Using Molecular Dynamics to Examine the Effect of a Novel L505H Mutation on Drug-Resistant B-Raf-V600E and to Understanding the Mechanistic Role of the Mobile Loop Region in Alkanesulfonate Monooxygenase

by

Nicole Ippolito Cronin

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
Requirements for the Degree of
Master of Science

Auburn, Alabama
December 12, 2015

Keywords: chemistry, biochemistry, molecular dynamics, drug discovery, B-Raf, monooxygenase

Approved by

Orlando Acevedo, Chair and Associate Professor of Chemistry
Holly Ellis, Associate Professor of Chemistry
Konrad Patkowski, Assistant Professor of Chemistry
Bradley Merner, Assistant Professor of Chemistr

Abstract

Two main projects were studied in this work: 1) the effect of type I and type II inhibitors on B-Raf V600E and the B-Raf V600E-L505H mutations 2) the mechanistic role of the mobile loop in SsuD in alkanesulfonate monooxygenase. Chapter 1 describes the background of the two different systems studied. Chapter 2 describes the theory of the computational methodology used in these studies. Chapter 3 reports the results of the simulation carried out on B-Raf V600E and B-Raf V600E-L505H in order to show the differences in inhibitor binding. Chapter 4 shows the computational results pertaining to the mobility of the dynamic loop for the SsuD component of alkanesulfonate monooxygenase and its role in catalytic activity.

Acknowledgements

I would like to thank my advisor Dr. Orlando Acevedo for helping me expand my knowledge of chemistry into the computational field. This would not have been possible without him. I would like to also thank our collaborator Dr. Holly Ellis for her help with the monooxygenase system. A big thanks should go to my committee members, Dr. Holly Ellis, Dr, Konrad Patkowski, and Dr. Bradley Merner for their support and work on my committee. My lab mates Dr. Symon Gathiaka, Dr. Kira Armacost, and Brittany Boykin have also been of invaluable support. I would like to thank my family and my husband for keeping me on track to gain an upper level degree in chemistry.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

EPK – Eukaryotic Protein Kinase

PKA – Protein Kinase A

MAPK – Mitogen Activated Protein Kinase

FMN, FMR – Reduced flavin

FMNOO-, FMP – Oxidized flavin

OCS - Octanesulfonate

SsuD – Alkanesulfonate Monooxygenase

SsuE – FMN Reductase

MD – Molecular Dynamics

aMD – Accelerated Molecular Dynamics

CpHMD – Constant pH Molecular Dynamics

CpHaMD – Constant pH Accelerated Molecular Dynamics

MC – Monte Carlo

AMBER – Assisted Model Building with Energy Refinement

GAFF – General Amber Force Field

RESP – Restrained Electrostatic Potential

BCC – Bond Charge Correction

CG – Conjugate Gradient

PCA – Principal Component Analysis

EDTA – Ethylenediaminetetraacetic acid

NTA – Nitriloacetic acid

DTA - Diethylenetriaminepentacetate

DHK - Dihydrokalafungin

DBT - Dibenzothiopene

Chapter 1 Introduction

B-Raf V600E-L505H

Protein Kinases

In the 1970's, a discovery was made that showed phosphorylation regulated the activity of glycogen phosphorylase. That discovery sparked the search for more understanding on how phosphorylation governed the protein function.[1] A total of 518 kinases have been discovered , and they make up nearly 1.7% of the human genome.[2]

Kinases are the mediator for signal transduction in cells in the majority of eukaryotic organisms.[3] Other functions of kinases are metabolism, transcription, movement, differentiation, and apoptosis just to name a few. Because protein kinases play such a crucial role in cellular function, they offer a good target for genomic treatements.[4] The eukaryotic protein kinase (EPK) super family's core residues consist of about 250 – 300 amino acids which make up the kinase domains, also known as the catalytic domains.[5] There are 12 conserved subdomains.[5] Each lobe of a kinase is made up of helical and β subdomains. Figure 1.1 shows the structure of protein kinase A (PKA).

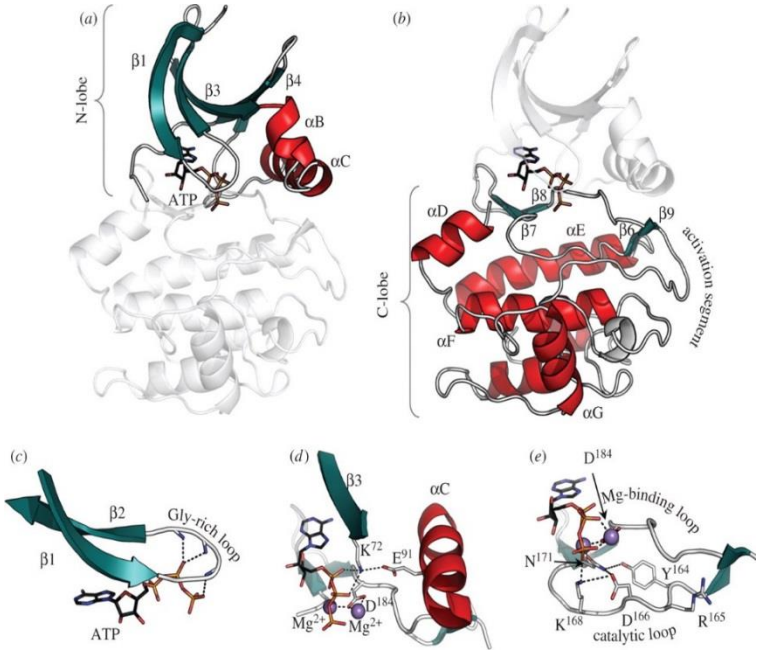


Figure 1.1 The structure of PKA is an example for the EPK superfamily. Helices are shown in red and β-strands are shown in teal. a) The N-lobe contains 5 β-strands and a αC-helix. b) The C-lobe is mostly helical and contains the activation segment with the β-sheet resting on active site core. c) The γ-phosphate of ATP positions near the glycine rich loop. d and e) This shows the catalytically pertinent residues for activation and catalysis. (Taken from figure 1 of Taylor, S. S., Keshwani, M. M., Steichen, J. M., & Kornev, A. P. (2012). *Philosophical Transactions of the Royal Society B: Biological Sciences*,*367*(1602), 2517–2528. 2012.0054[6])

Two lobes converge to form a deep active site cleft in the middle of the protein where ATP binds. Catalysis is governed by the opening and closing of the cleft. The N-lobe consists of five stranded anti-parallel β-sheets with a conserved αC-helix. β-strand 1 and β-strand 2 are linked by the glycine rich loop which anchors the ATP γ-phosphate. β-strand 3 contains the vital residue LysX which forms a salt bridge with GluX when the kinase is in the active conformation. The C-

terminal lobe of a kinase is mostly helical and contains the DFG-loop motif. The R-spine or hydrophobic spine is formed between β-strand 9 and the αF-helix. The spine connects the N and C lobes of the kinase. The orientation of the spine may help dictate kinase activity.[7]

The Mitogen Activated Protein Kinases, specifically the Ras/RAF/MEK/ERK signaling cascade

Mitogen Activated Protein Kinase, MAPK, pathways are present in all eukaryotic cells, and there are a variety of different pathways. In mammals, 14 MAPK's have been identified, and they are classified into two categories: conventional and atypical. Conventional MAPK pathways are made up of the extracellular signal-regulated kinases, and the atypical MAPK pathways do not conform to any specific function.[8, 9] Figure 1.2 shows the different MAPK pathways.

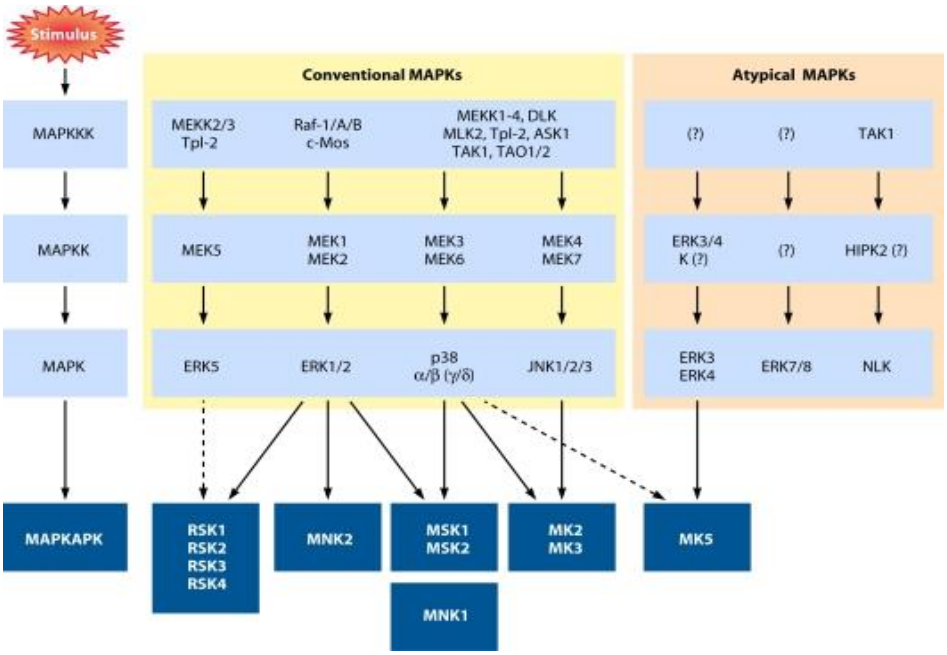


Figure 1.2 Representation of conventional and atypical kinase pathways. (Taken from figure 2 of Cargnello, M. and Roux, P. Microbiol. Mol. Biol. Rev. 2011 75(1), 50-83 [9])

Proteins in this pathway can be located in many different cellular compartments such as the nucleus, cytoplasm, or membrane. Because of this, the pathway gives a clear signal from the cytoplasm to the nucleus.[10] The proteins involved in this pathway are Ser/Thr kinases which are activated through phosphorylation events.[9]

The Ras/RAF/MEK/ERK pathway is a conventional MAPK pathway that regulates signaling as well as apoptosis. Figure 1.3 shows the extreme complexity of this conventional MAPK pathway.

