

UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Modeling of Protein Flexibility and Inter-Molecular Interactions: Applications to Computer-Aided Drug Design and Discovery

Permalink

<https://escholarship.org/uc/item/29g595sq>

Author

Ai, Rizi

Publication Date

2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Modeling of Protein Flexibility and Inter-Molecular Interactions: Applications to
Computer-Aided Drug Design and Discovery

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

Doctor of Philosophy

in

Genetics, Genomics and Bioinformatics

by

Rizi Ai

September 2012

Dissertation Committee:
Dr. Chia-en A. Chang, Chairperson
Dr. Thomas Girke
Dr. Zhenbiao Yang

Copyright by
Rizi Ai
2012

The Dissertation of Rizi Ai is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to express my deepest gratitude to my PhD advisor, Dr. Chia-en Chang, for her academic guidance, endless understanding, and constant encouragement for my five years' PhD study. I have been amazingly fortunate to have her as my advisor, who inspired me not only with her deep insight but also the passion for research. Her valuable guidance encouraged me to explore the unknown and shaped me to think independently, which is overwhelmingly important for my future careers. I hope that one day that I would become as an excellent scientist as Chia-en.

I greatly appreciated my dearest friends, the past and current members in Chang group, for their help and support: Dr. Myungshim Kang, Dr. Sepideh Yaghmaei, Dr. M. Qaiser Fatmi, Steven Ahrendt, Christopher Roberts, Mindy Huang, Zhiye Tang, So-Ning Lee, and Wai M. Thant. I also want to thank my collaborators, Dr. Michael Marsella, Dr. Huanbin Zhou, Dr. Wenbo Ma, Dr. Leonard Mueller, Dr. Michael Dunn, Dr. Dimitri Niks and Dr. Wenwan Zhong for their contributions to my work.

I am also thankful to my committee members, Dr. Thomas Girke and Dr. Zhenbiao Yang, for five years' guidance and valuable comments on my PhD qualify exam and annual report meeting.

Most importantly, none of this would have been possible without the love, support and strength of my family for all the years. I would like to express my heart-felt gratitude to my family, especially to my fiancé, Henry.

Last but not least, I want to take the opportunity to extend my greatest appreciation to University of California, Riverside and Graduate Program in Genetics, Genomics and Bioinformatics for providing me the opportunity to carry out my research and enjoy the wonderful time in Riverside.

Published materials are used for Chapter 2 and 3, which include:

Ai, R.; Fatmi, M. Q.; 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.

Ai, R.; Chang, C-E. A., Ligand-Specific Homology Modeling of Human Cannabinoid (CB1) Receptor. *Journal of Molecular Graphics and Modelling*. *In press*.

<http://dx.doi.org/10.1016/j.jmgm.2012.05.002>

Dedication

To my parents, Yan Ai and Yulin Liu, and my loving fiancé, Henry Song.

ABSTRACT OF THE DISSERTATION

Modeling of Protein Flexibility and Inter-Molecular Interactions: Applications to
Computer-Aided Drug Design and Discovery

by

Rizi Ai

Doctor of Philosophy, Graduate Program in Genetics, Genomics and Bioinformatics
University of California, Riverside, September 2012
Dr. Chia-en A. Chang, Chairperson

Computer-aided drug design (CADD) represents computational methods and resources that are used to facilitate the design and discovery of new therapeutic solutions. It has been applied at almost all stages in the drug discovery pipeline. CADD has the advantage of reducing the cost and time for the expensive and extensive laboratory-based experiment. In CADD, protein flexibility has long been recognized as a complicating factor. Proteins are in constant motions in their physiological environment and their conformational dynamics play important roles in various biological functions and regulating ligand binding. My research focuses on modeling of protein flexibility and inter-molecular interactions and their applications to CADD. Detailed projects include method development for analyzing protein conformational changes using Perl and applications of various computing tools to

study protein dynamics and protein-ligand interactions. In addition, ligand-specific molecular model was constructed on cannabinoid CB1 receptor, a G-protein coupled receptor, with protein flexibility considered. To discover novel antimicrobial agents against tryptophan synthase, another project was carried out using structure-based virtual screening method. Recently, nanoparticles have been applied to therapeutic use. To promote safer implementation of nanotechnology and reduce nanotoxicity for drug delivery, dynamics and interactions between human serum albumin and nanoparticles were studied.

Table of Contents

| | |
|-------------------------------------------------------------------------------------------------------------------|------|
| Acknowledgements | iv |
| Dedication | vi |
| Abstract | vii |
| List of Tables | xiii |
| List of Figures | xv |
| | |
| Chapter 1 Overview of Computer-Aided Drug Design and Discovery..... | 1 |
| 1.1. Drug discovery process and computer-aided drug design..... | 1 |
| 1.2. Current challenges in CADD | 3 |
| 1.3. Focus of this work | 4 |
| 1.4. References..... | 5 |
| | |
| Chapter 2 T-Analyst: a program for efficient analysis of protein conformational changes by torsion angles..... | 7 |
| 2.1. Abstract..... | 7 |
| 2.2. Introduction..... | 8 |
| 2.3. Methods..... | 9 |
| 2.3.1. Data manipulation and output | 9 |
| 2.3.2. Torsion angle correction..... | 10 |
| 2.3.3. Configurational entropy calculation and dihedral correlation..... | 10 |

| | |
|------------------------------------------------------------------------|----|
| 2.3.4. Clustering protein conformations..... | 13 |
| 2.3.5. Molecular dynamics simulations..... | 14 |
| 2.4. Analysis of MD simulations..... | 15 |
| 2.4.1. Analysis of protein conformations..... | 15 |
| 2.4.2. Clustering protein conformations..... | 18 |
| 2.4.3. Configurational entropy..... | 19 |
| 2.4.4. Dynamical cross-correlation between dihedral angles..... | 23 |
| 2.5. Conclusions..... | 26 |
| 2.6. References..... | 26 |
| Chapter 3 Ligand-Specific Homology Modeling of Human Cannabinoid (CB1) | |
| Receptor..... | 32 |
| 3.1. Abstract..... | 32 |
| 3.2. Introduction..... | 33 |
| 3.3. Methods..... | 36 |
| 3.3.1. Homology modeling..... | 36 |
| 3.3.2. Model refinement..... | 37 |
| 3.3.3. Modeling the binding site by four types of ligands..... | 40 |
| 3.3.4. Model verification..... | 43 |
| 3.3.5. Energy calculation..... | 44 |
| 3.3.6. Flexibility analysis of ACEA..... | 45 |
| 3.4. Results and Discussion..... | 45 |

| | |
|------------------------------------------------------------------------------------------|----|
| 3.4.1. CB1 receptor models..... | 45 |
| 3.4.1.1. CB1 model trained by HU-210..... | 49 |
| 3.4.1.2. CB1 model trained by ACEA..... | 50 |
| 3.4.1.2.1. Flexibility study of ACEA..... | 53 |
| 3.4.1.3. CB1 model trained by WIN55212-2..... | 56 |
| 3.4.1.4. CB1 model trained by SR141716A..... | 57 |
| 3.4.1.5. Model comparisons..... | 60 |
| 3.4.2. CB1 receptor model validation..... | 61 |
| 3.5. Conclusions..... | 64 |
| 3.6. References..... | 65 |
| Chapter 4 Discovery of Antimicrobial Agents for α -Subunit of Tryptophan Synthase | |
| by Molecular Docking Studies and In Vitro Assays..... | 75 |
| 4.1. Abstract..... | 75 |
| 4.2. Introduction..... | 76 |
| 4.3. Methods..... | 78 |
| 4.3.1. Protein conformations selection..... | 78 |
| 4.3.2. Docking based virtual screening..... | 79 |
| 4.3.3. In-silico ADME prediction..... | 81 |
| 4.3.4. In vitro test..... | 82 |
| 4.4. Results..... | 82 |
| 4.5. Conclusions..... | 87 |

| | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 4.6. References..... | 88 |
| Chapter 5 Molecular Dynamics Studies of Interactions between Poly(Acrylic Acid) -Coated Fe ₃ O ₄ Nanoparticles and Human Serum Albumin..... | |
| 5.1. Abstract..... | 91 |
| 5.2. Introduction..... | 92 |
| 5.3. Methods..... | 96 |
| 5.3.1. Preparing PAA coating compounds..... | 96 |
| 5.3.2. MD simulations..... | 97 |
| 5.4. Preliminary Results..... | 99 |
| 5.4.1. The binding poses of PAA coatings..... | 99 |
| 5.4.2. A and B-series PAA coatings in HSA binding site..... | 99 |
| 5.4.3. Protein flexibility analysis..... | 104 |
| 5.5. Conclusions..... | 107 |
| 5.6. References..... | 108 |
| Chapter 6 Future Work of Dynamics Studies of Interactions between Poly (Acrylic Acid)-Coated Fe ₃ O ₄ Nanoparticles and Human Serum Albumin..... | |
| 6.1 Further MD simulations..... | 113 |
| 6.2. Brownian dynamics simulations..... | 114 |
| 6.3. Molecular docking..... | 115 |
| 6.4. References..... | 115 |

List of Tables

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Table 3.1 | 38 |
| Sequence Alignment of Helix Regions of Multiple GPCR Sequences. | |
| Table 3.2 | 40 |
| Conformations of Helix 2 and 3 in CB1 Model Modified by Protein Threading. | |
| Table 3.3 | 47 |
| Average Interaction Energy of Known CB1 Binders, Compounds Structurally Similar to Binders and Random Compounds in the Four CB1 Models. (kcal/mol). | |
| Table 3.4 | 52 |
| Residues within 6 Å of Ligands in Ligand-CB1 Complex. | |
| Table 3.5 | 58 |
| Key Ligand-Aromatic Clustering in CB1 Models Optimized by WIN55212-2 and SR141716A. | |
| Table 3.6 | 58 |
| Key Aromatic Clustering in CB1 Models Optimized by WIN55212-2. | |

| | |
|-----------------------------------------------------------------------------------------------------|-----|
| Table 3.7 | 60 |
| Key Aromatic Clustering in CB1 Models Optimized by SR141716A. | |
| Table 3.8 | 63 |
| Interaction Energy of Cross Docking Results (kcal/mol). | |
| Table 4.1 | 84 |
| Weighted Average Interaction Energy and In-Silico ADME Prediction of Top Ranked Chemical Compounds. | |
| Table 4.2 | 85 |
| E.coli Cell Densities with 28 Compounds Measured by OD ₆₀₀ . | |
| Table 5.1 | 101 |
| Binding Modes and Docked Interaction Energy of A and B-series PAA NPs with Three Chain Lengths. | |

List of Figures

Figure 2.1.....11

Torsion angle distribution of side-chain χ_2 of Ile153 of TRPS. Torsion angle: (A) before angle correction and (B) after angle correction. Histogram of torsion angle distribution output by T-Analyst: (C) before angle correction and (D) after angle correction.

Figure 2.2.....17

Analysis of the hinge regions of HIV-1 protease. (A) An open conformation of HIV-1 protease obtained from the MD simulation. Flaps: residues 43–58, yellow; elbows: residues 37–42; hinge: Gly 40, red. (B) RMSF of $C\alpha$ atoms. (C) Distance between two flap tips: residues Ile 50 and Ile 50'. (D) Ψ angle distribution of Gly 40'; values within the dashed lines, between 720-880 ps, correspond to open flap conformations shown in (C)

Figure 2.3.....20

Clustering conformations of TRPS. (A-D) Ligand-free state, (E-H) ligand-bound state. Torsion angle distribution of side-chain χ_3 of Glu 49 in (A) ligand-free and (E) ligand-bound state. Torsion angle distribution of side-chain χ_2 of Ile 153 in (B) ligand-free and (F) ligand-bound state. (C) and (G): RMSD of Glu 49 and Ile 153 after

structure alignment. (D) and (H): Conformational groups computed by T-Analyst over the course of MD simulations. Each small diamond represents one frame.

Figure 2.4.....21

Snapshots of surrounding residues of Glu 49 in (A) ligand-bound and (B) ligand-free state. Grey dotted lines indicate stable hydrogen-bonds in (A) and green dotted lines indicate weak hydrogen-bonds in (B). Red dashed lines indicate the distance between Tyr173:OH and Glu49:CD. The circle labeled with IGP in (B) indicates the position of the ligand binding pocket.

Figure 2.5.....22

Entropy distributions in the binding site of TRPS. (A) Entropy distribution of Ψ torsion angles of residues within 8Å of ligand IGP and (B) side-chain torsion angles that directly interacting with IGP.

Figure 2.6.....25

Structure anatomy of HIV-1 protease and correlation maps of one subunit of HIV-1 protease in BAT coordinates and Cartesian coordinates. (A) Structure anatomy of HIV-1 protease. Color indicates distinct regions. Flaps: residues 43–58, red; flap tips: residues 49–52, yellow; flap elbows: residues 37–42, magenta; cantilevers: residues 59–75, green; fulcrums: residues 10–23, orange; loop area near flaps: residues 76–84,

grey; and interleaved β -strand motif forming the dimer interface: residues 1–4 and 96–99, blue/cyan. (B) Φ and Ψ torsion angle correlation map generated by T-Analyst using BAT coordinates. Yellow lines and boxes indicate correlations between I: fulcrum; II: flap elbow, III: flap with flap elbow, IV: flap, V: flap tip, VI: flap elbow with cantilevel, VII: flap with cantilevel, VIII: cantilevel, IX: fulcrum with loop, and X: loop. (C) Correlation map generated by the use of the backbone C α and N atoms and the Cartesian coordinates with the Bio3D package.

Figure 3.1.....35

Molecular structures of cannabinoid ligands in (1, 2) classical cannabinoids, (3, 4) non-classical cannabinoids, (5) endogenous cannabinoids, (7, 8) aminoalkylindoles and (9, 10) diarylpyrazoles. (1) (-)- Δ^9 -THC, (2) HU-210, (3) CP55940, (4) CP47497, (5) anadamide, (6) ACEA, (7) WIN55212-2, (8) JWH-015, (9) SR141716A, (10) AM281.

Figure 3.2.....41

CB1 receptor hybrid model built by template-based fragment assembly. (A1-3) Helix 1 built by template (A1) AA_{2A}R, (*A2) β_2 AR and (A3) individual helix 1 fragment; (B1-3) Helix 2 built by template (B1) AA_{2A}R; (B2) β_2 AR and (*B3) individual helix 2 fragment; (C1-3) Helix 3 built by template (*C1) AA_{2A}R, (C2) β_2 AR and (C3) individual helix 3 fragment; (D1-3) Helix 4 built by template (*D1) AA_{2A}R, (*D2) β_2 AR and (D3) individual helix 4 fragment; (E1-3) Helix 5 built by template (*E1) AA_{2A}R, (*E2) β_2 AR

and (E3) individual helix 5 fragment; (F1-3) Helix 6 built by template (*F1) AA_{2A}R, (F2) β₂AR and (F3) individual helix 6 fragment; (G1-3) Helix 7 built by template (*G1) AA_{2A}R, (G2) β₂AR and (G3) individual helix 7 fragment; (H) assembled CB1 receptor hybrid model. *, structures assembled into the hybrid model. For helix 4 and 5, part of the helix is from structure built by AA_{2A}R and the other part is by β₂AR.

Figure 3.3.....43

RMSD of residues that were within 12Å of ACEA during MD simulations of ACEA-CB1 receptor complex.

Figure 3.4.....46

CB1 binders docked to models built by (A1-3) HU-210, (B1-5) ACEA, (C1-2) WIN55212-2 and (D1-5) SR14176A

Figure 3.5.....47

Substructures of (A1-4) HU-210, (B1-3) ACEA, (C) WIN55212-2 and (D) SR141716A used for substructure search in the ZINC database.

Figure 3.6.....48

Molecular structures of 25 random non-binders from NCI diversity set II. (1) NSC1614, (2) NSC1620, (3) NSC5069, (4) NSC14304, (5) NSC29073, (6) NSC35930, (7) NSC41805,

(8) NSC47881, (9) NSC54509, (10) NSC55957, (11) NSC57608, (12) NSC67436, (13) NSC70980, (14) NSC76484, (15) NSC80997, (16) NSC91398, (17) NSC94600, (18) NSC166846, (19) NSC171303, (20) NSC197049, (21) NSC227186, (22) NSC288686, (23) NSC308848, (24) NSC401077, (25) NSC515893.

Figure 3.7.....49

Alignment of the four ligand-CB1 models. (A) Side view. (B) Top view. Seven transmembrane helices are indicated (H1-7). Colors represent models trained by HU-210 (red), ACEA (green), WIN55212-2 (blue) and SR141716A (cyan).

Figure 3.8.....51

Binding sites of CB1 receptor models trained by (A) HU-210, (B) ACEA, (C) WIN55212-2 and (D) SR141716A. Ligands are in ball-and-stick depiction. Residues that are directly contact with ligands are in opaque stick depiction. Residues that are within 6 Å of the ligands are in transparent stick depiction (see Table 3.4 for details).

Figure 3.9.....55

Entropy and distributions of ACEA dihedrals. (A) 17 dihedrals of ACEA analyzed by T-Analyst. (B) Configurational entropy of each ACEA dihedral angle. Blue and green lines show ACEA entropy of the free ligand state in a 10 Å × 10 Å × 10 Å cubic TIP3P

water box and in vacuum, respectively; red line shows the entropy in the ACEA–CB1 bound state. (C-E) Distributions of dihedral ϕ and ACEA conformations in (C) ACEA–CB1 bound state, (D) free state in water box and (E) free state in vacuum.

Figure 3.10.....61

Residue (A) F7.35(379) and (B) F3.37(381) in the binding sites of the four models. Red: HU-210 model; green: ACEA model; blue: WIN55212-2 model and cyan: SR141716A model.

Figure 4.1.....80

Side chain conformations of key residues in the binding site of TRPS.

Figure 4.2.....80

Side chain dihedral distributions of key residues in the binding site of TRPS.

Figure 4.3.....83

Docking pose of compound 1 in the binding site of TRPS.

Figure 4.4.....86

Cell Density of *E. coli* after 24 hrs incubation at 37°C

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 5.1 | 95 |
| Summary of ligand binding capacity and subdomains of HSA using PDB structure 1E7E. Long-chain fatty acids are depicted in VDW representation using VMD 1.8.7. The subdomains are depicted in blue (IA), yellow (IB), red (IIA), green (IIB), purple (IIIA) and cyan (IIIB). | |
| Figure 5.2 | 96 |
| Illustration of structures of NPs and two types of coatings. A. A series PAA-NPs. B: B series PAA-NPs. | |
| Figure 5.3 | 97 |
| Structures of (A1-3) A and (B1-3) B series PAA coatings for docking; with (A1, B1) two, (A2, B2) three or (A3, B3) four hydroxyl groups in the long PAA chains. | |
| Figure 5.4 | 100 |
| The center of drug binding site 2 of HSA for molecular docking | |
| Figure 5.5 | 103 |
| Interactions of (A) A-series and (B) B-series PAA coatings with drug binding site 2 of HSA during 10 ns MD simulations. | |

Figure 5.6.....103

Interactions and hydrogen bond networks in the binding site of HSA upon (A) A-series PAA and (B) B-series PAA binding.

Figure 5.7.....104

Illustration of flexible and rigid regions of HSA. Flexible region: entropy of each backbone Φ and ψ angle that is larger than 0 kcal/mol, in red; rigid region: entropy of each backbone Φ and ψ angles that is less than -0.3 kcal/mol in both HSA free and bound states, in blue.

Figure 5.8.....105

Entropy changes of side chain torsions within 6 Å of the PAA binding site. Blue dotted line indicates the entropy changes between A-series PAA bound state and free protein state. Red dotted line indicates the entropy changes between B-series PAA bound state and free protein state.

Figure 5.9.....106

Side chain torsion distributions of (A) Arg 410 and (B) Tyr 411 in (A1, B1) free state, (A2, B2) A-series PAA-NP bound state and (A3, B3) B-series PAA-NP bound state

Figure 6.1.....113

Chirality of A-series PAA coating.

Chapter 1

Overview of Computer-Aided Drug Design and Discovery

1.1. Drug discovery process and computer-aided drug design

Drug design and discovery is an expensive process due to the high research and development costs and extensive clinical testing. Typically, it takes about 10-15 years to develop one new medicine from the time that it is discovered to treat the patients. The average cost to research and develop a successful drug is estimated to be more than \$800 million.¹

Computer-aided drug design (CADD) represents computational methods and resources that are used to facilitate the design and discovery of new therapeutic solutions.² It plays a critical role in screening new potential drug candidates in medicinal chemistry and has been applied at almost all stages in the drug discovery pipeline. Using CADD can avoid performing expensive and extensive laboratory-based experiment and therefore reduce the cost and time significantly. For example, CADD allows the design and modification of ligands to optimize the binding on targets and a prediction of their binding affinities before carrying out experiment on

large database of compounds.^{3,4} With constant improvements in both computer power and algorithms, CADD become more and more powerful.

Methods utilized in CADD can be classified into two major categories: structure-based drug design and ligand-based drug design.⁵ Structure-based drug design depends on the three dimensional structures of receptors, often obtained by X-ray crystallography or NMR. Molecular docking is one of the most common methods used in structure-based drug design. The docking methods allow the quick prediction of the binding modes and binding affinities between ligands and receptor. In addition, virtual screening based on molecular docking is widely used to predict potential drugs from large dataset of compounds before carrying out expensive and time-consuming experiment.⁶ Without experimentally available three dimensional receptor structures, computational methods such as homology modeling, fold recognition or AB initio protein modeling can be used to computationally construct the protein structures based on the amino acid sequences.⁷ Moreover, ligand-based drug design, which relies on the biological and physicochemical properties of known molecules that can bind to the target, is another popular approach when the target structure is not available or cannot be predicted by computational methods. Quantitative structure-activity relationship (QSAR) and pharmacophore modeling typical methods for ligand-based drug design.⁸

1.2. Current challenges in CADD

Considering receptor flexibility has long been recognized as a complicating factor in CADD.⁹ Molecular receptors, such as proteins, are in constant motions in their physiological environment. The conformational dynamics of proteins play important roles in various biological functions and regulating ligand binding.^{6,10} During ligand binding, protein undergoes various conformational changes, from side chain rotations to backbone shift, towards the most adapted conformation to accommodate ligands. However, traditional docking programs only rely on a single, rigid receptor conformation which can be problematic to accurately predict the binding activity.¹¹ Docking programs allow the searching for the translational and rotational degrees of freedom of the ligand within the receptor binding site; but mostly only for ligand itself. Although some docking programs may allow certain protein side chain flexibilities, most of them still works better with rigid proteins.¹²⁻¹⁴ Computational power has increased dramatically over the last decade. However, fully allowing protein flexibility is still prohibited due to the high computational intensity.

Since most docking software prefers proteins to be rigid to avoid intensive computational effort, a promising strategy is to dock ligands into protein conformation ensembles obtained from MD simulations^{9,15}. Molecular dynamics (MD) simulations provide powerful tools for the exploration of the conformational energy landscape accessible to protein molecules because multiple conformations

are difficult to probe experimentally¹⁶⁻¹⁸ . Atomistic computer simulations of macromolecular receptors and their associated ligands play many roles in drug discovery, such as allosteric binding sites identification, virtual screening methodologies enhancement, and binding energies prediction.¹¹ To consider the protein flexibility, using multiple conformations could avoid single biased binding site for docking screening.

1.3. Focus of this work

To explore new methods to overcome the limitations of current CADD obstacles and facilitate the future drug design and discovery, my PhD research mainly focuses on modeling of protein flexibility and inter-molecular interactions and their applications to CADD. Detailed projects include method development for analyzing protein conformational changes using Perl, applications of various computing tools to study protein dynamics and protein-ligand interactions; ligand-specific molecular modeling on G-protein coupled receptors with protein flexibility considered; virtual drug screening against tryptophan synthase for antimicrobial agents; and molecular dynamics studies on interactions between protein and nano-particles.

1.4. References

1. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. *Journal of Health Economics*. 2003;22(2).
2. Song CM, Lim SJ, Tong JC. Recent advances in computer-aided drug design. *Briefings in Bioinformatics*. 2009;10(5).
3. Andricopulo AD, Salum LB, Abraham DJ. Structure-Based Drug Design Strategies in Medicinal Chemistry. *Current Topics in Medicinal Chemistry*. 2009;9(9):771-790.
4. Kolb P, Ferreira RS, Irwin JJ, Shoichet BK. Docking and chemoinformatic screens for new ligands and targets. *Current Opinion in Biotechnology*. 2009;20(4):429-436.
5. Wilson GL, Lill MA. Integrating structure-based and ligand-based approaches for computational drug design. *Future Medicinal Chemistry*. 2011;3(6).
6. Cheng T, Li Q, Zhou Z, Wang Y, Bryant SH. Structure-Based Virtual Screening for Drug Discovery: a Problem-Centric Review. *Aaps Journal*. 2012;14(1).
7. Huang H-J, Yu HW, Chen C-Y, et al. Current developments of computer-aided drug design. *Journal of the Taiwan Institute of Chemical Engineers*. 2010;41(6).
8. Acharya C, Coop A, Polli JE, MacKerell AD, Jr. Recent Advances in Ligand-Based Drug Design: Relevance and Utility of the Conformationally Sampled Pharmacophore Approach. *Current Computer-Aided Drug Design*. 2011;7(1).
9. Carlson HA, McCammon JA. Accommodating protein flexibility in computational drug design. *Molecular Pharmacology*. 2000;57(2):213-218.

10. Teague SJ. Implications of protein flexibility for drug discovery. *Nature Reviews Drug Discovery*. 2003;2(7).
11. Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. *Bmc Biology*. 2011;9.
12. May A, Sieker F, Zacharias M. How to efficiently include receptor flexibility during computational docking. *Current Computer-Aided Drug Design*. 2008;4(2):143-153.
13. Wong CF. Flexible ligand-flexible protein docking in protein kinase systems. *Biochimica Et Biophysica Acta-Proteins and Proteomics*. 2008;1784(1):244-251.
14. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. Vol. 18; 2008.
15. Amaro RE, Baron R, McCammon JA. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. Vol. 22; 2008.
16. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*. 2002;9(9):646-652.
17. Gumbart J, Wang Y, Aksimentiev A, Tajkhorshid E, Schulten K. Molecular dynamics simulations of proteins in lipid bilayers. Vol. 15; 2005.
18. Deng YQ, Roux B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. Vol. 113; 2009.

