slightly with synchronization of E2/Ubiquitin/NEDD8 component in the PCA of dBET57_9 CRL4A E3 ligase complex. By comparing the MD simulations and major motions computed from PCA for the PROTACs, it is reasonable to suggest that PROTAC with appropriate linker length (dBET6 and dBET23 with eight carbons vs dBET57 with two carbons) allows the ternary complex rearranging in a suitable spatial orientation and proximate lysine residues within the CES/D catalytic site, which in term is necessary for efficient ubiquitin transfer to occur and hence better degradation.

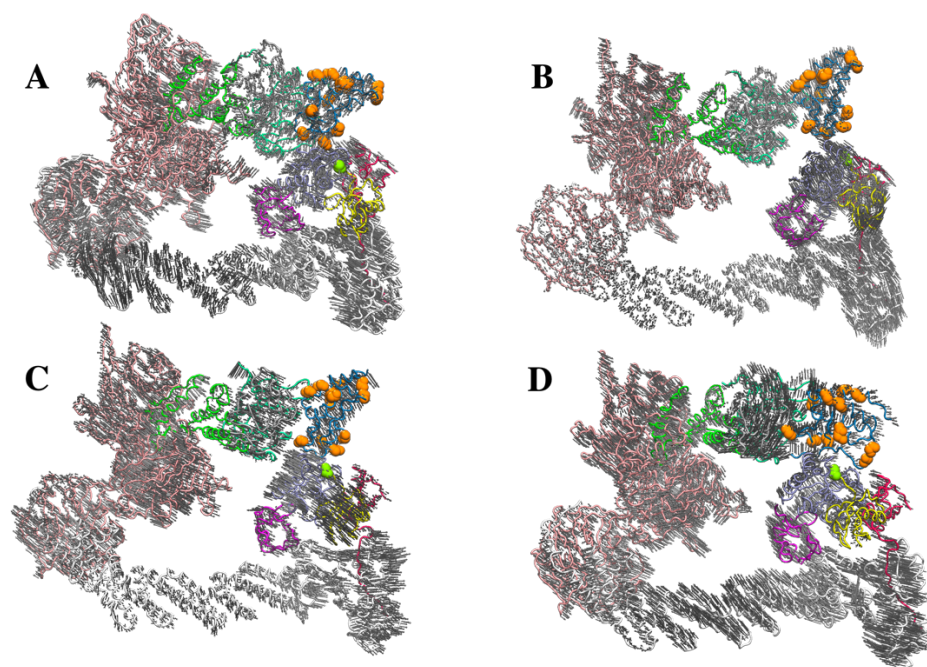


Figure 3.11. PCA major motion modes of predicted ternary complexes. A, dBET6. B, dBET70. C, dBET23. D, dBET57. DDB1 is shown red, CUL4A is shown in white, NEDD8 is shown in magenta, E2 is shown in gray, ubiquitin is shown in yellow. Rbx1 is shown in red, HBD of CRBN E3 is shown in light green, CTD of CRBN E3 is shown in mint, NTD of CRBN E3 is shown dark gray, BRD4^{BD1} is shown in blue.

3.5 Conclusions

In this work, we employed molecular docking, structural alignment and superposition, and MD simulations approaches to investigate BRD4^{BD1} PROTACs-mediated ternary complex stability and the underlying degradation capability among different PROTACs and revealed the functional relationship between the ternary complex formation and target protein ubiquitination. These BRD4^{BD1} PROTACs have the same warheads for the target protein and the E3 ligase, while having significant comparable degradability with different linker length. We started by generating ternary complex ensembles for five different PROTACs. The predicted ternary complex ensembles were carefully examined by filtering out “unreasonable” conformations. We then proceeded to integrate those “reasonable” ternary complexes into the CRL4A E3 ligase complexes by using structural based approach. Each ternary complex was aligned and superimposed with the CRL4A E3 ligase complex to further identify “degradable” ternary complex ensemble. The criterion was based on that at least one lysine residue of the target protein BRD4^{BD1} was proximate to the C-terminal of Ubiquitin. Furthermore, we performed MD simulations, MD-based energy calculation, and principal component analysis to study the ternary complex stability and to demonstrate the protein functional dynamics in the CRL4A E3 ligase complex. Our results revealed indispensable requirements during the ubiquitin catalytic process, in which the orientation of the target protein BRD4^{BD1} toward the E2/Ubiquitin/NEDD8 component but also lysine residue(s) reaching the CES/D catalytic cleft are necessary for efficient ubiquitination. We are interested to further validate our modeling findings by collaborating with experiment groups in our future study. To summarize, our multi-modeling platform

demonstrates a computational toolbox to provide explanations of PROTAC degradability and to broaden our knowledge in PROTAC design.

3.6 References

1. Lai AC, Crews CM. Induced protein degradation: an emerging drug discovery paradigm. *NATURE REVIEWS DRUG DISCOVERY* 2017;16(2):101-114.
2. Wang C, Zhang Y, Xing D, Zhang R. PROTACs technology for targeting non-oncoproteins: Advances and perspectives. *Bioorganic Chemistry* 2021;114:105109.
3. Salami J, Alabi S, Willard RR, Vitale NJ, Wang J, Dong H, Jin M, McDonnell DP, Crew AP, Neklesa TK and others. Androgen receptor degradation by the proteolysis-targeting chimera ARCC-4 outperforms enzalutamide in cellular models of prostate cancer drug resistance. *COMMUNICATIONS BIOLOGY* 2018;1.
4. Lu M, Liu T, Jiao Q, Ji J, Tao M, Liu Y, You Q, Jiang Z. Discovery of a Keap1-dependent peptide PROTAC to knockdown Tau by ubiquitination-proteasome degradation pathway. *EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY* 2018;146:251-259.
5. Liu H, Ding X, Liu L, Mi Q, Zhao Q, Shao Y, Ren C, Chen J, Kong Y, Qiu X. Discovery of novel BCR-ABL PROTACs based on the cereblon E3 ligase design, synthesis, and biological evaluation. *European Journal of Medicinal Chemistry* 2021;223:113645.
6. Sun Y, Zhao X, Ding N, Gao H, Wu Y, Yang Y, Zhao M, Hwang J, Song Y, Liu W. PROTAC-induced BTK degradation as a novel therapy for mutated BTK C481S induced ibrutinib-resistant B-cell malignancies. *Cell research* 2018;28(7):779-781.
7. You I, Erickson EC, Donovan KA, Eleuteri NA, Fischer ES, Gray NS, Toker A. Discovery of an AKT degrader with prolonged inhibition of downstream signaling. *Cell chemical biology* 2020;27(1):66-73. e7.
8. Xie S, Sun Y, Liu Y, Li X, Li X, Zhong W, Zhan F, Zhu J, Yao H, Yang D-H. Development of Alectinib-Based PROTACs as Novel Potent Degraders of Anaplastic Lymphoma Kinase (ALK). *Journal of medicinal chemistry* 2021;64(13):9120-9140.
9. Bond MJ, Chu L, Nalawansa DA, Li K, Crews CM. Targeted Degradation of Oncogenic KRAS(G12C) by VHL-Recruiting PROTACs. *ACS CENTRAL SCIENCE* 2020;6(8):1367-1375.
10. Gu S, Cui D, Chen X, Xiong X, Zhao Y. PROTACs: An Emerging Targeting Technique for Protein Degradation in Drug Discovery. *BIOESSAYS* 2018;40(4).

11. Zhao L, Zhao J, Zhong K, Tong A, Jia D. Targeted protein degradation: mechanisms, strategies and application. *Signal Transduction and Targeted Therapy* 2022;7(1):1-13.
12. Zhong Y, Chi F, Wu H, Liu Y, Xie Z, Huang W, Shi W, Qian H. Emerging targeted protein degradation tools for innovative drug discovery: From classical PROTACs to the novel and beyond. *European Journal of Medicinal Chemistry* 2022;114:142.
13. Zorba A, Chuong N, Xu Y, Starr J, Borzilleri K, Smith J, Zhu H, Farley KA, Ding W, Schiemer J and others. Delineating the role of cooperativity in the design of potent PROTACs for BTK. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA* 2018;115(31):E7285-E7292.
14. Gadd MS, Testa A, Lucas X, Chan K-H, Chen W, Lamont DJ, Zengerle M, Ciulli A. Structural basis of PROTAC cooperative recognition for selective protein degradation. *NATURE CHEMICAL BIOLOGY* 2017;13(5):514-521.
15. Nowak RP, DeAngelo SL, Buckley D, He Z, Donovan KA, An J, Safaei N, Jedrychowski MP, Ponthier CM, Ishoey M and others. Plasticity in binding confers selectivity in ligand-induced protein degradation. *NATURE CHEMICAL BIOLOGY* 2018;14(7):706-714.
16. Lin W, Chen T. General stepwise approach to optimize a TR-FRET assay for characterizing the BRD/PROTAC/CRBN ternary complex. *ACS Pharmacology & Translational Science* 2021;4(2):941-952.
17. Riching KM, Mahan S, Corona CR, McDougall M, Vasta JD, Robers MB, Urh M, Daniels DL. Quantitative live-cell kinetic degradation and mechanistic profiling of PROTAC mode of action. *ACS chemical biology* 2018;13(9):2758-2770.
18. Zengerle M, Chan K-H, Ciulli A. Selective small molecule induced degradation of the BET bromodomain protein BRD4. *ACS chemical biology* 2015;10(8):1770-1777.
19. Roy MJ, Winkler S, Hughes SJ, Whitworth C, Galant M, Farnaby W, Rumpel K, Ciulli A. SPR-measured dissociation kinetics of PROTAC ternary complexes influence target degradation rate. *ACS chemical biology* 2019;14(3):361-368.
20. Jiang F, Wei Q, Li H, Li H, Cui Y, Ma Y, Chen H, Cao P, Lu T, Chen Y. Discovery of novel small molecule induced selective degradation of the bromodomain and extra-terminal (BET) bromodomain protein BRD4 and BRD2 with cellular potencies. *Bioorganic & Medicinal Chemistry* 2020;28(1):115181.

21. Zhou B, Hu J, Xu F, Chen Z, Bai L, Fernandez-Salas E, Lin M, Liu L, Yang C-Y, Zhao Y. Discovery of a small-molecule degrader of bromodomain and extra-terminal (BET) proteins with picomolar cellular potencies and capable of achieving tumor regression. *Journal of medicinal chemistry* 2018;61(2):462-481.
22. Qin C, Hu Y, Zhou B, Fernandez-Salas E, Yang C-Y, Liu L, McEachern D, Przybranowski S, Wang M, Stuckey J. Discovery of QCA570 as an exceptionally potent and efficacious proteolysis targeting chimera (PROTAC) degrader of the bromodomain and extra-terminal (BET) proteins capable of inducing complete and durable tumor regression. *Journal of medicinal chemistry* 2018;61(15):6685-6704.
23. Bondeson DP, Smith BE, Burslem GM, Buhimschi AD, Hines J, Jaime-Figueroa S, Wang J, Hamman BD, Ishchenko A, Crews CM. Lessons in PROTAC design from selective degradation with a promiscuous warhead. *Cell chemical biology* 2018;25(1):78-87. e5.
24. Huang H-T, Dobrovolsky D, Paulk J, Yang G, Weisberg EL, Doctor ZM, Buckley DL, Cho J-H, Ko E, Jang J. A chemoproteomic approach to query the degradable kinome using a multi-kinase degrader. *Cell chemical biology* 2018;25(1):88-99. e6.
25. Vieux EF, Agafonov RV, Emerson L, Isasa M, Deibler RW, Simard JR, Cocozziello D, Ladd B, Lee L, Li H. A method for determining the kinetics of small-molecule-induced ubiquitination. *SLAS DISCOVERY: Advancing the Science of Drug Discovery* 2021;26(4):547-559.
26. Casement R, Bond A, Craigon C, Ciulli A. Mechanistic and structural features of PROTAC ternary complexes. *Targeted Protein Degradation: Methods and Protocols* 2021:79-113.
27. Yang C-Y, Qin C, Bai L, Wang S. Small-molecule PROTAC degraders of the Bromodomain and Extra Terminal (BET) proteins—A review. *Drug Discovery Today: Technologies* 2019;31:43-51.
28. Wang Q, Shao X, Leung ELH, Chen Y, Yao X. Selectively targeting individual bromodomain: drug discovery and molecular mechanisms. *Pharmacological Research* 2021;172:105804.
29. Lu J, Qian Y, Altieri M, Dong H, Wang J, Raina K, Hines J, Winkler JD, Crew AP, Coleman K. Hijacking the E3 ubiquitin ligase cereblon to efficiently target BRD4. *Chemistry & biology* 2015;22(6):755-763.
30. Winter GE, Buckley DL, Paulk J, Roberts JM, Souza A, Dhe-Paganon S, Bradner JE. Phthalimide conjugation as a strategy for in vivo target protein degradation. *Science* 2015;348(6241):1376-1381.

31. Smith BE, Wang SL, Jaime-Figueroa S, Harbin A, Wang J, Hamman BD, Crews CM. Differential PROTAC substrate specificity dictated by orientation of recruited E3 ligase. *Nature communications* 2019;10(1):1-13.
32. Bai N, Miller SA, Andrianov GV, Yates M, Kirubakaran P, Karanicolas J. Rationalizing PROTAC-Mediated Ternary Complex Formation Using Rosetta. *JOURNAL OF CHEMICAL INFORMATION AND MODELING* 2021;61(3):1368-1382.
33. Bai N, Riching KM, Makaju A, Wu H, Acker TM, Ou S-C, Zhang Y, Shen X, Bulloch D, Rui H. Modeling the CRL4A ligase complex to predict target protein ubiquitination induced by cereblon-recruiting PROTACs. *Journal of Biological Chemistry* 2022:101653.
34. Lv D, Pal P, Liu X, Jia Y, Thummuri D, Zhang P, Hu W, Pei J, Zhang Q, Zhou S. Development of a BCL-xL and BCL-2 dual degrader with improved anti-leukemic activity. *Nature communications* 2021;12(1):1-14.
35. Eron SJ, Huang H, Agafonov RV, Fitzgerald ME, Patel J, Michael RE, Lee TD, Hart AA, Shaulsky J, Nasveschuk CG. Structural characterization of degrader-induced ternary complexes using hydrogen–deuterium exchange mass spectrometry and computational modeling: implications for structure-based design. *ACS Chemical Biology* 2021;16(11):2228-2243.
36. Filippakopoulos P, Qi J, Picaud S, Shen Y, Smith WB, Fedorov O, Morse EM, Keates T, Hickman TT, Felletar I. Selective inhibition of BET bromodomains. *Nature* 2010;468(7327):1067-1073.
37. Fischer ES, Boehm K, Lydeard JR, Yang H, Stadler MB, Cavadini S, Nagel J, Serluca F, Acker V, Lingaraju GM and others. Structure of the DDBI-CRBN E3 ubiquitin ligase in complex with thalidomide. *NATURE* 2014;512(7512):49-53.
38. Drummond ML, Henry A, Li H, Williams CI. Improved Accuracy for Modeling PROTAC-Mediated Ternary Complex Formation and Targeted Protein Degradation via New In Silico Methodologies. *JOURNAL OF CHEMICAL INFORMATION AND MODELING* 2020;60(10):5234-5254.
39. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *Journal of molecular graphics* 1996;14(1):33-38.
40. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* 2018;46(W1):W296-W303.

41. Case D, Ben-Shalom I, Brozell S, Cerutti D, Cheatham III T, Cruzeiro V, Darden T, Duke R, Ghoreishi D, Gilson M. AMBER 2018; 2018. University of California, San Francisco.
42. Song D, Wang W, Ye W, Ji D, Luo R, Chen HF. ff14IDPs force field improving the conformation sampling of intrinsically disordered proteins. *Chemical Biology and Drug Design* 2017;89(1):5-15.
43. Wang B, Merz KM. A fast QM/MM (quantum mechanical/molecular mechanical) approach to calculate nuclear magnetic resonance chemical shifts for macromolecules. *Journal of chemical theory and computation* 2006;2(1):209-215.
44. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* 1983;79(2):926-935.
45. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *The Journal of chemical physics* 1995;103(19):8577-8593.
46. Ryckaert J-P, Ciccotti G, Berendsen HJ. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics* 1977;23(3):327-341.
47. Zheng S, Tan Y, Wang Z, Li C, Zhang Z, Sang X, Chen H, Yang Y. Accelerated rational PROTAC design via deep learning and molecular simulations. *Nature Machine Intelligence* 2022;4(9):739-748.
48. Fischer ES, Scrima A, Böhm K, Matsumoto S, Lingaraju GM, Faty M, Yasuda T, Cavadini S, Wakasugi M, Hanaoka F. The molecular basis of CRL4DDB2/CSA ubiquitin ligase architecture, targeting, and activation. *Cell* 2011;147(5):1024-1039.
49. Fischer ES, Böhm K, Lydeard JR, Yang H, Stadler MB, Cavadini S, Nagel J, Serluca F, Acker V, Lingaraju GM. Structure of the DDB1–CRBN E3 ubiquitin ligase in complex with thalidomide. *Nature* 2014;512(7512):49-53.
50. Petzold G, Fischer ES, Thomä NH. Structural basis of lenalidomide-induced CK1 α degradation by the CRL4CRBN ubiquitin ligase. *Nature* 2016;532(7597):127-130.
51. Mattioli F, Sixma TK. Lysine-targeting specificity in ubiquitin and ubiquitin-like modification pathways. *Nature structural & molecular biology* 2014;21(4):308-316.
52. Valimberti I, Tiberti M, Lambrugh M, Sarcevic B, Papaleo E. E2 superfamily of ubiquitin-conjugating enzymes: constitutively active or activated through phosphorylation in the catalytic cleft. *Scientific reports* 2015;5(1):1-13.

Chapter 4 CACNA1C Is A Biomarker and Therapeutic Target for Patients With Ovarian Cancer

4.1 Abstract

Ovarian cancer (OC) is the eighth-most common cause of death among women worldwide. In the absence of distinctive symptoms in the early stages, most women are diagnosed in advanced stages of the disease. Surgical debulking and systemic adjuvant chemotherapy remain the mainstays of treatment, with the development of chemoresistance in up to 75% of patients with subsequent poor treatment response and reduced survival. Therefore, there is a critical need to revisit existing, and identify potential biomarkers that could lead to the development of novel and more effective predictors for ovarian cancer diagnosis and prognosis. In this study, using a combination of Single gene mRNA sequencing data and corresponding clinical information obtained from The Cancer Genome Atlas Database (TCGA) and IHC from City of Hope (COH) cohort, we identified CACNA1C gene that encodes the voltage-dependent calcium ion transmembrane channel hCaV1.2, as an OC biomarker. We then performed homology modeling, molecular docking, and molecular dynamics simulation studies to investigate a Food and Drug Administration–approved calcium channel blocker Nifedipine (NIFE) in the hCaV1.2 Channel. Finally, we showed that NIFE suppressed the growth of different human ovarian cancer cell lines *in vitro*. Our findings provide direct evidence that Calcium Channel Blockers (CCB) could be promising clinical therapy to treat patients with advanced OC by affecting the tumor itself. Hence,

incorporating CCBs may provide a safe and effective means of improving clinical outcome in OC patients.

4.2 Introduction

Ovarian cancer (OC) is the third most common gynecologic malignancy worldwide, with approximately 21,750 newly estimated cases and 13,940 newly estimated death in USA, 2020.¹ Despite other cancers such as endometrial cancer (EC) having higher rates of incidence, OC is the deadliest of all female reproductive cancers.² This disease is rare in young women, especially under the age of 30 and its risks will increase with age, sharply after the age of 50. So, the average diagnosis age of OC is between 50 and 70.² Because the clinical manifestations of early OC are hidden or unspecific, OC is often called a silent killer, and approximately 70% of patients are diagnosed with an advanced stage.³ As reported, the 5-year survival rate for FIGO stage I is 90%, stage II is 65%, stage III is 34%, and stage IV is 15%.⁴ However, current improvements in treatment methods could only slightly improve OC survival rates.⁵ Indeed, there was an urgent need to clarify the potential therapeutical treatments of OC and identify novel biomarkers for early diagnosis and prognosis assessment.⁶ Calcium channels could be generally classified into two categories containing the voltage-gated calcium channels (VGCCs) and the ligand-gated calcium channels (LGCCs).⁷ As a type of VGCCs, calcium voltage-gated channel subunit alpha1C (CACNA1C) was coded by the $\alpha 1$ subunit Cav1.2 and it was reported to be involved in regulating cell-matrix adhesion, collagen fibril organization, cell adhesion, cellular response to amino acid stimulus, and negative regulation of cell proliferation.⁸ Previous meta-analysis results showed that CACNA1C was up regulated in brain tumors,

leukemia, breast cancer and other tumors, suggesting its regulatory roles in cancer progression.⁹ Bioinformatics analysis also revealed that mutation of CACNA1C was significantly associated with longer overall survival (OS) in EC patients.¹⁰ All of these indicated that CACNA1C could be served as a prognostic biomarker for human cancers. Therefore, this study aims to explore the prognostic values of CACNA1C in OC, to identify CACNA1C-related signal pathways and to further explore its associations with immunity, thus helping to provide some references for future basic CACNA1C related research in OC.

4.3 Methods and Material

Homology Modeling of the Human Calcium Channel (hCaV1.2)

The 3D molecular structure of hCaV1.2 channel was built using the homology modeling method.^{11,12} The recent reported cryo-EM structure of the α_1 subunit of the rabbit CaV1.1 channel (PDB:6JP5)¹³ was used as the template to build the hCav1.2 molecular structure in this study. In short, there are 37 sequence isoforms reported in the UniProt database (<https://www.uniprot.org/uniprot/Q13936>), and the isoform 1 (identifier: Q13936-1) was selected for the homology model construction. The BLAST search identified 50 models, and we selected the model rCaV1.1 having the sequence identity of 72.35% and with the highest GMQE value (0.41) for the construction. The structure of the constructed hCaV1.2, comprising residues L105 to L1649 is depicted in Figure 4.1.

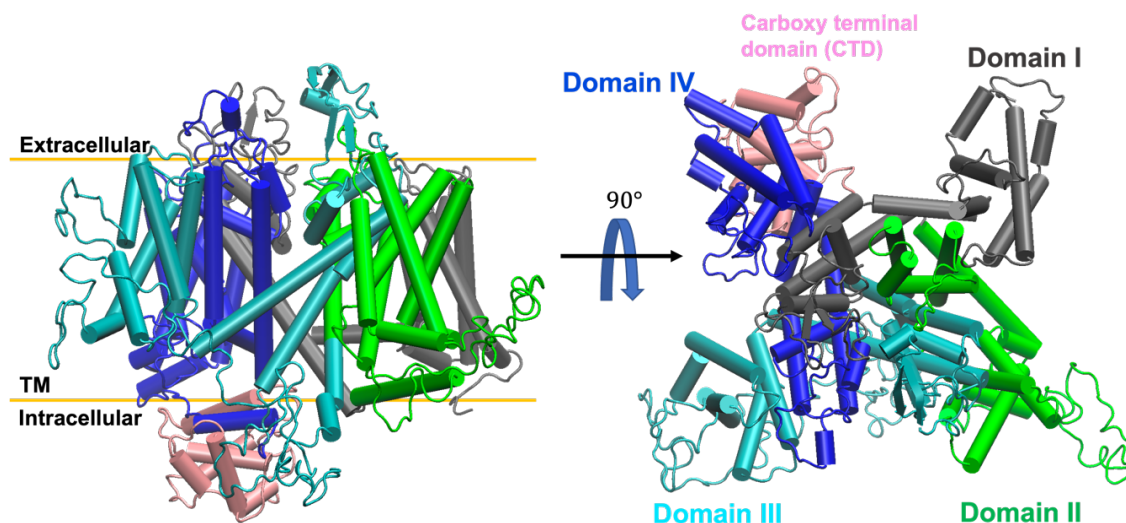


Figure 4.1. Structure of the constructed hCaV1.2 channel. Structure of the hCaV1.2 channel generated by homology modeling from lateral view (left) and rotate 90° from extracellular view (right). Scheme of cartoon representation for hCaV1.2 structure with four domains and carboxy terminal domains are indicated; domain I in gray, domain II in green, domain III in cyan, and domain IV in blue. Each domain is composed of six transmembrane helices (S1-S6), where the first four, S1-S4, formed the voltage-sensing domain (VSD) while the other two, S5-S6, formed the pore domain (PD). The extracellular segment that connects helices S5 and S6 formed two re-entrant short helices denominated P1 and P2.

Atomistic Molecular Dynamics Simulations of the hCaV1.2-Apo Structure

The constructed hCav1.2-apo molecular structure was subjected to energy minimization to clear possible clashes during the modeling step. Briefly, solvent molecules such as alcohol, DMSO *etc.*, were removed during the system preparation. Missing hydrogens were added using Tleap implemented in the AMBER 20 package.¹⁴ The molecular structure was minimized in a three-step process: hydrogen, side chain, and the full protein, with 500, 1000 and 5000 steps, respectively. The resulting minimized structure was used in the following molecular docking calculation.