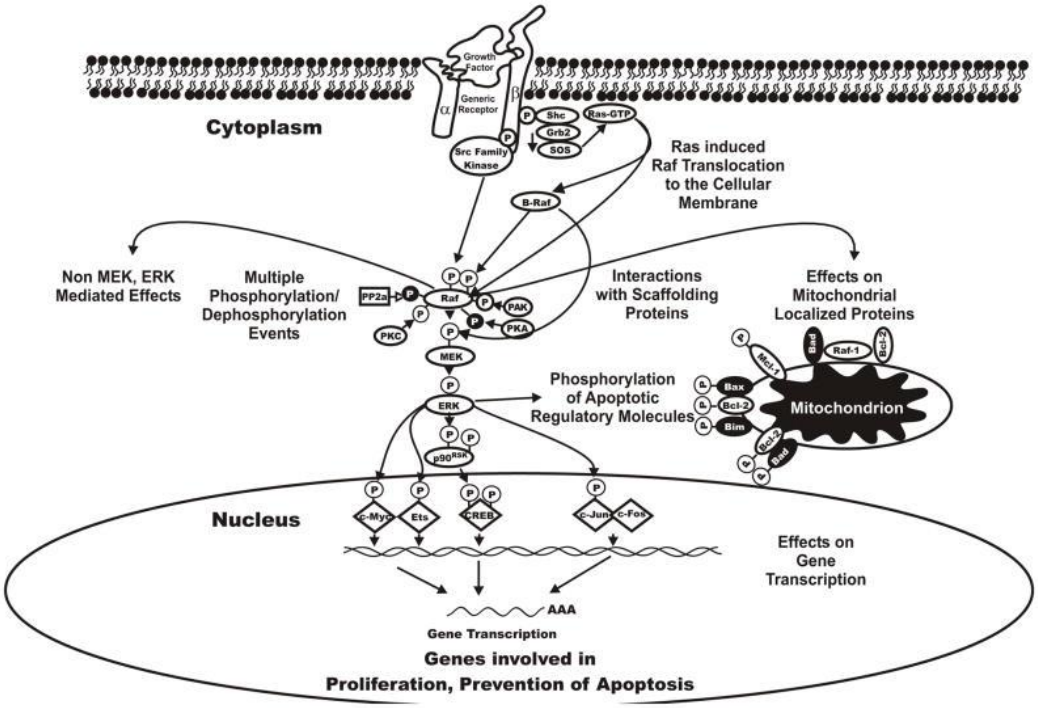


Figure 1.3 Overview of the RAF/MEK/ERK pathway. (Figure taken from figure 1 of J. A. McCubrey, L. S. Steelman, S. L. Abrams, et al. *Adv Enzyme Regul*, 2006, 46, 249-279 [11])

This pathway begins with the membrane bound, GTP-binding protein RAS, which is common to several pathways including RAF/MEK/ERK, PI3K/Akt, and RalEGF/Ral. Several Ras proteins have been identified: Ha-Ras, N-Ras, Ki-Ras 4A/4B of which Ki-Ras is the prime activator for the RAF/MEK/ERK pathway.[12] Ras usually under goes farnesylation to bind to the

membrane and following the binding of Ras to the correct growth factor, GDP is exchanged for GTP. A conformational change then occurs and Ras becomes active. This activation allows RAS to directly interact with its targets, such as Raf.

Raf has several mechanisms that govern the regulation. The first method of regulation is the interaction of Raf with Ras by recruitment of Raf to the membrane by GTPase. Ras binds with Raf in two domains: the CR1 domain and the cysteine rich domain. Phosphorylation of Raf is also another key mechanism in its regulation; however, some of these residues are already phosphorylated in the inactive state.[13] A 14-3-3 protein can bind to the Raf-1 CR2 region and may act as a stabilizer.[3] Another protein S621 becomes dephosphorylated by an unknown phosphatase, which then allows 14-3-3 to dissociate from Raf-1, allowing Raf-1 or A-Raf to become phosphorylated at S338, Y340, and Y341. These sites are replaced with aspartic acid in B-Raf (D492 and D493). An SRC family kinase is most likely responsible for the A-Raf activation; however, B-Raf looks to be SRC independent. Protein kinase C (PKC) can induce crosstalk between Raf/MEK/ERK which can also be an activating factor for Raf. Raf can be inactivated through phosphorylation on the CR(2) domain.[14] Akt and protein kinase A (PKA) phosphorylate S259 on Raf-1 causing inhibition. On B-Raf Akt will phosphorylate S364 and S428 to inhibit activity. They again bind with the 14-3-3 protein yielding inactivity. Dimerization of Raf heterodimers may also play an important role in activation.[11]

MEK is both a tyrosine and serine/threonine kinase. The activity is regulated by Raf phosphorylation on the serine residues in the catalytic domain. B-Raf is shown to be a greater activator to MEK than Raf-1/A-Raf and any of the Raf isoforms.[15] The downstream target of MEK is ERK. ERK 1/2 are serine-threonine kinases, and they are regulated through the phosphorylation

of MEK 1/2. ERK can phosphorylate many transcription factors and has many targets including NF-κB and ribosomal S6 kinase.[11]

B-Raf Structure and Function

The RAF protein is an ATP binding protein and is found in a homodimeric structure or heterodimeric structure with its isoforms. Three isoforms of Raf exist: A-Raf, B-Raf, and C-Raf.[16] In the inactive form of RAF, it is found in the cytoplasm with the N-terminal domain inhibiting the C-terminal domain. On activation, RAF moves to the plasma membrane where it drives phosphorylation at activating regions and phosphorylation at inhibiting regions. The main regions in Raf (Figure 1.4) involve a glycine rich loop (red), activation segment (blue), αC-helix (green), and catalytic loop(purple). The glycine rich loop helps anchor the ATP. The αC-helix can make or break the ATP binding site and also helps bind with MEK 1/2.

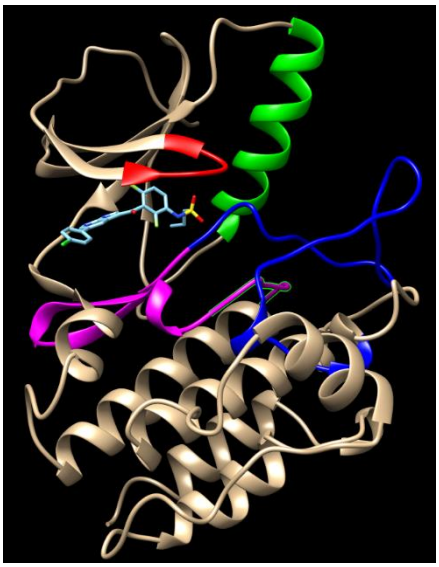


Figure 1.4 B-Raf V600E monomer bound with inhibitor PLX 4032. The red region specifies the glycine rich loop. The green region is the αC-helix. The purple region contains the catalytic loop, and the blue specifies the activation segment.

Figure 1.5 shows the crucial residues in the catalytic cleft. The activation segment contains two crucial amino acids, phenylalanine 595 and asparagine 594. When the Phe595 is pointed toward the active site region, Raf is in the inactive conformation. When Asp594 is pointed to the active site region it coordinates with Mg^{2+}, and Raf is in the active form.