Chapter 2

T-Analyst: A Program for Efficient Analysis of Protein Conformational Changes by Torsion Angles

2.1. Abstract

T-Analyst is a user-friendly computer program for analyzing trajectories from molecular modeling. Instead of using Cartesian coordinates for protein conformational analysis, T-Analyst is based on internal bond-angle-torsion coordinates in which internal torsion angle movements, such as side chain rotations, can be easily detected. The program computes entropy and automatically detects and corrects angle periodicity to produce accurate rotameric states of dihedrals. It also clusters multiple conformations and detects dihedral rotations that contribute hinge-like motions. Correlated motions between selected dihedrals can also be observed from the correlation map. T-Analyst focuses on showing changes in protein flexibility between different states and selecting representative protein conformations for molecular docking studies. The program is provided with instructions and full source code in Perl.

2.2. Introduction

The conformational dynamics of proteins play important roles in their functions and regulating ligand binding. A fundamental appreciation of how proteins work requires study of conformations and dynamics, as well as changes between states of protein motions, such as folded/unfolded and ligand-bound/-free states. For example, protein allosteric effects may be related by either or both conformational and dynamical changes.¹⁻³ Molecular dynamics (MD) simulations provide powerful tools for the exploration of the conformational energy landscape accessible to protein molecules because multiple conformations are difficult to probe experimentally.⁴⁻⁶ Moreover, recent computer-aided drug discovery studies have focused on protein flexibility in molecular docking processes.⁷⁻⁹ Since most docking software prefers proteins to be rigid to avoid intensive computational effort, a promising strategy is to dock ligands into protein conformation ensembles obtained from MD simulations.^{10,11} Several programs provide general tools or special modules for analysis of MD results and clustering conformations, but most are based on Cartesian coordinates.¹²⁻¹⁷ T-Analyst uses internal bond-angle-torsion (BAT) coordinates, which are efficient in capturing side-chain rotamers and most low-frequency motions.¹⁸⁻²⁰ Our program provides useful tool to analyze MD trajectories. For example, users can easily view proper rotameric states of dihedrals from the output files instead of plotting and correcting them manually. This program

allows for efficient analysis of MD simulations to study protein flexibility and extract structural information for virtual screening.

2.3. Methods

2.3.1. Data manipulation and output

T-Analyst reads NAMD, Amber or CHARMM trajectory files and Amber topology file. The CHARMM-type topology files can be converted to Amber topology files easily with the freely available CHAMBER program.²¹ The program implements `amber2accent` to transform Cartesian coordinates to BAT coordinates.²² To eliminate repeats, only heavy-atom side-chain torsion angles and ω , ϕ and ψ angles in backbones are considered. Users can choose angles and residues for analysis, and a dihedral distribution and its rotations during a simulation are output as `.agr`-format files, which can be viewed with Xmgr/Grace. Moreover, results generated by T-Analyst can be visualized by freely available packages such as VMD, Xmgr/Grace or R.^{12,23,24} Users can output a series of files, such as distributions of all or selected torsion angles in a protein, or sorted and/or unsorted standard deviations and entropy. T-Analyst also groups different conformations based on rotameric states of residues of interest and outputs coordinates of grouped conformations into different trajectory files. The program also computes pairwise cross-correlation coefficients for all pairwise dihedral angles that users selected. By examining the output correlation map, dihedrals that correlate with each other can be observed.

2.3.2. Torsion angle correction

To correctly reflect rotameric states and capture the angle periodicity, T-Analyst detects the angle population in margins, such as $\pm 180^\circ$ and $360^\circ/0^\circ$, depending on how users output the angles. If discontinuity in a dihedral distribution is detected (see Fig. 2.1A and C), the output angle range will be corrected; thus, energy wells can be illustrated and defined properly (Fig. 2.1B and D). For example, Fig. 2.1C illustrates a rotameric state of side-chain Ile 153 of ligand-bound tryptophan synthase (TRPS), whereby one energy well is split into two wells near -180° and $+180^\circ$. T-Analyst automatically detects the peak discontinuity at each edge of -180° and $+180^\circ$ and determines the range of each discontinued peak. Later, the whole population will be divided into left set (-180° to 0°) and right set (0° to $+180^\circ$) and be counted in each set. T-Analyst angle correction is then applied by moving the discontinued peak on the set with smaller population to the set with larger population. For example, in Fig. 2.1C, the set from 0° to $+180^\circ$ has the smaller population, so the peak near the $+180^\circ$ margin was shifted by -360° and merged with the other margin at -180° .

2.3.3. Configurational entropy calculation and dihedral correlation

T-Analyst calculates configurational entropy, S_{conf} , for each torsional degree of freedom by the Gibbs entropy formula: $TS_{\text{conf}(i)} = -RT \sum P_i \ln(P_i)$, where P_i is the probability distribution of angle i , R is the gas constant and T is the absolute

temperature. T-Analyst calls Xmgr/Grace to generate histogram for each degree of freedom. The bin size for each P_i is 0.5° for ω angle, 1° for ϕ and ψ angles and 5° for

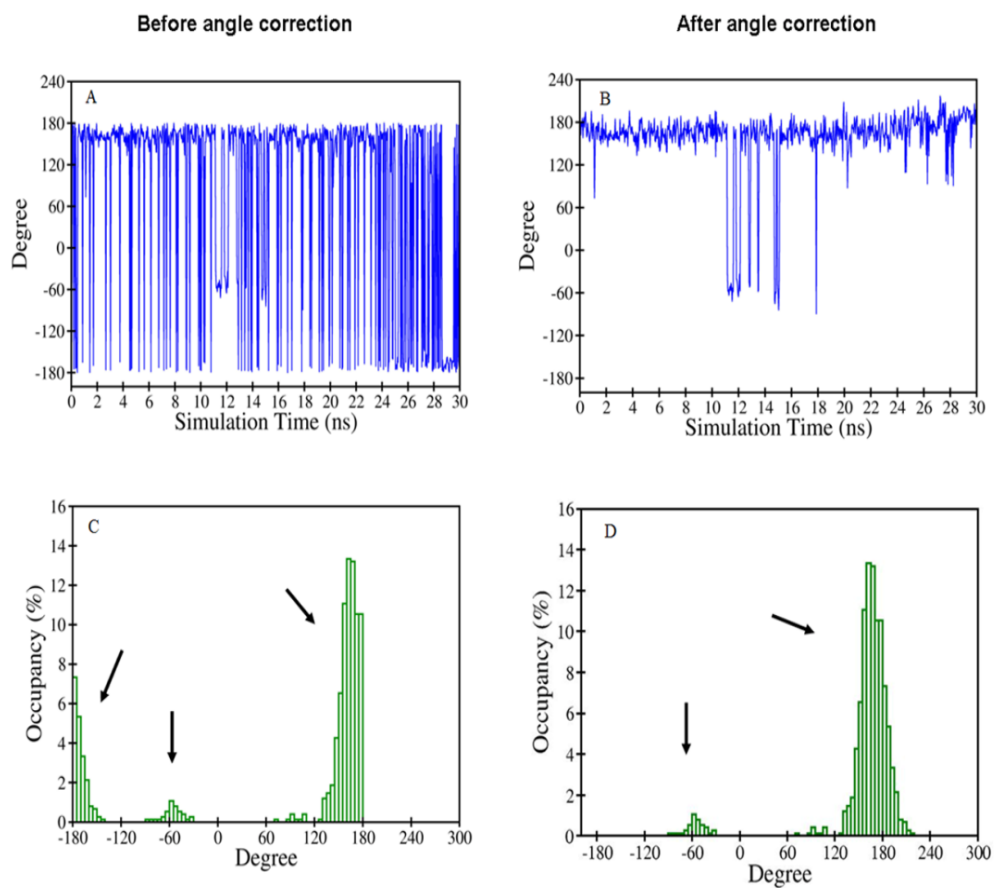


Figure 2.1 Torsion angle distribution of side-chain χ_2 of Ile153 of TRPS. Torsion angle: (A) before angle correction and (B) after angle correction. Histogram of torsion angle distribution output by T-Analyst: (C) before angle correction and (D) after angle correction.

side-chain dihedrals. The value of $TS_{\text{conf}(i)}$ has unit of kcal/mol, which allows for direct comparison with energy calculations. Summing $TS_{\text{conf}(i)}$ provides a quick approximation of the entropic contribution of a system, although coupling between torsions is ignored.²⁵

T-Analyst also computes the quasi-harmonic approximation (QH) from BAT coordinates. The covariance matrix \mathbf{C} can be computed, with the probability distribution functions approximated by a multidimensional Gaussian distribution function.²⁶ The configurational entropy from QH is computed by $TS_{\text{QH}} = 1/2 nRT + 1/2RT \ln[(2\pi)^n \det(\mathbf{C})]$, where n is the number of torsions. Although QH assumes that the probability distribution function is Gaussian, which is accurate for torsions that have only one rotameric state, TS_{QH} provides an upper bound limit for the configurational entropy.^{25,27} Moreover, the off-diagonal elements of the covariance matrix indicate the degree of significance of the coupling between the given torsions. Entropy computed from only the diagonal elements of the covariance matrix, $TS_{\text{QH_diag}}$, is also computed. If TS_{QH} equals $TS_{\text{QH_diag}}$, then there is no coupling among these torsions.

The extent to which pairs of dihedrals are correlated with one another can be assessed by examining the magnitude of their cross-correlation coefficients. T-Analyst computes a correlation matrix of dihedrals and calls levelplot function in R

to plot a correlation map. Users can select dihedral angles, e.g. backbone phi and psi angles of selected residues, to plot a correlation matrix. Typical characteristics of a correlation map include a line of strong cross-correlation along the diagonal (where matrix element $i = j$), and off-diagonal cross-correlations. The high diagonal values are set to 1.00. Off-diagonal correlations can be either positive or negative, and non-zero values may indicate potentially interesting correlations between two close proximity or non-contiguous regions of a protein system.

2.3.4. Clustering protein conformations

Our program clusters protein conformations on the basis of user-selected rotameric states of residues. Although RMSD-based clustering methods are mostly applied to group conformations with significant difference, small fluctuations are challenging to detect with classical RMSD-based clustering methods. This module is particularly useful for choosing representative conformations based on side-chain rotations of key residues. Users can input specific torsions with rotameric states of interest and the range of each torsion to run T-Cluster, the second part of T-Analyst. T-Analyst will provide all the combinations of groups for further analysis. The program is sensitive to dihedral rotations and can efficiently group user picked backbone or side-chain dihedrals into separate trajectory files. A report file is also generated to record information about each group.

2.3.5. *Molecular dynamics simulations*

Molecular dynamics simulations on ligand-free and ligand-bound TRPS were performed using the NAMD package.²³ Standard simulation procedures (e.g. ⁽²⁸⁾) were followed using the Amber 10 package and ff03 Amber force field and general Amber force field.^{6,29} Initial coordinates of TRPS were taken from PDB code 2J9X and 3CEP^{30,31}. Briefly, after preparation of the system by sequential steps of energy minimization and equilibration, the 30ns production runs were carried out at 298 K and 1 atm. The systems were solvated by a 12Å TIP3P water box. Snapshots of the atomic coordinates were recorded every 1 ps. As T-Analyst does not require too many frames to run the analysis, snapshots were saved every 20 ps for T-Analyst and 1500 frames were used. Molecular dynamics simulations on HIV-1 protease were initiated from crystallographic coordinates with a semi-open flap conformation (PDB code 1HHP).³² Amber ff99 force field was used for the protein. Aqueous solvation was modeled implicitly by using the Generalized Born approach³³ and temperature was maintained at 298 K by using Langevin dynamics. Standard simulation procedures (e.g. ⁽³⁴⁾) were followed with the Amber 9 package. Since the free protease predominantly populates the semi-open conformation, we took a 1.5 ns MD simulation and saved it as one 1500 frame trajectory which had one flap open state for our analysis.

2.4. Analysis of MD simulations

2.4.1. Analysis of protein conformations

Analysis of small backbone fluctuations, as well as conformational changes, involves investigating loop or side-chain motions. T-Analyst adopts torsion angle analysis, which allows for accurate expression of bond rotations. For example, Fig. 2.1A shows a rotameric state of side-chain Ile 153 of ligand-bound TRPS. Standard deviations of this torsion before and after angle correction are 136.4° and 49.4° , respectively. Large differences in standard deviations usually indicate changes in rotameric states. Of note, proper angle correction is necessary for computing accurate rotamers and their standard deviations.

With careful superposition, flexible regions in proteins can be revealed by computing root mean square deviation and fluctuation (RMSD and RMSF, respectively) in Cartesian coordinates.³⁵ Sometimes one or a few residues may serve as “hinges,” which mainly control movements of a flexible region. While the RMSD/RMSF method may not be sensitive enough to detect hinges, T-Analyst can illustrate torsion angles in hinges, which are associated with protein motions. User can compare multiple torsion distributions with protein motion to detect the hinge region. Fig. 2.2 A shows MD simulations of HIV-1 protease with open flaps, and the distance between flap tips is in Fig. 2.2 C. Highly flexible regions such as flaps and elbows are shown in the RMSF plot (Fig. 2.2 B) but determining whether any residues behave as hinges is not

straightforward.³⁴ Data from T-Analyst imply that the flap β -hairpins move relatively rigidly, but some residues of the elbow, such as Gly 40, act as hinges (Fig. 2.2D).

One common method to express protein flexibility is by showing their rotameric states computed from corrected torsion angle distribution. When comparing the rotameric states for proteins between different states, e.g. ligand-bound/-free states, folded/unfolded states, users can have valuable information regarding the protein conformational or flexibility changes between different states. Different ligand mechanisms, such as induced fit or conformational selection (population shift) can be studied.^{1,36}

Here we use TRPS, a protein with strong allosteric regulation of substrate binding, as an example. In both ligand-free and ligand-bound TRPS, the major peak of Glu 49 remains for both states to keep its favorite rotameric state (Fig. 2.3 A and E), which is often assumed.³⁷ However, instead of dropping the minor peak, the peak changed while the substrate binds. We found that the change may be due to different hydrogen-bonds formations. In the ligand-bound state, Glu 49 form stable hydrogen-bonds with ligand 3-indole-D-glycerol-3'-phosphate (IGP), Tyr 173 and Val 23, and a residue near Glu 49, Tyr 175, also forms a hydrogen-bond with IGP. These interactions stabilize the local environment for the ligand which is necessary for catalysis. In the ligand-free state, since there is no ligand to interact with the protein

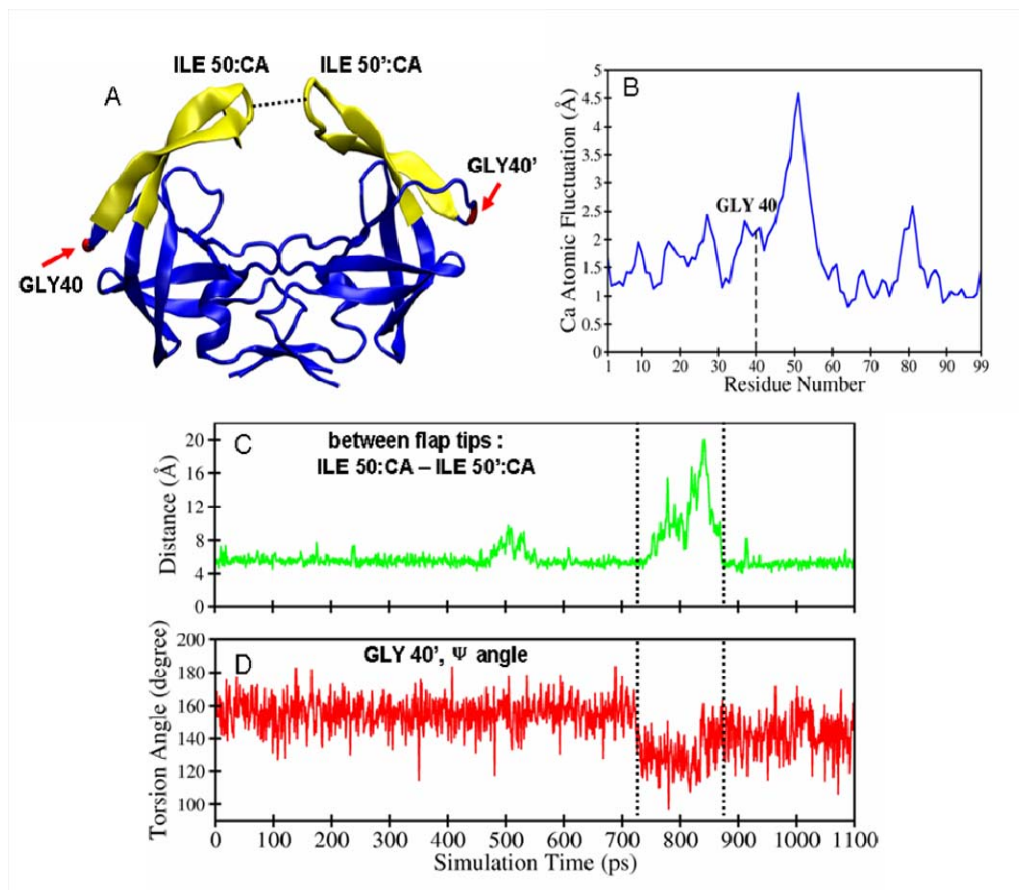


Figure 2.2 Analysis of the hinge regions of HIV-1 protease. (A) An open conformation of HIV-1 protease obtained from the MD simulation. Flaps: residues 43–58, yellow; elbows: residues 37–42; hinge: Gly 40, red. (B) RMSF of C α atoms. (C) Distance between two flap tips: residues Ile 50 and Ile 50'. (D) Ψ angle distribution of Gly 40'; values within the dashed lines, between 720-880 ps, correspond to open flap conformations shown in (C)

in the binding site, more free space is allowed for side-chains to move around, though less stable hydrogen-bonds between side-chains can still formed. Without the ligand, the hydrogen-bond between Glu 49 and Tyr 173 is absent and Tyr 173

flips away from the binding site. Further detailed analysis shows that the distance between Tyr173:OH and Glu49:CD changes from 3.6 ± 0.16 Å in ligand-bound state (red dashed lines in Fig. 2.4 A) to 17.5 ± 0.63 Å in ligand-free state (red dashed lines in Fig. 2.4 B). As a result, instead of completely diminishing one rotameric state of Glu 49, the interaction induces a new rotameric state to alter the local environment to accommodate the ligand. Similarly, the major peak of residue Ile 153 in Fig. 2.3B and F remains in the same position in both states, but it is more populated in the ligand-bound state. Upon IGP binding, the methyl group of Ile 153 interacts with the non-polar region of the ligand and stays in one peak. Therefore, the population of the rotameric states is shifted.

2.4.2. Clustering protein conformations

Docking potential ligands to target protein is one of the key steps in drug design and discovery process. During ligand-protein docking process, the position of side-chains in binding site can affect ligand-protein interaction significantly and directly, especially when using rigid protein conformations. Here we use the program to group conformations and then select representative conformations for performing docking studies to the α -subunit of TRPS. TRPS is a potential antibiotic target, and two torsion angles, Glu 49 and Ile 153, are known to directly involve in ligand-protein binding (see Fig. 2.3). Fig. 2.3A and B show the distribution of side-chain torsion angles in both residues in ligand-free state which is output by T-Analyst after angle

correction. Two populations are shown in each of the torsion angle distributions. T-Analyst can group the trajectories into four groups which are characterized by – group 1 (Glu 49-a, Ile 153-a), group 2 (Glu 49-a, Ile 153-b), group 3 (Glu 49-b, Ile 153-a), group 4 (Glu 49-b, Ile 153-b). Fig. 2.3 C and G show the RMSD distribution of Glu 49 and Ile 153 with simulation time. Fig. 2.3 D and H are the group distributions for the two torsion angles. Notably, although different protein conformations are usually clustered based on computed RMSD, the value is less sensitive to small scale conformational changes, such as side-chain rotations in a protein's active site. As illustrated in Fig. 2.3 C and G, there is no clue from RMSD that may be used for grouping. In contrast, T-Analyst also provides information regarding jumping between groups during a course of MD simulation, see Fig. 2.3 D and H.

2.4.3. Configurational entropy

Protein allosteric effects or post-translational modifications such as phosphorylation do not always involve substantial conformational changes. Recent experiments confirmed that in some cases, visual inspection of the active/inactive states may not reveal differences in the shape of the ligand binding site, but changes in protein dynamics.³ Therefore, the magnitude of configuration entropy computed from dihedral degrees of freedom provides a direct way to examine protein dynamics.

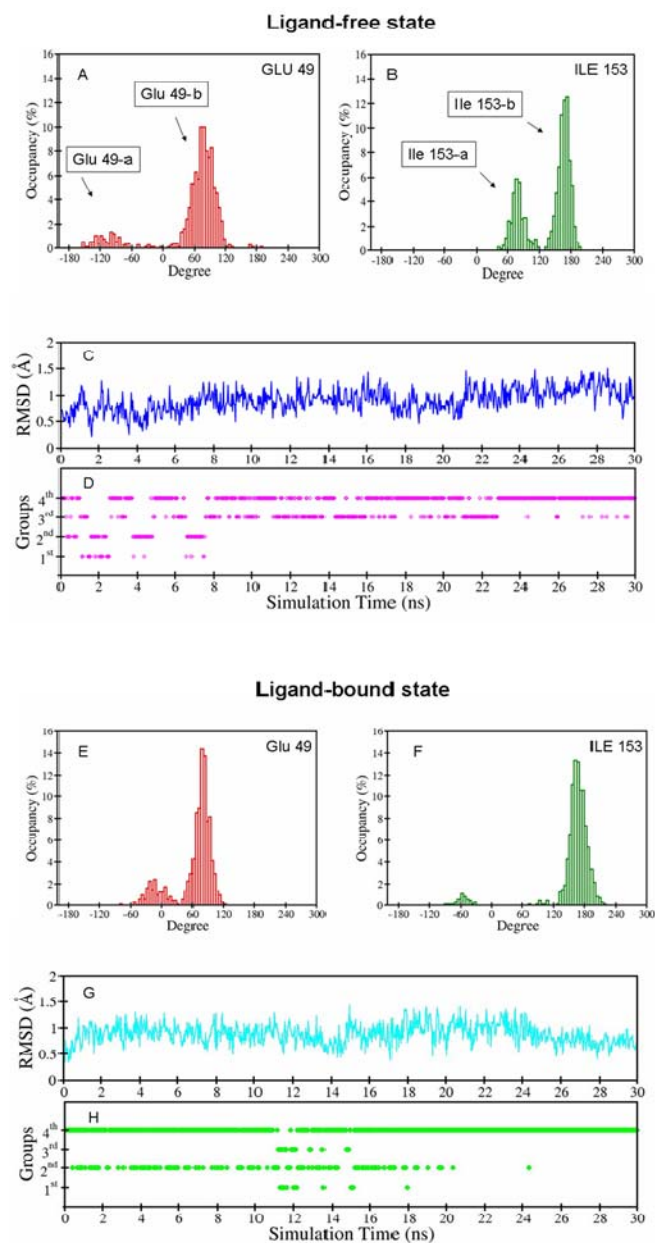


Figure 2.3 Clustering conformations of TRPS. (A-D) Ligand-free state, (E-H) ligand-bound state. Torsion angle distribution of side-chain χ_3 of Glu 49 in (A) ligand-free and (E) ligand-bound state. Torsion angle distribution of side-chain χ_2 of Ile 153 in (B) ligand-free and (F) ligand-bound state. (C) and (G): RMSD of Glu 49 and Ile 153 after structure alignment. (D) and (H): Conformational groups computed by T-Analyst over the course of MD simulations. Each small diamond represents one frame.

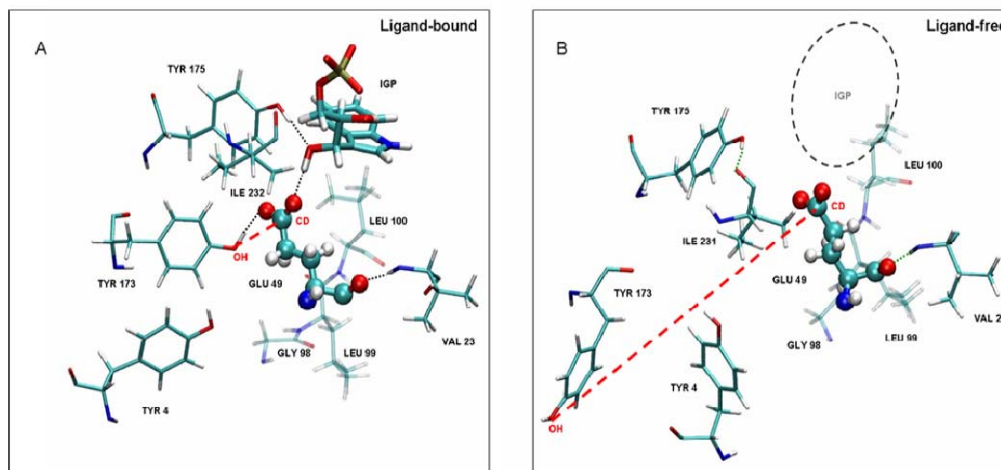


Figure 2.4 Snapshots of surrounding residues of Glu 49 in (A) ligand-bound and (B) ligand-free state. Grey dotted lines indicate stable hydrogen-bonds in (A) and green dotted lines indicate weak hydrogen-bonds in (B). Red dashed lines indicate the distance between Tyr173:OH and Glu49:CD. The circle labeled with IGP in (B) indicates the position of the ligand binding pocket.

The total entropy of a molecule in solution can be separated into two parts: solvent entropy associated with water motion and configurational entropy (S_{conf}) associated with molecular motion. The latter can be used to represent protein flexibility. Therefore, quantifying the configurational entropy, especially changes, could help explain important biochemical processes such as protein folding and ligand-protein binding. Fig. 2.5 shows S_{conf} calculated from selected backbone and side-chain dihedrals in the binding site of TRPS. T-Analyst suggested that the flexibility of a loop region changes significantly when the ligand binds into the pocket.

