UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Molecular Recognition: The Role of Conformation, Dynamics, and Allosteric Regulation in the Enzymatic Function of Proteins

Permalink

https://escholarship.org/uc/item/7sj4f358

Author

Bosken, Yuliana K

Publication Date

2021

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Molecular Recognition: The Role of Conformation, Dynamics, and Allosteric Regulation in the Enzymatic Function of Proteins

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

Doctor of Philosophy

in

Biochemistry and Molecular Biology

by

Yuliana K. Bosken

June 2021

Dissertation Committee:
Dr. Chia-en A. Chang, Chairperson
Dr. Michael F. Dunn
Dr. Leonard J. Mueller

Copyright by
Yuliana K Bosken
2021

The Dissertation of Yuliana K Bosken is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Chia-en Chang, no words can express my deep and sincere gratitude for her continuous guidance, support, and patience during my graduate study. She strongly encouraged my curiosity and delicately navigated my ideas in the proper direction. I greatly appreciate her prompt responses to all my questions and concerns, as her approachability created a great learning environment.

I am also extremely grateful to my committee members Dr. Michael F. Dunn and Dr. Leonard J. Mueller, for their valuable expertise, guidance, and feedback during our meetings.

I would also like to thank past and current group members, for their support, perspective, and feedback. Special thanks to Dr. Wanli You and Dr. Zhiye Tang for providing me with mentoring, training, and advice; Dr. Timothy Cholko and Kingsley Wu for being available for long and productive discussions.

My thanks go out to our collaborators, and specifically Dr. David Boehr, for allowing me to contribute to his work.

I would like to acknowledge the Graduate Program in Biochemistry and Molecular Biology for generous resources and support during my studies.

Last but not least, I would like to thank my family and friends for their support, my mom for her invaluable help and understanding, and most importantly my little daughter for her giant heart and patience.

The text of this dissertation, in part or in full, is a reprint of the material as it appears in the following publications:

Chapter 2: O'Rourke KF, Sahu D, Bosken YK, D'Amico RN, Chang C-eA, Boehr DD. Coordinated Network Changes across the Catalytic Cycle of Alpha Tryptophan Synthase. Structure 2019;27(9):1405-1415

Chapter 2: D'amico RN, Bosken YK, O'Rourke KF, Murray AM, Chang C-eA, Boehr DD. Substitution of a surface-exposed residue involved in an allosteric network enhances tryptophan synthase function in cells. Frontiers in Molecular Biosciences 2021;8 (accepted).

Chapter 4: Bosken YK, Cholko T, Lou Y-C, Wu K-P, Chang C-eA. Insights Into Dynamics of Inhibitor and Ubiquitin-Like Protein Binding in SARS-CoV-2 Papain-Like Protease. Frontiers in Molecular Biosciences 2020;7:174.

Among them Chapter 2 includes my contribution to both articles. Chapter 4 includes the article without being rewritten.

ABSTRACT OF THE DISSERTATION

Molecular Recognition: The Role of Conformation, Dynamics, and Allosteric Regulation in the Enzymatic Function of Proteins

by

Yuliana K Bosken

Doctor of Philosophy, Graduate Program in Biochemistry and Molecular Biology
University of California, Riverside, June 2021
Dr. Chia-en A. Chang, Chairperson

The broad objectives of this work are to provide and apply computational tools to reveal how structural and dynamic protein features affect substrate binding processes and enzymatic efficiency. New findings are also used to design enzyme mutants with enhanced catalytic activity. Static structures have been available for many proteins, however, only recently, advances in experimental and computational methods allow scientists to explore the relationship between protein dynamics and conformational changes involved in function. It has long been of interest to identify which mechanisms and rules proteins utilize for allosteric and synergetic regulations. Although experiments provide information from biochemical assays, it is challenging to understand why and how enzymes/proteins have the behavior we obtain from data. Computational methods bridge the gap by allowing an atomistic level insight into conformational changes and

protein-ligand interactions. While various experimental methods can find the overall effect on catalysis from a mutation distant from the active site, experiments cannot explain why. To answer the above questions, in this study, we have employed two model systems – tryptophan synthase and papain-like protease. We have applied a combination of various computational methods such as molecular dynamics simulations, molecular docking, binding energy calculations and pair-wise force distribution, to understand how dynamics and conformational changes are induced by ligand binding and protein mutants. Thorough understanding of allosteric regulations, ligand-protein interactions, and mutation effect assists designing potent inhibitors and more efficient enzymes.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Overview

As scientist have moved away from the initial "lock and key" hypothesis to the more dynamic "hand in glove" mechanism for ligand binding,[1] questions about the relationship between structure, dynamics and function have arisen. Molecular recognition is a key factor in understanding enzyme function Three models for ligand binding have been proposed: "lock and key," induced fit, and conformational selection. The "lock and key" model assumes that the specificity of the enzyme is due to presence of enzyme pocket/binding site which is geometrically fit to accommodate the substrate.[2] The induced fit model suggests that intramolecular structural rearrangements, commonly induced by an effector molecule or the ligand itself, allow ligand binding in the active site.[3-5] Conformational selection model likewise accounts for the dynamic nature of proteins but in contrast to the induced fit model, it postulates that functional proteins exist as an ensemble of conformational states and ligand binding is possible when the protein/enzyme adopts the complementary conformation.[6] The latter two models, which lie in the foundation of the concept of allosteric regulation,[7] emerged as more structural models of various proteins were available. Several methods have been established for determining 3D structure of proteins, nucleotides, and more than 170,000 structures are currently available in the Protein Data Bank. One limitation of most popular methods (X-ray crystallography and NMR) in structural biology is that they provide partial information on conformational ensembles of biomolecules. While X-ray crystallography

can determine the most populated conformations in apo and ligand-bound protein complex, and Nuclear magnetic resonance (NMR) can capture some states of the conformational ensemble, computational methods further allow to further understand the mechanism of conformational exchange at atomistic level.[8,9]

1.2 Molecular Dynamics Basics

Molecular dynamics (MD) is a computational approach utilizing statistical mechanics equations to simulate the motions of atoms within a biomolecular system. The method was developed decades ago[10-12], however more recently, with the increase of computational power, has MD become a widely used tool in studies of biological and non-biological systems.[13-15]

MD simulations shows how the atoms in the system, represented as hard balls, move with time, typically on the nanosecond time scale. The trajectory of the atom motions is obtained by the numerical, step-by-step solution of the Newton's equations of motion (Figure 1.1). The initial position of each atom is obtained from the PDB coordinates of the protein. Initial velocities are usually determined from a random distribution based on the required temperature and corrected so there is no momentum.

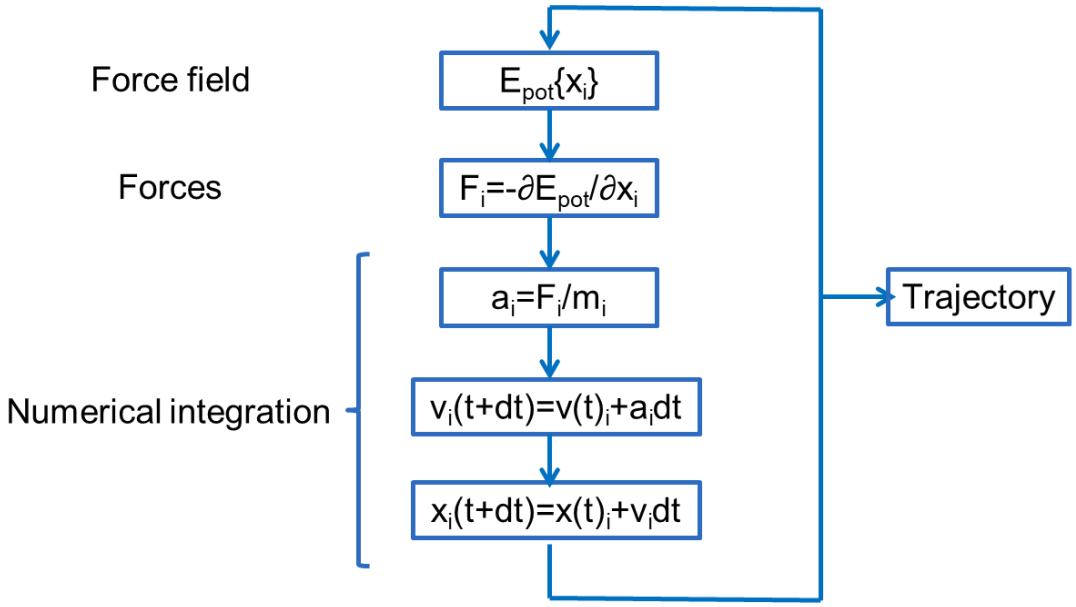


Figure 1.1. MD trajectory computation steps. The potential energy of the initial conformation is obtained, followed by force and acceleration calculations. The new velocity and position of each atom is obtained by integration.

1.2.1 Force Fields

The force field is a potential energy function and represents the sum of bonded and non-bonded terms (Figure 1.2):

$$E = \left(E_{bond} + E_{angle} + E_{torsion}\right) + \left(E_{eletrostatic} + E_{van\ der\ Waals}\right) \quad \text{Eq. 1.1}$$

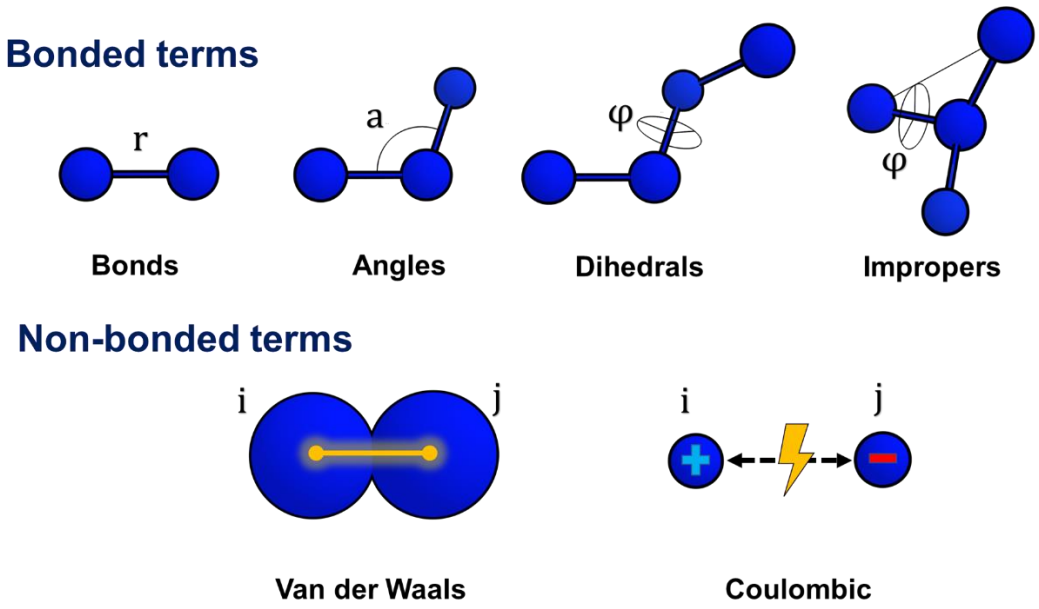


Figure 1.3 Bonded (bond, angle, torsion) and non-bonded (vdW and electrostatic) terms in potential energy function.

Each term can be further represented by the following equations:

$$E_{bond} = k_b(r - r_0)^2 \quad \text{Eq. 1.2}$$

$$E_{angle} = k_a(a - a_0)^2 \quad \text{Eq. 1.3}$$

$$E_{torsion} = \sum_p \frac{1}{2} V_n[1 + cos(p\theta - \varphi)] \quad \text{Eq. 1.4}$$

$$E_{electrostatic} = f_{ij}\left\{\epsilon_{ij}\left[\left(\frac{R_{min}}{rij}\right)^{12} - 2\left(\frac{R_{min}}{rij}\right)^{6}\right]\right\} \quad \text{Eq. 1.5}$$

$$E_{vdW} = f_{ij}\left\{\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}\right\} \quad \text{Eq. 1.6}$$

1.2.2 Temperature Control

Maintaining the correct temperature during equilibrium and production run in a constant temperature ensemble requires a thermostat. Currently, there are four thermostat algorithms used in MD simulations. Both Berendsen and Andersen thermostat controls can be thought as the systems are coupled to external heat bath which supplies or removes heat from the system as needed.[16,17] In the Andersen model, stochastic interruptions to the collision frequency will bring the system to energy levels appropriate to the required temperature, while the Berendsen algorithm will scale the velocities at each time step by a factor corresponding to the desired temperature. Nose-Hoover thermostat scales velocities by introducing a friction coefficient proportional to the degrees of freedom.[18] In this work, we use Langevin thermostat[19,20] in which atoms receive a random force and their velocities are lowered by applying constant friction force proportional to velocity.

$$ma = -\xi v + f(r) + f' \qquad \text{Eq. 1.7}$$

Where ξ is the friction force, and $f'$is the random force adding kinetic energy to the particle. The balance between the two added forces maintains the system at the desired temperature and generates the correct canonical ensemble.

1.2.3 Solvent Models

Solvent in molecular mechanics can be implicit or explicit. Implicit solvent model represents water as continuum and can speed up the simulation as interaction between solvent atoms are not calculated. Two methods are commonly used for implicit solvation.

The Poisson-Boltzmann equation (PBE) accurately models the electrostatic properties of a charged solute in an ionic solution,[21,22] however, due to its complexity and computational expense, it mostly used for energy calculation rather than for MD simulations. The generalized Born model is an approximation of PBE and models the solute atoms as charged spheres whose dielectric constant is lower than the external solvent.[23,24]

To obtain a realistic simulation of a protein system, explicit solvent model is preferred. In explicit solvation, water molecules are added to the system after minimization, and ions are used to neutralize the system charge or to obtain a desired salt concentration. There are several water models available but most popular which was also used in this work is TIP3P.[25]

1.3 Molecular Dynamics Simulation Protocol

For accurate and reproducible result, molecular dynamics requires a careful preparation of the initial structure obtained from Protein Data Bank. Figure 1.4 outlines the steps required to perform an MD simulation.

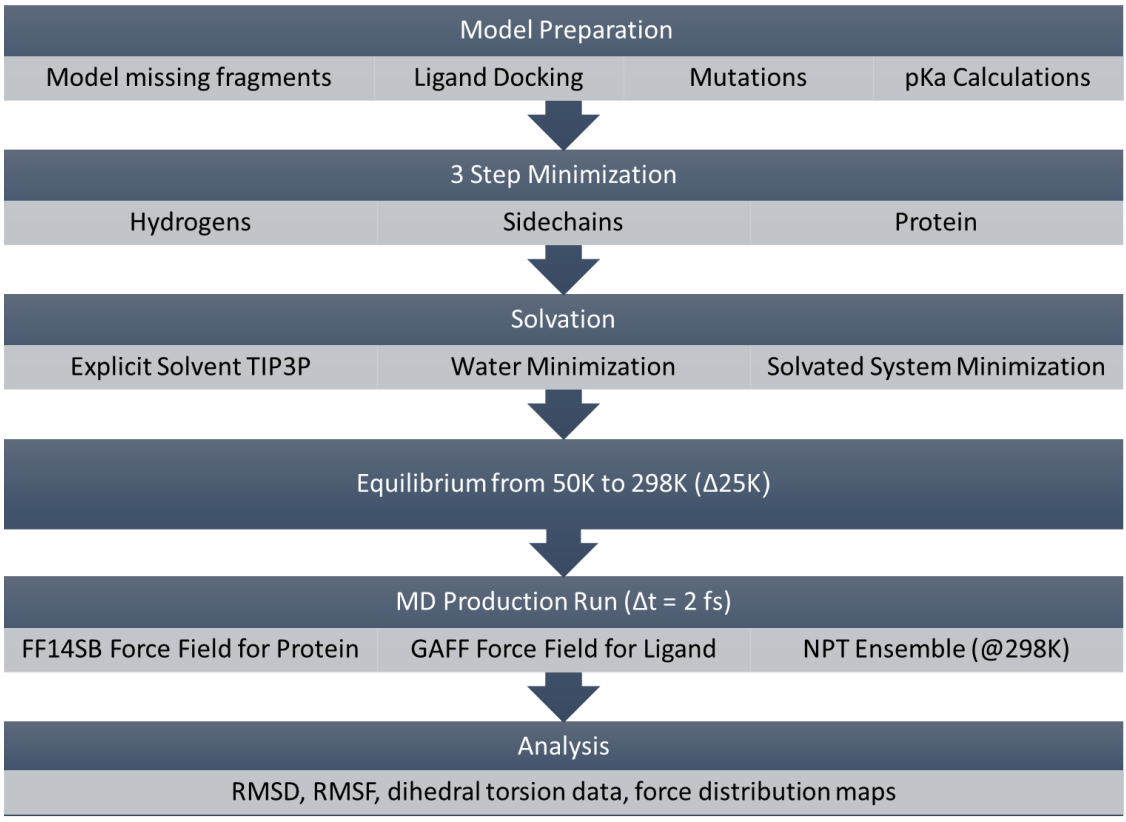


Figure 1.4. MD simulation protocol

1.3.1 Preparation of initial structure

Structures are obtained are obtained from Protein Data Bank. Coordinates for the systems used in this work are obtained from X-ray crystallography, they do not include hydrogen atoms and must be added. Some of the flexible regions such as loops may not be present either and require modeling. The most accurate method we use, is by alignment with a structure containing those coordinates. In case complete structure is not available, homology modeling can be used. Ligands, not present in the crystal structure, can also be modeled by alignment or docking. Any necessary substitutions need to be prepared

before further steps. Protonation states of polar amino acids are critical for analyzing amino acid interactions and are determined by calculating the pKa of residues.

1.3.2 Minimization and Solvation

We perform three-step minimization to ensure correct bond, angle and dihedral parameters and clear any steric clashes that may be produced during crystallization. Any modeled fragments need to be minimized before proceeding with these minimization steps. We first restrain all heavy atoms and minimize the hydrogen atoms. In the second step, we minimize the side chains of amino acids by applying the restraint to all backbone atoms. Finally, we minimize the whole protein. Once the protein is fully minimized, we proceed to solvate the system by adding water molecules and ions. The minimization is done in GB implicit solvent and therefore it is important to accurately apply the effective Born radius.

1.3.3 Equilibration and Production Run

Since the water molecules are added externally, they may not reach some regions within the protein. Equilibration of water alone is essential in order to properly position the water molecules within the solute systems and to stabilize the overall density of the system. While minimization can clear any close contacts, it does not necessary produce the most favorable residue conformations, therefore we cool the system to 50K and gradually increase the temperature at 25K intervals to allow the system to relax and slowly equilibrate to stabilize the force field parameters.

The production run is continued from the final coordinates and velocities obtained from the final equilibration at the desired temperature. The simulation is carried out at a constant pressure and temperature (NPT) with a 2 fs timestep. SHAKE algorithm is applied to restrain the heavy-atom-hydrogen bonds and Particle mesh Ewald summation is used to treat long-range electrostatics.[26-28]

1.4 Analysis

Analysis tools are selected based on our interest in the system. Common tools used for most simulation in this work are root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSD). RMSD is a measure of the deviation of the position of all atoms in reference to an initial position, while RMSF measures the displacement of a single atom averaged over time, thus showing highly fluctuating residues and regions. To quantify conformational changes, we use dihedral torsion and configurational entropy. T-Analyst[29] calculates the configurational entropy for protein residues using the Gibbs entropy formula:

$$S_{conf(i)} = -R\sum P_i \ln P_i, \qquad \text{Eq. 1.8}$$

where P_i is the probability distribution of dihedral torsion i and R is the gas constant. We use Force Distribution Analysis (FDA)[30] to identify significant contacts and interactions. FDA calculates the pair-wise forces for specified atoms over the simulation time and represents them as scalar values: negative for attraction and positive for repulsion.

1.3 References

1. Holyoak T. Molecular Recognition: Lock-and-Key, Induced Fit, and Conformational Selection. In: Roberts GCK, editor. Encyclopedia of Biophysics. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p 1584-1588.

2. Fischer E. Einfluss der Configuration auf die Wirkung der Enzyme. Berichte der deutschen chemischen Gesellschaft 1894;27(3):2985-2993.

3. Koshland DE. APPLICATION OF A THEORY OF ENZYME SPECIFICITY TO PROTEIN SYNTHESIS. Proceedings of the National Academy of Sciences of the United States of America 1958;44(2):98-104.

4. Koshland DE, Ray WJ, Erwin MJ. PROTEIN STRUCTURE AND ENZYME ACTION. Federation Proceedings 1958;17(4):1145-1150.

5. Changeux J-P, Edelstein S. Conformational selection or induced fit? 50 years of debate resolved. F1000 biology reports 2011;3:19-19.

6. Monod J, Wyman J, Changeux JP. ON NATURE OF ALLOSTERIC TRANSITIONS - A PLAUSIBLE MODEL. Journal of Molecular Biology 1965;12(1):88-&.

7. Changeux JP. 50th anniversary of the word "Allosteric". Protein Science 2011;20(7):1119-1124.

8. Ramanathan A, Savol A, Burger V, Chennubhotla CS, Agarwal PK. Protein Conformational Populations and Functionally Relevant Substates. Accounts of Chemical Research 2014;47(1):149-156.

9. Agarwal PK, Bernard DN, Bafna K, Doucet N. Enzyme Dynamics: Looking Beyond a Single Structure. Chemcatchem 2020;12(19):4704-4720.

10. Rahman A. Correlations in the Motion of Atoms in Liquid Argon. Physical Review 1964;136(2A):A405-A411.

11. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature 1977;267(5612):585-590.

12. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nature Structural Biology 2002;9(9):646-652.

13. Zahn D. Molecular dynamics simulation of ionic conductors: perspectives and limitations. Journal of Molecular Modeling 2011;17(7):1531-1535.

14. Hollingsworth SA, Dror RO. Molecular Dynamics Simulation for All. Neuron 2018;99(6):1129-1143.

15. Fu Q, Fang Z, Zhang Y, Yang L. Molecular Dynamics Simulation of a Jet in a Binary System at Supercritical Environment. Molecules 2019;24(1).

16. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. MOLECULAR-DYNAMICS WITH COUPLING TO AN EXTERNAL BATH. Journal of Chemical Physics 1984;81(8):3684-3690.

17. Andersen HC. Molecular dynamics simulations at constant pressure and/or temperature. The Journal of Chemical Physics 1980;72(4):2384-2393.

18. Nose S. A UNIFIED FORMULATION OF THE CONSTANT TEMPERATURE MOLECULAR-DYNAMICS METHODS. Journal of Chemical Physics 1984;81(1):511-519.

19. Lemons DS, Gythiel A. Paul Langevin's 1908 paper "On the Theory of Brownian Motion" ["Sur la théorie du mouvement brownien," C. R. Acad. Sci. (Paris) 146, 530–533 (1908)]. American Journal of Physics 1997;65(11):1079-1081.

20. Wu XW, Brooks BR. Self-guided Langevin dynamics simulation method. Chemical Physics Letters 2003;381(3-4):512-518.

21. Shestakov AI, Milovich JL, Noy A. Solution of the nonlinear Poisson-Boltzmann equation using pseudo-transient continuation and the finite element method. Journal of Colloid and Interface Science 2002;247(1):62-79.

22. Lu BZ, Zhang DQ, McCammon JA. Computation of electrostatic forces between solvated molecules determined by the Poisson-Boltzmann equation using a boundary element method. Journal of Chemical Physics 2005;122(21):7.

23. Tsui V, Case DA. Theory and applications of the generalized Born solvation model in macromolecular Simulations. Biopolymers 2001;56(4):275-291.

24. Onufriev AV, Case DA. Generalized Born Implicit Solvent Models for Biomolecules. In: Dill KA, editor. Annual Review of Biophysics, Vol 48. Volume 48, Annual Review of Biophysics. Palo Alto: Annual Reviews; 2019. p 275-296.

25. 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):926-935.

26. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational Physics 1977;23(3):327-341.

27. Sagui C, Pedersen LG, Darden TA. Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations. The Journal of Chemical Physics 2003.

28. Salomon-Ferrer R, Götz AW, Poole D, Grand SL, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. 2013.

29. Ai R, Qaiser Fatmi M, Chang C-eA. T-Analyst: a program for efficient analysis of protein conformational changes by torsion angles. Journal of Computer - Aided Molecular Design 2010;24(10):819.

30. Stacklies W, Seifert C, Graeter F. Implementation of force distribution analysis for molecular dynamics simulations. BMC Bioinformatics 2011;12(1):101.

Chapter 2 Modeling Coordinated Conformational Changes and Interaction Networks in Tryptophan Synthase

2.1 Introduction

The study of dynamic residue networks in proteins has emerged as indispensable part of understanding the various aspects of their structural and catalytic properties. In the classic model of allostery, enzyme function is regulated by conformational changes induced by an effector distal from the active site.[1-3] Population shifts in individual residues are possible via long- and short-range interactions and could be induced by perturbations such as ligand binding or site mutation. Propagation of the signal of such perturbation to the affected site is the basis for defining and mapping contact networks involved in the allosteric regulation of the functionality of proteins. Understanding protein residue networks allows for advances in protein engineering, function control and drug development.

We use the tryptophan synthase (TS) complex as a model system. This system has been studied for decades as a model for allosteric regulation and substrate channeling within protein complexes. The enzyme is a bienzyme nanomachine. Its catalytic activity is intimately related to allosteric signaling and metabolite transfer between its α- and β-subunits that are connected by a 25 Å-long channel. TS catalyzes the last two steps of tryptophan biosynthesis. It is a heterodimer but exists as a tetramer, and the two dimers connect linearly with two β subunits attached to each other. The enzyme α subunit catalyzes the first reaction, the cleavage of indole-3-glycerol phosphate (IGP), the indole

product is transferred through the hydrophobic channel to β active site where it is condensed with serine in a nine-step pyridoxal phosphate (PLP) dependent reaction to form tryptophan, and the glyceraldehyde-3-phosphate (G3P) is released from alpha subunit as a byproduct.[4,5] A wide range of allosteric interactions regulate the overall catalytic activity of the enzyme as to prevent the escape of the indole intermediate. Important feature of this allosteric regulation is the altering between open inactive and closed active conformations of the active sites of both subunits induced by the substrate bound to the α active site and the reaction intermediates formed at the β active site.[6]

Several computational approaches have been developed in an effort to map allosteric networks which commonly rely on tracing significant changes in residues conformation and interaction.[7-10] Similarly, in this study we evaluated pairwise forces to distinguish specific interaction changes within experimentally determined protein residue networks. In this work, we initially examined how changes in active site propagate to the surface of the enzyme. NMR chemical shift covariance analysis (CHESCA) mapped several amino acid networks in tryptophan synthase α subunit which showed coordinated response to perturbations and presence of substrate. In chemical shift covariance analysis (CHESCA), amino acid residues with covarying chemical shift changes across a set of perturbations are proposed to be involved in the same conformational change, and thus belong to the same amino acid interaction network.[11-13] Cross-correlation analysis identified a *CORE* network sustained throughout the catalytic cycle and two clusters (*CYCLE 1 and CYCLE 2*) with network changes highly coordinated across the enzyme states, that is: these residues were correlated to different residues depending on the enzyme state (Figure 2.1).

We applied molecular dynamics to simulate four states of the catalytic cycle of tryptophan synthase (TS) α subunit: apo state (E:apo), substrate-bound state (E:IGP), intermediates-bound state (E:indole:G3P), and product-bound state (E:G3P) to understand these changes. Dihedral data and force distribution analysis reveal how conformational exchange in individual residues affects interactions within these networks.

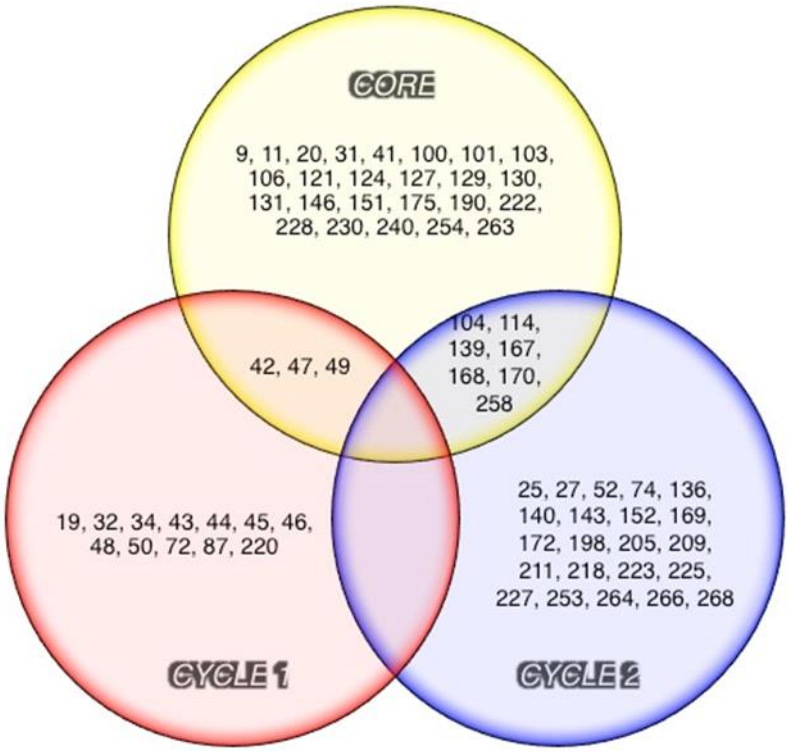


Figure 2.1. Venn diagram showing the overlap between the identified clusters. The *CORE* residues are those identified to be in the same cluster in the CHESCA-type analysis for the individual states (i.e. E:apo, E:IGP, E:indole:G3P, E:G3P). The *CYCLE 1* and *CYCLE 2* clusters do not overlap with each other, but some of these residues were identified also as *CORE* residues.

To further examine how changes outside of the active site may affect the overall dynamics within the identified residue networks, we analyzed the effects of substitution at a surface residue of α subunit, Ala198, that was found to be conformationally dynamic throughout the catalytic cycle of αTS,[14] and is distant from both the αTS active site and the αTS/βTS binding interface. Previous studies indicated that the A198G substitution induces a modest decrease in αTS catalytic activity, and changed the structure/dynamics of other network residues, according to NMR studies.[13] Contrary, a cell-based assay showed that A198W substitution induces more rapid bacterial growth in a Trp auxotrophic strain of *Escherichia coli*, likely by increasing catalytic efficiency of the full TS complex, while the turnover rate for the individual subunits remained unchanged. The simulated wild type and mutant systems show how conformation population changes induced by the distal substitution, enhance the interdomain communication and the channel accessibility.

2.2 Methods

2.2.1 Molecular Systems for Computer Simulations

The three-dimensional structure of *E. coli* tryptophan synthase α subunit was obtained from Protein Data Bank entry 1WQ5.[15] Missing and distorted portions of the structure – β2α2 loop, α2 helix and β6α6 loop (residues 52 to 77 and 175 to 195) were constructed by alignment with PDB entry 2CLK.[16]. The coordinates for the active conformation of Glu49 were taken from PDB 1QOQ.[17] Substrates were placed in the active site by alignment with PDB entries 1QOQ, 3CEP[18] and 2CLK for E:IGP, E:indole:G3P and E:G3P respectively.

Due to lack of crystal structure of dimeric *Escherichia coli* αβ TS complex, wild type and A198W mutant molecular systems were obtained from crystal structure of Salmonella Typhimurium TS PDB ID 2CLK[16] and corresponding residues were substituted to match the *E.coli* sequence. The coordinates for the active conformation of Glu 49 and the substrates were taken from PDB entry 1QOQ[17]

2.2.2 Molecular Dynamics Simulation

MD simulations were performed using standard Amber package with GPU acceleration.[19,20] The protein was parameterized using Amber Force Field FF14SB.[21] General Amber force field (GAFF) was applied to ligands and charges were assigned using AM1-bcc model.[22] All systems were prepared by a three-step minimization process (hydrogens, sidechains and all atoms), solvated with TIP3P water model with counter ions in a rectangular box with edges at minimum 12 Ǻ from any atom.[23] The solvated systems were minimized and equilibrated from 50 to 298K at 25K intervals. MD trajectories were collected over 200 ns at 1 ps interval with 2 fs timestep under constant pressure and temperature. Particle mesh Ewald was used for long-range electrostatics and SHAKE algorithm for fixed heavy atom – hydrogen bond lengths.[24,25] The systems were visualized and analyzed using Visual Molecular Dynamics (VMD)[26] and Molecular Operating Environment (MOE).[27]

2.2.2.1 Analysis of Catalytic Cycle States of α TS

For catalytic cycle states of α TS, the E:apo, E:IGP bound, E:G3P bound and the E:indole:G3P bound states, three independent 200 ns MD runs were carried out for each

system (different random seed, starting from the same equilibrated system). The trajectory output files were processed with PTRAJ software[28] to contain 20000 frames, each representing 0.01 ns timestep and dihedral data for Glu49 was collected for each state with T-Analyst software.[29] Three MD runs were analyzed for the E:apo, E:IGP bound and E:G3P bound states. For the E:indole:G3P state, only one trajectory was available in which the indole substrate remains in the active site. The three conformations of Glu49 were defined based on three distinct dihedral value sets exhibited in the analyzed trajectories. The dihedral value range criteria for each conformation were estimated based on the minimum and maximum value for each of those sets (Figure 2.2). The conformation of Glu49 with respect to the defined dihedral angle was assigned as follows: active conformation, dihedral angle values between 160° and 250°; transitional conformation, dihedral angle values between 30° and 160°; and inactive conformation, dihedral angle values between 250° and 360°. An undefined category was also included to account for seldom dihedral angle values between 0° and 30°. Occupancy was calculated as a percent relative to each individual run. The presence of the characteristic hydrogen bond for each conformation was further verified: active conformation, hydrogen bond between Glu49 and IGP; transitional conformation, hydrogen bond between Glu49 and Tyr173; inactive conformation, hydrogen bond between Glu49 and Tyr4.

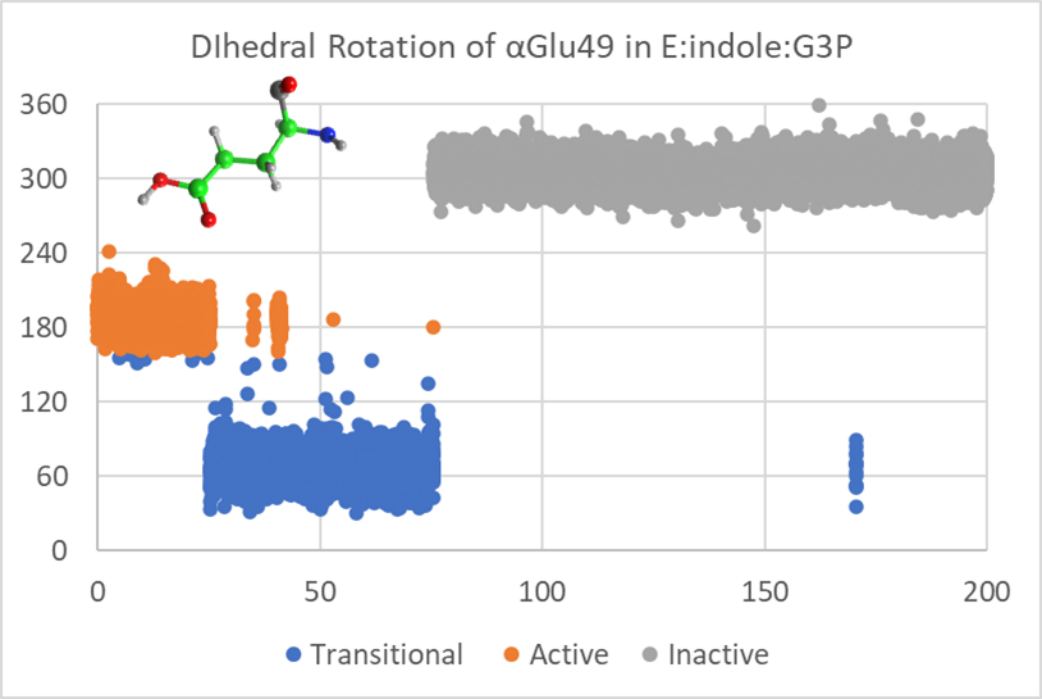


Figure 2.2 Representative plot for conformations observed for αGlu49 based on dihedral data for Cα-Cβ-Cγ-Cδ (colored in green).