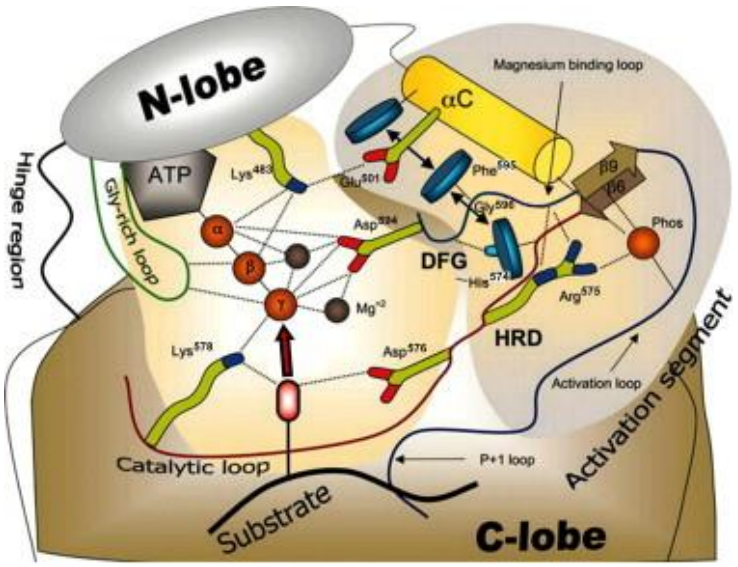


Figure 1.5 Diagram of the B-Raf active site. (Figure taken from figure 4 of Roskoski, R. *Biochemical and Biophysical Research Communications*, 2010, 399(3), 313-317 [17])

The activation segment will close off the active site preventing the solvent from entering because of favorable interactions with the glycine rich loop and αC-helix. The catalytic segment also contains a few key residues. Lysine 578 binds to the gamma-phosphate of ATP helping stabilize it further. Asparagine 576 abstracts a hydroxyl group which helps perform a nucleophilic attack on the oxygen from gamma-phosphorous atoms of MgATP.[17]

B-Raf is the most common isoform and is the main focus of this study. The B-Raf protein is mediated by equilibrium between active and inactive forms. The wild type B-Raf favors the

inactive conformation because of favoring interactions between the glycine rich loop and activation segment. However, a point mutation occurs at the Val600 amino acid resulting in the substitution of a valine to glutamate, and because of the polarity change acts as if the phosphorylating event already occurred.[18] This mutation bypasses the requirement for Ser602 and Thr599 phosphorylation by causing favorable interactions between the glycine rich loop and activation segment leading to uncontrollable phosphorylation to proteins downstream such as MEK.[19] The substitution of a neutral amino acid to a negatively charged amino acid also allows a net negative charge of the phosphorylating region. The glutamate is thought to potentially act as a phosphomimetic.[20] The V600E mutation may cause a 500 fold increase in kinase activity. This mutation is seen in 30-60% of melanomas, 30-50% thyroid cancers, and 5-20% colorectal cancers. This makes this particular protein a good target for inhibition for small molecule cancer drugs.[18]

Melanoma: Stages, Survival, and Current Treatments

Melanoma starts in the epidermis which is made up of three kinds of cells: squamous cells, basal cells, and melanocytes. Melanoma occurs when melanocytes, which are found in the lower layer of the epidermis, no longer function properly. It is the rarest form of skin cancer, and it is the most likely to invade other tissues and organs. It can appear anywhere on the skin. The cause of this cancer can be attributed to UV exposure which causes mutations. Melanoma may also be genetic, but the cause as to why it happens is unknown.[21]

Melanoma is represented in four stages each differing in degrees of severity. Stage 0 shows melanoma only in the epidermis, on the top layer of skin. It has not spread to any other layers. Stage 1A/1B shows melanoma only in the skin 1 mm-2 mm thick, and that it has not spread to other portions of the body. Stage 2A/2B, like stage 1, only has melanoma in the skin, but the

thickness of the spread increases to 2 mm-4 mm. Stage 3A/3B/3C shows melanoma has spread to lymph nodes and may be ulcerated depending on the severity. Stage 4 occurs when melanoma spreads to other parts of the body away from the point of origin.[22]

The common treatments for melanoma include chemotherapy, radiation therapy, and surgical treatments. Immunotherapies, such as Ipilimuab, help boost the immune system in order to fight the cancer cells. These therapies are useful for early stage cancer; however, they are a broad target therapy.[23] For higher stage cancer, a more direct therapy would be more efficient, and this leads to genomic treatments and the development of small molecule inhibitors. The drug that is used most often for melanoma today is the type I inhibitor, Vemurafenib, or PLX 4032.[24]

Difficulty in Inhibitor Design

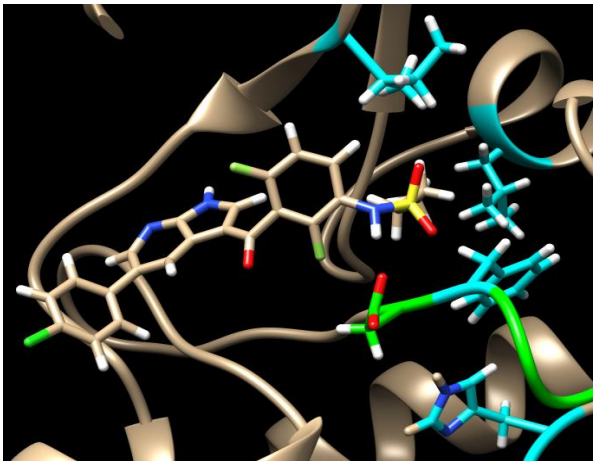
There is some difficulty in designing these inhibitors. If a mutation occurs in the binding region of the inhibitor, it may lead to a resistant strain of cancer in which the inhibitor is no longer effective. Resistance to PLX 4032 may involve several factors such as overexpression of PDGFR-B or increased apoptotic signaling.[25] The MAPK pathway may also be reactivated downstream from Raf.[17] The pathways from Ras to the two ERK proteins are extensive with the number around 1152, assuming all interactions occur between isoforms of the proteins. This number would decrease if some interactions would not occur.[17] Also, Raf can be bound with any of the isoforms either as a heterodimer or homodimer. Depending on the dimerization of Raf, otherwise if B-Raf dimerizes with C-Raf or some other isoform, it may activate ERK even with the inhibitor present. There is something known as paradoxical Raf activation which shows that cells exposed to an inhibitor may also increase Raf activation, hence the reason for some skin cell carcinomas especially with the type I inhibitors as wild type Raf activation is dependent on the dosage of the

inhibitor. Raf dimerization also has a large role to play in development of inhibitors; however, this study focuses on investigating inhibitors for the B-Raf homodimer.[16]

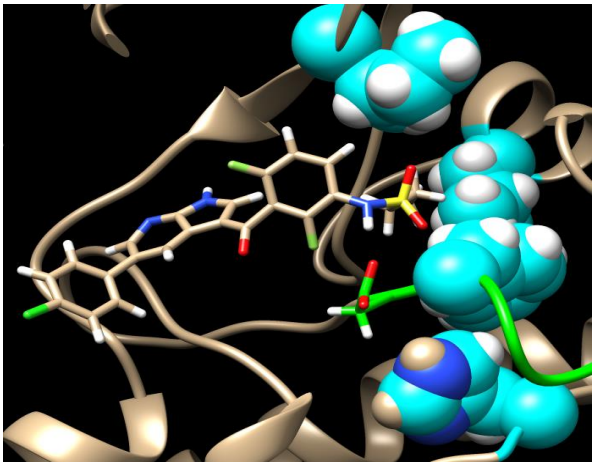
Inhibitor Type

With the recognition of these problems there are two types of inhibitors that are developed: type I and type II.[26,27] Type I inhibitors bind directly to the active site of a protein and specifically target ATP binding. Type II inhibitors may bind to the ATP binding site, but they may also bind to an allosteric site causing distortion of the ATP binding site region, thereby inhibiting the protein. Both are competitive ATP binding inhibitors. The mechanism by which these inhibitors work is not well understood; however, a theory known as the Oncogenic Shock Model states that pro-death signals decay slower than pro-survival signals, and this occurs when a small molecule inhibitor is introduced.[28] This results in death of the cancer cell. It is known that the type I inhibitors bind with the DFG-in conformation and type II inhibitors bind with the DFG-out conformation shown in figure 1.6.[29]