The calculations can provide quantitative information regarding entropy gain or loss. For example, in Fig. 2.5B, the side-chain entropies of Tyr 175 and Thr 183 drop notably, around 0.7 and 0.6 kcal/mol, respectively. In the ligand-bound state, Tyr 175 forms a stable hydrogen-bond with ligand IGP which stabilized the phenol ring, and Thr 183 also forms hydrogen-bonds with Asp 60 and Gly 61. In ligand-free state, both side-chains move more freely as there is no stable hydrogen-bond formed.

Although in most cases, one can observe entropy decrease upon ligand binding, there may be some exceptions which are of particular interest and worth further analysis. The entropy changes computed by T-Analyst provide guidance for users to

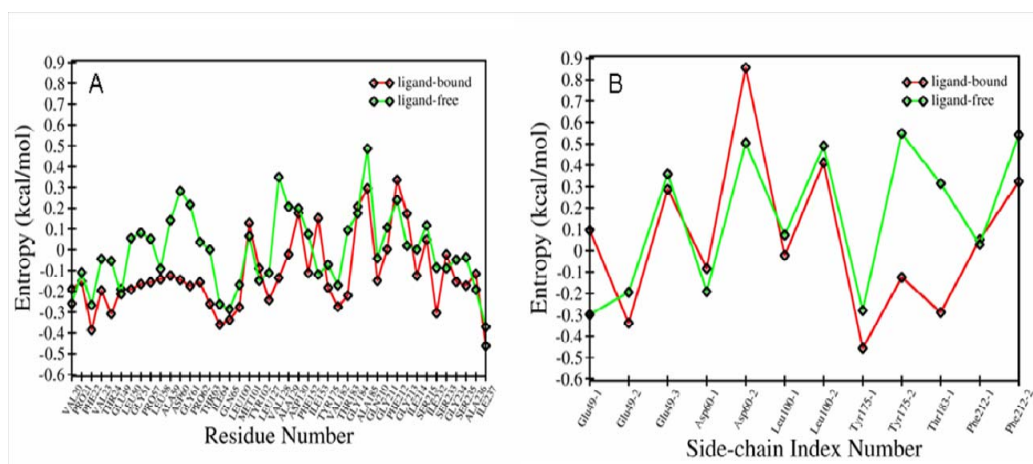


Figure 2.5 Entropy distributions in the binding site of TRPS. (A) Entropy distribution of Ψ torsion angles of residues within 8 Å of ligand IGP and (B) side-chain torsion angles that directly interacting with IGP.

pick up regions of a protein to do detail dynamic analysis. For example, T-Analyst showed an entropy gain in the second side-chain of Asp 60 in the presence of ligand, though not largely. Based on the information, we carried out further investigation near Asp 60, and found that 2 oxygen atoms in carboxyl group can form hydrogen-bonds with Tyr 102 and Thr 183 in both ligand-bound and ligand-free states. Interestingly, in the ligand-bound state, the 2 oxygen atoms of Asp 60 can flip very often but retain two stable hydrogen-bonds with Tyr 102 and Thr 183 alternatively. The presence of IGP provides a more hydrophilic environment around the carboxyl group of Asp 60; thus, the oxygen atoms can flip more freely. As a result, the local entropy increases without losing hydrogen-bonding and reducing electrostatic attraction. In contrast, when IGP is absent, the local environment around Asp 60 is mainly hydrophobic. Therefore, Asp 60 forms hydrogen-bond with Thr 183, the other hydrophilic residue, and the residues become less flexible.

2.4.4. Dynamical cross-correlation between dihedral angles

T-Analyst also performs cross-correlation analysis of a trajectory. The resulting cross-correlation map allows the identification of the correlated and anti-correlated motions involved in an entire protein or user selected dihedrals. Similar to calculating standard deviations, dihedral angles used to generate a correlation plot also need to run angle correction, or discontinuities in margins ($\pm 180^\circ$ or $360^\circ/0^\circ$) can cause errors when computing their cross-correlation coefficients.

Here we use HIV-1 protease as a test case, which continues to be one of the primary targets of anti-AIDS drug discovery. Understanding the dynamics of free HIV-1 protease has profound implications for designing new therapeutic agents, such as allosteric inhibitors. Fig. 2.6A shows the structure anatomy of the protein. Yellow boxes in the correlation map, computed by the backbone ϕ and ψ angles with T-Analyst (see Fig. 2.6B), show the highly correlated regions within the HIV protease monomer. For example, boxes I, IV, VIII and X indicate flexible and concerted motions within fulcrum, flap, cantilever and loop regions. Other boxes highlighted in Fig. 2.6 B indicate the correlations between flap elbow, flap tip, cantilever, fulcrum and loop regions. It has been suggested that motions of the flap tip and elbow are highly correlated, which is illustrated in our plot as well³⁴. Interestingly, correlations between the flap elbow and cantilever are identified (Box VI), and the motions of flap tip and cantilever are also correlated (Box VII), though not quite strongly. We compared our plot to another correlation map generated by the use of the backbone $C\alpha$ and N atoms and the Cartesian coordinates with the Bio3D package (Fig. 2.6 C)³⁸. Both plots suggested similar correlations. However, because of the nature of Cartesian coordinates, patterns of correlations are harder to distinguish in Fig. 2.6 C.

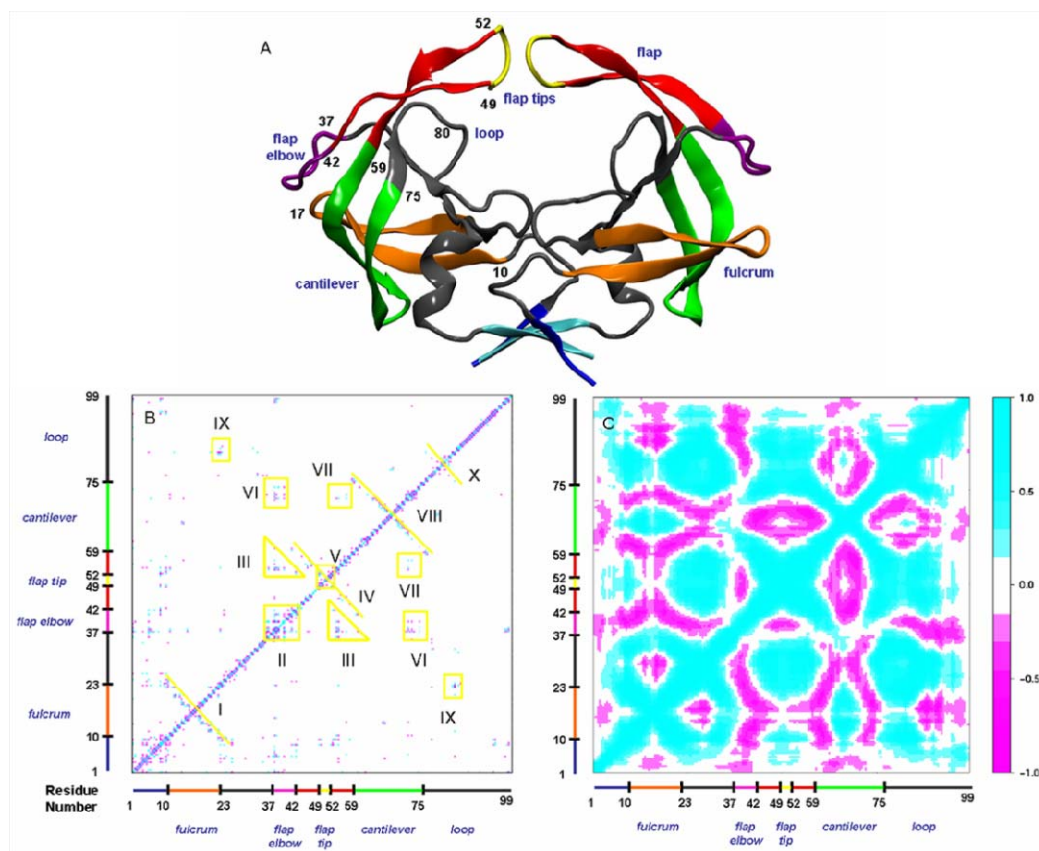


Figure 2.6 Structure anatomy of HIV-1 protease and correlation maps of one subunit of HIV-1 protease in BAT coordinates and Cartesian coordinates. (A) Structure anatomy of HIV-1 protease. Color indicates distinct regions. Flaps: residues 43–58, red; flap tips: residues 49–52, yellow; flap elbows: residues 37–42, magenta; cantilevers: residues 59–75, green; fulcrums: residues 10–23, orange; loop area near flaps: residues 76–84, grey; and interleaved β -strand motif forming the dimer interface: residues 1–4 and 96–99, blue/cyan. (B) Φ and Ψ torsion angle correlation map generated by T-Analyst using BAT coordinates. Yellow lines and boxes indicate correlations between I: fulcrum; II: flap elbow, III: flap with flap elbow, IV: flap, V: flap tip, VI: flap elbow with cantilevel, VII: flap with cantilevel, VIII: cantilevel, IX: fulcrum with loop, and X: loop. (C) Correlation map generated by the use of the backbone $C\alpha$ and N atoms and the Cartesian coordinates with the Bio3D package.

2.5. Conclusions

MD simulations provide invaluable conformational and dynamical landmarks useful for designing new experiments and for theoretical studies. The current analysis method we describe, T-Analyst, can help in examining protein motion, to identify structural and dynamic features, reveal changes of flexibility in different states, and group conformations based on dihedral rotamers. Analyzing the growing MD data may be the most time-consuming step in simulation studies, and our program facilitates this work. The program can be freely downloaded from <http://research.chem.ucr.edu/groups/chang/tools.htm>.

2.6. References

1. Cui Q, Karplus M. Allostery and cooperativity revisited. *Protein Science*. 2008;17(8):1295-1307.
2. Kern D, Zuiderweg ERP. The role of dynamics in allosteric regulation. *Current Opinion in Structural Biology*. 2003;13(6):748-757.
3. Tsai CJ, del Sol A, Nussinov R. Allostery: Absence of a change in shape does not imply that allostery is not at play. *Journal of Molecular Biology*. 2008;378(1):1-11.

4. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*. 2002;9(9):646-652.
5. Gumbart J, Wang Y, Aksimentiev A, Tajkhorshid E, Schulten K. Molecular dynamics simulations of proteins in lipid bilayers. *Current Opinion in Structural Biology*. 2005;15(4):423-431.
6. Deng YQ, Roux B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *Journal of Physical Chemistry B*. 2009;113(8):2234-2246.
7. May A, Sieker F, Zacharias M. How to efficiently include receptor flexibility during computational docking. *Current Computer-Aided Drug Design*. 2008;4(2):143-153.
8. Wong CF. Flexible ligand-flexible protein docking in protein kinase systems. *Biochimica Et Biophysica Acta-Proteins and Proteomics*. 2008;1784(1):244-251.
9. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Current Opinion in Structural Biology*. 2008;18(2):178-184.
10. Carlson HA, McCammon JA. Accommodating protein flexibility in computational drug design. *Molecular Pharmacology*. 2000;57(2):213-218.
11. Amaro RE, Baron R, McCammon JA. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *Journal of Computer-Aided Molecular Design*. 2008;22(9):693-705.

12. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. 1996;14(1):33-&.
13. Brooks BR, Brooks CL, Mackerell AD, et al. CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry*. 2009;30(10):1545-1614.
14. Christen M, Hunenberger PH, Bakowies D, et al. The GROMOS software for biomolecular simulation: GROMOS05. *Journal of Computational Chemistry*. 2005;26(16):1719-1751.
15. Van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*. 2005;26(16):1701-1718.
16. Case DA, Cheatham TE, Darden T, et al. The Amber biomolecular simulation programs. *Journal of Computational Chemistry*. 2005;26(16):1668-1688.
17. MacKerell AD, Bashford D, Bellott M, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. Vol. 102; 1998.
18. Chang CE, Gilson MK. Tork: Conformational analysis method for molecules and complexes. *Journal of Computational Chemistry*. 2003;24(16):1987-1998.
19. McClendon CL, Friedland G, Mobley DL, Amirkhani H, Jacobson MP. Quantifying Correlations Between Allosteric Sites in Thermodynamic Ensembles. *Journal of Chemical Theory and Computation*. 2009;5(9):2486-2502.
20. Killian BJ, Kravitz JY, Somani S, Dasgupta P, Pang YP, Gilson MK. Configurational Entropy in Protein-Peptide Binding: Computational Study of Tsg101 Ubiquitin E2

Variant Domain with an HIV-Derived PTAP Nonapeptide. *Journal of Molecular Biology*. 2009;389(2):315-335.

21. Crowley MF, Williamson MJ, Walker RC. CHAMBER: Comprehensive Support for CHARMM Force Fields Within the AMBER Software. *International Journal of Quantum Chemistry*. 2009;109(15):3767-3772.

22. Shao JY, 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;3:2312-2334.

23. Phillips JC, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD. Vol. 26; 2005.

24. Team RDC. R: A Language and Environment for Statistical Computing. In: Computing RFFS ed. Vienna, Austria; 2009.

25. Chang C-EA, McLaughlin WA, Baron R, Wang W, McCammon JA. Entropic contributions and the influence of the hydrophobic environment in promiscuous protein-protein association. *Proc Natl Acad Sci U S A*. 2008;105(21):7456-7461.

26. Levy RM, Karplus M, Kushick J, Perahia D. EVALUATION OF THE CONFIGURATIONAL ENTROPY FOR PROTEINS - APPLICATION TO MOLECULAR-DYNAMICS SIMULATIONS OF AN ALPHA-HELIX. *Macromolecules*. 1984;17(7):1370-1374.

27. Meirovitch H, Chelvaraja S, White RP. Methods for Calculating the Entropy and Free Energy and their Application to Problems Involving Protein Flexibility and Ligand Binding. *Current Protein & Peptide Science*. 2009;10(3):229-243.
28. Fatmi MQ, Ai R, Chang CEA. Synergistic Regulation and Ligand-Induced Conformational Changes of Tryptophan Synthase. *Biochemistry*. 2009;48(41):9921-9931.
29. Wang JM, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. Vol. 25; 2006.
30. Ngo H, Kimmich N, Harris R, et al. Allosteric regulation of substrate channeling in tryptophan synthase: Modulation of the L-Serine reaction in stage I of the ss-reaction by alpha-site ligands. Vol. 46; 2007.
31. Barends TRM, Domratcheva T, Kulik V, et al. Structure and mechanistic implications of a tryptophan synthase quinonoid intermediate. *Chembiochem*. 2008;9(7):1024-1028.
32. Berman HM, Battistuz T, Bhat TN, et al. The Protein Data Bank. Vol. 58; 2002.
33. Onufriev A, Bashford D, Case DA. Modification of the generalized Born model suitable for macromolecules. *Journal of Physical Chemistry B*. 2000;104(15):3712-3720.
34. Hornak V, Okur A, Rizzo RC, Simmerling C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proceedings of the National Academy of Sciences of the United States of America*. 2006;103(4):915-920.

35. Damm KL, Carlson HA. Gaussian-weighted RMSD superposition of proteins: A structural comparison for flexible proteins and predicted protein structures. *Biophysical Journal*. 2006;90(12):4558-4573.
36. Okazaki KI, Takada S. Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms. Vol. 105; 2008.
37. Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. Vol. 97; 2000.
38. Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. Bio3d: an R package for the comparative analysis of protein structures. Vol. 22; 2006.

Chapter 3

Ligand-Specific Homology Modeling of Human Cannabinoid (CB1) Receptor

3.1. Abstract

Cannabinoid (CB1) receptor is a therapeutic drug target, and its structure and conformational changes after ligand binding are of great interest. To study the protein conformations in ligand bound state and assist in drug discovery, CB1 receptor homology models are needed for computer-based ligand screening. The known CB1 ligands are highly diverse structurally, so CB1 receptor may undergo considerable conformational changes to accept different ligands, which is challenging for molecular docking methods. To account for the flexibility of CB1 receptor, we constructed four CB1 receptor models based on four structurally distinct ligands, HU-210, ACEA, WIN55212-2 and SR141716A, using the newest x-ray crystal structures of human β_2 adrenergic receptor and adenosine A_{2A} receptor as templates. The conformations of these four CB1-ligand complexes were optimized by molecular dynamics (MD) simulations. The models revealed interactions between CB1 receptor and known binders suggested by experiments and could successfully discriminate known ligands and non-binders in our docking assays. MD simulations were used to

study the most flexible ligand, ACEA, in its free and bound states to investigate structural mobility achieved by the rearrangement of the fatty acid chain. Our models may capture important conformational changes of CB1 receptor to help improve accuracy in future CB1 drug screening.

3.2. Introduction

Cannabinoid (CB1) receptor belongs to class A G-protein coupled receptors (GPCRs) that represent the largest membrane protein family and are of great pharmacological importance. Currently, nearly one-third of marketed pharmaceuticals target GPCRs.¹ CB1 receptor is a therapeutically useful target involved in a wide variety of physiological processes, including metabolic regulation, craving, pain, and anxiety.^{2,3} Licensed drugs target cannabinoid receptors for treating chemotherapy-induced nausea and vomiting, relieving neuropathic pain, and as an appetite stimulant for AIDS patients.⁴ Drugs targeting CB1 receptor are continually being developed.⁵⁻⁸

Because no crystal structure of CB1 receptor is available, computational methods have been used to model the receptor.⁹⁻¹² Up to late 2007, the CB1 receptor model was built based on bovine rhodopsin, a GPCR, because of its available high-resolution structure as a template for CB1 receptor modeling.^{10,11,13-16} In recent years, with more GPCR structures being crystallized, new templates such as human adenosine A_{2A} receptor (AA_{2A}R) and β_2 adrenergic receptor (β_2 AR) have become available.^{12,17,18}

The x-ray structures of different GPCRs share overall topology, but local structures can differ.¹⁹ Molecular dynamics (MD) simulations and docking methods have been used to further study the conformational changes of CB1 receptor and the interactions between ligands and CB1 receptor. MD simulations of CB1 receptor embedded in a lipid bilayer have been used to gain insight into the interhelical and protein–ligand interactions.^{17,20,21} Because we lack ligand–CB1 receptor experimental structures, mutation experiments are commonly used to help determine functional residues that affect ligand binding and can be guidelines to evaluate modeling results.^{9,22-25}

Cannabinoid ligands are highly diverse structurally. CB1 agonists can be classified into four groups: classical cannabinoids (1, 2, Fig. 3.1), non-classical cannabinoids (3, 4, Fig. 3.1), endogenous cannabinoids (5, Fig. 3.1) and aminoalkylindoles (7, 8, Fig. 3.1). CB1 antagonists/inverse agonists are diarylpyrazoles (9, 10, Fig. 3.1) or in other chemical series.²⁶ Representative ligands in each group are in Figure 3.1. The groups differ greatly in constitution of rings and hydrocarbon chains. Although the endogenous cannabinoids do not have a ring conformation, the flexible long hydrocarbon chains can adopt conformations with high affinities.

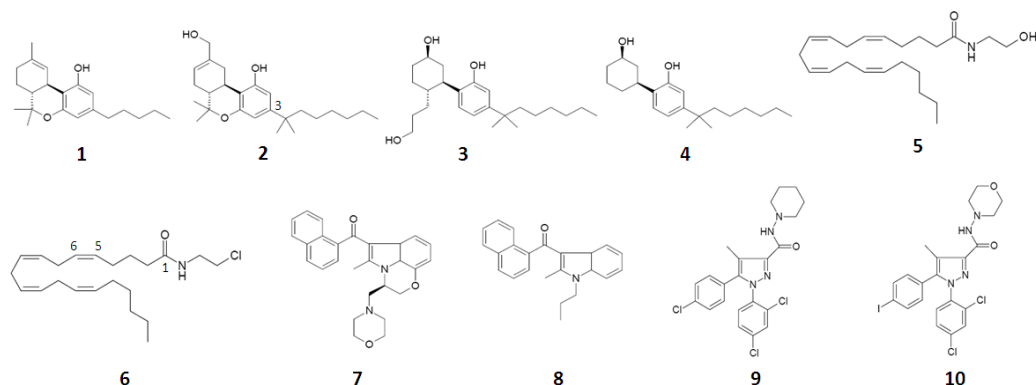


Figure 3.1. Molecular structures of cannabinoid ligands in (1, 2) classical cannabinoids, (3, 4) non-classical cannabinoids, (5) endogenous cannabinoids, (7, 8) aminoalkylindoles and (9, 10) diarylpyrazoles. (1) (-)- Δ^9 -THC, (2) HU-210, (3) CP55940, (4) CP47497, (5) anandamide, (6) ACEA, (7) WIN55212-2, (8) JWH-015, (9) SR141716A, (10) AM281.

In this study, in order to study the conformational changes in the ligand bound states and consider the flexibility of CB1 receptor for drug screening, we constructed four CB1 receptor homology models based on four structurally different tight binders: (-)-11-hydroxydimethylheptyl- Δ^8 -tetrahydrocannabinol (HU-210) (2, Fig. 3.1), arachidonyl-2-chloroethylamide (ACEA) (6, Fig. 3.1), (R)-(+)-[2,3-dihydro-5-methyl-3-[(4-morpholinyl)methyl]pyrrolo[1,2,3-de]-1,4-benzoxazin-6-yl](1-naphthalenyl)methanone (WIN55212-2) (7, Fig. 3.1) and N-(piperidin-1-yl)-5-(4-chlorophenyl)-1-(2,4-dichlorophenyl)-4-methyl-1H-pyrazole-3-carboxamide (SR141716A) (9, Fig. 3.1). The classical cannabinoid agonist HU-210 is a

structural analog of (-)- Δ^9 -tetrahydrocannabinol (THC) (1, Fig. 3.1) but with higher binding affinity. Because the classical and non-classical cannabinoids share overall structural features, HU-210 was selected to represent ligands in these two groups. ACEA is a selective CB1 agonist that is an endogenous cannabinoid analog but with higher binding affinity than a natural ligand anandamide (AEA) (5, Fig. 3.1). WIN55212-2 is a typical aminoalkylindole, and SR141716A is the first reported CB1 antagonist to display nanomolar CB1 receptor affinity.⁷ We first constructed homology models using the newest GPCR crystal structures, β_2 AR and AA_{2A}R, as templates and then used MD simulations and protein threading to train the models. To reveal changes in ligand flexibility in the free and bound state, motions of key ACEA dihedral angles were analyzed.²⁷ Since SR141716A is an antagonist to CB1 receptor, whole-protein MD simulations was used to adjust the model for antagonist/inverse agonist binding. The models can successfully discriminate known binders, compounds that are structurally similar to the binders and randomly chosen compounds by molecular docking and re-scoring.

3.3. Methods

3.3.1. Homology modeling

Construction of CB1 receptor homology models involved several essential steps. First, human brain CB1 protein sequence (GI: 237681175) was downloaded from the NCBI protein database (<http://www.ncbi.nlm.nih.gov/protein>). Two templates, β_2 AR (PDB

code: 3KJ6) and AA_{2A}R (PDB code: 3EML), were selected to construct CB1 model.^{28 29}

For multiple sequence alignment, GPCRs with similar protein sequences with CB1 receptor were identified by gapped BLAST search and HHSearch in template library of SWISS-MODEL.^{30 31} Multiple sequence alignment was performed at T-Coffee (<http://www.ebi.ac.uk/Tools/t-coffee/>) using sequences of bovine rhodopsin (PDB code: 1F88 and 1U19), β_2 AR (PDB code: 2R4R, 2RH1 and 3KJ6) and AA_{2A}R (PDB code: 3EML) (Table 3.1).³²⁻³⁶ To ensure that the alignment correctly placed the sequences of transmembrane helices (TMHs), PsiPred for secondary-structure prediction of CB1 receptor was used as guidelines to correct alignment.³⁷ Finally, with the two selected templates and the multiple sequence alignment, SWISS-MODEL (<http://swissmodel.expasy.org/>) *Alignment Mode* was used for CB1 receptor homology modeling.³⁸ Two models were built initially; one is based on template β_2 AR and the other is based on template AA_{2A}R.

3.3.2. Model refinement

To further refine the transmembrane region of CB1 model, each helix and loop fragment was also constructed individually by SWISS-MODEL. The shape of each individually built helix in the transmembrane region was compared with models generated with the full-length CB1 protein sequence by two templates (AA_{2A}R and β_2 AR). Based on the templates and PsiPred secondary-structure prediction, the best-shaped helix and loop pieces were isolated and assembled into a new hybrid

model (Fig. 3.2).³⁷ For example, compared helix 1 built by template AA_{2A}R (Fig. 3.2 A1), template β₂AR (Fig. 3.2 A2) and individual built helix 1 fragment (Fig. 3.2 A3), the beginning part of helix 1 built by template AA_{2A}R (Fig. 3.2 A1) did not show well-shaped helix compared to the other two structures which are both well-constructed.