The Force Distribution Analysis (FDA) tool was used to identify significant contacts and persistent interactions throughout 200 ns trajectories resaved every 0.1 ns. The pairwise atom forces are represented as a scalar value with negative values indicating attraction and positive values showing repulsion.[30]

2.2.2.2 Analysis of WT and A198W αβ TS

For the wild type and A198W αβ TS, four independent simulations were performed for each system. The trajectory output files were processed with PTRAJ software to contain 2000 frames, each representing 0.1 ns timestep. Dihedral data was collected with T-Analyst software.[28,31] Dihedral entropy was also calculated using T-Analyst and residues

with difference in entropy between the wild type and mutant higher than 0.2 kcal/mol were selected for further analysis (sidechain and backbone dihedral angles were considered). Force Distribution Analysis tool was used to identify significant contacts and persistent interactions throughout 200 ns trajectories.

2.3 Results and Discussion

2.3.1 Interaction Changes across the αTS Catalytic Cycle

CHESCA analysis suggested that changes to the amino acid interaction networks are highly coordinated and cooperative, including network changes near and far from the active site. To better understand the long-range connections in *E. coli* αTS and how they may change across the catalytic cycle, we conducted 200-ns MD simulations for the four states of the catalytic cycle of αTS: apo state, substrate-bound state, intermediates-bound state, and product-bound state (Figure 2.3). Dihedral and pair-wise force analysis of MD simulations suggest that these network changes are coupled to the reorientation of Glu49 in preparation for catalytic turnover and then product release.

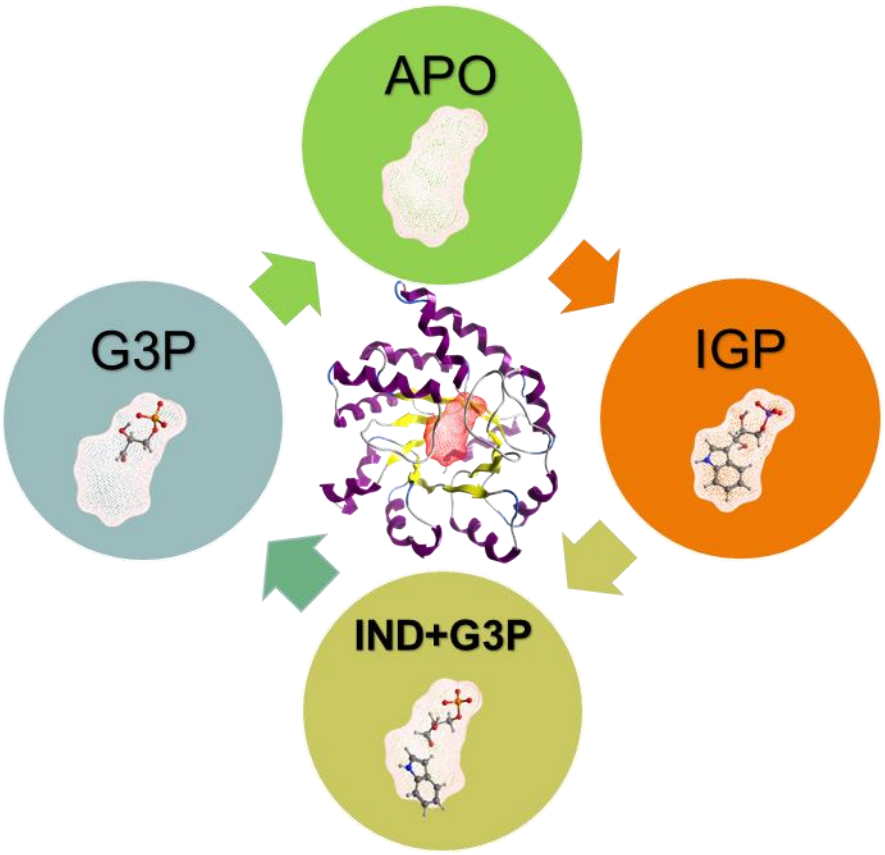


Figure 2.3. Four catalytic states of αTS: apo state, substrate-bound state (IGP), intermediates-bound state (indole and G3P), and product-bound state (G3P).

We first focused on chemical shift correlation results for residue pairs Leu11-Glu49 and Val20-Glu49 both of which change as a function of the catalytic cycle. The interactions between Leu11-Glu49 and Val20-Glu49 are strongest and weakest in different complexes, but the changes in these interactions are somehow coordinated (Figure 2.4).

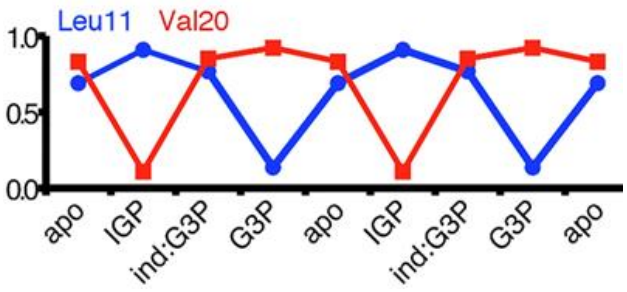


Figure 2.4. The chemical shift correlations (R) between Leu11 and Glu49 (top, blue), Val20 and Glu49 (top, red), Asp27 and Glu49 (bottom, blue), and Leu34 and Glu49 (bottom, red) change as a function of the αTS catalytic state.

Dihedral data collected from the four simulated states showσ that Glu49 exhibits three major conformations, which we designated active, inactive, and transitional (Figures 2.5 and 2.6). The three conformations of Glu49 were defined based on three distinct dihedral values sets exhibited in the MD trajectories: active, transitional, and inactive conformations had Cα-Cβ-Cγ-Cδ dihedral angle ranges of 160° to 250°, 30° to 160° and 250° to 360°, respectively (Figure 2.2). These results are consistent with previous CHESPA results also suggesting three different enzyme conformations.[12] The Glu49 conformations are also defined by different hydrogen bonding interactions (Figure 2.6A). The active conformation of Glu49 is present primarily in the E:IGP complex (Figure 2.5), in which a hydrogen bond forms between the Glu49 carboxyl group and IGP's C3' hydroxyl. In the transitional conformation, Glu49 forms a hydrogen bond instead with Tyr173, and in the inactive conformation Glu49 is associated more with the N-terminal helix through its interaction with Tyr4. It should be noted that these hydrogen bonds are not mutually exclusive. In some frames, Glu49 forms a hydrogen bond with Tyr173 and

IGP simultaneously. The E:apo and E:indole:G3P states fluctuate into all three Glu49 conformations, but the transitional and inactive conformations are favored. In E:G3P, Glu49 stays almost entirely in the inactive conformation (Figure 2.5).

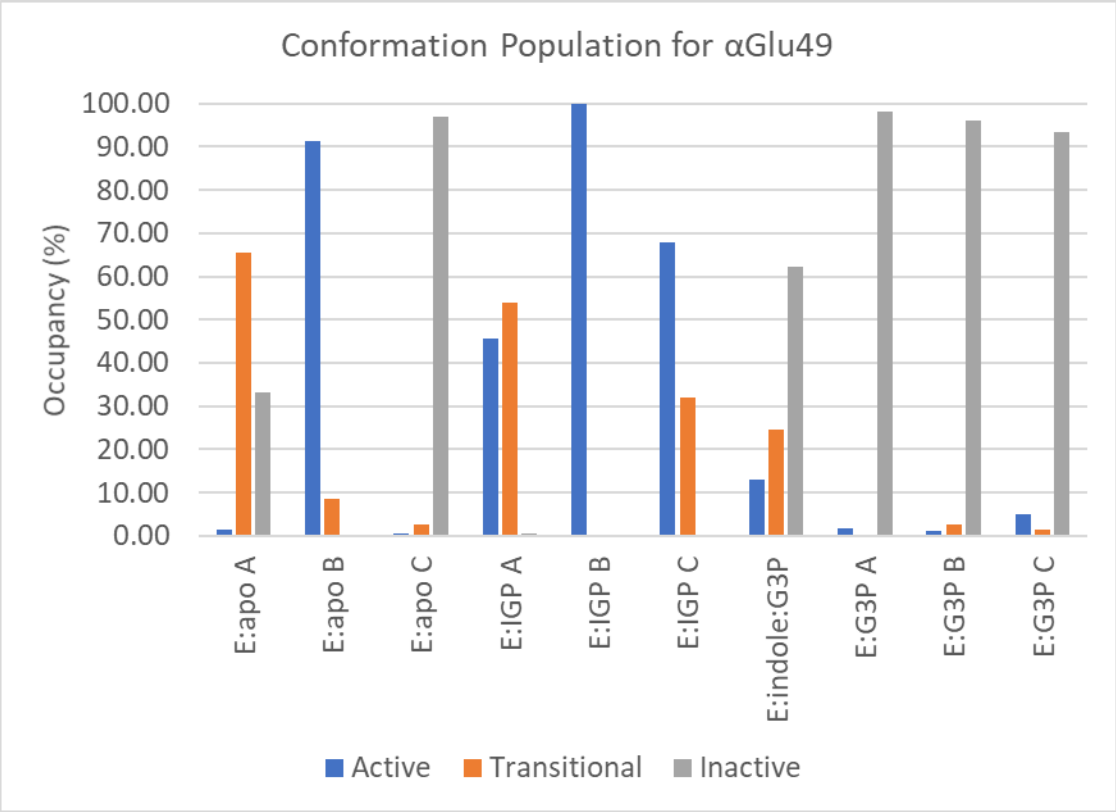


Figure 2.5. Conformation population for catalytic residue Glu49. The active conformation is preferred in the E:IGP state, and the inactive conformation is preferred in the E:G3P state. The E:apo and E:indole:G3P states access all three conformations but prefer the transitional and inactive conformations. Occupancy was calculated as a percent relative to each individual run.

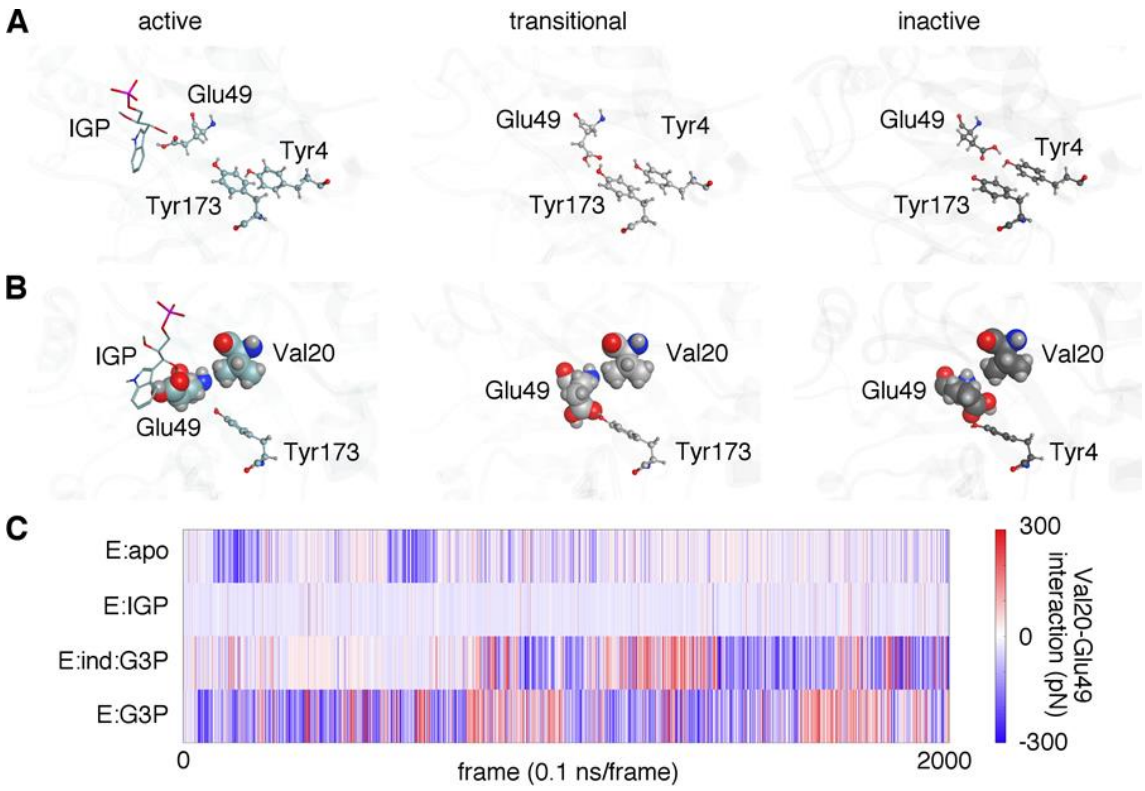


Figure 2.6. The catalytic residue Glu49 changes its interactions throughout the αTS catalytic cycle. (A) In the active, transitional, and inactive conformations, Glu49 makes hydrogen bond interactions with IGP, Tyr173 and Tyr4, respectively. (B) The conformation and interactions of Glu49 will also affect how it interacts with nearby residue Val20. The change in hydrogen bonding interactions will likewise affect other residues, including Leu11 that is on the same N-terminal α helix as Tyr4. (C) Pairwise force distribution analysis of the MD simulations shows that the Val20-Glu49 interaction is weakest in the E:IGP state (i.e. near 0 pN), and strongest in the E:G3P (i.e. either strong attraction, ~-300 pN, or strong repulsion, ~300 pN), potentially explaining why the chemical shift correlation between these two residues is lowest and highest in the E:IGP and E:G3P states, respectively.

The NMR data indicated that the interaction between Val20 and Glu49 changes across the catalytic cycle. The MD simulations help to rationalize this finding. Val20 forms a persistent van der Waals (VDW) interaction with Glu49 in the apo, E:indole:G3P and E:G3P states, however the presence of IGP disrupts this interaction (Figure 2.6B). Force distribution analysis shows that the hydrogen bond between Glu49 and IGP results in little or no interaction between Glu49 and Val20 (Figure 2.6C). The disruption of this interaction likely explains the poor chemical shift correlation for these residues in the E:IGP state (Figure 2.4). As Glu49 shifts to its transitional conformation, it interacts with Tyr173, and the hydrophobic contact with Val20 is established. This VDW contact appears strongest when the Glu49-Tyr4 hydrogen bond is present (i.e. in the inactive conformation). This result potentially explains why the NMR chemical shift correlation between Val20 and Glu49 is highest in the E:G3P state.

We further applied the force distribution analysis (FDA) tool to better understand how interactions involving the *CYCLE 1* and *CYCLE 2* cluster residues change during the catalytic cycle (Figures 2.1 and 2.7). In FDA, pairwise forces are calculated between each atom pair during an MD simulation, instead of a total force acting on one atom, which would average to zero over time. FDA provides insight into mechanical linkages and signal propagations, even in the absence of larger conformational changes. The pairwise forces are represented as a scalar value with negative and positive values representing attractive and repulsive forces, respectively . In our analysis, we averaged residue interactions over all atoms, so changes to specific atom-atom contacts might potentially be offset by changes to other atom-atom interactions. In addition, when averaging values

over time, persistent repulsive or attractive interactions throughout the trajectories result in strong negative or positive values, respectively. While alternating positive and negative force values may result in a small time-averaged values, they still reflect changes in interactions and some of the observed changes were illuminating. For example, the averaged pairwise forces for Val20-Glu49 across the catalytic cycle (i.e. apo, -32.79 pN; E:IGP, -19.52 pN; E:indole:G3P, -25.93 pN; E:G3P, -40.49 pN) follow a similar pattern as the chemical shift correlations (Figure 2.4) with the strongest interaction and highest chemical shift correlation occurring at the E:G3P state.

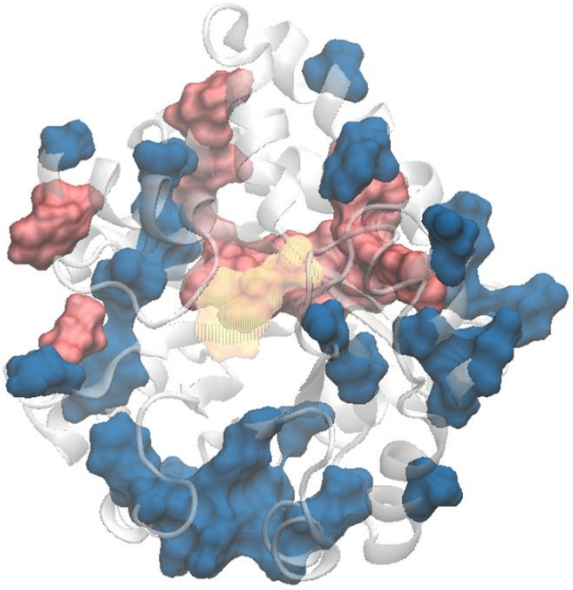
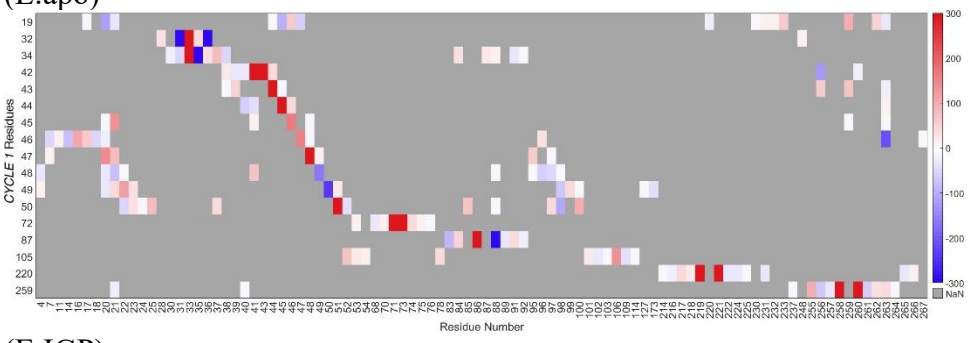


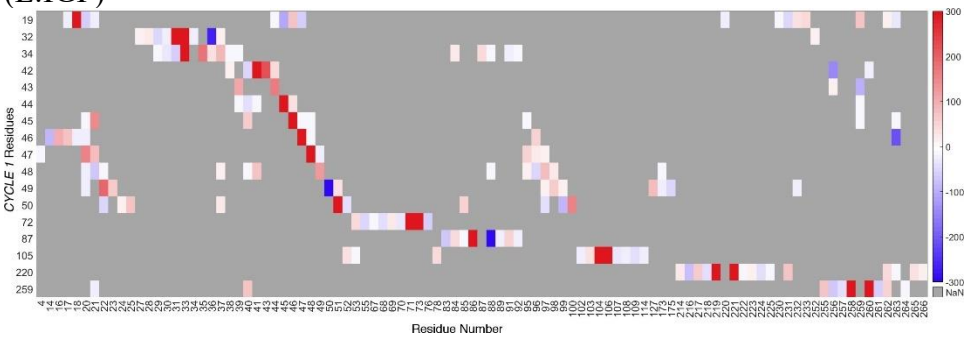
Figure 2.7. Topology of *CYCLE 1* and *CYCLE 2* clusters. *CYCLE 1* (red) involves residues mainly near the active site of αTS; *CYCLE 2* (blue) spans the surface of αTS.

In general, the interactions between *CYCLE 1* cluster residues were maintained throughout the catalytic cycle, displaying similar pairwise forces across the catalytic cycle, with some notable exceptions (Figure 2.8). In addition to Glu49, *CYCLE 1* residues Phe19, Leu34 and Val259 all show extended networks of interactions, and the nature of these interactions changes across the catalytic cycle (Figures 2.8 and 2.9). Dihedral angle analyses show that changes in overall force distribution may be directly or indirectly affected by conformational exchange of surrounding and interacting residues. For example, Ala43 and Val259 had an overall repulsive interaction in all states, except for the E:IGP state (Figure 2.9A). These two residues are part of a small hydrophobic cluster including amino acid residues Ala43, Ile237, Phe258, Val259, and Met262, connected to the active site by their interactions with Phe19 (Figure 2.9B). Changes in the Val259 conformation were not observed, however, it is in relatively consistent contact with Ile237 in the E:IGP state (Figure 2.9C). According to dihedral angle analysis, Ile237 maintains one conformation in this state and its interaction with Val259 possibly relieves the repulsion between Ala43 and Val259. In all other states, Ile237 appears to populate three different conformations, and its contact with Val259 is much weaker (Figures 2.9C and 2.9D). In this context, it is worth noting that the chemical shift correlation for Ala43-Val259 was highest for the E:IGP state ($R>0.99$), also suggesting their close relationship in this state.

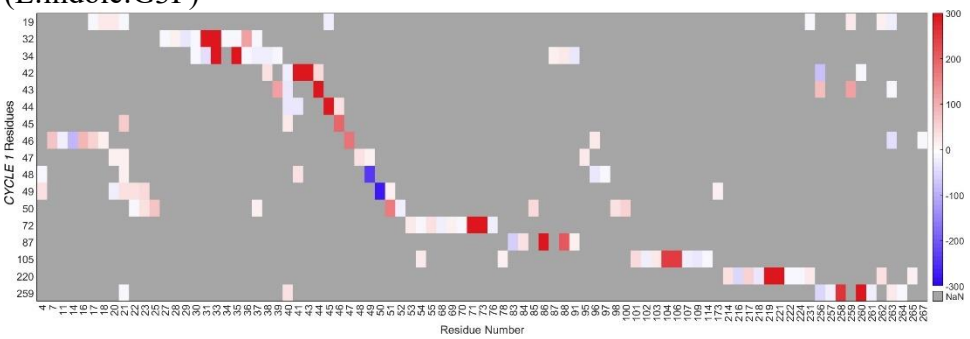
A (E:apo)



B (E:IGP)



C (E:indole:G3P)



D (E:G3P)

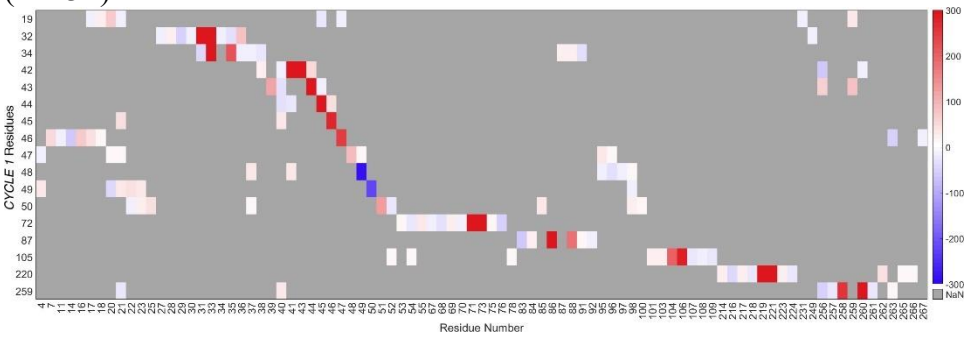


Figure 2.8. Heat maps for pairwise force distribution analysis (in pN) for *CYCLE 1* residues in (**A**) E:apo, (**B**) E:IGP, (**C**) E:indole:G3P and (**D**) E:G3P states.

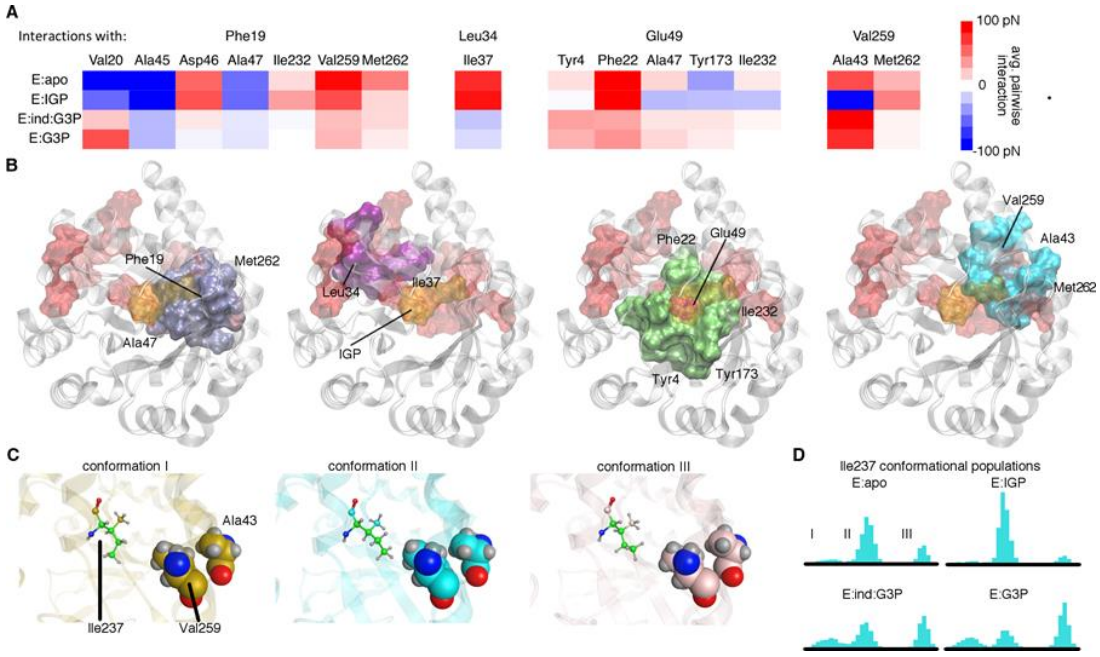


Figure 2.9. *CYCLE 1* residues change their interactions across the αTS catalytic cycle according to MD simulations. (A) Select average pairwise interactions for *CYCLE 1* residues Phe19, Leu34, Glu49 and Val259. Attractive and repulsive interactions are shown in varying degrees of blue and red, respectively. (B) FDA and dihedral angle analyses indicate that *CYCLE 1* residues Phe19, Leu34, Glu49 and Val259 change their interactions across the catalytic cycle. Changes in these interactions likely affect other *CYCLE 1* residues. *CYCLE 1* residues are plotted in red, and residues associated with Phe19, Leu34, Glu49 and Val259 are plotted in blue, magenta, green and cyan, respectively. IGP is plotted in yellow. (C) Ile237 can access different conformations, which affects the Ala43-Val259 interaction. The Ile237 Cα-Cβ-Cγ1-Cδ dihedral angle is highlighted in green. (D) The conformational populations of Ile237 across the catalytic cycle (conformation I, 30° to 130°; conformation II, 130° to 210°; conformation III, 260° to 330°).

Other notable changes involve interactions between Phe19 and residues 44-47, which are mainly characterized by hydrogen bonds between backbone atoms of Phe19 and Ala45 and Asp46 (Figure 2.10). The overall dynamics of these hydrogen bonds appear to relate to the position of the Glu49 sidechain and its interaction with Val20, which in turn affects the overall conformation of the β1 strand relative to the β2 strand and the adjacent turn (α1β2 loop).

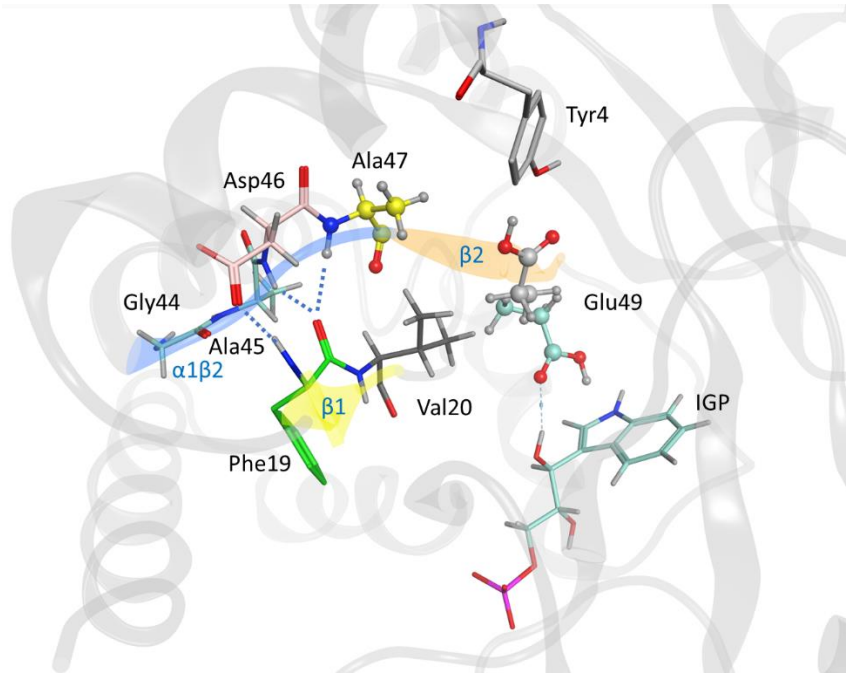
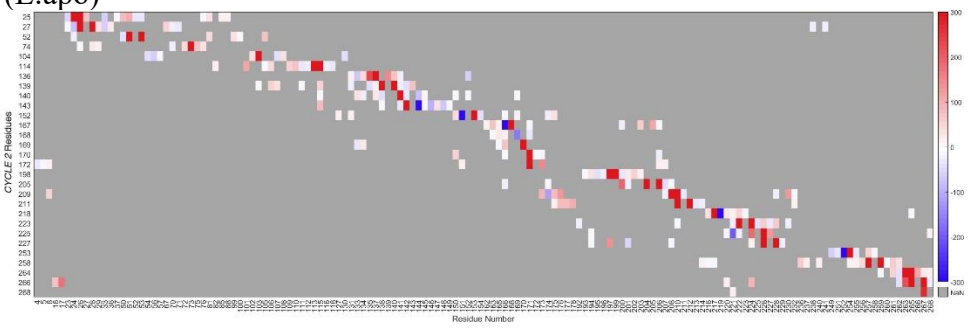


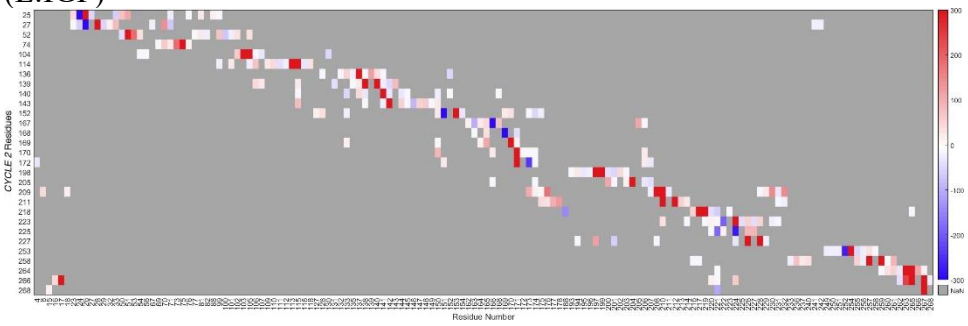
Figure 2.10. Interactions involving Phe19 and residues on the β1 strand (yellow), α1β2 loop (blue) and β2 strand (orange). Phe19, Gly44, Ala45, Asp46, Ala47 and Glu49 are *CYCLE 1* residues, and Val20, Ala47, Glu49 are *CORE* residues.

The interactions between *CYCLE 2* cluster residues were also generally maintained in all four states of the enzyme with only a few interactions changing substantially across the catalytic cycle (Figures 2.11 and 2.12A). For example, Leu25 and Asp27 had a repulsive interaction in the E:apo state (68.70 pN), but an attractive interaction in the E:IGP state (-54.0 pN). The interactions between the backbone atoms of the two residues mainly contribute to the averaged force value. A salt bridge between Asp27 and Arg70 of the α2 helix appears to affect the relative position of Asp27 that in the E:apo state is closer to Leu25 causing van der Waal repulsion between Thr24, Leu25 and Asp27 (Figures 2.12A and 2.13). The dynamics of the α2 helix are likely affected by the motion of the β2α2 loop, responding to the presence of the different ligands. The change in interactions paralleled the change in the chemical shift correlations; the chemical shift correlation for Leu25-Asp27 was higher in the E:IGP state (R 0.98) compared to the E:apo state (R 0.73).

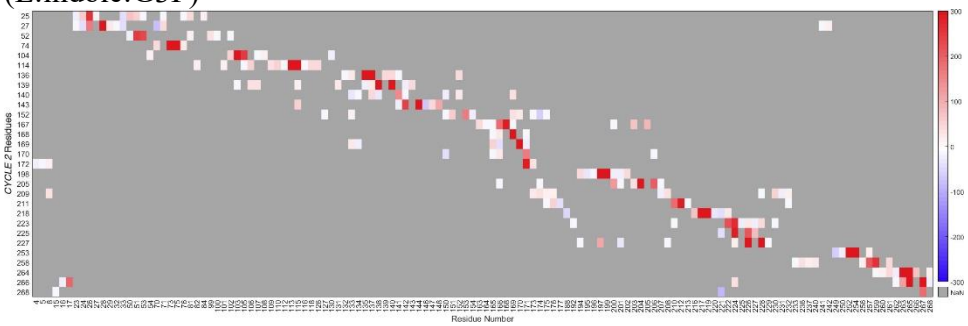
A (E:apo)



B (E:IGP)



C (E:indole:G3P)



D (E:G3P)

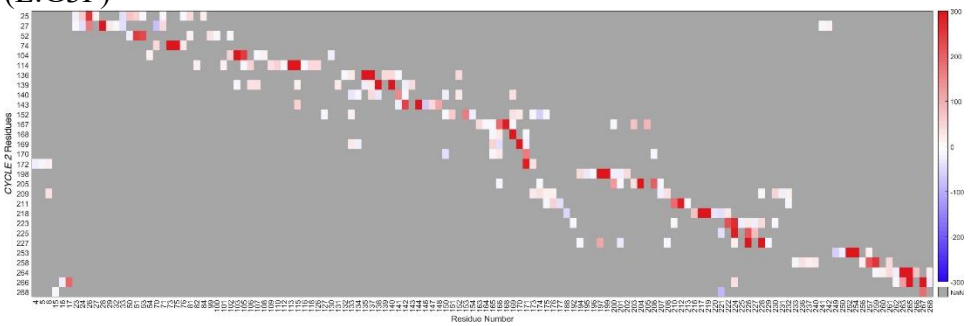


Figure 2.11. Heat maps for pairwise force distribution analysis (in pN) for *CYCLE 2* residues in (**A**) E:apo, (**B**) E:IGP, (**C**) E:indole:G3P and (**D**) E:G3P states.

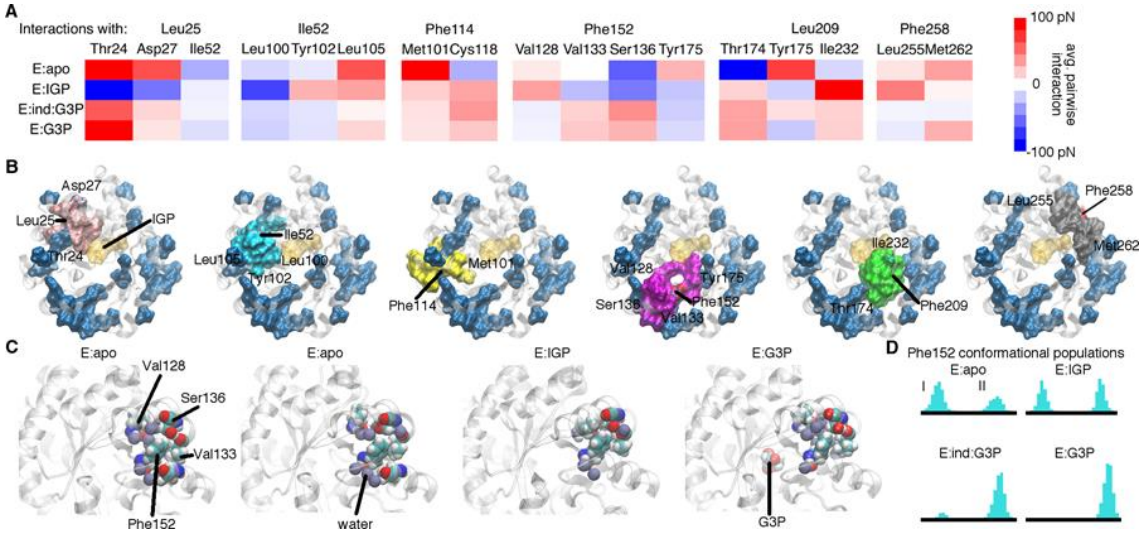


Figure 2.12 *CYCLE 2* residues change their interactions across the αTS catalytic cycle according to MD simulations. (A) Select average pairwise interactions involving *CYCLE 2* residues Leu25, Ile52, Phe114, Phe152, Leu209 and Phe258. Attractive and repulsive interactions are shown in varying degrees of blue and red, respectively. (B) *CYCLE 2* residues are plotted in blue, and residues associated with Leu25, Ile52, Phe114, Phe152, Leu209 and Phe258 are plotted in peach, cyan, yellow, magenta, green and grey, respectively. IGP is plotted in yellow. (C) Interactions involving Val128, Val133, Ser136 and Phe152 change across the catalytic cycle, which can also affect interactions with water. In the E:IGP state, water is excluded at this site, allowing for rotation of Phe152. In the E:G3P state, conformational exchange involving Phe152 does not occur. (D) The conformational populations of Phe152 across the catalytic cycle as defined by the Phe152 Cα-Cβ-Cγ-Cδ1 dihedral angle (conformation I, 50° to 150°; conformation II, 240° to 320°), which indicates a rotation of the phenyl ring. There is conformational exchange in the E:apo and E:IGP states, but limited exchange in the E:indole:G3P and exchange is absent in the E:G3P, indicating population shifts over the catalytic cycle.