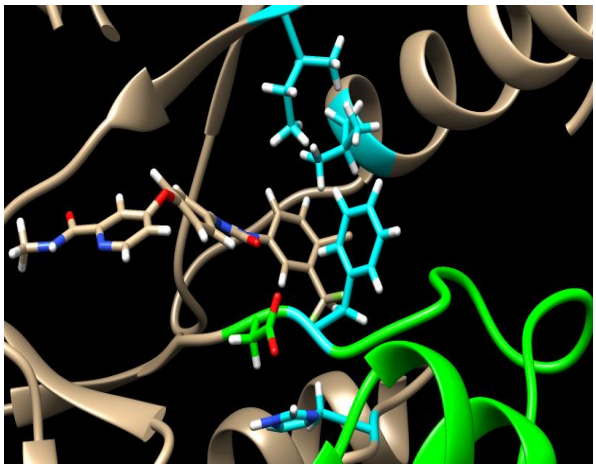
A) PLX 4032 with BRAF V600E- DFG-in



B) PLX 4032 with BRAF V600E- R-spine



C) Sorafenib with BRAF V600E – DFG-out



D) Sorafenib with BRAF V600E R-spine

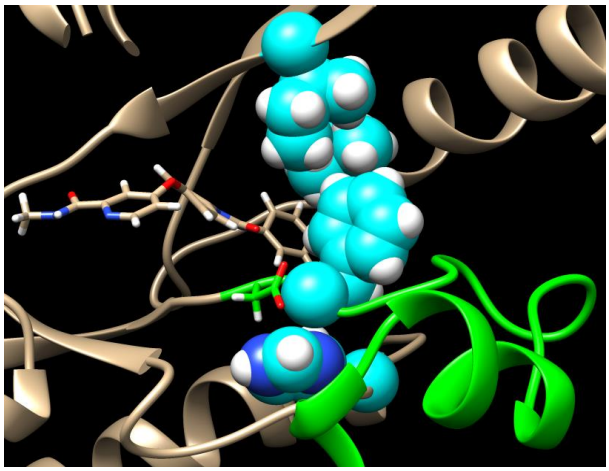


Figure 1.6 DFG-in and DFG-out conformations. A,B) BRAF V600E is bound with PLX 4032, a type I inhibitor. A) Occupation of Asp in the active site for the DFG-in conformation. B) R-Spine is broken. C,D) Sorafenib, a type II inhibitor, bound with BRAF V600E. C) Phe occupies binding site for the DFG-out conformation. D) Shows attached R-spine.

Type II inhibitors are also known to be more kinase specific than type I inhibitors because of their interaction with the activation segment of the kinase. Both types of inhibitor should have one to three hydrogen bonds with the enzyme.[30]

B-Raf V600E-L505H Mutation

B-Raf V600E was noticed to develop a resistance to the FDA approved drug PLX 4032 in about 1% of cancer cells.[31] Because of this, the cancer comes back drug resistant. It has been found that another mutation occurs at the 505 amino acid.[32] Leucine, a nonpolar residue, mutates into a histidine, a polar, positively charged residue, and this mutation is the cause of drug resistance. Several experimental studies were done in order to show that this mutation occurs as well as show

how it affects inhibitors. First, exome sequencing was done and showed a mutation from GA to AT in the DNA sequence which was linked to the 505 amino acid. In resistant cells, this mutation occurred about 20% of the time. IC50 values for the drug PLX 4032 increased with this mutation, and a western blot analysis showed that downstream activation was occurring even with the inhibitor present. Finally, further experiments showed that this mutation had little effect on MEK 1/2 inhibitors and on combination inhibitors. Because type I inhibitors mimic the binding domain, this mutation may cause them to become ineffective. However, type II inhibitors should leave the conformational loop open; therefore, this mutation should have little effect on these inhibitors.[32]

Accelerated molecular dynamic (aMD) studies were carried out to determine the effectiveness of the different types of inhibitors. Simulations were carried out on B-Raf V600E, B-Raf V600E-L505H, and B-Raf V600E-L505H(P) (protonated His505) using type I and type II inhibitors for a total of 500 ns. Chapter 3 describes the results of the aMD simulations.

Alkanesulfonate Monooxygenase

The Two-Component Monooxygenase System

Flavoproteins can catalyze many reactions due to a flavin's multiple redox states. Flavins usually bind covalently but also can be a substrate for the enzyme. Two component flavin dependent enzymes have a variety of purposes such as catalyzing the oxidation of aromatic and polycyclic compounds for carbon sources, biosynthesis of antibiotics, bioluminescence, oxidation of long chain alkanes, and desulfurization of compounds.[33] All of these systems rely on a reductase component and oxygenase component. The reductase components catalyze the reduction of reduced flavin and differ in their specificity for NADH, NADPH or the ability to use pyridine

nucleotide. The flavin may either be a prosthetic group or a substrate which dictates the mechanism for flavin transfer.[33]

Monooxygenases react with dioxygen to oxygenate the substrate. Several monooxygenases have been characterized and their reactions are shown in figure 1.7.

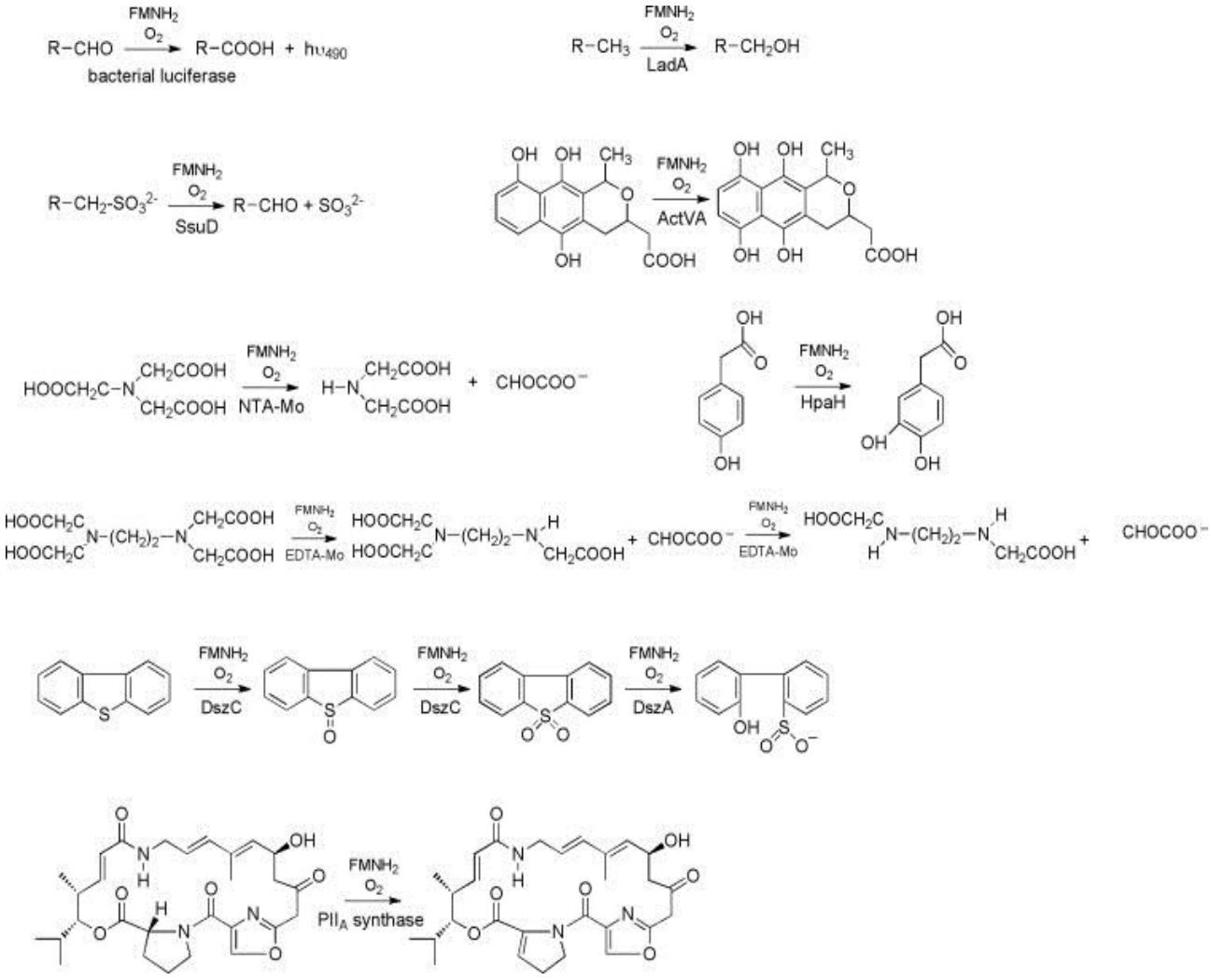


Figure 1.7 The reactions of multiple monooxygenase enzymes all of which require a reductase and are reduced flavin dependent. (Figure taken from figure 1 of Ellis, H. *Archives of Biochemistry and Biophysics* 01 March 2010 [34])