After generating the backbone, sidechains were added by use of SCWRL4.⁴⁰ Quick conjugate gradients energy minimization and MD simulations were performed on CB1 hybrid model by NAMD/VMD in AMBER force field.⁴¹⁻⁴³ Because this study focuses on the ligand binding site, extensive loop conformational search was not performed.^{44,45}

Table 3.1 Sequence Alignment of Helix Regions of Multiple GPCR Sequences.

| Helix | Start residue | Sequence Alignment | | End residue |
|-------|---------------|--------------------|-----------------------------------------------------------|-------------|
| 1 | 35 | 1U19 | PWQFSMLAAYMFLIMLGFPI ^{NT} FTLYVTVQ | 65 |
| | 23 | 3EML | VY---ITVELAIAV-LA-ILGNVLCWAVWL | 48 |
| | 32 | 3KJ6 | VWVVGMGIVMSLIV-LAIVFGNVLVITAIK | 61 |
| | 113 | CB1 | PSQQLAIAVLSLTLGTFTVLENLIVLCVILH | 143 |
| | | | : : . * . * . . | |
| 2 | 72 | 1U19 | PLN-YILLNLA ^{VD} LFMVFG-GFTTTLYTSL | 100 |
| | 55 | 3EML | VTN-YFV ^{VS} LAAADI ^{AV} GL-AIPFA--ITI | 81 |
| | 68 | 3KJ6 | VTN-YFITS ^{LA} CADLVMGLA-VV ^{PF} GAAHIL | 96 |
| | 150 | CB1 | RPSYHF ^{IG} SLAVADLLG ^{SV} IFVYSFIDFHVF | 180 |
| | | | . : : . ** ** : . . | |

| | | | | |
|---|-----|------|------------------------------------------|-----|
| 3 | 106 | 1U19 | FGPTGCNLEGGFFATLGGEIALWSEVVLAIERYVVVC | 141 |
| | 87 | 3EML | AACHGCLFIACFVVLVTQSSIFSILAI AIDRYIAIR | 122 |
| | 102 | 3KJ6 | FGNFWCEFWTSIDVLCVTASIEELCVIAVDRYFAIT | 137 |
| | 186 | CB1 | RNVFLFKLG--GVTASFTASVGSIFLTAIDRYISIH | 219 |
| | | | : : : * * : : * . : | |
| 4 | 151 | 1U19 | ENHAIMGVAFTHVMALACAA-PPLVG | 175 |
| | 133 | 3EML | GTRAKGIIAICWVLSFAIGL-TPMLG | 157 |
| | 148 | 3KJ6 | KNKARVILIMVIVSGLTSFLPIQMH | 173 |
| | 230 | CB1 | RPKAVVAFCLMWTIAIVIAVL-PLLG | 254 |
| | | | : * . : * : : | |
| 5 | 201 | 1U19 | NESFVIY-MFVVHFIIPLIVIFFCYGLVFTVKEAAAQQ | 238 |
| | 188 | 3EML | PMNYMVFYFNFFACVLVPLLLMLGVYLRIFLAARRQ---- | 222 |
| | 197 | 3KJ6 | NQAYAIA-SSIVSFYVPLVIMVYVSRVFQEAQRQ---- | 230 |
| | 272 | CB1 | DETYLMFW-IGVT-SVLLLFIVYAYMYILWKAHSHAVRM | 308 |
| | | | . : * : . . * : . . : | |
| 6 | 246 | 1U19 | KAEKEVTRMVIIMVIAFLICWLEYAGVAFYIFT | 278 |
| | 389 | 3EML | KEVHAA-KSLAIVGLFALCWLFLHIINCFTFF | 420 |
| | 268 | 3KJ6 | KEHKAL-KTLGIIMGTFTLWLEFFIVNIVHVI | 299 |
| | 337 | CB1 | MDIRLA-KTLVLILVVLIIWGLLAIMVYDVF | 368 |
| | | | : : : : : : * * * : | |
| 7 | 285 | 1U19 | GPIFMTI-PAFFAKTSAVYNEVYIIM | 309 |
| | 427 | 3EML | APLWLMYLAIVLSHTNSVYVNFYIAY | 452 |
| | 306 | 3KJ6 | KEVY-ILL-NWIGYVNSGFNELYIC- | 328 |
| | 376 | CB1 | KTVF-AFC-SMLCLLNSTVNFIIYAL | 399 |
| | | | : : . : * * . * * | |
| 8 | 312 | 1U19 | KQFRNCMVTTL | 322 |
| | 455 | 3EML | REFRQTFRKII | 465 |
| | 331 | 3KJ6 | PDFRIAFQELL | 341 |
| | 402 | CB1 | KDLRHAFRSMF | 412 |
| | | | : : * : : | |

GPCR sequences included rhodopsin (PDB code: 1U19), β_2 adrenergic receptor (β_2 AR; PDB code: 3KJ6), human adenosine A_{2A} receptor (AA_{2A} R; PDB code: 3EML) and CB1 receptor, shown in helix regions only. Conservancy of the aligned sequence is represented by consensus symbols: "*" for identical residues (red); "." for conserved substitutions (cyan); and "." for semi-conserved substitutions (yellow).

Therefore, helix 1 structure constructed by template β_2 AR (Fig. 3.2 A2) was selected for CB1 model. In addition, helix 2 and 3 were further modified by MD simulations and protein threading. Finally, we adopted the conformation generated by protein threading at WURST (<http://www.zbh.uni-hamburg.de/wurst/>) (Table 3.2) ³⁹.

3.3.3. Modeling the binding site by four types of ligands

To more accurately construct the binding site of CB1 receptor, CB1 models were further optimized by four tight binders of CB1 receptor, HU-210, ACEA, WIN55212-2 and SR141716A, using MD simulations. Before optimizing, the CB1 model underwent MD simulations, with explicit water molecules included only in the protein binding site. After that, HU-210 was docked to the binding site of CB1 model using Vdock to obtain the initial conformation for MD simulations.⁴⁶ Antechamber in AmberTools was used to assign parameters to the ligand–protein complex.⁴⁷ MD simulations were carried out using the NAMD package implemented in VMD, with Amber ff99SB and

Table 3.2. Conformations of Helix 2 and 3 in CB1 Model Modified by Protein Threading.

| | PDB Code |
|--------------------------|----------|
| Residue Tyr172 to Asp176 | 3DAH |
| Residue Phe177 to Phe180 | 1SFS |
| Residue His181 to Ser185 | 1UDL |
| Residue Arg186 to Phe189 | 1SFS |
| Residue Leu190 to Gly194 | 1W85 |

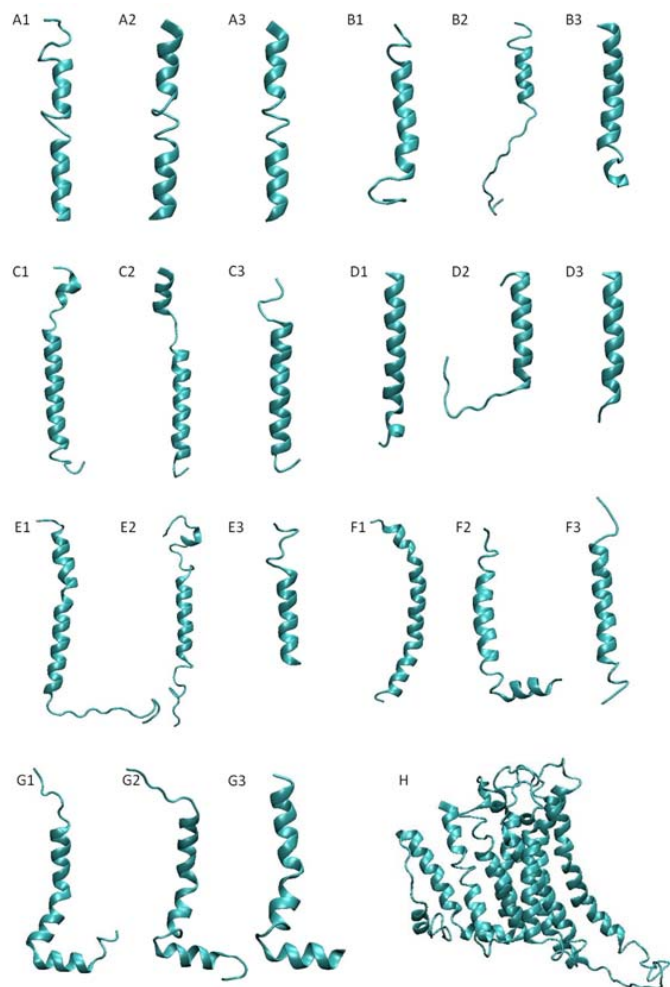


Figure 3.2. CB1 receptor hybrid model built by template-based fragment assembly. (A1-3) Helix 1 built by template (A1) AA_{2A}R, (*A2) β_2 AR and (A3) individual helix 1 fragment; (B1-3) Helix 2 built by template (B1) AA_{2A}R; (B2) β_2 AR and (*B3) individual helix 2 fragment; (C1-3) Helix 3 built by template (*C1) AA_{2A}R, (C2) β_2 AR and (C3) individual helix 3 fragment; (D1-3) Helix 4 built by template (*D1) AA_{2A}R, (*D2) β_2 AR and (D3) individual helix 4 fragment; (E1-3) Helix 5 built by template (*E1) AA_{2A}R, (*E2) β_2 AR and (E3) individual helix 5 fragment; (F1-3) Helix 6 built by template (*F1) AA_{2A}R, (F2) β_2 AR and (F3) individual helix 6 fragment; (G1-3) Helix 7 built by template (*G1) AA_{2A}R, (G2) β_2 AR and (G3) individual helix 7 fragment; (H) assembled CB1 receptor hybrid model. *, structures assembled into the hybrid model. For helix 4 and 5, part of the helix is from structure built by AA_{2A}R and the other part is by β_2 AR.

general Amber force field (GAFF) for CB1 receptor and ligands, respectively.⁴¹⁻⁴³ All MD simulations were performed under the NVT ensemble at 300 K with a time step of 1 fs. Residues within 12 Å of HU-210 were set as mobilized. When CB1 receptor was in the ligand free state, several polar residues, such as K3.28(192), formed hydrogen bonds with themselves and neighboring residues. To more efficiently relax the sidechain conformations, residues within 12 Å of HU-210 were heated to 500 K shortly (around 0.1 ns) each time to rearrange sidechain conformations and re-form proper ligand–CB1 receptor intermolecular interactions. The system was then followed by 1000 steps quick energy minimization and longer MD simulations in 300 K to prevent secondary structure changes. MD simulations were performed at 500 K and 300 K alternately and in total 1.2 ns at 500 K and 5.2 ns at 300 K. Afterwards, quick conjugate gradients energy minimization (1000 to 3000 steps) in AMBER force field was used to obtain a local energy minimum conformation of the CB1 receptor model for molecular docking studies.⁴³

We optimized the other three CB1 models using ACEA, WIN55212-2 and SR141716A with the same procedure described in *Method 2.3*. Root-mean-square deviation (RMSD) of residues that were within 12Å of ACEA during MD simulations of ACEA-CB1 complex is shown in Fig. 3.3. Since SR14176A is an antagonist for CB1 receptor, training the CB1 model based on SR141716A toward inactive state was necessary. Since both WIN55212-2 and SR141716A share certain structural

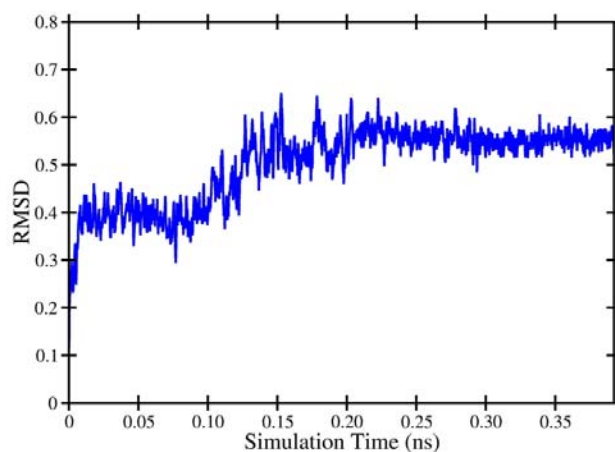


Figure 3.3. RMSD of residues that were within 12Å of ACEA during MD simulations of ACEA-CB1 receptor complex.

similarities and were identified to share the same microdomain in CB1 receptor⁹, we constructed inactive CB1 receptor model using the initial model trained by WIN55212-2. Four ns whole-protein MD simulations were used to adjust the model for antagonist/inverse agonist binding.

3.3.4. Model verification

Four CB1 receptor models were then tested by docking known binders, compounds that are structurally similar to the known binders and random compounds with Vdock.⁴⁶ AutoDock 4 and AutoDock Vina were also used; however, both programs failed to dock ACEA ligand, so we reported only results from Vdock.^{48,49} Structures of

CB1 binders are listed in Fig. 3.4. Compounds with similar structures to each known binder were selected by substructure search and similarity search (similarity = 50%) from the ZINC database (Fig. 3.5).⁵⁰ The resulting compounds were further screened to avoid highly similar structures. For example, compounds that are structurally similar to HU-210 were selected using substructure search (Fig. 3.5 A1-4) and similarity search (similarity = 50%) from the ZINC database and 59 of them (Table 3.3) were picked to dock to the CB1 model trained by HU-210. In addition, 25 random compounds (Fig. 3.6) were chosen from the NCI diversity set II (http://dtp.nci.nih.gov/branches/dscb/div2_explanation.html).

3.3.5. Energy calculation

In addition to docking, energy calculation was applied for re-scoring. In total, 3000 steps of energy minimization were used to relax the binding sites of the best-docked ligand–CB1 complexes. Total energy (E) was calculated by NAMDenergy in VMD.⁴² Binding energy between protein and ligand was calculated as follows:

$$\Delta E_{bind} = E_{complex} - (E_{protein} + E_{ligand})$$

Because ligands that are similar to HU-210 are smaller than other types of CB1 ligands, for the CB1 model optimized by HU-210, 500 to 1000 steps, instead of 3000 steps, of energy minimization were used before energy calculation. The minimization was terminated when the dihedral energy reached a plateau.

3.3.6. Flexibility analysis of ACEA

To study motions of the unsaturated acyl chain of ACEA, we used T-Analyst²⁷ to analyze the flexibility of each rotatable bond for trajectories from MD simulations. We performed three 3 ns MD simulations for ACEA in different environments: ACEA–CB1 bound state, free ACEA in a 10 Å × 10 Å × 10 Å cubic TIP3P water box, and free ACEA in vacuum. Frames were saved every 3 ps, for a total of 1000 frames. The MD simulations all started from the same initial ACEA conformation, with a folded structure from the ligand bound state. The configurational entropy of 17 dihedral angles of ACEA was computed using T-Analyst.

3.4. Results and Discussion

3.4.1. CB1 receptor models

In considering ligand-induced conformational changes in the binding site of CB1 receptor and the highly diverse structures of CB1 ligands, we constructed four CB1 receptor models based on four types of ligands: HU-210, ACEA, WIN55212-2 and SR141716A. The four CB1 ligands are in similar positions in the binding sites (Fig. 3.7). However, because the compounds have various scaffolds, the backbones of the four CB1 receptor models are slightly different to accommodate the ligands.

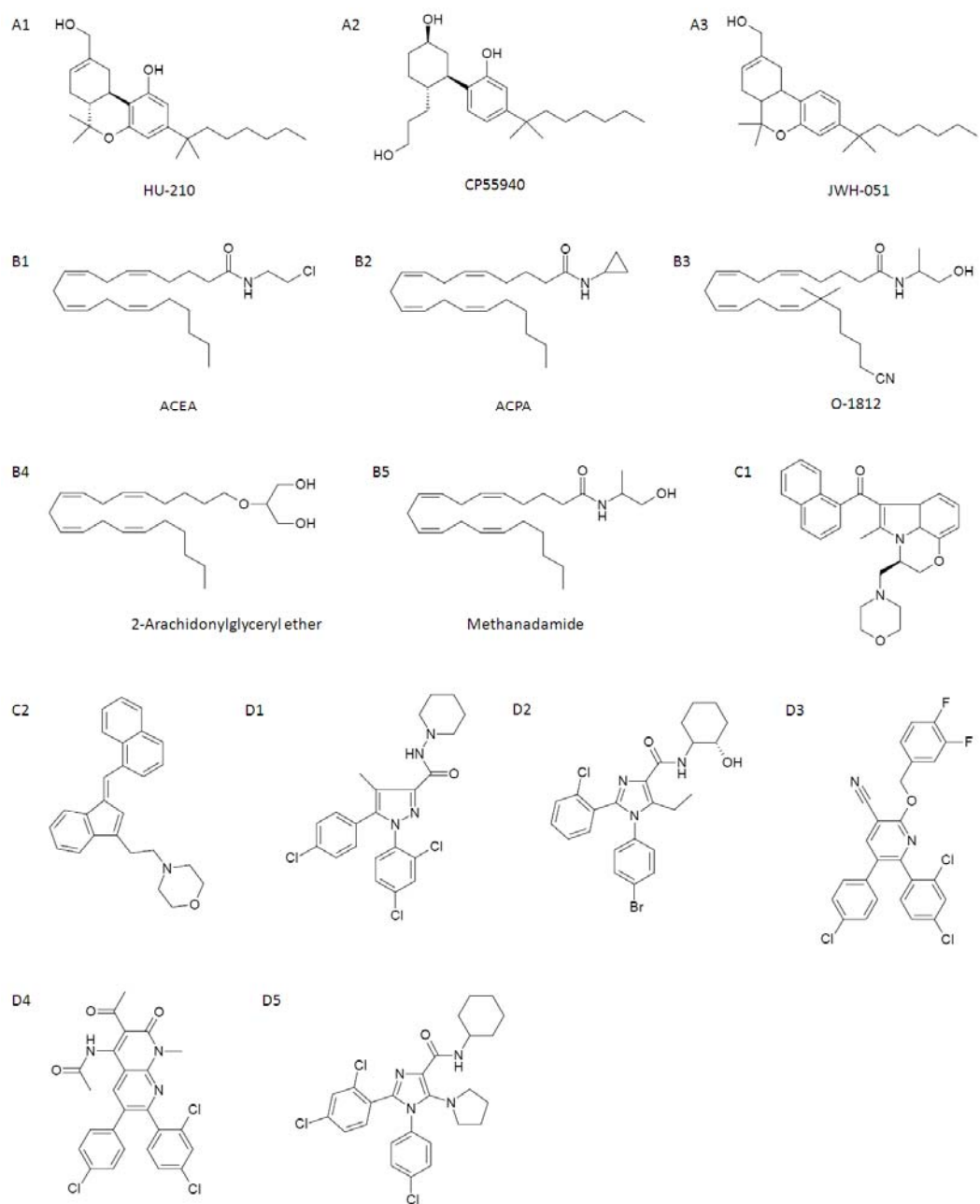


Figure 3.4. CB1 binders docked to models built by (A1-3) HU-210, (B1-5) ACEA, (C1-2) WIN55212-2 and (D1-5) SR14176A

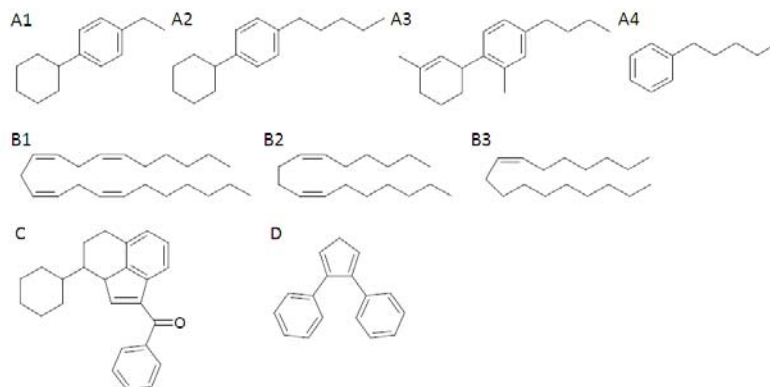


Figure 3.5. Substructures of (A1-4) HU-210, (B1-3) ACEA, (C) WIN55212-2 and (D) SR141716A used for substructure search in the ZINC database.

Table 3.3. Average Interaction Energy of Known CB1 Binders, Compounds Structurally Similar to Binders and Random Compounds in the Four CB1 Models. (kcal/mol).

| Model | Binders | | Compounds with Similar Binder Structures* | | Random Compounds** | |
|------------------|----------------|--------------|-------------------------------------------|-------------------|--------------------|-------------------|
| | No. | Binding E. | No. | Binding E. | No. | Binding E. |
| HU-210 Model | 3 ^a | -50.43 ±1.37 | | | 22 | -31.58 ±9.39 |
| | | | 59 | -36.73 ±7.57 | 3 | Failed to dock*** |
| ACEA Model | 5 ^a | -62.42 ±2.81 | 15 | -48.94 ±13.56 | 23 | -44.43 ±13.23 |
| | | | 5 | Failed to dock*** | 2 | Failed to dock*** |
| WIN55212-2 Model | 2 ^b | -62.12±5.82 | 36 | -45.12 ±9.76 | 24 | -44.63 ±14.6 |
| | | | 3 | Failed to dock*** | 1 | Failed to dock*** |
| SR141716A Model | 5 ^c | -66.53±6.10 | 38 | -50.59 ±11.98 | 20 | -37.83 ±11.83 |
| | | | 27 | Failed to dock*** | 5 | Failed to dock*** |

*, compounds structurally similar to binders were selected by substructure and similarity search (similarity = 50%) from ZINC database; **, random compounds were chosen from the NCI diversity set II; ***, docking failed because the docking software failed to dock the ligand in the binding site of the protein or the ligand was docked in an inappropriate position. a, b and c refer to references 21,50 and 51.

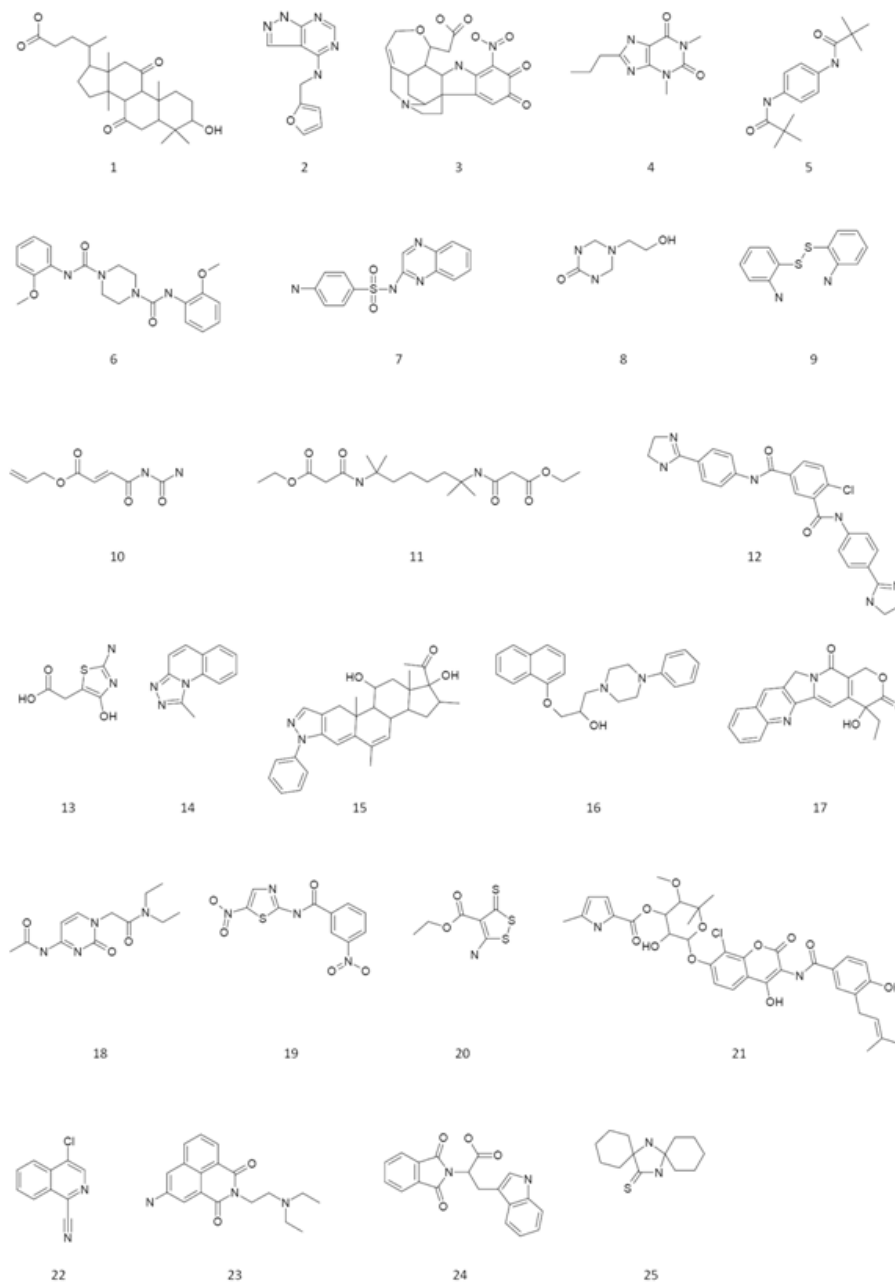


Figure 3.6. Molecular structures of 25 random non-binders from NCI diversity set II. (1) NSC1614, (2) NSC1620, (3) NSC5069, (4) NSC14304, (5) NSC29073, (6) NSC35930, (7) NSC41805, (8) NSC47881, (9) NSC54509, (10) NSC55957, (11) NSC57608, (12) NSC67436, (13) NSC70980, (14) NSC76484, (15) NSC80997, (16) NSC91398, (17) NSC94600, (18) NSC166846, (19) NSC171303, (20) NSC197049, (21) NSC227186, (22) NSC288686, (23) NSC308848, (24) NSC401077, (25) NSC515893.

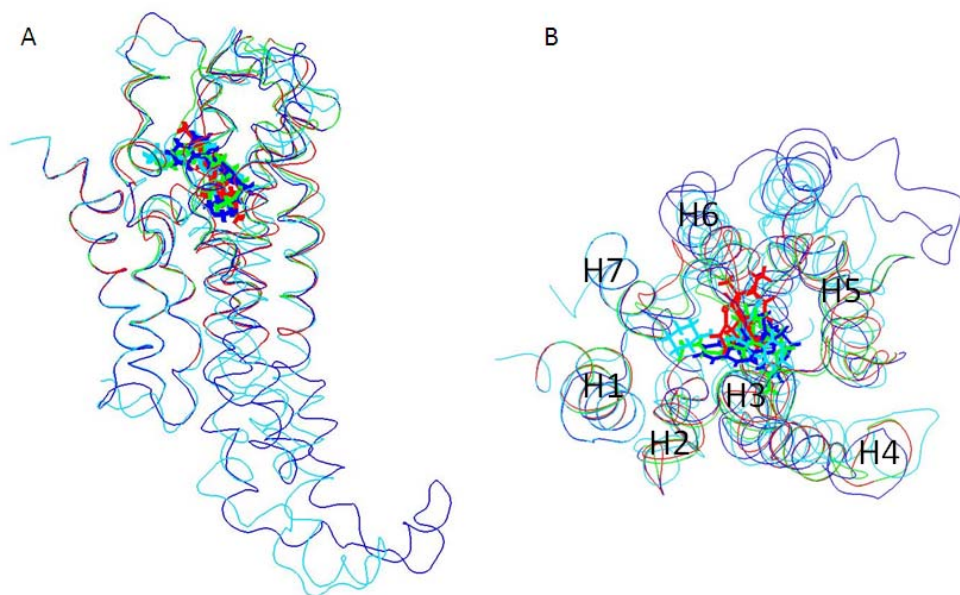


Figure 3.7. Alignment of the four ligand-CB1 models. (A) Side view. (B) Top view. Seven transmembrane helices are indicated (H1-7). Colors represent models trained by HU-210 (red), ACEA (green), WIN55212-2 (blue) and SR141716A (cyan).