Molecular Docking Calculation of NIFE Compound in the hCaV1.2 Channel

The molecular docking calculation of the NIFE and the modelled hCaV1.2 structure were performed using Molecular Operating Environment (MOE).¹⁵ The constructed hCaV1.2 molecular structure was selected from the energy minimized step, as described in the previous section. The 3D DHP-based structure, NIFE, was obtained from the PubChem database (ID:4485). During the molecular docking calculation, ligand binding sites were first searched by the implemented MOE-site finder tool using AlphaShapes.¹⁶ There were 116 sites found from the searching step and three sites (#3, #21 and #73 among 116) within the pore of the hCaV1.2 were determined as the subsequent docking site. (Figure 4.2). AM1BCC charge model¹⁷ was used to assign charges for NIFE. The docking algorithm in MOE allowed induced fit and free side chain rotation. London dG and GBVI/WSA dG scoring functions were selected. The resulting five binding modes from each site for the ligand with the lowest scorings were determined for the subsequent binding mode analysis. (Table 4.1)

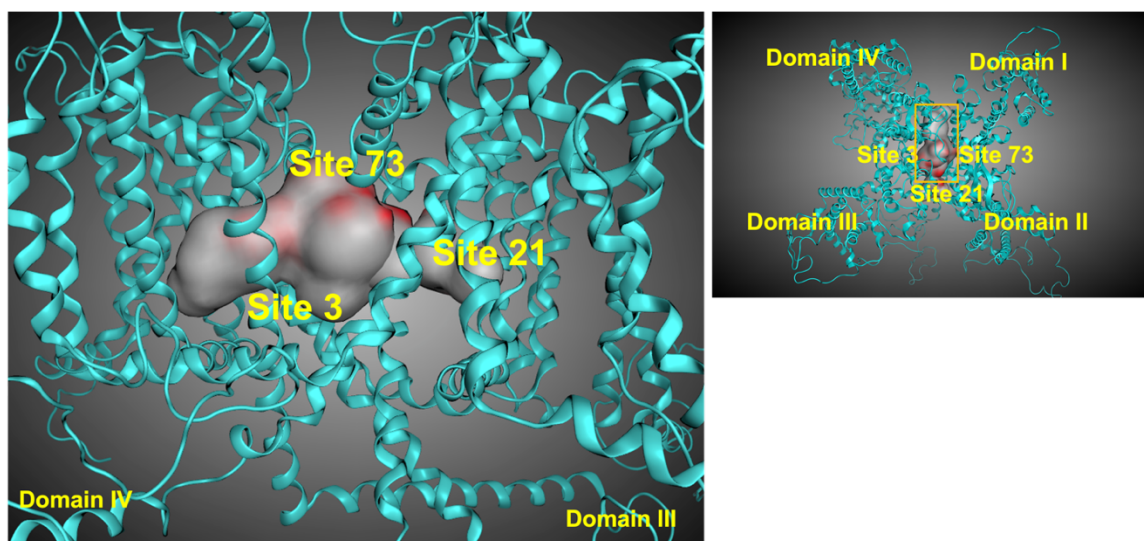


Figure 4.2. Ligand binding sites in the structure of the constructed hCaV1.2 channel. Three sites were found within the pore of the hCaV1.2 channel. Site numbers were indicated by the order found in the site finder tool.

Table 4.1: Binding modes with the relative scorings of NIFE.

Binding mode	Score (kcal/mol) (Nifedipine)
Site 3 mode 1	-6.89
Site 3 mode 2	-6.82
Site 3 mode 3	-6.76
Site 3 mode 4	-6.75
Site 3 mode 5	-6.67
Site 21 mode 1	-7.30
Site 21 mode 2	-7.22
Site 21 mode 3	-7.17
Site 21 mode 4	-7.10
Site 21 mode 5	-7.00
Site 73 mode 1	-7.72
Site 73 mode 2	-7.29
Site 73 mode 3	-7.24
Site 73 mode 4	-7.20
Site 73 mode 5	-7.10

Calcium Ion Channel Membrane Assembly

The CHARMM-GUI Membrane Builder was used to build the membrane system in both the apo-form and NIFE-bound structures priori to MD simulations.¹⁸⁻²² We selected the binding mode 2 in the site 21 as our structure from the docking result. The heterogeneous lipid bilayer was created by choosing palmitoyl-oleoyl-phosphatidylcholine (POPC) lipid with the ratio of x length and y length in 200 Å in the inner and outer leaflets of the membrane. The systems were hydrated with 20.0 Å thickness water on the top and lower leaflets. The system size along the X and Y dimensions are set as 190 Å and 150 Å for Z dimensions. We added ionic concentration of 150 mM sodium chloride solution into the rectangle simulation box. We used Ligand Reader and Modeler to generate the parameter

and force field for ligand NIFE.²³ The fully assembled systems comprised about 513,000 atoms for both structures and were subject to NPT ensemble simulation setup at 300 K temperature. Parameters and inputs in the overall systems, including the calcium channel protein, lipid bilayer, water, ionic concentration, and ligand in the NIFE-bound structure were assigned using AMBER force field generator, and AMBER *14FFSB* force field for protein and GAFF2 for ligand were selected in the force field options.

Classical MD Simulation

The MD simulation protocol generated by the CHARMM-GUI employs two consecutive NVT and NPT equilibration steps after energy minimization. The minimization step included 2500 steps of steepest descent with 5000 minimization cycles, in which proteins, ligands and lipid head groups were restrained. In the NVT equilibration step, simulations were performed with a time step at 1 fs for 250 ps, followed by NPT equilibration step with a time step at 2 fs for 2 ns. Heating of the systems from 50 K to 300 K with an increment of 50K was achieved in 200 ps (time step of 2 fs) at each temperature. Long-range electrostatic interactions were computed by the particle mesh Ewald method²⁴. The SHAKE constraint was applied all atoms including hydrogens²⁵. The Langevin thermostat with a damping constant of 2 ps⁻¹ was applied to maintain a temperature of 300 K. A cut-off of 12 Å was used for the vdW and short-range electrostatic interactions. All the simulations were performed under the periodic boundary conditions. The final production MD runs were performed under NPT ensemble for 200 ns (time step 2 fs), and each frame was collected at 10 ps time interval.

Gene Expression and Survival Analyses Using TCGA

Whole-transcriptome sequences of 302 primary ovarian cancer was download from TCGA (TCGA-BRCA, <https://portal.gdc.cancer.gov/projects/TCGA-BRCA>). The “high” and “low” groups were segregated based on median mRNA expression values. Kaplan–Meier survival analysis was used to determine the survival differences between “high” and “low” mRNA expression groups, which were visualized by Kaplan–Meier plots and compared using Cox regression analysis, with *P*-values calculated by log-rank test using the Survival package in R [Ref]. The survival differences were considered to be statistically significant when *P*-values were < 0.05.

Cell Culture, Chemicals, and Reagents

NIFE was obtained from Fujifilm Wako (Richmond, VA). Cisplatin was from Teva Pharmaceuticals (Parsippany, NJ). Dulbecco's Modified Eagle Medium (DMED or RPMI) and HL-60 cells were purchased from American Type Culture Collection (ATCC) (Manassas, VA, USA). Fetal bovine serum (FBS), and penicillin–streptomycin was purchased from Sigma (St. Louis, MO, USA). IMDM contains 4 mM L-glutamine, 4500 mg/L glucose, and 1500 mg/L sodium bicarbonate, and is supplemented with 10% (v/v) FBS, and 1% (W/V) antibiotics. The live cells were incubated at 37°C in a 5% CO₂ incubator (Thermo Scientific, Waltham, MA, USA).

Cell Treatment

The experiments were conducted with three replicates for each control or treatment group. Uniform cell density of 5×10^4 cells/mL was maintained for all the treatments and controls. Cells were treated with 312 μ M of NIFE or cisplatin or left untreated and

incubated for 72h at 37°C in a humidified 5% CO₂ incubator (Thermo Scientific, Waltham, MA, USA).

Cell Proliferation Assay

Cell viability was assessed with CellTiter 96° AQueous One Solution Cell Proliferation Assay kit from Promega (Madison, WI, USA) with few modifications. Briefly, 100 μ L aliquots of the treated or untreated cell suspension were seeded into 96-well polystyrene tissue culture plates and 20 μ L of assay reagent was added to each well. After 60 min of incubation at 37°C, the absorbance was read at 490 nm with a 96 well plate reader from BMG LABTECH GmbH (Ortenberg, Germany).

Immunohistochemistry

Formalin-fixed paraffin-embedded tissue blocks were sectioned at a thickness of 4 μ m and put on positively charged glass slides. IHC stains were performed on Ventana Discovery Ultra (Ventana Medical Systems, Roche Diagnostics, Indianapolis, USA) IHC Automated Stainer. Briefly, the slides were deparaffinized, rehydrated and incubated with endogenous peroxidase activity inhibition reagents and antigen retrieval solution. The anti-human Cav1.2 polyclonal primary antibody (1:400; NOVUS BIOLOGICALS INC, CO, USA) was incubated following by DISCOVERY anti-rabbit HQ and DISCOVERY anti-HQ-HRP (Ventana) incubation. The stains were visualized with DISCOVERY ChromoMap DAB Kit and counterstained with hematoxylin (Ventana) and cover slipped. After IHC staining, whole slide images were acquired with a Ventana iScan HT Scanner (Roche) and viewed by iScan image viewer software.

Assessment of Immunostaining

Images were scanned by a Hamamatsu nano zoomer scanner and analyzed using Visiopharm software. In each image we used custom APP to measure or count the area of an image analyzed, and the number of cells positive for each marker within that area. This quantitative information was further calculated.

Statistical Analysis

All data were derived from ≥ 3 independent experiments. Statistical analyses were conducted using GraphPad Prism 9.4.1 (GraphPad Software, Inc., San Diego, CA). Values were calculated as mean $\pm$ standard error of the mean. Significant differences between the groups were determined using the unpaired one-way analysis of variance test. $P < 0.05$ was considered to indicate a statistically significant difference.

4.4 Results and Discussion

CACNA1C mRNA Gene Expression TCGA OV Cohort Validation

To explore clinical associations of CACNA1C, we validated the mRNAseq gene expression among 302 TCGA (The Cancer Genome Atlas) Ovarian (OV) patient cohort. The high mRNA expression of CACNA1C was shown statistically significant correlation with overall survival with p-value = 0.00369 and to predict patient clinical outcomes. (Figure 4.3)

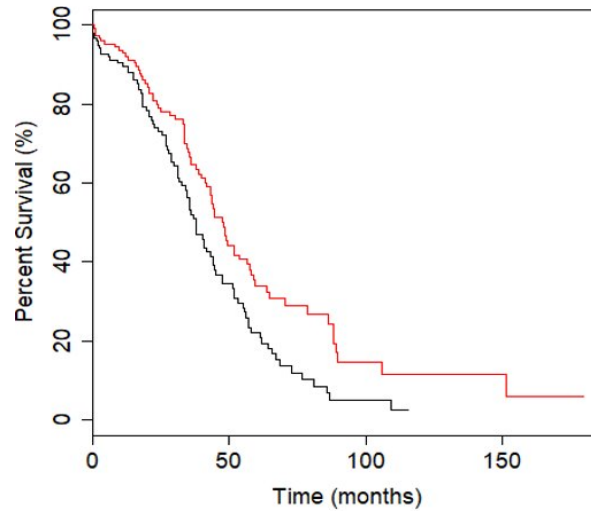


Figure 4.3. CACNA1C mRNA gene expression mature strands was analyzed as reads-per-million-per-reads; data were downloaded from TCGA using TCGA Assembler.¹³ Of 302 mRNAseq. Kaplan-Meier Overall survival analysis shows statistically significant correlated with patient outcome with p-value= 0.00369.

COH Patient Characteristics

Two antibodies were applied to stain OC patients tissue samples and the patients samples were compared with normal OC tissue samples shown in Figure 4.4.

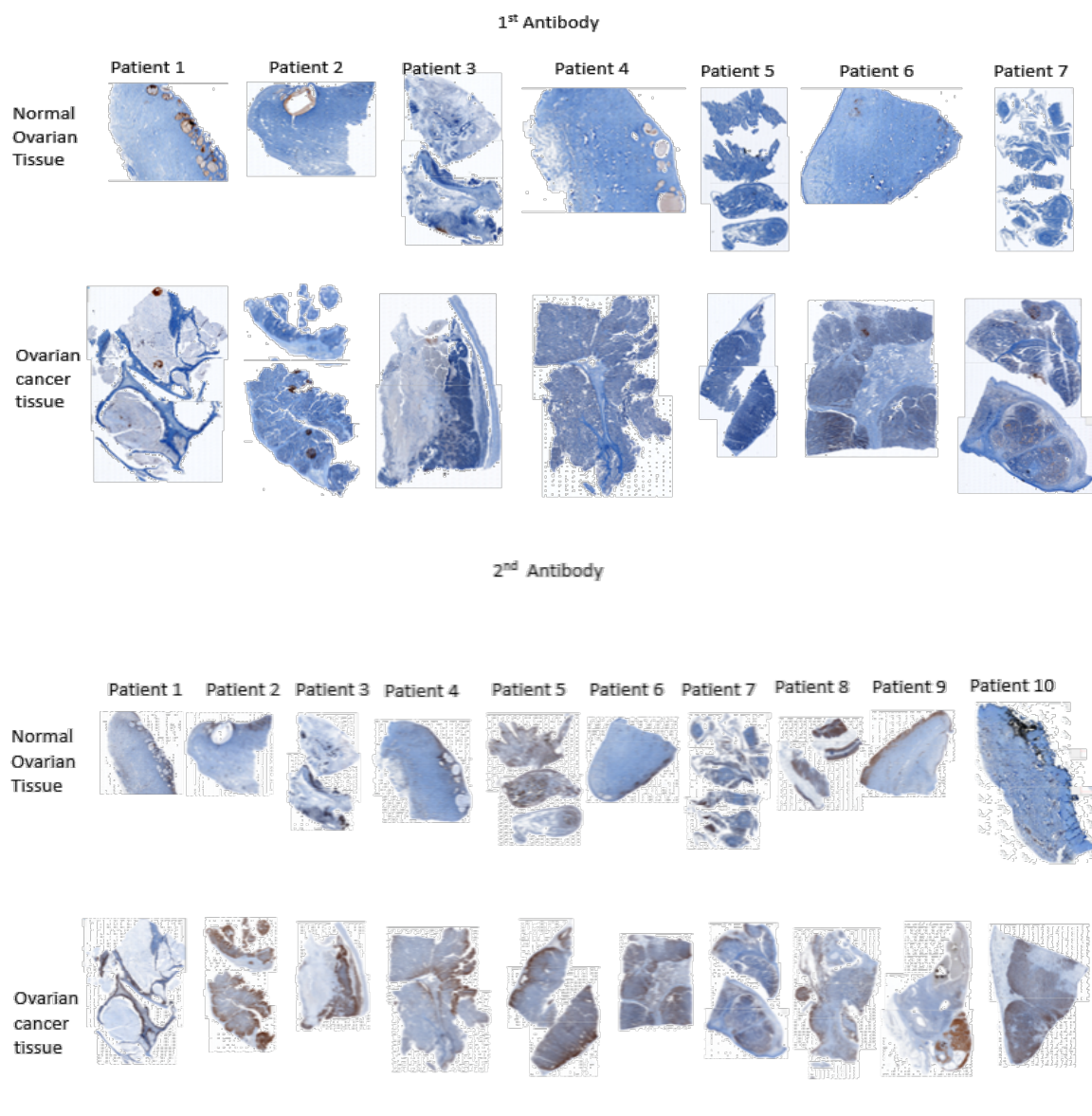


Figure 4.4: IHC of hCaV1.2 in paired OC sections

NIFE Kills OC Cells

Functional experiments on NIFE, calcium antagonist widely used in the clinic, showed that it had inhibitory effects on OC cell proliferation (Figure 4.5). Notably, the effects of NIFE

were much more significant than the standard of care Cisplatin (CIS) at the same dose (312 μ m) on 3 different human OC cell lines including the CIS-resistant cell line A2780R.

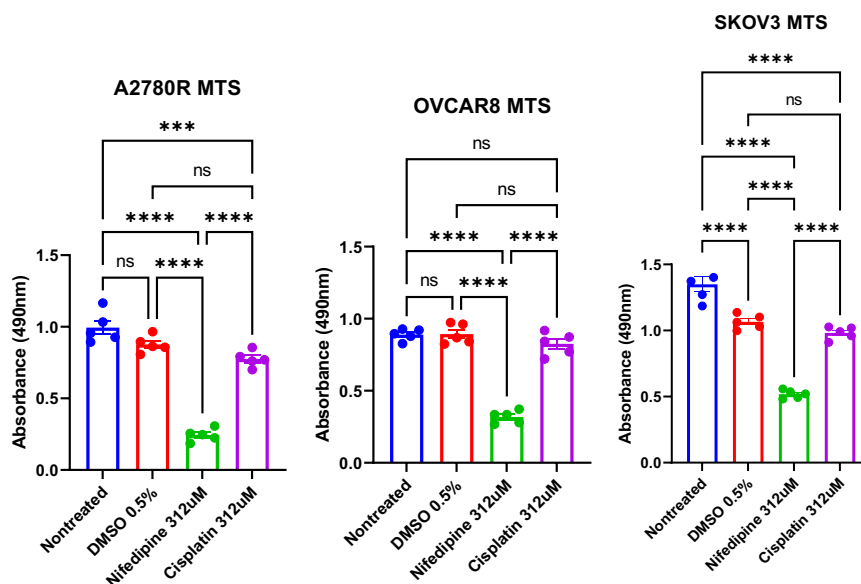


Figure 4.5. Cytotoxic effect of NIFE on human ovarian cancer cell lines (A2780R, OVCAR8 and SKOV3). Cells were treated with 312 μ M of NIFE for 72h. Cell viability was tested by cell proliferation assay using a spectrophotometer at 490 nm. Results were expressed as means of three independent experiments $\pm$ standard error mean. p-values less than 0.05 were considered statistically significant.

Molecular Docking Calculations of NIFE in hCaV1.2 Constructed Structure

To investigate and understand the binding mechanism of compounds in human CaV1.2 channel, we applied molecular docking in current study. Due to the lack of the molecular structure of human CaV1.2 channel in the protein data bank, we built our hCaV1.2 molecular structure using the homology modeling.¹¹ The structure of the constructed hCaV1.2, comprising residues L105 to L1649 is depicted in Figure 4.1. The 3D dihydropyridine-based structure, NIFE, was obtained from the PubChem database and assigned partial charges using AM1BCC. To dock the ligand into the constructed hCaV1.2

model, we performed MOE-site finder to search for possible binding sites. Among a total of 116 searched sites, three sites (site #3, #21 and #73) located in the pore domain of the channel were considered for the subsequent docking study. Site #3 and #21 located in the fenestration enclosed by domains III and IV, except that site #3 covered a wider range in the pore domain. Site #73 located in the center cavity, slightly off the pore axis toward the fenestration formed by the domains III and VI. It is surprising that our site-finder search method identified these three sites, which were also observed in other calcium channel studies.^{13,26-29} Ligand docking calculations were performed in these three sites and the resulting five binding modes from each site for NIFE with the lowest scorings were determined for the subsequent binding mode analysis. (Table 4.1) Scores calculated from site #21 were the lowest among three sites for the ligand.

The overall binding modes based on the scoring consideration were subjected to further analysis. (Figure 4.6) In the NIFE binding modes, all 15 poses (five from each site) located in the fenestration enclosed by domains III and IV. (Figure 4.6 A). According to the previous reports and other crystal structures,^{13,26} poses locating in the fenestration enclosed by domains III and IV, close to helical segments S5_{III}, S6_{III}, S6_{IV} and P1_{III} were further analyzed.

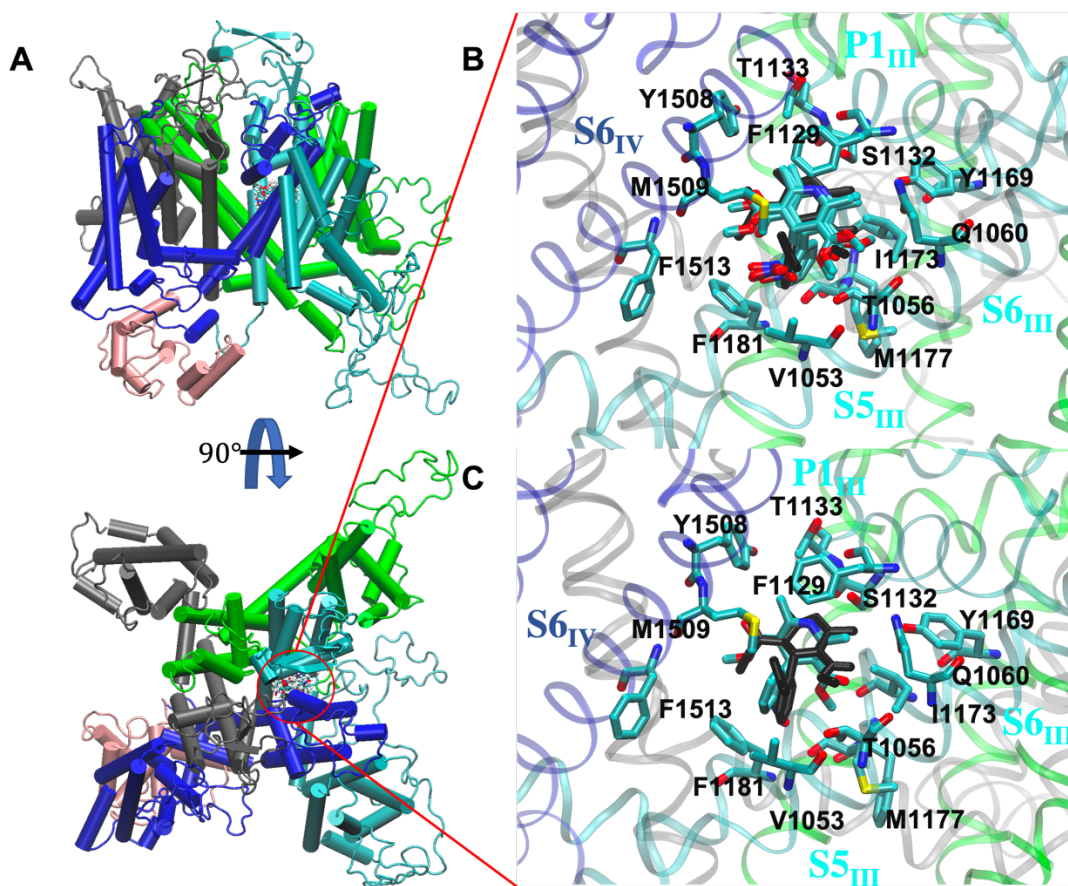


Figure 4.6. NIFE binds to the fenestration enclosed by domain III (cyan) and IV (blue) of hCaV1.2. A. 15 binding modes were overlaid in the fenestration enclosed by domain III (cyan) and IV (blue). B. Four binding modes were superimposed with the crystal structure (black) (PDB ID: 6JP5). C. One binding mode was superimposed with the crystal structure (black), nitrophenyl ring rotating 180°.

In the NIFE binding modes, four molecular poses (site #21 mode 2 and 3, site #73 mode 1 and 3, bold in Table 4.1) were found to closely agree with the reported crystal structure (PDB ID: 6JP5).¹³ (Figure 4.6 B) In the structure, the whole molecule was embedded in a hydrophobic core formed by Phe1129 and Thr1133 on P1_{III}, Val1053 on S5_{III}, Tyr1169, Ile1173, Met1177 and Phe1181 on S6_{III}, and Tyr1508, Met1509 and Phe1513 on S6_{IV}. The N1 amine in the dihydropyridine group formed a H-bond with the hydroxyl group of

Ser1132 on P1_{III}. Thr1056 and Gln1060 on S5_{III} each formed a H-bond with two oxygens from the C3-ester in the NIFE. In addition to the polar interactions, non-polar interactions were also observed; Phe1129 was in parallel with the dihydropyridine group, forming π - π interaction. The phenyl ring in the nitrophenyl group was T-shaped stacking to the Phe1181. Interestingly, we observed one binding mode (site #21 mode 1) (Figure 4.6 C) that was also overlaid with the crystal structure, except the nitrophenyl ring rotating 180°. The whole ligand molecule was still accommodated in the hydrophobic pocket as described above.

Atomistic Molecular Dynamics Simulations of the Apoprotein and NIFE Bound Complex

To gain a better understanding of the NIFE interactions in the hCaV1.2 complex for further drug development, we performed MD simulations to examine protein conformational changes before and after NIFE binds. We therefore applied MD simulations for both NIFE-hCaV1.2 complex and apoprotein (hCaV1.2 protein only). Since the hCaV1.2 ion channel is a membrane bound protein, the structure was embedded in a POPC bilayer with water molecules and ions. (Figure 4.7) (See methods for details) The time-dependent root-mean-square-deviation (RMSD) results showed that apoprotein reached the equilibrium at 80 ns, whereas NIFE bound complex reached the equilibrium at 20 ns in the 200 ns MD simulations. (Figure 4.8) Inlet in the Figure 4.8 is the RMSD results for residues in the ligand bound pocket in both systems. It showed that NIFE bound pocket reached the equilibrium at 50 ns in the NIFE bound structure, but it took about 100 ns for the apoprotein

to reach equilibrium. Our RMSD results implied that the NIFE bound structure stabilized and reached a local energy minimum quicker than the apoprotein structure, but two structures reached stability in our 200 ns MD simulations.