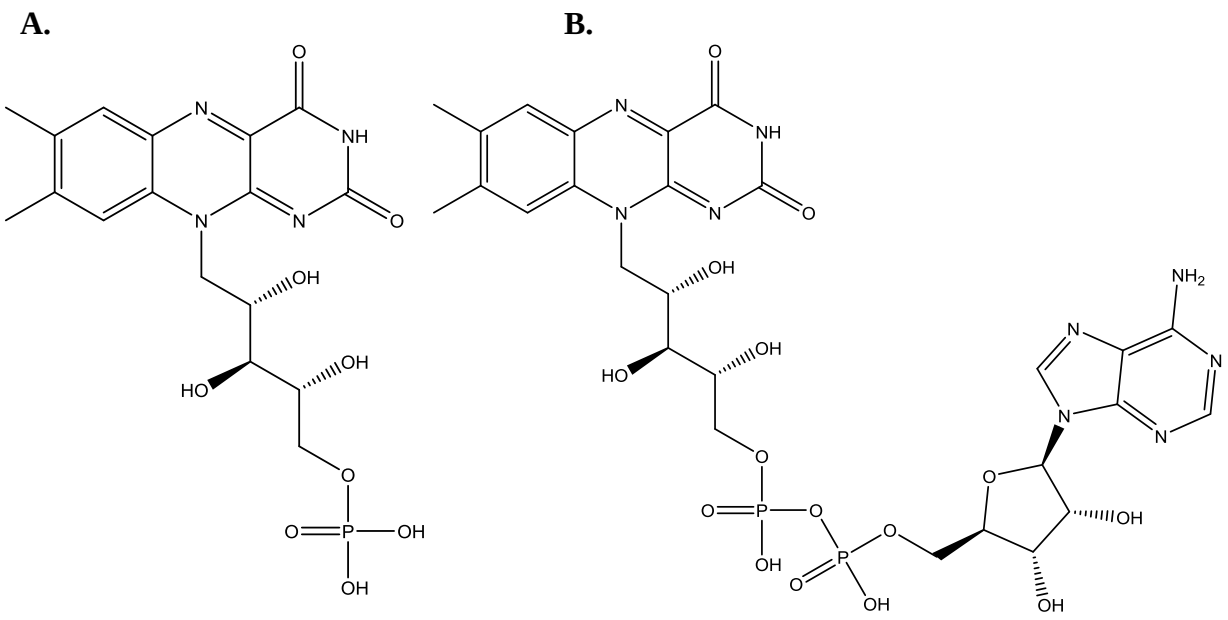


Figure 1.8 A. Flavin Mononucleotide (FMN) B. Flavin Adenine Dinucleotide (FAD)

Bacterial luciferase has been widely studied and is probably the most characterized monooxygenase. In its reaction, it uses a long chain aliphatic aldehyde, reduced flavin, and dioxygen to form a carboxylic acid and light.[35] LadA, located in *Geobacillus thermodenitrificans*, catalyzes the reaction of long chain alkanes ranging from C15-C36 to primary alcohols.[36] Monooxygenases can also catalyze the degradation of chelating agents such as ethylenediaminetetraacetic acid (EDTA) and nitriloacetic acid (NTA) for sources of carbon, nitrogen and energy. There are two known enzymes that do this: NtaA (NTA-mo) and EmoA (Emo-Mo).[37,38] EmoA is an EDTA degrading monooxygenase that oxidizes EDTA to etheylenediaminetriacetate, and then finally to ethtylenediaminediacetate. It can also use NTA and diethylenetriaminpentacetate (DTA), and may or may not have a metal complex. This monooxygenase is FAD or FMN dependent (figure 1.8). Monooxygenases characterized for the synthesis of antibiotics are Act-Va and SnaA. Act-Va is involved in actinorhodin biosynthesis and through hydroxylation of dihydrokalafungin (DHK) leads to the actinorhodin product.[39] SnaA

catalyzes the synthesis of polyunsaturated cyclic peptide pristinamycin (II_A), and this enzyme is different because it involves oxidation rather than oxygenation.[40] Some monooxygenases are involved in sulfur acquisition and actually make up a 3-compenent system. DszC and DszA are both FMN dependent monooxygenases. First DszC converts dibenzothiophene (DBT) to DBT sulfone. DszA then converts DBT sulfone to 2-hydroxybiphenyl-2-sulfinite.[41] Alkanesulfonate monooxygenase, the two-component system pertaining to this study, is involved in sulfur acquisition.

The Interaction between SsuE and SsuD

Some bacterial organisms rely on inorganic sulfur for the biosynthesis of sulfur compounds; however, during periods of sulfur starvation these organisms have found a way to generate proteins that use alternate sulfur sources.[42] Two operons are expressed in organisms at this time to allow sulfur uptake: tau and ssu. The tau operon is a taurine dyhydroxygenase, and an ABC-type transporter that is involved in uptake of taurine. The ssu operon encodes for SsuE and SsuD and is also an ABC-type transporter for sulfur transport.[43] *Escherichia coli* is a bacterial organism that has developed a way to acquire sulfur during starvation periods by synthesizing a two component system from the ssu operon: FMN reductase (SsuE) and alkanesulfonate monooxygenase (SsuD).

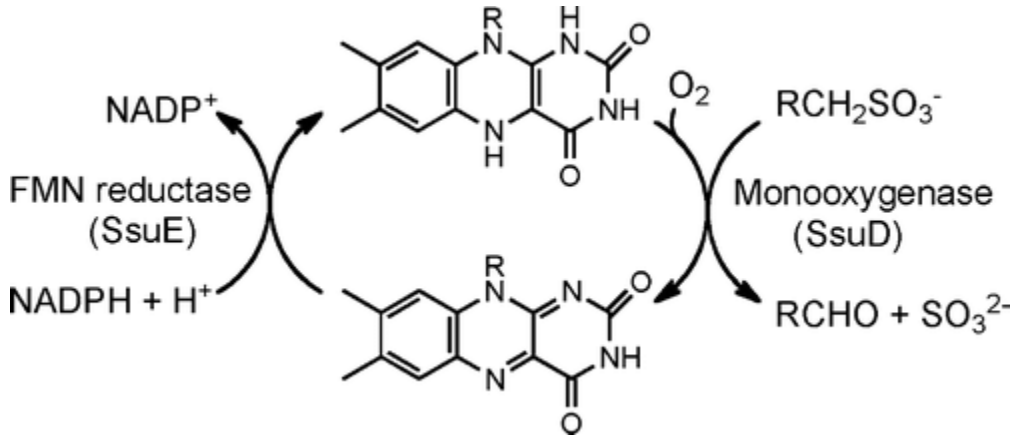


Figure 1.9 The overall reaction for the two SsuE and SsuD. NADPH is converted to NADP+ to produce reduced FMN which reacts with SsuD to form the aldehyde and sulfite products. (Figure taken from scheme 1 from Carpenter, R. et al. *Biochemistry* 2011 *50* (29), 6469-6477 [44])

The flavin-reductase (SsuE) catalyzes the reduction of FMN by NADPH. This then transfers the reduced flavin to the alkanesulfonate monooxygenase (SsuD) which catalyzes the production of sulfite and aldehyde from the alkanesulfonate. SsuE does not show much sequence similarity with other flavin reductases; however, the mechanism of reduction is comparable.[45] Steady-state kinetic assays showed that SsuE follows a sequential order in its mechanism with NADPH as the first substrate to bind and NADP+ as the first to leave.[45] The system then moves to a rapid equilibrium ordered mechanism which shows that protein-protein interactions are significant for the flavin transfer to SsuD.[45] These protein interactions were seen as stable through affinity chromatography and spectroscopic analysis which means a strong possibility for direct flavin transfer.[46]

Structure and Mechanism of SsuD

SsuD is the enzyme that catalyzes the desulfonation reaction and requires the reduced flavin from SsuE. It is a homotetramer that contains four active sites, one in each subunit. The subunit is composed of a TIM-barrel fold with the active site at the C-terminal end of the β-barrel, which is typical for the TIM-barrel proteins.[47]

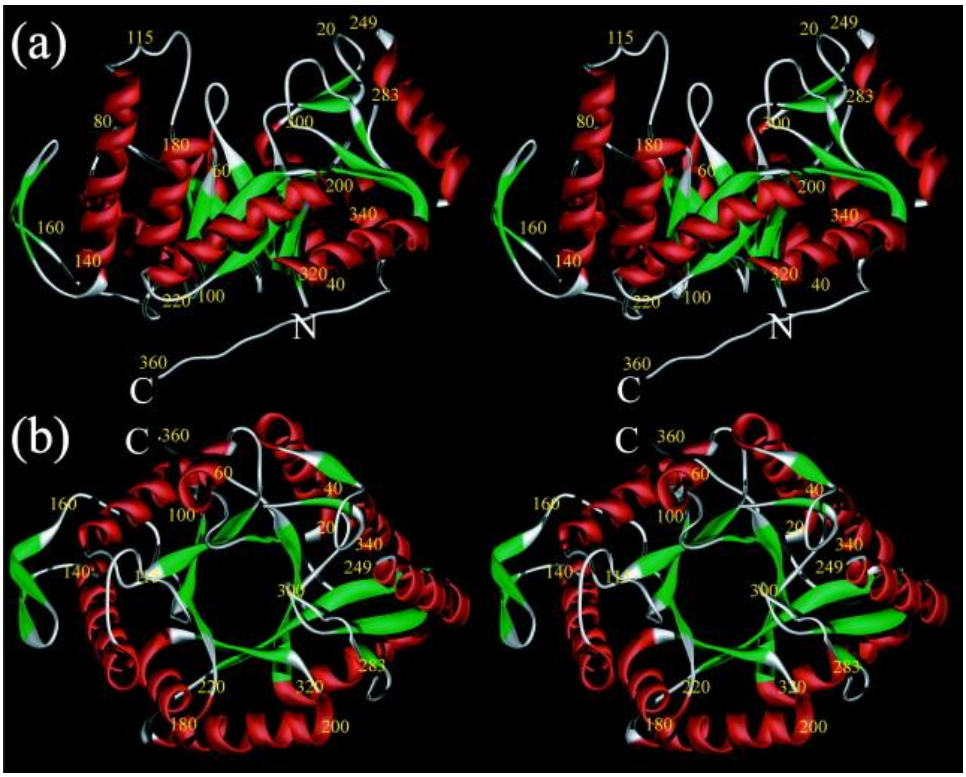


Figure 1.10 Structure of SsuD enzyme. View perpendicular to β-barrel (a) and view along β-barrel axis (b) (Figure taken from figure 3 of Eichhorn, E. et al *J. Mol. Biol.* 2000, 324, 457-68[47])