3.4.1.1. CB1 model trained by HU-210

HU-210 is an analog to (-)- Δ^9 -THC, with higher binding affinity than natural THC, so it was selected to represent structurally similar classical and non-classical cannabinoids. In this model, HU-210 binds to the TMH3-6-7 region of CB1 receptor. The alkyl chain of HU-210 points to the inside of the binding cavity and locates near I6.46(354), C6.47(355), W6.48(356), L6.51(359) and L6.52(360), and the tricyclic scaffold of HU-210 is toward F2.64(177), F3.25(189), K3.28(192), M6.55(363), F7.35(379) and S7.39(383) (Fig. 3.8 A). Mutation study²² showed K3.28(192) is an important residue

involved in HU-210 binding, and the phenolic oxygen of HU-210 formed a hydrogen bond with K3.28(192) in our complex conformation. In addition, the binding affinity of HU-210 was reduced 50- to 100-fold when mutating S7.39(383) to alanine.²⁴ In our model, K3.28(192) acts as a hydrogen bond donor to F3.25(189) which locates near the tricyclic scaffold of HU-210 and helps stabilize HU-210 binding. The pyranyl oxygen of HU-210 acts as a hydrogen bond acceptor to S7.39(383). Besides, C6.47(355) is in close contact with the end of the alkyl chain of HU-210, which is also consistent with previous docking and site-directed mutagenesis study.⁵¹

3.4.1.2. CB1 model trained by ACEA

ACEA is a selective CB1 agonist that is structurally similar to endogenous cannabinoids such as AEA but with higher binding affinity. The long unsaturated acyl chain distinguishes itself from other groups of cannabinoid ligands. In our model, ACEA binds to TMH2-3-6-7 region and adopted a folded J-shape to form hydrophobic intermolecular contacts with CB1 receptor. U-shaped endocannabinoids conformations were also found in others studies.^{11 16} ACEA is located at the lipophilic region of the binding site and in the vicinity of residues such as K3.28(192), S7.39(383), F3.25(189), F7.35(379), F7.37(381), Y5.39(275), F3.36(200) and

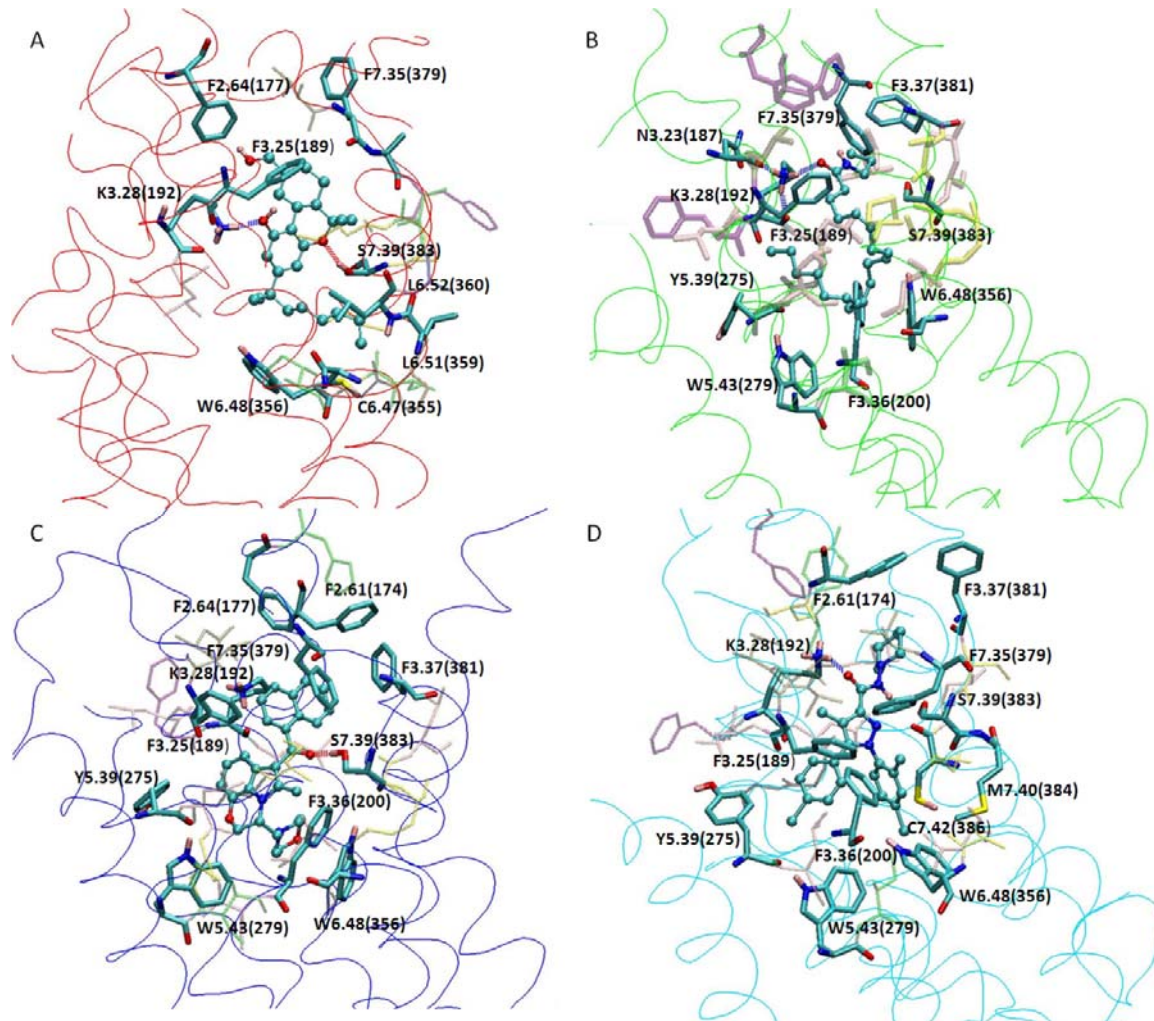


Figure 3.8. Binding sites of CB1 receptor models trained by (A) HU-210, (B) ACEA, (C) WIN55212-2 and (D) SR141716A. Ligands are in ball-and-stick depiction. Residues that that are directly contact with ligands are in opaque stick depiction. Residues that are within 6 Å of the ligands are in transparent stick depiction (see Table 3.4 for details).

Table 3.4. Residues within 6 Å of Ligands in Ligand-CB1 Complex.

| Model | Residues |
|------------------|--------------------------------------------------------------------------------------------------------------------|
| HU210 model | 174, 177, 178, 187 to 194, 276, 280, 353 to 363, 376, 378 to 384, 386 |
| ACEA model | 126, 171, 172, 174, 177, 187 to 196, 200, 275 to 277, 279, 280, 356, 357, 360, 363, 379, 381 to 384, 386, 387, 389 |
| WIN55212-2 model | 174, 177, 178, 187 to 195, 200, 201, 275 to 277, 279, 280, 356, 357, 360, 363, 379, 381, 383, 384, 386, 387 |
| SR14176A model | 126, 171 to 174, 177, 187 to 195, 200, 275, 276, 279, 280, 355, 356, 360, 363, 378, 379, 381 to 384, 386, 387 |

W6.48(356). According to mutation study, K3.28(192) is a critical residue for AEA binding²², and the carbonyl oxygen of ACEA formed a hydrogen bond with K3.28(192) in our model. Moreover, K3.28(192) formed another two hydrogen bonds; one, an intra-molecular hydrogen bond with its own carbonyl oxygen and the other with N3.23(187) (Fig. 3.8B). Mutation study revealed that the binding affinity of AEA decreased an approximate 13-fold by mutating Y5.39(275) to phenylalanine, but mutating Y5.39(275) to isoleucine could abolish ligand binding and receptor signaling.²³ In addition, mutation of F3.25(189) to alanine moderately (approximately 6-fold) decreased AEA binding affinity.⁹ In our model, Y5.39(275) locates very close to ACEA and should contribute to the ligand binding, as suggested by experiments. The C5=C6 double bond of ACEA interacts with F3.25(189), which was also reported previously.⁹

3.4.1.2.1. Flexibility study of ACEA

ACEA is significantly more flexible than most drug-like compounds. A large penalty in configuration entropy is expected, which opposes binding. However, the unsaturated acyl chain of ACEA can adopt many energetically equivalent conformations that may help compensate potential entropy loss due to rigidifying the flexible compound.⁵² To reveal the most rigid/flexible regions of ACEA and how the conformational changes of ACEA may affect binding, we therefore examined the changes in ACEA flexibility in three states: ACEA–CB1 bound state, free ACEA in a water box and free ACEA in vacuum. Notably, docking such a highly flexible ligand can be challenging, because the acyl chain of ACEA may cause an insufficient search for docking programs to find the best ligand-binding mode.

MD simulations were performed on ACEA in different environments: ACEA–CB1 bound state, free ACEA in a $10 \text{ \AA} \times 10 \text{ \AA} \times 10 \text{ \AA}$ cubic TIP3P water box, and free ACEA in vacuum. Conformations were analyzed by T-Analyst. In free states, ACEA could adopt many conformations, such as extended linear structures, folded J-shape, and U-shaped conformations. In contrast, in the bound state, ACEA adopted a curved J-shape and had fewer conformations.

We computed the configurational entropy of each ACEA dihedral angle for each simulation (see Fig. 3.9 A, B): the values for the ACEA unsaturated acyl chain were

significantly lower in the bound state than the free states in water or in vacuum. Therefore, the unsaturated acyl chain of ACEA became significantly rigid after binding because of the confined binding site. Interestingly, the free ACEA in the aqueous environment was less flexible than that in vacuum, which shows diminished hydrophobic effects. The entropy of dihedrals 9 and 11 was significantly lower in the free state in water than in the bound state. Figure 3.9 C-E shows the distribution of dihedral 9 in the bound state and free states in water and in vacuum. In the bound state, dihedral 9 samples one more conformation than in its free form in water, where the binding site of CB1 receptor provides a more non-polar environment than in the aqueous environment. In analyzing the MD trajectories, we found that dihedral 9 acts as a “hinge” in the middle of ACEA to keep the free ACEA partially folded in water; while the long chains on both sides of the “hinge” move flexibly. In the bound state, the protein binding site confines the movement of the long acyl chain, but it brings new ACEA conformations by rearranging the “hinge”, dihedral 9, to adjust the conformations of long chains on both sides. However, observed from the conformations of ACEA displayed in Fig. 3.9 C-E, although dihedral 9 can adopt more conformations in the ACEA-CB1 bound state, the entire ligand is rigid as compared to the ACEA free states in a water box or in vacuum.

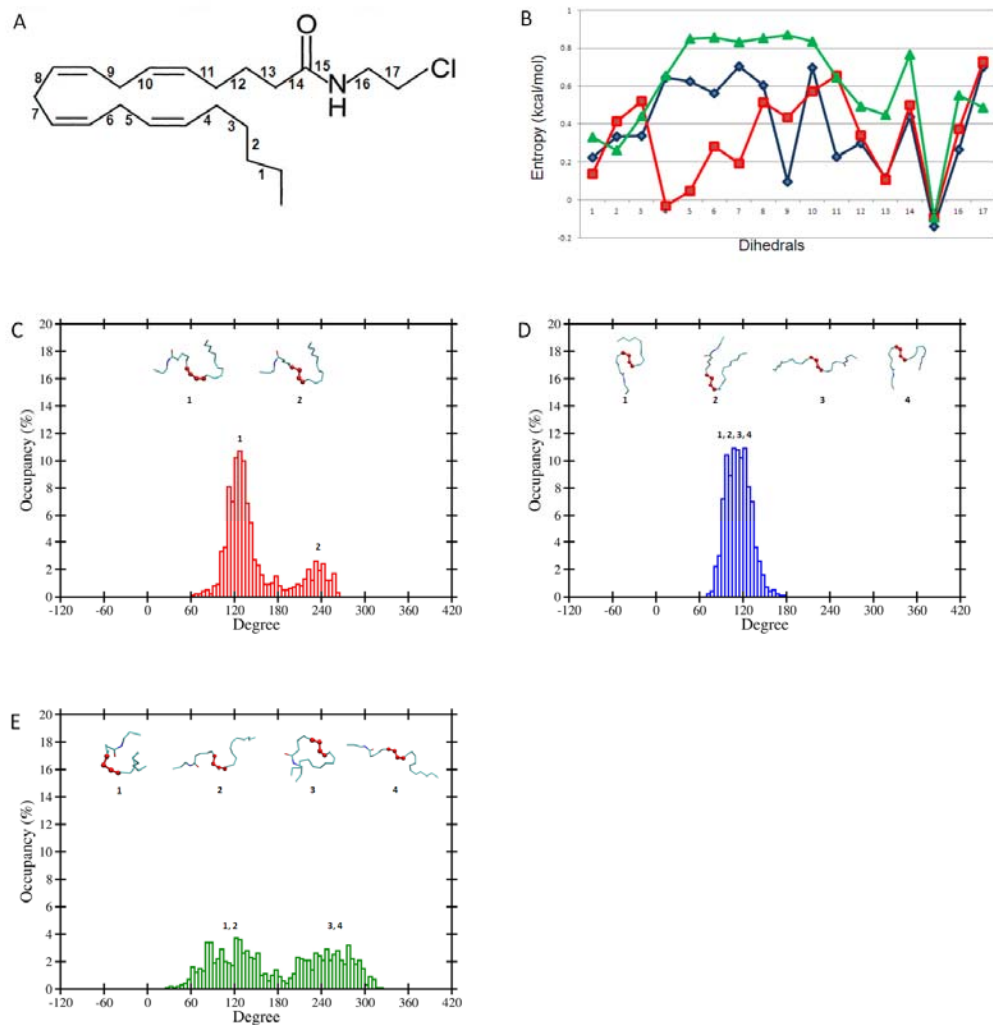


Figure 3.9. Entropy and distributions of ACEA dihedrals. (A) 17 dihedrals of ACEA analyzed by T-Analyst. (B) Configurational entropy of each ACEA dihedral angle. Blue and green lines show ACEA entropy of the free ligand state in a $10 \text{ \AA} \times 10 \text{ \AA} \times 10 \text{ \AA}$ cubic TIP3P water box and in vacuum, respectively; red line shows the entropy in the ACEA–CB1 bound state. (C–E) Distributions of dihedral 9 and ACEA conformations in (C) ACEA–CB1 bound state, (D) free state in water box and (E) free state in vacuum.

3.4.1.3. CB1 model trained by WIN55212-2

WIN55212-2 is a typical aminoalkylindole, with relatively more ring scaffolds and a larger structure than HU-210 and ACEA. Therefore aromatic stacking can be important for binding WIN55212-2 and its analogs to CB1 receptor.⁹ Different from models built by HU-210 and ACEA, this model shows no hydrogen bonds between WIN55212-2 and K3.28(192), which agrees with mutation results.²² In our model, WIN55212-2 is located at TMH 3-5-6-7 region. An aromatic microdomain constitutes the binding region for WIN55212-2 which includes F2.61(174), F2.64(177), F3.25(189), F3.36(200), Y5.39(275), W5.43(279), W6.48(356), F7.35(379) and F3.37(381) (Tables 3.5, 3.6 and Fig. 3.8C). The arrangement of residues in the binding site can be further supported by mutagenesis results, which indicated that the aromaticity of Y5.39(275) is crucial for WIN55212-2 binding.²³ McAllister and co-workers suggested that WIN55212-2 binds within TMH3-4-5-6 aromatic microdomain and directly interacts with F3.36(200), W5.43(279) and W6.48(356) with aromatic stacking.⁹ In our model, residues that are in close contact with WIN55212-2 are as follows: for the naphthyl ring, F2.61(174), F2.64(177), F3.25(189), F7.35(379) and F3.37(381); for the indole ring, F3.25(189), Y5.39(275) and W5.43(279); and for the morpholinyl moiety, F3.36(200), W5.43(279) and W6.48(356). Our model showed different but more toward the “Aroyl-up1” WIN55212-2 binding mode compared to the binding conformations proposed by Shim and Howlett.⁵³ Experimental studies suggested that F3.36(200) and

W5.43(279) play critical roles in providing bulky groups for WIN55212-2 binding.²⁵ When the aromatic residues F3.36(200), W5.43(279), and W6.48(356) were replaced by alanine, the binding of both WIN55212-2 and SR141716A to CB1 receptor was significantly reduced.⁵⁴ In addition to F3.36(200), F7.35(379) directly interacts with WIN55212-2 by aromatic stacking, and a few more aromatic stacking interactions were observed among F2.61(174), F2.64(177), F7.35(379) and F3.37(381) in our model (Tables 3.5 and 3.6).

3.4.1.4. CB1 model trained by SR141716A

SR141716A is the first reported CB1 antagonist that displayed nanomolar CB1 receptor affinity and stabilized the receptor in its inactive state.⁷ Our SR141716A model was trained toward inactive state by 4 ns MD simulations on the ligand-CB1 complex. In this model, SR141716A locates at the TMH 3-4-5-6-7 region and binds to the same aromatic microdomain with WIN55212 (Tables 3.5 and 3.7). Overall, the binding site of this model is composed of F2.61(174), F3.25(189), K3.28(192), F3.36(200), Y5.39(275), W5.43(279), W6.48(356), F7.35(379), F7.37(381), S7.39(383) and M7.40(384) (Fig. 3.8D). McAllister and co-workers suggested that SR141716A binds within TMH3-4-5-6 aromatic microdomain in

Table 3. 5. Key Ligand-Aromatic Clustering in CB1 Models Optimized by WIN55212-2 and SR141716A.

| | WIN55212-2 | | | | SR141716A | | | |
|------------|-----------------------|-----------------------|------------------|----------|-----------------|----------|-----------------|----------|
| | NAP ^a | | IND ^b | | MC ^c | | DC ^d | |
| | <i>d</i> ^e | α ^f | <i>d</i> | α | <i>d</i> | α | <i>d</i> | α |
| F2.61(174) | 7.99 | 69.57 | 13.45 | 78.96 | - | - | - | - |
| F2.64(177) | 6.15 | 87.40 | 10.90 | 53.29 | - | - | - | - |
| F3.25(189) | 8.19 | 88.53 | 7.98 | 88.23 | 6.22 | 81.98 | 7.91 | 49.68 |
| Y5.39(275) | 10.97 | 78.51 | 5.81 | 38.26 | 6.18 | 33.85 | 10.32 | 75.99 |
| W5.43(279) | 12.81 | 24.83 | 7.24 | 86.44 | 7.00 | 52.09 | 7.18 | 87.56 |
| F7.35(379) | 4.27 | 32.87 | 8.69 | 89.48 | 7.66 | 82.97 | 6.89 | 73.34 |
| F3.37(381) | 6.26 | 88.69 | 11.78 | 30.42 | - | - | - | - |

^a Naphthyl ring. ^b Indole ring. ^c Monochlorophenyl ring. ^d Dichlorophenyl ring. ^e The distance between the centroids, in Å. ^f The angle between the ring planes, in degrees.

Table 3. 6. Key Aromatic Clustering in CB1 Models Optimized by WIN55212-2.

| | F2.64(177) | | F7.35(379) | | F3.37(381) | |
|------------|-----------------------|-----------------------|------------|----------|------------|----------|
| | <i>d</i> ^a | α ^b | <i>d</i> | α | <i>d</i> | α |
| F2.61(174) | 6.86 | 64.06 | 9.25 | 79.72 | 6.16 | 82.31 |
| F2.64(177) | - | - | 6.54 | 60.75 | 7.78 | 34.70 |
| F7.35(379) | - | - | - | - | 5.03 | 58.22 |

^a The distance between the centroids, in Å. ^b The angle between the ring planes, in degrees.

CB1 inactive state and directly involves aromatic stacking interactions with F3.36(200), Y5.39(275) and W5.43(279), as well as hydrogen bonding with K3.28(192).⁹ Hurst and co-workers hypothesized that, in ligand free state, the salt bridge between K3.28(192) and D6.58(366) appears to be important to position K3.28(192) for ligand interaction.^{55,56} In our model, with SR141716A binding, K3.28(192) forms a hydrogen bond with the carbonyl oxygen of SR141716A and also two more hydrogen bonds with nearby N3.23(187) and S2.60(173). F3.25(189) and Y5.39(275) directly form single aromatic stacking interactions with the monochlorophenyl ring of SR141716A. W5.43(279) interacts with both the monochlorophenyl and the dichlorophenyl ring of SR141716A, which is consistent with the modeling studies by McAllister and co-workers.⁹ Besides, F3.36(200), W6.48(356) and F7.35(379) interact with the dichlorophenyl ring of SR141716A. F7.35(379) directly stacks with F3.25(189) (Table 3.7). Mutagenesis study showed that mutation of F3.36(200), W5.43(279) or W6.48(356) to alanine significantly reduced the binding of SR141716A to the CB1 receptor.^{9,54,55} Studies suggest that the interaction between F3.36(200) and W6.48(356), representing a “toggle switch”, is an important constrain that keeps CB1 receptor in its inactive state and the salt bridge between K3.28(192) and D6.58(366) stabilizes CB1 in its inactive state.^{54,55} In our model, F3.36(200) and W6.48(356) form a parallel-displaced stacking (Table 3.7) and both constitute the aromatic microdomain for SR141716A binding. Moreover,

C7.42(386) is right located at the dichlorophenyl ring of SR14176A in our model while the introduction of a bulky group on C7.42(386) may inhibit SR141716A binding.⁵⁷

3.4.1.5. Model comparisons

Although the ligand binding sites of the four models were similar, different residues and various sidechain rearrangements characterized each binding pocket. For example, the slightly different sidechain positions of F7.35(379) and F7.37(381) could result in binding to dissimilar ligands (Fig. 3.10). As well, the backbones that constructed each model varied, so mobilizing sidechains in the binding site might not be enough for CB1 receptor to accept different types of ligands. The models also

Table 3. 7. Key Aromatic Clustering in CB1 Models Optimized by SR141716A.

| | | d^a | α^b |
|------------|------------|-------|------------|
| F2.61(174) | F2.64(177) | 6.75 | 75.19 |
| F2.61(174) | F3.37(381) | 6.40 | 53.83 |
| F3.25(189) | F7.35(379) | 5.13 | 23.82 |
| F3.36(200) | W6.48(356) | 7.42 | 6.42 |

^a The distance between the centroids, in Å. ^b The angle between the ring planes, in degrees.

illustrated the flexibility of CB1 receptor and showed that multiple models were necessary for ligand screening; typical molecular docking programs cannot mobilize protein backbone structures.

3.4.2. CB1 receptor model validation

CB1 receptor models were assessed by docking and scoring known binders, compounds that are structurally similar to the binders and random compounds to determine whether the models could successfully rank known binders. We first performed molecular docking to place ligands into the binding site and quick energy minimization was then followed. To rank ligands, interaction energy between ligand

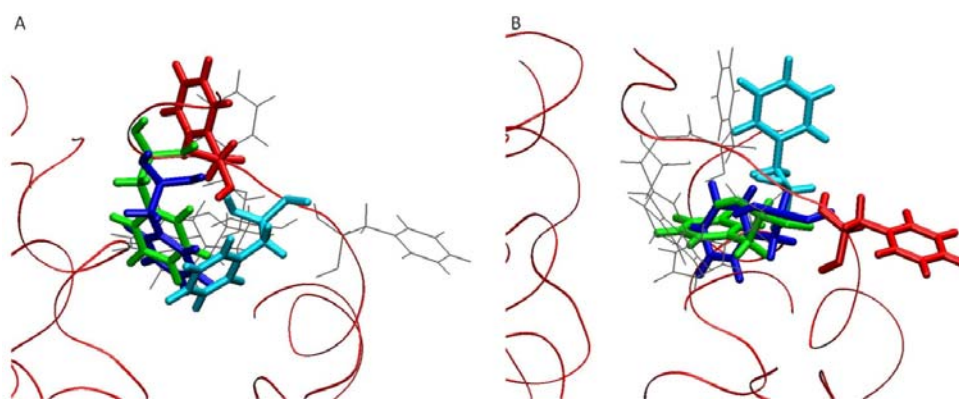


Figure 3.10. Residue (A) F7.35(379) and (B) F3.37(381) in the binding sites of the four models. Red: HU-210 model; green: ACEA model; blue: WIN55212-2 model and cyan: SR141716A model.

and protein was calculated. The average binding energies for known binders, compounds with similar structure to the binders and random compounds (Table 3.3) validated that the four CB1 receptor homology models can distinguish known binders from the unknown compounds. Note that the energy computed does not include entropic contribution, but the binders clearly showed considerably stronger energetic attraction.

To gain insight into the ligand-induced conformational changes in the binding site of CB1 receptor, cross docking was performed by docking binders of one model to other models. Similar to with model validation, energy minimization was applied, and interaction energy was calculated to obtain average interaction energy for comparisons (Table 3.8). The four CB1 receptor models could reasonably accommodate all CB1 binders, which suggests that although the models were optimized by four structurally distinguishable ligands, the binding sites maintain key characters for CB1 ligand binding. The SR141716A model has the most unique nature in failing to bind most analogs of THC and AEA but only accepts analogs of WIN55212-2, which may due to their structural similarity. Of note, SR141716A is a known CB1 antagonist. Although simulating the active and inactive CB1 receptor structures is beyond the scope of this study, the distinct SR141716A model suggests possibly substantial conformational changes between the active and inactive states of CB1 receptor.