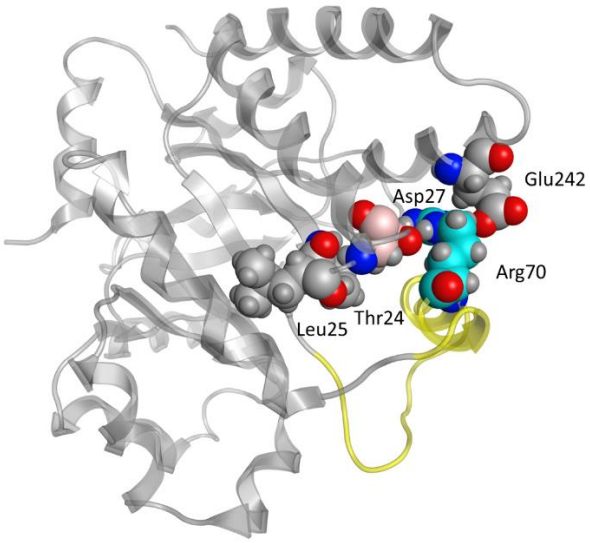


Figure 2.13. Interactions involving Leu25 and Asp27. The α2 helix and β2α2 loop are colored in yellow.

The interaction between Ser136 and Phe152 likewise changed from an attractive interaction in the first part of the catalytic cycle (i.e. E:apo, E:IGP) to a repulsive interaction in the second part of the catalytic cycle (i.e. E:indole:G3P, E:G3P). Rotation of the aromatic ring of Phe152 was observed in the E:apo and E:IGP states (Figure 2.12). While conformational exchange of Phe152 does not necessarily directly correlate with attractive forces between Ser136 and Phe152, rotation indicates more degrees of freedom for the aromatic ring of Phe152. In the E:IGP state, the hydrophobic contact of Phe152 with Val128 and Val133 allows for conformational exchange by excluding water from the site and allowing Phe152 to rotate. In the other three states, the limited contact between Val128 and Phe152 potentially allows a water layer to form between these

residues (Figure 2.12C), helping to force Phe152 towards the methyl and backbone atoms of Ser136 and neighboring residues. In the E:apo state, this repulsion is relieved by rotation of the Asp136 hydroxyl group, while in the E:indole:G3P and E:G3P states, such conformational exchange was not observed (Figure 2.12C). These effects might help to explain the changes to the Ser136-Phe152 chemical shift correlations across the catalytic cycle; this chemical shift correlation is weak in the E:apo state (R 0.65), but even weaker in the E:indole:G3P (R 0.44) and E:G3P (R 0.32) states.

As with the *CYCLE 1* residues, we also noted that some *CYCLE 2* residues changed their interactions with residues outside the cluster. We selected six residues (Leu25, Ile52, Phe114, Phe152, Leu209 and Phe258) showing expanded networks of interactions, reaching towards and into the active site (Figure 2.12). Ligand changes at the active site may propagate changes to these interactions, affecting the *CYCLE 2* cluster residues. Many of these interactions also involve *CYCLE 1* (e.g. Ile52-Leu105) and *CORE* (e.g. Ile52-Leu100, Phe114-Met101, Phe152-Tyr175, Phe258-Ile240) residues (Figure 2.12), suggesting that the *CYCLE* and *CORE* clusters may act in concert. It is also noted that the *CORE* and *CYCLE* clusters are not mutually exclusive, and many *CORE* residues are positioned next to or even between *CYCLE* cluster residues. Altogether, our results suggest that the interactions within the *CYCLE 1* and *CYCLE 2* clusters do not necessarily change across the catalytic cycle, but these residues are especially responsive (directly or indirectly) to changing interactions and surrounding chemical environments.

2.3.2 A198W Mutation Induced Conformational Changes

We further examined the effects of the mutation at a surface residue of αTS subunit. As the increase in catalytic efficiency affected only the full TS complex and not the individual subunits, we focused our analysis on interacting residues at the interdomain surface rather than α and β active sites. While the NMR data provided some insight into communication between position 198 and the αTS/βTS interface, the NMR data was limited in providing full context for the αβ TS complex. As such, we analyzed MD trajectories for αβ heterodimers, including for wild type (WT) and A198W TS. Differences in the interactions at the mutation site are subtle but they appear to translate into more significant changes in the interdomain communication at the αβ surface Briefly, the MD simulations indicated that local changes at the substitution site propagated structural dynamic changes throughout αTS (Figure 2.14), including decreasing the overall flexibility of active-site loop 6 (αL6). It is it important to note that large scale conformational changes captured in NMR experiments on millisecond timescale are induced by side chain motions and backbone fluctuations which could be detected within nanoseconds timeframe. Detailed atomistic analysis of our trajectories explains how changes in forces and interactions lead to different dynamics and conformational behavior in the WT and the mutant system. Our data shows that such changes led to the establishment of contacts at the αTS/βTS interface at four key positions in the A198W variant within the 200 ns simulation time, while such links were not formed or maintained in the WT enzyme on the same timescale (Figure 2.15). Furthermore, interactions observed for key residues within the indole channel (βTyr279,

βPhe280) in the mutant system suggest that substrate channeling may be more efficient in the A198W variant. We detail some of the conformational dynamics and contact changes induced by the A198W substitution below.

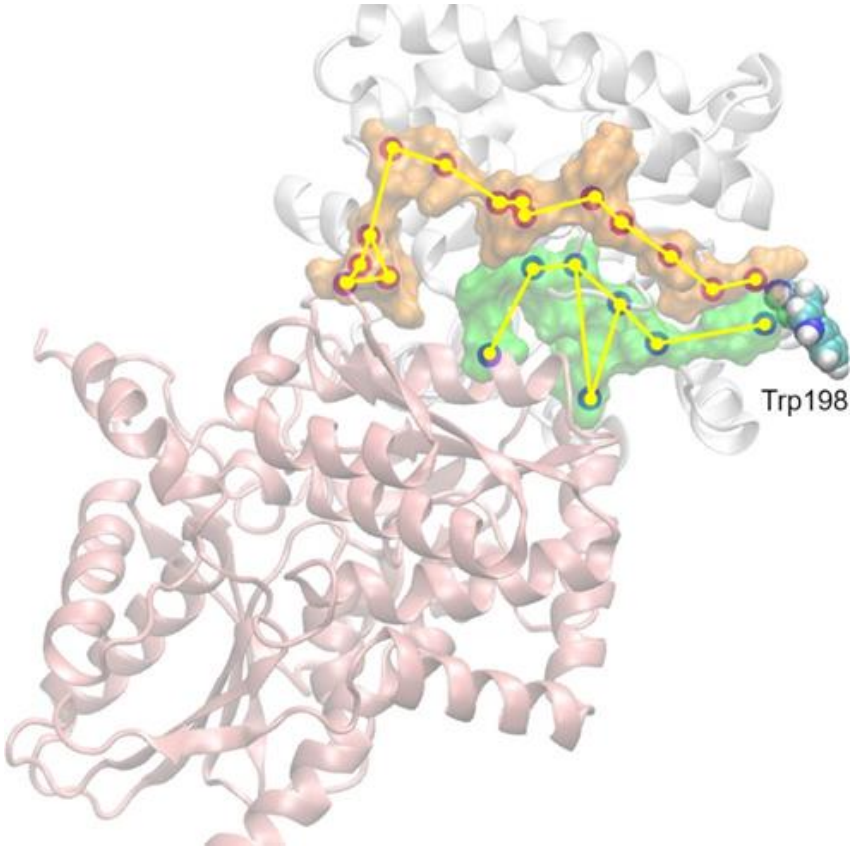


Figure 2.14. Propagation of interactions changes induced by the A198W substitution were examined in two directions leading to the αTS (white) – βTS (pink) interface. αC of residues with affected conformation are traced, based on pair-wise force distribution analysis, from the substitution site as follows: in orange – αAsn194, αLeu193, αLeu191, αAla189, αArg188, αGlu186, αAla185, αSer235, αLys239, αGlu242, αAsn66, αGln65, βSer161, βGly162; in green – αHis195, αPro156, βIle20, αLeu177, αPhe212, αThr183, βArg175.

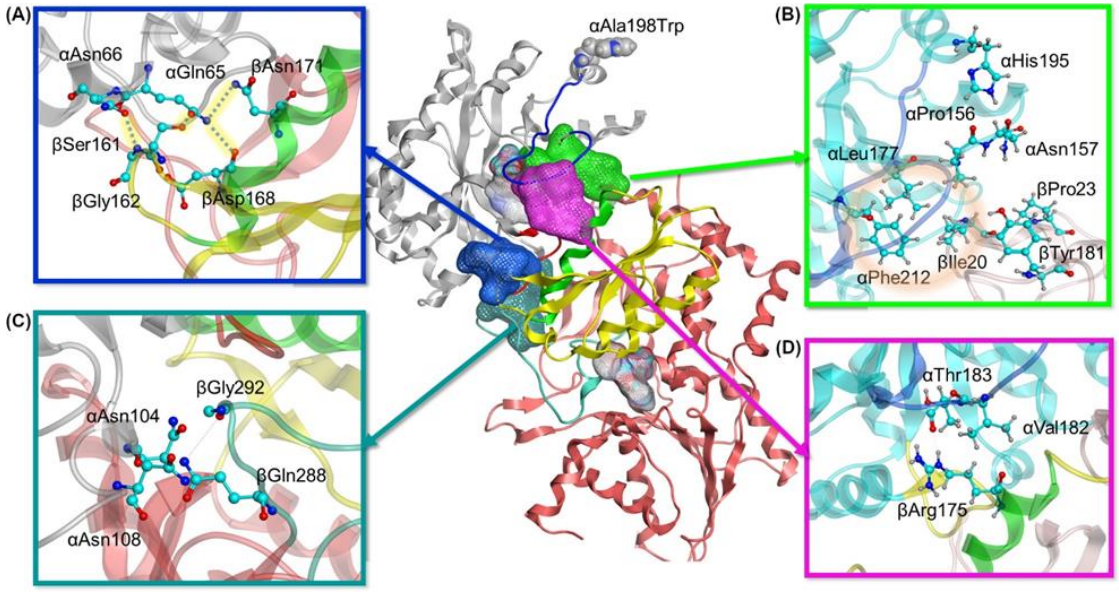


Figure 2.15. **.** Molecular modeling reveals four positions with stronger interdomain interactions in A198W TS. (A) Electrostatic link between alpha residues αGln65, αAsn66 and beta residues βSer161, βGly162, βAsp168 and βAsn171. (B) Hydrophobic cluster between αLeu177, αPhe212, and βIle20 observed in A198W TS. (C) αAsn104 forms consistent hydrogen bonds with βGln 288/βGly 292 in A198W TS. (D) Consistent hydrogen bond between αThr183 and αArg175 may contribute to closed αL6 conformation.

The most evident effect of the mutation, observed in the MD simulations, is the displacement of alpha helix 6 (αH6) in the mutant in the direction of alpha loop 6 (αL6) or the N-terminal direction shift (Figure 2.16). In the mutant this N-terminal direction shift brings αHis195 closer to the turn between alpha sheet 5 (αS5) and αH5 resulting in a dynamic but consistent hydrogen bond with αPro156 and recurrent contact with αAsn157. A group of hydrophobic residues αLeu177, αPhe212, and βIle20 are also in

consistent close contact. While it is not clear if this set of interactions induces the helix shift or vice versa, lack of it, as observed in the WT system, results in a formation of a small cluster between αAsn157, βIle20, βPro23 and βTyr181, while the interactions between βIle20 and alpha Leu177 and Phe212 are interrupted. (Figures 2.15B and 2.17) This collection of interactions may be important as the absence of the hydrophobic contact between βIle20 and αLeu177 and αPhe212 may contribute to αL6 disordering and moving to open conformation providing possible means of substrate escape from alpha subunit. (Figure 2.17D).

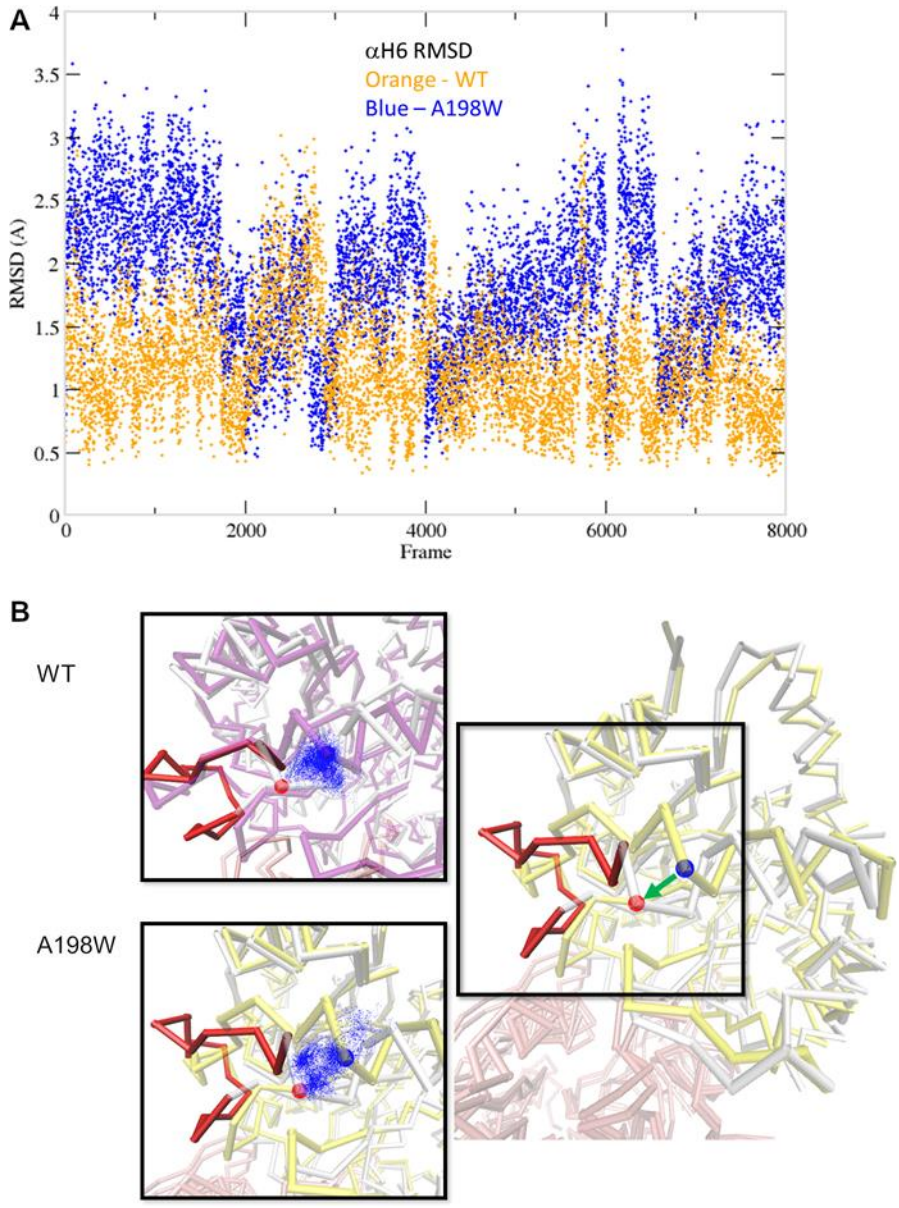


Figure 2.16. Displacement pattern of αH6 leads to conformation change in local and distant residues. (**A**) Root-mean-square deviation (RMSD) for αC of αH6 (residues 194 – 203). (B) αHis195 αC position distribution over 8000 analyzed frames represented by blue dots. The initial position of the atom is shown in blue and the most distant position in red. αL6 is colored in red.

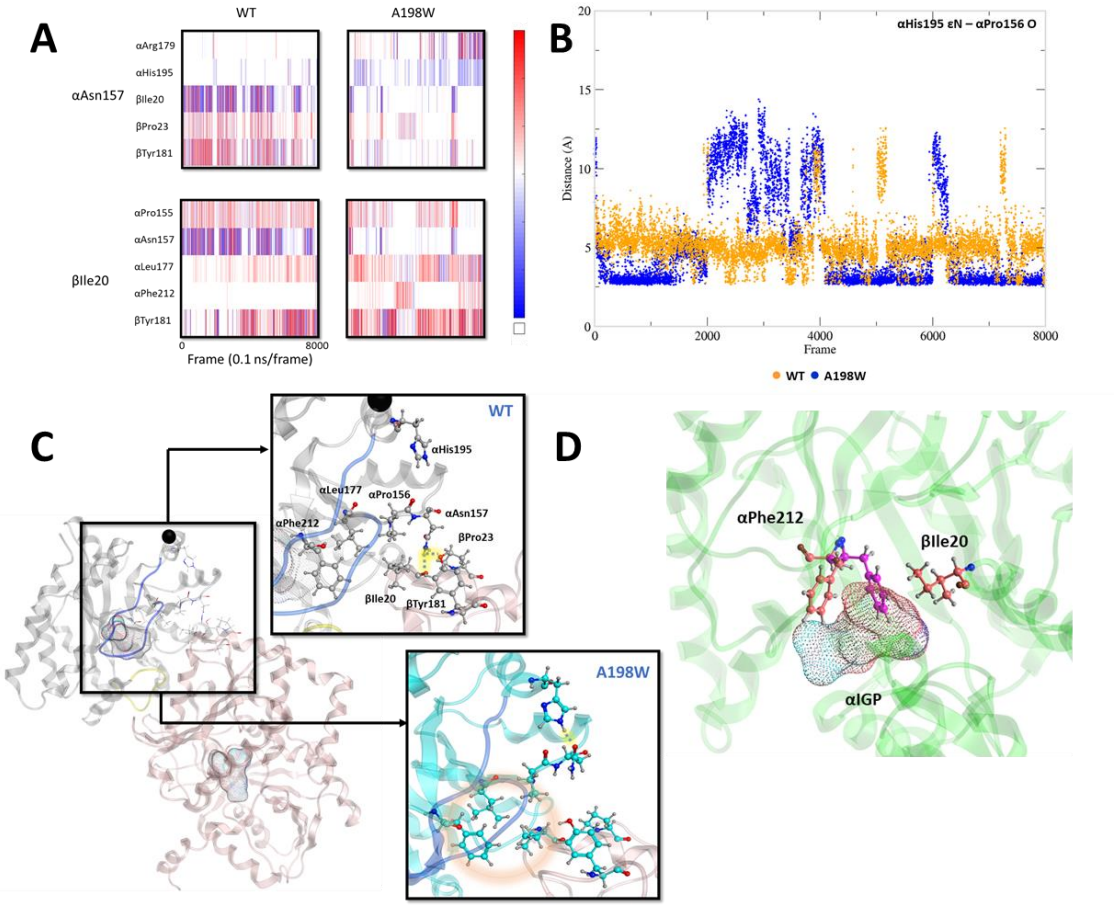


Figure 2.17. Local interaction changes are induced by A198W substitution. (A) Pair-wise force distribution analysis shows different interactions for αAsn157 and βIle20 in WT and A198W. (B) Hydrogen bond between αHis195 and αPro156 is formed and maintained in three of the A198W MD simulations analyzed, while such hydrogen bond is rarely observed in WT. (C) Hydrophobic interactions between residues βIle20 (βTS in pink), αLeu177 and αPhe212 in A198W are inconsistent in WT. (D) Lack of such interactions may lead to rotation of the sidechain of αPhe212 providing escape route for substrate (initial conformation shown in magenta, rotation in pink).

Differences in the dynamics of residues part of loop 6 are more challenging to trace due to its high flexibility. Changes in interactions were observed for αAsn194 as its sidechain forms a more stable hydrogen bond with backbone oxygen of αAla226 of αH7 in the mutant compared to wild type. A backbone hydrogen bond between αLeu191 O and αGln210 N is strong in mutant and rarely observed in wild type (Figure 2.18). These findings agree with suppressed conformational exchange of αGln210 and αAla226 on the millisecond timescale, at least according to the NMR studies on αTS alone. Consistent contact is observed between αAla185, αSer215 and αSer235 which may impact the closed conformation of αL6, as in the wild type, these interactions are interrupted.

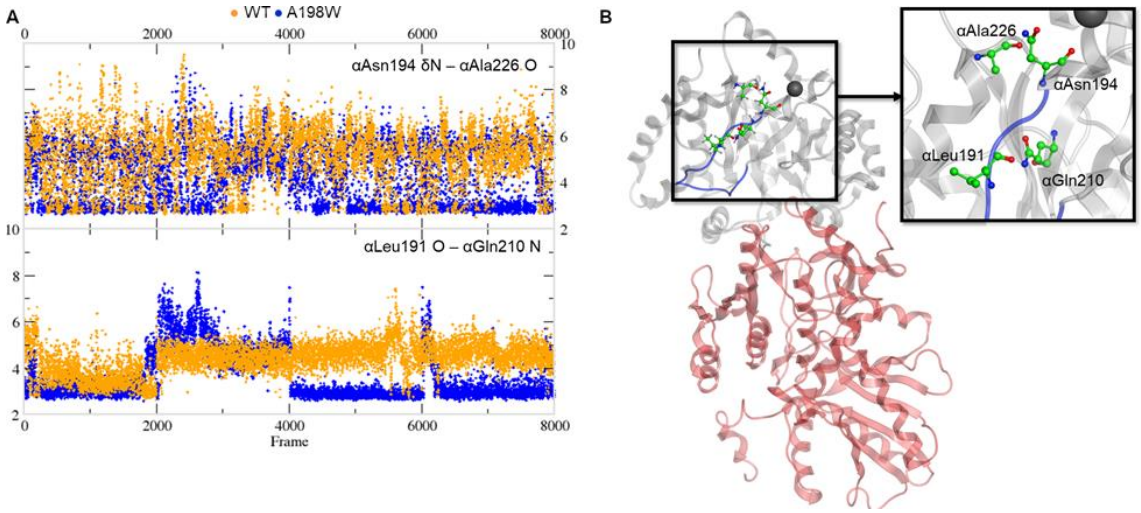


Figure 2.18. Hydrogen bonds between αAsn194 and αAla226, and αLeu191 and αGln210 (A and B) are more consistent in A198W and may further contribute to the ordered closed conformation of αL6 (B) colored in blue. Mutation site is indicated by black circle, αTS is colored in white and βTS is colored in pink.

An extremely critical aspect of the dynamics of αL6 in the mutant system is its consistent interactions with βH6, an important part of COMM domain which is well known to play a key role in the interdomain communication and the activation of αTS.[5] Pairwise force data and hydrogen bond analysis show a consistent hydrogen bond between αThr183 O and βArg175 guanidino group in over 70% of the frames analyzed (Figures 2.15D and 2.19). This hydrogen bond was not maintained in any of the trajectories of the simulated WT system. In addition to this hydrogen bond, a hydrophobic interaction is established between αVal182 and the hydrocarbon portion of the βArg175 sidechain. Loss of these interactions as observed in wild type system and in one of the trajectories for the mutated system results in displacement and disordering of αL6. In A198W trajectory, with the absent hydrogen bond between αThr183 and βArg175, a hydrophobic interaction between αVal182 and beta residues Ser178 and Gly179 is maintained but it is not sufficient to prevent the loop displacement.

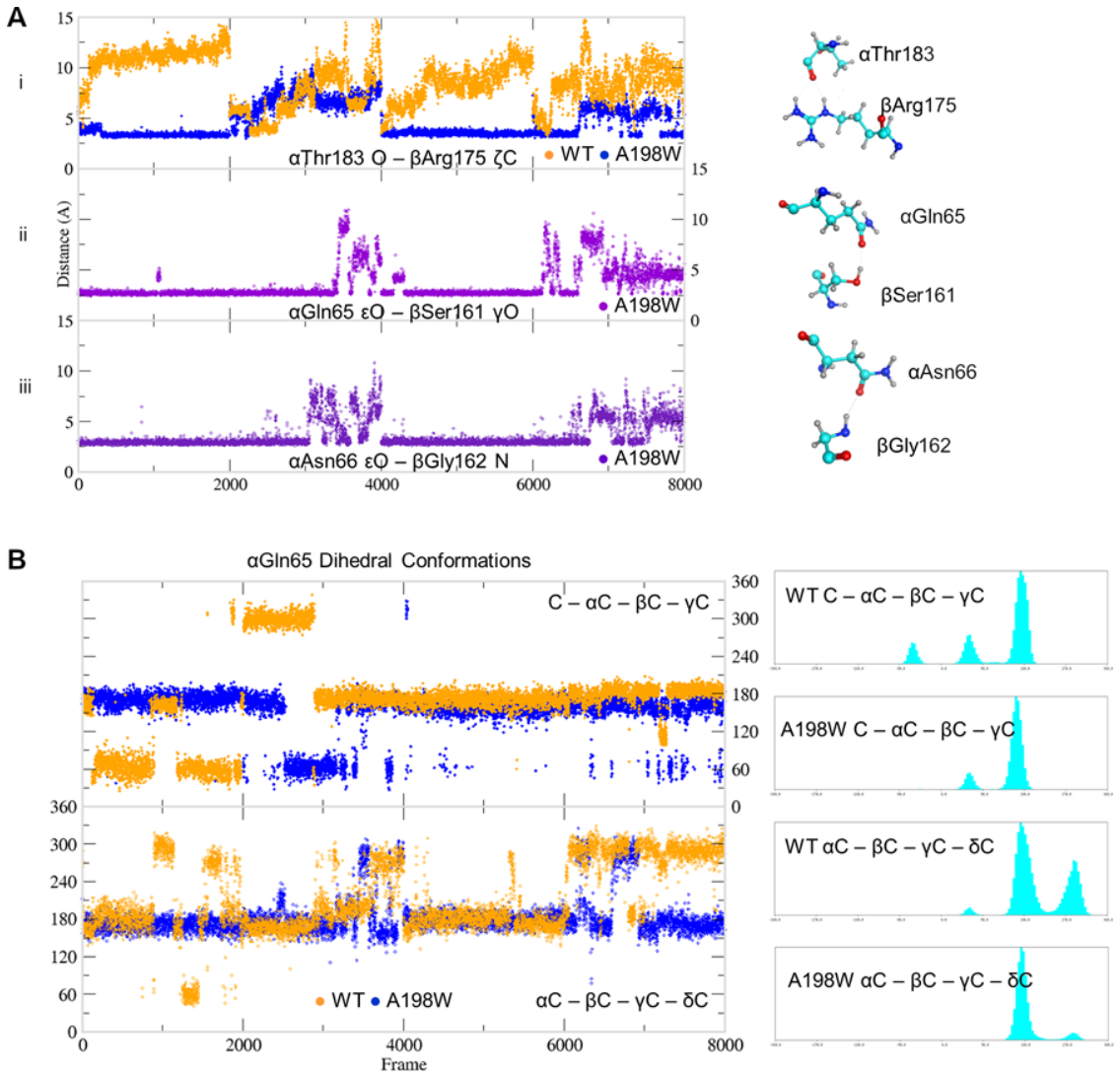


Figure 2.19. Hydrogen bonds at αTS/ βTS interface in A198W (A) are more consistent and affect the conformation populations of the involved residues. Dihedral analysis shows one predominant conformation for αGln65 in A198W and multiple conformations in WT due to lack of stable interactions (B).

Another series of interactions correlates with the events occurring at the αL6 and βH6 interface - interactions of αGln65 and αAsn66 with beta residues βSer161 and βGly162 (Figure 2.15A). Dihedral analysis shows one distinct conformation for the C – αC – βC – γC angle of αGln65 and αAsn66 in the mutant (Figure 2.19B). This conformation is associated with a hydrogen bond between αGln65 and βSer161 (68% in mutant and less than 3% in wild type). In addition, electrostatic (salt-bridge like) interaction between αGln65 and beta residues βAsp168 and βAsn171 is observed in the A198W variant. βAsp168 and βAsn171 are also part of beta helix 6 (βH6). A hydrogen bond between Oδ of αAsn66 and backbone of βGly162 is observed in over 50% of all frames for the mutant system, and less than 20% in the wild type. (figure 2.19A). Another hydrogen bond between the backbone nitrogen of αGln65 and αGly61 is more stable in the mutant. These results are interesting considering that the A198W substitution induced substantial chemical shift changes in both αGly61 and αAla67. Unfortunately, the backbone resonances of αGln65 and αAsn66 are unassigned. Nonetheless, the NMR results indicate that there are substantial structural dynamic changes in this region even in the absence of βTS. The A198W substitution also induced millisecond conformational exchange in αAla67, suggesting that this region may be seeking alternative binding interactions.
The interactions of αGln65 and αAsn66 with the small loop between βS5 and βH6 and αThr183-βArg175 hydrogen bond appear to affect the position of βH6 relative to alpha subunit. In A198W variant, the displacement occurs in the direction of α subunit whereas in the WT, lack of consistent interactions leads to displacement in the opposite direction away from αTS (Figure 2.20). Difference in the interactions, possibly associated with this

movement, are observed in the two systems for the indole channel lining residues βTyr279 and βPhe 280 (Figure 2.21). Force analysis shows that both residues in the wild type have stronger interactions with residues from βH6. A more consistent link between βTyr279 and βLys167 is observed in the wild type. A hydrogen bond between βTyr 279 and βAsn171 is also observed in one of the WT simulations. Such interactions position the sidechain ring of βTyr 279 within the tunnel possibly interfering with the indole transfer.[32] Similarly, βPhe280 in WT enzyme shows stronger interactions with βCys170 of βH6 which brings the sidechain of phenylalanine within the indole channel. In the mutant, these residues favor interactions with beta residues 306-308 on the channel "wall." A hydrogen bond between βTyr 279 and αAsp56 is observed in 40% of MD frames in the A198W variant compared to 10% in WT TS. Such interaction may stabilize the aromatic ring of tyrosine in an "open channel" conformation.

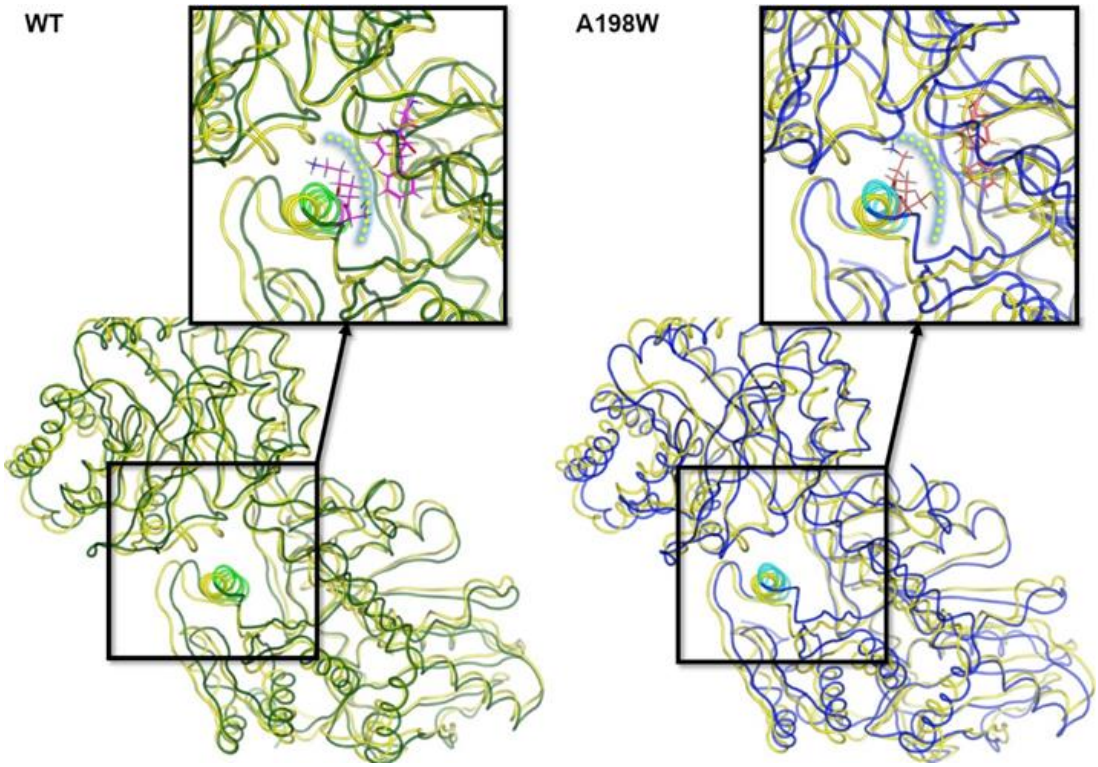


Figure 2.20. Displacement of βH6 towards αTS leads to widening of indole channel which is further supported by weaker interactions between the channel lining residues βTyr279 and βPhe280 with residues of βH6 in A198W. Initial frame of MD simulations for both systems is in yellow trace and consequent displacement indicated in green for WT and blue for A198W, βH6 is colored in light green and light blue, respectively.

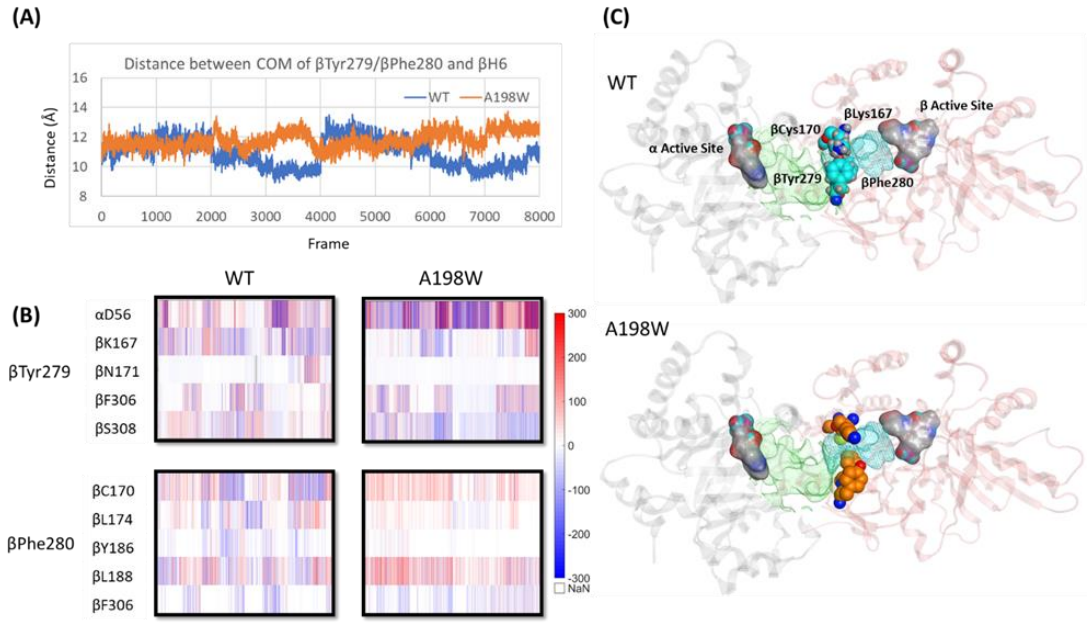


Figure 2.21. Displacement of βH6 towards αTS leads to widening of indole channel which is further supported by weaker interactions between the channel lining residues βTyr279 and βPhe280 with residues of βH6 in A198W. (A) Distance between center of mass of residues βTyr279/βPhe280 and βH6 lining the affected tunnel portion. (B) Interactions for residues βTyr279 and βPhe280 show different trends in WT and A198W. (C) Interactions between βTyr279 and βLys167 position the aromatic ring within the indole channel. Active sites in both subunits (grey – α subunit, pink – β subunit) are represented as atom colored surface. Indole path is shown in green surface and the affected portion of the channel is mapped in cyan.

Other interactions induced by the A198W substitution may also be important for αTS-βTS communication, although the mechanism is unclear. For example, αAsn104 also appears to favor interactions with βTS in the A198W variant (Figure 2.15C). In WT, αAsn104 forms a very consistent hydrogen bond with αAsp130 (>60%) and in much

smaller extent with the backbone of βIle278. In A198W mutant, a dynamic hydrogen bond network between αAsn104 and βGln288/βGly292 was established, while such link was not formed in the wild type; these beta residues form part of the allosteric "metal binding loop" – βL8 (residues 287-310).[33] This consistent interdomain interaction in the mutant was further aided by additional set of hydrogen bonds between αAsn108 and residues from βL8 but most hydrogen bonds are sustained for less than 20 % of the frames. NMR studies indicated that the A198W substitution induced chemical shift perturbations for αAsn104 and αAsp130.