SsuD involves the oxygenoletic cleavage of carbon sulfur bonds and can catalyze a variety of alkanesulfonate substrates. Figure 1.11 shows the mechanism by which the reduced flavin can react with oxygen (I). This first step yields a neutral flavin radical and superoxide (II), which can form the C4a-peroxide (IV). C4a-peroxide is a nucleophile but upon prontonation it becomes an

electrophile and a C4a-hydroperoxide (V).[48] Once the peroxide is generated, it then undergoes a Baeyer-Villiger rearrangement to generate the aldehyde and sulfite products.

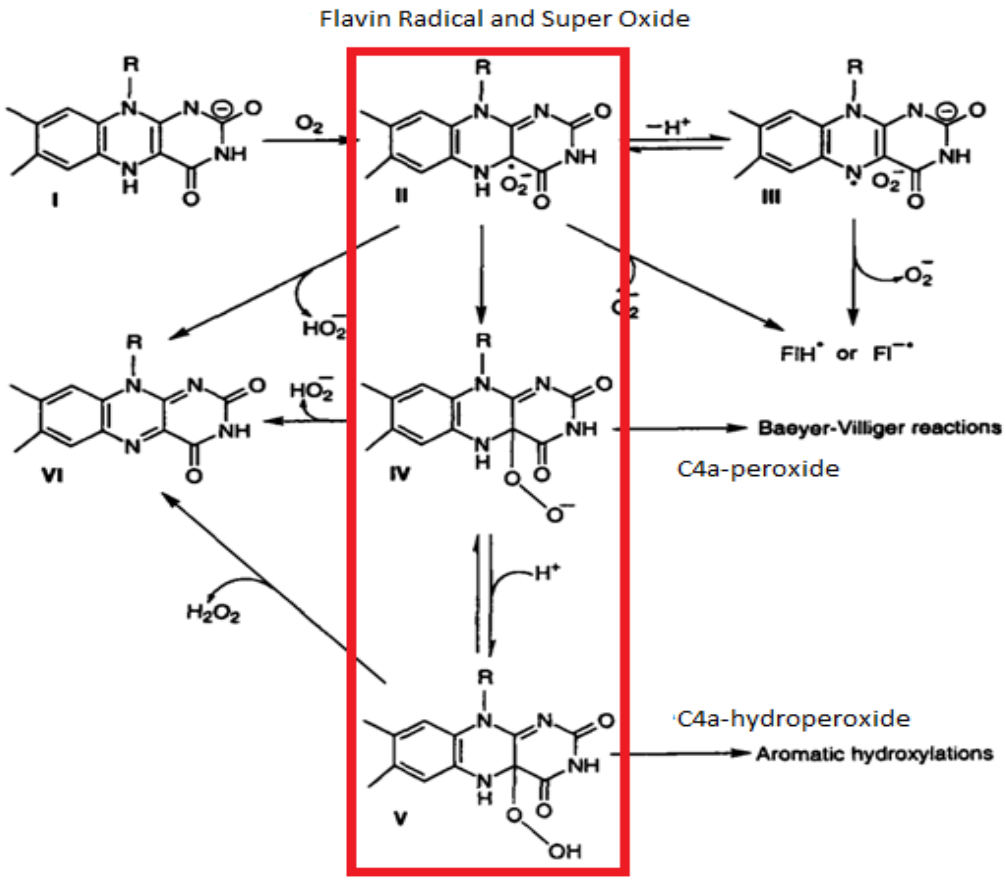


Figure 1.11 The possible reactions of reduced flavin with dioxygen. The route for generation of the C4a-hydroperoxyide II,IV, IV (Figure adapted from scheme 3 of Massey, V. Biochem Soc Trans. 2000, 28(4), 283-296 [48])

Once the peroxyflavin is generated, it is then able to form the aldehyde and sulfite products. There are two theories on how this happens. First, an active site base will abstract a carbon from the alkanesulfonate to produce a carbanion intermediate (IIA). The carbanion intermediate then makes a nucleophilic attack on the hydroperoxyflavin (IIIA). This generates an unstable 1-hydroxyalkanesulfonate (IVA) which then decomposes to the aldehyde and sulfite products (VA).

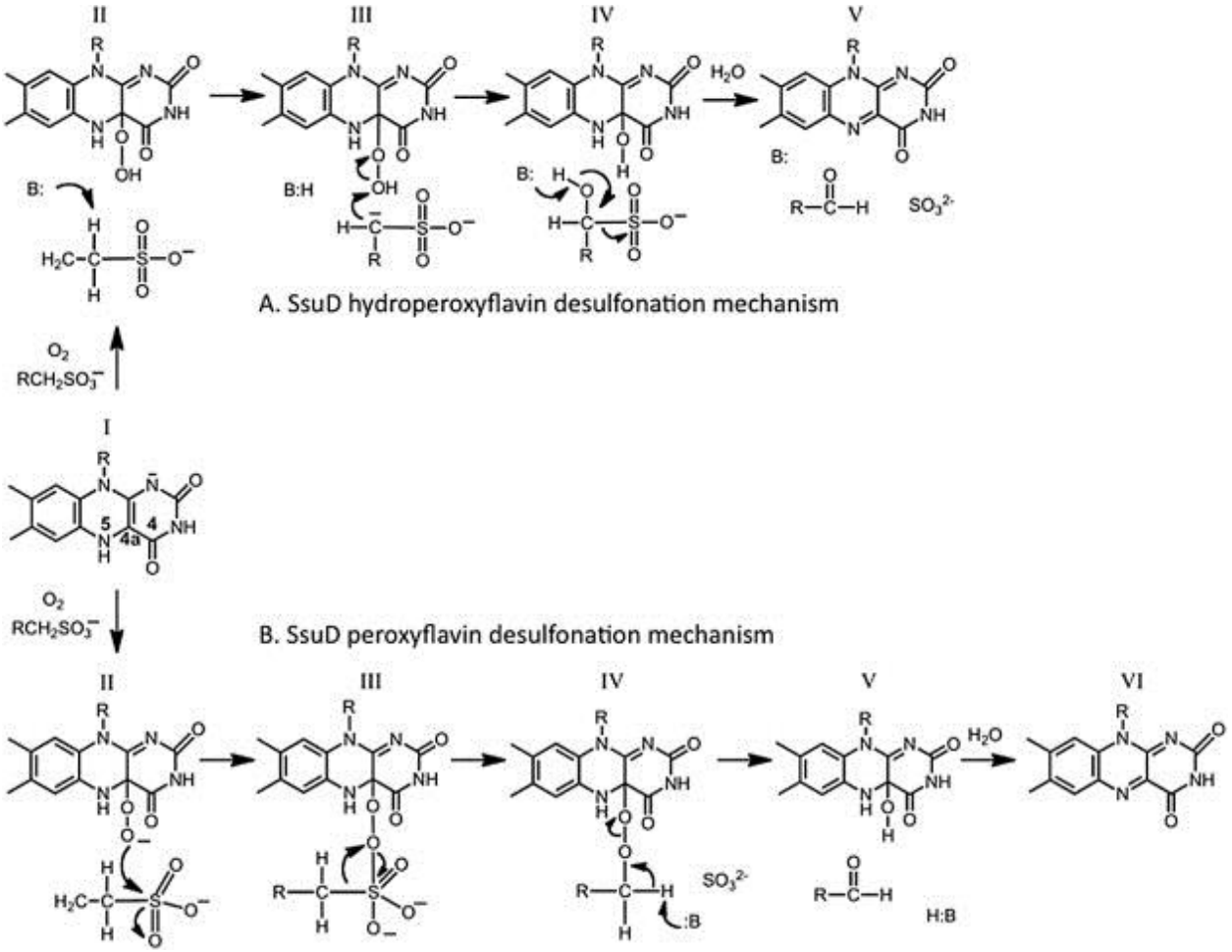


Figure 1.12 Two possible mechanism by which the sulfite and aldehyde products are generated in SsuD. (Taken from figure 7 of Ellis, H. 2011, Bioorganic Chemistry, 39(5-6), 178-184 [43])