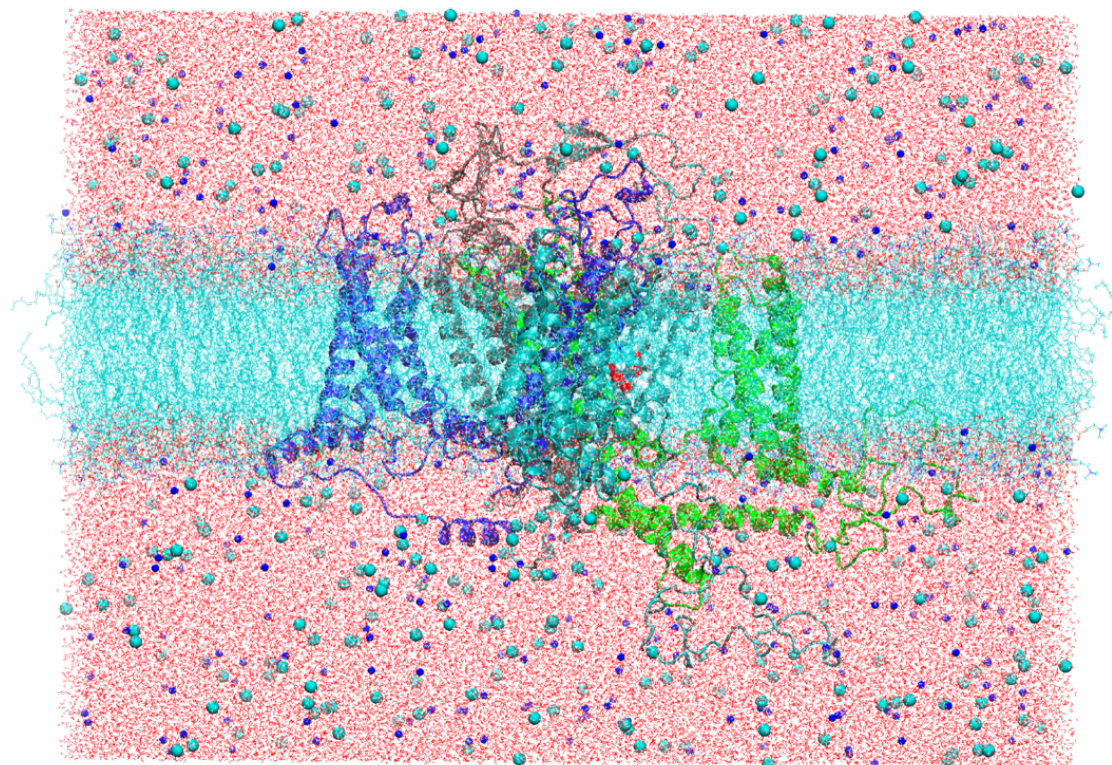


Figure 4.7. Snapshot of the constructed hCaV1.2 ion channel embedded in a POPC bilayer with water molecules, sodium, and chloride ions. Four subdomains (gray for domain I, green for domain II, cyan for domain III, and blue for domain IV) are shown. Ligand NIFE (VDW representation) is encapsulated between the domain III and IV, which is circled in red in the figure.

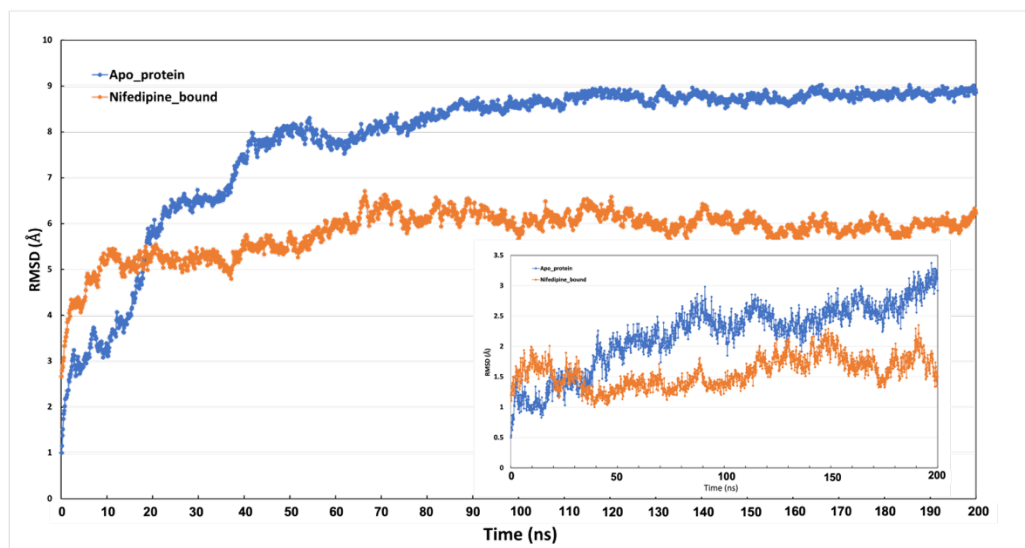


Figure 4.8. Comparison of the RMSD over 200 ns MD simulations between the Apoprotein and NIFE bound structures. Inlet is the RMSDs of ligand bound pocket.

The exploration of the rotameric states of the residues provides useful information regarding the protein flexibility and protein conformational changes and dynamics. To better understand the interactions between the NIFE and the residues, we applied T-analyst program to study the residues' rotameric states in the binding pocket in both the NIFE-hCaV1.2 complex and the apoprotein.³⁰ We focused on the residues surrounding the NIFE binding pocket; Val1053, Phe1129, Ser1132, Phe1176, Met1178, Phe1181, Met1509, as these key residues were reported to be involved in mutagenesis experiments with known LTCC antagonists.²⁹ (Figure 4.9) The rotameric state of two key residues Phe1129 and Ser1132 on P1_{III} were reported and compared (Figures 4.9 C, D for apoprotein and Figures 4.9 F, G for NIFE-bound structure). Phe1129 maintained one rotameric state in the whole MD simulations by forming a strong π - π interaction with the dihydropyridine ring in the NIFE, (Right panel, Figure 4.9 B and F) while it was observed to flip its phenyl ring 90° in the apoprotein structure during the simulation. (Left panel, Figure 4.9 A and C) Previous

observation showed that Ser1132 has a strong binding affinity with the DHP antagonists,^{31,32} hence, its hydroxyl group formed a stable H-bond with the N1 amine in the dihydropyridine group of NIFE in the NIFE bound structure, (Figure 4.9 G) but rotated away in the apoprotein structure, alternatively forming H-bonds with Glu1060 on S5_{III} and Try1169 on S5_{IV}. (Figure 4.9 D) Phe1181 on S6_{III} formed a stable T-shaped aryl-aryl interaction with the phenyl ring in the nitrophenyl group of NIFE, as indicated by one rotameric state, (Figure 4.9 H) but rotated alternatively to fill the ligand binding pocket when NIFE was absent. (Figure 4.9 E). It was observed that there are three rotameric states of Val1053 in the apoprotein structure (Figure 4.10 A), whereas there was one rotameric state of Val1053 (Figure 4.10 E) in the NIFE bound structure, indicating that Val1053 was more flexible in the apoprotein as it lost its ligand bound coordination when ligand was absent. The hydrophobic residues Phe1176 and Met1178 on S6_{III} had each one rotameric state in the NIFE bound structure, (Figure 4.10 F and G, respectively) as they were observed to be pushed outwards of the binding pocket with the presence of the NIFE. Two rotameric states were observed in these two residues when there was a vacancy in binding pocket in the apo structure. (Figure 4.10 B and C) Met1509 on S6_{IV} was also observed to rotate alternatively to fill the binding pocket in the apoprotein structure, (Figure 4.10 D) but was pushed outwards when NIFE was present. (Figure 4.10 H) In the NIFE bound state, these residues coordinated with the NIFE and stabilized the local environment in the binding pocket, while in the apoprotein structure, these residues are free to move and rotate since more free space is allowed in the binding pocket. Our investigation of hCaV1.2 ion channel dynamics suggested that NIFE interacted with the side chain of residues Val1053,

Phe1129, Phe1176, Met1178, Phe1181 and Met1509 in the pocket via mainly the hydrophobic interactions. In particular, the NIFE showed a strong H-bond interaction with Ser1132, alongside persistent π - π stacking interactions with the Phe1129 and Phe1181. These residues were able to coordinate with the NIFE, yet it was observed that it is highly flexible in the apoprotein structure, which may construct a pocket that binds to compounds sharing similar scaffold as NIFE. This analysis provides useful information for further modifications of the NIFE and discovery of new inhibitors targeting hCaV1.2 ion channel.

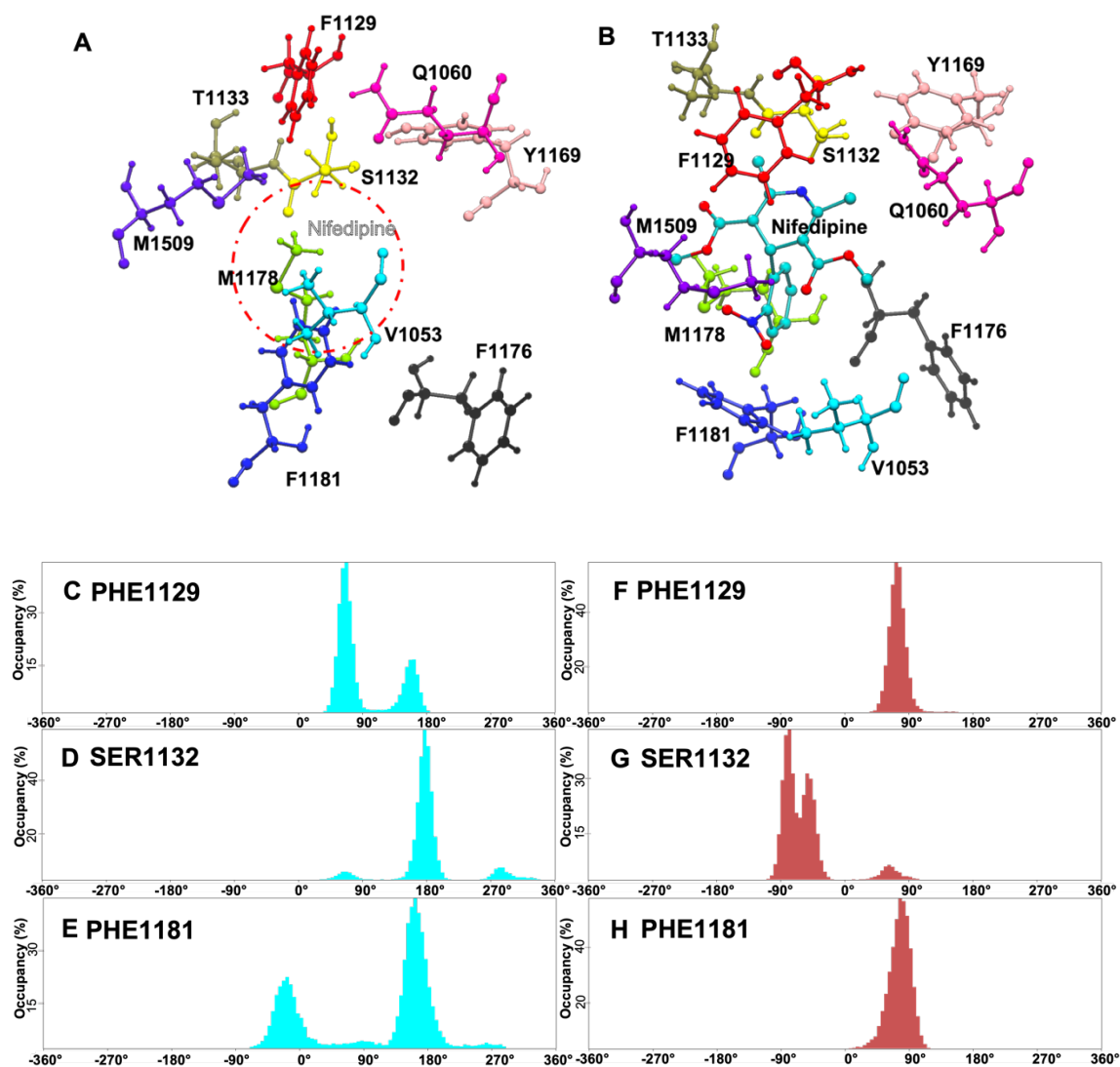


Figure 4.9. Snapshot of residues around the NIFE binding pocket in apoprotein structure (A) and NIFE bound structure (B). The dash line circle in (A) indicates the position of the NIFE bound in the pocket. (C)-(H) Clustering conformations of residues surrounding the ligand binding pocket. Each histogram can represent one rotameric state. (C)/(F) (Phe1129), (D)/(G) (Ser1132), E/H (Phe1181). Left panel is for apoprotein structure and right panel is for NIFE bound structure.

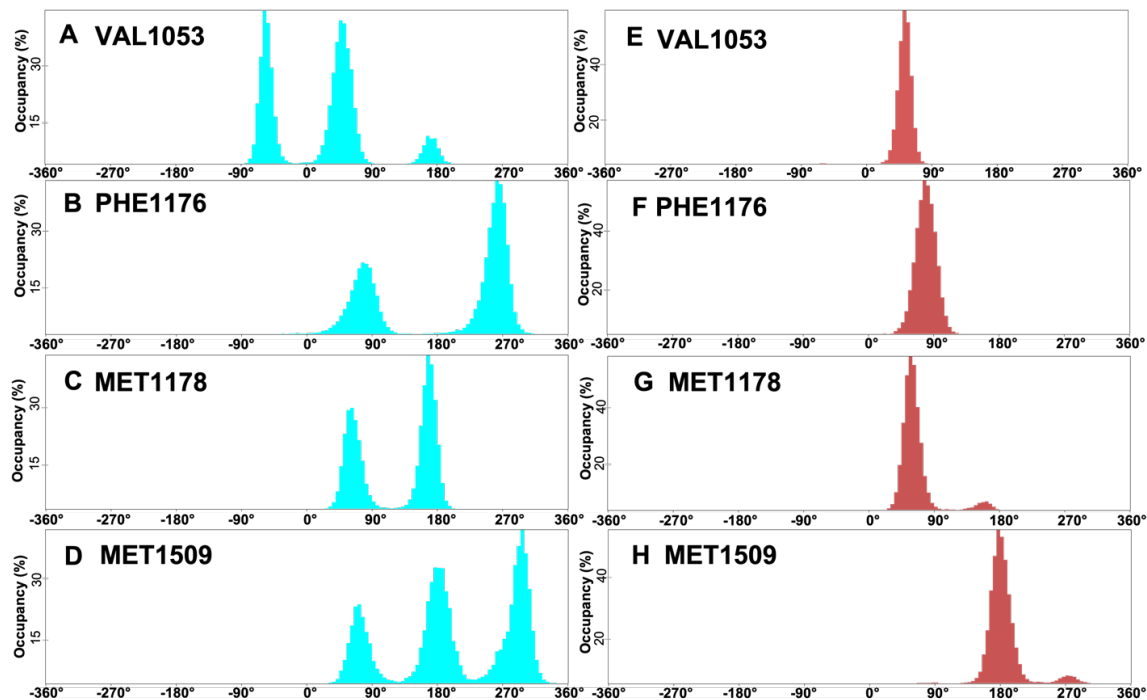


Figure 4.10. Clustering conformations of residues surrounding the ligand binding pocket. Each histogram can represent one rotameric state. (A)/(E) (Val1053), (B)/(F) (Phe1176), C/G (Met1178), D/H (Met1509). Left panel is for apoprotein structure and right panel is for NIFE bound structure.

To explore the activation/inactivation state caused by the effect of the NIFE bound to the structure, we measured the permeation path of the hCaV1.2 ion channel in both the apoprotein and NIFE bound structures. (Figure 4.11) Previous studies have shown that DHP antagonists selectively inhibit L-type calcium channel under persistent depolarized conditions, which linked to sealed intracellular gate.³³ We applied HOLE program³⁴ to calculate the radius of the calcium permeation path along the channel pore. Green dots indicated a widen intracellular gate and selective filter in the apoprotein, whereas red dots indicated a narrow intracellular gate and selective filter. It is suggestive that the channel

pore is less permeable for calcium ions to pass through since the NIFE bound structure exhibit depolarized conformation and is in inactivated state.

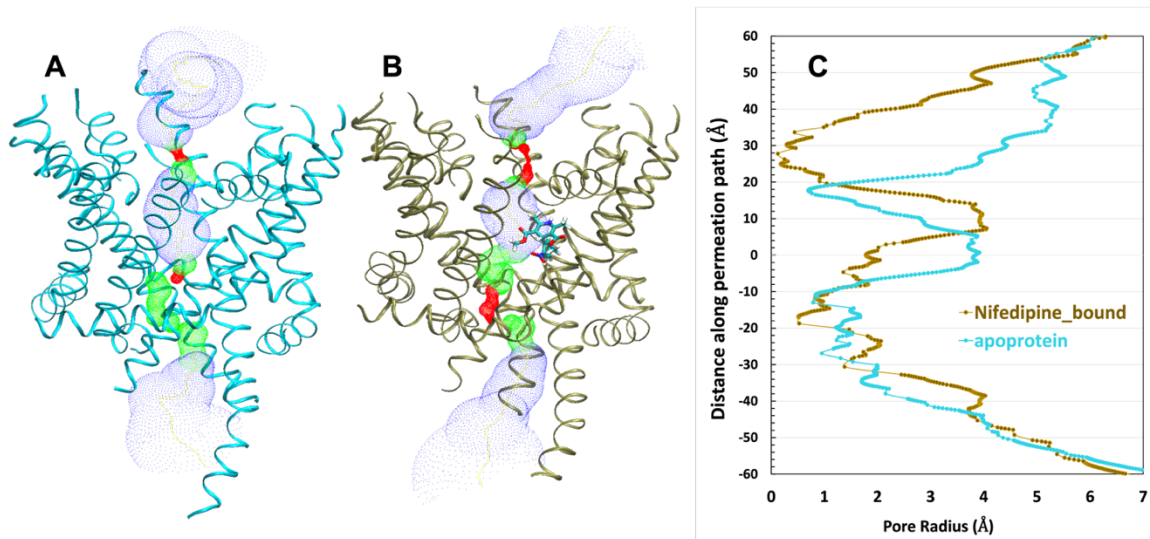


Figure 4.11. The permeation path of hCaV1.2 apoprotein (A) and NIFE bound (B) structures calculated by HOLE³⁴ are illustrated by dots in the panel. The pore radii < 1.15 Å are colored in red, between 1.15 Å and 2.3 Å are colored in green and > 2.3 Å are colored in blue. Measurement is according to¹³ (C) Pore radii along the permeation path.

4.5 Conclusions

Calcium and its relevant intracellular calcium channels are chiefly associated with the regulation of cardiovascular and neuronal systems. Recently, an increasing number of studies have shown that calcium channels are also responsible for tumorigenesis and progression. However, the underlying mechanisms and the involved signaling transduction pathways remain elusive. In this study, we combined bioinformatics, molecular modeling, and *in vitro* experiment to study a potential therapeutical treatment and biomarker CACNA1C gene, which encodes the voltage-dependent calcium ion transmembrane

channel (hCaV1.2). We show that NIFE, a calcium channels blocker (CCB) had a tumor suppressive effect based on binding model predicted by computer assisted molecular modeling and *in vitro* validation using human OC cell line. Further validation using large The Cancer Genome Atlas ovarian public cohort, we found the CACNA1C mRNA expression strongly correlated with poor patient survival outcome. Here, we also show strong correlation of CACNA1C protein expression using immunohistochemistry correlating with COH ovarian carcinomas patients' disease progression. This research demonstrates to the broader community that targeting OC via CCBs may be therapeutically beneficial in OC.

4.6 References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA: a cancer journal for clinicians* 2022;72(1):7-33.
2. Momenimovahed Z, Tiznobaik A, Taheri S, Salehiniya H. Ovarian cancer in the world: epidemiology and risk factors. *International journal of women's health* 2019;287-299.
3. Feeney L, Harley IJ, McCluggage WG, Mullan PB, Beirne JP. Liquid biopsy in ovarian cancer: Catching the silent killer before it strikes. *World Journal of Clinical Oncology* 2020;11(11):868.
4. Suh DH, Kim TH, Kim J-W, Kim SY, Kim HS, Lee TS, Chung HH, Kim Y-B, Park NH, Song YS. Improvements to the FIGO staging for ovarian cancer: reconsideration of lymphatic spread and intraoperative tumor rupture. *Journal of gynecologic oncology* 2013;24(4):352-358.
5. Kurnit KC, Fleming GF, Lengyel E. Updates and new options in advanced epithelial ovarian cancer treatment. *Obstetrics and gynecology* 2021;137(1):108.
6. Wu J, Sun H, Yang L, Deng Y, Yan Y, Wang S, Yang G, Ma H. Improved survival in ovarian cancer, with widening survival gaps of races and socioeconomic status: a period analysis, 1983-2012. *Journal of Cancer* 2018;9(19):3548.
7. Phan NN, Wang CY, Chen CF, Sun Z, Lai MD, Lin YC. Voltage-gated calcium channels: Novel targets for cancer therapy. *Oncology letters* 2017;14(2):2059-2074.
8. Lazary J, Eszlari N, Kriko E, Tozser D, Dome P, Deakin JW, Juhasz G, Bagdy G. Genetic analyses of the endocannabinoid pathway in association with affective phenotypic variants. *Neuroscience Letters* 2021;744:135600.
9. Wang C-Y, Lai M-D, Phan NN, Sun Z, Lin Y-C. Meta-analysis of public microarray datasets reveals voltage-gated calcium gene signatures in clinical cancer patients. *PloS one* 2015;10(7):e0125766.
10. Qiao Z, Jiang Y, Wang L, Wang L, Jiang J, Zhang J. Mutations in KIAA1109, CACNA1C, BSN, AKAP13, CELSR2, and HELZ2 are associated with the prognosis in endometrial cancer. *Frontiers in genetics* 2019;10:909.
11. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* 2018;46(W1):W296-W303.

12. Bienert S, Waterhouse A, de Beer TA, Tauriello G, Studer G, Bordoli L, Schwede T. The SWISS-MODEL Repository—new features and functionality. *Nucleic acids research* 2017;45(D1):D313-D319.
13. Zhao Y, Huang G, Wu J, Wu Q, Gao S, Yan Z, Lei J, Yan N. Molecular basis for ligand modulation of a mammalian voltage-gated Ca²⁺ channel. *Cell* 2019;177(6):1495-1506. e12.
14. Case DA, Aktulga HM, Belfon K, Ben-Shalom I, Brozell SR, Cerutti DS, Cheatham III TE, Cruzeiro VWD, Darden TA, Duke RE. Amber 2021. University of California, San Francisco; 2021.
15. Inc CCG. Molecular operating environment (MOE). Chemical Computing Group Inc. 1010 Sherbooke St. West, Suite# 910, Montreal ...; 2016.
16. Edelsbrunner H. The union of balls and its dual shape. 1993. p 218-231.
17. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics and modelling* 2006;25(2):247-260.
18. Jo S, Kim T, Im W. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PloS one* 2007;2(9):e880.
19. Jo S, Lim JB, Klauda JB, Im W. CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophysical journal* 2009;97(1):50-58.
20. Lee J, Hitzenberger M, Rieger M, Kern NR, Zacharias M, Im W. CHARMM-GUI supports the Amber force fields. *The Journal of chemical physics* 2020;153(3):035103.
21. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of computational chemistry* 2008;29(11):1859-1865.
22. Wu EL, Cheng X, Jo S, Rui H, Song KC, Dávila-Contreras EM, Qi Y, Lee J, Monje-Galvan V, Venable RM. CHARMM-GUI membrane builder toward realistic biological membrane simulations. Wiley Online Library; 2014.
23. Kim S, Lee J, Jo S, Brooks III CL, Lee HS, Im W. CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. Wiley Online Library; 2017.
24. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *The Journal of chemical physics* 1995;103(19):8577-8593.

25. Ryckaert J-P, Ciccotti G, Berendsen HJ. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics* 1977;23(3):327-341.
26. Rangel-Galván M, Rangel A, Romero-Méndez C, Dávila EM, Castro ME, Caballero NA, Meléndez Bustamante FJ, Sanchez-Gaytan BL, Meza U, Perez-Aguilar JM. Inhibitory Mechanism of the Isoflavone Derivative Genistein in the Human CaV3. 3 Channel. *ACS Chemical Neuroscience* 2021;12(4):651-659.
27. Gao S, Yan N. Structural Basis of the Modulation of the Voltage-Gated Calcium Ion Channel Cav1. 1 by Dihydropyridine Compounds. *Angewandte Chemie International Edition* 2021;60(6):3131-3137.
28. Mosa FE, Suryanarayanan C, Feng T, Barakat K. Effects of selective calcium channel blockers on ions' permeation through the human Cav1. 2 ion channel: A computational study. *Journal of Molecular Graphics and Modelling* 2021;102:107776.
29. Xu L, Li D, Tao L, Yang Y, Li Y, Hou T. Binding mechanisms of 1, 4-dihydropyridine derivatives to L-type calcium channel Ca v 1.2: a molecular modeling study. *Molecular BioSystems* 2016;12(2):379-390.
30. Ai R, Fatmi MQ, Chang C-e, A. T-Analyst: a program for efficient analysis of protein conformational changes by torsion angles. *Journal of computer-aided molecular design* 2010;24(10):819-827.
31. Yamaguchi S, Okamura Y, Nagao T, Adachi-Akahane S. Serine residue in the IIIS5-S6 linker of the L-type Ca²⁺ channel α 1C subunit is the critical determinant of the action of dihydropyridine Ca²⁺ channel agonists. *Journal of Biological Chemistry* 2000;275(52):41504-41511.
32. Yamaguchi S, Zhorov BS, Yoshioka K, Nagao T, Ichijo H, Adachi-Akahane S. Key roles of Phe1112 and Ser1115 in the pore-forming IIIS5-S6 linker of L-type Ca²⁺ channel α 1C subunit (CaV 1.2) in binding of dihydropyridines and action of Ca²⁺ channel agonists. *Molecular pharmacology* 2003;64(2):235-248.
33. Sanguinetti MC, Kass RS. Voltage-dependent block of calcium channel current in the calf cardiac Purkinje fiber by dihydropyridine calcium channel antagonists. *Circulation Research* 1984;55(3):336-348.
34. Smart OS, Neduvilil JG, Wang X, Wallace B, Sansom MS. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *Journal of molecular graphics* 1996;14(6):354-360.