The cross-docking results suggested that CB1 ligands may be able to bind in each optimized model. However, if a certain type of ligand was docked to a CB1 model optimized by other types of ligands, the key interactions between the ligand and CB1receptor were rarely observed. As a result, if the scaffold of a ligand is decided, choosing a model optimized by that class of ligand is preferred. For example, if compounds with similar scaffolds as WIN55212-2 need to be screened, choosing CB1

Table 3.8. Interaction Energy of Cross Docking Results (kcal/mol).

| | HU-210 Model | ACEA Model | WIN55212-2 Model | SR14176A Model |
|------------------------------|-----------------|----------------|---------------------|-----------------------------|
| HU-210 | -48.85 | -43.85 | -45.19 | Failed to dock ^a |
| CP55940 | -51.25 | -58.17 | -64.86 | -56.7 |
| JWH-051 | -51.18 | -63.12 | -62.15 | Failed to dock |
| ACEA | -62.64 | -60.54 | -66.67 | -62.29 |
| ACPA | -38.82 | -58.79 | -60.63 | Failed to dock |
| O-1812 | -55.07 | -64.64 | -68.03 | Failed to dock |
| 2-Arachidonylglycerol ether | -52.24 | -62.57 | -73.32 | Failed to dock |
| Methanadamide | -47.36 | -65.56 | -63.74 | Failed to dock |
| WIN55212-2 | -65.91 | -62.93 | -66.23 | -61.87 |
| CHEBI309167 | -48.81 | -60.34 | -58 | -51.39 |
| SR14176A | -62.62 | Failed to dock | Failed to dock | -61.66 |
| Binder 1 similar to SR14176A | -48.04 | -64.22 | -62.72 | -64.57 |
| Binder 2 similar to SR14176A | -43.23 | -58.2 | -58.38 | -61.19 |
| Binder 3 similar to SR14176A | -45.13 | -56.31 | -63.51 | -65.05 |
| Binder 4 similar to SR14176A | -56.29 | -62.24 | -64.28 | -75.32 |

Interaction energy was calculated after 3000 steps of energy minimization. ^a Docking failed either because the docking software failed to dock the ligand in the protein binding or the ligand was docked in an inappropriate position.

model optimized by WIN55212-2 should be preferred. Compared with other CB1 ligands, HU-210 is relatively small in size, so the model optimized by HU-210 has the smallest binding pocket. Thus, the HU-210 model is the most ideal to screen high-affinity binders with less bulky structures. For screening compounds with linear forms, the ACEA model is the most suitable because it can accept highly flexible long acyl chains. However, ligands with flexible long chains are challenging for docking programs to find the correct binding modes, which must be carefully considered when screening ligands.

3.5. Conclusions

To study ligand-induced conformational changes in the CB1 receptor binding site, we constructed four homology models based on HU-210, ACEA, WIN55212-2 and SR141716A. Each of these four ligands represents one group of structurally diverse cannabinoid ligands. The interactions in the binding site of each model were carefully studied to ensure that our models reproduce known interactions suggested by experiments. Molecular docking results showed that our models can distinguish known binders from compounds with similar structures to binders and random compounds. Although all models can accept most CB1 ligands, they have preferences for different ligand scaffolds. Therefore, determining which of the four models is the most suitable one when screening a particular type of ligand can help achieve the most accurate results. The coordinates of the CB1 models are available upon request.

3.6. References

1. Kimple, A. J.; Bosch, D. E.; Giguere, P. M.; Siderovski, D. P., Regulators of G-Protein Signaling and Their G alpha Substrates: Promises and Challenges in Their Use as Drug Discovery Targets. *Pharmacological Reviews* 2011, *63* (3), 728-749.
2. Mackie, K., Cannabinoid receptors as therapeutic targets. *Annual Review of Pharmacology and Toxicology* 2006, *46*, 101-122.
3. Pacher, P.; Batkai, S.; Kunos, G., The endocannabinoid system as an emerging target of pharmacotherapy. *Pharmacological Reviews* 2006, *58* (3), 389-462.
4. Pertwee, R. G., Emerging strategies for exploiting cannabinoid receptor agonists as medicines. *British Journal of Pharmacology* 2009, *156* (3), 397-411.
5. Ben-Shabat, S.; Hanus, L. O.; Katzavian, G.; Gallily, R., New cannabidiol derivatives: Synthesis, binding to cannabinoid receptor, and evaluation of their antiinflammatory activity. *Journal of Medicinal Chemistry* 2006, *49* (3), 1113-1117.
6. Bourne, C.; Roy, S.; Wiley, J. L.; Martin, B. R.; Thomas, B. F.; Mahadevan, A.; Razdan, R. K., Novel, potent THC/anandamide (hybrid) analogs. *Bioorganic & Medicinal Chemistry* 2007, *15* (24), 7850-7864.
7. Reggio, P. H., Endocannabinoid Binding to the Cannabinoid Receptors: What Is Known and What Remains Unknown. *Current Medicinal Chemistry* 2010, *17* (14), 1468-1486.
8. Hanus, L. O.; Mechoulam, R., Novel Natural and Synthetic Ligands of the Endocannabinoid System. *Current Medicinal Chemistry* 2010, *17* (14), 1341-1359.

9. McAllister, S. D.; Rizvi, G.; Anavi-Goffer, S.; Hurst, D. P.; Barnett-Norris, J.; Lynch, D. L.; Reggio, P. H.; Abood, M. E., An aromatic microdomain at the cannabinoid CB1 receptor constitutes an agonist/inverse agonist binding region. *Journal of Medicinal Chemistry* 2003, *46* (24), 5139-5152.
10. Shim, J. Y.; Welsh, W. J.; Howlett, A. C., Homology model of the CB1 cannabinoid receptor: Sites critical for nonclassical cannabinoid agonist interaction. *Biopolymers* 2003, *71* (2), 169-189.
11. Salo, O. M. H.; Lahtela-Kakkonen, M.; Gynther, J.; Jarvinen, T.; Poso, A., Development of a 3D model for the human cannabinoid CB1 receptor. *Journal of Medicinal Chemistry* 2004, *47* (12), 3048-3057.
12. Durdagi, S.; Papadopoulos, M. G.; Zoumpoulakis, P. G.; Koukoulitsa, C.; Mavromoustakos, T., A computational study on cannabinoid receptors and potent bioactive cannabinoid ligands: homology modeling, docking, de novo drug design and molecular dynamics analysis. *Molecular Diversity* 2010, *14* (2), 257-276.
13. Montero, C.; Campillo, N. E.; Goya, P.; Paez, J. A., Homology models of the cannabinoid CB1 and CB2 receptors. A docking analysis study. *European Journal of Medicinal Chemistry* 2005, *40* (1), 75-83.
14. Gonzalez, A.; Duran, L. S.; Araya-Secchi, R.; Garate, J. A.; Pessoa-Mahana, C. D.; Lagos, C. F.; Perez-Acle, T., Computational modeling study of functional microdomains in cannabinoid receptor type 1. *Bioorganic & Medicinal Chemistry* 2008, *16* (8), 4378-4389.

15. Chang, C-E.; Ai, R.; Gutierrez, M.; Marsella, M. J., Homology Modeling of Cannabinoid Receptor: Discovery of Cannabinoid Analogues for Therapeutic Use. In *Computational Drug Discovery and Design*, Baron, R., (Ed.) Springer: 2012; Vol. 819.
16. Tuccinardi, T.; Ferrarini, P. L.; Manera, C.; Ortore, G.; Saccomanni, G.; Martinelli, A., Cannabinoid CB2/CB1 selectivity. Receptor modeling and automated docking analysis. *Journal of Medicinal Chemistry* 2006, *49* (3), 984-994.
17. Shim, J. Y., Transmembrane Helical Domain of the Cannabinoid CB1 Receptor. *Biophysical Journal* 2009, *96* (8), 3251-3262.
18. Latek, D.; Kolinski, M.; Ghoshdastider, U.; Debinski, A.; Bombolewski, R.; Plazinska, A.; Jozwiak, K.; Filipek, S., Modeling of ligand binding to G protein coupled receptors: cannabinoid CB(1), CB(2) and adrenergic beta(2)AR. *Journal of Molecular Modeling* 2011, *17* (9), 2353-2366.
19. Shim, J. Y., Understanding Functional Residues of the Cannabinoid CB1 Receptor for Drug Discovery. *Current Topics in Medicinal Chemistry* 2010, *10* (8), 779-798.
20. Lynch, D. L.; Reggio, P. H., Cannabinoid CB1 receptor recognition of endocannabinoids via the lipid bilayer: molecular dynamics simulations of CB1 transmembrane helix 6 and anandamide in a phospholipid bilayer. *Journal of Computer-Aided Molecular Design* 2006, *20* (7-8), 495-509.
21. Shim, J.-Y.; Bertalovitz, A. C.; Kendall, D. A., Identification of Essential Cannabinoid-binding Domains STRUCTURAL INSIGHTS INTO EARLY DYNAMIC EVENTS

IN RECEPTOR ACTIVATION. *Journal of Biological Chemistry* 2011, 286 (38), 33422-33435.

22. Song, Z. H.; Bonner, T. I., A lysine residue of the cannabinoid receptor is critical for receptor recognition by several agonists but not WIN55212-2. *Molecular Pharmacology* 1996, 49 (5), 891-896.

23. McAllister, S. D.; Tao, Q.; Barnett-Norris, J.; Buehner, K.; Hurst, D. P.; Guarnieri, F.; Reggio, P. H.; Harmon, K. W. N.; Cabral, G. A.; Abood, M. E., A critical role for a tyrosine residue in the cannabinoid receptors for ligand recognition. *Biochemical Pharmacology* 2002, 63 (12), 2121-2136.

24. Kapur, A.; Hurst, D. P.; Fleischer, D.; Whitnell, R.; Thakur, G. A.; Makriyannis, A.; Reggio, P. H.; Abood, M. E., Mutation studies of Ser7.39 and Ser2.60 in the human CB1 cannabinoid receptor: Evidence for a serine-induced bend in CB1 transmembrane helix 7. *Molecular Pharmacology* 2007, 71 (6), 1512-1524.

25. Sitkoff, D. F.; Lee, N.; Ellsworth, B. A.; Huang, Q.; Kang, L. Y.; Baska, R.; Huang, Y. T.; Sun, C. Q.; Pendri, A.; Malley, M. F.; Scaringe, R. P.; Gougoutas, J. Z.; Reggio, P. H.; Ewing, W. R.; Pellemounter, M. A.; Carlson, K. E., Cannabinoid CB(1) receptor ligand binding and function examined through mutagenesis studies of F200 and S383. *European Journal of Pharmacology* 2011, 651 (1-3), 9-17.

26. Howlett, A. C.; Barth, F.; Bonner, T. I.; Cabral, G.; Casellas, P.; Devane, W. A.; Felder, C. C.; Herkenham, M.; Mackie, K.; Martin, B. R.; Mechoulam, R.; Pertwee, R.

- G., International Union of Pharmacology. XXVII. Classification of cannabinoid receptors. *Pharmacological Reviews* 2002, 54 (2), 161-202.
27. Ai, R.; Fatmi, M. Q.; 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.
28. Bokoch, M. P.; Zou, Y. Z.; Rasmussen, S. G. F.; Liu, C. W.; Nygaard, R.; Rosenbaum, D. M.; Fung, J. J.; Choi, H. J.; Thian, F. S.; Kobilka, T. S.; Puglisi, J. D.; Weis, W. I.; Pardo, L.; Prosser, R. S.; Mueller, L.; Kobilka, B. K., Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature* 2010, 463 (7277), 108-U121.
29. Jaakola, V. P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y. T.; Lane, J. R.; Ijzerman, A. P.; Stevens, R. C., The 2.6 Angstrom Crystal Structure of a Human A(2A) Adenosine Receptor Bound to an Antagonist. *Science* 2008, 322 (5905), 1211-1217.
30. Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J. H.; Zhang, Z.; Miller, W.; Lipman, D. J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 1997, 25 (17), 3389-3402.
31. Soding, J., Protein homology detection by HMM-HMM comparison (vol 21, pg 951, 2005). *Bioinformatics* 2005, 21 (9), 2144-2144.
32. Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M., Crystal

structure of rhodopsin: A G protein-coupled receptor. *Science* 2000, 289 (5480), 739-745.

33. Okada, T.; Sugihara, M.; Bondar, A. N.; Elstner, M.; Entel, P.; Buss, V., The retinal conformation and its environment in rhodopsin in light of a new 2.2 angstrom crystal structure. *Journal of Molecular Biology* 2004, 342 (2), 571-583.

34. Rasmussen, S. G. F.; Choi, H.-J.; Rosenbaum, D. M.; Kobilka, T. S.; Thian, F. S.; Edwards, P. C.; Burghammer, M.; Ratnala, V. R. P.; Sanishvili, R.; Fischetti, R. F.; Schertler, G. F. X.; Weis, W. I.; Kobilka, B. K., Crystal structure of the human beta(2) adrenergic G-protein-coupled receptor. *Nature* 2007, 450 (7168), 383-U4.

35. Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H. J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C., High-resolution crystal structure of an engineered human beta(2)-adrenergic G protein-coupled receptor. *Science* 2007, 318 (5854), 1258-1265.

36. Notredame C.; Higgins DG.; Heringa J., T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* 2000, 302 (1), 205-17.

37. Jones, D. T., Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology* 1999, 292 (2), 195-202.

38. Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T., The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 2006, 22 (2), 195-201.

39. Torda, A. E.; Procter, J. B.; Huber, T., Wurst: a protein threading server with a structural scoring function, sequence profiles and optimized substitution matrices. *Nucleic Acids Research* 2004, 32, W532-W535.
40. Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L., Improved prediction of protein side-chain conformations with SCWRL4. *Proteins-Structure Function and Bioinformatics* 2009, 77 (4), 778-795.
41. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 1996, 14 (1), 33-38.
42. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 2005, 26 (16), 1781-1802.
43. Ponder, J. W.; Case, D. A., Force fields for protein simulations. *Advances in Protein Chemistry* 2003, 66, 27-85.
44. Bertalovitz, A. C.; Ahn, K. H.; Kendall, D. A., Ligand Binding Sensitivity of the Extracellular Loop Two of the Cannabinoid Receptor 1. *Drug Development Research* 2010, 71 (7), 404-411.
45. Shim, J.-Y.; Rudd, J.; Ding, T. T., Distinct second extracellular loop structures of the brain cannabinoid CB(1) receptor: Implication in ligand binding and receptor function. *Proteins-Structure Function and Bioinformatics* 2011, 79 (2), 581-597.

46. Kairys, V.; Gilson, M. K., Enhanced docking with the mining minima optimizer: Acceleration and side-chain flexibility. *Journal of Computational Chemistry* 2002, *23* (16), 1656-1670.
47. Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B. P.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvai, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A. and Kollman, P. A. (2010), AMBER 11, University of California, San Francisco.
48. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry* 1998, *19* (14), 1639-1662.
49. Trott, O.; Olson, A. J., Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *Journal of Computational Chemistry* 2010, *31* (2), 455-461.
50. Irwin, J. J.; Shoichet, B. K., ZINC - A free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling* 2005, *45* (1), 177-182.

51. Picone, R. P.; Khanolkar, A. D.; Xu, W.; Ayotte, L. A.; Thakur, G. A.; Hurst, D. P.; Abood, M. E.; Reggio, P. H.; Fournier, D. J.; Makriyannis, A., (-)-7'-isothiocyanato-11-hydroxy-1',1'-dimethylheptylhexahydrocannabinol (AM841), a high-affinity electrophilic ligand, interacts covalently with a cysteine in helix six and activates the CB1 cannabinoid receptor. *Molecular Pharmacology* 2005, 68 (6), 1623-1635.
52. van der Stelt, M.; van Kuik, J. A.; Bari, M.; van Zadelhoff, G.; Leeftang, B. R.; Veldink, G. A.; Finazzi-Agro, A.; Vliegthart, J. F. G.; Maccarrone, M., Oxygenated metabolites of anandamide and 2-arachidonoylglycerol: Conformational analysis and interaction with cannabinoid receptors, membrane transporter, and fatty acid amide hydrolase. *Journal of Medicinal Chemistry* 2002, 45 (17), 3709-3720.
53. Shim, J. Y.; Howlett, A. C., WIN55212-2 docking to the CB1 cannabinoid receptor and multiple pathways for conformational induction. *Journal of Chemical Information and Modeling* 2006, 46 (3), 1286-1300.
54. McAllister, S. D.; Hurst, D. P.; Barnett-Norris, J.; Lynch, D.; Reggio, P. H.; Abood, M. E., Structural mimicry in class A G protein-coupled receptor rotamer toggle switches - The importance of the F3.36(201)/W6.48(357) interaction in cannabinoid CB1 receptor activation. *Journal of Biological Chemistry* 2004, 279 (46), 48024-48037.
55. Reggio, P. H., Toward the Design of Cannabinoid CB1 Receptor Inverse Agonists and Neutral Antagonists. *Drug Development Research* 2009, 70 (8), 585-600.

56. Hurst, D. P.; Lynch, D. L.; Barnett-Norris, J.; Hyatt, S. M.; Seltzman, H. H.; Zhong, M.; Song, Z. H.; Nie, J. J.; Lewis, D.; Reggio, P. H.,
N-(Piperidin-1-yl)-5-(4-chlorophenyl)-1-(2,4-dichlorophenyl)-4-methyl-1H
-pyrazole-3-carboxamide (SR141716A) interaction with LYS 3.28(192) is crucial for its
inverse agonism at the cannabinoid CB1 receptor. *Molecular Pharmacology* 2002, *62*
(6), 1274-1287.
57. Fay, J. F.; Dunham, T. D.; Farrens, D. L., Cysteine residues in the human
cannabinoid receptor: Only C257 and C264 are required for a functional receptor,
and steric bulk at C386 impairs antagonist SR141716A binding. *Biochemistry* 2005, *44*
(24), 8757-8769.

Chapter 4

Discovery of Antimicrobial Agents for α -Subunit of Tryptophan Synthase by Molecular Docking Studies and In Vitro Assays

4.1. Abstract

Tryptophan biosynthesis pathway is essential for bacterial growth but absent in higher animals and humans. Drugs that can inhibit bacterial biosynthesis of tryptophan offer a potential new class of antibiotics. In this work, we reported a structure-based virtual screening targeting α -subunit of tryptophan synthase, an enzyme that catalyzes the final two steps in the biosynthesis of L-tryptophan. National Cancer Institute (NCI) Diversity Set of around 1800 compounds was screened by computational docking and *in silico* ADME study. Four of 28 *in silico* hit compounds showed promising activity in whole-cell minimum inhibitory concentration test. Three hits were determined to target tryptophan biosynthesis pathway. Our study shows that using virtual screening can efficiently enrich the hits rate, which supports the use of computational methods in future antimicrobial drug discovery efforts.

4.2. Introduction

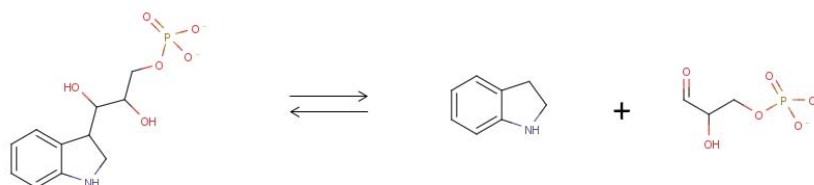
Infectious diseases are responsible for greater than 25% of the world's annual death rate and bring a large burden of disease and economic impact.¹ For billions of years, bacteria encountered various naturally occurred antimicrobial agents which made bacteria developed antibiotic resistance mechanisms to survive. Some bacteria produce biofilms to make them extremely antibiotic resistant which can be very dangerous patients.² Bacteria resistance to almost all available antimicrobial agents continues to increase progressively which affected the effectiveness of current developed antibiotics over the past 50 years.^{3,4} New antibiotics are very much needed especially for resistant pathogenic microorganisms.⁵ Besides bacteria resistance, new antibiotics are continually needed to tackle the new diseases caused by new emerging pathogens which are still not eliminated by any antibiotic, e.g., AIDS, *Pseudomonas aeruginosa*, Hantavirus.⁶

Tryptophan (Trp) is critical for all living organisms. For higher organisms, Trp is an essential amino acid that can be obtained through diet. Bacteria, yeast, molds and plants can synthesize Trp through tryptophan synthase (TRPS) biosynthesis pathway encoded by the Tryptophan Operon.⁷ Enzymes that are essential to Trp biosynthesis pathway can be new drug targets for antibiotics. Herbicides have been developed targeting TRPS for agricultural use, while not existing drugs inhibiting Trp biosynthesis pathway in human pathogens.⁸ Auxotrophic mutants of pathogenic

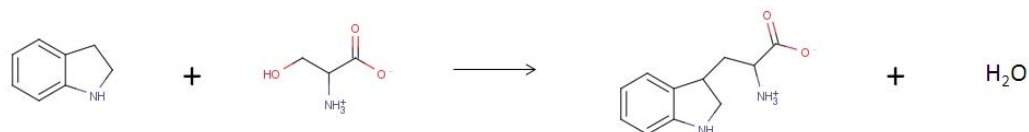
bacteria that defects in Trp biosynthesis lost virulence within a host organism. Although host organism can acquire Trp from diet, the concentration of Trp in the host is too low for pathogenic bacteria to retain virulence.^{9,10} Therefore, drugs that can inhibit the Trp biosynthesis pathway offer as a new class of antibiotics.

TRPS is a $\alpha_2\beta_2$ bienzyme complex that catalyzes the last two steps of L-tryptophan biosynthesis. The α -subunit catalyzes the cleavage of 3-indole-D-glycerol 3'-phosphate (IGP) to indole and D-glyceraldehyde 3'-phosphate (G3P) (Scheme 4.1A). Indole is then channeled to β -subunit via a 25 Å long tunnel. Within the β -site, indole and L-Serine react in the presence of pyridoxal 5'-phosphate (PLP) cofactor in a two-stage reaction to produce L-Tryptophan (Scheme 4.1B). In Stage I, L-Ser forms an external aldimine, E(Aex₁), which converts to the α -aminoacrylate aldimine, E(A-A). In stage II, indole reacts with E(A-A) to give L-Trp.¹¹⁻¹³

Scheme 4.1. α -site reaction



Scheme 4.2. β -site reaction



The conformational dynamics of proteins play important roles in regulating ligand binding. To consider the protein flexibility, using multiple TRPS conformations could avoid single biased binding site for docking screening. Long-equilibrium molecular dynamics (MD) simulations of TRPS from previous studies provided TRPS conformational ensembles in its ligand-free state.¹⁴ Choosing representing snapshots from MD trajectories are critical. In this study, we employed T-Analyst¹⁵ to cluster MD ensembles based on the sidechain conformations of key residues in the TRPS binding site and sample representative conformations for flexible-receptor docking.

4.3. Methods

4.3.1. Protein conformations selection

Target conformations were selected both from x-ray crystal structures and MD simulations, which include ligand-free and ligand-bound states. In total, seven TRPS conformations were selected as target conformations, which include three from crystal structures and four from MD simulations. X-ray crystal structures were taken

from PDB codes: 3PR2, 3CEP, 2CLO.^{12, 16, 17} 60 ns molecular dynamics (MD) simulations were performed on TRPS $\alpha\beta$ -ligand-free state in our previous study.¹⁴ In order to select representative TRPS conformations for virtual screening, T-Analyst was used to analyze the dynamics of protein-ligand complex. Based on the side chain configurational entropy and standard deviations, large conformational changes were observed for 10 key residues (Fig. 4.1) in the protein binding site upon ligand binding. MD ensembles was then grouped based on the side chain rotations (Fig. 4.2) of the 10 key residues, which included Glu 49, Ile 64, Leu 100, Leu 127, Ile 153, Tyr 175, Leu 177, Arg 179, Thr 183 and Phe 212. Finally, four most populated clusters were selected and one representative TRPS conformation from each cluster was sampled for flexible-target docking.

4.3.2. Docking based virtual screening

NCI diversity set of around 1800 compounds were screened against seven target frames using the Linux version of Vdock.¹⁶ Two independent docking runs were carried out on each TRPS conformation with different random number seeds. While in theory, the two independent rounds of virtual screening can be replaced by a single run with larger number of runs. In practice, two independent rounds with different random number seeds increase the varieties of docking poses. A 10 Å cubic grid box was placed at the ligand binding site at α -site.

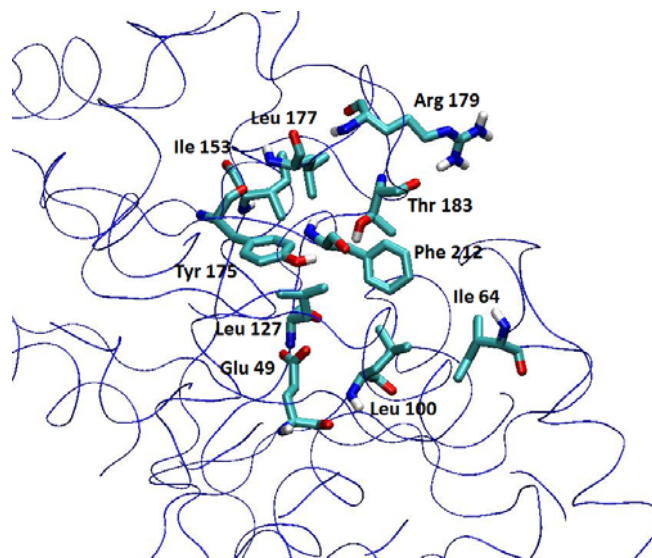


Figure 4.1 Side chain conformations of key residues in the binding site of TRPS.

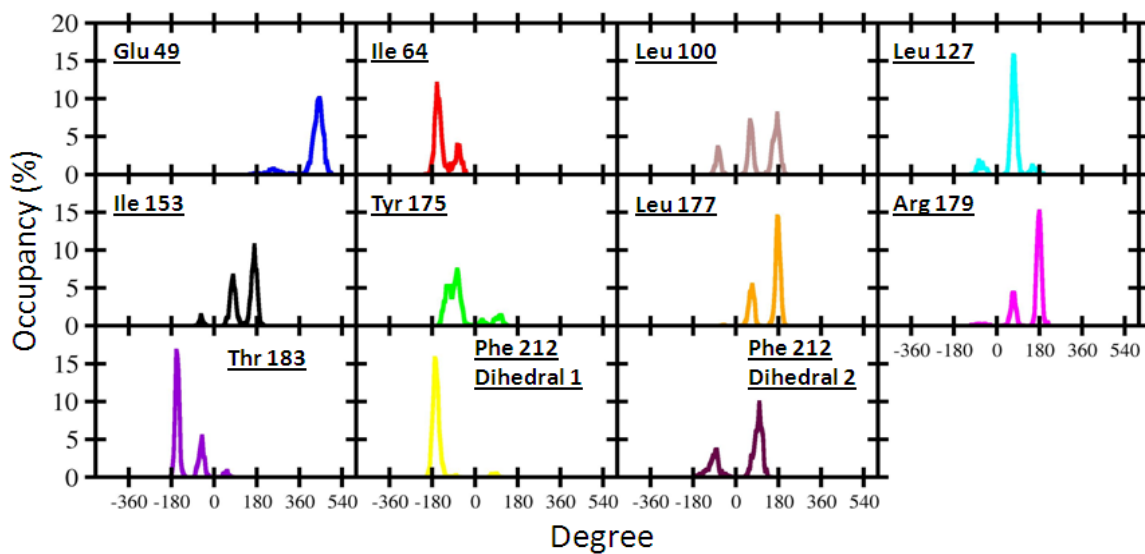


Figure 4.2 Side chain dihedral distributions of key residues in the binding site of TRPS.