We briefly analyzed two other variants A198K and A198V, following up experimental results showing improper folding. The MD simulations for the A198K and A198V variants indicated a dramatic movement of αL6 into a position that completely opens the active site, and which is likely not conducive for catalysis (Figure 2.22); this aL6 conformation was not observed for either WT or A198W αTS. The large conformational change in aL6 may help to explain the large chemical shift and peak intensity changes in the A198K and A198V NMR spectra.

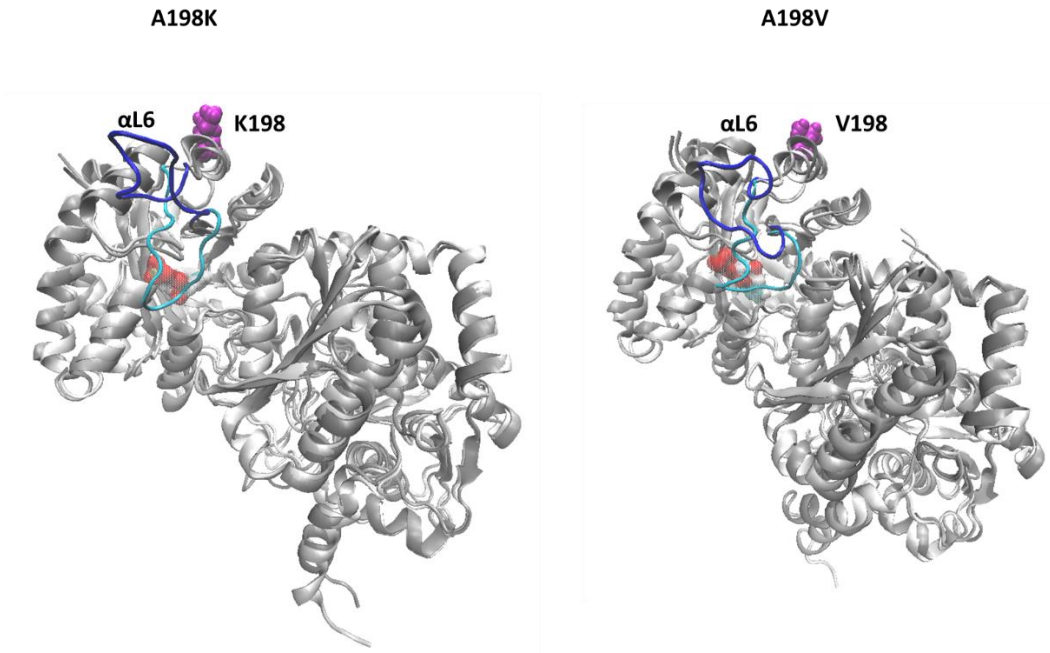


Figure 2.22. Molecular dynamics simulations suggest that the A198K and A198V induce a large conformational change in the active site αL6 (light and dark blue). Shown are snapshots from the MD simulations for the A198K (left) and A198V (right) variants with IGP indicating the active site.

Lastly, to further explore the significance of the aforementioned residues in terms of conservation, we aligned αβ TS complex to analogs in six different bacterial species. Most of the amino acids involved in the αTS-βTS interaction and/or form the indole channel, outlined in our MD work are conserved (Figures 2.23 and 2.24) or have a positive substitution to maintain interactions.

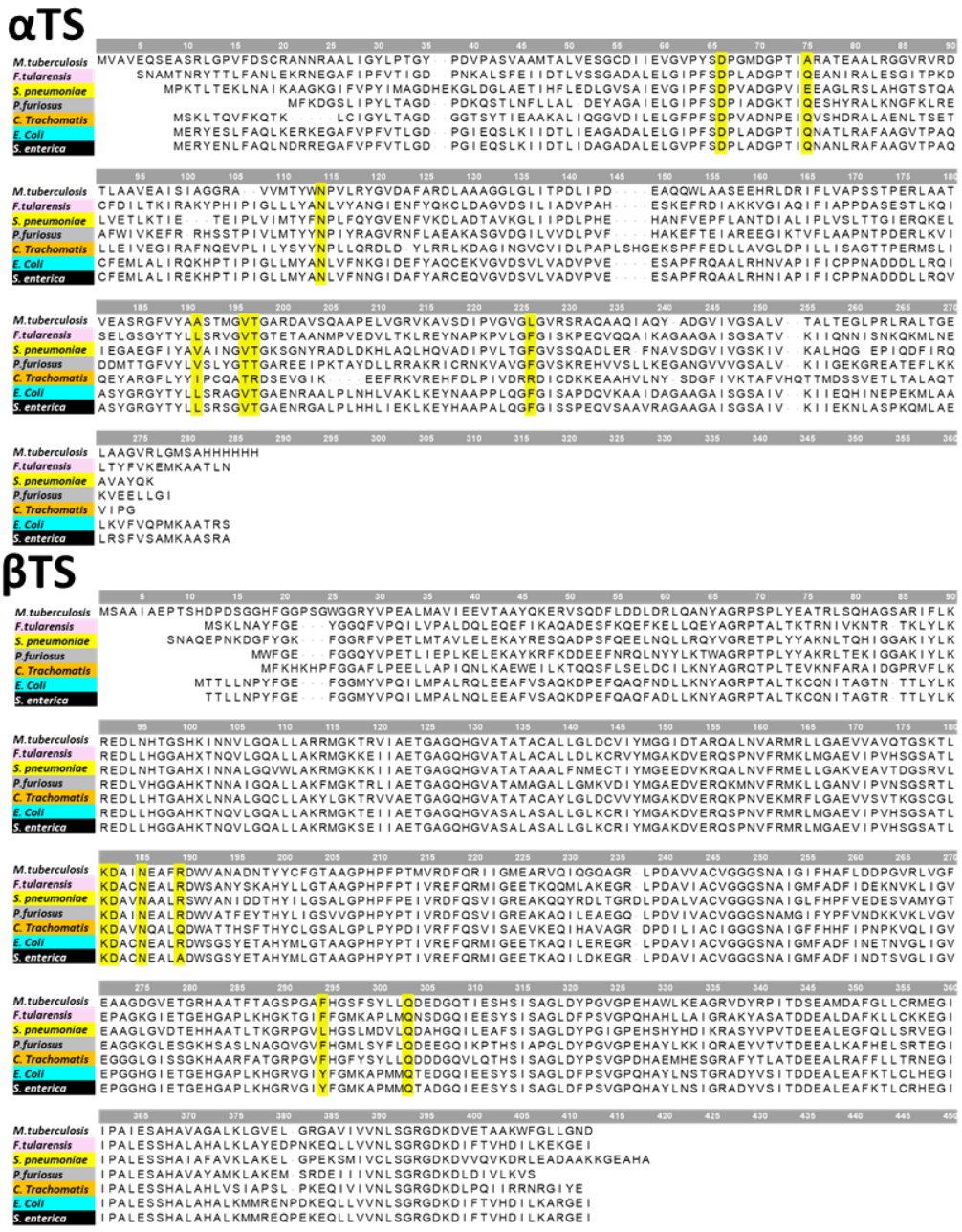


Figure 2.23 Sequence alignment for α and β chains of TS from *M. tuberculosis* (white), *F. tularensis* (pink), *S. pneumoniae* (yellow), *P. furiosus* (grey), *C. trachomatis* (orange), *E. coli* (teal), and *S. enterica* (black). Conserved residues, showing conformational changes due to the A198W substitution are highlighted.

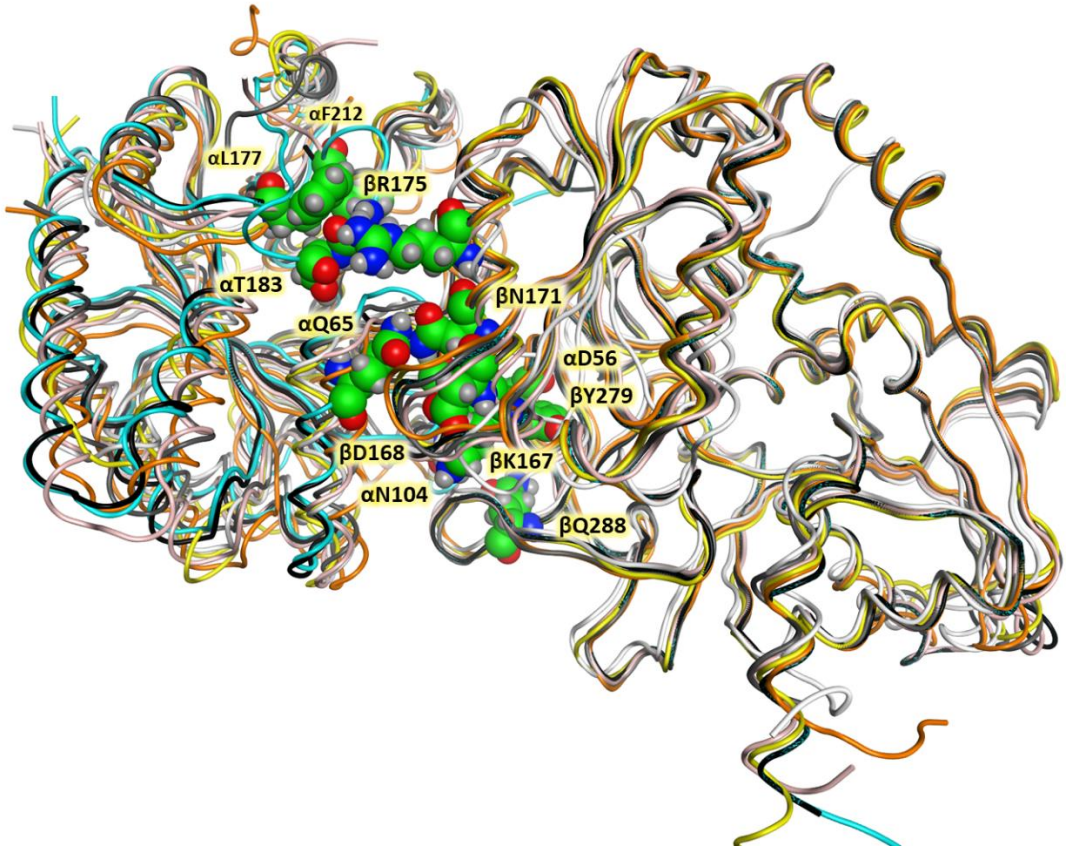


Figure 2.24 Structural alignment of for α and β chains of TS from *M. tuberculosis* (white), *F. tularensis* (pink), *S. pneumoniae* (yellow), *P. furiosus* (grey), *C. trachomatis* (orange), *E. coli* (teal), and *S. enterica* (black). Six structures of αβ TS (PDB ID: 6DWE, 5KZM, 5KIN, 1WDW, 6V82, 1QOQ) were aligned with our *E. coli* model (see Methods). Highlighted residues are shown as VDW model.

2.4. Conclusions

In this work, we combined several experimental and computational analysis tools in an effort to delineate protein residues network in our model system TS which heavily relies on allosteric communication for its functionality. NMR and CHESCA studies identified several networks which may affect the overall enzyme function. Conformational changes

within these networks are highly coordinated and cooperative, and likely drive the enzyme through the structural and dynamic changes necessary to carry the enzyme along its catalytic cycle.

We applied molecular dynamics and several analysis tools to gain an atomistic level insight into conformational changes occurring within and away from active site in the presence of different substrates throughout the catalytic cycle. Our data revealed three conformations for catalytic residue αGlu49. These three conformations appear to correlate with the presence of substrate in the active site and affect the interactions of the residue. Local changes in the active site led to conformational exchanges that propagate to surface residues and affect the overall dynamics of the protein. We studied the two networks defined by CHESCA analysis, *CYCLE1 and CYCLE2*, and found key residues with population shifts affecting the overall dynamics of each cluster. *CYCLE 1* residues were shown to responding more directly to ligand changes at the active site, while the *CYCLE 2* residues responded to structural changes outside the active site, especially those involving the outer a-helices, nonetheless these population shifts were in response to conformational changes occurring within the ligand binding site.

Our analysis also revealed how local changes at the mutation site propagate through a series of interactions and may lead to significant changes away from the perturbed site. Contacts at the interdomain surface were established at four key positions (Figure 2.15) in the mutant within the 200 ns simulation time, while such links were not formed or maintained in the wild type on the same timescale. Stronger interactions between

αPhe212 and βIle20 was shown to prevent substrate escape. The consistent hydrogen bond between αThr183 and βArg175 appears to be important in maintaining α-L6 conformation. It was previously shown αThr183 has critical importance in the interdomain communication,[34] and our results show that the residue is the main link between αL6 and βH6. The electrostatic group formed between the N terminal end of βH6 and alpha residues αGln65 and αAsn66 proved to enhance the position of βH6 relative to α subunit. This position of the helix may also affect the channel accessibility. Stronger interactions of βTyr279 and βPhe280 with βLys167, βAsn171, βCys170 of βH6 may indicate that the helix is closer therefore allowing continuous contact. Previous mutation studies reveal that αAsn104 is critical in alpha-beta association and communication[35] and in the mutated system, interactions between αAsn104 and αAsn108 and metal binding loop are more consistent.

The combined methodology in this work deepens understanding of allosteric regulation and molecular recognition. Overall, our results indicate that amino acid interaction networks may not only be important for function within an enzyme subunit, but they may be used to bridge communication between subunits in a multienzyme complex. This finding reveals new possibilities of controlling function in proteins, engineering more efficient enzymes, developing new drugs and improving the efficacy of existing drugs.

2.5 References

1. Cui Q, Karplus M. Allostery and cooperativity revisited. Protein Science 2008;17(8):1295-1307.

2. Dokholyan NV. Controlling Allosteric Networks in Proteins. Chemical Reviews 2016;116(11):6463-6487.

3. Motlagh HN, Wrabl JO, Li J, Hilser VJ. The ensemble nature of allostery. Nature 2014;508(7496):331-339.

4. Kulik V, Hartmann E, Weyand M, Frey M, Gierl A, Niks D, Dunn MF, Schlichting I. On the Structural Basis of the Catalytic Mechanism and the Regulation of the Alpha Subunit of Tryptophan Synthase from Salmonella typhimurium and BX1 from Maize, Two Evolutionarily Related Enzymes. Journal of Molecular Biology 2005;352(3):608-620.

5. Dunn MF. Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase bienzyme complex. Archives of Biochemistry and Biophysics 2012;519(2):154-166.

6. 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.

7. Wang J, Jain A, McDonald LR, Gambogi C, Lee AL, Dokholyan NV. Mapping allosteric communications within individual proteins. Nature Communications 2020;11(1):13.

8. Botello-Smith WM, Luo Y. Robust Determination of Protein Allosteric Signaling Pathways. Journal of Chemical Theory and Computation 2019;15(4):2116-2126.

9. Johnson QR, Lindsay RJ, Shen TY. CAMERRA: An analysis tool for the computation of conformational dynamics by evaluating residue-residue associations. Journal of Computational Chemistry 2018;39(20):1568-1578.

10. Feher VA, Durrant JD, Van Wart AT, Amaro RE. Computational approaches to mapping allosteric pathways. Current Opinion in Structural Biology 2014;25:98-103.

11. Selvaratnam R, Chowdhury S, VanSchouwen B, Melacini G. Mapping allostery through the covariance analysis of NMR chemical shifts. Proceedings of the National Academy of Sciences 2011;108(15):6133.

12. Axe JM, Boehr DD. Long-Range Interactions in the Alpha Subunit of Tryptophan Synthase Help to Coordinate Ligand Binding, Catalysis, and Substrate Channeling. Journal of Molecular Biology 2013;425(9):1527-1545.

13. Axe JM, Yezdimer EM, O'Rourke KF, Kerstetter NE, You W, Chang C-eA, Boehr DD. Amino Acid Networks in a (β/α)8 Barrel Enzyme Change during Catalytic Turnover. Journal of the American Chemical Society 2014;136(19):6818-6821.

14. O'Rourke KF, Axe JM, D'Amico RN, Sahu D, Boehr DD. Millisecond Timescale Motions Connect Amino Acid Interaction Networks in Alpha Tryptophan Synthase. Frontiers in Molecular Biosciences 2018;5:92.

15. Nishio K, Morimoto Y, Ishizuka M, Ogasahara K, Tsukihara T, Yutani K. Conformational changes in the alpha-subunit coupled to binding of the beta 2-subunit of tryptophan synthase from Escherichia coli: crystal structure of the tryptophan synthase alpha-subunit alone. Biochemistry 2005;44(4):1184-92.

16. Ngo H, Kimmich N, Harris R, Niks D, Blumenstein L, Kulik V, Barends TR, Schlichting I, Dunn MF. 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. Biochemistry 2007;46(26):7740-7753.

17. Weyand M, Schlichting I. Crystal structure of wild-type tryptophan synthase complexed with the natural substrate indole-3-glycerol phosphate. Biochemistry 1999;38(50):16469-80.

18. Barends TR, Domratcheva T, Kulik V, Blumenstein L, Niks D, Dunn MF, Schlichting I. Structure and mechanistic implications of a tryptophan synthase quinonoid intermediate. Chembiochem 2008;9(7):1024-8.

19. Salomon-Ferrer R, Gotz AW, Poole D, Le Grand S, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. J Chem Theory Comput 2013;9(9):3878-88.

20. Case DA, Ben-Shalom IY, Borzell SR, Cerutti DS, Cheatham TE, Cruzeiro VWD, Darden TA, Duke RE, Ghoreishi MK, Gilson MK and others. AMBER 2018. University of California, San Francisco2018.

21. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput 2015;11(8):3696-713.

22. Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J Comput Chem 2002;23(16):1623-41.

23. 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):926-935.

24. 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):327-341.

25. Sagui C, Pedersen LG, Darden TA. Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations. Journal of Chemical Physics 2004;120(1):73-87.

26. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14(1):33-8, 27-8.

27. *Molecular Operating Environment (MOE)*. 2018.01. 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7: Chemical Computing Group ULC; 2018.

28. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. Journal of Chemical Theory and Computation 2013;9(7):3084-3095.

29. Ai R, Fatmi MQ, Chang CEA. T-Analyst: a program for efficient analysis of protein conformational changes by torsion angles. J Comput Aid Mol Des 2010;24:819-827.

30. Stacklies W, Seifert C, Graeter F. Implementation of force distribution analysis for molecular dynamics simulations. BMC Bioinformatics 2011;12:101.

31. Ai R, Qaiser Fatmi M, Chang C-eA. T-Analyst: a program for efficient analysis of protein conformational changes by torsion angles. Journal of Computer - Aided Molecular Design 2010;24(10):819.

32. Weyand M, Schlichting I. Crystal Structure of Wild-Type Tryptophan Synthase Complexed with the Natural Substrate Indole-3-Glycerol Phosphate. Biochemistry 1999;38:16469.

33. Weyand M, Schlichting I. Structural basis for the impaired channeling and allosteric inter-subunit communication in the beta A169L/beta C170W mutant of tryptophan synthase. Journal of Biological Chemistry 2000;275(52):41058-41063.

34. Yang XJ, Miles EW. Threonine-183 and adjacent flexible loop residues in the tryptophan synthase-alpha subunit have critical roles in modulating the enzymatic-activities of the beta-subunit in the alpha-2-beta-2 complex. Journal of Biological Chemistry 1992;267(11):7520-7528.

35. Hiraga K, Yutani K. Roles of hydrogen bonding residues in the interaction between the alpha and beta subunits in the tryptophan synthase complex - Asn-104 of the alpha subunit is especially important. Journal of Biological Chemistry 1997;272(8):4935-4940.

Chapter 3 Discovery of Antimicrobial Agents Targeting the α Subunit of Tryptophan Synthase

3.1 Introduction

Multidrug-resistant pathogens are an increasing danger to public health, with more than 2.8 million infections reported annually in the United States, resulting in more than 35 000 deaths.[1] The biochemical mechanisms by which bacterial cells resist treatment have been classified into four general types, including modification of the drug, alteration of the target site, decreased permeability and overall structural or functional adaptations.[2,3] These genetically dictated adjustments in bacteria render many of the currently used antibiotics ineffective and lead to more severe symptoms and high hospitalization and death rates in infected patients.[3,4] The US Center for Disease Control and Prevention has listed clinically resistant *Salmonella* strains at the top of the 10 serious public health threats, with markedly increasing resistance to commonly used antibiotics.[1] Drug-resistant *Escherichia coli* strains are also of concern because of their high occurrence in humans and animals.[5] Therefore, the development of novel drugs targeting vital bacterial pathways is essential to successfully fight bacterial infections.

One critical pathway in microorganisms is L-tryptophan (L-Trp) biosynthesis.[6,7] In bacteria, yeast, molds, and plants, the tryptophan operon encodes for enzymes involved in metabolism of this amino acid. Auxotrophic mutants of pathogenic bacteria with defects in L-Trp biosynthesis lose virulence within a host organism.[8] For higher organisms, L-Trp is an essential amino acid and can only be acquired from diet.[9] Antigenic and

inflammatory signals lead to activation of the indoleamine 2,3 dioxygenase pathway responsible for L-Trp catabolism, thus resulting in L-Trp depletion upon microbial infection.[10,11] With the bacterial survival capacity under L-Trp starvation, the limited supply of the amino acid in host cells, and the lack of comparable biosynthetic machinery in higher animals and humans, components of the L-Trp biosynthetic pathway are a highly specific drug target.[12,13]

Our study focused on inhibiting the function of the α subunit of tryptophan synthase (TRPS), a bienzyme complex that catalyzes the last two steps of tryptophan biosynthesis. TRPS is a heterodimer, but in a biological setting, it exists as a tetramer, and the two dimers connect linearly with two β subunits attached to each other **(**Figure 3.1A**)**. The enzyme α subunit catalyzes the first reaction, the cleavage of indole-3-glycerol phosphate (IGP), then the indole product is transferred through a 25-Å hydrophobic channel to the β site, where it is condensed with L-serine in a pyridoxal phosphate (PLP)-dependent reaction to form tryptophan. The glyceraldehyde-3-phosphate is released from the α subunit as a byproduct **(**Figure 3.1B and 3.1C**)**.[14] A wide range of interactions between α and β subunits of TRPS, predominantly between residues of the α loop 2 (αL2) and β COMM domain, regulate the overall catalytic activity of the enzyme.[15] An important aspect of this allosteric regulation is the shifting between open inactive and closed active conformations of the active sites of both subunits induced by the ligands bound to the α active site and the covalent intermediates formed at the β active site.[16-18] The closed conformation of the flexible α loop 6 (αL6) has a crucial role in ligand binding and catalytic activity of the TRPS α subunit.[15,19]

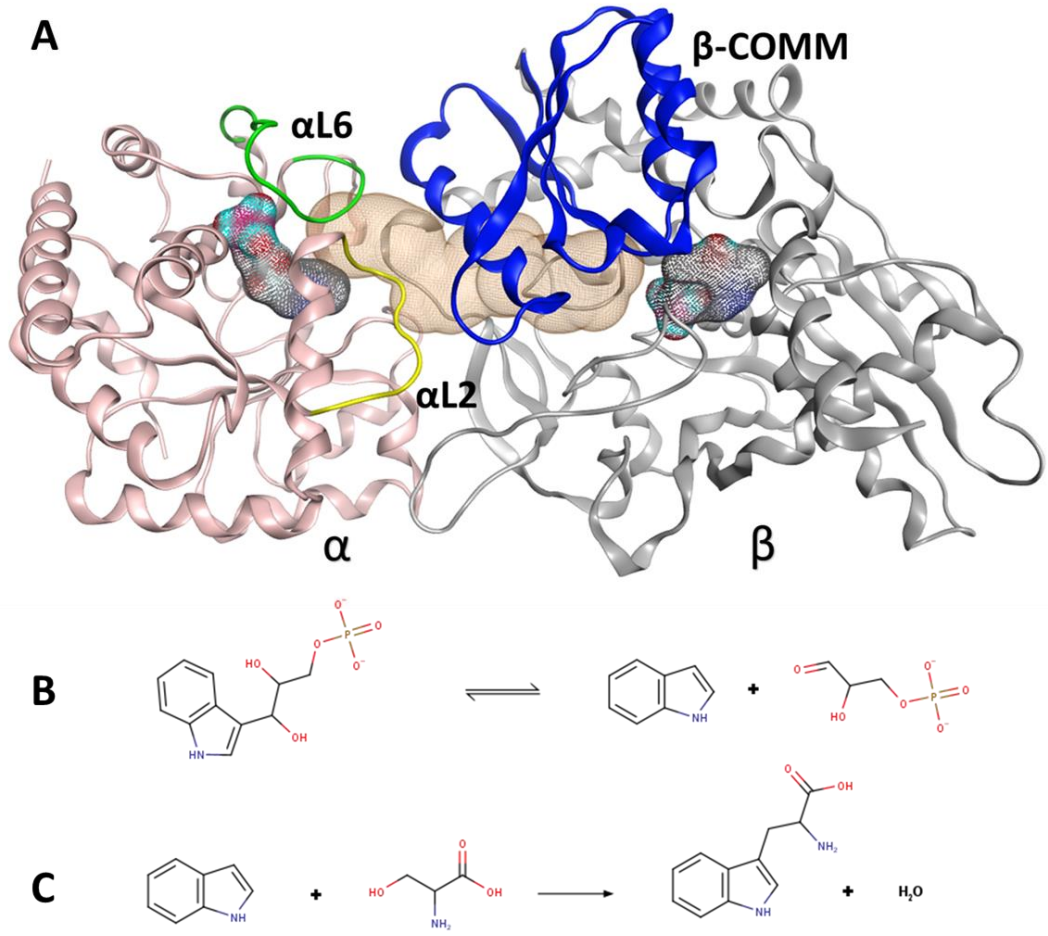


Figure 3.1. Tryptophan synthase (TRPS) overview. (A) Ribbon representation of TRPS αβ complex (α – pink, β – grey). Flexible loop 2 (αL2) and loop 6 (αL6) are in yellow and green, respectively, and the communication domain (β-COMM) is in blue. Active sites are indicated in atom color surface representation. Tunnel, connecting α and β subunits is in brown. (B) α subunit reaction – cleavage of indole-3-glycerol phosphate (IGP). (C) β subunit reaction – condensation of indole with L-serine (Ser) to produce L-Trp. MarvinSketch was used for drawing the reactions.[20]

This study is a continuation of the work of a former group member, in which the dynamic nature of proteins was incorporated to perform *in silico* screening of 1800 compounds obtained from the National Cancer Institute (NCI) Diversity Set I. Protein conformational changes play an important role in ligand binding [21,22]; therefore, four conformations from previously performed molecular dynamics (MD) simulations and three experimental structures were selected for docking screening. Ranking of compounds was based on the lowest binding energy from docking results and scores from their absorption, distribution, metabolism, and excretion (ADME) properties. A whole-cell minimum inhibitory concentration (MIC) assay was used to evaluate in vitro inhibition of the selected 28 compounds and one compound, 3-amino-3-imino-2-phenyldiazenylpropanamide (Compound 1), was identified to significantly prevent bacterial growth. In addition to the α subunit binding site assessed in the docking experiment, crystallization of the protein in complex with Compound 1 (PDB ID 6XIN) showed an alternative binding mode for the ligand at the αβ interface of the enzyme. Modeling work further supported the binding ability of Compound 1 in both sites. MD simulations were performed to identify specific ligand-protein interactions and provide prospects for improving this lead compound. The compound is a promising lead for further development of a strong antimicrobial agent targeting the L-Trp metabolic pathway.

3.2 Materials and Methods

3.2.1 Molecular Systems

The coordinates of a fully closed active conformation of TRPS were obtained from the crystal structure of PDB ID 4HN4[17] with the α subunit ligand F9F removed and α-aminoacrylate ligand in the β subunit. Compound 1 was docked in the active site of the α subunit by using the Molecular Operating Environment (MOE)[23] with an induced fit option and free side chain rotation allowed. Two docking poses were selected with respect to the position of the propenamide portion of the compound. The interface binding mode was obtained by alignment with the newly crystallized structure of the complex PDB ID 6XIN.

3.2.2 Molecular Dynamics Simulations

We performed three 200-ns MD simulations of TRPS in complex with the strongest inhibitor, (3-amino-3-imino-2-phenyldiazenylpropanamide (Compound 1), to evaluate two binding sites: the active site of the α subunit and the αβ subunit inter-domain interface. MD simulations were carried out using the standard Amber package with GPU acceleration.[24,25] The protein was parameterized by using Amber Force Field FF14SB.[26] General Amber force field (GAFF) was applied to ligands and charges were assigned by using the AM1-BCC model.[27] All systems were prepared by a three-step minimization process: hydrogens, sidechains, and finally the whole system. The systems were solvated by using an explicit TIP3P water model in a rectangular box with edges at a minimum of 12 Å from any atom.[28] We added 18 positive counter ions (Na^+) to neutralize the overall

system charge. The solvated systems were then minimized, followed by 750-ps water equilibration at 298 K and a whole-system equilibration from 50 to 298 K at 25-K intervals for a total of 1.3 ns (50 to 275 K, 100 ps each; 400 ps at 298 K). MD trajectories were collected over 200 ns at a 1-ps interval with a 2-fs timestep under constant pressure and temperature. Particle mesh Ewald was used for long-range electrostatics and the SHAKE algorithm was applied for fixed heavy atom–hydrogen bond lengths.[29,30] The systems were visualized and analyzed by using Visual Molecular Dynamics [31] and MOE. The trajectory output files were processed with PTRAJ software[32] to contain 20000 frames, each representing 0.01-ns timestep.

3.3 Results and Discussion

To investigate specific interactions between TRPS and Compound 1, we performed three MD simulations initiated by complex structures reflecting two binding poses, A and B, for Compound 1 in the active site of the α subunit and a pose C in the inter-domain interface shown in our crystal structure (Figure 3.2A). We used the Molecular Operating Environment (MOE) program to dock Compound 1 in the α active site to ensure that the position of Compound 1 suggested by the Vdock program could be reproduced by a different docking algorithm. Consistent with the Vdock results, the phenyl ring, in the 10 highest-scoring poses with MOE, occupied the small cleft where the phosphate group of the natural substrate IGP binds (Figure 3.2). We observed two orientations for the propanamide moiety from our docking results; therefore, two initial conformations, poses A and B, were used for the two MD runs. The MD simulations showed that the hydrogen bond between αAsp60 and the amino/imino groups of Compound 1 is more favorable

because after dihedral rotation in pose A, the same hydrogen bonds are formed as in pose B (Figure 3.2A, green dotted line). As a result, both poses A and B held two stable hydrogen bonds between the carboxyl oxygen of compound 1 and αTyr175 as well as the amino group of the ligand and αAsp60 (Figure 3.2B). At the interface, αAsp130 coordinates the amido and imino groups, whereas the amino and carboxyl groups establish a hydrogen bond with the backbone oxygen of βIle278 and the backbone nitrogen of βGly281, respectively (Figure 3.2B).

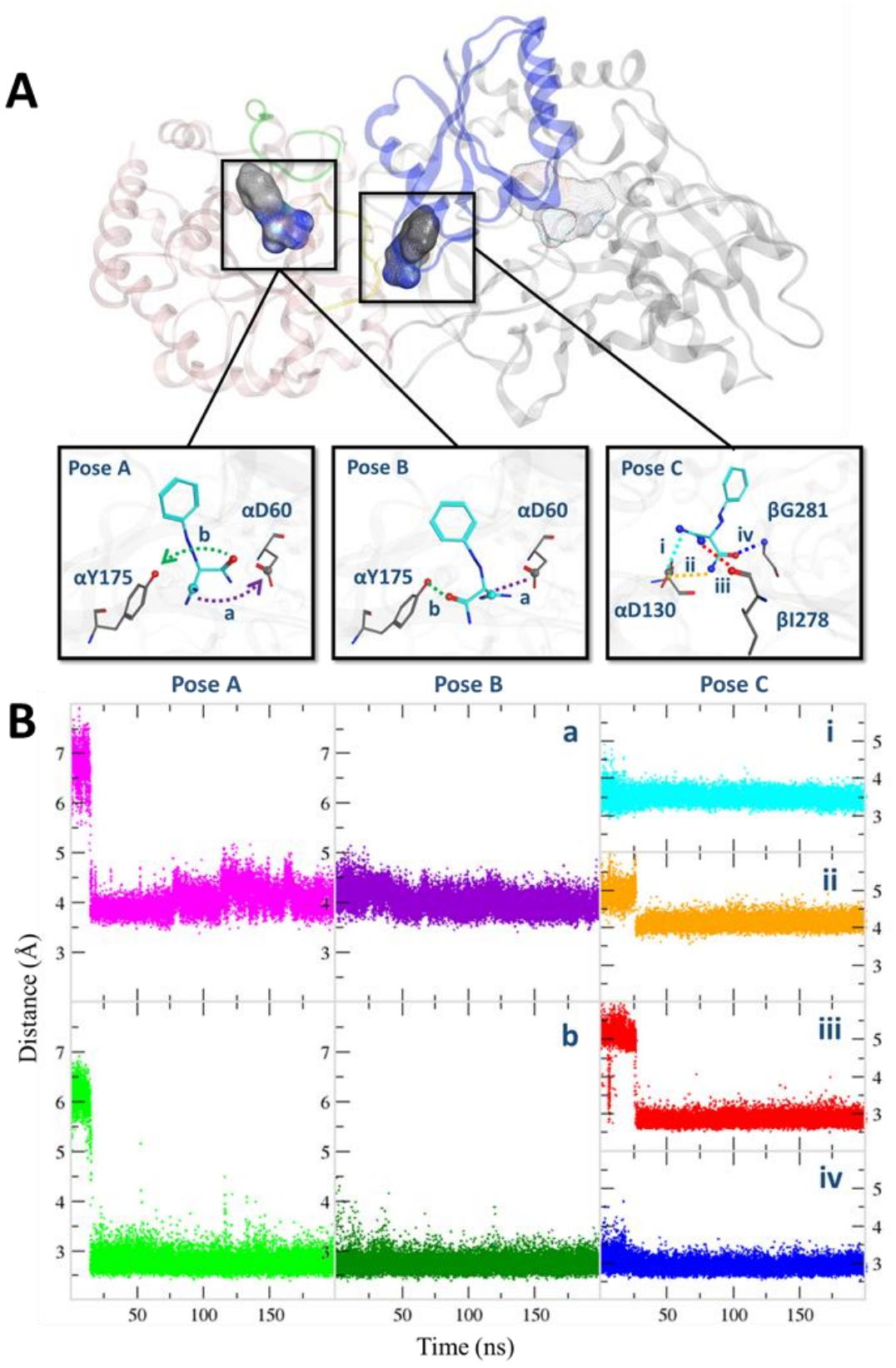

A
Pose A
αD60
b
αY175
a
Pose B
αD60
αY175
b
a
Pose C
βG281
i
iv
ii
iii
αD130
βI278
B
Pose A
Pose B
Pose C
a
b
i
ii
iii
iv
Distance (Å)
Time (ns)
50
100
150
3
4
5
6
7

Figure 3.2. Compound 1 conformations and interactions with protein in the α-subunit active site and interdomain surface. (A) Three selected initial conformations of Compound 1 for molecular dynamics (MD) simulations. Compound 1 can bind to two distinct binding sites: the active site (poses A and B) of the α subunit, and that in the αβ inter-domain interface (pose C). The two major conformations in the active site are characterized by a 180° rotation of the amido group resulting in stable hydrogen bonds with the nearby residues. (B) Intermolecular hydrogen bonds between Compound 1 and TRPS during 200-ns MD runs. Hydrogen bonds between Compound 1 and residues D60 (purple, a) and Y175 (green, b) in the active site of α subunit. At the αβ inter-domain interface, hydrogen bonds are observed between residues D130 of α subunit (cyan, i and orange, ii), I278 of β subunit (red, iii) and G281 of β subunit (blue, iv).

Root-mean-square deviation (RMSD) for protein backbone showed that the systems are well equilibrated, with no significant deviations (Figure 3.3A). After initial adjustments for each binding pose, the ligand was stable in the binding pocket for all three systems, as indicated by the calculated RMSD for the heavy atoms of the compound (Figure 3.3B).