4.7 Acknowledgements

Research reported in this publication included work performed in the Pathology Core supported by the National Cancer Institute of the National Institutes of Health under grant number P30CA033572. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes

Chapter 5: GeomBD3: Brownian Dynamics Simulation Software for Biological and Engineered Systems

5.1 Abstract

GeomBD3 is a robust Brownian dynamics simulation package designed to easily handle natural or engineered systems in diverse environments and arrangements. The software package described herein allows users to design, execute, and analyze BD simulations. The simulations use all-atom, rigid molecular models that diffuse according to overdamped Langevin dynamics and interact through electrostatic, Lennard-Jones, and ligand desolvation potentials. The program automatically calculates molecular association rates, surface residence times, and association statistics for any number of user-defined association criteria. Users can also extract molecular association pathways, diffusion coefficients, intermolecular interaction energies, intermolecular contact probability maps, and more using the provided supplementary analysis scripts. We detail the use of the package from start to finish and apply it to a protein-ligand system and a large nucleic acid biosensor. GeomBD3 provides a versatile tool for researchers from various disciplines that can aid in rational design of engineered systems or play an explanatory role as a complement to experiments. GeomBD version 3 is available on our website at <http://chemcha-gpu0.ucr.edu/geombd3/> and KBbox: <https://kbbox.hits.org/toolbox/methods/molecular-simulation/geombd/>.

5.2 Introduction

Biological and chemical processes *in vivo* and *in vitro* typically start out with molecular encounters, in which two or more molecules must find one another. These diffusional

processes can take place on temporal and spatial scales up to milliseconds and micrometers,¹⁻³ rendering them beyond the reaches of methods such as molecular dynamics simulation. However, such processes must be studied to elucidate the kinetics and mechanisms of important reactions. Moreover, environmental, structural and orientational modifications to complex systems can dramatically affect reactions taking place within, so the ability to rapidly assess these variables is extremely valuable. Understanding driving forces and mechanisms for binding assists the design of nano-scale molecular systems. In addition, most experimental techniques lack the ability to resolve subtle differences in device functionality that result from nano-scale modifications. Molecular simulation tools can bridge the gap and guide rational design of novel biotechnologies.^{4,5}

Brownian dynamics (BD) has been used to investigate long-time scale molecular processes for several decades.^{6,7} Many software packages are available for studying a variety of problems related to bimolecular association.⁸ MacroDox, UHBD, BrownDye, Simulation of Diffusional Association (SDA), BD_BOX, BROMOC, ReaDDy, Smoldyn, SEEKR, and BDpack are used to probe ligand–protein associations with a flexible or rigid biomolecular model.⁹⁻¹⁸ Molecular interactions can be affected by their environments, several BD programs also consider some environmental factors, such as molecular crowders. However, not all-important system features can be modeled by existing programs. For example, nanostructures involving multi-enzyme cascades, systems with important surface properties, and systems with ligands under diffusive flux in a particular direction cannot be simulated with existing BD packages.

The BD simulations carried out in our software, GeomBD3 (GeomBD version 3), are specifically designed for this purpose. GeomBD3 simulates the molecular recognition process of two molecules over long distances and time scales in a vast array of different systems and system sizes. Systems may include many separate molecules and may utilize a range of various boundary conditions and initial ligand positions. The package simulates multiple replicas of a ligand in parallel and generates multiple trajectories of the same ligand. The software is easily capable of handling molecular models up to hundreds of nanometers in scale, thus allowing simulation of simple protein-ligand systems or very large nanoscale technologies, such as biosensors or surface-tethered multi-enzyme arrays. For example, the software has been used to study enzyme bioconjugates, understand DNA hybridization in biosensors,¹⁹ and replicate SPR experiments under diffusive flux.²⁰ The package allows users to arrange multi-enzyme in space or on a membrane and begin ligand diffusion from a pocket of an enzyme to bind with another enzyme. In many biosensing and catalyst applications, the device surface properties are of high importance, as they can control substrate adsorption and diffusion and thus overall efficiency.^{21,22} GeomBD3 provides the ability to examine different experimental settings *in silico* to rapidly evaluate multiple design variables.

Here, we present a detailed description of GeomBD3 along with two applications of the software and basic analysis of the results. Users are guided through the general simulation workflow, from set-up to execution to final analysis. We then present applications of the program to a simple and well-studied protein-ligand system consisting of

acetylcholinesterase and acetylcholine, then to a larger, more complex nucleic acid biosensor system.

5.3 Methods

Implementation

GeomBD3 source code is written entirely in C++, whereas several pre- and post-simulation scripts are available in both C++ and Python3 to aid in simulation set-up or analysis. Simulation set-up, execution, and analysis are carried out using standard PDB, PQR, and DCD file types. The source code can be divided into two categories: pre-simulation code (*Parameterize.py* & *Gridder*) for parameterization of the molecular models and calculation of potential energy grids, and simulation code (GBD) for running, recording, and monitoring the Brownian dynamics. All this code can be found on our group GitHub at <http://github.com/chang-group>.

Simulation Algorithm

The basic workflow of the software package (Figure 5.1) is as follows. Initially, users should obtain PDB format molecular models of the system. GeomBD3 uses all-atom, rigid molecular models with no internal degrees of freedom. Generally, systems consist of a ligand and one or more other molecules that may or may not be receptors of the ligand. The ligand is the molecule which diffuses during the simulation, and the receptor(s) is(are) its intended binding partner(s). The ligand may range in size from a protein to a small drug molecule, but it must be chosen as the smaller of the binding partners since simulation time scaling linearly with ligand size. The package also allows to include multiple conformations of a ligand. Next, the PDBs are converted into PQR type files by using the

Parameterize.py script. These PQRs are then fed into the *Gridder* programs that pre-compute as much of the potential energy calculations as possible to save time. Users are not required to generate all three available potential grids and should choose which are needed for the study.

The remaining steps involve calculating the Brownian dynamics trajectories and monitoring output. An input configuration file should be supplied containing the system PQR files, desired grid files, and simulation parameters. Upon running the *GBD* code, first, several independent ligand replicates are positioned according to the desired starting conditions. If ligand replica lies on the potential energy grids, the intermolecular forces are calculated by the finite difference method. As an example, forces in the x dimension F_x are calculated as

$$F_x = \frac{U(gx+1,gy,gz)-U(gx-1,gy,gz)}{2d} \quad \text{Eq. 5.1}$$

where the potential U is a function of an atom's location on the grid specified by gx , gy , and gz , and d is the grid spacing. During simulation, the potential at the coordinates of each ligand atom is approximated by trilinear interpolation between grid points. The linear interpolation of the force may lead to some inaccuracy for highly charged systems. However, the accuracy gained by using the nonlinear interpolation is not very significant, compared with the increased computational time, especially for large systems. A comparison between linear and cubic spline interpolation shows 40% increase in the CPU time.²³ A random force R with the properties

$$\langle R(t) \rangle = 0 \quad \text{Eq. 5.2}$$

$$\langle R(t)R(t') \rangle = \delta(t - t') \quad \text{Eq. 5.3}$$

generated from a Gaussian process with a zero mean is also applied at this time, and ligands are moved to their next positions according to the overdamped Langevin equation

$$r_{t+\Delta t} = r_t - \frac{D}{k_B T} F \Delta t + \sqrt{2D\Delta t} R \quad \text{Eq. 5.4}$$

where r is the ligand position, D is the appropriate diffusion coefficient, $k_B T$ is Boltzmann's constant times the absolute temperature, F is the 3-dimensional force calculated from its x-, y- and z-components such as Eq.1, and Δt is a linearly scaled timestep. The translational and rotational diffusion coefficients are calculated based on the hydrodynamic radius r_h of either the ligand only or the ligand and receptor. The former case is only used when the receptor is attached to a surface i.e., not diffusing.

$$D_{trans} = \frac{k_B T}{6\pi\gamma} \left(\frac{1}{r_{h,lig}} + \frac{1}{r_{h,rec}} \right) \quad \text{Eq. 5.5}$$

$$D_{rot} = \frac{k_B T}{8\pi\gamma} \left(\frac{1}{r_{h,lig}^3} + \frac{1}{r_{h,rec}^3} \right) \quad \text{Eq. 5.6}$$

Hydrodynamic interactions (HI) are not included. It has been observed by Northrup et al that the effect of HI on k_{ON} in similar modeling condition is $\sim 6\%$ which is not significant.²⁴ At each new position, each binding criterion is checked and, if satisfied, a binding event is counted and stored. The bound replica of the ligand is then reset according to the starting conditions to begin a new trajectory. In this way, the numbering of ligand trajectories being simulated is constant. These steps are repeated until certain simulation parameters converge below a defined threshold, or the user manually terminates the program.

Relevant data pertaining to association rates, binding probabilities, site/surface-specific binding dwell times, ligand-receptor interactions, and simulation convergence are

automatically output to a log file. Simulation trajectories are output as a DCD-format file for viewing and analysis.

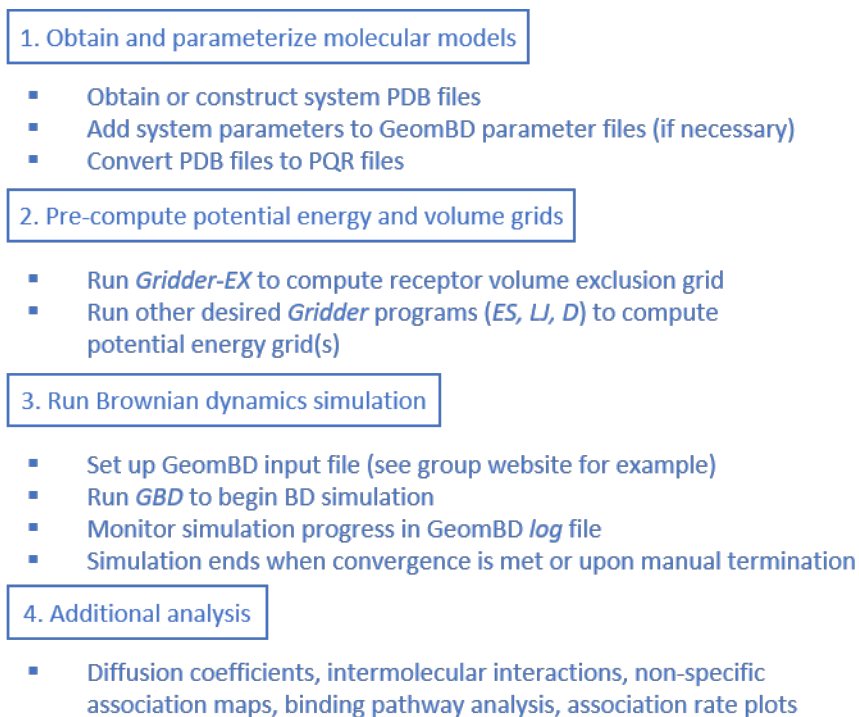


Figure 5.1: Diagram of GeomBD3 workflow.

Simulation Set-Up

The initial step in the simulation workflow is to obtain PQR format input files containing the molecular structures of the ligand and receptor. These PQR files can be generated by the script *Parameterize.py*, which takes in PDB files and converts them to PQR format.

Any file in standard PDB format is acceptable, such as those downloadable from the Protein Data Bank.²⁵ GeomBD3 also handles non-standard systems, if PDB molecular models can be created by the user and the parameters for new molecules are added to the parameter files. The parameter files are simple to read and edit, making addition of new parameters very straightforward. Instructions on preparing a simulation of a novel system

are provided in the user manual at <http://chemcha-gpu0.ucr.edu/geombd3/>. The PQR files contain the coordinates, partial charges, and vdW radii of each atom, which are later used in energy calculations. The values of partial charges for amino acids and standard nucleotides were taken from the AMBER ff14sb²⁶ while for other molecules using any software that performs quantum or semi-empirical charge calculation. For instance, partial charges for Ach (Case4) were calculated using AM1-BCC method. The vdW radii were taken from the GAFF²⁷ force fields. For example, the input pdb files of SAM and gold surfaces were modeled by AMBER GAFF force field using xleap package and then converted to pqr files.

After obtaining parameterized PQR files, users should run the series of *Gridder* programs to pre-compute the potential energy at all points in a grid-based representation of the receptor molecule. The following *Gridder* programs are currently available:

Gridder-EX.cc - This creates a grid (the EX grid) defining the excluded volume of the simulation space due to the presence of the receptor. It defines the space occupied by the receptor and prevents unphysical interpenetration of atoms. The calculation maps each atom of the receptor to a grid point, then searches a distance equal to the atom's LJ radii in x, y, and z dimensions. Every grid falling within this search range is occupied by an atom. The EX grid should always be generated before running a simulation.

Gridder-ES.cc - This creates a grid (the ES grid) holding the electric field strength of the receptor molecule at each point calculated by the screened Coulomb method.²⁸ We also considered Adaptive Poisson-Boltzmann Solver (APBS) method for highly charged molecules. However, the tests were not feasible due to very high computation cost and

memory. During simulation, the partial charge, q_n , on each ligand atom can be multiplied by the appropriate grid value, $\varphi_{SC,grid}(g)$, to give the electrostatic potential between that atom and the receptor.

$$\varphi_{SC,grid}(g) = \sum_m \frac{k_e q_m e^{-k r_{gm}}}{r_{gm} \epsilon} \quad \text{Eq. 5.7}$$

$$E_{SC,grid}(g, q_n) = q \varphi_{sc,grid}(g) \quad \text{Eq. 5.8}$$

$$\kappa = \sqrt{\frac{\sum_i (c_i q_i^2)}{\epsilon \epsilon_0 k_B T}} \quad \text{Eq. 5.9}$$

where k_e is Coulomb's constant, g is a grid coordinate, q_m and q_n are the charges on the m^{th} receptor atom and n^{th} ligand atom, respectively, r_{gm} is the distance between a grid point g and m^{th} receptor atom, and ϵ is the dielectric of water and ϵ_0 is vacuum permittivity. κ is the inverse Debye length, where c_i and q_i are the molar concentration and formal charge of ionic species i .

Gridder-LJ.cc - This creates grids (the LJ grids) storing the 12-6 Lennard-Jones potential E_{LJ} between the receptor and each atom type in the ligand. The default padding and spacing values are 12 Å and 0.5 Å, respectively.

$$E_{LJ} = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad \text{Eq. 5.10}$$

where ϵ is the potential well depth, σ is the interatomic distance corresponding to zero potential, and r is the interatomic distance.

Gridder-D.cc - This creates a grid holding the desolvation potential of the ligand. This potential is based on the degree of desolvation of each atom and the atom's desolvation parameter, S . The function used in GeomBD3 borrows the functional form and parameters

used in AutoDock4.2,²⁹ which is similar to the concept developed by Stouten, *et al.*³⁰ Here, Eq. 5.10 is the energy calculation required to hydrate an atom using volume-summing method and Eq. 5.11 calculates atomic solvation parameter for each atom type required.

$$E_{desolv} = \sum_{j=1}^N S_i V_j e^{\frac{-r_{ij}^2}{2\sigma^2}} \quad \text{Eq. 5.11}$$

$$S_i = solpar + 0.01097q_i \quad \text{Eq. 5.12}$$

where S_i is the desolvation parameter of ligand atom i , V_j is the fragmental volume of receptor atom j , r_{ij} is the distance between them, σ is 3.5 Å and N is the number of atoms in the receptor molecule. In S_i , *solpar* is an empirically derived part of the desolvation parameter specific to each atom type and q_i is the partial charge on the atom.

Each grid has a user-definable *padding* and *spacing* value. The *padding* variable is the distance the grid will extend in each dimension beyond the minimum/maximum atomic coordinates which can be configured by the user. The default *padding* values are 40, 12, and 12 Å for the ES, LJ, and D grids, respectively. The *spacing* variable is the distance between adjacent grid points and has a default 0.5 Å for all grids in the discussed cases. After observing highly similar results if the grid spacing is ≤ 0.5 Å, we suggest 0.375 Å and 0.5 Å may perform well in most systems. The users can use other grid sizes as well. Additional grids of each crowder molecule can be included to represent crowding effects.

Configuring and Running the Simulation

Once all the desired pre-computed grids have been made, the simulation configuration file should be set. The file consists of keywords followed by either a number or string. A user manual with detailed description of each keyword and guidance on selecting appropriate

values is provided on our website at <http://chemcha-gpu0.ucr.edu/geombd3/>. Here we will discuss only a brief overview of the file.

The generated PQR files should be specified after the *ligand* and *receptor* keywords. This is followed by the specification of the grid files to be loaded. The *grid* keyword followed by a specific string denoting the grid type e.g., *es* for electrostatic, should be given first. Then the path to the appropriate grid file should follow.

The next set of keywords specify the simulation parameters. These include *ligandStart*, *boundary*, *boundWall*, and *fixedReceptor*. For example, with *ligandStart* and *boundary*, users can select periodic or reflective boundary conditions, and the bounding cell may be either cuboid, hexagonal prismatic, or spherical. In addition, ligands may be started from an xy-aligned plane, a random point, or a specified point. Alternatively, simulations can be run under NAM (after Northrup, Allison, and McCammon) conditions,²⁴ in which ligands start from a spherical plane and exit at a larger spherical plane. The value of the inner spherical plane must be large enough to satisfy these conditions: the interaction between ligand and receptor can be considered centrosymmetric; and the ligand has no angle dependence with isotropic reactivity. These options are made available so that simulations can be made to match experimental and natural conditions, or so that more appropriate comparisons can be made between simulated and theoretical results. Association rate constants in most simulations are calculated as

$$k_{on} = \frac{1}{ct_{avg}} \quad \text{Eq. 5.13}$$

where t_{avg} is the average association time and c is the ligand concentration estimated by the simulation volume. The concentration can be varied by using a different simulation volume.⁷ Under NAM conditions, the association rate constant is calculated as

$$k_{on} = \frac{4\pi r D \beta}{1 - (1 - \beta)^{\frac{k_d(b)}{k_d(q)}}} \quad \text{Eq. 5.14}$$

where r is the ligand-receptor distance required for binding, D is the relative diffusion coefficient of ligand and receptor, β is the fraction of ligands that bound before exiting, and b and q are the radii of the starting and exiting spherical planes, respectively.²⁴ Collision probability, β , is determined on the basis of simulations which yield similar results and the starting position of the ligand must be angle independent. The *biasForce* keyword allows users to apply a biasing force along one dimension to mimic flow of the ligands inside the system. For example, such conditions were used to simulate surface plasmon resonance experimental conditions to investigate the effect of flow on ligand association to HIV protease.²⁰ Due to the external flow effect in the system, calculation of rate constant using Eq. 5.13 or Eq. 5.14 is not possible. The ligand diffusive flux resulting from this force can be calculated as

$$\mathcal{T} = \frac{c F_{bias}}{f} \quad \text{Eq. 5.15}$$

where c is the ligand concentration, F_{bias} is the biasing force, and f is the friction force on the diffusing ligand.

The *bindgroups* keyword defines the criteria that must be met to achieve binding between ligand and receptor. This should be carefully set based on visualization of the molecular models and knowledge of the system/process being simulated. Each binding criterion

consists of an XYZ coordinate, an index number of a ligand atom, and a distance. Taken altogether, this represents the distance between the ligand atom and specified coordinate below which the two molecules are considered bound. For example, the input file line

bindgroups 5.0 10.0 15.0 6 12.0

specifies a condition in which ligand atom number 6 must be less than 12.0 Å from the cartesian point 5.0 10.0 15.0 to record a binding event. The user can also include multiple binding sites with different combinations using “AND” and “OR” terms in *bindgroups*, which may further affect the binding pathways and rate. Notably, the *bindgroups* and distance between a ligand and *bidgroup* are system dependent. If an experimental k_{ON} is available, the value may be used to benchmark the *bindgroup* and distance as the values determine whether a binding is successful.

Simulations can be set to automatically terminate under specific conditions, or they can be manually terminated by the user at any time. Most often, simulations are set to terminate when some value of interest has converged below a certain threshold. For example, if mean binding times t_{avg} are desired, GeomBD3 can be set to terminate when the SEM (Standard Error of Mean) of binding times is below a threshold set by the *convergence* input keyword. In such a case, setting *convergence* to 0.01 would cause the program to stop when the $\text{SEM}/t_{\text{avg}}$ was < 0.01 . The calculation of t_{avg} gives insight to diffusion-limited rate constants, trends in reaction rates, and diffusion rates in chemically diverse environments. Because the program simulates a constant number of ligand trajectories, when many replicates are being run e.g., 5000, the initial data in the log file is necessarily biased toward fast-binding

trajectories. Generally, one needs to sample several times *Nreplicates* binding events before results are reasonably converged.

Simulation Output

Five files are automatically output by GeomBD3. The log file (*.log* extension) displays updated information about the number of association events, the average association time, the SD (Standard Deviation) and SEM of association times, convergence of the mean association time, ligand exit events, dwell time, and ligand-receptor interaction energies. The ligand's time experiencing non-zero LJ potential is considered as dwell time. Since the ligand is diffusing under strong short-range forces from the receptor, it can be considered dwelling on the surface. All information about binding events is displayed separately for each specified binding criterion.

The output PQR and DCD files are needed for viewing the trajectory and can be loaded together into any program which supports these file formats e.g., VMD,³¹ Pymol,³² *etc.* The PQR contains the initial system configuration, that is, the starting positions of all ligand replicates; it does not contain the receptor coordinates. The DCD file containing the simulation frames should be loaded into this PQR file. Users will see trajectories of all ligand replicates displayed at once. Individual trajectories can be viewed by selecting the appropriate representation in the chosen software. The receptor PQR or PDB used as simulation input should be loaded separately to view ligand-receptor interactions. Trajectory visualization tips and analysis are given in the user manual at <http://chemcha-gpu0.ucr.edu/geombd3/>.

Simulation algorithms are parallelized with the Intel compiler's Cilk+ extension. GeomBD3 performance depends on the system, substrate, and number of substrate replicates. An example of the HIVp-xk263 system is included to show the efficiency of the program. The system is a box having HIVp (3128 atoms) placed at the bottom and 100 replicas of xk263 (84 atoms each replica) starting as a plane from the top. Simulations were run on 24 CPU cores of Intel(R) Xeon(R) CPU E5-2620 v2 with a step rate of 2308.3 steps/sec. Considering each step is of 50 fs, a 7 μ s simulation of this system takes around 17hrs.

Restarting GeomBD3 Simulations

There are two additional output files with the extensions *.t* and *.crd*. The former contains the restart times and binding information, while the later contains the coordinates for each ligand replicate. These files can be used to restart a simulation by restoring the coordinates and trajectory lifetimes of all ligands replicates as well as the statistics for all binding criteria. This provides the option of restarting a simulation at any point.

5.4 Example Applications

Case 1: Ligand Association to a Surface-Bound Protein Under Diffusive Flux

In another application, GeomBD3 was used to study the effect of ligand flow and ambient surface properties. The simulation consisted of surface-bound HIV protease (HIVp) as the receptor and xk263 as the ligand. We analyzed association time and probability while varying the flow rate of xk263, the surface properties, and HIVp's orientation with respect to the surface. We also examined the effect of HIVp's accessible surface for xk263 by allowing it to bind non-specifically to the protein interface.

The simulation cell was a rectangular prism containing HIVp bound to a self-assembled monolayer (SAM) surface (open conformation) and xk263 (PDB Id: 1HVR). The HIVp used in the model is an open conformer derived from MD simulation of a closed conformer (PDB Id: 1HHP). The SAM surface with CH₃ terminal groups provides a hydrophobic surface environment. The ligand replicas started to diffuse from an yz-aligned plane at the -x-direction boundary with a drift in the +x-direction (Figure 5.2). The flux effect was introduced by boosting the force value with the biasForce in the x component of the overdamped Langevin equation only. Boundaries in the y-direction were periodic, and a termination boundary was used +x-direction. We considered two orientations for HIVp and seven flux rates for xk263 leading to 56 different system configurations. We analyzed diffusion coefficients, xk263 interactions with SAM surface and the distribution of xk263 with respect to the flow.

We found that the weak vdW attractions of surface with xk263 increases the binding time in the absence of diffusion flux. With the increase in flux, the association time decreases even in the presence of SAM surface. The simulated average association time of xk263 with HIVp in the presence of SAM but no flow is 1.491 μ s. The remaining association time data for each system according to flux can be found in the original research publication²⁰. With the increase in ligand flux, the concentration in middle region of the system increases and leads to less ligand-surface interaction. We have also found that after placing the HIVp 20 Å above the surface made the bottom of HIVp accessible for the xk263 interaction. This exposed an attractive non-specific association site, increasing association time by 10%-20% in the static environment.

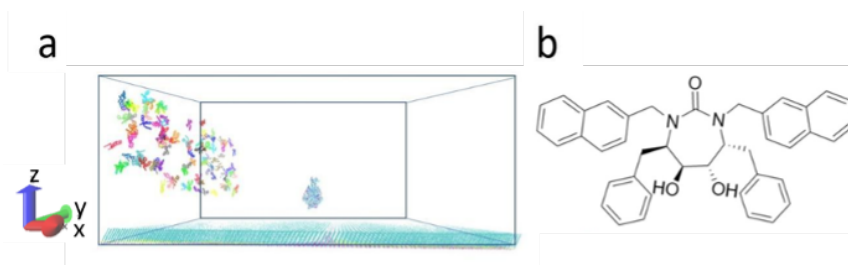


Figure 5.2: (a) HIVp is placed 20 Å above the SAM parallel to the yz-plane and (b) Structure of xk263 ligand.