Following the docking, for each virtual screening run, the docking pose with lowest interaction energy of each compound was selected. The interaction energy was then compared at the two separate runs. The run with the lower interaction energy (E_{int}) was then output for further weighted average energy calculations. As a result, docking interaction energy of each compound was weighted given the conformation occupancy in MD simulations using the equation,

$$\text{Weighted Average } E_{int} = [E_{int_3PR2} + E_{int_3CEP} + E_{int_2CLO} + 4 \times (OCC_{MD_1} \times E_{int_MD_1} + OCC_{MD_2} \times E_{int_MD_2} + OCC_{MD_3} \times E_{int_MD_3} + OCC_{MD_4} \times E_{int_MD_4})] / 7$$

where E_{int} is the docking interaction energy. OCC stands for the cluster occupancy for each representative TRPS conformation in MD trajectories. $OCC_{MD_1} = 36\%$, $OCC_{MD_2} = 11.33\%$, $OCC_{MD_3} = 44.67\%$, $OCC_{MD_4} = 8\%$

4.3.3. *In-silico ADME prediction*

The ranked NCI diversity set I was then followed by rapid absorption, distribution, metabolism, and excretion (ADME) predictions using QikProp (version 3.5, Schrödinger, LLC, New York, NY, 2012). Three descriptors were used for screening the drug-like compounds: Lipinski's rule of five, aqueous solubility and other properties that within 95% range of similar values for known drugs.

4.3.4. *In vitro* test

Given the ranked docking interaction energy and drug-likeness prediction, top 28 compounds were selected and tested by *Escherichia coli* (*E. coli*) whole-cell minimum inhibitory concentration (MIC). *E. coli* strain DH5a were grown in 5 ml of rich medium broth in 37°C shaker overnight. Bacteria cells were collected by centrifuging, followed by washing once in 5 ml of M9 minimal medium (Chart 4.1), and then grow in 5 ml of M9 minimal medium supplemented with 50 µg/ml chemical compounds. The activity of the 28 hits against whole *E. coli* growing in M9 liquid culture was determined by bacteria cell density, light absorption at OD₆₀₀. To determine specific binding and test inhibition recovery rate, tryptophan was added to the *E. coli* media at two concentrations, 0.45 mg/ml and 0.82 mg/ml, respectively. The concentrations for the chemical compounds in the minimal media with low and high concentrations of Trp are 0.18 mg/ml and 0.17 mg/ml, respectively. 0.10 mg/ml Ampicillin was used as positive control for the growth comparison of *E. coli*.

4.4. Results

The weighted average interaction energy compounds from docking results and the *in silico* ADME prediction results of the 28 top ranked chemical compounds were showed in Table 4.1. The weighted average interaction energy of one TRPS binder, F6, is -50.91 kcal/mol. Comparing to F6, the weighted average interaction energies of the

28 picked compounds are lower than -37 kcal/mol. High aqueous solubility of the compounds was preferred in this study. All 28 compounds satisfy the Lipinski's rule of

Chart 4.1 Protocols for making M9 minimal media recipe (1000ml)

-
1. Make M9 salts
add 800ml H₂O
64g Na₂HPO₄·7H₂O
15g KH₂PO₄
2.5g NaCl
5.0g NH₄Cl
Stir until dissolved
Adjust to 1000ml with distilled H₂O
Sterilize by autoclaving
-
2. Measure ~700ml of distilled H₂O (sterile)
-
3. Add 200ml of M9 salts
-
4. Add 2ml of 1m MgSO₄ (sterile)
-
5. Add 20ml of 20% glucose
-
6. Add 100ul of 1M CaCl₂ (sterile)
-
7. Adjust to 1000ml with distilled H₂O
-

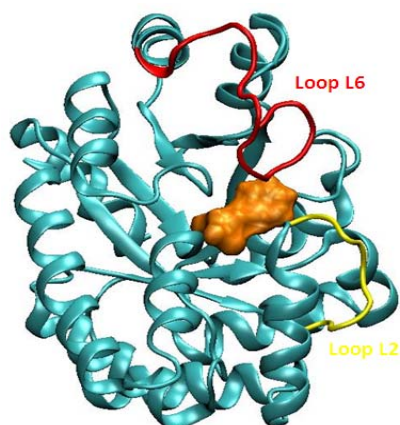


Figure 4.3. Docking pose of compound 1 in the binding site of TRPS.

Table 4.1. Weighted Average Interaction Energy and In-Silico ADME Prediction of Top Ranked Chemical Compounds.

| Compounds | W. Avg. E_{int} (kcal/mol) * | In-silico ADME Prediction** | | |
|-----------|-----------------------------------|-----------------------------|----------|--------------|
| | | QPlogS** | #stars** | RuleOfFive** |
| 1 | -41.00 | -0.82 | 0 | 0 |
| 2 | -41.20 | -1.84 | 0 | 0 |
| 3 | -42.70 | -1.13 | 0 | 0 |
| 4 | -41.40 | -1.62 | 0 | 0 |
| 5 | -41.39 | -1.42 | 0 | 0 |
| 6 | -39.30 | -0.55 | 0 | 0 |
| 7 | -40.21 | -1.86 | 0 | 0 |
| 8 | -37.72 | -0.72 | 0 | 0 |
| 9 | -37.48 | -0.20 | 0 | 0 |
| 10 | -37.67 | -0.57 | 0 | 0 |
| 11 | -40.50 | -0.23 | 0 | 0 |
| 12 | -40.58 | -1.49 | 0 | 0 |
| 13 | -42.06 | -0.49 | 0 | 0 |
| 14 | -44.61 | -1.67 | 0 | 0 |
| 15 | -42.87 | -1.58 | 0 | 0 |
| 16 | -40.34 | -1.50 | 0 | 0 |
| 17 | -41.13 | -1.69 | 0 | 0 |
| 18 | -44.65 | -0.42 | 0 | 0 |
| 19 | -48.31 | -1.73 | 0 | 0 |
| 20 | -37.19 | -0.27 | 0 | 0 |
| 21 | -40.71 | -0.73 | 0 | 0 |
| 22 | -43.70 | -1.84 | 0 | 0 |
| 23 | -38.44 | -0.70 | 0 | 0 |
| 24 | -40.92 | -0.59 | 0 | 0 |
| 25 | -39.61 | -0.90 | 0 | 0 |
| 26 | -42.10 | -1.59 | 0 | 0 |
| 27 | -38.68 | -0.07 | 0 | 0 |
| 28 | -42.20 | -0.01 | 1 | 0 |

* W. Avg. E_{int} , weighted average interaction energy. The weighted average interaction energy of one TRPS binder, F6, is -50.91 kcal/mol.

** Refer to QikProp, version 3.5, Schrödinger, LLC, New York, NY, 2012

Table 4.2. E.coli Cell Densities with 28 Compounds Measured by OD₆₀₀.

| Compounds | OD600 after 8 hrs incubation | OD600 after 27 hrs incubation | OD600 after 21.5 hrs incubation |
|-----------------------------|-----------------------------------|-----------------------------------|------------------------------------|
| - Control* | 1.609 (initial density, 0.172) | 1.945 (initial density, 0.013) | 1.323 (initial density, 0.054) |
| + Control (Ampicillin)** | 0.068 | 0.162 | 0.022 |
| 1 | - | 0.158 | - |
| 2 | 2.330 | - | - |
| 3 | 1.579 | - | - |
| 4 | - | 1.698 | - |
| 5 | 1.700 | - | - |
| 6 | 1.422 | - | - |
| 7 | 1.848 | - | - |
| 8 | - | 1.441 | - |
| 9 | - | 1.945 | - |
| 10 | - | 2.522 | - |
| 11 | - | - | - |
| 12 | - | - | - |
| 13 | 1.455 | - | - |
| 14 | 0.357 | - | 0.202 |
| 15 | - | 1.849 | - |
| 16 | - | 1.862 | - |
| 17 | 1.554 | - | - |
| 18 | 0.567 | - | 0.034 |
| 19 | - | 1.840 | - |
| 20 | - | 1.651 | - |
| 21 | - | 1.620 | - |
| 22 | 1.814 | - | - |
| 23 | - | 1.738 | - |
| 24 | - | 2.039 | - |
| 25 | 1.613 | - | - |
| 26 | 1.598 | - | - |
| 27 | - | 1.793 | - |
| 28 | - | - | 0.038 |

* Minimal media contains only E.coli. No chemical compound was added

** Minimal media contains E.coli and 0.10 mg/ml Ampicillin as positive control

five which are considered drug-like. Table 4.2 shows the cell density measured by OD₆₀₀ of the 28 compounds. Figure 4.3 displays the docking pose of compound 1 in the binding site of α -subunit of TRPS.

Four of the 28 compounds showed distinct inhibition activities in *E. coli* MIC test. After adding Trp back to the media, three compounds showed inhibition recovery (Fig 4.4). After 24 hours incubation, compared to *E. coli* negative control media (OD₆₀₀

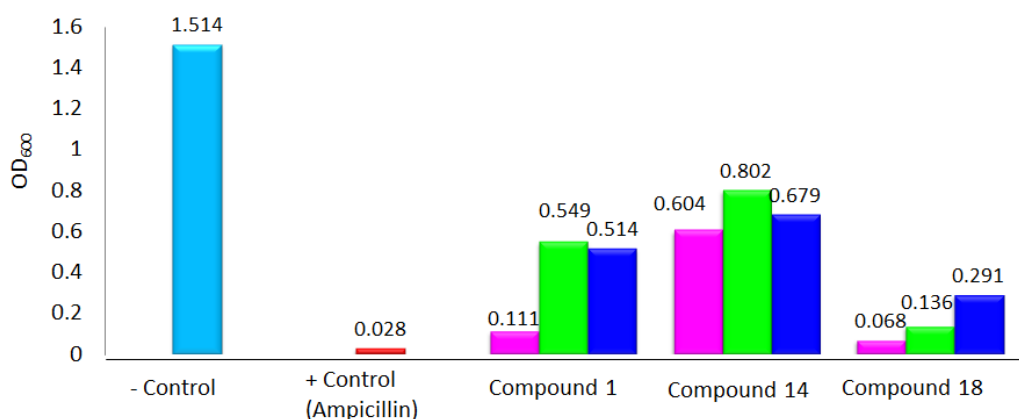


Figure 4.4. Cell Density of *E. coli* after 24 hrs incubation at 37°C

:1.514), compound 18 showed highest inhibition activity (OD₆₀₀: 0.068), followed by compound 1 (OD₆₀₀: 0.111). Although compared to the other two compounds, compound 14 had the lowest inhibition rate (OD₆₀₀: 0.604), it inhibited 60 percent the growth of *E. coli*. After adding Trp solution with two concentrations back to the

media, bacteria inhibited by compound 1 resumed highest growth rate and resulted in more than 5-fold bacteria cell density growth (OD_{600} : 0.549/0.514) compared to inhibition activity (OD_{600} :0.111). Compound 14 and 18 also showed bacteria growth recovery at different rates.

4.5. Conclusions

Three novel inhibitors that target the Trp biosynthesis pathway were discovered and identified by virtual screening and *in vitro* experiment assays. During virtual screening, protein flexibility was considered for flexible docking. Seven protein conformations were selected based on the conformations of key residues in the binding site of TRPS. During whole-cell *E. coli* MIC test, four compounds show apparent inhibition activities. To test specific binding, Trp solution was added back to the *E.coli* media to check the growth recovery rate. Three compounds were finally identified specifically targeting TRP biosynthesis pathway. NMR Spectroscopy study and fluorescence assay will be performed by our collaborators for further determination.

4.6. References

1. Morens DM, Folkers GK, Fauci AS. The challenge of emerging and re-emerging infectious diseases (vol 430, pg 242, 2004). *Nature*. 2010;463(7277).
2. Demain AL, Sanchez S. Microbial drug discovery: 80 years of progress. *Journal of Antibiotics*. 2009;62(1):5-16
3. Demain AL. Antibiotics: Natural Products Essential to Human Health. *Medicinal Research Reviews*. 2009;29(6):821-842.
4. Hancock REW. The end of an era? *Nature Reviews Drug Discovery*. 2007;6(1).
5. Spizek J, Novotna J, Rezanka T, Demain AL. Do we need new antibiotics? The search for new targets and new compounds. *Journal of Industrial Microbiology & Biotechnology*. 2010;37(12):1241-1248.
6. DaSilva E, Iaccarino M. Emerging diseases: a global threat. *Biotechnology Advances*. 1999;17(4-5).
7. 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.
8. Finn J, Langevine C, Birk I, Birk J, Nickerson K, Rodaway S. Rational herbicide design by inhibition of tryptophan biosynthesis. *Bioorganic & Medicinal Chemistry Letters*. 1999;9(16):2297-2302.

9. Smith DA, Parish T, Stoker NG, Bancroft GJ. Characterization of auxotrophic mutants of *Mycobacterium tuberculosis* and their potential as vaccine candidates. *Infection and Immunity*. 2001;69(2):1142-1150.
10. Ocallaghan D, Maskell D, Liew FY, Easmon CSF, Dougan G. CHARACTERIZATION OF AROMATIC-DEPENDENT AND PURINE-DEPENDENT *SALMONELLA-TYPHIMURIUM* - ATTENUATION, PERSISTENCE, AND ABILITY TO INDUCE PROTECTIVE IMMUNITY IN BALB/C MICE. *Infection and Immunity*. 1988;56(2):419-423.
11. Barends TRM, Dunn MF, Schlichting I. Tryptophan synthase, an allosteric molecular factory. *Current Opinion in Chemical Biology*. 2008;12(5):593-600.
12. Ngo H, Kimmich N, Harris R, et al. Allosteric regulation of substrate channeling in tryptophan synthase: Modulation of the L-Serine reaction in stage I of the ss-reaction by alpha-site ligands. Vol. 46; 2007.
13. Dunn MF. Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase henzym complex. *Archives of Biochemistry and Biophysics*. 2012;519(2):154-166.
14. Fatmi MQ, Chang C-eA. The Role of Oligomerization and Cooperative Regulation in Protein Function: The Case of Tryptophan Synthase. *Plos Computational Biology*. 2010;6(11).
15. Ai R, Fatmi MQ, Chang CEA. 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.

16. Kairys V, Gilson MK. Enhanced docking with the mining minima optimizer: Acceleration and side-chain flexibility. *Journal of Computational Chemistry*. 2002;23(16):1656-1670.

Chapter 5

Molecular Dynamics Studies of Interactions between Poly(Acrylic Acid)-Coated Fe₃O₄ Nanoparticles and Human Serum Albumin

5.1. Abstract

Nanoparticles (NPs) have been developed as important technologies to deliver conventional drugs, recombinant proteins and vaccines. One major advantage is that NPs target drug delivery to specific site of disease which could avoid undesirable side effects compared to current medical treatment. Superparamagnetic iron oxide NPs are demonstrated highly useful for cancer therapy and other biomedical applications, but their drug delivery mechanisms and health impacts are not very clear. Potential toxicity associated with NPs brought the attention recently due to their quantum size effect and the large surface area to volume ratio. Exploring the possible binding sites of NPs on the surface of protein can predict biological consequences of absorption and therefore benefit the study of nanotoxicity. In this study, we are using molecular dynamics simulations to study the possible binding sites and interactions of Fe₃O₄ NPs to human serum albumin. Two types of PAA-coatings, with similar molecular structures but resulting in different association rates, were studied to examine their

binding mechanism. The initial results preliminarily explained the mechanisms for different kinetics. Our study could help promote safer implementation of nanotechnology and reduce nanotoxicity for drug delivery and other therapeutics in the future.

5.2. Introduction

Nanoparticles (NPs) have been developed as new technologies to deliver conventional drugs, recombinant proteins, vaccines and more recently, nucleotides and become one of the most important areas of drug research.¹ This new technology emerged as an important strategy that could selectively deliver drugs to specific pathological sites in the last decade. NPs are microscopic particles, varies from 10 – 1000 nm in diameter.² Various types of NPs formulated from different materials serve as drug delivery vehicles to treat different diseases, such as polymeric NPs, solid-lipid NPs, ceramic NPs, magnetic NPs and metal based NPs. Different ways can be used to load drugs on to NPs, for example, encapsulation, entrapment and surface attachment.¹ The widespread use of NPs makes them useful therapeutical and diagnostic tools for drug delivery, ranging from cancer therapeutics, efficient imaging diagnostics, antimicrobial actions, vaccine delivery. Due to their small size, NPs can efficiently penetrate through small capillaries into individual cell and allow efficient drug accumulation at the target site. The site-specific delivery of drugs by NPs could avoid undesirable side effects and toxicity compared to current medical treatment.³

Chemotherapeutic drugs such as carboplatin, paclitaxel, doxorubicin and etoposide, have been successfully loaded onto NPs and these NP systems are very potent against various cancers.¹

Although nanotechnology is among the fastest growing areas of improving drug design and targeting, imaging and diagnosis, potential nanotoxicity related problems brought the attention recently.^{4,5} Because of the quantum size effect and the large surface area to volume ratio, NPs have unique properties which may lead to unwanted effects. For example, the increase of lung inflammation was reported due to the pro-inflammatory effects.⁶ Diesel NPs have been found to damage the cardiovascular system in a mouse model.⁷ The entrapment of NPs in liver and spleen could affect the function of the organs. A proper mechanism is needed for NP excretions from the body.⁸ A significant problem is posed when considering the dose response relationships in the toxicology of NPs. For NPs, the concentration number and the resulting large total surface area determine the interactions with biological systems; therefore to be more reasonable parameters for doses in terms of exposure.^{6,9} Unclear mechanisms of endocytosis and degradation pathways still need further understood, although they are critical for clinical transition.¹⁰ Besides, nanotoxicity associated with NPs can result in increased reactive oxygen species (ROS) generation, protein, DNA and membrane injury, oxidative stress, loss of enzyme activity, auto-antigenicity, brain and peripheral nervous system injury.¹¹

Iron oxide NP is one of the most promising nanotechnological platforms. It is specifically relevant to cancer therapy and provides an important alternative to conventional chemotherapy, radiation and surgery.¹² Iron oxide NP is usually composed of three main components: an iron core, a polymer coating and functional moieties. The iron core is superparamagnetic which is highly useful for magnetic resonance imaging (MRI) for medical diagnosis and therapeutics. Polymer coatings are usually used to cover and protect the surface of iron core from agglomeration and oxidation. They also help transform NPs into biomedical nanotools for *in vivo* applications. Different moieties attached to the coatings are used to target macromolecules, therapeutics and for imaging tags.¹³

HSA is the most abundant protein in human plasma and representing the main determinant of plasma oncotic pressure.¹⁴ HSA displays an extraordinary ligand-binding capacity and is able to bind up to nine fatty acids (FAs) (Fig. 5.1).¹⁵ It provides a depot for many endogenous compounds including nonesterified FAs, bilirubin, hemin, thyroxine, and exogenous drugs, such as warfarin, diazepam, ibuprofen, and various types of nano-particles.^{16,17} HSA affects pharmacokinetics of many drugs, provides the metabolic modification of some ligands, renders potential toxins harmless, accounts for most of the anti-oxidant capacity of human plasma, and displays (pseudo-)enzymatic properties.^{18,19}

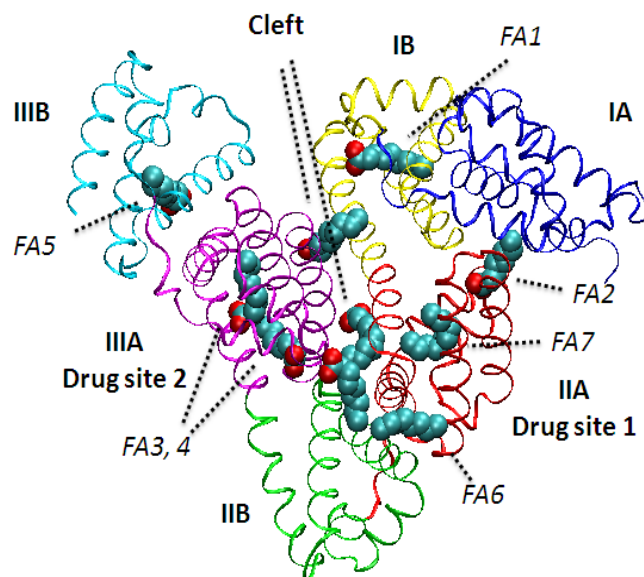


Figure 5.1. Summary of ligand binding capacity and subdomains of HSA using PDB structure 1E7E. Long-chain fatty acids are depicted in VDW representation using VMD 1.8.7. The subdomains are depicted in blue (IA), yellow (IB), red (IIA), green (IIB), purple (IIIA) and cyan (IIIB).

The binding site of PAA-coated Fe_3O_4 NPs was identified to be part of subdomain IIIA of HSA, the so-called drug binding site 2. Competitive binding between the corresponding drug, ibuprofen, and the NPs was observed by our collaborator.²⁰ The drug binding site 2 is a pre-formed hydrophobic cavity with distinct polar features and has a single main polar patch at the pocket entrance. Drugs with acidic or electronegative features are selective for this binding site.²¹ The binding can improve the solubility of the drug in serum and help its transportation in the circulation system.²⁰

Two types of PAA coatings (Fig. 5.2) for Fe_3O_4 NPs were synthesized by our collaborator. Both types of PAA coatings contain long carboxyl chains with negative charges at pH 7.0. The only difference between two types of PAA coatings lies at the end of the long chains where one has a ring scaffold and the other is with a hydroxyl group. The dissociation constant (K_D) of A and B series PAA coatings to HSA are 5.12 μM and 1.44 μM , respectively. Interestingly, the association rates are more than 250-fold difference between two PAA-coated NPs. The B-Series PAA NPs bind with fast kinetics, while A-series PAA NPs bind with slow kinetics.

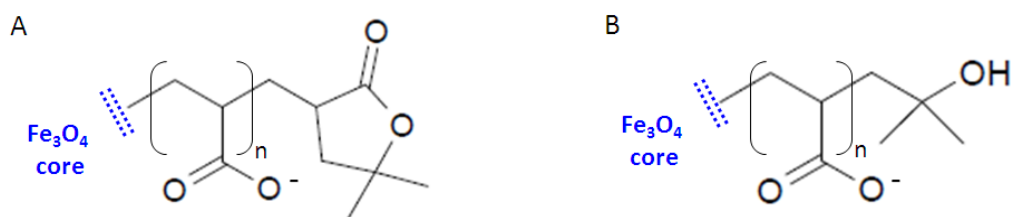


Figure 5.2 Illustration of structures of NPs and two types of coatings. A. A series PAA-NPs. B: B series PAA-NPs.

5.3. Methods

5.3.1. Preparing PAA coatings

Molecular docking was used to obtain initial PAA-HSA complex conformations for MD simulations by Vdock.²² 2D structures of PAA coatings with 3 different chain lengths were sketched with MDL ISIS/Draw 2.5. VEGA ZZ 2.3.1 was used to generate

3D structures and Vconf 2.0 was applied for conformational searches to identify the most stable conformations of PAA coatings.^{23,24} Based on the binding modes and interaction energy, complexes containing A and B series PAA coatings with 2 carboxyl groups in the long chains (Fig. 5.3 A1, B1) were initially selected for MD simulations.

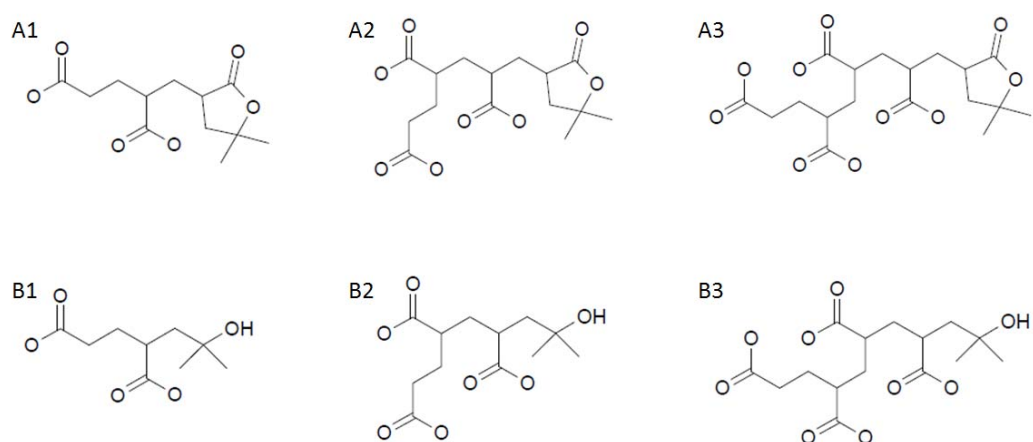


Figure 5.3 Structures of (A1-3) A and (B1-3) B series PAA coatings for docking; with (A1, B1) two, (A2, B2) three or (A3, B3) four hydroxyl groups in the long PAA chains.