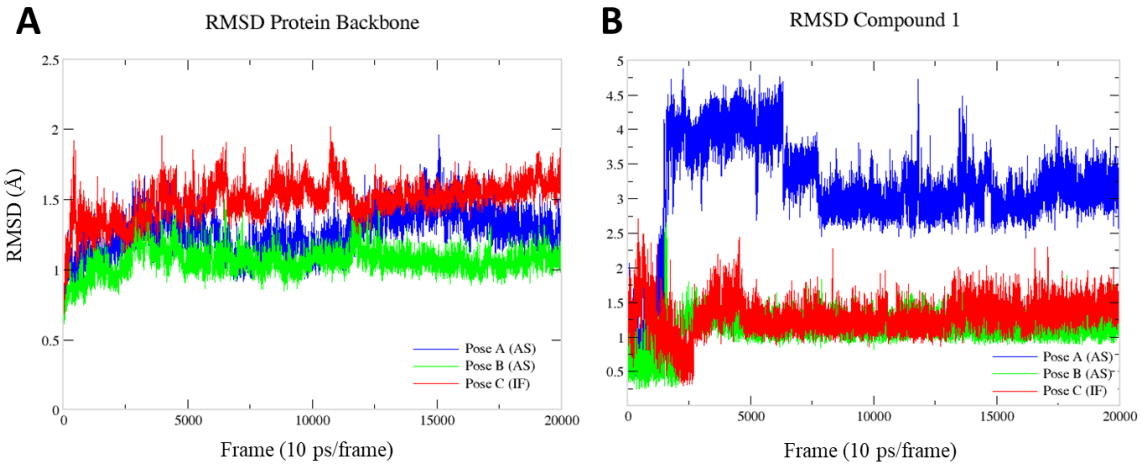


Figure 3.3. Root-mean square deviation (RMSD) for protein backbone and compound 1 within binding site. (A) All three compound 1 bound systems are well equilibrated and significant deviations were not observed. (B) Compound 1 is stable in both binding sites. The initial spike in pose A is due to the rotation of propanamide moiety.

We also compared root-mean square fluctuation data (RMSF) for the three systems to analyze the overall dynamics of the protein and identify more flexible regions, specifically αL6 (Figure 3.4). A closed ordered conformation of this loop is essential for retaining the substrate in the binding site.[18,19] In pose A and interface pose C, the upper portion of αL6 (residues 186 to 191) is more flexible; nonetheless a consistent hydrogen bond between αAsp60 and αThr183 indicates that αL6 remains closed (Figure 3.5). Pose B shows a highly ordered conformation for αL6, comparable to the natural substrate IGP-bound complex, whereas the αL6 appears more dynamic in the ligand-free enzyme (Figure 3.4B and 3.4C). This ordered state of the loop further indicates the stability of Compound 1 within α subunit active site. Conformational fluctuation of the β subunit is similar in all systems. Because the motions of the β subunit are allosterically regulated by the α subunit ligand, substrate binding should alter the overall dynamics of the β subunit as well.[33,34]

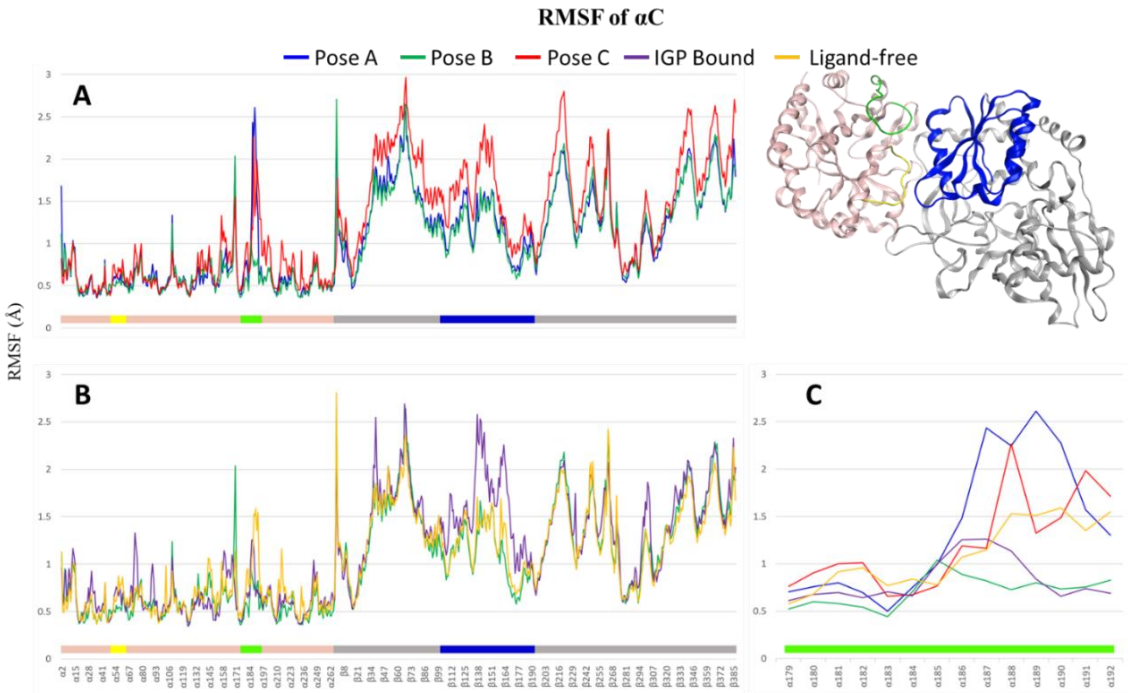


Figure 3.4. Conformational fluctuation of TRPS with Compound 1. RMSF data calculated relative to the initial position of αC of each residue. The X-axis presents the residue numbers of the complex, the color bar indicates the structural region of the enzyme (α-subunit – pink, αL2 – yellow, αL6 – green, β-subunit – grey, βCOMM domain – blue). (A) RMSF computed from three 200-ns MD runs, with Compound 1 bound to the active site of α subunit (poses A and B) or the αβ-inter-domain interface (pose C), show similar dynamics. αL6 (green) is more flexible in poses A and C, and the loop remains stable in pose B. (B) Comparison of RMSF computed from pose B, MD simulation in complex with natural substrate IGP and ligand-free TRPS. (C) αL6 fluctuation in ligand-free TRPS and TRPS in complex with Compound 1 (3 poses) and IGP.

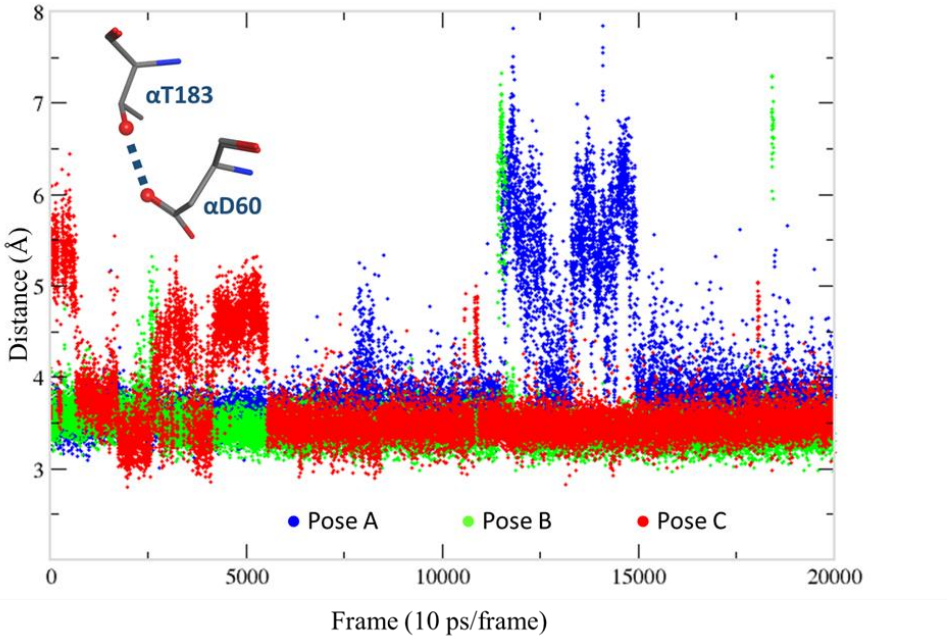


Figure 3.5. Hydrogen bond distances between αAsp60 and αT183 for three simulated binding poses. αL6 remains closed in pose B and shows slight fluctuation in poses A and C.

Further modifications to Compound 1 based on interactions revealed in our modeling work may improve binding in the active site. Adding a hydrogen bond donor, such as carboxyl substitution at the meta or para position of the phenyl ring could likely increase hydrogen bond interactions, specifically with αGly184, αGly213 and αSer235, which are shown to coordinate the phosphate group in IGP. The addition of an aromatic ring or cyclic hydrocarbon to the imino nitrogen could further increase affinity by engaging in hydrophobic interactions with Phe22 and Leu100. By using MOE to evaluate our design strategies, five Compound 1 analogs were docked to the active site of the α subunit (Figure 3.6A). Docking results showed that the addition of a ring improved the binding score by more than 2 kcal/mol as compared with Compound 1 (Figure 3.6B). However, the addition of a carboxyl group to the phenyl ring at para position without an additional ring structure only slightly improved the docking score, which suggests that the binding pocket has unutilized space for further investigation.

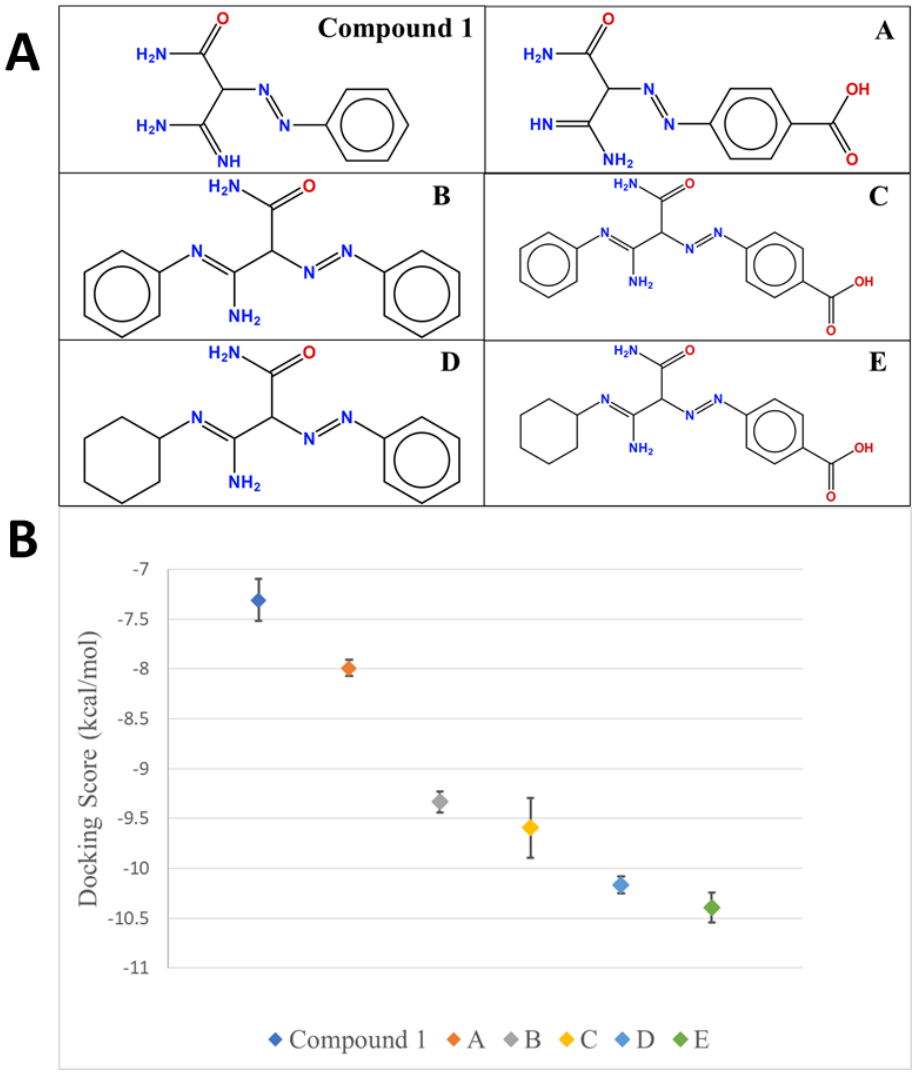


Figure 3.6. Docking of modified Compound 1. (A) Compound 1 substitutions. (B) Average score of the five lowest-scoring docking poses modeled by the Molecular Operating Environment (MOE) program.

3.4 Conclusions

Antibiotic resistance poses a great challenge in treating bacterial infections and presents a serious public health threat. The development of novel drugs to fight these infections is vital to prevent further health crises. The tryptophan biosynthesis pathway is essential for bacterial growth. In this work, we aimed to identify an inhibitor for the α subunit of TRPS. With combined computational and experimental approaches, we identified a propanamide-based compound with satisfactory pharmacokinetic properties. In vitro MIC testing showed that the selected compound significantly reduced bacterial growth, and equally important, adding L-Trp to the medium recovered bacterial growth, indicating that the compound specifically targets the L-Trp biosynthetic pathway. In addition to the known binding pocket in the α subunit, MD simulations and crystallization of the TRPS-Compound 1 complex revealed another stable binding site of Compound 1 in the αβ-inter-domain interface. Our MD runs for three different binding poses showed that TRPS was stable when Compound 1 was bound to either binding site. TRPS showed similar dynamic fluctuations when its natural substrate IGP or compound 1 (pose B) was bound to the active site. We further modified Compound 1 to enhance binding, and our docking results showed an improved binding score after the addition of substituents. In summary, our work shows that compound 1 is a good lead for further improvement and development of a potent antimicrobial drug.

3.5 References

1. 2019 Antibiotic resistance threats in the United States. Center for Disease Control and Prevention <https://www.cdc.gov/drugresistance/biggest-threats.html>.

2. Hawkey PM. The origins and molecular basis of antibiotic resistance. British Medical Journal 1998;317(7159):657-660.

3. Munita JM, Arias CA. Mechanisms of Antibiotic Resistance. Microbiology Spectrum 2016;4(2):24.

4. Angelo KM, Reynolds J, Karp BE, Hoekstra RM, Scheel CM, Friedman C. Antimicrobial resistance among nontyphoidal salmonella isolated from blood in the united states, 2003-2013. Journal of Infectious Diseases 2016;214(10):1565-1570.

5. Rasheed MU, Thajuddin N, Ahamed P, Teklemariam Z, Jamil K. Antimicrobial drug resistance in strains of Escherichia coli isolated from food sources. Revista Do Instituto De Medicina Tropical De Sao Paulo 2014;56(4):341-346.

6. Wellington S, Nag PP, Michalska K, Johnston SE, Jedrzejczak RP, Kaushik VK, Clatworthy AE, Siddiqi N, McCarren P, Bajrami B and others. A small-molecule allosteric inhibitor of Mycobacterium tuberculosis tryptophan synthase. Nature Chemical Biology 2017;13(9):943-+.

7. Lott JS. The tryptophan biosynthetic pathway is essential for Mycobacterium tuberculosis to cause disease. Biochemical Society Transactions 2020;48(5):2029-2037.

8. 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.

9. Rose WC, Haines WJ, Warner DT. The amino acid requirements of man: v. The rôle of lysine, arginine, and tryptophan. Journal of Biological Chemistry 1954;206(1):421-430.

10. MacKenzie CR, Heseler K, Muller A, Daubener W. Role of indoleamine 2,3-dioxygenase in antimicrobial defence and immuno-regulation: Tryptophan depletion versus production of toxic kynurenines. Current Drug Metabolism 2007;8(3):237-244.

11. Munn D, Mellor A. Indoleamine 2,3 dioxygenase and metabolic control of immune responses. Trends in Immunology 2013;34(3):137-143.

12. Zhang YJJ, Reddy MC, Ioerger TR, Rothchild AC, Dartois V, Schuster BM, Trauner A, Wallis D, Galaviz S, Huttenhower C and others. Tryptophan Biosynthesis Protects Mycobacteria from CD4 T-Cell-Mediated Killing. Cell 2013;155(6):1296-1308.

13. Naz S, Farooq U, Ali S, Sarwar R, Khan S, Abagyan R. Identification of new benzamide inhibitor against α-subunit of tryptophan synthase from Mycobacterium tuberculosis through structure-based virtual screening, anti-tuberculosis activity and molecular dynamics simulations. Journal of Biomolecular Structure and Dynamics 2019;37(4):1043-1053.

14. Creighton TE. A steady-state kinetic investigation of the reaction mechanism of the tryptophan synthetase of Escherichia coli. Eur J Biochem 1970;13.

15. Schneider TR, Gerhardt E, Lee M, Liang PH, Anderson KS, Schlichting I. Loop closure and intersubunit communication in tryptophan synthase. Biochemistry 1998;37(16):5394-5406.

16. 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.

17. Niks D, Hilario E, Dierkers A, Ngo H, Borchardt D, Neubauer TJ, Fan L, Mueller LJ, Dunn MF. Allostery and Substrate Channeling in the Tryptophan Synthase Bienzyme Complex: Evidence for Two Subunit Conformations and Four Quaternary States. 2013.

18. Dunn MF. Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase bienzyme complex. Archives of Biochemistry and Biophysics 2012;519(2):154-166.

19. Naz S, Farooq U, Khan S, Sarwar R, Mabkhot YN, Saeed M, Alsayari A, Bin Muhsinah A, Ul-Haq Z. Pharmacophore model-based virtual screening, docking, biological evaluation and molecular dynamics simulations for inhibitors discovery against alpha-tryptophan synthase from Mycobacterium tuberculosis. Journal of Biomolecular Structure & Dynamics 2021;39(2):610-620.

20. MarvinSketch. Volume ChemAxon: https://www.chemaxon.com.

21. Axe JM, Yezdimer EM, O'Rourke KF, Kerstetter NE, You W, Chang C-eA, Boehr DD. Amino Acid Networks in a (β/α)8 Barrel Enzyme Change during Catalytic Turnover. Journal of the American Chemical Society 2014;136(19):6818-6821.

22. O'Rourke KF, Sahu D, Bosken YK, D'Amico RN, Chang CEA, Boehr DD. Coordinated Network Changes across the Catalytic Cycle of Alpha Tryptophan Synthase. Structure 2019;27(9):1405-+.

23. *Molecular Operating Environment (MOE)*. 2018.01. 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7: Chemical Computing Group ULC; 2018.

24. D.A. Case IYB-S, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman. Amber 2018: University of California, San Francisco.; 2018.

25. Salomon-Ferrer R, Götz AW, Poole D, Grand SL, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. 2013.

26. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. 2015.

27. Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. Journal of Computational Chemistry 2002;23(16):1623-1641.

28. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics 1998.

29. Sagui C, Pedersen LG, Darden TA. Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations. The Journal of Chemical Physics 2003.

30. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational Physics 1977;23(3):327-341.

31. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. Journal of Molecular Graphics 1996;14(1):33-38.

32. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. Journal of Chemical Theory and Computation 2013;9(7):3084-3095.

33. Fatmi MQ, Ai R, Chang C-eA. Synergistic Regulation and Ligand-Induced Conformational Changes of Tryptophan Synthase. Biochemistry 2009;48(41):9921-9931.

34. Fatmi MQ, Chang C-eA. The Role of Oligomerization and Cooperative Regulation in Protein Function: The Case of Tryptophan Synthase. PLoS Comput Biol 2010;6(11):e1000994.

Chapter 4 Insights into Dynamics of Inhibitor and Ubiquitin-like Protein Binding in SARS-CoV-2 Papain-like Protease

4.1 Introduction

Covid-19, caused by a novel form of coronavirus, has created a global health crisis due to the lack of vaccines and anti-viral drugs. Over the past two decades, coronaviruses such as the Severe Acute Respiratory Syndrome coronavirus (SARS-CoV-1 or CoV1) and Middle East Respiratory Syndrome coronavirus (MERS-CoV) have caused mass human fatality. In late 2019, the novel form of coronavirus, known as SARS-CoV-2 (CoV2), spread rapidly from Wuhan, China to all continents of the world within months, causing widespread mortality and worldwide panic.[1] The only way to curtail the spread of the virus thus far has been through strict, indefinite quarantine of millions of people. Clearly, development of anti-viral drugs capable of inhibiting CoV2 is of paramount importance.

CoV2 contains a Papain-like protease (PLpro) that is vital for viral replication.[2] PLpro is responsible for the proteolytic processing of the product of open reading frame 1a (ORF1a) in the replicase gene of CoV2, a large viral polyprotein containing nonstructural proteins which form the replicase complex.[3] PLpro exists as a monomer in biological settings and has the USP fold, typical for the ubiquitin-specific proteases (USP) family in humans, which is topologically organized into four domains – UBL, thumb, palm, and fingers (Figure 4.1A).[4] The peptide bond cleavage in the active site is catalyzed by a conserved catalytic triad comprised of residues Cys111, His272 and Asp286.[5] In addition, PLpro possesses deubiquitinating and deISGylating capabilities[6] which interfere with

critical signaling pathways leading to the expression of type I interferons, resulting in antagonistic effect on host innate immune response.[7,8] Therefore, inhibition of PLpro activity can halt viral replication and disrupt its role in host immune response evasion, making it an excellent anti-viral drug target.

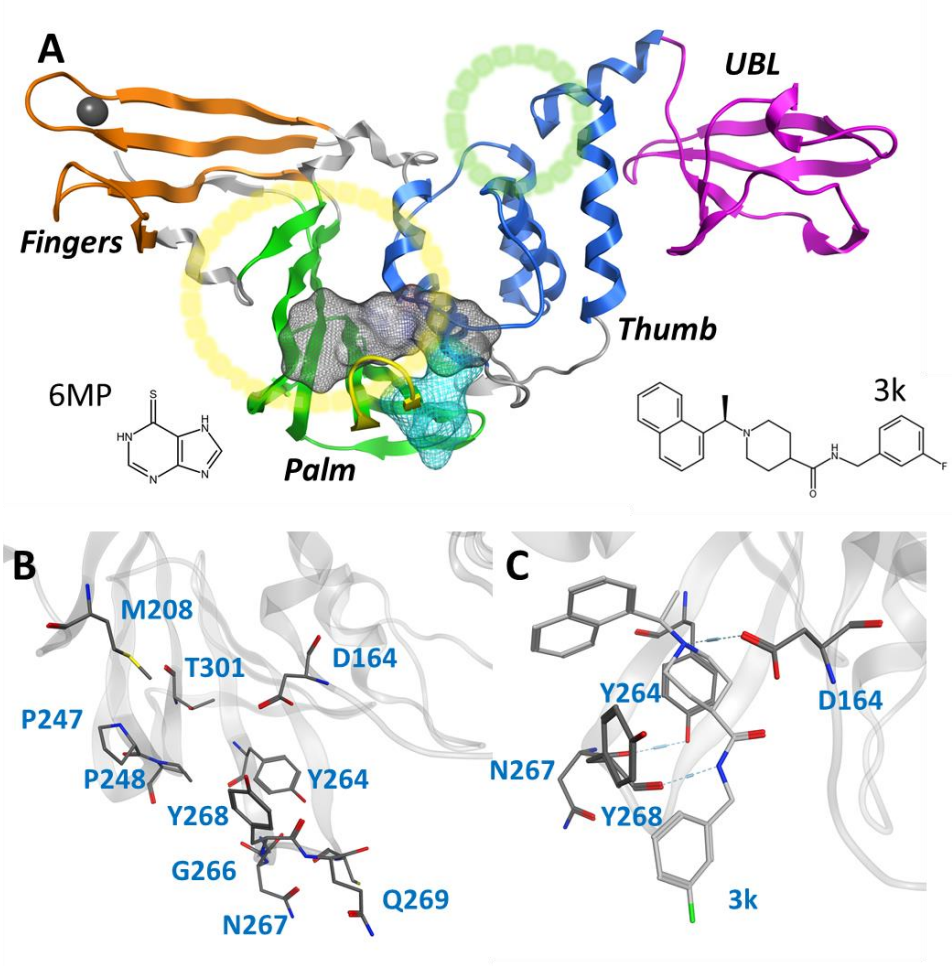


Figure 4.1. Cartoon representation of the entire CoV2 PLpro structure and close-ups of regions important to ligand binding. (A) PLpro with the four domains and other major regions indicated as follows: fingers – orange, palm – green (BL2 loop – yellow), thumb – blue, UBL – magenta, SUb1 and SUb2 – yellow and green circles, respectively. The putative 3k binding site is shown as a grey surface and the active site as a teal surface. 6MP was docked to the putative 3k site and active site. (B) important binding site residues. (C) 3k (light grey) engaging in hydrogen bonds with D164 and Y268, and the important BL2 loop-stabilizing hydrogen bond between Y264 and N267.

CoV2 PLpro exhibits a high sequence similarity to CoV1 PLpro (Figure 4.2); in particular, the binding site and active site residues are nearly identical. We have introduced a leading naphthalene-based inhibitor, 3k, and chemotherapy agent 6-mercaptopurine (6MP) (Figure 4.1A), which successfully inhibited CoV1 PLpro and MERS-CoV PLpro, respectively,[9-11] to assess their binding affinity for CoV2 PLpro. The 3k binding site is adjacent to the catalytic triad and sterically inhibits the binding of ubiquitin (Ub) and Interferon-stimulated gene 15 (ISG15) by occupying the space normally reserved for their C-terminal (LXGG cleavage site) at ubiquitin binding subsite 1 (SUb1) (Figure 4.1A). Compounds capable of binding to this site therefore exhibit high inhibitory capabilities.

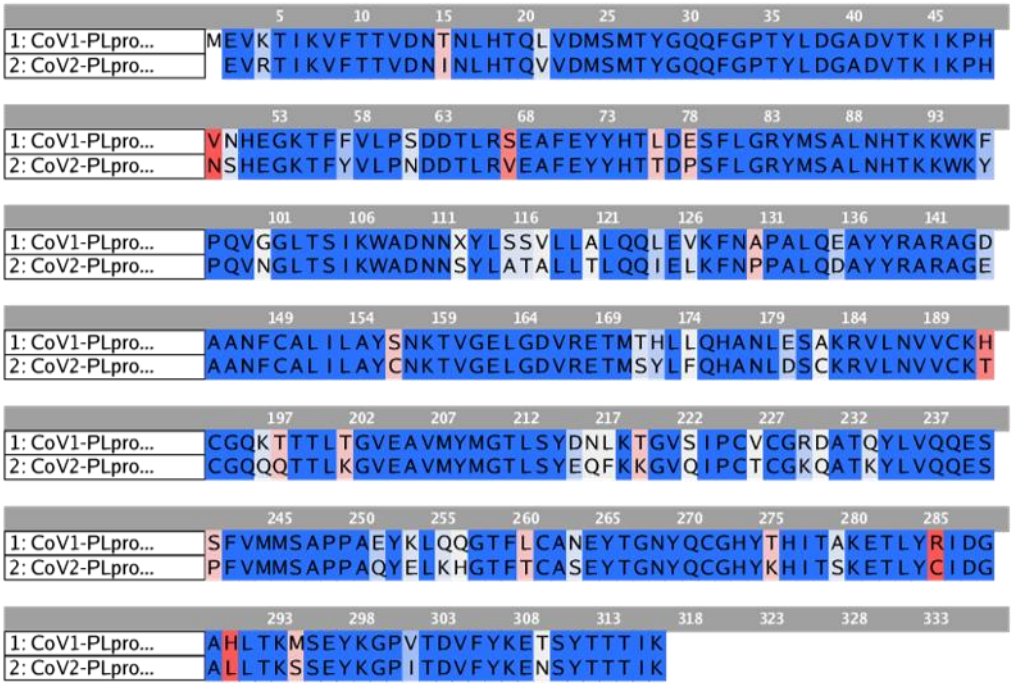


Figure 4.2 Amino acid sequence alignment of CoV1 and CoV2 PLpro

In this work, we carried out several molecular dynamics (MD) simulations of ligand-free and ligand-bound CoV2, CoV1, and MERS-CoV PLpro (Table 4.1). Based on detailed examination of the CoV2 3k binding site, we provide guidance and suggestions for optimization of compounds targeting this site. Moreover, by simulating 3k bound to CoV2 and CoV1 PLpro, we show that it exhibits a highly similar binding mode in both proteins, suggesting that 3k and similar compounds should have an inhibitory effect on CoV2 PLpro. After analyzing the binding mode and binding site, we constructed and docked new ligands based on the 3k scaffold which showed improved binding affinity over the current molecule. Additionally, we carried out experimental assays to validate 3k binding to CoV2 PLpro and inhibit enzymatic function. We show that the overall dynamics of ligand-free PLpro in all analyzed systems is highly similar, with comparable flexibility in BL2 loop, zinc-binding region and UBL domain. Our detailed description of 3k binding in the protein provides insight into the essential interactions necessary for successful fragment-based drug design. Additionally, we provide well-sampled dynamics of the available CoV2 PLpro crystal structures for wider use as a guide to potential drug binding sites or in docking and drug screening studies.

Table 4.1. Summary of all simulations performed. All ligand-free proteins were simulated twice under identical conditions except for the initial random number seed, first for 1 μs, followed by a 500 ns secondary run to confirm consistency in the observed dynamics. Similarly, all ligand-bound proteins were simulated three times for at least 200 ns. Where necessary, secondary and tertiary runs are referred to by a dash and number after the main designation e.g., MD1-2 means the second run of simulation MD1.

Summary of Simulations			
Simulation Index	**PDB**	**Protein system**	**Length**
MD1 (1, 2)	6W9C	CoV2 PLpro	1 μs, 500 ns
MD2 (1, 2)	6WRH	CoV2 PLpro	1 μs, 500 ns
MD3 (1, 2)	4OW0	CoV1 PLpro	1 μs, 500 ns
MD4 (1, 2)	4RNA	MERS-CoV PLpro	1 μs, 500 ns
MD5a (1-3)	6W9C	CoV2 PL pro complexed w/ 3k (pose A)	1 μs, 500 ns, 200 ns
MD5b (1-3)	6W9C	CoV2 PL pro complexed w/ 3k (pose B)	3 × 200 ns
MD5c (1-3)	6W9C	CoV2 PL pro complexed w/ 3k (pose C)	3 × 200 ns
MD5d (1-3)	6W9C	CoV2 PL pro complexed w/ 3k (pose D)	3 × 200 ns
MD6 (1-3)	4OW0	CoV1 PLpro complexed w/ 3k	1 μs, 500 ns, 200 ns
MD7a (1-3)	6W9C	CoV2 PLpro complexed w/ 6MP (in putative site)	3 × 200 ns
MD7b (1-3)	6W9C	CoV2 PLpro complexed w/ 6MP (in active site)	3 × 200 ns

4.2 Methods

4.2.1 MD Simulation Protocol

MD simulations were prepared and run using the *Amber18* molecular dynamics package with GPU acceleration.[12,13] Force fields ff14SB[14] and the general Amber force field (GAFF2)[15] were used on proteins and ligands, respectively. Ligands 3k are 6MP were parameterized using Amber's *antechamber* program with the AM1-BCC charge assignment method.[16] All systems were solvated with a rectangular box of explicit TIP3P water extending 12 Å beyond the solute edges, and 1 or 2 explicit counterions were added to neutralize the overall system charge. Systems were minimized in four steps. First, using Generalized Born implicit solvent,[17] we minimized the hydrogen atoms, then protein sidechains, and finally the entire protein for 500, 1000, and 5000 steps, respectively. Next, the entire solvated structure was minimized for 5000 steps. Solvated systems were equilibrated in the isothermic-isobaric (NPT) ensemble from 50 to 275 K in 25 K increments for 100 ps each, and finally at 298 K for 500 ps. Production simulations were performed in the NPT ensemble at 298K using the Langevin thermostat with a 2 fs timestep. A 12 Å cutoff distance was used for direct non-bonded energy calculations and long-range electrostatics were calculated by the particle mesh Ewald method.[18] The SHAKE algorithm[19] was employed to constrain all bonds involving hydrogen. Raw trajectories were saved every 2 ps and then processed using Amber's *cpptraj*[20] for analysis.

4.2.2 Selection of Initial Structures for MD Simulation

Initial coordinates for CoV2 PLpro simulations were obtained from two crystal structures of the ligand-free protein, PDB IDs 6W9C[21] and 6WRH.[22] The ligand-bound complexes for CoV2 were obtained by docking ligands into a protein conformation selected from MD1; details are provided in the following subsection. CoV1 PLpro simulations began from a crystal structure of a 3k-bound complex, PDB 4OW0.[9] Ligand 3k was manually removed from the binding site for our ligand-free CoV1 PLpro simulation. MERS-CoV PLpro was simulated only in the ligand-free state, starting from crystal structure 4RNA.[23] For simplicity, we have indexed these simulations as shown in Table 4.1.

4.2.3 Ligand Docking to CoV2 PLpro

Force distribution analysis tool (FDA)[24] was used to identify the residues interacting with 3k in CoV1 PLpro (Figure 4.3), and since these residues are identical in CoV2 PLpro, we used them as a ligand docking site. To choose a CoV2 PLpro conformation that was highly similar to the CoV1 3k-bound crystal structure, we found the CoV2 frame from MD1-1 with minimum RMSD between key binding site residues to use for docking (Figure 4.3). The ligands were docked to this single PLpro conformation using *Molecular Operating Environment* (MOE).[25] Four poses (Figure 4.4) of 3k in this site were selected for MD simulations. To obtain poses A and D, we used the induced fit docking option with constrained/tethered side chain rotations allowed; poses B and C were obtained using the same induced fit option with free sidechain rotation allowed. Poses A and B closely resemble the Cov1 PLpro–3k crystal structure, PDB 4OW0. Pose C is a rotamer

of A with a 180° rotation of the naphthalene moiety, while in pose D the piperidine moiety is rotated 180° with respect to A. In addition to the 3k binding site, 6MP was also docked to the active site following the same method. The designed ligands reflecting our suggested modifications to the 3k scaffold were docked in the 3k site by the same protocol to the same CoV2 PLpro conformation as 3k.

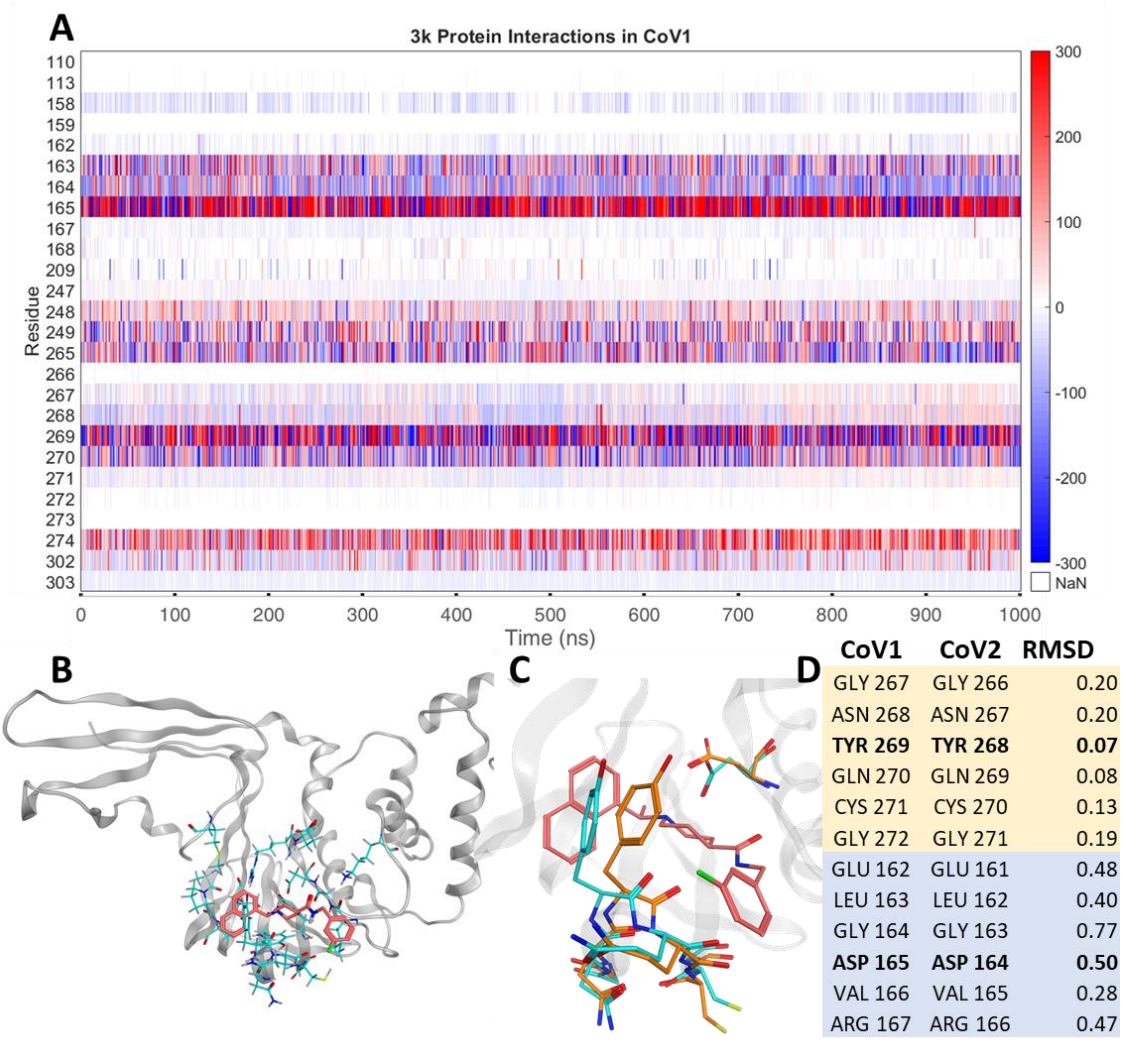


Figure 4.3. Residues selected as the docking site. (A) Residue-wise force distribution heatmap indicates protein residues interacting with ligand 3k in CoV1 PLpro. (B) Conformation of key residues in CoV1 PLpro (orange) and the conformation of the same residues selected for docking in CoV2 PLpro (teal). (C) The docking site residues selected based on force analysis displayed in CoV2 PLpro. (D) RMSD values between binding site residues in the minimized CoV1 PLpro crystal structure and CoV2 PLpro conformation used for docking. Force values are in piconewtons.