Case 2: DNA Hybridization in a Biosensor

GeomBD3 is also useful for simulations of large or complex engineered systems that span hundreds of nanometers. As a second GeomBD3 application, we present previously published data¹⁹ from a simulation of DNA hybridization (binding) taking place in a biosensor (Figure 5.3). The sensor consisted of ssDNA probes inserted into a self-assembled monolayer (SAM) of functionalized alkanethiol molecules. With GeomBD3, it was possible to rapidly study the sensor under different probe surface densities and equipped with different functional groups on the self-assembled monolayer (SAM). These simulations used one or more ssDNA probes inserted into a slab of SAM as the receptor and a ssDNA molecule complementary to the probe strand as the ligand. The simulation space was bounded by a cuboid box with the biosensor placed at its bottom. Periodic boundary conditions were used in the x and y dimensions and a hard wall at the top. The x and y dimensions of the box were varied to encompass different amounts of SAM, yielding three different probe surface densities. Additionally, hybridization with clusters of five tightly packed probes was also studied on the $-\text{CH}_3$ and $-\text{COO}^-$ SAMs. The probe density was varied while also varying SAM functional groups between $-\text{OH}$, $-\text{COO}^-$, and $-\text{CH}_3$, yielding 15 different system permutations in total.

Varying the surface properties by changing SAM functional groups can dramatically alter DNA-surface interactions and hybridization rates. In some cases, DNA adsorbs and diffuses along the SAM 2-dimensionally, which may slow down or speed up hybridization depending on the strength of the interaction. This surface mediated hybridization is defined as 2-d whereas if the diffusion is directly to the probe through the solvent, then it is 3-d hybridization. Using the included script *Pathfinder.cc*, we analyzed the DCD trajectories to find the fraction of hybridization taking place 2-dimensionally, through a surface-mediated mechanism, which we label as γ . Our results indicated that when surface interactions are strong, 2-dimensional hybridization was significantly slower than the 3-dimensional hybridization.

In the biosensor modeled with a cluster of probes (Figure 5.3), each probe (labeled 1-5; 5 at center) represented a possible binding site. GeomBD3 accepts multiple binding criteria and can independently track binding data for each criterion. We quantified the hybridization time and fraction of hybridization taking place at each probe to understand how tight probe packing may influence accessibility. The results are presented in Table 5.1, and they indicated that the center probes are nearly inaccessible when the spacing between neighbors is less than or equal to the target DNA length. The reversal of hybridization efficiency at high densities experimentally observed agrees with our simulation results.³³

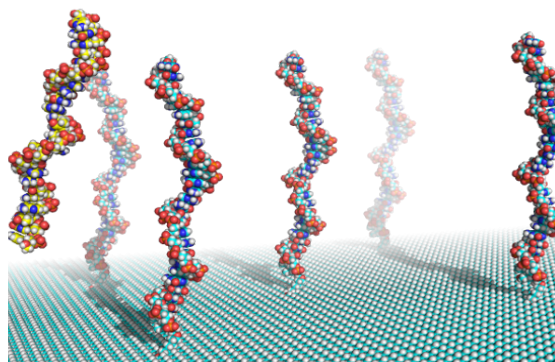


Figure 5.3. GeomBD3 molecular model of a DNA biosensor with multiple probes (blue strands) and the target DNA (yellow strand) which binds to the probes.

Additional Details for Case 2: DNA Hybridization in a Biosensor

Simulations used 0.5M monovalent salt. An additional simulation was carried out at 0.35M monovalent salt for the original research article, but that case is not discussed in this application note. Biosensor models were represented by one large grid file for each potential type (LJ, ES), allowing neighboring probes to feel one another. Gold atom LJ parameters were: $R_{ij} = 3.072$ and $\epsilon = 0.1544$. No partial charges were assigned to the gold. Periodic boundaries were used in X and Y dimensions, with an impassable boundary in the Z dimension. Three hybridization (binding) conditions were defined: 1) inter-strand C.O.M distance $< 7\text{\AA}$, 2) five or more base pair overlap of probe and target strands, and 3) any Watson-Crick base pair partners $< 7\text{\AA}$ apart. The energies given in Table 5.1 are averages over all DNA-surface interaction orientations. We describe the different inter-probe spacings rather than the densities. The densities are the same for all three cases since each model sensor contains five probes on the same amount of SAM. Only the inter-probe spacing changed, since the goal was to understand when hybridization becomes hindered as probe pack more tightly.

Sample parameterized (.pqr) file of acetate ion

```

ATOM 1 C ACT 1 0.5840 0.5250 0.0690 -0.2001 1.9080
ATOM 2 H ACT 1 0.7820 1.3180 0.8250 0.0070 1.3680
ATOM 3 H ACT 1 1.2380 0.7020 -0.8170 0.0070 1.3680
ATOM 4 H ACT 1 0.8310 -0.4710 0.5040 0.0070 1.3680
ATOM 5 C ACT 1 -0.8780 0.5650 -0.3740 0.9016 1.9080
ATOM 6 O ACT 1 -1.4310 -0.5200 -0.7210 -0.8613 1.7000
ATOM 7 O ACT 1 -1.4720 1.6840 -0.3820 -0.8613 1.7000
END

```

Table 5.1: Hybridization time, hybridizing probe site, fraction of 2D hybridization (γ), and DNA-SAM interaction energies on the two different SAMs used for probe cluster simulations.

	inter-probe separation (nm)	τ_{sim} (μs)	hybridizing probe (%)					γ	ΔE_{elec} (kcal/mol)	ΔE_{vdW} (kcal/mol)
			1	2	3	4	5			
CH₃	5.0	22.6 $\pm$ 0.7	23	23	26	26	2	0.13	-0.35 $\pm$ 0.26	-3.71 $\pm$
	10.0	14.4 $\pm$ 0.1	22	20	27	23	7	0.20		1.60
	15.0	15.7 $\pm$ 0.4	21	13	24	27	14	0.24		
COO⁻	5.0	20.6 $\pm$ 0.4	15	30	29	24	2	0.03	0.21 $\pm$ 0.33	0.00 $\pm$
	10.0	15.4 $\pm$ 0.2	23	19	28	26	4	0.03		0.00
	15.0	16.1 $\pm$ 0.5	15	19	24	22	20	0.04		

Case 3: Multienzyme Cascade With and Without Crowders

Biocatalytic cascades can guide complex, efficient and selective intracellular transformations. However, the intermediate substrate transfer efficiency and diffusion rate can be affected by the environment within the cell, as nearly 20% to 40% of the cytoplasmic volume is occupied by various small molecules or proteins.³⁴ These small molecules and proteins crowd the environment and influence reactions taking place around them. Popielec *et al.* discussed the effect of crowders on catalytic activity on various factors.³⁵ Here, we introduced crowders simply to include volume exclusion and non-specific interactions for substrate. As a third GBD3 simulation case, we aimed to investigate the overall catalytic efficiency in a multienzyme cascade complex with and without crowders in a cell

environment. We report the average association time ($t_{\text{association}}$) for substrate to target enzyme, the association probability (p), and the timer a substrate associated with the proteins (t_{dwell}).

We selected phosphoribosyl anthranilate isomerase (PRAI, named E1) and indoleglycerol phosphate synthase (IGPS, named E2) from *E.coli* as our multienzyme cascade model system, and ribonuclease as our crowder since it is one of the most abundant macromolecules in bacteria.³⁶ E1 and E2 are involved in the third step of tryptophan biosynthesis pathway,³⁷ in which the product of E1, 1,2-Carboxyphenylamino)-1-deoxy-d-ribulose 5-phosphate (CdRP) serves as the substrate of E2 to make IGP³⁸. The distance between E1 and E2 is manually placed 50 Å apart based on their centers of mass. The orientation of the initial substrate diffusing position in E1 (along x-axis) and the binding pocket in E2 (along z-axis) is 90° relative to each other. We selected this orientation setup based on its fused crystal structure (PDB: 1PII). We considered two simulation systems in which ribonuclease occupied either 20% (19 crowders randomly placed, Figure 5.4 A) or 0% (0 crowder, diluted environment, Figure 5.4 B) of the total volume of the 341 Å-radius spheres. As in case 1, the spherical boundary is set to be impenetrable, so that the simulation substrate concentration is fixed.

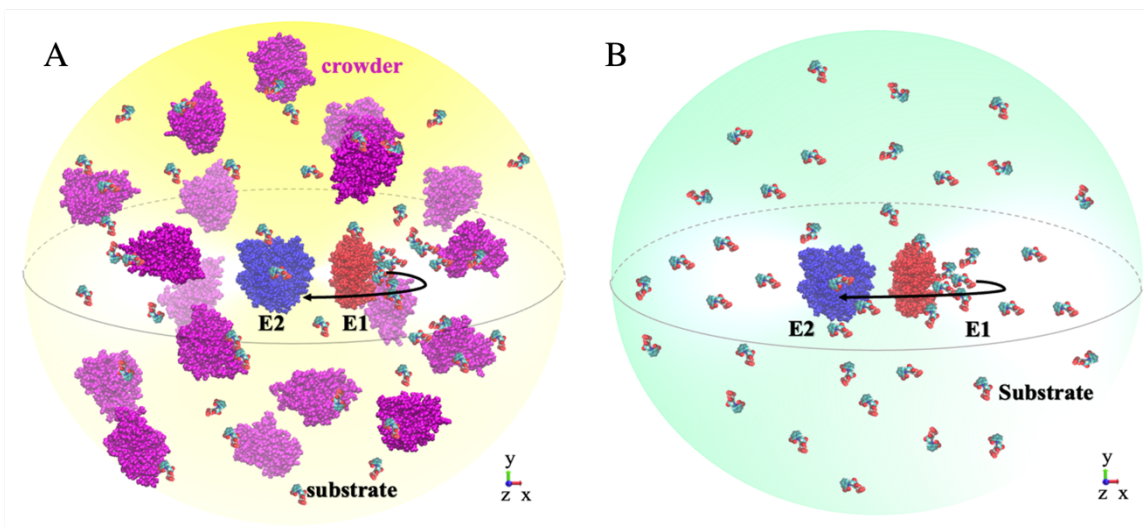


Figure 5.4. (A) Model with crowders (magenta). (B) Model with diluted environment. The E1 (PRAI, red) and E2 (IGPS, blue) are biocatalytic cascade enzymes in the tryptophan biosynthesis pathway. Substrate (CdRP) was produced in the PRAI and catalyzed in the IGPS to make IGP. The crowders were placed randomly in the sphere, in which the radius of the sphere was set to 341 Å corresponding to 10.0 μ M concentration.

Our simulated $t_{\text{association}}$ values were $6.38 \pm 0.20 \mu\text{s}$ and $3.83 \pm 0.13 \mu\text{s}$ in the crowded and diluted environments, respectively. Having 20% of the volume occupied by the crowders, the substrate diffusion rate (conceptually similar to inverse association time) is ~ 1.7 fold slower than that in the diluted environment. This indicates that the substrate interacts with the crowders non-specifically when it diffuses towards the target enzyme. We also evaluated the time the substrate spent on the surface of the enzymes and crowders before association, t_{dwell} , which is shown in the last column of the Table 5.2. The time spent on the surface of the crowders is directly proportional to the time needed to reach the target enzyme. Additionally, we measured the direct binding percentage. We traced the substrate diffusion by assigning a spherical region A with an 80 Å cutoff radius centered on the target enzyme binding site. If a substrate binds without hitting the A sphere, the event was

considered a direct binding. If a substrate exited the A sphere and came back to bind, the event was considered an indirect binding. The low substrate direct binding probability was observed in the crowded environment due to the intermolecular forces between substrate and crowders (Table 5.2). Our results indicated that when the cell environment was occupied by crowders, the substrate diffusion rate was slower than the diluted environment.

Table 5.2: Average association time $\langle t_{\text{assoc}} \rangle$, probability of direct binding (p_{direct}), and surface dwell time $\langle t_{\text{dwell}} \rangle$ of CdRP in the crowding and diluted models.

BD index	$\langle t_{\text{assoc}} \rangle \pm \text{SEM} (\mu\text{s})$	$p_{\text{direct}} (\%)$	$\langle t_{\text{dwell}} \rangle \pm \text{SEM} (\mu\text{s})$
A1 (20% crowder)	6.38 $\pm$ 0.20	2.0	2.964 $\pm$ 0.092
A2 (0% crowder)	3.83 $\pm$ 0.13	3.7	0.079 $\pm$ 0.005

Case 4: Protein-Ligand Binding

One of the most common and straightforward uses of GeomBD3 is to calculate the rate of protein-ligand binding. Here we demonstrate such a simulation applied to the acetylcholinesterase-acetylcholine (AChE-ach) system. The ligand and receptor molecular models were derived from an X-ray crystal structure of AChE from Torpedo Californica (PDB ID: 2ACE).³⁹ The total charge on ach and AChE were 1 e and 8 e , respectively. Acetylcholine replicates were started from random points in a $6.54 \times 10^7 \text{ \AA}^3$ spherical simulation space with the stationary grid representation of AChE at the center. When ligands passed the boundary of the sphere, they were returned to their previous position and a new diffusion step was generated. The sphere essentially acted as an impenetrable wall containing a ligand concentration of 25.4 μM . The binding site is chosen as point (1.3, 56.7, 74.2), which corresponds to an atom in the gorge leading to the catalytic binding site.⁴⁰ Here, we are only interested in calculating the association (binding) rate. The

simulation was set to terminate when the SEM of the binding time has converged below 0.01 of the mean binding time. The timestep is scaled between 30 and 500 Å with a fine setting of 0.025 ps and coarse setting of 0.5 ps. Since AchE is known to be driven by electrostatics,⁴¹ we ran simulations under three different conditions: full electrostatics (0.0 M salt), 0.5 M implicit monovalent salt, and no electrostatics. All simulations contained LJ and ligand desolvation potential grids.

Our calculated k_{on} values using equation 12 were 94.0, 6.58, or $2.95 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$ with 0.0 M implicit ions, 0.5 M implicit ions, or with no ES potential, respectively. This shows the dramatic decrease in association rate as electrostatics are screened by ions or completely turned off. Raymond et al had also observed the same trend.⁴² These rates will suffer some disagreement from experimental rates due to the user's choice of binding condition(s) and the rigid molecular representation of a flexible system.⁴³ We also analyzed the non-specific association of ach to AchE with the program *ProbDX*. This program accepts a PDB-format trajectory and maps ligand coordinates to the nearest grid point, then outputs a DX-format density map which can be visualized (Figure 5.5). This revealed that the ligand spends much of its time associated to the protein surface and identified multiple association hot spots. All are located on one side of the enzyme owing to its strong dipole. Additionally, the GeomBD3 log file contains the surface dwell time i.e., the time the ligand spent on the receptor surface before associating.

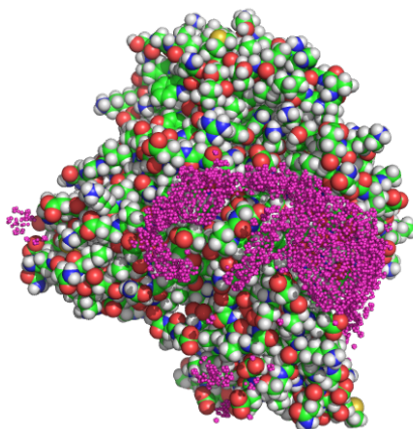


Figure 5.5: Acetylcholinesterase with common ligand association sites found by *ProbDX* indicated by pink dots. Higher dot density indicates higher probability of association.

5.5 Conclusions

GeomBD3 provides researchers with a robust Brownian dynamics software package for modeling molecular association under a wide range of macromolecular arrangements and environments. The program uses rigid all-atom molecular models of a diffusing ligand and stationary grid-based representation of one or more receptors. Molecular recognition is simulated by diffusing a molecule according to overdamped Langevin dynamics and evaluating user-defined binding criteria to track binding events. The program is capable of quickly simulating binding kinetics for drug-protein systems as well as identifying binding pathways and non-specific adsorption sites as demonstrated by our application to the acetylcholinesterase enzyme. GeomBD3 can also elucidate the effects of specific system properties on molecular recognition to reveal phenomena unseen in experiments or to help interpret results. We demonstrate this in our application of the program to a nucleic acid biosensor with varied probe surface density and surface properties. This also demonstrates the applicability of the program to extremely large, non-standard engineered systems for which the program's parameter files can be easily customized. An online interface for

molecule parameterization and grid construction is currently being planned. Thus, the GeomBD3 package can efficiently meet the needs of researchers in a wide range of disciplines as both an explanatory and predictive tool.

5.6 References

1. Berg OG, von Hippel PH. Diffusion-controlled macromolecular interactions. *Annual review of biophysics and biophysical chemistry* 1985;14(1):131-158.
2. Mereghetti P, Kokh D, McCammon JA, Wade RC. Diffusion and association processes in biological systems: theory, computation and experiment. Springer; 2011.
3. Schreiber G, Haran G, Zhou H-X. Fundamental aspects of protein– protein association kinetics. *Chemical reviews* 2009;109(3):839-860.
4. Gao Y, Roberts CC, Toop A, Chang CeA, Wheeldon I. Mechanisms of enhanced catalysis in enzyme–DNA nanostructures revealed through molecular simulations and experimental analysis. *ChemBioChem* 2016;17(15):1430-1436.
5. Bellucci L, Delfino F, Tozzini V. In silico design, building and gas adsorption of nano-porous graphene scaffolds. *Nanotechnology* 2020;32(4):045704.
6. Huber GA, McCammon JA. Brownian Dynamics Simulations of Biological Molecules. *Trends Chem* 2019;1(8):727-738.
7. Roberts CC, Chang C-eA. Modeling of enhanced catalysis in multienzyme nanostructures: effect of molecular scaffolds, spatial organization, and concentration. *Journal of chemical theory and computation* 2015;11(1):286-292.
8. Bruce NJ, Ganotra GK, Richter S, Wade RC. KBbox: A toolbox of computational methods for studying the kinetics of molecular binding. *Journal of chemical information and modeling* 2019;59(9):3630-3634.
9. Northrup S, Laughner T, Stevenson G. MacroDox macromolecular simulation program. Tennessee Technological University, Department of Chemistry, Cookeville, TN 1999.
10. Madura JD, Briggs JM, Wade RC, Davis ME, Luty BA, Ilin A, Antosiewicz J, Gilson MK, Bagheri B, Scott LR. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. *Computer Physics Communications* 1995;91(1-3):57-95.
11. Huber GA, McCammon JA. Browndye: a software package for Brownian dynamics. *Computer Physics Communications* 2010;181(11):1896-1905.
12. Martinez M, Bruce NJ, Romanowska J, Kokh DB, Ozboyaci M, Yu X, Öztürk MA, Richter S, Wade RC. SDA 7: A modular and parallel implementation of the

simulation of diffusional association software. *Journal of computational chemistry* 2015;36(21):1631-1645.

13. Długosz M, Zieliński P, Trylska J. Brownian dynamics simulations on CPU and GPU with BD_BOX. *Journal of computational chemistry* 2011;32(12):2734-2744.

14. De Biase PM, Markosyan S, Noskov S. BROMOC suite: Monte Carlo/Brownian dynamics suite for studies of ion permeation and DNA transport in biological and artificial pores with effective potentials. *Journal of computational chemistry* 2015;36(4):264-271.

15. Andrews SS. Smoldyn: particle-based simulation with rule-based modeling, improved molecular interaction and a library interface. *Bioinformatics* 2017;33(5):710-717.

16. Votapka LW, Jagger BR, Heyneman AL, Amaro RE. SEEKR: simulation enabled estimation of kinetic rates, a computational tool to estimate molecular kinetics and its application to trypsin–benzamidine binding. *The Journal of Physical Chemistry B* 2017;121(15):3597-3606.

17. Saadat A, Khomami B. Matrix-free Brownian dynamics simulation technique for semidilute polymeric solutions. *Physical Review E* 2015;92(3):033307.

18. Schöneberg J, Noé F. ReaDDy-a software for particle-based reaction-diffusion dynamics in crowded cellular environments. *PloS one* 2013;8(9):e74261.

19. Cholko T, Chang C-eA. Modeling Effects of Surface Properties and Probe Density for Nanoscale Biosensor Design: A Case Study of DNA Hybridization near Surfaces. *The Journal of Physical Chemistry B* 2021;125(7):1746-1754.

20. Kaushik S, Chia-en AC. Molecular Mechanics Study of Flow and Surface Influence in Ligand–Protein Association. *Frontiers in molecular biosciences* 2021;8.

21. Cholko T, Kaushik S, Chia-en AC. Dynamics and molecular interactions of single-stranded DNA in nucleic acid biosensors with varied surface properties. *Physical Chemistry Chemical Physics* 2019;21(29):16367-16380.

22. Gong P, Levicky R. DNA surface hybridization regimes. *Proceedings of the National Academy of Sciences* 2008;105(14):5301-5306.

23. English NJ, MacElroy J. Atomistic simulations of liquid water using Lekner electrostatics. *Molecular Physics* 2002;100(23):3753-3769.

24. Northrup SH, Allison SA, McCammon JA. Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *The Journal of Chemical Physics* 1984;80(4):1517-1524.
25. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic acids research* 2000;28(1):235-242.
26. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of chemical theory and computation* 2015;11(8):3696-3713.
27. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *Journal of computational chemistry* 2004;25(9):1157-1174.
28. Hassan SA, Mehler EL. A critical analysis of continuum electrostatics: the screened Coulomb potential–implicit solvent model and the study of the alanine dipeptide and discrimination of misfolded structures of proteins. *Proteins: Structure, Function, and Bioinformatics* 2002;47(1):45-61.
29. Huey R, Morris GM, Olson AJ, Goodsell DS. A semiempirical free energy force field with charge-based desolvation. *Journal of computational chemistry* 2007;28(6):1145-1152.
30. Stouten PF, Frömmel C, Nakamura H, Sander C. An effective solvation term based on atomic occupancies for use in protein simulations. *Molecular Simulation* 1993;10(2-6):97-120.
31. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *Journal of molecular graphics* 1996;14(1):33-38.
32. Schrodinger L. The PyMOL molecular graphics system, Version, 2.0 Schrödinger. LLC. Available at: <https://pymol.org/2/support.html>; 2010.
33. Peterson AW, Heaton RJ, Georgiadis RM. The effect of surface probe density on DNA hybridization. *Nucleic acids research* 2001;29(24):5163-5168.
34. Ellis RJ. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Current opinion in structural biology* 2001;11(1):114-119.
35. Popielec A, Ostrowska N, Wojciechowska M, Feig M, Trylska J. Crowded environment affects the activity and inhibition of the NS3/4A protease. *Biochimie* 2020;176:169-180.

36. Ishihama Y, Schmidt T, Rappsilber J, Mann M, Hartl FU, Kerner MJ, Frishman D. Protein abundance profiling of the Escherichia coli cytosol. BMC genomics 2008;9(1):1-17.
37. Crawford IP. Evolution of a biosynthetic pathway: the tryptophan paradigm. Annual review of microbiology 1989;43(1):567-600.
38. Braus GH, Luger K, Paravicini G, Schmidheini T, Kirschner K, Hütter R. The role of the TRP1 gene in yeast tryptophan biosynthesis. Journal of Biological Chemistry 1988;263(16):7868-7875.
39. Raves ML, Harel M, Pang Y-P, Silman I, Kozikowski AP, Sussman JL. Structure of acetylcholinesterase complexed with the nootropic alkaloid,(-)-huperzine A. Nature structural biology 1997;4(1):57-63.
40. Axelsen PH, Harel M, Silman I, Sussman JL. Structure and dynamics of the active site gorge of acetylcholinesterase: Synergistic use of molecular dynamics simulation and X-ray crystallography. Protein Science 1994;3(2):188-197.
41. Ripoll DR, Faerman CH, Axelsen PH, Silman I, Sussman JL. An electrostatic mechanism for substrate guidance down the aromatic gorge of acetylcholinesterase. Proceedings of the National Academy of Sciences 1993;90(11):5128-5132.
42. Tan RC, Truong TN, McCammon JA, Sussman JL. Acetylcholinesterase: electrostatic steering increases the rate of ligand binding. Biochemistry 1993;32(2):401-403.
43. Cheng S, Song W, Yuan X, Xu Y. Gorge motions of acetylcholinesterase revealed by microsecond molecular dynamics simulations. Scientific reports 2017;7(1):1-12.

5.7 Acknowledgements

This study was supported by the US National Science Foundation (MCB-1932984), US National Institutes of Health (GM-109045) and NSF national supercomputer centers (TG-CHE130009).

Chapter 6: Multiscale Modelling of Enzymatic Efficiency in Multienzyme Complex Contributed From Inter-Enzyme Distance, Orientation and Side Reactions: Insights From All-Atom Molecular Dynamics and Brownian Dynamics Simulations

6.1 Abstract

Molecular recognition is one of the most fundamental principles of chemistry and biology. Biomolecular complexes need to colocalize sequential catalytic and chemical processes such that the intermediates are precisely positioned to diffuse efficiently. The relative distance and orientation between the active sites within these complexes are predicted to affect the intermediate binding efficiency differently under different cell environments. Yet, the origin of catalytic efficiency in multienzyme complex remains elusive. In this work, Brownian dynamics simulations were employed to model an intermediate step in the tryptophan biosynthesis pathway in *E.coli*. We focus on the PRAI and IGPS as multienzyme constructs and the CdRP as the intermediate. We reported the average association time, dwell time and direct binding percentage when the substrate concentration is 10.0 μM as an estimated physiological concentration in cells. Our results suggest the catalytic efficiency can be achieved by optimizing the inter-enzyme distance and by properly aligning the orientation of the active sites. In addition, we found the intermolecular interactions between the substrate and macromolecules largely affect the catalytic efficiency. The catalytic enhancement in multienzyme complex depends on the distance between enzymes when the substrate leaking phenomenon existed. Our work highlights the importance of the enzyme spatial organization in regulating catalytic efficiency, broadens implications in catalytic processes under the complex cell

environments and brings biophysical insights into the enzyme-substrate association and understanding of target ligand/substrate kinetics.