5.3.2. MD simulations

10ns Molecular dynamics (MD) simulations were carried out with Amber11 and NAMD2.6 simulation packages.^{25,26} The initial structural coordinate for HSA was obtained from crystal structure (PDB code: 1AO6). The Amber ff99SB force field and general Amber force field (GAFF) were applied to parameterize the HSA protein and

PAA coatings, respectively.^{27,28} The initial PAA-HSA conformation for MD simulations was obtained by docking PAA ligand to the HSA binding site 2 using Vdock.²⁴ Vcharge 1.0 was used to assign formal charges to PAA coatings.²⁹ The Antechamber package in AmberTools was used to assign parameters to the ligand–protein complex and create topology and coordinate files.³⁰ 17 disulfide bridges were assigned using Tleap. Energy minimization was carried out for the system by the sander program.³⁰ HSA has a negative net charge per molecule of about 15, at pH 7.0. To electronically neutralize the system, 80 Na⁺ and 65 Cl⁻ ions were added based on the ionic concentration in human serum. All complexes were solvated in a rectangular box of 12 Å TIP3P water with the tleap program, and the system has about 88000 atoms.³¹ The initial energy minimization for water molecules was carried out by the sander program. NAMD 2.6 was then used for sequential steps of further energy minimization, equilibration, and production runs. During equilibration, the system was gradually heated from 50 K to 300 K, in a 50 K interval, for a total of 525 ps. The resulting trajectories were saved every 1 ps. The NPT ensemble was applied, and periodic boundary conditions were used throughout the MD simulations. A temperature of 300 K was maintained with use of a Langevin thermostat, with a damping constant of 2 ps⁻¹, and the hybrid Nose-Hoover Langevin piston method was used to control the pressure at 1 atm. The SHAKE algorithm was used to constrain the length of all bonds involving hydrogen atoms during MD simulations with a time step of 2 fs.³² Particle Mesh Ewald was used to consider the

long-range electrostatic interactions beyond cutoff limit.³³ The nonbonded interactions were truncated at a distance of 14 Å with a switching beginning at 12 Å. T-Analyst was employed to analyze MD simulations results and compare conformational changes of HSA in free and bound states.³⁴

5.4. Preliminary Results

5.4.1. The binding poses of PAA coatings

The initial HSA-PAA complex for MD simulations was obtained by molecular docking. The center of the binding site is shown in Figure 5.4. Table 5.1 displays the detailed docking results of PAA with 3 different chain lengths to HSA. Based on the docking poses and the PAA orientations in the protein binding site, PAA coatings with two hydroxyl groups were initially selected for MD simulations.

5.4.2. A and B-series PAA coatings in HSA binding site

PAA coatings in drug binding site 2 of HSA are surrounded by all six helices of subdomain IIIA (Fig. 5.5). The binding site is composed of a large pre-formed hydrophobic cavity with distinct polar features. (Fig. 5.6) The entrance to the pocket is more exposed to the solvent which may facilitate the binding of large NPs.

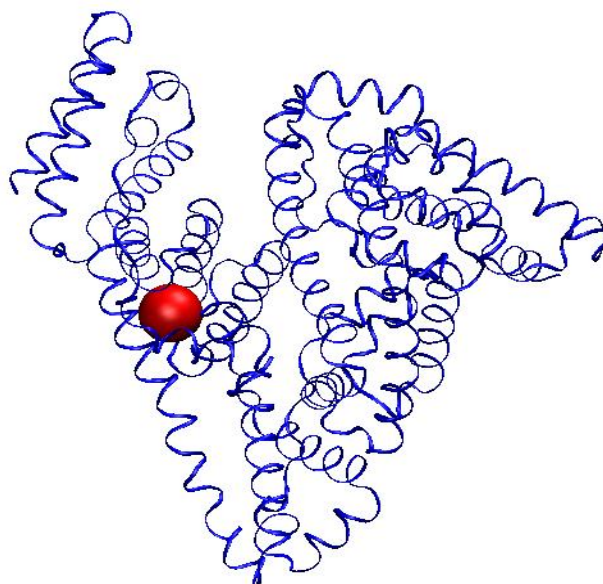





Figure 5.4. The center of drug binding site 2 of HSA for molecular docking

The binding pocket has a single main polar patch, including Arg 410, Tyr 411, Lys 414 and Ser 489, which locates closely at one side of the entrance. In HSA free state, Arg 410 points out of the binding site and partially blocks the entrance. Tyr 411 also presents at the entrance to the binding site. While in HSA bound state, these two residues rotated and, with all polar patch residues, formed hydrogen bonds with the PAA coating and stabilized its binding. The basic and polar residues on the hydrophobic interior walls make the drug binding site 2 selective for electronegative compounds. PAA contains large numbers of carboxyl groups with negative charges and help itself adapt to the binding site.

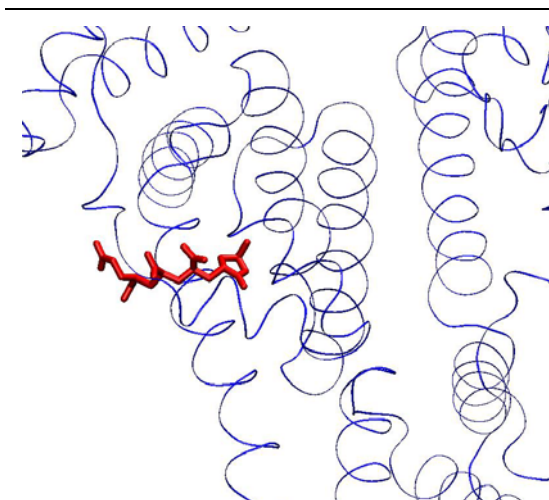
Table 5.1. Binding Modes and Docked Interaction Energy of A and B-series PAA NPs with Three Chain Lengths.

| | |
|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | <p>A-series PAA Two hydroxyl groups in PAA chain</p> <p>Correct orientation $E_{int}^* = -36.32$ kcal/mol</p> |
|  | <p>B-series PAA Two hydroxyl groups in PAA chain</p> <p>Correct orientation $E_{int} = -30.18$ kcal/mol</p> |
|  | <p>A-series PAA Three hydroxyl groups in PAA chain</p> <p>Correct orientation $E_{int} = -40.89$ kcal/mol</p> |



B-series PAA
Three hydroxyl groups in PAA chain

Incorrect orientation
 $E_{\text{int}} = -47.00$ kcal/mol



A-series PAA
Four hydroxyl groups in PAA chain

Correct orientation
 $E_{\text{int}} = -43.71$ kcal/mol



B-series PAA
Four hydroxyl groups in PAA chain

Incorrect orientation
 $E_{\text{int}} = -32.05$ kcal/mol

* E_{int} , Interaction energy output from docking results

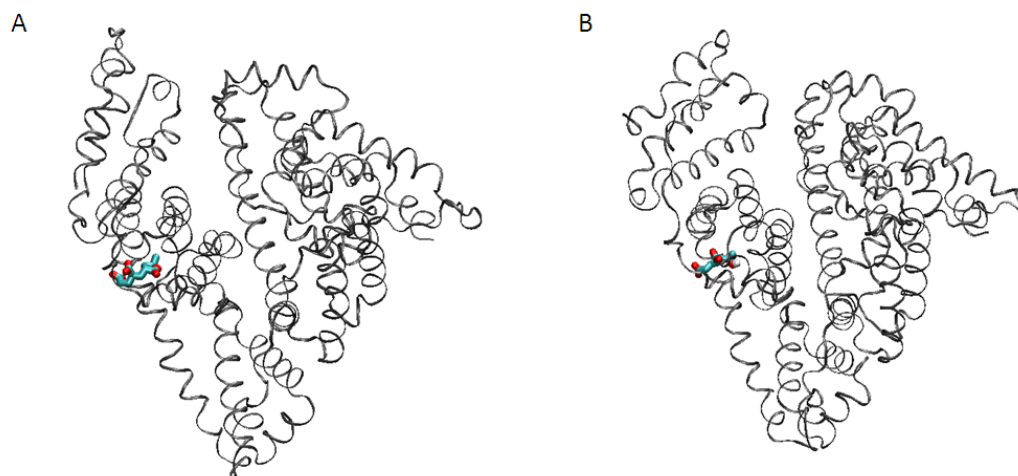


Figure 5.5. Interactions of (A) A-series and (B) B-series PAA coatings with drug binding site 2 of HSA during 10 ns MD simulations.

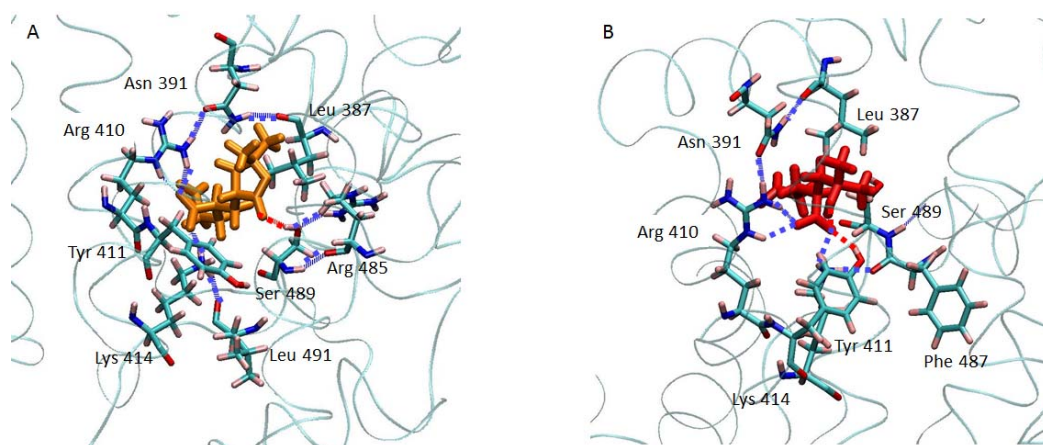


Figure 5.6. Interactions and hydrogen bond networks in the binding site of HSA upon (A) A-series PAA and (B) B-series PAA binding.

5.4.3. Protein flexibility analysis

T-Analyst was used to analyze the protein flexibility. The magnitude of configuration entropy computed from protein backbone and sidechain dihedral degrees of freedom provides a direct way to examine protein dynamics. Figure 5.7 illustrated the flexible and rigid regions corresponding to the entropy of each backbone Φ and ψ angle that is larger than 0 kcal/mol or less than -0.3 kcal/mol, respectively, in both HSA free and bound states. The protein flexibility study results can be provided for future Brownian dynamics studies.

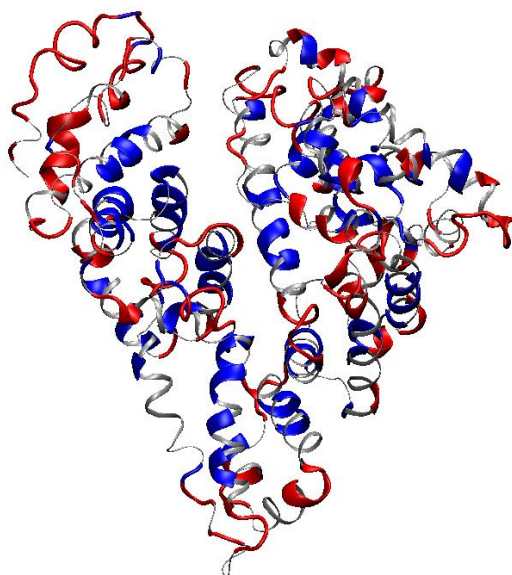


Figure 5.7. Illustration of flexible and rigid regions of HSA. Flexible region: entropy of each backbone Φ and ψ angle that is larger than 0 kcal/mol, in red; rigid region: entropy of each backbone Φ and ψ angles that is less than -0.3 kcal/mol in both HSA free and bound states, in blue.

By analyzing entropy changes between A or B-series PAA bound state and free protein state, protein dynamics changes upon ligand binding are revealed. Side chain conformations in the protein binding site were studied and compared between different states. Figure 5.8 showed the side chain entropy changes of residues that within 6 Å of PAA. In-depth study of entropy changes of each dihedral may initially explain the underlying mechanisms of association rates differences between two types of PAA coatings.

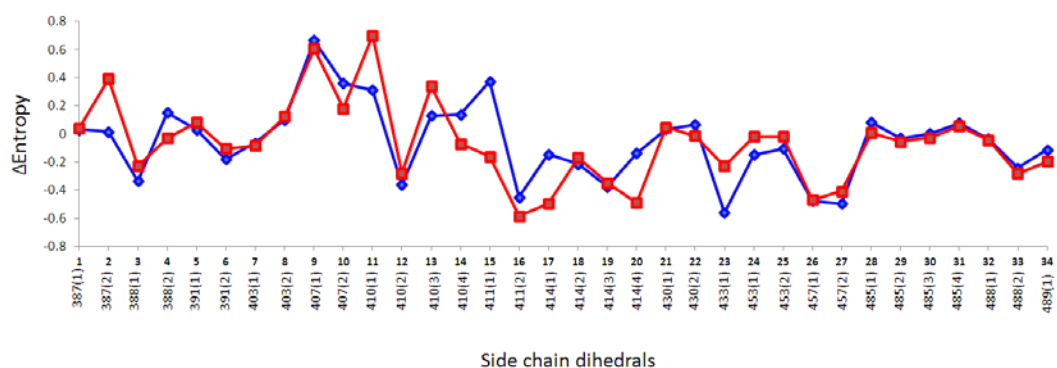


Figure 5.8 Entropy changes of side chain torsions within 6 Å of the PAA binding site. Blue dotted line indicates the entropy changes between A-series PAA bound state and free protein state. Red dotted line indicates the entropy changes between B-series PAA bound state and free protein state.

For example, figure 5.9 plotted the side chain distributions of Tyr 411 (Fig. 5.9 A1-3) and Arg 410 (Fig. 5.9 B1-3) in free state (Fig. 5.9 A1, B1), A-series PAA bound state

(Fig. 5.9 A2, B2) and B-series PAA bound state (Fig. 5.9 A3, B3) in the 10 ns MD simulations. In free state, the side chain of Tyr 411 (Fig.5.9 A1) populated around -70° ; while in A-series PAA bound state (Fig. 5.9 A2), the distribution split into two populations; one shifted to around -160° and the other was around -110° . This mechanism can be described as induced-fit. Therefore during A-series PAA binding,

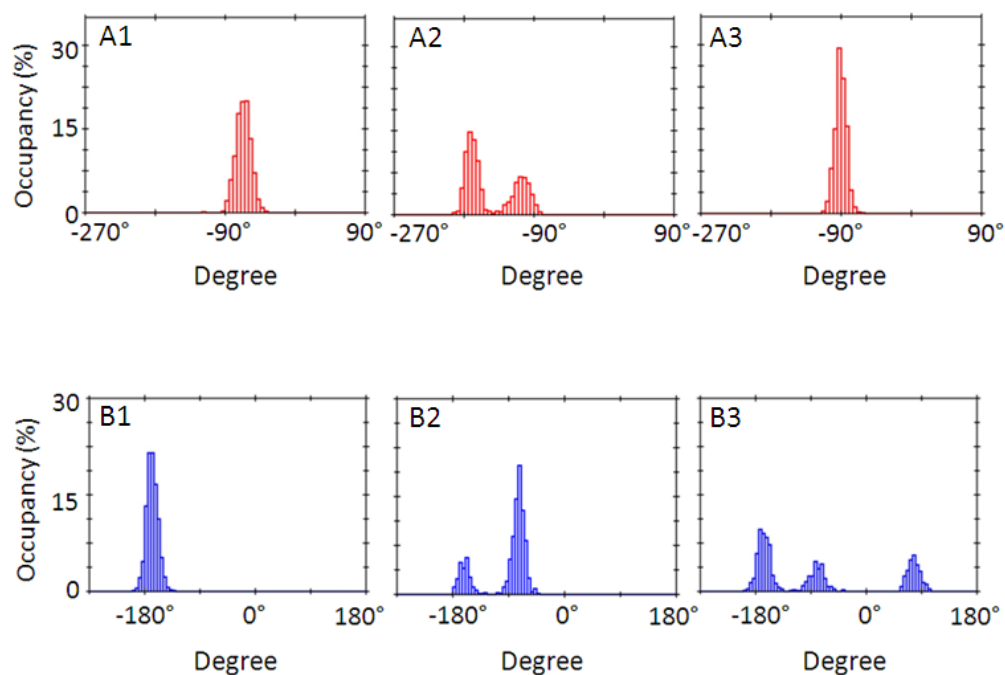


Figure 5.9. Side chain torsion distributions of (A) Arg 410 and (B) Tyr 411 in (A1, B1) free state, (A2, B2) A-series PAA-NP bound state and (A3, B3) B-series PAA-NP bound state

the side chain dihedral of Tyr 411 needs to adjust and rotate from around -70° to around -160° or -110° to form interactions with ligand. This underlying mechanism may contribute to one of the possible factors for the slow association rate. In B-series PAA bound state (Fig.5.9 A3), the population only shifted slightly and the side chain distribution almost remains the same. The pre-organized binding site for B-series PAA binding may contribute to the faster association rate. Several more residues were found with the same pattern of side chain organizations upon PAA binding.

5.5. Conclusions

The interactions between HSA and PAA-coated NPs were studied using MD simulations. Two types of PAA-coatings with similar binding affinity but large differences in association rates were simulated. T-Analyst was used to analyze and compare the MD dynamics results between A and B-series PAA bound and free protein state. Our preliminary result suggested that the pre-organized binding site for B-series PAA may contribute to the fast kinetics; while to accommodate A-series PAA binding, residues in the protein binding site requires time to rotate and adjust conformations which may partially explain the reasons for the slower association rate.

5.6. References

1. Parveen S, Misra R, Sahoo SK. Nanoparticles: a boon to drug delivery, therapeutics, diagnostics and imaging. *Nanomedicine-Nanotechnology Biology and Medicine*. 2012;8(2):147-166.
2. Sahoo SK, Labhasetwar V. Nanotech approaches to delivery and imaging drug. *Drug Discovery Today*. 2003;8(24):1112-1120.
3. Yih TC, Al-Fandi M. Engineered nanoparticles as precise drug delivery systems. *Journal of Cellular Biochemistry*. 2006;97(6).
4. Han X, Corson N, Wade-Mercer P, et al. Assessing the relevance of in vitro studies in nanotoxicology by examining correlations between in vitro and in vivo data. *Toxicology*. 2012;297(1-3).
5. Oberdoerster G. Nanotoxicology: in Vitro-in Vivo Dosimetry. *Environmental Health Perspectives*. 2012;120(1).
6. Oberdorster G. Toxicology of ultrafine particles: in vivo studies. *Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences*. 2000;358(1775).
7. Cassee FR, Campbell A, Boere AJF, et al. The biological effects of subacute inhalation of diesel exhaust following addition of cerium oxide nanoparticles in atherosclerosis-prone mice. *Environmental Research*. 2012;115.

8. Gibaud S, Demoy M, Andreux JP, Weingarten C, Gouritin B, Couvreur P. Cells involved in the capture of nanoparticles in hematopoietic organs. *Journal of Pharmaceutical Sciences*. 1996;85(9).
9. Brown DM, Wilson MR, MacNee W, Stone V, Donaldson K. Size-dependent proinflammatory effects of ultrafine polystyrene particles: A role for surface area and oxidative stress in the enhanced activity of ultrafines. *Toxicology and Applied Pharmacology*. 2001;175(3).
10. Kim J-E, Shin J-Y, Cho M-H. Magnetic nanoparticles: an update of application for drug delivery and possible toxic effects. *Archives of Toxicology*. 2012;86(5).
11. Nel A, Xia T, Madler L, Li N. Toxic potential of materials at the nanolevel. *Science*. 2006;311(5761).
12. Rosen JE, Chan L, Shieh D-B, Gu FX. Iron oxide nanoparticles for targeted cancer imaging and diagnostics. *Nanomedicine-Nanotechnology Biology and Medicine*. 2012;8(3).
13. Yigit MV, Moore A, Medarova Z. Magnetic Nanoparticles for Cancer Diagnosis and Therapy. *Pharmaceutical Research*. 2012;29(5).
14. Peters T, Jr. All about albumin: Biochemistry, genetics, and medical applications. *All about albumin: Biochemistry, genetics, and medical applications*. 1996.
15. Simard JR, Zunszain PA, Hamilton JA, Curry S. Location of high and low affinity fatty acid binding sites on human serum albumin revealed by NMR drug-competition analysis. *Journal of Molecular Biology*. 2006;361(2):336-351.

16. Aggarwal P, Hall JB, McLeland CB, Dobrovolskaia MA, McNeil SE. Nanoparticle interaction with plasma proteins as it relates to particle biodistribution, biocompatibility and therapeutic efficacy. *Advanced Drug Delivery Reviews*. 2009;61(6):428-437.
17. Kratz F. Albumin as a drug carrier: Design of prodrugs, drug conjugates and nanoparticles. *Journal of Controlled Release*. 2008;132(3):171-183.
18. Yang F, Bian C, Zhu L, Zhao G, Huang Z, Huang M. Effect of human serum albumin on drug metabolism: Structural evidence of esterase activity of human serum albumin. *Journal of Structural Biology*. 2007;157(2):348-355.
19. Fanali G, di Masi A, Trezza V, Marino M, Fasano M, Ascenzi P. Human serum albumin: From bench to bedside. *Molecular Aspects of Medicine*. 2012;33(3):209-290.
20. Li N, Zeng S, He L, Zhong W. Exploration of Possible Binding Sites of Nanoparticles on Protein by Cross-Linking Chemistry Coupled with Mass Spectrometry. *Analytical Chemistry*. 2011;83(18):6929-6934.
21. Ghuman J, Zunszain PA, Petitpas I, Bhattacharya AA, Otagiri M, Curry S. Structural basis of the drug-binding specificity of human serum albumin. *Journal of Molecular Biology*. 2005;353(1):38-52.
22. Kairys V, Gilson MK. Enhanced docking with the mining minima optimizer: Acceleration and side-chain flexibility. *Journal of Computational Chemistry*. 2002;23(16):1656-1670.

23. Pedretti A, Villa L, Vistoli G. VEGA - An open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming. *Journal of Computer-Aided Molecular Design*. 2004;18(3).
24. Chang CE, Gilson MK. Tork: Conformational analysis method for molecules and complexes. Vol. 24; 2003.
25. Picone RP, Khanolkar AD, Xu W, et al. (-)-7'-isothiocyanato-11-hydroxy-1',1'-dimethylheptylhexahydrocannabinol (AM841), a high-affinity electrophilic ligand, interacts covalently with a cysteine in helix six and activates the CB1 cannabinoid receptor. *Molecular Pharmacology*. 2005;68(6):1623-1635.
26. Phillips JC, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD. Vol. 26; 2005.
27. 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).
28. Wang JM, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. Vol. 25; 2006.
29. Chen W, Huang J, Gilson MK. Identification of symmetries in molecules and complexes. *Journal of Chemical Information and Computer Sciences*. 2004;44(4).
30. Case DA, Cheatham TE, Darden T, et al. The Amber biomolecular simulation programs. *Journal of Computational Chemistry*. 2005;26(16):1668-1688.

31. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. COMPARISON OF SIMPLE POTENTIAL FUNCTIONS FOR SIMULATING LIQUID WATER. *Journal of Chemical Physics*. 1983;79(2).
32. Ryckaert JP, Ciccotti G, Berendsen HJC. NUMERICAL-INTEGRATION OF CARTESIAN EQUATIONS OF MOTION OF A SYSTEM WITH CONSTRAINTS - MOLECULAR-DYNAMICS OF N-ALKANES. *Journal of Computational Physics*. 1977;23(3).
33. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A SMOOTH PARTICLE MESH EWALD METHOD. *Journal of Chemical Physics*. 1995;103(19).
34. Ai R, Fatmi MQ, Chang CEA. 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.

Chapter 6

Future Work of Molecular Dynamics Studies of Interactions between Poly(Acrylic Acid)-Coated Fe₃O₄ Nanoparticles and Human Serum Albumin

6.1 Further MD simulations

Further MD simulations will be performed to study the interactions between Human serum albumin (HSA) and Poly(Acrylic Acid) (PAA)-coated Nanoparticles (NPs). The chirality of PAA coating (Fig. 6.1) and more varieties of chain lengths of PAA coatings will be included to further explore the interactions. In addition to 10 ns, longer MD simulations are needed to collect enough samples to analyze the dynamics and interactions between PAA NPs and HSA.

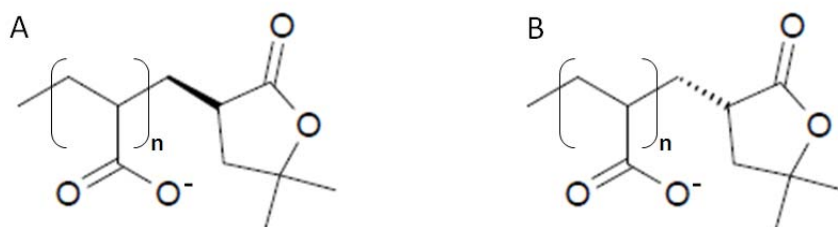


Figure 6.1. Chirality of A-series PAA coating.

6.2. Brownian dynamics simulations

To study the long-range diffusion mechanism of NPs to HSA, Brownian dynamics (BD) simulations will be employed using SDA and in-house BD program.¹ The NP diffusion mechanisms and driving forces will be studied. Since different sizes of PAA NPs may affect their diffusions, we are going to build several coarse grain models representing various NP sizes and study their diffusion mechanisms. For the coarse grain model, one bead will be assigned to represent the iron core; and for the PAA coatings, different ways to assign beads will be tested and analyzed. The rigid part of HSA, suggested in Chapter 5, will be fixed for long-range diffusion studies to avoid computational time.

Following long-range diffusions, short-range associations and interactions between HSA and PAA-coated NPs will be studied. NPs will be placed on the surface of the protein. Since the long-range diffusion results will indicate the most populated spots that NPs would like to stay on the surface of the protein, these spots will provide the initial positions of PAA-coated NPs related to HSA. During short-range BD simulations, rigid part of protein, which is based on the protein dynamics analysis by T-Analyst, will also be fixed.

6.3. Molecular docking

Since HSA has up to nine binding sites, molecular docking will be used to explore other possible bindings of NPs to HSA in addition to drug binding site 2. Exploring the possible binding sites of NPs to protein can predict biological consequences of absorption and therefore benefit the study of nanotoxicity research.

6.4. References

1. Gabdoulline RR, Wade RC. Simulation of the diffusional association of Barnase and Barstar. *Biophysical Journal*. 1997;72(5).