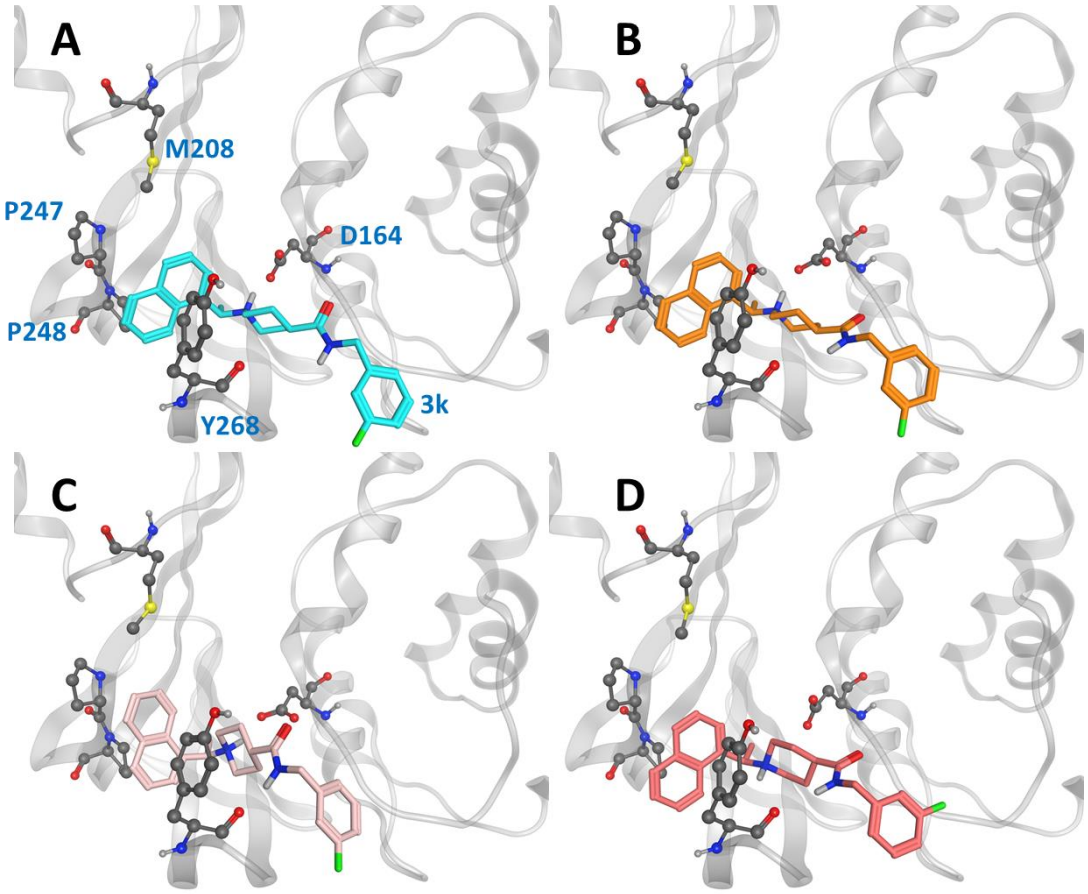


Figure 4.4. Ligand poses A-D from which the CoV2 PLpro-3k complex simulations began. Key binding site residues are shown in grey to show the difference in relative orientation of the ligand in each pose.

4.2.4 Simulation Analysis

4.2.4.1 Trajectory Visualization and Dihedral Analysis.

The simulations were visualized using *Visual Molecular Dynamics* (VMD)[26] and MOE. Dihedral angle populations and entropy were calculated using *T-Analyst*[27].

4.2.4.2 Cartesian Principal Component Analysis.

To observe major protein motions, we performed principal component analysis[28,29] of α-carbon atoms in the 1 μs trajectory of ligand-free CoV2 PLpro. PCA reduces the high-dimensional data set of all α-carbon motions throughout the MD trajectory to its principal components (PCs), the directions which contain the largest motions. We used the average α-carbon positions as references. The first and second largest PCs were analyzed to reveal the dominant motions.

4.2.4.3 MM/PBSA.

We used the MM/PBSA method[30] to evaluate the intermolecular interactions between ligands and PLpro. From a total of 20,000 MD frames making up the 200 ns ligand-bound trajectories, system conformations were analyzed every 2 ns. This method computes the energy (E) of a system from theprotein, ligand, and protein-ligand complex, and computes the interaction energy as $\Delta\langle E\rangle = \langle E_{complex}\rangle - \langle E_{protein}\rangle - \langle E_{ligand}\rangle$. $\langle E\rangle$ denotes the computed average energy from a given MD trajectory. The default values of a solute dielectric of 1.0 and solvent dielectric of 80.0 were used. The total binding energy term was computed as $E_{MM/PBSA} = E_{MM} + G_{PB} + Gnp$, where E_{MM} includes standard molecular mechanics force field terms, G_{PB} is the solvation energy computed by solving the Poisson Boltzmann (PB) equation, and Gnp is the nonpolar energy estimated from the solvent accessible surface area (A) as $\gamma A + b$, where γ is the surface tension and b is a correction term. We used $\gamma = 0.005$ kcal mol^{-1} Å^{-2} and b = 0.0 kcal mol^{-1} , and estimated the surface area with the LCPO method.[31]

4.3 Results and Discussion

We analyzed 1 μs trajectories of ligand-free CoV2, CoV1 and MERS-CoV PLpro to uncover the overall protein dynamics of the novel coronavirus protease and to make comparisons to older coronavirus PLpro for which inhibitors have been developed. In addition, we simulated ligand-bound trajectories of CoV1 and CoV2 PLpro to assess potential effectiveness of one naphthalene-based and one thiopurine inhibitor – 3k and 6MP, respectively – in the 2019 coronavirus. We showed that 3k formed stable interactions with CoV2 PLpro, suggesting that the compound can bind to the protein, which was verified by experimental assays. Moreover, we designed and docked new ligands based on the 3k-scaffold to CoV2 PLpro and show that they achieve lower binding free energies. Protein flexibility, entropy, and conformational changes were analyzed in the ligand-free protein simulations to characterize the overall protein dynamics and to assess similarities and differences relevant to inhibitor or Ub binding in CoV2 PLpro. The ligand-bound MD simulations were analyzed for a detailed characterization of ligand binding modes by analyzing residue-wise interactions, binding energy, and ligand-induced conformational changes.

4.3.1 Structure and Dynamics of Ligand-free CoV2 PLpro and Implications for Drug Discovery

Dynamic regions of potential importance to small molecule drug or Ub binding in CoV2 PLpro include portions of the thumb domain (containing SUb2), the fingers region (adjacent to SUb1) and the BL2 loop (directly adjacent to the 3k binding site). Principal

component analysis (PCA) shows that the dominant overall motion of CoV2 PLpro occurs due to high flexibility of the fingers domain – especially the zinc-binding region, the BL2 loop, and the UBL domain (Figure 4.5). The fingers domain is the most mobile region of PLpro, and has been shown to crystallize in different conformations [5]. Because this region is highly flexible and challenging for a small molecular inhibitor to bind tightly, it is not considered as an ideal druggable site.

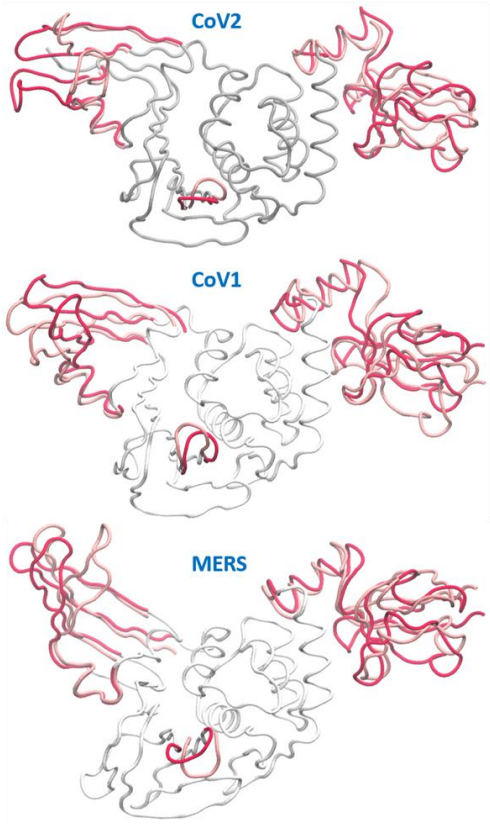


Figure 4.5. PCA shows the most flexible regions of CoV2, Cov1 and MERS PLpro highlighted in red and pink to show the two extremes of a region's range of motion (motion exaggerated 3-fold for clarity). From left to right, red/pink protein indicates the fingers domain, BL2 loop, and UBL domain.

This study focuses on the binding site of naphthalene inhibitors (Figure 4.1A), a druggable site reported in previous studies[9] that is directly adjacent to the PLpro active site to prevent off-target binding to the highly similar active site of human proteins.[32] Flexibility of the BL2 loop, which can result in an open or closed conformation, indicates potential of this binding site to accommodate compounds with new scaffolds or different derivatives of 3k, which may include larger substitutions to strengthen binding with underutilized regions. One such region is the hydrophobic portion lined by residues Met208, Pro247 and Pro248 (Figure 4.1B). Closer to the BL2 loop, Gly266 may be able to provide inhibitor binding specificity through hydrogen bond formation . The portion of the binding site extending just past the BL2 loop in the direction of the UBL domain presents substantial space to engage PLpro residues with larger ligands (Figure 4.1B).

One very prominent motion of CoV2 PLpro is partial rotation of the UBL domain and its relative position to "ridge" helix (Asp62 – His73) in the SUb2 region (Figure 4.6). The function of the UBL domain is unknown, and although some studies suggest that it has no effect on function of PLpro,[33] we observed one noteworthy interaction involving this domain. Transposition of UBL towards the thumb domain results in hydrophobic interactions between Pro59 of the UBL domain and Pro77 and Thr75; Thr75 then interacts with Phe69 of the "ridge" helix and can alter the latter residues conformation. Mutating this Phe was shown to affect the binding affinity of ISG15 and K48 linked diUb in CoV1 PLpro,[34] so the conformational dynamics of this residue may also be important in CoV2. Since CoV1 exhibits this same interaction between UBL residues and this Phe residue, we compared the conformation populations for Phe69/70 (residue numbering

differs by 1 between Cov2/CoV1) between the two PLpros. Notably, Cov1 contains Leu at position 75 (rather than Thr75 as in Cov2) and its concerted motion with Phe70 yields four different conformations. The Phe69-Thr75 interaction in CoV2 affects Phe69 to a lesser extent, resulting in just two distinct conformations of the same sidechain (Figure 4.7). In contrast to the dynamic SUb2 region, SUb1, the binding site for distal Ub, does not show any significant structural fluctuation. Ub-interacting hydrophobic residues Met208 and Pro247 are exposed to the solvent to potentially engage in ligand interactions (Figure 4.8), which may be an alternative method to disrupt Ub binding at SUb1 in aside from blocking its C-terminal from the LXGG cleavage site.

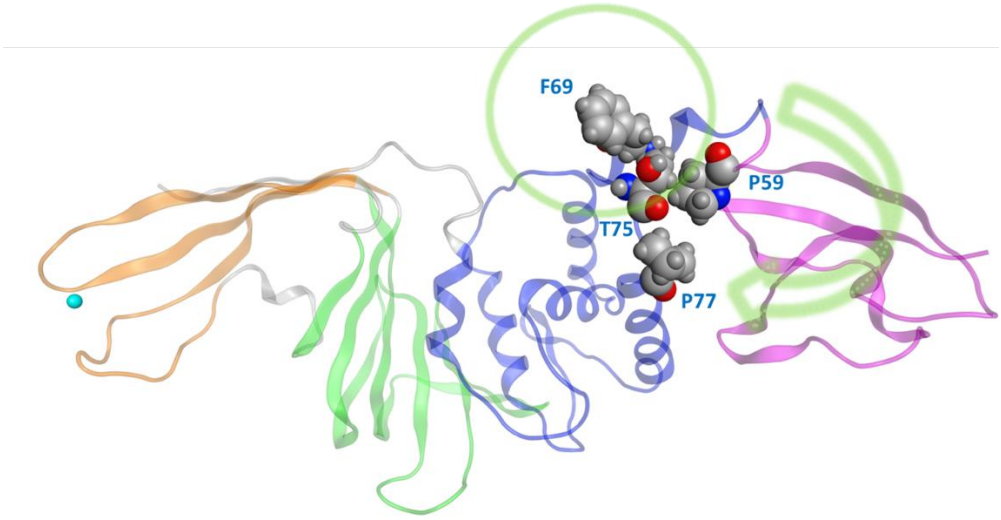


Figure 4.6. The movement of UBL in CoV2 allows for interactions between UBL residue Pro59 and the thumb domain residues Thr75 and Pro77. These interactions subsequently result in different rotameric states for the nearby key Ub-interacting residue Phe69/Phe70 in CoV2/CoV1. Green arrow indicates the major motion of UBL; green circle indicates SUb2.

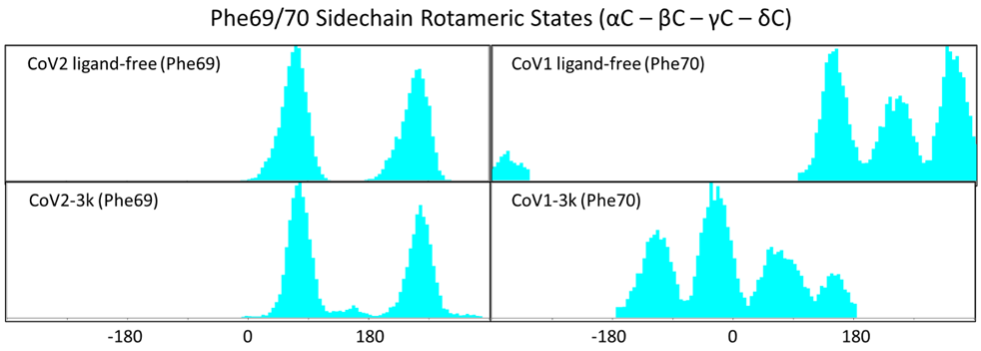


Figure 4.7. Conformation populations of the Ub-interacting residue Phe69/70 sidechain in ligand-free and 3k-bound CoV2 and CoV1 PLpro. CoV2 shows only two conformations, whereas CoV1 shows four, which may cause unique interactions between CoV2 PLpro and Ub or Ub-like proteins at SUb2. This difference is seen in both ligand-free and -bound systems.

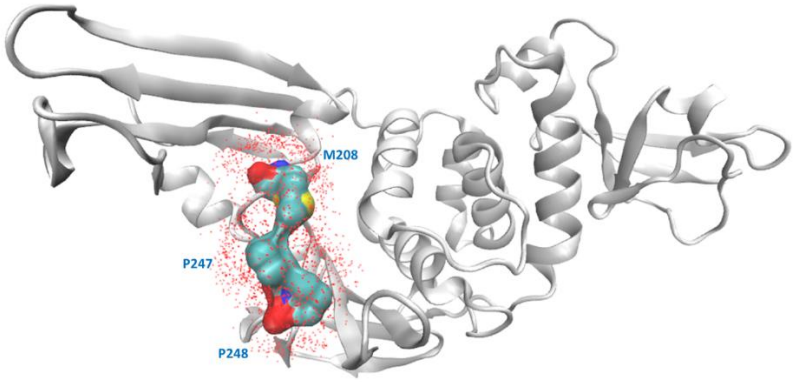


Figure 4.8. The hydrophobic region of SUb1 (multi-color surface), which contains ligand-interacting residues M208, P247 and P248 in the putative bind site, is exposed to solvent in ligand-free systems. The residues here are available to establish new

hydrophobic ligand interactions, which may be able to disrupt Ub or Ub-like protein binding at this site.

Overall, the dynamics of the CoV2, CoV1 and MERS-CoV ligand-free PLpro is quite similar. Figure 4.5 shows the first principal component of overall motion for all three systems, which reveals similarly high mobility in the zinc-binding domain, BL2 loop and UBL domain. CoV2 simulations MD2 and MD3 showed similar flexibility to CoV1 in most of the highly flexible regions during the entire course of the 1 μs MD simulations. Notably, in MD1, the initial crystal structure conformation shows a unique conformation of Asn267 and Tyr268 (Figure 4.9), resulting in larger root-mean-square fluctuation (RMSF) and dihedral entropy values than those computed for the other ligand-free PLpros (Figures 4.10), as well as additional rotameric states (Figure 4.11). Around 420 ns into MD1-1 and just 20 ns into MD1-2 the residues change conformation to ones highly similar to those in MD2 (CoV1) and MD3 (CoV2), at which point the RMSFs become nearly identical (Figure 4.12). This unique conformation of key ligand-binding residue Tyr268[35] is not preorganized for protein-ligand complex formation, thus it may incur a cost in conformational energy or entropy which can affect inhibitor binding. In terms of dihedral entropy as well, CoV1 and CoV2 are quite similar. The entropy calculations for the backbone torsion show only a few regions with higher conformational sampling in CoV2, mainly in the zinc binding region of the fingers domain and BL2 loop (Figure 4.10). In MERS-CoV, the amino acid composition of the BL2 loop is entirely different from CoV2 with the exception of two flanking Gly residues.[23] Although the entropy and RMSF show similar flexibility of the loop, its overall conformation relative to the palm

domain is more open than in CoV1 and CoV2. BL2 remains in this open conformation, which appears to be stabilized by hydrophobic interactions of Gln270, Glu273, Thr274 and His278 sidechains, for the entirety of MD4 (Figure 4.13).

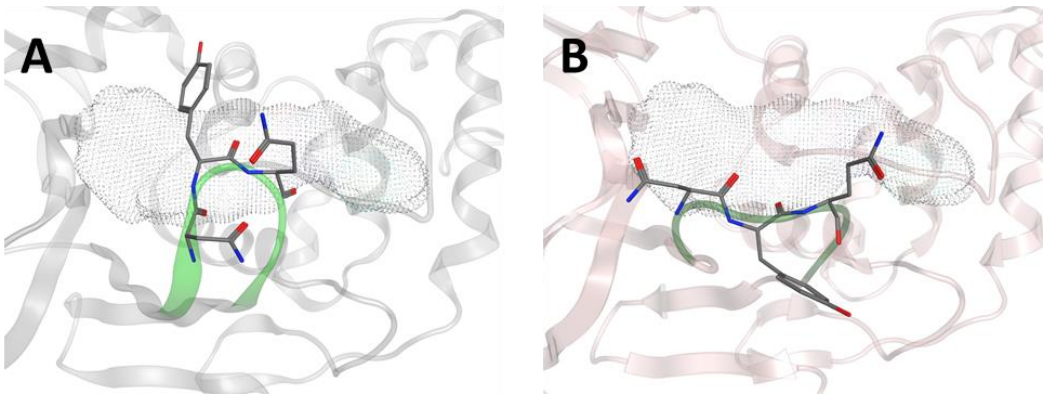


Figure 4.9. Conformation of Asn267 and Tyr268, 3k binding site is indicated with dotted surface. (A) common conformation of these residues observed in CoV2 (6WRH) and CoV1 simulations. (B) Unique conformation observed only in CoV2 (6W9C) simulation.

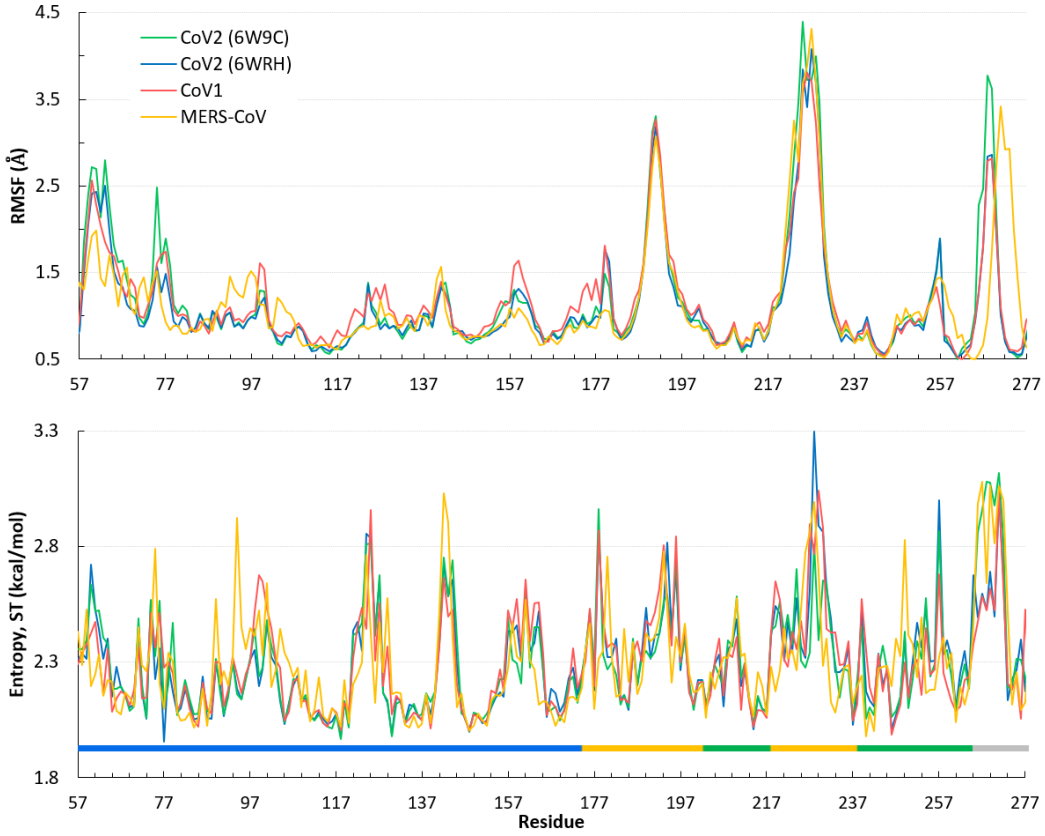


Figure 4.10. Quantifying the overall dynamics and conformational flexibility of PLpro. Top: RMSF of alpha C atoms over the 1 μs trajectory of all four ligand-free PLpros. The spike at the BL2 loop (~ residues 265-275) is larger for CoV2 (6W9C) and MERS-CoV because of their more open conformations during all or part of the simulations. Residues 1-56 (UBL domain) and 300-311 (C-terminal) have been omitted for clarity. Bottom: Dihedral entropy of the psi angle for CoV2, CoV1 and MERS-CoV systems. PLpro regions indicated by color bar: thumb – blue, fingers – yellow, palm – green, BL2 loop – grey. Entropy calculated at 298 K.

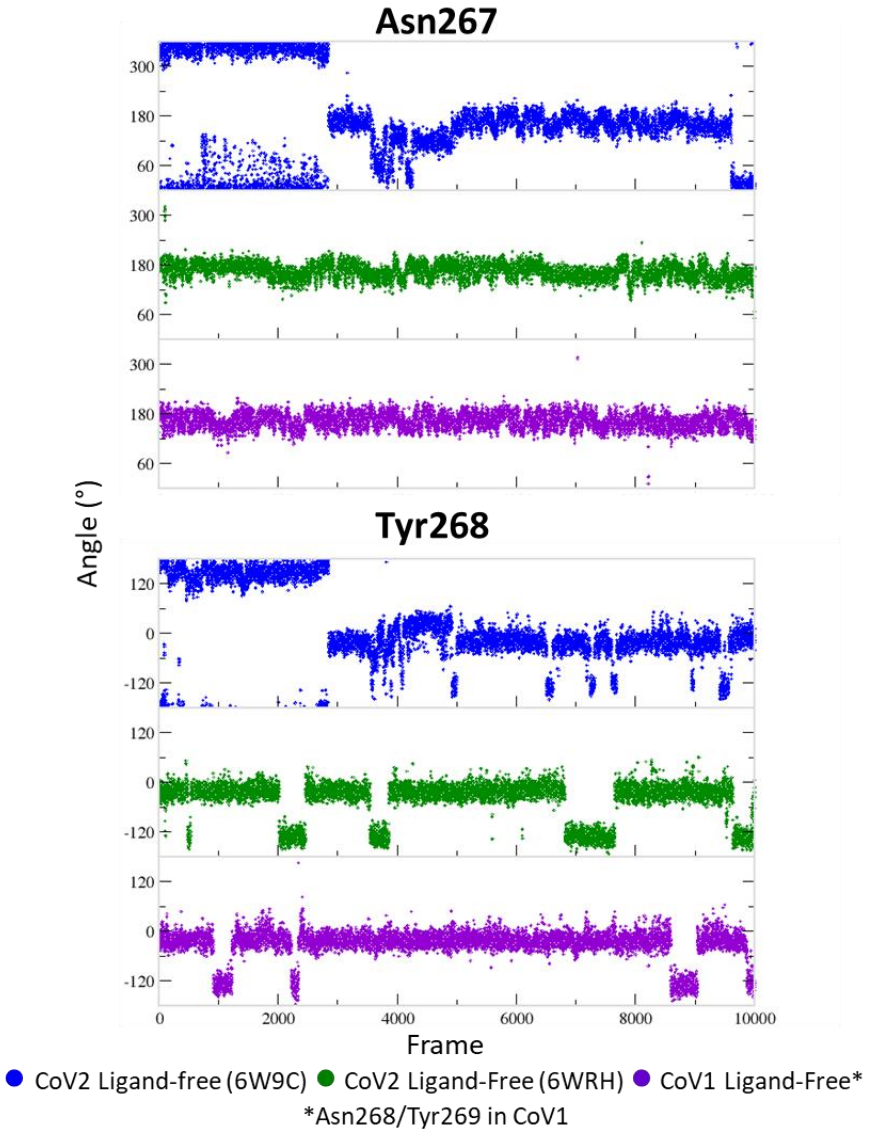


Figure 4.11. Psi angle in BL2 loop residues Tyr268 and Asn267 over simulation time for ligand-free CoV2 PLpro starting from the two different CoV2 crystal structures, 6W9C and 6WRH, and in CoV1 PLpro. The unique outward conformation of these residue in 6W9C results in rotameric states of key ligand-interacting residue Tyr268 that may have implications on ligand binding in the putative site.

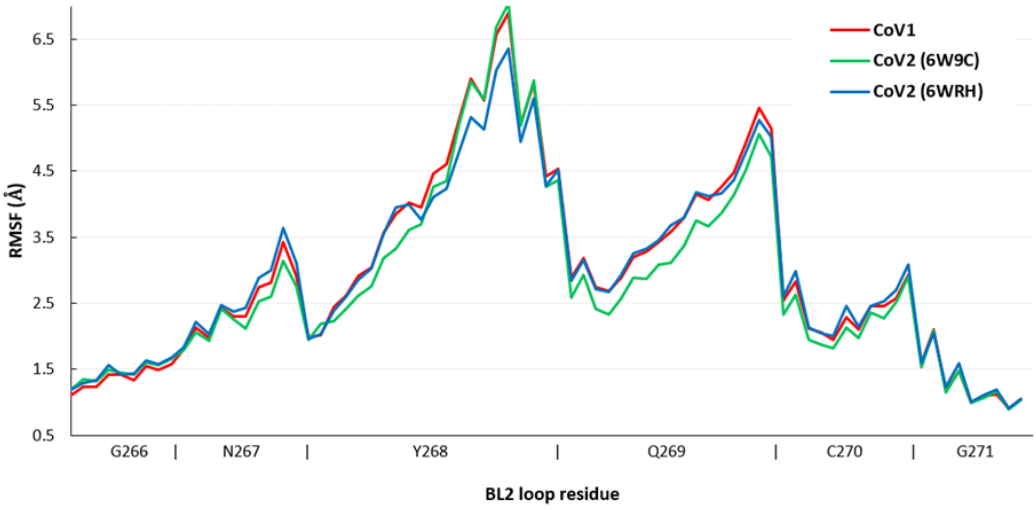


Figure 4.12. Root-mean-square fluctuation (RMSF) for all atoms in the BL2 loop during only the last 500 ns of simulation of ligand-free CoV1 PLpro and the two CoV2 PLpro simulations beginning from different crystal structures (6W9C and 6WRH). In the last 500 ns of simulation, after the unique starting conformations of Asn267 and Tyr268 in crystal structure 6W9C go away, the dynamics of the loops in all three systems become highly similar.

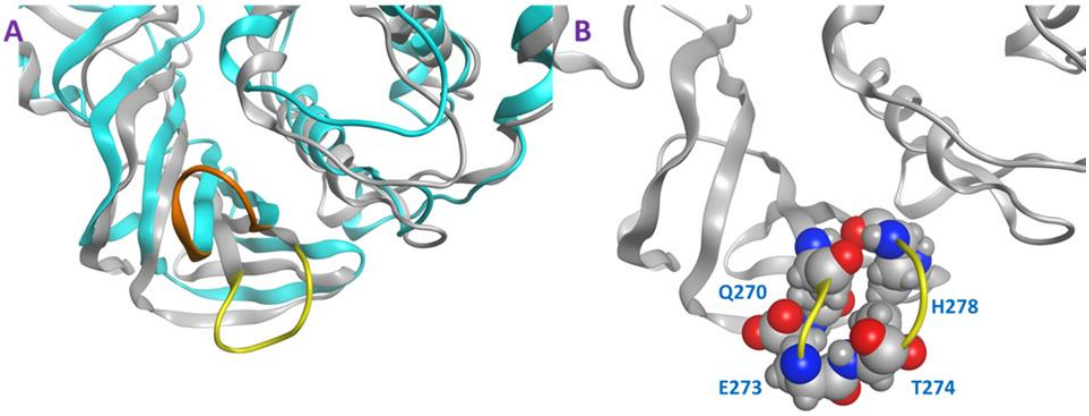


Figure 4.13. Depiction of the open conformation of the MERS-CoV BL2 loop a) BL2 loop of MERS-CoV PLpro (yellow) compared to CoV2 PLpro (orange) showing the more open conformation in MERS-CoV. b) Hydrophobic interactions between the labeled residues (shown as van der Waals spheres) hold the MERS-CoV BL2 loop in the open conformation.

4.3.2 Comparison of Ligand-free and Ligand-bound Structures CoV1 and CoV2 PLpro

Revealing detailed protein conformational changes after ligand binding provides insight to the key binding interactions relevant to drug development. We compared ligand-free and ligand-bound systems to identify how binding shifts the populated conformations of surrounding residues.

Simulations show the CoV2 PLpro BL2 loop having significant flexibility in ligand-free proteins. Residues Asn267, Gln269, and most importantly Tyr268, account for most of this motion, which resembles opening and closing of the loop (Figures 4.14A and 4.14B).

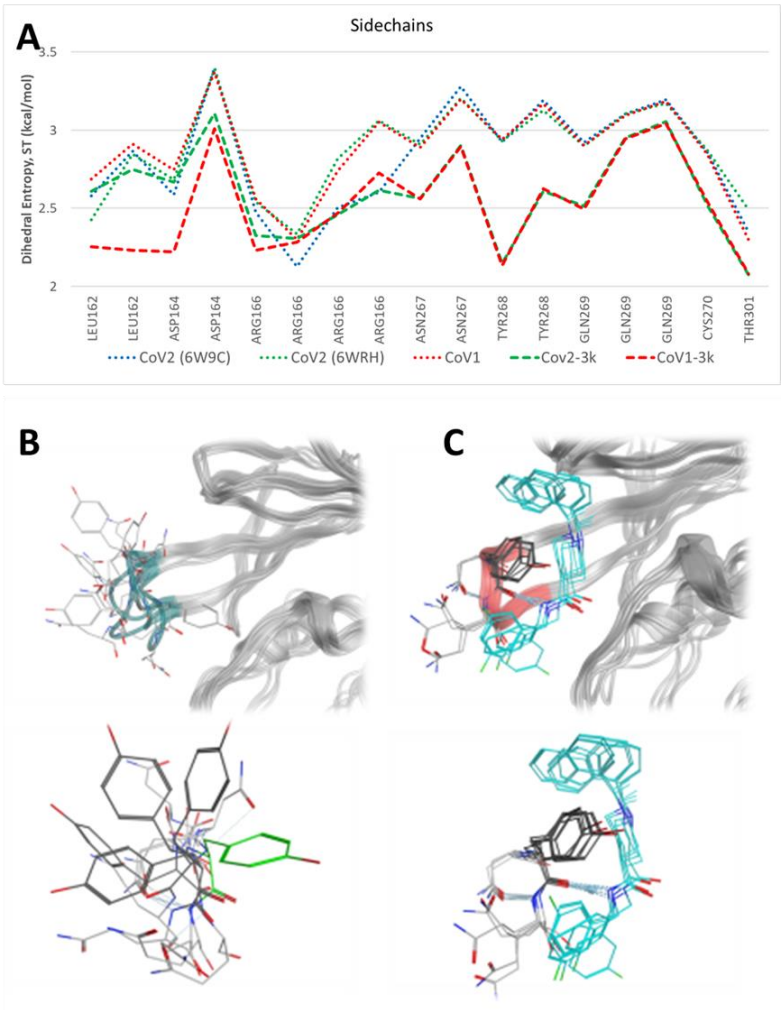


Figure 4.14. Plot of entropy for 3k binding site residues and pictures of their conformations. (A) Sidechain dihedral angle entropy for 3k binding site residues in ligand-free and 3k-bound CoV1 and CoV2 PLpro shows the stabilization of these residues after ligand binding. (B) An overlay of several MD frames shows the range of conformations adopted by BL2 loop (dark green) residues in the ligand-free state. Tyr268 in the ligand-bound conformation is shown in light green. (C) The conformational sampling of these residues is dramatically reduced upon binding of 3k (teal). Entropy calculated at 298K.

MD5a-d all show that the BL2 loop in CoV2 PLpro is highly stabilized by ligand binding, as most residues interacting with the ligand are confined to a single conformation (Figure 4.14C). Most notably, the sidechain and backbone rotation of key residue Tyr268 is minimized through a hydrogen bond and strong vdW interactions with the ligand, as detailed in next subsection. The very same ligand-induced stabilization of the BL2 loop is seen for CoV1 PLpro (Figure 4.15). The central portion of this binding pocket, which houses the piperidine, carboxyl and amide moieties of 3k, is narrower and may already be maximized in terms of inhibitor binding potential. Two key hydrogen bonds form here (Figure 4.1C), and it has been shown that substitutions larger than a methyl group or hydrogen at the benzylic-naphthyl or benzyl position, respectively, on naphthalene inhibitors lowered their effectiveness.[9]

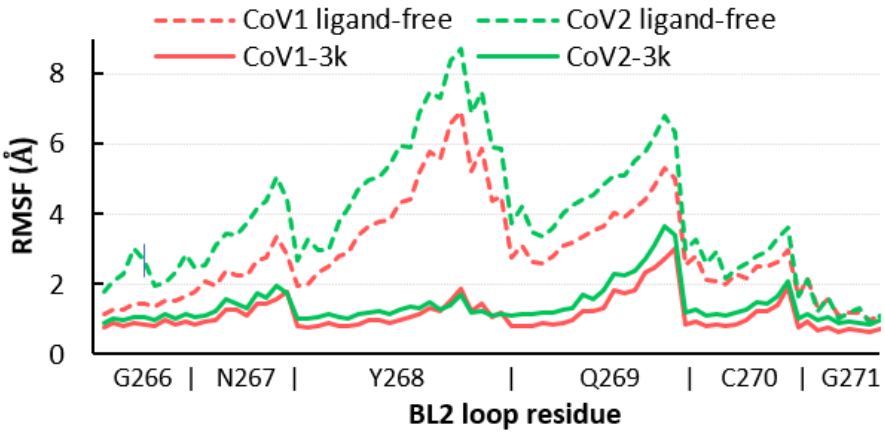


Figure 4.15. RMSF of all atoms in the BL2 loop in ligand-free and 3k-bound CoV1 and CoV2 PLpro. In both systems, ligand binding induces a closed, ordered BL2 loop conformation resulting in dramatically reduced mobility of this region. The x-axis indicates the range of atoms in each BL2 loop residue.