6.2 Introduction

In biological systems, multienzyme cascades have achieved enzymatic efficiency, stereospecificity, and stereoselectivity by controlling the relative catalytic sites and physical distance between different biomolecules.¹⁻³ Alternatively, nature has evolutionally brought multiple enzymes in closed proximity to form multienzyme complex⁴⁻⁶ and colocalized them to produce desired compounds, or protected substrate/intermediates from being utilized by other macromolecules via substrate or metabolic channeling and transporting.⁷⁻¹² There has been a well-established theory that the evolution tends to combine different functions, particularly in the subsequential steps of metabolic pathways, into multifunctional enzyme complex.¹³ Inspired by nature's ingenuity, enzyme colocalization and utilizing multienzyme complex have become effective strategies in enzyme engineering, drug development and other commercial applications in simplified cell-free based and diluted systems.¹⁴⁻²¹ Understanding the molecular recognition in the substrate binding process is critical in understanding the protein/enzymes functions and provides insight into the enzyme design and drug development.

Inside living cells, however, there are concentrated media of biomolecules, making the sequential biochemical reactions complicated, which vary differently from *in vitro* environments. Macromolecules such as proteins, nucleic acids, lipids, small molecules, metabolites, and solvents are constantly moving and interacting with each other, exerting complications in understanding the molecular recognition mechanism and biomolecular

diffusion. Moreover, due to the cellular heterogeneity, spatial organization of enzymes and substrate concentrations vary in distinct cellular compositions, compartmentalization, and conditions, which further complicate the underlying concept of multienzyme complex in regulating the catalytic rates and diffusion pathways of biomolecules.

With the improvement of the experimental techniques, many experiments have been performed to investigate and measure biomolecules' diffusion constants under *in vivo* conditions.²²⁻²⁶ However, experimental measurements are limited to describe elementary processes and to understand the fundamental mechanisms involving in macromolecules, their timescales, and spatial distributions. Simulation methods can be applied to study the *in vivo* molecular diffusion with less expensive efforts. Atomistic molecular dynamics (MD) simulations are ideal to illustrate the molecular recognition process and protein dynamics by considering all atoms and explicit solvent. However, conditions, such as the average micrometer cell size and various sizes/shapes of biomolecules in the cell environment limit the use of MD simulations, partially due to the computational cost, and are system dependent. Without sacrificing much atomistic details and retaining sufficient accuracy, while being able to extend the space and timescale and to minimize the computational cost, Brownian Dynamics (BD) simulations can be applied to directly study substrate diffusion and non-specific interactions governed by physical properties in a cell-like environment.²⁷⁻³⁸ Bauler *et al.* modelled two enzyme spheres with oppsited charges between enzyme and substrate. Their results indicated the success of reactions were strongly determined by the orientation and distance between the active sites.³⁹ Trovato *et al.* advaced to bacteirial cytoplasm system, wherein size dependence of diffusion coffection,

anomalous exponents and viscosity effect were shown to largely affect molecular diffusion.⁴⁰ Working with the wild-type and single or double mutations of the globular domain of mouse linker, Ozturk *et al.*'s BD simulations results predicted the relative effect of single mutations related to protein binding and showed in good agreement with experimental study.⁴¹

Several studies and reviews have shown and accepted the viewpoint that the overall catalytic enhancement in multienzyme constructs on scaffolds are due to the “proximity effect”.^{14,16,21,42-46} However, other counterarguments have also stated that distance dependency in the multienzyme cascades is not the major contribution to the catalytic enhancement, but other factors, such as pH microenvironment near binding sites, multienzyme clustering and intrinsic enzyme dynamics, could also contribute to the overall catalytic enhancement.⁴⁷⁻⁵² For example, in tryptophan biosynthesis pathway, phosphoribosyl anthranilate isomerase (PRAI) and indoleglycerol phosphate synthase (IGPS) are sequential enzymes, which are responsible for synthesizing tryptophan in the third and fourth steps in the tryptophan biosynthetic pathway. In mesophile *E.coli*, PRAI and IGPS are fused together as a bifunctional enzyme complex,⁵³ while these two enzymes are separated as monofunctional enzyme in thermophile *T.mari*.^{54,55} (Figure 6.1) Existing studies showed that the catalytic rate in the PRAI is ~10 times faster than its succeeding IGPS, but the catalytic rate of the IGPS has no advantages when fused with PRAI over its respective monomeric construct.^{56,57} It is hypothetical that it could be the genomic level as the loop is closed so that to protect substrate, instead of functional enhancement. Yet, it reminds unknowns at many stages of how coordinated networks, protein dynamics and

structural arrangement affect enzymatic efficiency in a monomeric enzyme and in a multienzyme complex. To resolve the paradox and understand the diffusion mechanism in multienzyme cascades, in this work, we employed multi-scale simulations to investigate the catalytic efficiency in a multienzyme complex. We first studied the protein dynamic in the fused complex PRAI:IGPS, then we monitored the inter-enzyme distance by varying the relative distance of these two enzymes. We also included crowders in the systems based upon the distance dependent condition. In addition, we investigated competing side-reaction problem in the cell-like environment, mimicking conditions in which many intermediates are unstable and degraded during reaction processes. Lastly, we monitored the relative orientation of the catalytic sites of two enzymes.

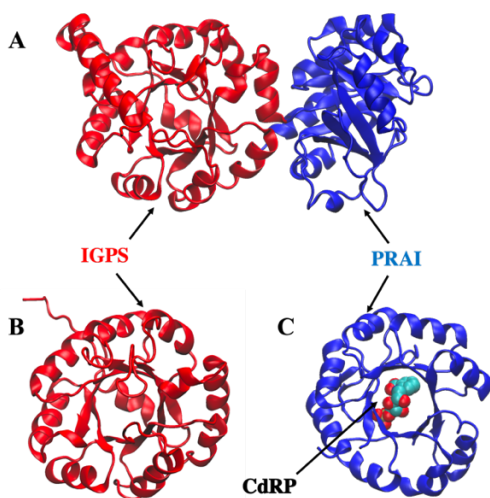


Figure 6.1. A. *Escherichia coli* (*E.coli*), mesophile, bifunctional TrpCF fusion protein (PDB 1PII, *trpCF*), phosphoribosyl anthranilate isomerase (PRAI), blue, Indoleglycerol phosphate synthase (IGPS), red. B. *Thermotoga maritima* (*T.mari*), thermophile, monomeric IGPS (PDB 1J5T, *trpC*). C. *Thermotoga maritima* (*T.mari*), thermophile, monomeric PRAI (PDB 1LBM, *trpF*), 1,2-carboxyphenylamino-1-deoxy-d-ribulose 5-phosphate (CdRP) in the catalytic site of PRAI.

6.3 Methods and Materials

Structure Preparation and Molecular Dynamics Simulations

The *E.coli* crystal structure TrpCF fusion enzyme (PDB ID: 1PII) was downloaded from the PDB website.⁵³ Because the TrpCF fusion enzyme structure used here did not have a bound substrate, we aligned CdRP-PRAI complex (species: *T.mari* and PDB ID: 1LBM)⁵⁴ with TrpCF to place substrate CdRP in the catalytic site of PRAI. AM1-BCC charge model was used to calculate the CdRP charge. Simulation was performed using ff14SB force field⁵⁸ for the enzymes and GAFF⁵⁹ for substrate and with molecular simulation package AMBER18. The sequential minimization steps were performed for the hydrogen atoms, side chains and the entire system complexes to clean any clashes. A standard AMBER protocol was used.^{60,61} The system was solvated in a TIP3P water box⁶² with a padding of 12 Å from the edge of the box to the whole system. 14 Na⁺ ions were added to neutralize the system in the water box. Heating of the system from 50 K to 300 K with an increment of 25 K was achieved in 200 ps (time step of 2 fs) for each temperature. The system was subsequently equilibrated at 300 K for 500 ps with all atoms relaxed. Long-range electrostatic interactions were computed by the particle mesh Ewald method.⁶³ The SHAKE constraint was applied all atoms including hydrogens.⁶⁴ The Langevin thermostat with a damping constant of 2 ps⁻¹ was applied to maintain a temperature of 300 K. The NPT ensemble was used for MD production run for 400 ns, and each frame was collected at 1 ps time interval. To get substrate dissociation, we proceeded to accelerated MD (aMD) simulations with different levels of alpha boost potential ($\alpha=0,1,2,3,4,6$) (see Table 6.1 for

details). The simulation time step and trajectory frame interval were the same as classical MD mentioned above.

Table 6.1: List of all aMD simulations for substrate 1,2-carboxyphenylamino-1-deoxy-d-ribulose 5-phosphate (CdRP) dissociation from PRAI and association to IGPS. Red: The substrate successfully dissociated from the pocket of PRAI. Blue: The substrate dissociated from the PRAI and associated to the IGPS successfully. Black: The substrate did not successfully dissociate from the PRAI pocket. (*a), different boost potentials.

Accelerate d MD index	aMD with different boost potentials (*a)	Simulation time (ns)	Dissociation of substrate CdRP	Final substrate location	Note
1	crp_amd0a_1.dcd	200ns	No	Pocket of PRAI	n/a
2	crp_amd0a_2.dcd	200ns	No	Pocket of PRAI	n/a
3	crp_amd0a_3.dcd	200ns	No	Pocket of PRAI	n/a
4	crp_amd1a_1.dcd	200ns	No	Pocket of PRAI	n/a
5	crp_amd1a_2.dcd	200ns	No	Pocket of PRAI	n/a
6	crp_amd1a_3.dcd	200ns	No	Pocket of PRAI	n/a
7	crp_amd2a_1.dcd	200ns	No	Pocket of PRAI	n/a
8	crp_amd2a_2.dcd	200ns	No	Pocket of PRAI	n/a
9	crp_amd2a_3.dcd	200ns	No	Pocket of PRAI	n/a
10	crp_amd3a_1.dcd	200ns	No	Pocket of PRAI	IGPS loop is open
11	crp_amd3a_2.dcd	200ns	No	Pocket of PRAI	n/a
12	crp_amd3a_3.dcd	200ns	No	Pocket of PRAI	IGPS loop is open
13	crp_amd4a_1.dcd	200ns	Yes	Return to pocket of PRAI	substrate dissociated and reassociated in the pocket
14	crp_amd4a_2.dcd	200ns	No	Pocket of PRAI	IGPS loop is partially open
15	crp_amd4a_3.dcd	200ns	No	Pocket of PRAI	n/a
16	crp_amd4a_4+a1.d cd	400ns	No	Pocket of PRAI	n/a
17	crp_amd4a_4+a2.d cd	400ns	No	Pocket of PRAI	n/a
18	crp_amd4a_4+a3.d cd	400ns	No	Pocket of PRAI	n/a
19	crp_amd4a_5.dcd	200ns	No	Pocket of PRAI	n/a
20	crp_amd4a_6.dcd	200ns	No	Pocket of PRAI	n/a
21	crp_amd4a_7+a1.d cd	440ns	Yes	Edge of PRAI	n/a

22	crp_amd4a_7+a2.d cd	440ns	Yes	Return to pocket of PRAI	substrate moved to edge, left the pocket of PRAI, but diffused back to pocket
23	crp_amd4a_7+a3.d cd	400ns	Yes	Edge of IGPS	Dissociated from PRAI, diffused, and stayed near the edge of IGPS
24	crp_amd4a_8+a1.d cd	400ns	Yes	Edge of IGPS	Dissociated from PRAI, diffused, and stayed near the edge of IGPS
25	crp_amd4a_8+a2.d cd	440ns	Yes	Interface, near to the PRAI side	Dissociated from PRAI and stayed at the interface of enzyme
26	crp_amd4a_8+a3.d cd	400ns	No	Pocket of PRAI	n/a
27	crp_amd6a_1.dcd	200ns	No	Pocket of PRAI	IGPS loop is open
28	crp_amd6a_2.dcd	200ns	No	Pocket of PRAI	n/a
29	crp_amd6a_3+a1.d cd	400ns	Yes	Pocket of IGPS	Dissociated and associated successfully to IGPS
30	crp_amd6a_3+a2.d cd	380ns	Yes	Back of the enzyme's interface	Dissociated, diffused to the back of the interface of enzyme
31	crp_amd6a_3+a3.d cd	400ns	Yes	Return to edge of PRAI	Dissociated and diffused back to PRAI pocket

32	crp_amd6a_4+a1.d cd	360ns	Yes	pool	Dissociated, diffused to the near edge of IGPS, then diffused away
33	crp_amd6a_4+a2.d cd	400ns	Yes	Edge of IGPS	Dissociated and diffused to the edge of IGPS
34	crp_amd6a_4+a3.d cd	400ns	Yes	Edge of IGPS	Dissociated and diffused to the edge of IGPS

Brownian Dynamics (BD) Simulations

Given the fact that the loop of the IGPS is closed in the TrpCF fusion enzyme crystal structure, and BD simulations is performed in a rigid-body construct, one frame with loop open in the IGPS's binding pocket was extracted from the 400 ns aMD simulation trajectory (aMD index 29 in Table 6.1). The center of mass of the two enzymes were considered, we manually separated two enzymes from the fusion enzyme (35 Å) and moved IGPS 50 Å, 100 Å and 150 Å way with respective to PRAI (Figure 6.2 A), while keeping enzymes' catalytic site orientations unchanged. To simulate cell-like environment, 19 ribonucleases (PDB ID 4IBN)⁶⁵ serving as crowders were added and held rigid randomly in the spherical volume, which is equivalent to 20% of the occupied simulation volume.⁶⁶ To mimic a condition in which substrates are consumed in a cell-like environment (for example, compete site-reaction), four crowder molecules having termination sites (two are ~100 Å away from the PRAI's staring point and two are ~200 Å away from the PRAI's starting point) were selected (Figure 6.2 B, orange circle). To

investigate the structural orientation effects when two catalytic sites may face each other or be further apart (Figure 6.2 C), we varied the relative orientations of the two enzymes. The catalytic site in each enzyme was rotated in opposite directions by a 90 degrees interval, generating six simulated systems in the cell-environment.

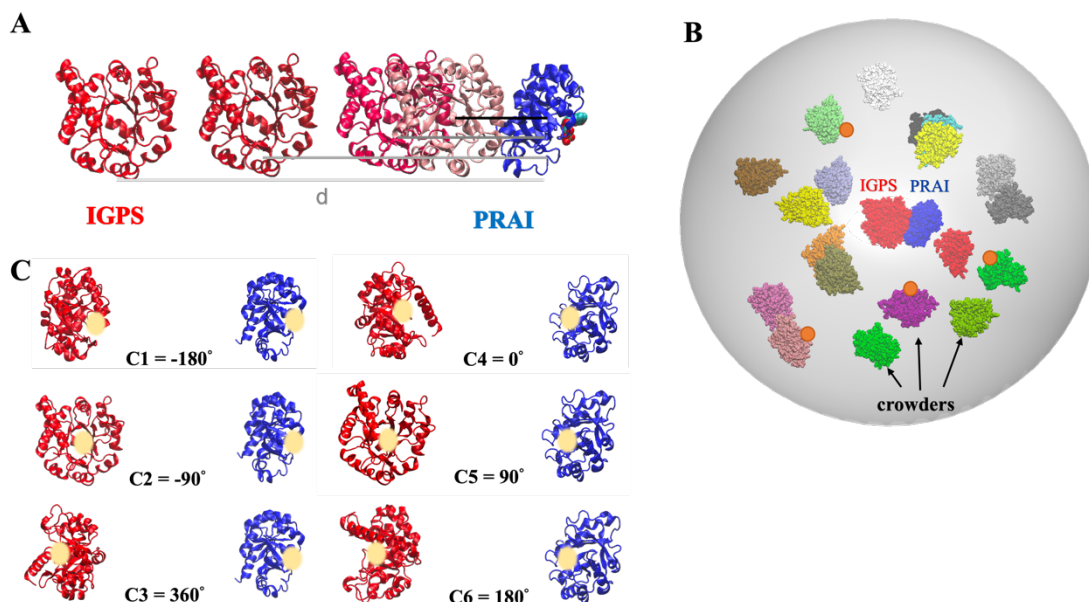


Figure 6.2. Schematic of simulation setups, A. Distance is varied from ~ 35 Å (fusion) to 150 Å, based on the center of mass of the two enzymes. B. 20% crowding model, TrpCF (fusion) is centered in the sphere. The crowder proteins are placed in the simulation sphere randomly. Four crowders having termination sites are marked with orange circle. C. Relative orientation of the two active sites at a given 100 Å apart.

The GeomBD3 package was used to carry out BD simulations.⁶⁷ The simulation temperature was 298.15 K. Simulation space was spherical with a radius of 341 Å, which is equivalent to the estimated substrate concentration of 10.0 μM in cells. Ionic strength was 0.1 M NaCl. Simulation time step was 0.05 ps. Substrate binding was successful when the oxygen atom of CdRP and residue K55 of IGPS was within 8.0 Å (Figure 6.3), and the simulation was terminated. The total charge for the CdRP is -3.0 and the fusion enzyme

TrpCF is -11.0. The charge for PRAI is -7.0 and the charge for IGPS is -4.0. The charge for the crowder enzyme ribonuclease is +2.0.

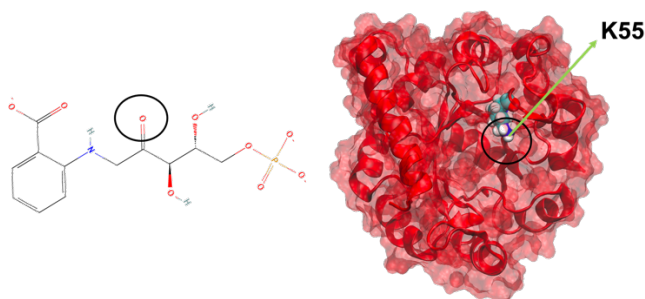


Figure 6.3. Left. 2D chemical structure of 1,2-carboxyphenylamino-1-deoxy-d-ribulose 5-phosphate (CdRP). Right. Indoleglycerol phosphate synthase (IGPS) with K55 shown in vdW representation.

The simulation program employed pre-computed grid representations for the enzymes' excluded volume potential, screened electrostatic potential, van der Waal (vdW) potential and desolvation potentials. The grids were extended up to 15 Å and 40 Å for the vdW and screened electrostatic potentials, respectively, for substrate diffusion. The substrate diffused in the implicit solvent model using overdamped Langevin equation

$$r_i(t + \Delta t) = r_i(t) - \frac{D_i}{k_b T} \frac{\partial E}{\partial r_i} \Delta t + \sqrt{2D_i \Delta t} R \quad \text{Eq. 6.1}$$

where D_i is the translation/rotational diffusion coefficient, K_b is Boltzmann's constant, T is the simulation temperature. $\partial E / \partial r_i$ is the potential gradient, Δt is the time step, R is stationary Gaussian random number with a zero mean.

In each simulation, 100 independent substrate replicas diffused from the same initial position of the PRAI's catalytic site with different random number seed. Once a replica satisfied the binding criteria, it terminated, and a new independent replica generated from

the initial position. Each simulation was terminated once the standard error of mean (SEM) for association time is 5% or below and with at least 500 successful bindings.

6.4 Results and Discussion

We began with aMD simulations for the CdRP–PRAI:IGPS fusion complex to examine how CdRP left the catalytic pocket of PRAI and associate to the catalytic site of IGPS with atomistic details, where CdRP started in the catalytic site of PRAI. Using conformations obtained from aMD runs, we carried out BD simulations to investigate CdRP diffusion with different enzyme arrangements and micro-environments. We analyzed the effects of the inter-enzyme distances, rotational orientations of PRAI and IGPS, presence of crowders and presence of side reactions. Each simulation set began with the fused PRAI:IGPS complex system, which is unique in *E. coli*. Keeping the two enzymes' relative orientations, we examined the effects from three different inter-enzyme distances for CdRP association to IGPS, with center of mass distance 50Å, 100Å and 150Å, followed by adding the environmental factors, such as crowders and crowders with competing side reactions. Interestingly, the fused PRAI:IGPS complex did not appear to provide advantages for substrate diffusion. To examine the hypothesis that the fused PRAI:IGPS complex may not yield the optimal orientation of the two catalytic sites, we selected inter-enzyme distance 100Å by rotating the enzymes to investigate effects from various relative orientations of the two pockets. This study focuses on evaluating CdRP association time. The dwell time and direct binding probability were also recorded to provide information regarding possible increase in local substrate concentration.

Modeling CdRP Diffusion From PRAI to IGPS With aMD Simulations.

We investigated the processes of the intermediate CdRP traveling from PRAI to the next enzyme IGPS in the tryptophan biosynthesis pathway. To ensure that a catalytic product may easily dissociate from the enzyme, binding free energy between the product and enzyme are usually not strong. However, because conformational changes are required for CdRP dissociation processes, it will need excessively long classical MD simulations to sample a whole dissociation process. We therefore utilized an enhanced sampling method aMD with explicit solvent to examine the processes in atomistic details. CdRP made multiple attempts to leave the catalytic site of PRAI, where 12 successful dissociations were sampled among a total of 34 aMD runs (Table 6.1). Although the intermolecular attractions retained CdRP on the surface of PRAI:IGPS complex, not all surface diffusion resulted in final substrate binding to the catalytic site of IGPS. (Figure 6.4) Notably, during CdRP surface diffusion, although the enzyme complex did not change the overall conformations, local side-chain rearrangements allowed formation of favorable interactions that kept CdRP in a few locations on the surface of PRAI:IGPS. CdRP could stay within the local energy minima for longer than 400 ns in the aMD runs, and we therefore terminated the simulation when CdRP was trapped in a local energy minimum.

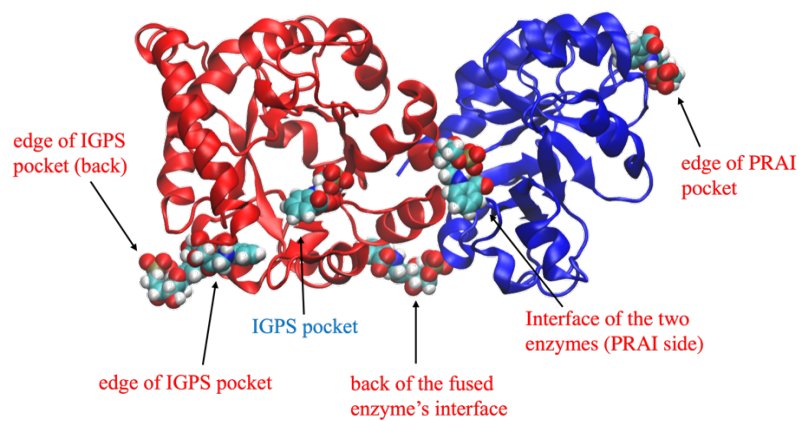


Figure 6.4. Successful CdRP dissociations from aMD simulations with the final locations on the surface of the PRAI:IGPS fusion complex. One successful CdRP association to the IGPS pocket is shown in blue.

Figure 6.5 shows one successful unbinding and binding event illustrated by the distance between the center of the IGPS pocket and CdRP. Starting from locating CdRP in the catalytic site of PRAI (Figure 6.5 A), after local fluctuations for ~ 100 ns, CdRP moved to the edge of the PRAI binding pocket (Figure 6.5 B). Once CdRP left the pocket, it quickly moved to a groove comprised by a loop and helix of PRAI (at ~ 145 ns, Figure 6.5 C) and stayed locally for longer than 100 ns (an example shown at 200 ns in Figure 6.5 D). At ~ 264 ns, CdRP moved away from the surface only 3 ns, rebound back to a cleft between two α -helices of IGPS near the surface of PRAI:IGPS fusion complex, and was trapped in the region for another 100 ns (Figures 6.5 E and F). Once CdRP left the local energy minimum, it took the substrate only 15 ns to bind to the catalytic site of IGPS (Figure 6.5 G). Of note, although our aMD sampled preferred locations while CdRP diffusing from PRAI to IGPS, we did not observe a major diffusion pathway(s) along the surface. In contrast, other systems such as malate dehydrogenase (MDH) and citrate synthase (CS)

metabolons have a substrate channeling guiding mechanism,^{34,35,68} or tryptophan synthase (TRPS) utilizes a physical tunnel connecting two enzyme subunits to prevent substrate leaking.^{9,69} The simulations also showed that the loops near the catalytic sites of both enzymes needed to open up to allow CdRP leaving or binding to the pocket.