The second theory and the most likely route starts with the peroxyflavin intermediate. The peroxyflavin attacks the sulfate on alkanesulfonate to form an organosulfate adduct (IIB). A Baeyer-Villiger rearrangement of the sulfite adduct then releases the sulfite product and a peroxyalkane (IIIB). An active site base then abstracts a hydrogen atom from the C1 alkane leading to heterolytic cleavage of the oxygen-oxygen bond of the alkane flavin adduct (IVB). This forms the aldehyde product and C4a-hydroxyflavin products (VB).[43]

The Role of Active Site Residues

SsuD has low sequence similarity with its family members; however, the structure and active site are similar to bacterial luciferase and LadA. Residues pertinent to catalysis for SsuD are proposed to be: His11, His228, His333, Tyr331, Arg226, Arg297, and Cys54. [46, 43, 49, 50, 47]

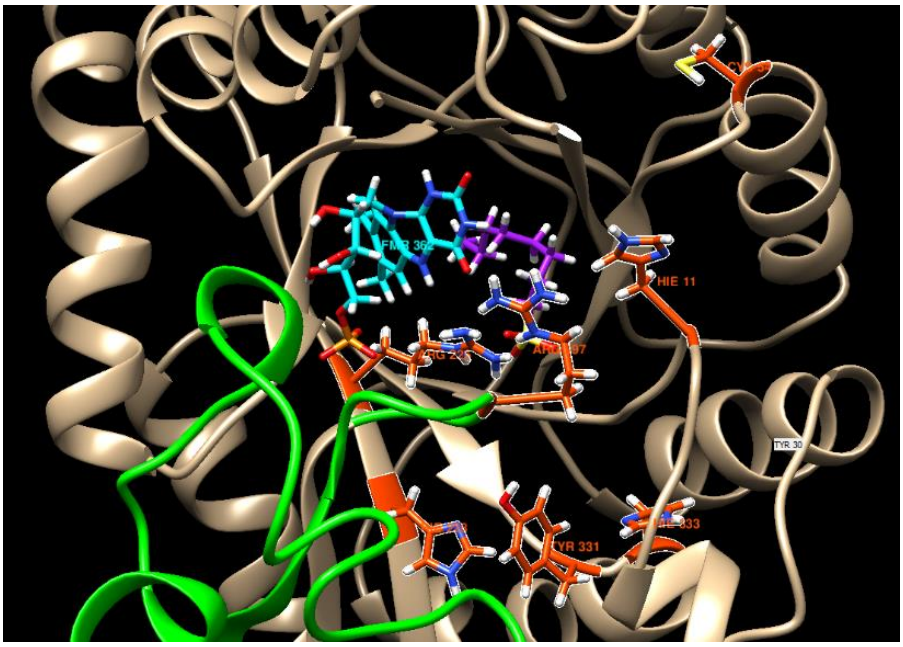


Figure 1.13 SsuD bound with reduced flavin and octanesulfonate showing active site residues His11, His228, His333, Tyr331, Arg226, Arg297, and Cys54.

Three histidines occupy the active site of SsuD, one of these histidines is thought to be the base in the reaction mechanism. The closer relative bacterial luciferase had a decrease in bioluminescence when the active site His was mutated, and because this residue is conserved across the family, it is likely it plays a similar role in SsuD.[51] The tyrosine may be involved in helping maintain active site conformation through hydrogen bonding. In LadA, mutation of this residue completely eliminated activity.[36] Arginine 226 is believed to be the acid in the active site by donating the proton for reprotonation of the hydroxyflavin intermediate which triggers opening

of the enzyme and product release.[52] Arginine 297 is located on the mobile loop and believed to anchor the phosphate on the peroxyflavin intermediate.[44] The cysteine residue initially was believed to take part in sulfur acquisition as it is the only cysteine residue in the enzyme and near the active site. Mutagenesis studies, however, showed that this cysteine was more likely to play a role in active site conformation and is not involved in flavin binding.[49]

The Mobile Loop

The dynamic region in monooxygenase ranges from residues 241-300. This loop is thought to be important in the conversion of reduced flavin (FMN) to oxidized flavin (FMNOO-). Monooxygenase contains a triose phosphate isomerase (TIM)-barrel fold which has conserved residues with SsuD homologs of different organisms. It is believed that this loop may protect reactive intermediates from premature release and bulk solvent.[53]

The Arg297 residue has been shown to be important to activity of SsuD. A molecular dynamics simulation of 20 ns showed a salt bridge between Arg297-Asp111/Glu20. The Arg297-Asp111 salt bridge was proposed to favor the closed conformation in order to close off substrate from bulk solvent. The other salt bridge between Arg297-Glu20 was proposed to occur during the open confirmation of SsuD.[54] However, a 300 ns molecular dynamics study showed that the Arg297-Asp 111 salt bridge formed less than 10% of the time when the reduced flavin was bound. Also, the salt bridge between Arg297-Glu20 was formed less than 5% of the time. The 300 ns simulation showed that the Arg297 salt bridges were not relative to change in conformation of SsuD; however, the Arg297 is still believed to play an important role in catalytic activity.[55]

Deletion studies were performed on the mobile loop of SsuD and showed that the sulfite products could not be generated without the loop; however, the studies also showed the deletion

of the loop did not affect flavin binding.[56] The conserved nature of the Arg297 residue and the location it has on the mobile loop suggest it plays a role in catalysis. Mutation studies showed that when this residue was changed, catalytic activity was lost.[44] The purpose of this study was to investigate the motility of the mobile loop and interactions of all residues with emphasis on the interaction of Arg297 with the substrates over a 1000 ns accelerated molecular dynamics simulation using Amber for the apo SsuD, FMN-octanesulfonate bound SsuD, and $FMNOO^{-}$ octanesulfonate bound SsuD. Chapter 4 elaborates on the results of these simulations.

Chapter 2 Methods

Molecular Docking with Small Molecules

Small molecule docking was done in this study on B-Raf V600E, B-Raf V600E-L505H, and B-Raf V600E-L505H(P) in order to obtain a starting conformation for non-crystalized inhibitors as well as gain an estimation on the free energy of binding. The results are discussed in later chapters. The purpose of this section is to provide an overview on small molecule docking.

Molecular docking predicts binding of macromolecules usually with a smaller molecule. Docking is used in drug discovery to help cut costs as it reduces product cost as well as personnel time involved in synthesis. Companies can sort through thousands of molecules efficiently and pick the best candidates to test experimentally.

Binding affinities can be obtained through higher level calculations which are more computationally expensive but make fewer approximations. The most rigorous method is free energy perturbation (FEP).[57] These calculations can show the relaxation in the protein/solvent system; however, it is far too computationally expensive to achieve convergence for large systems.[58] A more commonly used method is the molecular-mechanics-Poisson-Boltzmann/surface area calculation (MM-PBSA). These calculations use molecular mechanics energy and solvation energy with Poisson-Boltzmann.[59] Another method is the MM-GBSA which combines the Generalized Borne (GB) solvation energy.[60] These calculations use entropic estimations through quasiharmonic or normal mode analysis to predict binding energies.[61] Many docking programs also exist to calculate the predicted structure as well as potential binding affinity. These programs make approximations on solvation; however, they are fast and efficient for large numbers of molecules.[62] These programs include Surflex[63], AutoDock[64], Gold[65], and AutoDock Vina[66].

Docking attempts to reproduce chemical potentials. Typically, a scoring function is used to reproduce those potentials and determine the free energy of binding.

$$\Delta G = \Delta H - T\Delta S \quad (2.1)$$

$$\Delta G' = -RT\,(\ln C\ \frac{1}{k_a}) \quad (2.2)$$

The free energy is related to H, enthalpy, S, entropy. The binding affinity is related to the free energy of binding through the dissociation constant, k_a. R is the gas constant, C is the concentration, and T is temperature. In some calculations, entropy may be ignored to speed up the process.[67]

High resolution crystal structures allow for structure determined pharmaceutical design. Docking allows high throughput scanning of many potential drugs. Homology modeling can fill in small gaps that a crystal structure may have because of high motility in a region. Docking starts with an unbound molecule and ligand. The docked structure and binding affinity is then computed. In order to determine accuracy of the docking program, a crystal structure should be compared with the predicted structure if the calculations are lower level.