Residues involved in consistent interactions with the ligand show a significant difference in dihedral entropy. Hydrogen bonds formed with 3k substantially restrict conformational exchange for the associated residues. Asp164 and Tyr268 appear to be a key aspect in the 3k-CoV2 PLpro interactions, which is reflected by the decreased dihedral entropy (Figure 4.16).

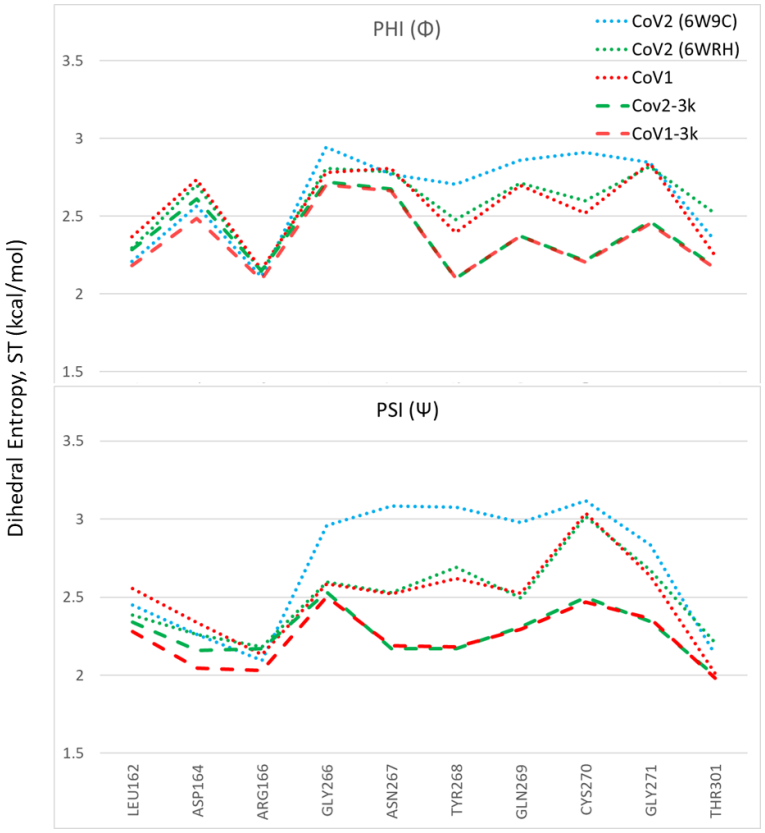


Figure 4.16. Dihedral entropy calculated at 298 K for phi and psi torsions of ligand-interacting residues in ligand-free and 3k-bound CoV1 and CoV2 PLpro.

Comparison of MD5a-d to MD6 (Table 4.1) reveals that 3k binds very similarly in CoV2 and CoV1 PLpro, inducing a closed, ordered conformation of the BL2 loop around the ligand. Moreover, the RMSD values of 3k over 200 ns in MD5a and MD6 of 1.06 and 0.95 Å, respectively, reveal similar stability in the CoV2 and CoV1 putative binding sites. The naphthalene moiety occupies the hydrophobic cleft of the pocket and the fluorophenyl ring protrudes from the opposite end of the pocket while retaining a high degree of mobility relative to the rest of the compound. The high similarity of these binding modes indicates strong potential of naphthalene inhibitors to have an inhibitory effect on CoV2 PLpro through a similar mechanism as in CoV1 PLpro.

4.3.3 Ligands Binding Modes in CoV2 PLpro and Strategies for Drug Design

Because the putative naphthalene inhibitor binding site of CoV2 PLpro is comprised by the same residues as in CoV1 PLpro, we examined one of the most effective second-generation naphthalene inhibitors of CoV1 PLpro,[9] 3k, to reveal structural information regarding binding to CoV2 PLpro for future structure-based drug development. After analyzing free and ligand-bound CoV1 PLpro simulations, we docked 3k to one CoV2 PLpro conformation to obtain four different binding poses (Figure 4.4) and ran three simulations for each pose (MD5a-d). Poses A and B were nearly the same, except B was docked with unconstrained side chain rotations allowed, so 3k starts slightly rotated with respect to A. MD5c and MD5d start with a 180° rotation of the naphthalene or piperidine moiety, respectively, compared to MD5a. Ultimately, MD5a, b and d all establish the same major interactions with PLpro. The majority of our discussion focuses on MD5a , where the initial conformation (Figure 4.4A) is the most similar to the CoV1 PLpro-3k

crystal structure. Our results indicated that 3k binds strongly and suggest that the ligand can inhibit the enzymatic function of CoV2 PLpro.

In the hydrophobic portion of the binding site, the naphthalene moiety sits stably between residues Pro247 and Pro248 to one side and Tyr268 on the other (Figure 4.17) However, additional space exists in this pocket to engage more residues. Specifically, it may be possible to increase the hydrophobic interactions here with a methyl (or larger) substitution on the naphthalene to further engage in vdW interactions with Pro248 or Tyr264 (Figure 4.17, blue dots). Pro248 and Met208 can also be further engaged in hydrophobic interactions with substitutions at the appropriate positions on naphthalene (Figure 4.17, yellow dots), or potentially even a substitution of the entire naphthalene moiety for a larger aromatic structure such as anthracene or phenanthrene (Figure 4.17). MD5c (Figure 4.4C) , in which the naphthalene in the initial 3k conformation is flipped 180° relative to MD5a, provides support for this idea, as the flipped moiety is seen making closer contact with residues Met208 and Pro248, resulting in greater attraction to these residues (Figure 4.18) and slightly lower overall binding energy (Table 4.2) than in MD5a. Lastly, a hydrogen bond donor or acceptor substitution at the correct naphthalene position (Figure 4.17, blue dots) may be able to engage with the Gly266 backbone.

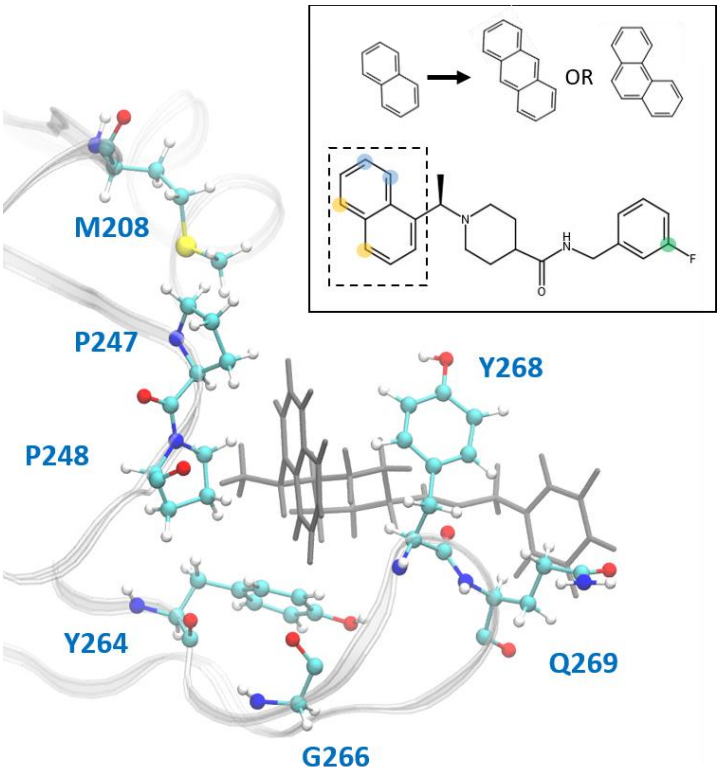


Figure 4.17. Ligand 3k (grey) in the CoV2 PLpro binding site. Residues with which new interactions are achievable or current ones can be strengthened are labeled and shown in ball-and-stick representation. Top-right: 2D molecular structure of ligand 3k indicating proposed substitution positions for increased binding affinity. Substitutions at the yellow positions may be capable of additional hydrophobic contacts with Pro247, Pro248, or Met208. Substitutions at blue positions may be capable of additional hydrophobic contacts with Pro248 or Tyr264, or hydrogen bonds with the backbone carboxyl of Gly266. Finally, substitutions at the green position in combination with an extended benzene linkage may be capable of increased attractive interactions with Gln269 or other nearby residues. The naphthalene moiety is indicated by the dashed box, with the proposed anthracene or phenanthrene substitutions indicated above.

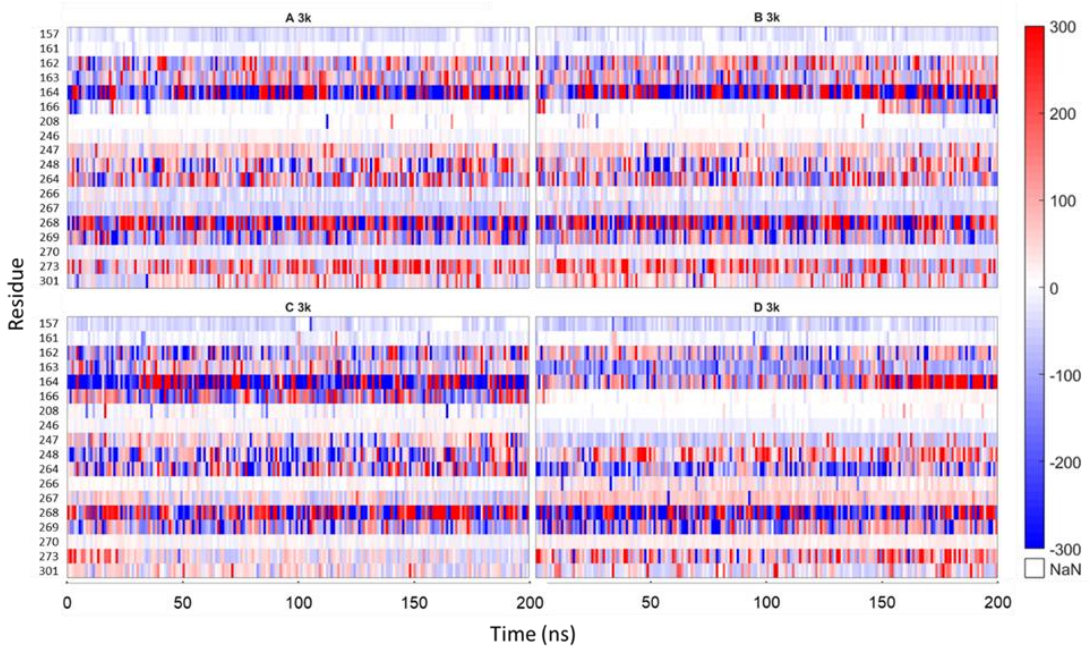


Figure 4.18. Residue-wise protein-ligand interaction forces, where blue indicates attractive forces and red indicates repulsion, from one 200 ns CoV2 PLpro-3k complex simulation beginning from each of the four binding poses, A-D. The interactions with residues Tyr268 and Asp164 are dominant for all four. Pose D, despite starting with its hydrogen bond donor pointed away from Asp164, establishes this important interaction in the latter half of the simulation. Force values are in piconewtons.

Table 4.2. MM/PBSA energy breakdowns for the binding energy from simulations of the four different starting poses (A-D) of CoV2 PLpro-3k and CoV1 PLpro-3k. $\Delta E_{elec+PB}$ is the electrostatic plus polar solvation energy contributions, and ΔE_{vdW+np} is the van der Waals plus non-polar solvation energy contribution. $\Delta E_{MM/PBSA}$ is the binding energy predicted by MM/PBSA. Energies are in kcal/mol; values are ± SD.

MM/PBSA Binding Energies					
	CoV1-3k	**CoV2-3k A**	**CoV2-3k B**	**CoV2-3k C**	**CoV2-3k D**
$\mathbf{\Delta E_{elec+PB}}$	11.9	11.8	11.6	11.0	14.7
$\mathbf{\Delta E_{vdW+np}}$	-28.9	-26.2	-28.1	-28.7	-27.0
$\mathbf{\Delta E_{MM/PBSA}}$	-17.0 ± 3.9	-14.5 ± 4.3	-16.5 ± 4.5	-17.7 ± 3.6	-12.3 ± 5.4

3k engages in two strong hydrogen bonds with the protein: one to Asp164 and the other to the backbone of Tyr268. Notably, even in MD5d, which began with the piperidine nitrogen and its hydrogen pointed in the opposite direction of Asp164, the entire moiety rotates after 25-140 ns (varying between the three runs) to establish the hydrogen bond with this residue. Previous studies found that bulky ligand substitutions that occupy this portion of the pocket decreased inhibition.[1] A possible explanation is that the specific ligand orientation needed to maintain both of these strong hydrogen bonds was not attainable due to the additional bulk. Moreover, we observed a consistent intra-protein hydrogen bond between Tyr264 and Asn267 in ligand-bound CoV2 that could be disrupted by larger ligand substitutions here, which may destabilize the closed BL2 loop. Indeed, analysis of this interaction shows very high correlation between formation of the hydrogen bond and a closed loop conformation (Figure 4.19). A small hydrophobic

pocket formed by Tyr264, Tyr273 and Thr301 accommodates the methyl group at the benzylic-naphthyl position.

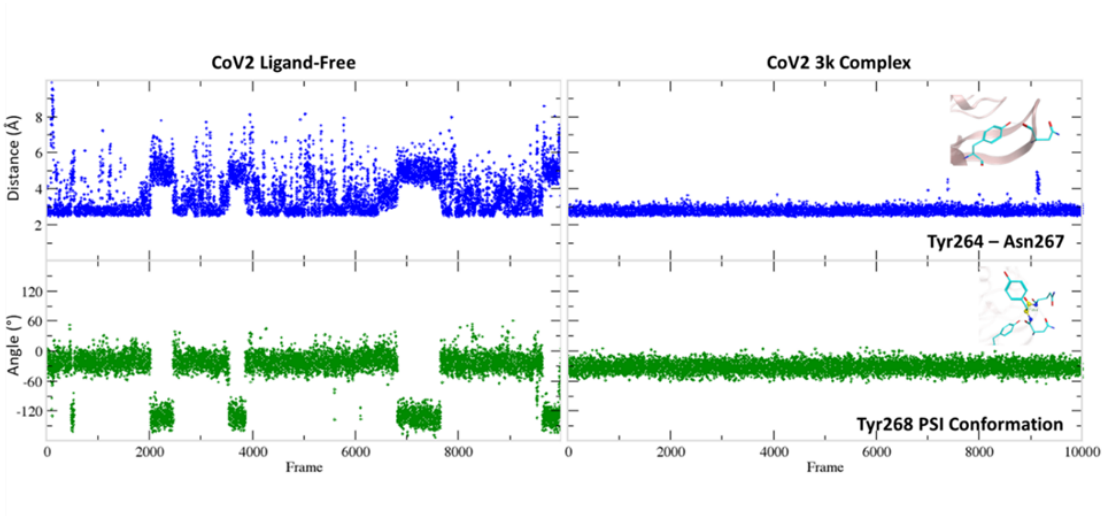


Figure 4.19. Tyr264-Asn267 hydrogen bond length and Tyr268 psi angle over time. In ligand-free CoV2 PLpro (left), a short hydrogen bond length between Tyr264 and Asn267 (top, blue) has a strong correlation with a closed BL2 loop conformation (bottom, green). After ligand binding (right), this distance is consistently short and the loop conformation remains closed.

The fluorophenyl ring of 3k appears to interact mostly with the hydrocarbon portion of Gln269, but also engages in vdW interactions with Tyr268 (Figure 4.17). However, because of the openness of PLpro in this region, the position of the ring fluctuates widely, and it rotates freely with the fluorine observed at several positions consistent with 360-degree rotation. Increasing ligand engagement with PLpro residues is achievable in this region, although previous attempts at doing so in CoV1 PLpro had mixed results. Substitutions on the benzyl ring in first generation naphthalene inhibitors found that

anything bulkier than methyl at the ortho position decreased inhibition;[9] however, the linkage between the amide and benzene ring was one carbon shorter than in the second generation, possibly causing the added bulk to disrupt one of the two important hydrogen bonds with Tyr268 or Asp164. With a longer linkage to the benzene in second-generation naphthalene inhibitors, various benzene substitutions were tested, but showed no clear trend in effectiveness. Ultimately, the fluorine substitution at the meta position, as seen in 3k, showed the best result. Extending the linkage between the amide and benzene ring by one additional carbon was found to weaken inhibition, providing evidence that benzene ring primarily contributes to binding through vdW interactions with Tyr268 and Gln269, and so needs to be close to those residues. This is consistent with our observations and residue-wise force calculations as well (Figure 4.18).

One method to increase binding affinity in this region may be through increasing the hydrophobic surface area of the benzyl end of the ligand. This can be achieved either through substitution of methyl or larger hydrocarbon groups onto the benzene ring, or by replacing the benzene with a bulkier group, such as naphthalene. Although, as previously stated, it has been shown that both increasing ligand bulkiness near the benzene end and extending the linkage to benzene can sometimes lead to decreased inhibition, changing these two factors simultaneously has not been tested. A longer linkage may accommodate increased bulk, while the added hydrophobic mass can still reach residues Tyr268 and Gln269 for attractive interactions. Moreover, since no clear trend in ligand effectiveness from substitutions on the benzyl ring has been found, we suggest exploration of the available space in this portion of the binding site.

To validate some of our proposed modifications to the current naphthalene-based scaffold, we docked these modified ligands to the same CoV2 PLpro conformation used to dock 3k. First, to investigate the potential for making additional hydrophobic contacts in the cleft near SUb1, we substituted anthracene or phenanthrene to the naphthalene position. Results for both substitutions show more favorable docking scores than for 3k, with anthracene showing slightly better performance than phenanthrene (Figure 4.20). The favorable contacts arise from interaction with Asp166, a residue that rotates freely in the CoV2-3k simulations, indicating that it may be available to form a stable interaction with ligands capable of reaching it. Additionally, both the anthracene and phenanthrene-substituted ligands maintained all the other essential interactions we identified for 3k. In an attempt to increase polar interactions, we added a hydroxyl to the naphthalene moiety to form a hydrogen bond with Gly266 or other hydrogen bond acceptors in the area. We found that the hydrogen bond with Gly266 does indeed form as expected, with a distance of just 1.74 Å. However, in this conformation, the important hydrogen bonding group on the ligand that usually interacts with Asp164 is slightly out of position (Figure 4.20). Despite this, the binding is still more favorable than that of 3k. Also, notably, our MD simulations of CoV2-3k that started without the 3k-Asp164 hydrogen bond quickly formed that hydrogen bond after the simulation began, providing evidence that the same will likely happen in the case of the new ligand. Taken altogether, the MD and docking results show that the 3k scaffold should be capable of exploiting bulkier hydrophobic groups or polar groups at the naphthalene end to establish both favorable hydrophobic and polar contacts while maintaining the essential residue interactions that made 3k

successful in CoV1. These are but a few of many possible enhancements to the naphthalene-based scaffold, and they require further validation through MD simulation or experimental assays. However, this serves as a proof-of-concept for future CoV2 PLpro design directions.

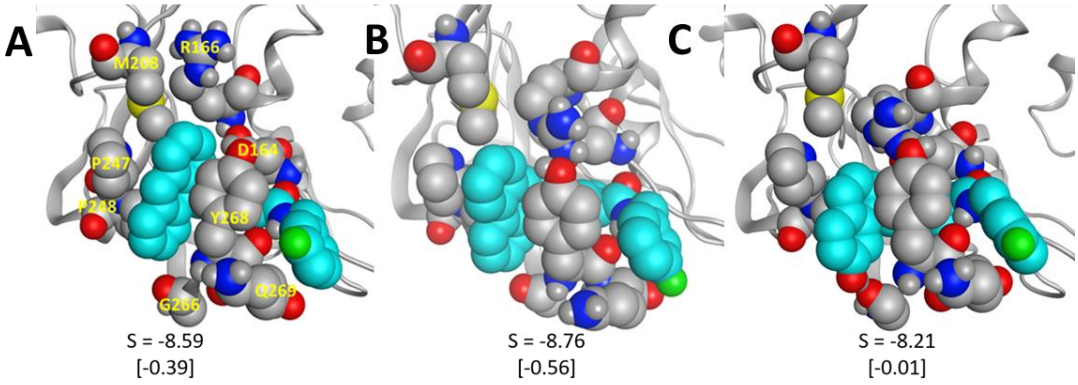


Figure 4.20. The modified ligands (teal) docked to CoV2 PLpro. (A) 3k with anthracene substituted for naphthalene, (B) 3k with phenanthrene substituted for naphthalene, and (C) 3k with a hydroxyl substituted on the napthyl moiety; the hydrogen bond with Gly266 is clearly visible. S is the docking score given by MOE, and the difference between the pictured docked conformation and the best 3k score is shown in brackets.

Pair-wise force distribution analysis (Figure 4.18) and interaction energies (Table 4.3) indicate that the binding mode of 3k in CoV2 is dependent on both strong vdW interactions and hydrogen bonds (Figure 4.21), highly similar to that of 3k in CoV1 (Table 4.2). The interaction with Tyr268 is a dominant one in all ligand-bound simulations. The residue engages in a hydrogen bond donated by the amide nitrogen of 3k and a T-shaped pi-stacking interaction with the naphthalene moiety (Figure 4.17),

which is seen with all naphthalene-based inhibitors [5]. The hydrogen bond between 3k and Asp164 (Figures 4.1C and 4.21) is another strong protein-ligand interaction shared in CoV1 and CoV2. All major ligand interactions with binding site residues are shown in Figure 4.18 and listed in Table 4.3.

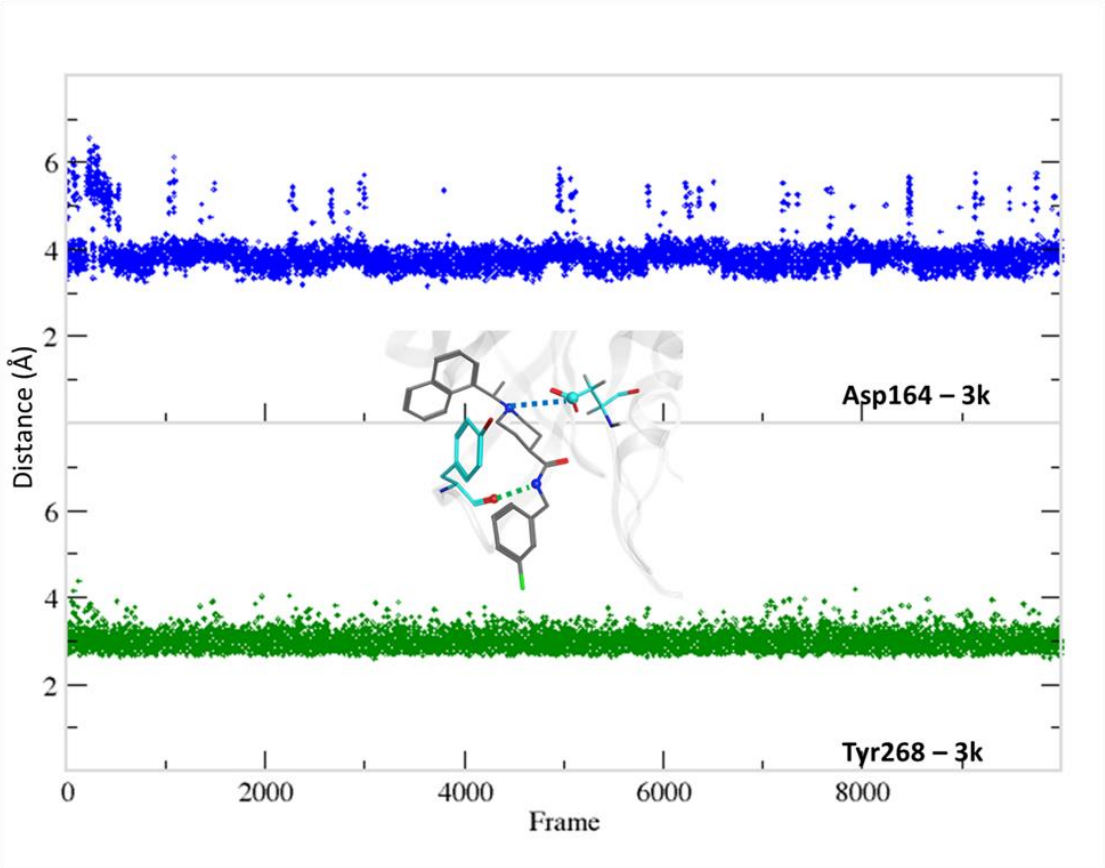


Figure 4.21. Distance of the two key ligand-protein hydrogen bonds between 3k and Asp164 or Tyr268 over 200 ns of simulation. The minimal fluctuation in distance indicates the stability of the bonds and the very limited space in this part of the binding pocket.

Table 4.3. Total, electrostatic (elec), and van der Waals (vdW) residue-wise protein-ligand interaction energies from 3k-bound CoV2 and CoV1 PLpro simulations, in kcal/mol. Energies were calculated every 200 ps and then averaged over the 200 ns trajectory. Values are ± SD.

	CoV2 - pose A			CoV2 - pose B			CoV2 - pose C		
Res	**Total**	**vdW**	**elec**	**Total**	**vdW**	**elec**	**Total**	**vdW**	**elec**
164	-77.7 ± 16.1	-2.4	-75.4	-74.6 ± 9.8	-3.03	-71.6	-83.5 ± 5.8	-2.5	-81.0
268	-21.5 ± 2.1	-9.6	-11.9	-21.6 ± 2.0	-9.5	-12.1	-21.1 ± 2.0	-9.6	-11.6
167	-16.6 ± 12.5	-0.2	-16.4	-18.3 ± 10.6	-0.2	-18.0	-16.5 ± 12.1	-0.2	-16.3
267	-7.4 ± 7.2	-1.5	-5.9	-7.9 ± 8.1	-1.3	-6.5	-5.8 ± 9.4	-1.3	-4.5
248	-7.3 ± 3.2	-3.8	-3.5	-7.5 ± 3.0	-3.9	-3.5	-7.4 ± 3.0	-3.4	-3.9
163	-5.5 ± 3.4	-1.9	-3.5	-4.5 ± 3.1	-1.8	-2.7	-3.2 ± 3.16	-2.0	-1.2
269	-3.9 ± 10.7	-4.4	0.6	-2.7 ± 10.4	-4.5	1.9	-5.5 ± 11.2	-4.2	-1.2
249	-3.0 ± 2.7	-0.1	-2.9	-2.7 ± 3.2	-0.1	-2.6	-3.3 ± 2.8	-0.1	-3.3
274	3.7 ± 8.5	-0.0	3.8	7.5 ± 7.7	-0.0	7.5	7.2 ± 7.1	-0.0	7.2
165	5.3 ± 9.3	-0.22	6.0	2.0 ± 8.8	-0.3	2.4	3.14 ± 8.2	-0.2	3.3

	CoV2 - pose D			**CoV1**		
Res	**Total**	**vdW**	**elec**	**Total**	**vdW**	**elec**
164	-56.3 ± 14.7	-3.3	-53.0	-74.3 6.8	-3.2	-71.1
268	-23.7 ± 2.8	-10.2	-13.5	-20.7 ±2.3	-9.4	-11.3
167	-15.5 ± 12.2	-0.2	-15.2	-19.1 ± 9.8	-0.3	-18.8
267	-5.5 ± 8.0	-1.6	-3.9	-7.8 ± 7.3	-1.4	-6.3
248	-7.2 ± 2.6	-4.0	-3.2	-7.4 ± 2.9	-4.0	-3.5
163	-6.8 ± 4.0	-2.0	-4.9	-6.2 ± 2.7	-1.9	-4.4
269	-4.9 ± 7.5	-4.6	-0.3	-3.1 ± 2.8	-4.6	1.5
249	-2.5 ± 2.9	-0.1	-2.4	-3.1 ± 10.1	-0.1	-3.0
274	8.0 ± 9.2	-0.0	8.0	4.2 ± 6.7	-0.0	4.2
165	3.8 ± 7.5	-0.3	4.2	8.3 ± 6.2	-0.4	8.7

In addition to 3k, 6-mercaptopurine (6MP) from the thiopurine class of inhibitors has been reported to reversibly inhibit CoV1[10] and MERS-CoV[11] PLpro activity. To assess its potential for inhibiting CoV2 PLpro, we docked 6MP in the active site of the enzyme and the putative ligand binding site (Figure 4.1A) based on the proposed binding poses from existing studies. We ran three independent simulations of the two complexes for 200 ns each (MD7a and MD7b). The compound dissociated from the putative ligand binding site

within 80 ns or less and no stable intermolecular interactions were established. The compound stayed within the active site during two of the three MD7b simulations; however, it remained highly mobile and unstable in the pocket (Figure 4.22). Because the ligand was unstable in both binding sites, we did not compute interaction energies between 6MP and Cov2 PLpro. Our analysis suggests that ligand 6MP is a weak binder and may be a poor inhibitor of Cov2 PLpro.

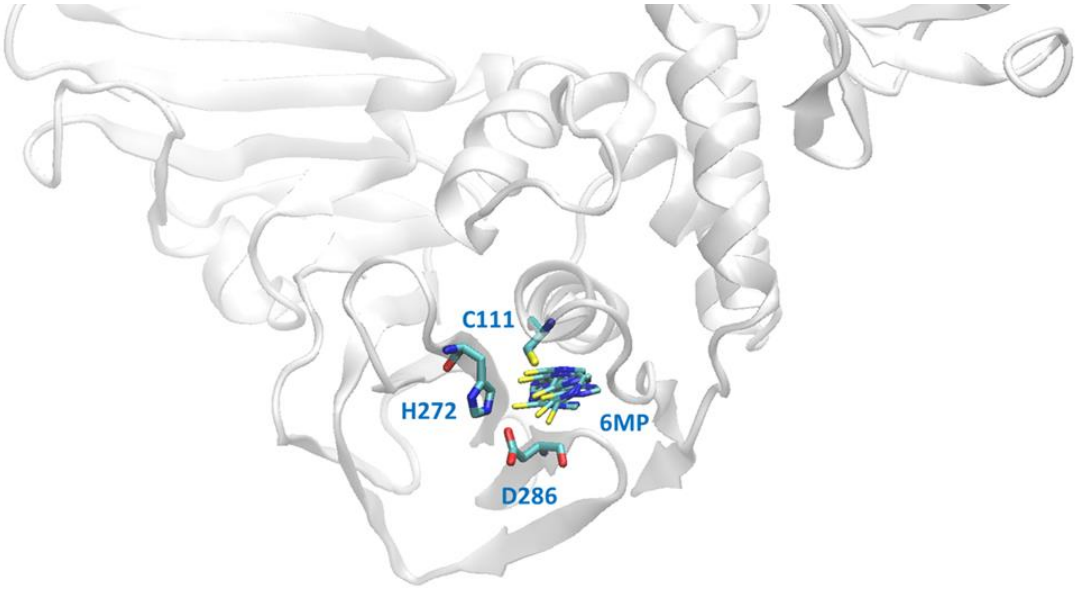


Figure 4.22. Several conformations, shown as superimposed MD simulation frames, were observed for ligand 6MP in the active site of CoV2 PLpro, indicating it is a weak binder and does not appear to be a promising inhibitor.

4.4 Conclusions

By analyzing the dynamics of ligand-free and ligand-bound CoV2 PLpro, we have gained insight to the important dynamics and intramolecular interactions relevant to its function

and the development of small molecule inhibitory drugs. The BL2 loop, zinc binding region, and UBL domain are the most mobile protein regions, and CoV2 PLpro overall dynamics are extremely similar to those of CoV1 PLpro. SUb1 contains hydrophobic residues that contact the ligand in the 3k binding site, while SUb2 is adjacent to the highly mobile UBL domain and is affected by contacts to its Ub-interacting residues brought about by UBL domain rotation.

We docked two ligands, 3k and 6MP, known to inhibit CoV1 and MERS-CoV PLpro, respectively, into CoV2 PLpro to assess their ability as CoV2 inhibitors and identify opportunities for further optimization of the ligand scaffolds. We found that not only can 3k bind strongly to CoV2 PLpro, but that there is likely room for further optimization of binding affinity by exploitation of space in the small hydrophobic cleft near Pro247, Pro248 and Tyr264, or by making additional residue contacts in the open pocket region at the opposite end of the binding site. 6MP was unable to bind stably in the 3k site and dissociated quickly in all three simulations. It associated for longer to the active site; however, even when it remained bound, the compound was unstable.

Our results show that naphthalene-based inhibitors or similar compounds should have an inhibitory effect on CoV2 PLpro, and we have provided detailed suggestions for how this ligand scaffold can be furthered improved by engaging residues in underutilized space of the binding site. This study also generates an ensemble of CoV2 PLpro conformations that illustrate potential inhibitor-protein interactions for structure-based inhibitor design and elucidates protein dynamics relevant to Ub or Ub-like protein binding.

4.5 References

1. Provisional Death Counts for Coronavirus Disease (COVID-19). 2020.

2. Harcourt BH, Jukneliene D, Kanjanahaluethai A, Bechill J, Severson KM, Smith CM, Rota PA, Baker SC. Identification of Severe Acute Respiratory Syndrome Coronavirus Replicase Products and Characterization of Papain-Like Protease Activity. 2004.

3. Wertz IE, Murray JM. Structurally-defined deubiquitinase inhibitors provide opportunities to investigate disease mechanisms. Drug Discovery Today: Technologies 2019;31:109-123.

4. Ye Y, Scheel H, Hofmann K, Komander D. Dissection of USP catalytic domains reveals five common insertion points. 2009.

5. Baez-Santos YM, St John SE, Mesecar AD. The SARS-coronavirus papain-like protease: Structure, function and inhibition by designed antiviral compounds. Antiviral Research 2015;115:21-38.

6. Sulea T, Lindner HA, Purisima EO, Ménard R. Deubiquitination, a New Function of the Severe Acute Respiratory Syndrome Coronavirus Papain-Like Protease? J Virol. Volume 792005. p 4550-1.

7. Bekes M, van Noort GJV, Ekkebus R, Ovaa H, Huang TT, Lima CD. Recognition of Lys48-Linked Di-ubiquitin and Deubiquitinating Activities of the SARS Coronavirus Papain-like Protease. Molecular Cell 2016;62(4):572-585.

8. Devaraj SG, Wang N, Chen Z, Chen Z, Tseng M, Barretto N, Lin R, Peters CJ, Tseng C-TK, Baker SC and others. Regulation of IRF-3-dependent Innate Immunity by the Papain-like Protease Domain of the Severe Acute Respiratory Syndrome Coronavirus. 2007.

9. Baez-Santos YM, Barraza SJ, Wilson MW, Agius MP, Mielech AM, Davis NM, Baker SC, Larsen SD, Mesecar AD. X-ray Structural and Biological Evaluation of a Series of Potent and Highly Selective Inhibitors of Human Coronavirus Papain-like Proteases. Journal of Medicinal Chemistry 2014;57(6):2393-2412.

10. Chou CY, Chien CH, Han YS, Prebanda MT, Hsieh HP, Turk B, Chang GG, Chen X. Thiopurine analogues inhibit papain-like protease of severe acute respiratory syndrome coronavirus. Biochemical Pharmacology 2008;75(8):1601-1609.

11. Cheng KW, Cheng SC, Chen WY, Lin MH, Chuang SJ, Cheng IH, Sun CY, Chou CY. Thiopurine analogs and mycophenolic acid synergistically inhibit the papain-like

protease of Middle East respiratory syndrome coronavirus. Antiviral Research 2015;115:9-16.

12. D.A. Case IYB-S, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman. Amber 2018: University of California, San Francisco.; 2018.

13. Götz AW, Williamson MJ, Xu D, Poole D, Grand SL, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. 2012.

14. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. 2015.

15. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. Journal of computational chemistry 2004;25(9):1157-1174.

16. Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. Journal of Computational Chemistry 2002;23(16):1623-1641.

17. Chen J, Brooks CL, Khandogin J. Recent advances in implicit solvent-based methods for biomolecular simulations. Current Opinion in Structural Biology 2008;18(2):140-148.

18. Sagui C, Pedersen LG, Darden TA. Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations. The Journal of Chemical Physics 2003.

19. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational Physics 1977;23(3):327-341.

20. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. Journal of Chemical Theory and Computation 2013;9(7):3084-3095.

21. The crystal structure of papain-like protease of SARS CoV-2 FAU - Osipiuk, J., Jedrzejczak, R., Tesar, C., Endres, M., Stols, L., Babnigg, G., Kim, Y., Michalska, K., Joachimiak, A., Center for Structural Genomics of Infectious Diseases (CSGID) CRDT - 2020/03/22 12:00 AID - 10.2210/pdb6w9c/pdb [doi].

22. The crystal structure of Papain-Like Protease of SARS CoV-2 , C111S mutant FAU - Osipiuk, J., Tesar, C., Jedrzejczak, R., Endres, M., Welk, L., Babnigg, G., Kim, Y., Michalska, K., Joachimiak, A., Center for Structural Genomics of Infectious Diseases (CSGID) CRDT - 2020/04/29 12:00 AID - 10.2210/pdb6wrh/pdb [doi].