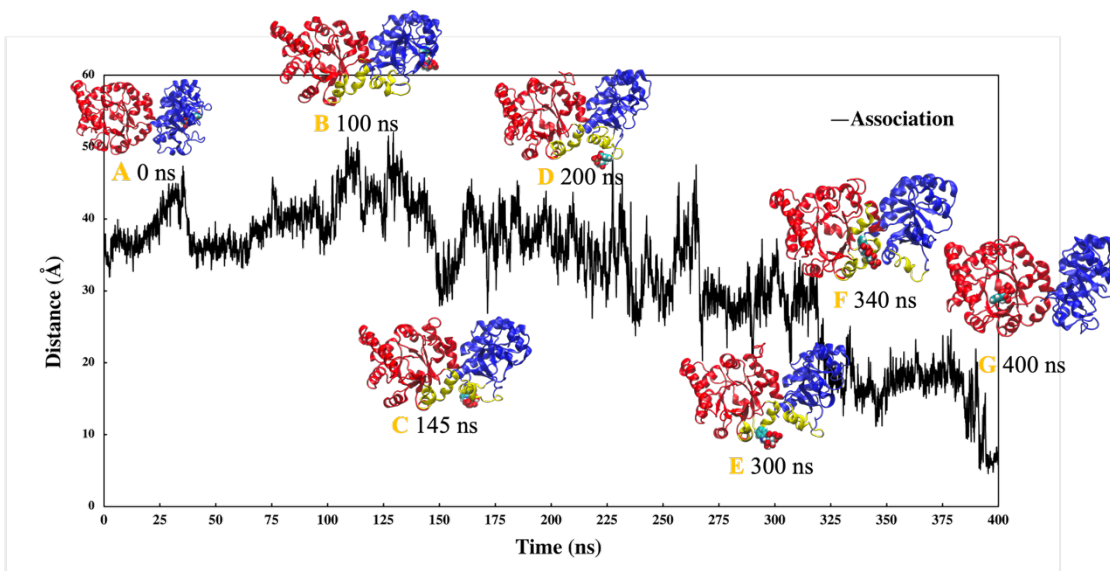


Figure 6.5. The association pathway of CdRP in the PRAI:IGPS fusion complex in a 400 ns aMD.

Modeling Substrate Diffusion in Test-Tube and Cell-Like Environments Using BD Simulations.

To investigate factors affecting inter-enzyme diffusion that may alter the overall product synthesis rates, we used BD simulations to obtain sufficient sampling for modelling each system. Existing studies showed a significant advantage in catalytic enhancement by reducing inter-enzyme distance using engineered scaffolds.^{14,70} We therefore suspected that the PRAI:IGPS fused complex might yield the most efficient CdRP inter-enzyme transportation because the fused complex provided possibly the smallest inter-enzyme distance between the two sequential enzymes, PRAI and IGPS. To examine the effects on

inter-enzyme distance, we manually separated the fused PRAI:IGPS complex in the BD runs and analyzed the CdRP association time. The relative orientation of the two enzymes remained the same. Because the BD simulations applied rigid-body Brownian movements, a conformation from an aMD trajectory was selected, where the active sites of the two enzymes were accessible for CdRP binding/unbinding without changing the enzyme conformation. For example, the crystal structure of IGPS shows a closed loop near the catalytic site; however, the loop must switch to an open conformation to allow CdRP binding. We therefore selected a loop open conformation in the IGPS from the aMD simulations.

When diffusing in the solvent, our BD movement yielded the translational diffusion coefficient of CdRP, $D_{\text{substrate}} = \sim 7.37 \pm 0.74 \times 10^{-6} \text{ cm}^2/\text{s}$ (Table 6.2) using the radius of gyration of CdRP $\sim 3.71 \text{ \AA}$. The modeled substrate diffusion coefficient is in a ballpark with analytical values, $D_{\text{analytical}} = 5.0 \times 10^{-6} \text{ cm}^2/\text{s}$, which also validates our simulation setting and BD algorithm. Using $D_{\text{substrate}}$ to approximate the diffusion time without considering any surface diffusion, CdRP directly diffusing $50 \text{ \AA}$, $100 \text{ \AA}$ and $150 \text{ \AA}$ from one enzyme to another takes lesser than 25ns, 70 ns and 160ns, respectively. However, CdRP took significantly longer time than the approximated diffusion time to arrive the catalytic site of IGPS, and the medium association time was all longer than $8.4 \mu\text{s}$ (Table 6.3). Interestingly, the fused PRAI:IGPS complex did not assist a faster substrate association, and the most efficient binding happened when PRAI and IGPS were separated with inter-enzyme distance $50 \text{ \AA}$. Increasing the inter-enzyme distance induced an initial delay in the association distribution (Figure 6.6) however, the overall probability of direct substrate-

diffusion association remained similar. Because all three molecules, PRAI, IGPS and CdRP, carry negative charge, hence, no long-range electrostatic steered to accelerate CdRP binding. Shorter inter-enzyme distance or even fusing PRAI:IGPS does not promote more direct-diffusion association.

Table 6.2. Calculation of diffusion coefficient for different BD simulations.

BD index	Diluted	BD index	Test-tube environment
Distance	D_T (10⁻⁶ cm²/s)	Distance	D_T (10⁻⁶ cm²/s)
A1 (fusion)	7.37±0.74		
A2 (d = 50 Å)	7.29±0.74	B1 (fusion)	6.27±1.27
A3 (d = 100 Å)	7.40±0.97	B2 (d = 50 Å)	6.30±1.24
A4 (d = 150 Å)	7.24±0.94	B3 (d = 100 Å)	6.82±1.12
Orientation	D_T	B4 (d = 150 Å)	6.07±1.26
(d = 100 Å)	(10⁻⁶ cm²/s)		
C1 (-180°)	7.38±0.61		
C2 (-90°)	7.40±0.97		
C3 (360°)	7.33±0.65		
C4 (0°)	7.38±0.54		
C5 (90°)	7.13±0.65		
C6 (180°)	7.18±0.75		

Table 6.3. Medium Association time (T_{asso}), Binder Percentage in IGPS and in the near Competitors for Different Inter-enzyme Distance Setups in three Simulation Conditions

Simulation conditions	Simulation index	Interenzyme distance	Medium $T_{\text{asso}}(\mu\text{s})$	Probability direct substrate diffusion %	Consumed by side reactions
Test-tube environment	A1	PRAI:IGPS (fusion)	10.4	1.24	N/A
	A2	50Å	8.4	1.20	N/A
	A3	100Å	9.6	n/a	N/A
	A4	150Å	11.5	n/a	N/A
Cell-like environment with side reactions	B1	PRAI:IGPS	3.2	0.78	47.5%
	B2	50Å	3.4	n/a	44.4%
	B3	100Å	3.7	n/a	44.3%
	B4	150Å	4.4	n/a	45.1%

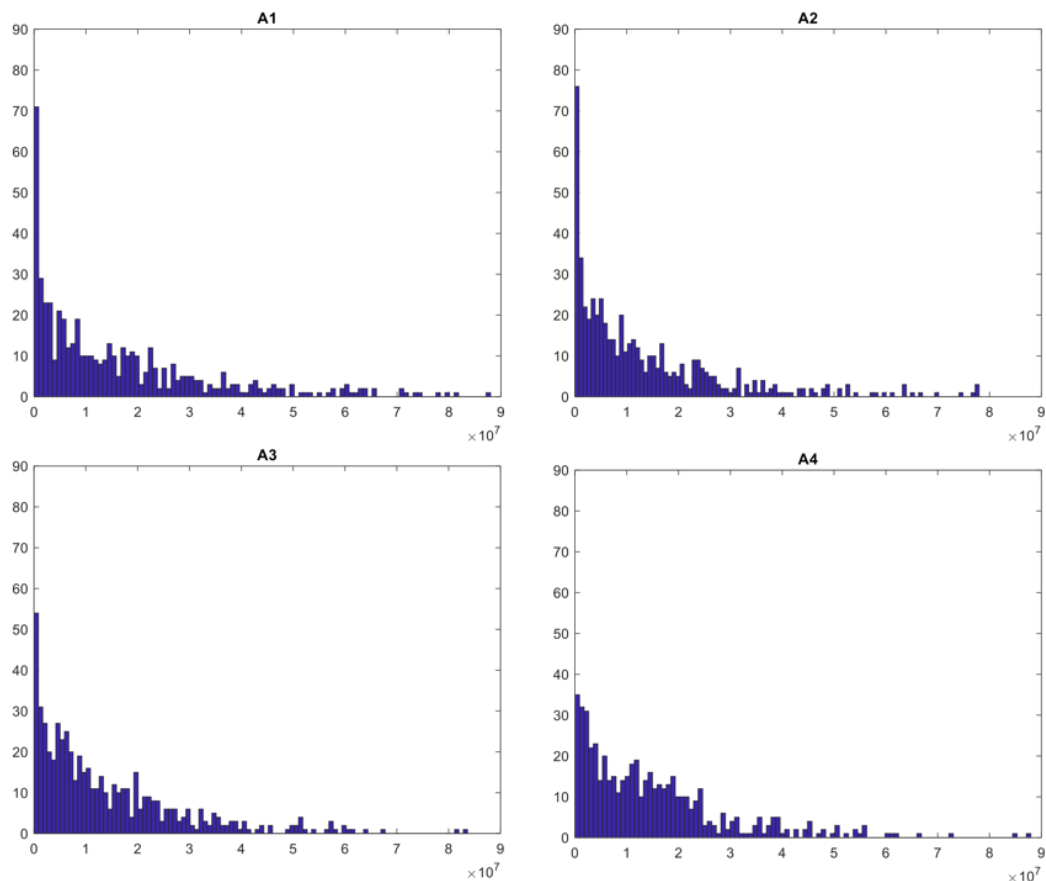


Figure 6.6. Binding time distribution for PRAI and IGPS enzymes at the different inter-enzyme distances.

Notably, here the substrate direct-diffusion ignores possible surface diffusion which retains a substrate near the catalytic site to increase local effective substrate concentration, which may also increase the overall catalytic rate. We therefore measured (Table 6.4). Interestingly, with 50 Å inter-enzyme distance, CdRP stayed the longest, in average of 3.25 ns on IGPS surface, where IGPS has several concave areas to keep CdRP especially in the regions close to PRAI in the current orientation setting (Figure 6.7 grey and yellow). Because the negatively charged nature of IGPS, non-polar interactions are the major driving force to retain CdRP locally. The closer proximity with 50 Å inter-enzyme distance

provides higher possibility of CdRP to reach the grey/yellow regions, resulting in longer dwell time. However, these locations are occluded when IGPS is fused with PRAI, leading to shorter dwell time of CdRP in the fused PRAI:IGPS complex.

Table 6.4: Direct Binding Percentage and Dwell time for Different Inter-Enzyme Distance Setups in Three Simulation Conditions

Simulation Conditions	BD sim index	P_{direct}	Dwell time (ns)
Test-tube environment	A1 (fusion)	1.24	1.50
	A2 (50Å)	1.20	3.25
	A3 (100Å)	-	2.00
	A4 (150Å)	-	1.25
Cell-like environment with substrate leaking	B1 (fusion)	0.78	1.25
	B2 (50Å)	-	2.75
	B3 (100Å)	-	1.00
	B4 (150Å)	-	2.00

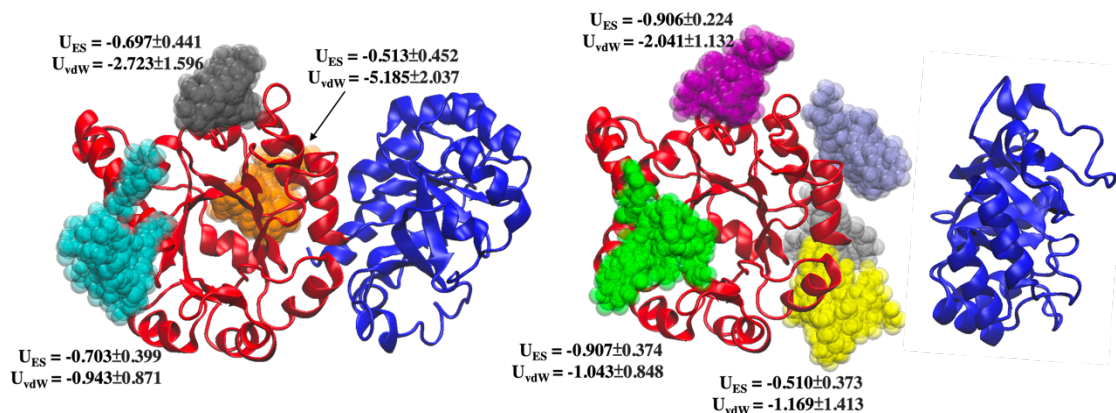


Figure 6.7. Substrate binding map on the surface of the enzymes. Substrate binding energy (kcal/mol) in different surfaces of the IGPS enzyme. PRAI:IGPS fused complex system (left), 50Å inter-enzyme system (right).

Our results suggest that in test tube environments, fusing IGPS and PRAI is not an optimal setting to reduce inter-enzyme diffusion time. Because the two catalytic sites are not facing each other in the fused IGPS:PRAI, the local geometry condition is not optimal for CdRP transportation. Existing studies also suggested that orientations of the multienzyme complexes affected the substrate diffusion and their reaction rates.^{15,35,71-74} To understand the relative enzyme orientation in this case, we chose 100 Å inter-enzyme distance and rotated the two enzymes to model the relative positions of the two catalytic sites. As anticipated, the fastest binding and the highest direct-diffusion association events (13.6%), occurred when the two catalytic sites facing each other (BD setting index C4 in Table 6.5). With rotation of the enzymes, CdRP had fewer chances to directly diffusing to IGPS and the association time was delayed (Figure 6.8). However, because of the concave catalytic site region of IGPS, the relative rotation of -90° or +90° produced different distributions in association time. CdRP can more easily access the binding site with +90° orientation (Figure 6.2 C), resulting in significantly more direct diffusion association (7.4%),

compared with that of -90° orientation (1.6%). The study demonstrates the importance of geometric details of the enzymes used for spatial organization for optimizing the overall catalytic rates.

Table 6.5: Medium Association time (T_{asso}) and Dwell time for Different Orientation Setups in Diluted Simulation Condition

BD sim index	$\langle T_{\text{asso}} \rangle \pm \text{std } (\mu\text{s})$	Medium $T_{\text{asso}}(\mu\text{s})$	Dwell time (ns)
C1 (-180°)	14.5 $\pm$ 15.1	9.7	1.50
C2 (-90°)	14.0 $\pm$ 14.1	9.6	2
C3 (360°)	16.2 $\pm$ 15.9	10.8	1
C4 (0°)	12.6 $\pm$ 14.4	8.2	0.75
C5 (90°)	14.6 $\pm$ 14.9	10.6	1
C6 (180°)	14.8 $\pm$ 15.0	10.0	1

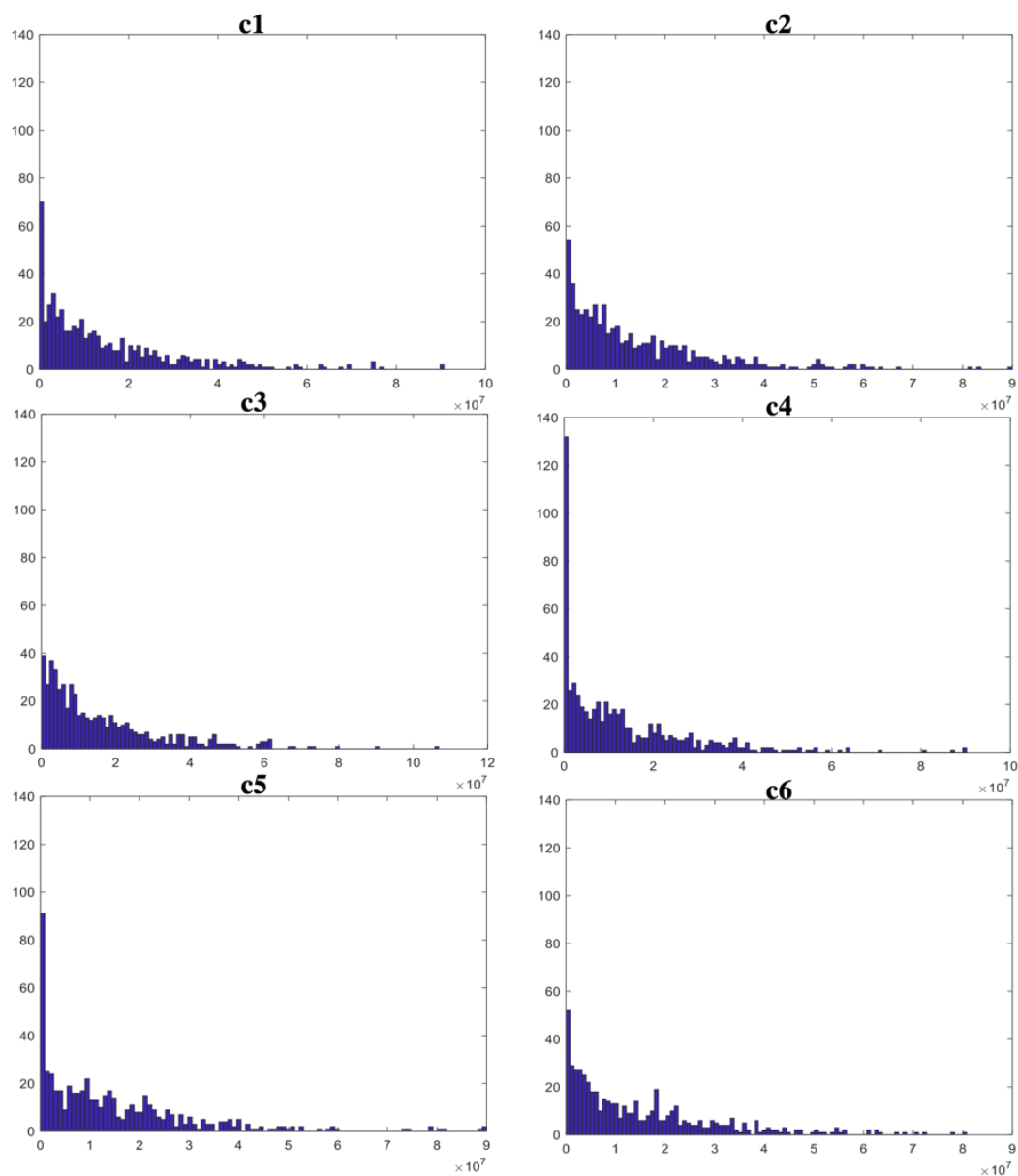


Figure 6.8. Binding time distribution for PRAI and IGPS enzymes at the different binding pocket orientation.

Notably, even catalytic sites may be occluded and become unavailable, other factors can still bring advantages in the overall reactions. For example, although the active sites of acetylcholinesterase (AChE) are partially occluded when assembling four AChE in the compact structure, electrostatic forces still substantially enhance the associate rate of the

natural substrate in the tetramer form relative to the monomer.⁷⁵ Nature considers all the features to optimize the system, and the tetramer is the most important functional form under physiological conditions. Moreover, assembling multiple enzymes to a multi-enzyme complex may accelerate conformational changes and/or allow sophisticated allosteric regulation, such as the dimeric unit of tryptophane synthase.^{76,77} However, previous catalytic studies in test-tube environment showed that fused PRAI:IGPS did not enhance catalytic efficiency and suggested that the complex did not facilitate conformational changes needed in the system.⁵⁷ Although the rigid-body BD studies cannot reiterate conclusions from experiments regarding protein dynamics and catalysis, our studies suggest no advantages in substrate diffusion in fused PRAI:IGPS complex, compared with the arrangement with inter-enzyme distance 50 Å in the test-tube environment. Nevertheless, without fusing the two enzymes, it can be challenging in a cell environment to co-localize the two enzymes to bring them in a proximity. In brief, our studies demonstrated that maximum intermediate substrate transfer was achieved with small (<50 Å) inter-enzyme separation and catalytic sites oriented toward one another.

Modeling Substrate Diffusion in Cell-Like Environments With Side-Reactions.

Unlike dilute environments used in most biochemistry assays, approximately 20%-40% of the cytoplasmic volume is occupied by various molecules. In addition, competing side reactions happen which can also consume the same substrate. Of note, CdRP has a phosphate group where cellular environments have multiple enzymes to perform dephosphorylation. Therefore, it is of particular interest to investigate the effects on substrate diffusion when IGPS and PRAI are fused or in close enzyme proximity in cell-

like environment. To create a cell-like environment, we randomly placed 19 ribonucleases in our simulation to occupy 20% of the total volume.

Although the simulation volume was reduced because of the volume occupied by the crowders (~20% reduced volume), the crowding effect in fact dominated by slowing down the substrate diffusion. Theoretically speaking, the reduced volume caused the substrate traveling time shorter from PRAI to IGPS in confined space when interactions between substrate and macromolecules do not exist. However, trajectory analysis revealed that substrates bound to the crowders anisotropically for a certain period before reentering the bulk solution and continue to diffuse. (Figure 6.9) Hence, the substrate translational diffusion coefficients were also reduced. It is suggestive that substrates' shorter traveling time was balanced out by the longer substrate-crowder interaction time. The results indicated that the crowder biomolecules having non-specific interactions were the driving force for the enhancement of the average association time.⁷⁸⁻⁸³

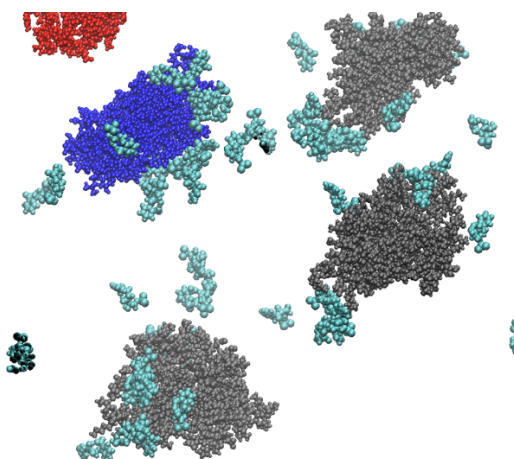


Figure 6.9. Screenshot of non-specific interactions between substrate, enzymes and crowders.

The fusion PRAI:IGPS complex has the shortest inter-enzyme distance between active sites, but with the longest average association times with respect to its counterparts in the systems, indicating that the proximity channeling provides only temporary benefit for substrate traveling distance but does not have a long-lasting catalytic enhancement.^{48,50,52,84} In other words, by simply combining two enzymes as a multienzyme cascade or put them in close proximity does not help to enhance catalytic efficiency in the cell-like environment.

Of note that the hydrodynamic interaction (HI) was excluded in our BD simulations since the enzymes, crowders were held rigid, and substrate CdRP is relatively small compared with other macromolecules in the system. Moreover, the potential energy grids were statically positioned for substrate diffusion during the simulations. It is yet an important physical property affecting molecular diffusion, and hence required detailed analysis when including macromolecules.⁸⁵⁻⁸⁷

In the cell-like environment with compete side reactions, a couple crowders in the cell-like environment were enabled to consume the substrate during the simulations, mimicking cases in which some substrates can be degraded by other enzymes or simply competitors. In these simulations, two crowders having the termination site were placed at ~ 100 Å and two crowders having the termination sites were placed at ~ 200 Å away from PRAI starting point. Previous studies in combining mathematical modeling with *in vivo E.coli* experiment and applying mesoscopic modeling with metal-organic framework (MOF) experiment

supported our simulation results that the catalytic efficiency can be achieved with proximity effect when side reaction condition or competitor exist.^{51,72,84}

6.5 Conclusions

In this work, we accessed the catalytic efficiency with spatial organization of PRAI and IGPS enzymes in the tryptophan biosynthesis pathway. To achieve this goal, we applied all-atom multiscale simulations to investigate the enhancement in catalysis when the inter-enzyme distance and the orientation of catalytic sites are varied in different cell environments. There has been a subject of yet to be addressed question about the “proximity effect”. Our work has shown that the multienzyme complex in regulating the catalytic efficiency is cell environment dependent. We showed that two enzymes fused as multienzyme cascades can enhance reactions when compete side reactions exist. In the diluted or cell-like closed environments, the fusion enzyme has the longest association time with respect to its counterpart system. Hence, we believe that simply bringing two enzymes in close proximity or fuse two enzymes in a low enzyme concentration environment or in a closed environment with other macromolecules are present cannot increase the reaction rate. The reaction rate is largely affected when crowding molecules are present. The dominant driving force for the association time enhancement is due to non-specific interactions between the crowders and substrate. Our simulation results from varying the catalytic sites of enzymes also suggested that proper catalytic sites alignment can provide beneficial in reaction rate. Our study provides an alternative explanation computationally to understand the multienzyme catalysis in confined space and highlights the importance of spatial organization in regulating the catalytic efficiency. Interesting directions for the

further study will be to include the enzymes' dynamics and the hydrodynamic interactions in the simulations.

6.6 References

1. Jandt U, You C, Zhang Y-P, Zeng A-P. Compartmentalization and metabolic channeling for multienzymatic biosynthesis: practical strategies and modeling approaches. *Fundamentals and Application of New Bioproduction Systems* 2013:41-65.
2. Sperl JM, Sieber V. Multienzyme Cascade Reactions-Status and Recent Advances. *Acs Catalysis* 2018;8(3):2385-2396.
3. Schrittwieser JH, Velikogne S, Hall ML, Kroutil W. Artificial biocatalytic linear cascades for preparation of organic molecules. *Chemical Reviews* 2018;118(1):270-348.
4. Crawford IP. Evolution of a biosynthetic pathway: the tryptophan paradigm. *Annual review of microbiology* 1989;43(1):567-600.
5. Tran L, Broadhurst RW, Tosin M, Cavalli A, Weissman KJ. Insights into protein-protein and enzyme-substrate interactions in modular polyketide synthases. *Chemistry & biology* 2010;17(7):705-716.
6. Patel M, Korotchkina L. Regulation of the pyruvate dehydrogenase complex. *Biochemical Society Transactions* 2006;34(2):217-222.
7. Hyde C, Ahmed S, Padlan E, Miles EW, Davies D. Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from *Salmonella typhimurium*. *Journal of Biological Chemistry* 1988;263(33):17857-17871.
8. Dunn MF, Aguilar V, Brzovic P, Drewe Jr WF, Houben KF, Leja CA, Roy M. The tryptophan synthase bienzyme complex transfers indole between the. alpha.-and. beta.-sites via a 25-30. Å long tunnel. *Biochemistry* 1990;29(37):8598-8607.
9. Dunn MF, Niks D, Ngo H, Barends TRM, Schlichting I. Tryptophan synthase: the workings of a channeling nanomachine. *Trends in Biochemical Sciences* 2008;33(6):254-264.
10. Wheeldon I, Minter SD, Banta S, Barton SC, Atanassov P, Sigman M. Substrate channelling as an approach to cascade reactions. *Nature chemistry* 2016;8(4):299-309.
11. Lin J-L, Palomec L, Wheeldon I. Design and analysis of enhanced catalysis in scaffolded multienzyme cascade reactions. *Acs Catalysis* 2014;4(2):505-511.
12. Huang X, Holden HM, Raushel FM. Channeling of substrates and intermediates in enzyme-catalyzed reactions. *Annual review of biochemistry* 2001;70(1):149-180.