23. Lee H, Lei H, Santarsiero BD, Gatuz JL, Cao SY, Rice AJ, Patel K, Szypulinski MZ, Ojeda I, Ghosh AK and others. Inhibitor Recognition Specificity of MERS-CoV Papain-like Protease May Differ from That of SARS-CoV. Acs Chemical Biology 2015;10(6):1456-1465.

24. Stacklies W, Seifert C, Graeter F. Implementation of force distribution analysis for molecular dynamics simulations. BMC Bioinformatics 2011;12(1):101.

25. *Molecular Operating Environment (MOE)*. 2018.01. 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7: Chemical Computing Group ULC; 2018.

26. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. Journal of Molecular Graphics 1996;14(1):33-38.

27. Ai R, Qaiser Fatmi M, Chang C-eA. T-Analyst: a program for efficient analysis of protein conformational changes by torsion angles. Journal of Computer - Aided Molecular Design 2010;24(10):819.

28. Analysis of a complex of statistical variables into principal components. Volume 24. US: Warwick & York; 1933. p 417-441.

29. Hotelling H. Relations Between Two Sets of Variates. In: Kotz S, Johnson NL, editors. Breakthroughs in Statistics: Methodology and Distribution. New York, NY: Springer New York; 1992. p 162-190.

30. Wang C, Greene DA, Xiao L, Qi R, Luo R. Recent Developments and Applications of the MMPBSA Method. Frontiers in Molecular Biosciences 2018;4:87.

31. Weiser J, Shenkin PS, Still WC. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). Journal of Computational Chemistry 1999;20(2):217-230.

32. Kemp M. Chapter Three - Recent Advances in the Discovery of Deubiquitinating Enzyme Inhibitors. In: Lawton G, Witty DR, editors. Progress in Medicinal Chemistry. Volume 55: Elsevier; 2016. p 149-192.

33. Clasman JR, Báez-Santos YM, Mettelman RC, O'Brien A, Baker SC, Mesecar AD. X-ray Structure and Enzymatic Activity Profile of a Core Papain-like Protease of MERS Coronavirus with utility for structure-based drug design. Scientific Reports 2017;7(1):1-13.

34. Ratia K, Kilianski A, Baez-Santos YM, Baker SC, Mesecar A. Structural Basis for the Ubiquitin-Linkage Specificity and deISGylating Activity of SARS-CoV Papain-Like Protease. Plos Pathogens 2014;10(5):15.

35. Chaudhuri R, Tang S, Zhao G, Lu H, Case DA, Johnson ME. Comparison of SARS and NL63 papain-like protease binding sites and binding site dynamics: inhibitor design implications. Journal of molecular biology 2011;414(2):272-288.

Chapter 5 Molecular Dynamics Studies of the Conformational Changes and Protein Dynamics in Alpha Subunit of Tryptophan Synthase and Implications for Enzyme Engineering

5.1 Introduction

The use of enzymes has increasingly become a better option for many chemical processes in various industries such as food and beverage, textile and paper technology, biofuels, and agriculture. These biotechnological methods reduce significantly manufacturing cost and environmental pollution. Enzymes have also transformed the medical research approach for various conditions and led to the development of a new class of drugs that allow for highly specific treatment, prevention, and control of many injuries and diseases. Widespread use of enzymes in biotechnology has prompted the need of increasing enzymatic efficiency. More often than not, a wild-type enzyme discovered in nature is not suitable for an industrial process. There is a need to engineer and optimize enzyme performance in terms of its properties such as activity, selectivity , stability, and others in order for the enzymatic process to make sense commercially. Generally, this process involves three steps, selecting the target amino acids, accomplishing the selected modifications, and finally evaluating for improved enzyme properties. [1-7]

Two methods are commonly employed in selecting the mutations, each has its limitations, however they are not mutually exclusive. Directed evolution involves repeated cycles of random mutagenesis and/or gene recombination to generate molecular diversity, followed by high throughput screening and selection for positive mutants.[8,9]

This method mimics natural evolution and normally produces better results to rational design. The approach can be problematic as even protein libraries with millions of members still sample only a tiny fraction of the possible sequence space for an average protein. Instead of looking into bigger libraries and more screening to address this issue, many researchers are working on new methods to create smaller and higher quality libraries. Rational design is a more oriented knowledge-based library design, and relies extensively on information on protein sequence, structure, and function, as well as computational predictive algorithms to preselect promising target sites which translates into dramatically reduced library sizes.[10]

Motivated by the growing popularity and need of enhanced proteins, our study aims to reveal how substrate binding processes and residue conformation exchange affect enzymatic efficiency using α subunit of *S.enterica* tryptophan synthase (TS) and its homolog BX1 in maize as model systems. TS is well understood functionally and structurally. Crystal structures of the complex bound to native substrate and number of intermediates and inhibitors are readily available. TS catalyzes the last two steps of tryptophan biosynthesis. It is e heterodimer but exists as a tetramer, and the two dimers connect linearly with two β subunits attached to each other. The enzyme α subunit catalyzes the first reaction, the cleavage of indole-3-glycerol phosphate (IGP), the indole product is transferred through a 25 Å hydrophobic channel to β site where it is condensed with serine in a pyridoxal phosphate (PLP) dependent reaction to form tryptophan, and the glyceraldehyde-3-phosphate (G3P) is released as a by-product from α site.[11] A wide range of allosteric interactions between α and β subunits of TS regulate the overall

catalytic activity of the enzyme as to prevent the diffusion of the indole intermediate into the cell environment.[12] A structural and catalytic homolog of α subunit is found in maize. Bx1 gene in maize encodes for indole-3-glycerol-phosphate lyase (BX1), which also cleaves IGP to produce indole which is the intermediate for the synthesis of compounds which are part of plants defensive mechanism against insects and pathogens.[13] Its activity is independent of beta site. The α active site is well conserved, similarity matrix shows 50% sequence similarity, however, the catalytic rate of the BX1 α subunit is 14 times faster than the TS complex (Figure 5.1).[14] Furthermore, the isolated α subunit of bacterial TS catalyzes the cleavage of IGP 1400 times slower than BX1.[11,15] This raises questions about the further role of the interactions between the two subunits in the overall catalytic efficiency of the enzyme as well as the structural and conformational composition of α subunit.

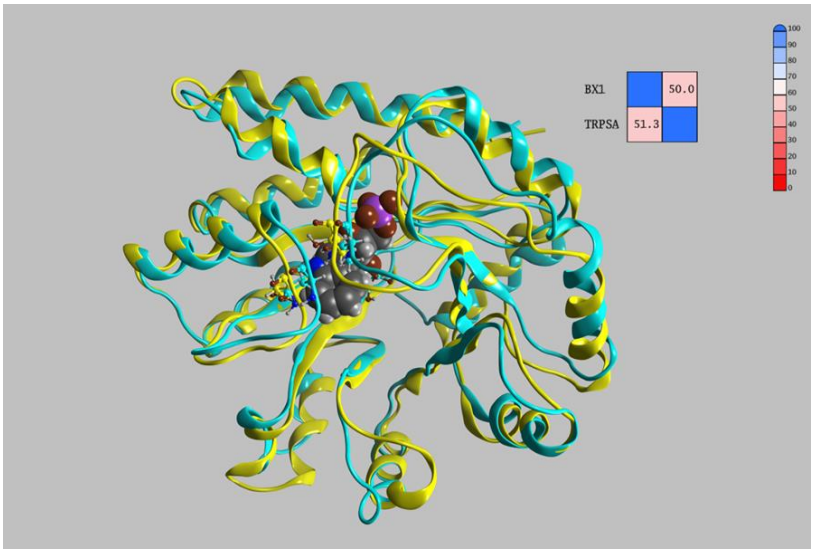


Figure 5.1 Structural alignment of αTS (teal) and BX1 (yellow). 50% sequence similarity shown in the upper right corner.

We performed several conventional and accelerated MD simulations of the wild type α subunit of TS to examine the range of conformations of catalytically important residues. According to the proposed catalytic mechanism for α subunit reaction, the active conformation of the ligand bound enzyme requires a hydrogen bond between Glu49 and hydroxyl group of IGP at the cleavage site. More recent study proposes that a water molecule mediates the retro-aldol cleavage of IGP, but our data supports the originally proposed reaction mechanism (Figure 5.2).[11,16] It has been further been suggested that the change in the conformation of catalytic Glu49 may be important in activating αTS.[17] NH of the indole ring forms a hydrogen bond with Asp60 carboxyl group. The proper positioning of Asp60 is supported by the hydrogen bond between Asp60 and Thr183 which is possible upon αL6 loop closure (Figure 5.2). Based on our data we speculate that the active conformation of Glu49 is energetically costly, and we propose a set of three mutations for which our MD data shows stabilization of the active conformation of Glu49.

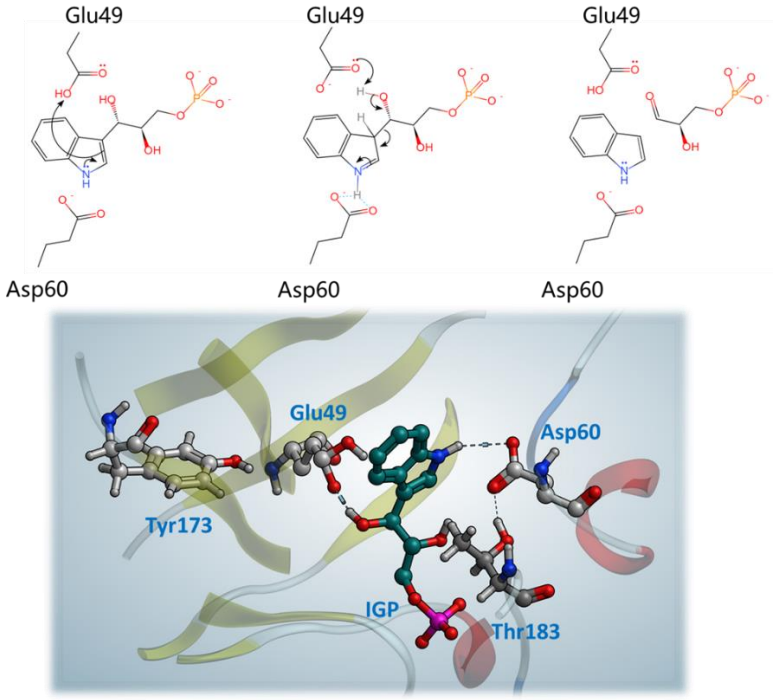


Figure 5.2 The mechanism of αTS reaction (retro-aldol cleavage) and the active conformation of Glu49.

5.2 Methods

We performed six independent conventional MD runs for αTS wild type, and two variants as follows: vTS1 contains three mutations – F22Y, G98V and Y173F; vTS2 contains the same three mutations as vTS1, two crystal water molecules were removed from the active site. For vTS3 containing two mutations – F22Y and G98V, we carried out three MD simulations (Figure 5.3). The coordinates of the wild type αTS were obtained from the crystal structure PDB ID 1QOQ. Protonation states were assigned based on MOE residues pKa calculation and notably, the pKa for Glu49 was 7.8. MD simulations were carried out using the standard Amber package with GPU

acceleration.[18,19] The protein was parameterized by using Amber Force Field FF14SB.[20] General Amber force field (GAFF) was applied to ligands and charges were assigned by using the AM1-BCC model.[21] All systems were prepared by a three-step minimization process: hydrogens, sidechains, and finally the whole system. The systems were solvated by using an explicit TIP3P water model in a rectangular box with edges at a minimum of 12 Å from any atom.[22] We added seven positive counter ions (Na^+) to neutralize the overall system charge. The solvated systems were then minimized, followed by 750-ps water equilibration at 298 K and a whole-system equilibration from 50 to 298 K at 25-K intervals for a total of 1.3 ns (50 to 275 K, 100 ps each; 400 ps at 298 K). MD trajectories were collected over 200 ns at a 1-ps interval with a 2-fs timestep under constant pressure and temperature. Particle mesh Ewald was used for long-range electrostatics and the SHAKE algorithm was applied for fixed heavy atom–hydrogen bond lengths.[23,24] The systems were visualized and analyzed by using Visual Molecular Dynamics [25] and MOE. The trajectory output files were processed with PTRAJ software[26] to contain 20000 frames, each representing 0.01-ns timestep. We further carried out four 500 ns accelerated MD simulation for wild type αTS to sample more conformations.

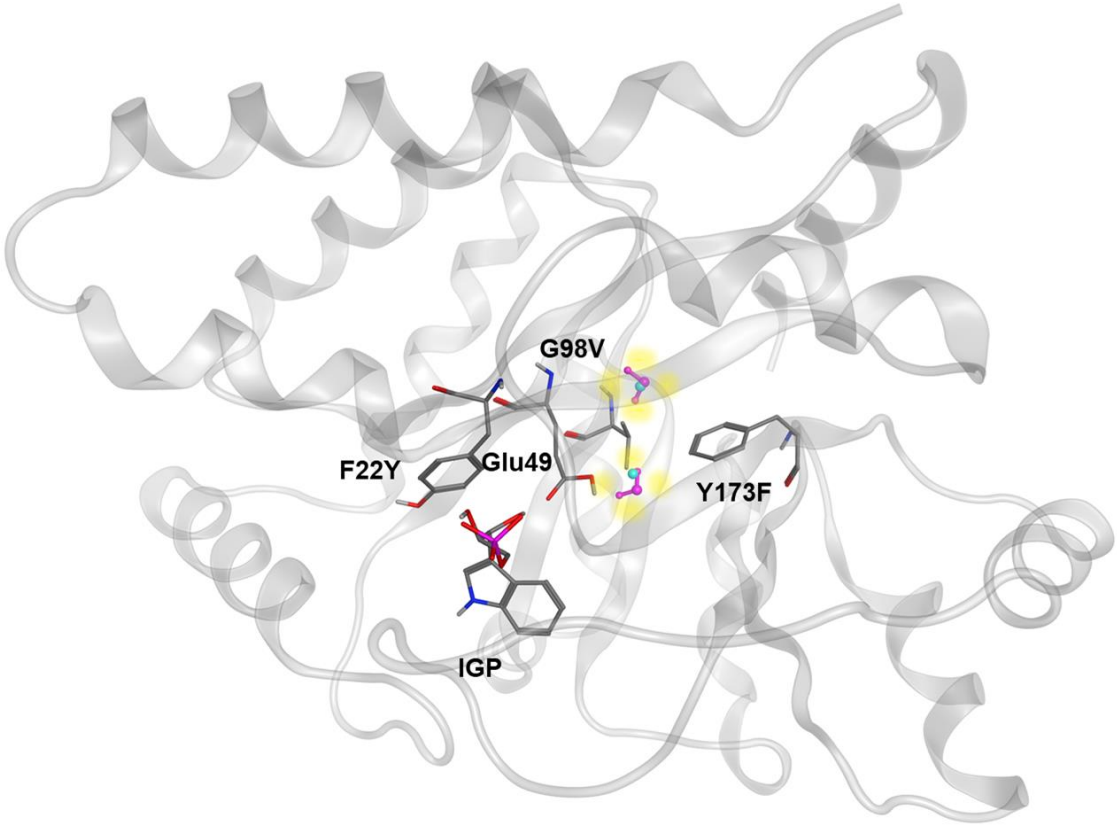


Figure 5.3. Structure of αTS, with substrate IGP, catalytic Glu49, and substituted residues shown in sticks. Crystal water waters removed in vTS2 are highlighted in yellow.

5.3 Results and Discussion

Our work, described here, focuses on the active conformation of catalytic Glu49. We analyzed six 200 ns conventional (cMD), and four 500 ns accelerated MD (aMD) simulations for the presence of the hydrogen bond between Glu49 carboxyl oxygen and C3'-OH of IGP (Figure 5.4). Based on the pKa computation, Glu49 is protonated, which is also consistent with the proposed reaction mechanism of αTS, in which Glu49 mediates a proton transfer between C3'-OH and C3 of IGP. We were not able to sample a

stable active conformation for Glu49 and consistent hydrogen bond with the substrate in cMD, however aMD showed that the hydrogen bond is indeed possible. The number of frames, in which the hydrogen bond was present did not increase significantly, however, the reversible movement of Glu49 towards and away from substrate in the aMD was convincing (Figure 5.5).

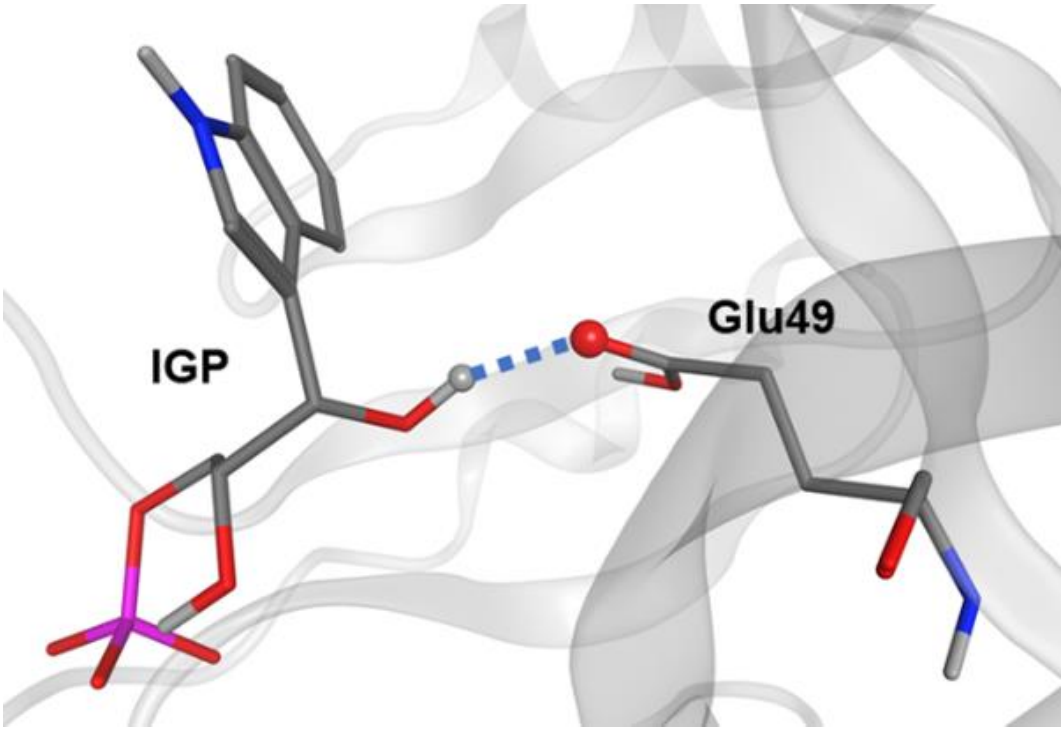


Figure 5.4 Hydrogen bond between Glu49 and IGP.

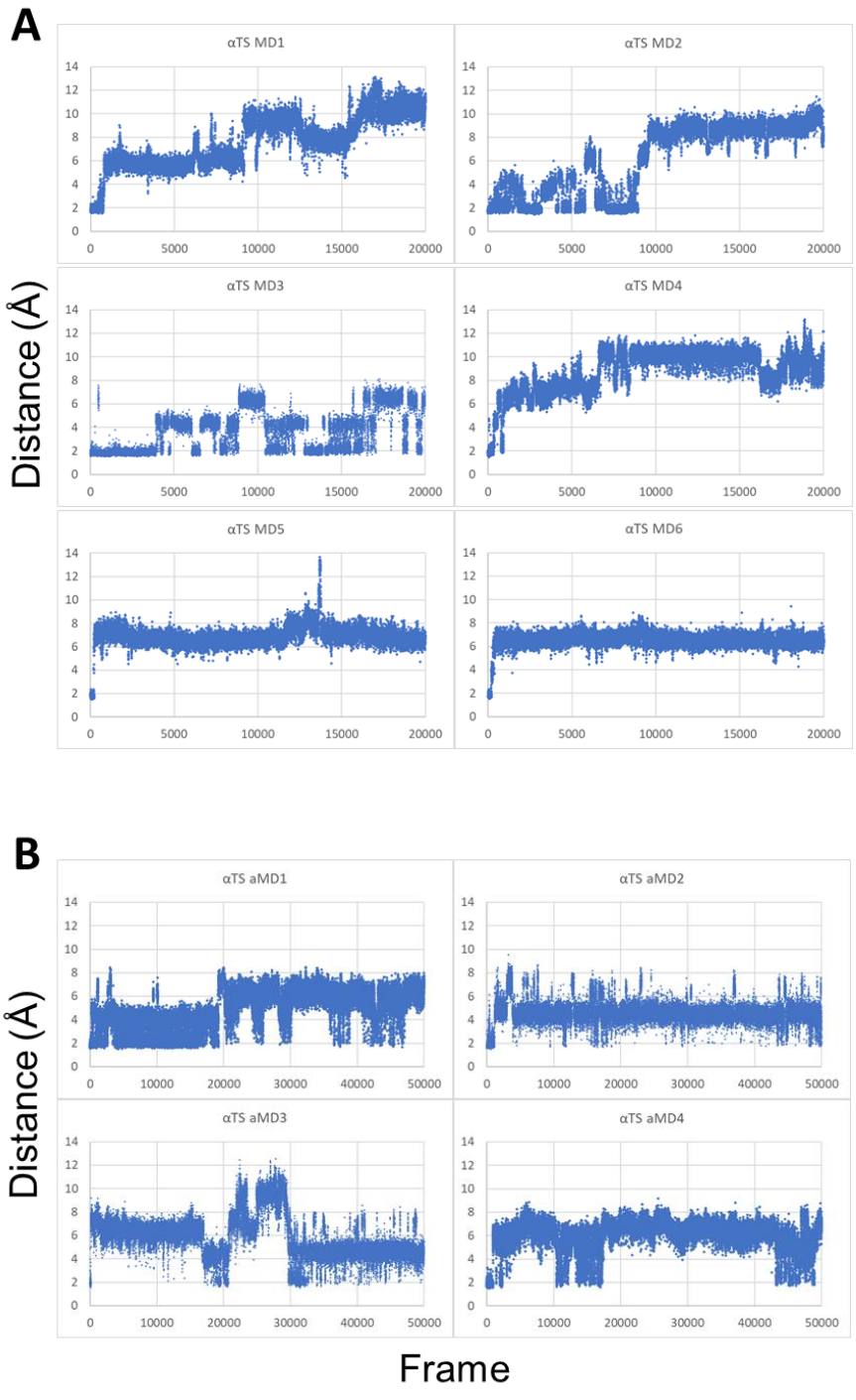


Figure 5.5 Distance between selected atoms of Glu49 and IGP in αTS in cMD (A) and aMD (B).

We further investigated what interactions prevent the catalytic residue to form a hydrogen bond with IGP. We found out that a water molecule present in the crystal structure, moves between IGP and Glu49 sidechain, disrupts the hydrogen bond between them and it forces the sidechain of the residue towards Tyr173, resulting in what appears to be a more favorable hydrogen bond between Glu49 and Tyr173 and another water molecule within the small tunnel leading to N-terminal of the protein (Figure 5.6). We compared the local amino acid composition to our model BX1 and investigated three site mutations. We simulated a variant with three substitutions at positions 22, 98 and 173 in two settings: one with the crystal waters disrupting the interactions between Glu49 and the ligand, and the other with those waters removed. We further simulated a variant containing only two substitutions to evaluate the significance of Y173F variant. We compared all trajectories for the total number of frames, in which the distance between the hydrogen of the hydroxyl group at C3' of IGP and the unprotonated carboxyl oxygen of Glu49 is small enough (≤ 3 Å) to facilitate the desired hydrogen bond, as a percent relative to all frames for each variant system combined (Table 5.1). Our data shows that removal of water molecules from the binding pocket, along with the three substitutions, has the most significant effect (4 fold increase) on the conformation of the catalytic residue and its interactions with the ligand IGP and could possibly affect the overall catalytic rate. The Y173F substitution eliminates the competing hydrogen bond, and the G98V blocks the entry of solvent. The vTS2 simulation did not show any significant change from the wild type. By substituting F22Y, we intended to allow for a hydrogen bond between Tyr22 and Thr24, which would position the aromatic ring of the residue

closer to the binding pocket “wall” and prevent van der Waals interaction with the hydrocarbon portion of Glu49 sidechain (Figure 5.7).

Table 5.1. MD frames with distance between C3’-OH of IGP and carboxyl oxygen of Glu49 (see Figure 5.4) presented as percent of all frames for each variant.

System	**Total number of frames**	**Frames with distance ≤ 3 Å(%)**
αTS	120,000	13.0
vTS1	120,000	26.2
vTS2	120,000	53.3
vTS3	60,000	20.4
aMD αTS	200,000	7.1

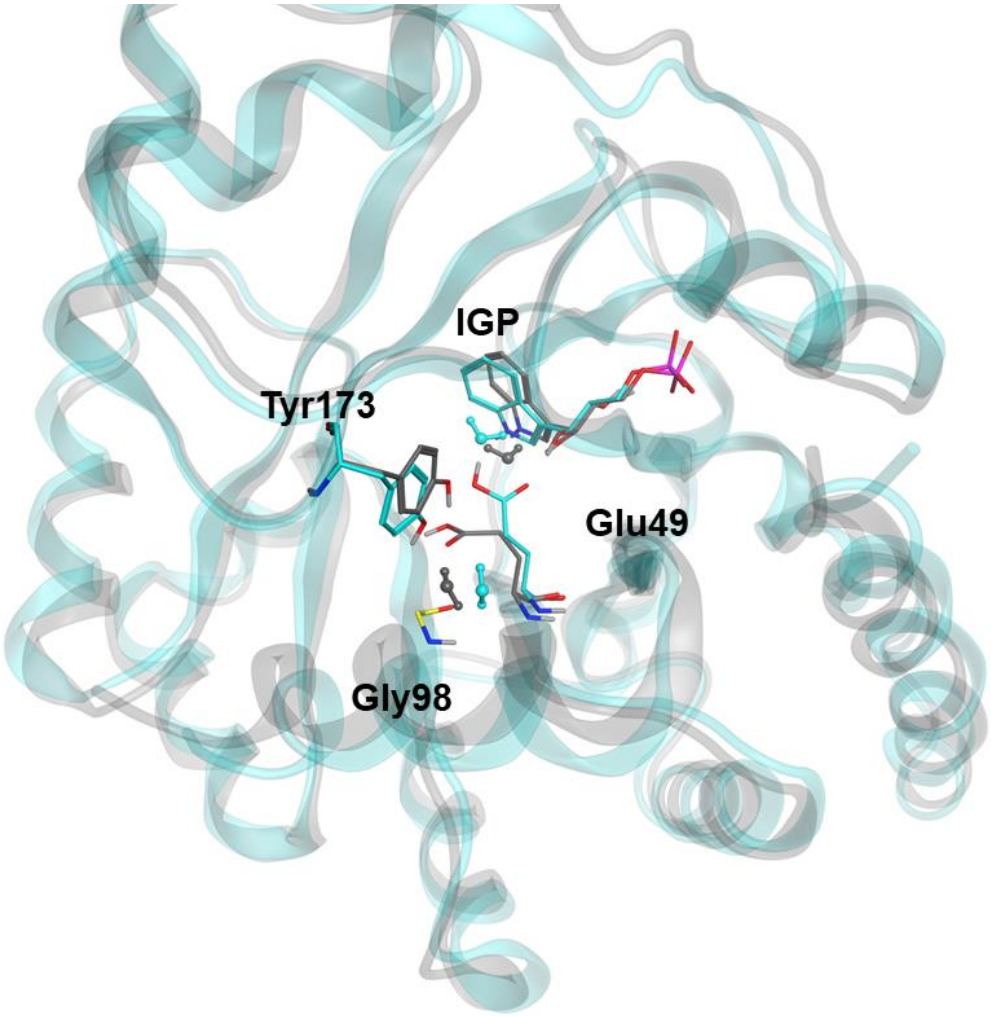


Figure 5.6. Rotation of Glu49 is caused by a water molecule moving between the residue and IGP. Two conformations are shown, before the hydrogen bond disruption (teal) and after (gray).

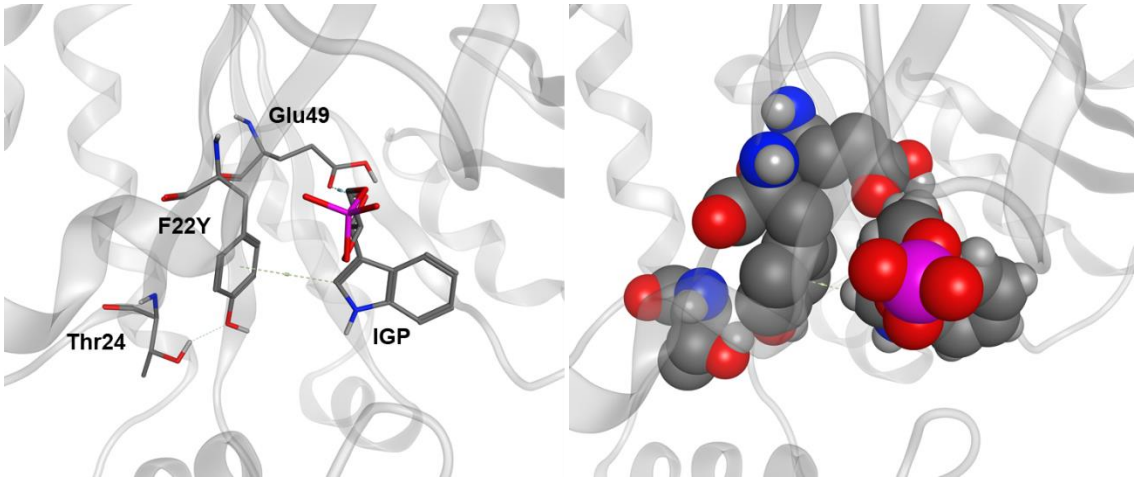


Figure 5.7. Interactions between the substituted Phe22Tyr and Thr24, shown as stick and vdW model.

5.4 Conclusion

We carried out several molecular dynamics to investigate important interactions which may have implications in the catalytic properties of the enzyme. We found out that a water molecule disrupts the interactions between Glu49 and IGP. We designed a variant with three substitutions, which in the absence of water molecules in the binding pocket, showed to substantially stabilized the active conformation of Glu49 and improved its interactions with the substrate IGP. We suggest that this set of mutations would enhance the catalytic efficiency of stand alone αTS. Further analysis of these substitutions is necessary to obtain more conclusive details for the significance of each substitution. Dihedral torsion analysis could further show the conformational populations of residues in the binding pocket and those involved in catalysis. MD simulations for each individual

site mutation may also be necessary as to understand the importance of each substituent. Nonetheless, experimental validation of our design will be crucial.

5.5 References

1. Singh R, Kumar M, Mittal A, Mehta PK. Microbial enzymes: industrial progress in 21st century. 3 Biotech 2016;6:15.

2. Gurung N, Ray S, Bose S, Rai V. A Broader View: Microbial Enzymes and Their Relevance in Industries, Medicine, and Beyond. Biomed Research International 2013;2013:18.

3. Choi JM, Han SS, Kim HS. Industrial applications of enzyme biocatalysis: Current status and future aspects. Biotechnology Advances 2015;33(7):1443-1454.

4. Kamini NR, Hemachander C, Mala JGS, Puvanakrishnan R. Microbial enzyme technology as an alternative to conventional chemicals in leather industry. Current Science 1999;77(1):80-86.

5. Jemli S, Ayadi-Zouari D, Hlima HB, Bejar S. Biocatalysts: application and engineering for industrial purposes. Critical Reviews in Biotechnology 2016;36(2):246-258.

6. Sharma A, Gupta G, Ahmad T, Mansoor S, Kaur B. Enzyme Engineering: Current Trends and Future Perspectives. Food Reviews International 2021;37(2):121-154.

7. Kapoor S, Rafiq A, Sharma S. Protein engineering and its applications in food industry. Critical Reviews in Food Science and Nutrition 2017;57(11):2321-2329.

8. Arnold FH. Directed Evolution: Bringing New Chemistry to Life. Angewandte Chemie (International ed. in English) 2018;57(16):4143-4148.

9. Almhjell PJ, Boville CE, Arnold FH. Engineering enzymes for noncanonical amino acid synthesis. Chemical Society Reviews 2018;47(24):8980-8997.

10. Arnold FH. Combinatorial and computational challenges for biocatalyst design. Nature 2001;409(6817):253-257.

11. Kulik V, Hartmann E, Weyand M, Frey M, Gierl A, Niks D, Dunn MF, Schlichting I. On the Structural Basis of the Catalytic Mechanism and the Regulation of the Alpha Subunit of Tryptophan Synthase from Salmonella typhimurium and BX1 from Maize, Two Evolutionarily Related Enzymes. Journal of Molecular Biology 2005;352(3):608-620.

12. Dunn MF. Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase bienzyme complex. Archives of Biochemistry and Biophysics 2012;519(2):154-166.

13. Frey M, Stettner C, Paré PW, Schmelz EA, Tumlinson JH, Gierl A. An herbivore elicitor activates the gene for indole emission in maize. Proceedings of the National Academy of Sciences 2000;97(26):14801-14806.

14. Frey M, Chomet P, Glawischnig E, Stettner C, Grün S, Winklmair A, Eisenreich W, Bacher A, Meeley RB, Briggs SP and others. Analysis of a Chemical Plant Defense Mechanism in Grasses. Science 1997;277(5326):696-699.

15. Kriechbaumer V, Weigang L, Fießelmann A, Letzel T, Frey M, Gierl A, Glawischnig E. Characterisation of the tryptophan synthase alpha subunit in maize. BMC Plant Biology 2008;8(1):44.

16. 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.

17. Barends TRM, Dunn MF, Schlichting I. Tryptophan synthase, an allosteric molecular factory. Current Opinion in Chemical Biology 2008;12(5):593-600.

18. D.A. Case IYB-S, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman. Amber 2018: University of California, San Francisco.; 2018.

19. Salomon-Ferrer R, Götz AW, Poole D, Grand SL, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. 2013.

20. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. 2015.

21. Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. Journal of Computational Chemistry 2002;23(16):1623-1641.

22. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics 1998.

23. Sagui C, Pedersen LG, Darden TA. Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations. The Journal of Chemical Physics 2003.

24. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational Physics 1977;23(3):327-341.

25. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. Journal of Molecular Graphics 1996;14(1):33-38.

26. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. Journal of Chemical Theory and Computation 2013;9(7):3084-3095.

Chapter 6 Conclusion and Future Directions

6.1 Conclusion

We provided an incredibly detailed analysis for various inter- and intra-molecular interactions of both systems studied in this work – bacterial tryptophan synthase and viral papain-like protease. Our results for both systems are in a remarkably good agreement with experimental data and further show that molecular modeling is a reliable approach in studying biological systems. Increasing computational power allows for longer time scale simulations which in turn permits sampling of a much larger conformational space. Our findings can be used to create potent inhibitors, gain better control of protein function and design more efficient enzymes.

6.2 Future Work

6.2.1 Enhancing Catalytic Efficiency of α Subunit of Tryptophan Synthase.

We were not able to observe consistent behavior of α subunit loop 6 in the presence of the ligand, in our simulations of *S.enterica* tryptophan synthase, the loop opens with IGP in the binding site and the substrate is exposed to solvent. Since BX1 enzyme functions as a single subunit we need to further examine the amino acid content and behavior of the loop to find out the mechanism of opening and closing in the absence of β subunit. Investigation of the interactions between individual residues, similarly to our analysis of tryptophan synthase in *E.coli* (Chapter 2), may help us understand the dynamics of the loop better and present us with further possibilities into enhancing the catalytic efficiency of α subunit.

6.2.2 Inhibition of Papain-like Protease in Coronaviruses

We continue to study papain-like protease and potential drug candidates, both covalent and non-covalent leads. We screen libraries for possibly repurposing currently available drugs. We are planning to further expand our investigation to include PLpro interactions in a complex with ubiquitin and ISG15. Gaining a better insight into these protein-protein interactions may reveal further drug target sites and provide better understanding of the mechanism by which this viral protein is able to invade the host immune system. Zhang et al. have engineered highly specific and effective Ub variants inhibiting the function of MERS PLpro. Crystal structures for the Ub variants in complex with the viral PLpro are also available.[1] We aim to perform MD simulations for these complexes and compare protein-protein interactions with PLpro in complex with wild type Ub.

6.3 References

1. Zhang W, Bailey-Elkin B, Knaap RCM, Khare B, Dalebout TJ, Johnson GG, van Kasteren PB, McLeish NJ, Gu J, He WG and others. Potent and selective inhibition of pathogenic viruses by engineered ubiquitin variants. Plos Pathogens 2017;13(5):21.