13. Hütter R, Niederberger P, DeMoss JA. Tryptophan biosynthetic genes in eukaryotic microorganisms. *Annual review of microbiology* 1986;40:55-77.
14. Fu J, Liu M, Liu Y, Woodbury NW, Yan H. Interenzyme substrate diffusion for an enzyme cascade organized on spatially addressable DNA nanostructures. *Journal of the American Chemical Society* 2012;134(12):5516-5519.
15. Lim SI, Yang B, Jung Y, Cha J, Cho J, Choi E-S, Kim YH, Kwon I. Controlled orientation of active sites in a nanostructured multienzyme complex. *Scientific reports* 2016;6(1):1-9.
16. Quin MB, Wallin K, Zhang G, Schmidt-Dannert C. Spatial organization of multi-enzyme biocatalytic cascades. *Organic & biomolecular chemistry* 2017;15(20):4260-4271.
17. Sperl JM, Sieber V. Multienzyme Cascade Reactions Status and Recent Advances. *Acs Catalysis* 2018;8(3):2385-2396.
18. Hwang ET, Lee S. Multienzymatic cascade reactions via enzyme complex by immobilization. *ACS Catalysis* 2019;9(5):4402-4425.
19. Abdallah W, Hong X, Banta S, Wheeldon I. Microenvironmental effects can masquerade as substrate channelling in cascade biocatalysis. *Current Opinion in Biotechnology* 2022;73:233-239.
20. Dubey NC, Tripathi BP. Nature inspired multienzyme immobilization: strategies and concepts. *ACS Applied Bio Materials* 2021;4(2):1077-1114.
21. Ngo TA, Nakata E, Saimura M, Morii T. Spatially organized enzymes drive cofactor-coupled cascade reactions. *Journal of the American Chemical Society* 2016;138(9):3012-3021.
22. Li C, Wang G-F, Wang Y, Creager-Allen R, Lutz EA, Scronce H, Slade KM, Ruf RA, Mehl RA, Pielak GJ. Protein 19F NMR in Escherichia coli. *Journal of the American Chemical Society* 2010;132(1):321-327.
23. Sharaf NG, Barnes CO, Charlton LM, Young GB, Pielak GJ. A bioreactor for in-cell protein NMR. *Journal of Magnetic Resonance* 2010;202(2):140-146.
24. Verkman AS. Diffusion in cells measured by fluorescence recovery after photobleaching. *Methods in enzymology*. Volume 360: Elsevier; 2003. p 635-648.
25. Dix JA, Verkman A. Crowding effects on diffusion in solutions and cells. *Annu. Rev. Biophys.* 2008;37:247-263.

26. Dey D, Marciano S, Nunes-Alves A, Kiss V, Wade RC, Schreiber G. Line-FRAP, a versatile method to measure diffusion rates in vitro and in vivo. *Journal of molecular biology* 2021;433(9):166898.
27. Tozzini V. Multiscale modeling of proteins. *Accounts of chemical research* 2010;43(2):220-230.
28. Trovato F, Nifosi R, Di Fenza A, Tozzini V. A minimalist model of protein diffusion and interactions: the green fluorescent protein within the cytoplasm. *Macromolecules* 2013;46(20):8311-8322.
29. Zhou H-X, Bates PA. Modeling protein association mechanisms and kinetics. *Current opinion in structural biology* 2013;23(6):887-893.
30. Gabdoulline RR, Wade RC. Brownian dynamics simulation of protein-protein diffusional encounter. *Methods* 1998;14(3):329-341.
31. Gabdoulline RR, Wade RC. Biomolecular diffusional association. *Current opinion in structural biology* 2002;12(2):204-213.
32. Mereghetti P, Kokh D, McCammon JA, Wade RC. Diffusion and association processes in biological systems: theory, computation and experiment. Volume 4: Springer; 2011. p 1-6.
33. Shen T, Wong CF, McCammon JA. Atomistic Brownian dynamics simulation of peptide phosphorylation. *Journal of the American Chemical Society* 2001;123(37):9107-9111.
34. Elcock AH, McCammon JA. Evidence for electrostatic channeling in a fusion protein of malate dehydrogenase and citrate synthase. *Biochemistry* 1996;35(39):12652-12658.
35. Huang YmM, Huber GA, Wang N, Minter SD, McCammon JA. Brownian dynamic study of an enzyme metabolon in the TCA cycle: Substrate kinetics and channeling. *Protein Science* 2018;27(2):463-471.
36. Ouporov IV, Knull HR, Huber A, Thomasson KA. Brownian dynamics simulations of aldolase binding glyceraldehyde 3-phosphate dehydrogenase and the possibility of substrate channeling. *Biophysical Journal* 2001;80(6):2527-2535.
37. Huber GA, McCammon JA. Brownian dynamics simulations of biological molecules. *Trends in chemistry* 2019;1(8):727-738.

38. Kekenés-Huskey PM, Scott CE, Atalay S. Quantifying the influence of the crowded cytoplasm on small molecule diffusion. *The Journal of Physical Chemistry B* 2016;120(33):8696-8706.
39. Bauler P, Huber G, Leyh T, McCammon JA. Channeling by proximity: the catalytic advantages of active site colocalization using Brownian dynamics. *The journal of physical chemistry letters* 2010;1(9):1332-1335.
40. Trovato F, Tozzini V. Diffusion within the cytoplasm: a mesoscale model of interacting macromolecules. *Biophysical journal* 2014;107(11):2579-2591.
41. Öztürk MA, Wade RC. Computation of FRAP recovery times for linker histone–chromatin binding on the basis of Brownian dynamics simulations. *Biochimica et Biophysica Acta (BBA)-General Subjects* 2020;1864(10):129653.
42. Wilner O, Weizmann Y, Gill R, Lioubashevski O, Freeman R, Willner I. *Nat. Nanotechnol* 2009;4:249-254.
43. You C, Myung S, Zhang YHP. Facilitated substrate channeling in a self-assembled trifunctional enzyme complex. *Angewandte Chemie International Edition* 2012;51(35):8787-8790.
44. Kuchler A, Yoshimoto M, Luginbühl S, Mavelli F, Walde P. Enzymatic reactions in confined environments. *Nature nanotechnology* 2016;11(5):409-420.
45. Zhang Y, Ge J, Liu Z. Enhanced activity of immobilized or chemically modified enzymes. *ACS catalysis* 2015;5(8):4503-4513.
46. Rajendran A, Nakata E, Nakano S, Morii T. Nucleic-Acid-Templated Enzyme Cascades. *ChemBioChem* 2017;18(8):696-716.
47. Idan O, Hess H. Origins of activity enhancement in enzyme cascades on scaffolds. *ACS nano* 2013;7(10):8658-8665.
48. Zhang Y, Tsitkov S, Hess H. Proximity does not contribute to activity enhancement in the glucose oxidase–horseradish peroxidase cascade. *Nature communications* 2016;7(1):1-9.
49. Chado G, Stoykovich M, Kaar J. *ACS Catal.* 2016, 6, 5161–5169. DOI.
50. Zhang Y, Hess H. Toward rational design of high-efficiency enzyme cascades. Volume 7: ACS Publications; 2017. p 6018-6027.
51. Kuzmak A, Carmali S, von Lieres E, Russell A, Kondrat S. Can enzyme proximity accelerate cascade reactions? *Sci Rep* 9: 455. 2019.

52. Sweetlove LJ, Fernie AR. The role of dynamic enzyme assemblies and substrate channelling in metabolic regulation. *Nature communications* 2018;9(1):1-12.
53. Wilmanns M, Priestle JP, Niermann T, Jansonius JN. Three-dimensional structure of the bifunctional enzyme phosphoribosylanthranilate isomerase: indoleglycerolphosphate synthase from *Escherichia coli* refined at 2.0 Å resolution. *Journal of molecular biology* 1992;223(2):477-507.
54. Henn-Sax M, Thoma R, Schmidt S, Hennig M, Kirschner K, Sterner R. Two ($\beta\alpha$) 8-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates. *Biochemistry* 2002;41(40):12032-12042.
55. Hennig M, Darimont B, Jansonius J, Kirschner K. The catalytic mechanism of indole-3-glycerol phosphate synthase: crystal structures of complexes of the enzyme from *Sulfolobus solfataricus* with substrate analogue, substrate, and product. *Journal of molecular biology* 2002;319(3):757-766.
56. Hommel U, Eberhard M, Kirschner K. Phosphoribosyl anthranilate isomerase catalyzes a reversible amadori reaction. *Biochemistry* 1995;34(16):5429-5439.
57. Eberhard M, Tsai-Pflugfelder M, Bolewska K, Hommel U, Kirschner K. Indoleglycerol phosphate synthase: phosphoribosyl anthranilate isomerase: Comparison of the bifunctional enzyme from *Escherichia coli* with engineered monofunctional domains. *Biochemistry* 1995;34(16):5419-5428.
58. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C, . ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of chemical theory and computation* 2015;11(8):3696-3713.
59. Wang B, Merz KM. A fast QM/MM (quantum mechanical/molecular mechanical) approach to calculate nuclear magnetic resonance chemical shifts for macromolecules. *Journal of chemical theory and computation* 2006;2(1):209-215.
60. Bosken YK, Cholko T, Lou YC, Wu KP, Chang CEA. Insights Into Dynamics of Inhibitor and Ubiquitin-Like Protein Binding in SARS-CoV-2 Papain-Like Protease. *Frontiers in Molecular Biosciences* 2020;7.
61. Wu KY, Doan D, Medrano M, Chang C-eA. Modeling structural interconversion in Alzheimers' amyloid beta peptide with classical and intrinsically disordered protein force fields. *Journal of Biomolecular Structure and Dynamics* 2021.
62. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* 1983;79(2):926-935.

63. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *The Journal of chemical physics* 1995;103(19):8577-8593.
64. Ryckaert J-P, Ciccotti G, Berendsen HJ. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics* 1977;23(3):327-341.
65. Nguyen T-N, You D-J, Kanaya E, Koga Y, Kanaya S. Crystal structure of metagenome-derived LC9-RNase H1 with atypical DEDN active site motif. *FEBS letters* 2013;587(9):1418-1423.
66. Ishihama Y, Schmidt T, Rappsilber J, Mann M, Hartl FU, Kerner MJ, Frishman D. Protein abundance profiling of the Escherichia coli cytosol. *BMC genomics* 2008;9(1):1-17.
67. Cholko T, Kaushik S, Wu KY, Montes R, Chang C-eA. GeomBD3: Brownian Dynamics Simulation Software for Biological and Engineered Systems. *Journal of Chemical Information and Modeling* 2022.
68. Elcock AH, Huber GA, McCammon JA. Electrostatic channeling of substrates between enzyme active sites: comparison of simulation and experiment. *Biochemistry* 1997;36(51):16049-16058.
69. Miles EW, Rhee S, Davies DR. The molecular basis of substrate channeling. *Journal of Biological Chemistry* 1999;274(18):12193-12196.
70. Lin J-L, Zhu J, Wheeldon I. Synthetic protein scaffolds for biosynthetic pathway colocalization on lipid droplet membranes. *ACS Synthetic Biology* 2017;6(8):1534-1544.
71. Roberts CC, Chang CEA. Modeling of enhanced catalysis in multienzyme nanostructures: Effect of molecular scaffolds, spatial organization, and concentration. *Journal of Chemical Theory and Computation* 2015;11(1):286-292.
72. Cao Y, Li X, Xiong J, Wang L, Yan L-T, Ge J. Investigating the origin of high efficiency in confined multienzyme catalysis. *Nanoscale* 2019;11(45):22108-22117.
73. Liu Y, Ogorzalek TL, Yang P, Schroeder MM, Marsh ENG, Chen Z. Molecular orientation of enzymes attached to surfaces through defined chemical linkages at the solid-liquid interface. *Journal of the American Chemical Society* 2013;135(34):12660-12669.
74. Xu S, Minter SD. Investigating the impact of multi-heme pyrroloquinoline quinone-aldehyde dehydrogenase orientation on direct bioelectrocatalysis via site specific enzyme immobilization. *Acs Catalysis* 2013;3(8):1756-1763.

75. Gorfe AA, Chia-en AC, Ivanov I, McCammon JA. Dynamics of the acetylcholinesterase tetramer. *Biophysical journal* 2008;94(4):1144-1154.
76. D'Amico RN, Bosken YK, O'Rourke KF, Murray AM, Admasu W, Chang C-EA, Boehr DD. Substitution of a surface-exposed residue involved in an allosteric network enhances tryptophan synthase function in cells. *Frontiers in molecular biosciences* 2021;8:679915.
77. Axe JM, Yezdimer EM, O'Rourke KF, Kerstetter NE, You W, Chang CEA, Boehr DD. Amino acid networks in a $(\beta/\alpha)_8$ barrel enzyme change during catalytic turnover. *Journal of the American Chemical Society* 2014;136(19):6818-6821.
78. Kang M, Roberts C, Cheng Y, Chang C-eA. Gating and intermolecular interactions in ligand-protein association: coarse-grained modeling of HIV-1 protease. *Journal of chemical theory and computation* 2011;7(10):3438-3446.
79. Ostrowska N, Feig M, Trylska J. Modeling crowded environment in molecular simulations. *Frontiers in Molecular Biosciences* 2019:86.
80. Ostrowska N, Feig M, Trylska J. Crowding affects structural dynamics and contributes to membrane association of the NS3/4A complex. *Biophysical Journal* 2021;120(17):3795-3806.
81. Skóra T, Vaghefikia F, Fitter Jr, Kondrat S. Macromolecular crowding: how shape and interactions affect diffusion. *The Journal of Physical Chemistry B* 2020;124(35):7537-7543.
82. Długosz M, Trylska J. Diffusion in crowded biological environments: applications of Brownian dynamics. *BMC biophysics* 2011;4(1):1-9.
83. Schreiber G, Haran G, Zhou H-X. Fundamental aspects of protein– protein association kinetics. *Chemical reviews* 2009;109(3):839-860.
84. Castellana M, Wilson MZ, Xu Y, Joshi P, Cristea IM, Rabinowitz JD, Gitai Z, Wingreen NS. Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nature biotechnology* 2014;32(10):1011-1018.
85. Ando T, Skolnick J. Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. *Proceedings of the National Academy of Sciences* 2010;107(43):18457-18462.
86. McGuffee SR, Elcock AH. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS computational biology* 2010;6(3):e1000694.

87. Dwyer JD, Bloomfield VA. Brownian dynamics simulations of probe and self-diffusion in concentrated protein and DNA solutions. *Biophysical journal* 1993;65(5):1810-1816.

Chapter 7 Conclusion and Future Work

7.1 Conclusion

Molecular recognition process between biomolecules has first been described as a “lock-and-key” mechanism involving interactions between rigid bodies and structural complementarity. Later an “induced-fit” mechanism was described, as binding of ligand to proteins require some degree of flexibility and their conformational rearrangement. Emerging evidences have indicated that more dynamics “hand in glove” mechanism for protein ligand binding, and conformational selection is considered a dominant mechanism of protein ligand binding.^{1,2} Many questions about the structural-functional-dynamics relationship remain to be addressed in many fields, such as enzyme, bioengineering, material sciences, and drug discovery. Increasing computational power and improved simulation architecture allow one to *in silico* study molecular recognition process and protein functional dynamics with microscopic perspective in biomolecules. In this dissertation project, we provided detailed analysis for protein ligand binding interactions and protein’s functional dynamics in many biological systems, including intrinsically disordered peptide $\alpha\beta 42$, ovarian cancer biomarker hCaV1.2, BRD4^{BD1} mediated by PROTAC in the ubiquitin degradation, and a fusion enzyme PRAI:IGPS in tryptophan biosynthesis pathway. Our findings elucidated important protein functional dynamics and demonstrate critical protein ligand binding interactions.

7.2 Future Work

The linker properties of the PROTAC are considered crucial factor in modulating the ternary complex formation,³ and residence time ($1/k_{\text{off}}$, the dissociation rate constant), in addition to binding affinity, correlates closely to target of interest degradation. Surface plasmon resonance (SPR), which has long been a core tool for measuring and quantifying protein binding affinities, association/dissociation rate constants, and complex formation,⁴⁻⁶ is an attractive technique for better investigating key parameters that impact formation of stable ternary complex. We further expand our *in silico* modeling approaches and collaborate with SPR team to identify and verify structural features of PROTACs that lead to stable ternary complex. We will study anaplastic lymphoma kinase (ALK), a membrane-bound tyrosine kinase that is an important drug target for a variety of cancers.⁷ Two groups of compounds: known ALK-CRBN PROTACs, TL13-12, TL13-112, MS4077, and MS4078 will be included in this work.^{8,9} Because the target protein ALK is a membrane-bound protein, we first examine and mimic the SPR setting by placing ALK onto a POPC membrane using the HIS-tag attachment and simulate for a few 500ns classical MD or accelerate MD runs. This step will provide enhanced sampling as there is not structure reported yet. MD trajectories will be analyzed by the K-Means clustering algorithm and ~3-5 highly populated distinct conformations will be selected.¹⁰ Each selected conformation will be modeled with the GeomBD3 program to find possible CRBN binding positions to the membrane-bound ALK.¹¹⁻¹³ The selected conformations are held rigid for BD simulations, so this step can quickly illustrate possible ternary orientations when the membrane is present. Next, we will dock those ALK-CRBN-PROTACs to ALK and

CRBN using programs MOE, or AutoDock. Top ~50 docked complexes will be run for a short cMD with an explicit water model before taking the structures for protein-protein docking. *Different from existing PROTAC sampling methods*¹⁴⁻¹⁷ that sample compound conformations without the presence of protein and waters, here we provide the protein and water environments to the linkers and use cMD to model the binary conformations more accurately. The binary complexes are re-ranked and clustered. Then the two proteins will be docked with the presence of a linker. Two pairs, ALK-ceritinib/CRBN-(pomalidomide+linker) and ALK-(ceritinib+linker)/CRBN-pomalidomide will be used in protein-protein docking with MOE, ROSETTA or HADDOCK.¹⁷⁻²⁰ Because the docking steps do not include a membrane, the docked ternary complexes will be analyzed with our BD results, and only the overall orientation like those in BD will be kept. Notably, our strategies consider the linker space in protein-protein docking and can efficiently eliminate unrealistic ternary structures in cell-like membrane environment. We will then superimpose the docked ternary conformations from the two pairs using the linker region to re-compose the PROTAC in the complex: ALK-(ceritinib+linker+pomalidomide, aka PROTAC)-CRBN. A MD-based scoring approach with MM/GBSA-type energy calculations²¹ for the ternary complexes will be used to estimate the stability of the ternary complex, which will be compared directly with the binding affinity measured by SPR.^{22,23} Then, ternary complex with good binding affinity will be used to superimpose to the degradation machinery (CRBN-based CRL4A ligase complex) reported by this recent work.²⁴ This step helps to obtain suitable orientation and identify key lysine residues in the ALK, which directly links to the degradation. This multi-level simulation protocol allows

sampling of the ternary conformations with emphasis on the structure orientation on the membrane environment, modeling of the E2/E3 Ub conjugation cascade that is proximate the lysine(s) in the target protein and evaluation of their binding stability efficiently and accurately. Features critical to stabilize the complex, iteratively informed and validated by experiments, will guide the PROTAC design.

7.3 References

1. Vogt AD, Di Cera E. Conformational selection is a dominant mechanism of ligand binding. *Biochemistry* 2013;52(34):5723-5729.
2. Boehr DD, Nussinov R, Wright PE. The role of dynamic conformational ensembles in biomolecular recognition. *Nature chemical biology* 2009;5(11):789-796.
3. Bemis TA, La Clair JJ, Burkart MD. Unraveling the Role of Linker Design in Proteolysis Targeting Chimeras: Miniperspective. *Journal of Medicinal Chemistry* 2021;64(12):8042-8052.
4. Homola J, Yee SS, Gauglitz G. Surface plasmon resonance sensors: review. *SENSORS AND ACTUATORS B-CHEMICAL* 1999;54(1-2):3-15.
5. Morrill PR, Millington RB, Lowe CR. Imaging surface plasmon resonance system for screening affinity ligands. *JOURNAL OF CHROMATOGRAPHY B-ANALYTICAL TECHNOLOGIES IN THE BIOMEDICAL AND LIFE SCIENCES* 2003;793(2):229-251.
6. Scarano S, Mascini M, Turner APF, Minunni M. Surface plasmon resonance imaging for affinity-based biosensors. *Biosensors & Bioelectronics* 2010;25(5):957-966.
7. Hallberg B, Palmer R. The role of the ALK receptor in cancer biology. *Annals of Oncology* 2016;27:iii4-iii15.
8. Xie S, Sun Y, Liu Y, Li X, Li X, Zhong W, Zhan F, Zhu J, Yao H, Yang D-H. Development of Alectinib-Based PROTACs as Novel Potent Degraders of Anaplastic Lymphoma Kinase (ALK). *Journal of medicinal chemistry* 2021;64(13):9120-9140.
9. Zhang C, Han X-R, Yang X, Jiang B, Liu J, Xiong Y, Jin J. Proteolysis targeting chimeras (PROTACs) of anaplastic lymphoma kinase (ALK). *European journal of medicinal chemistry* 2018;151:304-314.
10. Roe DR, Cheatham III TE. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation* 2013;9(7):3084-3095.
11. Cholko T, Chang CEA. Modeling Effects of Surface Properties and Probe Density for Nanoscale Biosensor Design: A Case Study of DNA Hybridization near Surfaces. *Journal of Physical Chemistry B* 2021;125(7):1746-1754.

12. Cholko T, Barnum J, Chang CEA. Amyloid-beta (A beta 42) Peptide Aggregation Rate and Mechanism on Surfaces with Widely Varied Properties: Insights from Brownian Dynamics Simulations. *Journal of Physical Chemistry B* 2020;124(27):5549-5558.
13. Fischer ES, Boehm K, Lydeard JR, Yang H, Stadler MB, Cavadini S, Nagel J, Serluca F, Acker V, Lingaraju GM and others. Structure of the DDBI-CRBN E3 ubiquitin ligase in complex with thalidomide. *NATURE* 2014;512(7512):49-53.
14. Drummond ML, Williams CI. In silico modeling of PROTAC-mediated ternary complexes: validation and application. *Journal of chemical information and modeling* 2019;59(4):1634-1644.
15. Drummond ML, Henry A, Li H, Williams CI. Improved Accuracy for Modeling PROTAC-Mediated Ternary Complex Formation and Targeted Protein Degradation via New In Silico Methodologies. *JOURNAL OF CHEMICAL INFORMATION AND MODELING* 2020;60(10):5234-5254.
16. Zaidman D, Prilusky J, London N. PROsettaC: Rosetta based modeling of PROTAC mediated ternary complexes. *Journal of chemical information and modeling* 2020;60(10):4894-4903.
17. Bai N, Miller SA, Andrianov GV, Yates M, Kirubakaran P, Karanickolas J. Rationalizing PROTAC-Mediated Ternary Complex Formation Using Rosetta. *Journal of Chemical Information and Modeling* 2021;61(3):1368-1382.
18. van Zundert GCP, Bonvin A. Modeling Protein-Protein Complexes Using the HADDOCK Webserver "Modeling Protein Complexes with HADDOCK". *Protein Structure Prediction, 3rd Edition* 2014;1137:163-179.
19. Porter KA, Desta I, Kozakov D, Vajda S. What method to use for protein-protein docking? *Current Opinion in Structural Biology* 2019;55:1-7.
20. Scafuri N, Soler MA, Spitaleri A, Rocchia W. Enhanced Molecular Dynamics Method to Efficiently Increase the Discrimination Capability of Computational Protein-Protein Docking. *Journal of Chemical Theory and Computation* 2021;17(11):7271-7280.
21. Wang C, Greene DA, Xiao L, Qi R, Luo R. Recent Developments and Applications of the MMPBSA Method. *Frontiers in Molecular Biosciences* 2018;4.
22. Liao J, Nie X, Unarta IC, Ericksen SS, Tang W. In Silico Modeling and Scoring of PROTAC-Mediated Ternary Complex Poses. *Journal of Medicinal Chemistry* 2022;65(8):6116-6132.

23. Li W, Zhang J, Guo L, Wang Q. Importance of Three-Body Problems and Protein–Protein Interactions in Proteolysis-Targeting Chimera Modeling: Insights from Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling* 2022;62(3):523-532.
24. Bai N, Riching KM, Makaju A, Wu H, Acker TM, Ou S-C, Zhang Y, Shen X, Bulloch D, Rui H. Modeling the CRL4A ligase complex to predict target protein ubiquitination induced by cereblon-recruiting PROTACs. *Journal of Biological Chemistry* 2022:101653.