Docking with AutoDock Vina

Vina uses "machine learning" rather than a physics based scoring. The scoring function for AutoDock Vina starts with the conformational dependent part, which is the sum of the inter and intra molecular forces.

$$c = c_{inter} + c_{intra} \quad (2.3)$$

$$c = \sum_{i<j} f_{(ti)tj}(r_{ij}) \quad (2.4)$$

The conformational part is the sum over all atoms moving relatively to each other with the exception of atoms separated by three covalent bonds. Atom type i is assigned to t_i. The symmetric set of interaction function, $f_{(ti)tj}$, and interatomic distance, r_{ij}, are defined. From the intermolecular part from 2.4, the free energy of binding is predicted where g is a smooth function (2.5).

$$s = g(c_{inter1}) \tag{2.5}$$

The derivation of Vina's scoring function uses knowledge based potentials and empirical scoring functions. The empirical information is gathered from the conformational preference of the complex as well as experimental affinity measurements. The derivation of the scoring function has not been published but is related to the scoring function of Xscore.[68] In Vina's scoring function, hydrogen atoms are not considered explicitly and are omitted from equation 2.5. The interaction function, f_{titj}, is defined relative to the surface distance where R_t is the van der Waals radius of atom type t:

$$d_{ij} = r_{ij} - R_{ti} - R_{tj} \tag{2.6}$$

The interaction function, f_{titj}, and the interatomic distance, r_{ij}, is equal to the sum of the steric interactions, h_{titj}, times the surface distance, d_{ij}:

$$f_{titj}(r_{ij}) = h_{titj}(d_{ij}) \tag{2.7}$$

The scoring function includes weighting in the steric interaction sum function, h_{titj}. Metals in the vina scoring function are treated as hydrogen bond donors and the interaction function is cut off at r_{ij} = 8 Å.

The weighting of each term is as follows:

Table 1. Scoring function weights and terms for AutoDock Vina

Weights	Term
-0.0356	$gauss_1$
-0.00516	$gauss_2$
.840	Repulsion
-0.0351	Hydrophobic
-0.587	Hydrogen Bonding
0.0585	Number of Rotatable Bonds

The definition of the steric terms are defined:

$$gauss_1(d) = e^{-(d/.5\text{Å})^2} \tag{2.8}$$

$$gauss_2(d) = e^{-((d-3\text{Å})/2\text{Å})^2} \tag{2.9}$$

$$repulsion\ (d) = \begin{cases} d^2, if\ d < 0 \\ 0, if\ d \geq 0 \end{cases} \tag{2.10}$$

$$hydrophobic\ (d) = \begin{cases} 1, if\ d < .5\text{Å} \\ 0, if\ d > 1.5\text{Å} \end{cases} \tag{2.11}$$

$$hydrogen\ bonding\ (d) = \begin{cases} 1, if\ d < -.7\text{Å} \\ 0, if\ d > 0 \end{cases} \tag{2.12}$$

Equation 2.11 and 2.12 have linear interpolation in between the values.

The conformational independent function, g, was a chosen parameter where

$$g(c_{inter}) = \frac{c_{inter}}{1+wN_{rot}} \quad (2.13)$$

N_{rot} is the number of rotatable bonds for heavy atoms and w is the weight of the atoms.

The free energy of binding is determined as follows:

$$\Delta G_{binding} = \Delta G_{gauss} + \Delta G_{rep} + \Delta G_{hbond} + \Delta G_{Nrot} \quad (2.14)$$

The free energy of binding is useful for showing relative binding energies among large quantities of molecules tested; however, more rigorous methods of docking should be used in order to obtain binding energies that are more accurate.

AutoDock Vina developers opted to use an iterated local search global optimizer. This algorithm uses a stepwise approach including a mutation and local optimization. Each step is accepted using the metropolis criterion. The quasi-Newton Broyden-Fletcher-Goldforb-Shaano (BFGS) method was used as a local optimizer. This optimizer uses the derivatives of the scoring function with respect to orientation, position of the small molecule, and torsions of rotatable bonds. With the development of AutoDock Vina, multiple processors are able to be used in order to speed up calculations.[66]

Molecular Dynamics

Molecular dynamics (MD) is a useful tool in the study of biological systems. It can help merge the bridge of experiment and theory. MD allows for the motions of a molecule to be recorded in respect to time by using derivations of Newton's laws of motion. Alder and

Wainwright were the very first to use the theory back in the 1950's. They used the theory of MD to calculate equilibrium properties on hard spheres.[69] Rahaman continued to build on this theory and did simulations on liquid argon in the 1960's.[70] The first studies on proteins were done by McCammon in the 1970s's.[71] Since the 70's, the simulations for molecular dynamics has greatly improved, and it has become an extremely useful tool in the study of macromolecules.

There are two types of MD families: classical, and quantum. Classical MD treats atoms as classical objects, the ball and stick theory. Quantum MD accounts for quantum nature of chemical bond. The electron density of valence electrons is calculated quantumly; however, ions are treated as classical models. This method is too computationally expensive to use for most researchers.[72]

In this study, molecular dynamics simulations using Amber were done for both B-Raf and alkanesulfonate monooxygenase. In chapter 3, the molecular dynamics simulations were used in hopes to determine what type of inhibitor works for B-Raf V600E, B-Raf V600E-L505H, and B-Raf V600E-L505H(P). In chapter 4, molecular dynamics was used in order to determine the difference in the changes in conformation for alkanesulfonate monooxygenase. In this section, the theory of molecular dynamics and the Amber package are outlined.

Equations of Motion and Algorithms for Molecular Dynamics

Molecular dynamics shows the evolution and the state of a macromolecule, and this evolution is known as the trajectory. The trajectory can be obtained from Newton's equations.[73]

$$F = ma = m\frac{dv_i}{dt} \tag{2.15}$$

$$F_i = \frac{m_i d^2(r_i(t))}{dt^2} \; where \; r_i = \big(x_i(t)\big), (y_i(t)), (z_i(t)) \tag{2.16}$$

Equation 2.16 show the evolution of the particles in phase space where a, v, and r are acceleration, velocity, and position of the particles. MD trajectories are defined as position and velocity vectors, and the equations of motion are discretized and solved numerically. Equation 2.16, a second order differential equation, must be integrated so that the initial forces, velocities, and positions are calculated. The velocities are driven by a finite time interval, which is known as the Verlet algorithm.[74] The velocities of the system can define the kinetic energy and the temperature of the system. The positions are expressed as $r_i(t + \Delta t)$ at time $t + \Delta t$ in positions of already known time. The time step, Δt, should be chosen to be small enough to read all periods of vibrations. Large time steps may not read all vibrations and could cause the system to collapse.[72]

The Verlet algorithm is derived from the Taylor expansion of $r_i(t)$ and is commonly used in molecular dynamic simulations because of its simplicity:

$$r_i(t + \Delta t) \cong 2r_i(t) - r_i(t - \Delta t) + \Delta t^2 a(t) \qquad (2.17)$$

The computational expense of the Verlet algorithm increases with the number of particles squared.[75] Velocities, however, do not appear in this algorithm. This algorithm uses the position at time (t) and position from the previous step to calculate the new position. The Verlet algorithm may have precision problems because a small term is added to a large term, and velocities are difficult to obtain and are not expressed explicitly. To help with this, the velocity-Verlet algorithm was developed.[76] Equations 2.18 and 2.19 show this algorithm calculates positions, velocities, and acceleration at the same time.

$$r(t + \Delta t) = r(t) + \Delta t v(t + \tfrac{1}{2}\Delta t) \quad (2.18)$$

$$v(t + \Delta t) = v(t) + \tfrac{1}{2}\Delta t[a(t) + a(t + \Delta t)] \quad (2.19)$$

This algorithm is implemented in three stages. Equation 2.20 shows that the acceleration must first be calculated at $t + \Delta t$, and then the velocity must be calculated at $t + \frac{1}{2}\Delta t$.

$$v\left(t + \tfrac{1}{2}\Delta t\right) = v(t) + \tfrac{1}{2}\Delta t a(t) \quad (2.20)$$

The new forces calculated from the current positions at $t + \Delta t$ are then determined as in equation 2.21:

$$v(t + \Delta t) = v\left(t + \tfrac{1}{2}\Delta t\right) + \tfrac{1}{2}\Delta t a(t + \Delta t) \quad (2.21)$$

This algorithm should be faster and more efficient than the Verlet algorithm. It should also conserve energy, momenta, be time reversible, and be able to handle a larger time step.[77]

A variation of the Verlet algorithm is the Leap Frog algorithm. This algorithm includes velocities as shown in equations 2.22 and 2.23

$$v\left(t + \tfrac{1}{2}\Delta t\right) = v\left(t - \tfrac{1}{2}\Delta t\right) + a(t)\Delta t \quad (2.22)$$

$$r(t + \Delta t) = r(t) + \Delta t v(t + \tfrac{1}{2}\Delta t) \quad (2.23)$$

In this algorithm, velocities are first calculated and leap frog over the positions to give values at $t + \frac{1}{2}\Delta t$. The positions then leap over the velocities to give values at $t + \Delta t$.[77]

The SHAKE algorithm is another algorithm derived from the Verlet algorithm.[78] The equations of motions are solved while simultaneously satisfying holomeric constraints which can be expressed as:

$$f(q_1, q_2, q_3, \dots t) = 0, q - represent\ coordinate\ of\ particle \qquad (2.24)$$

If the system has 3N number of particles and k constraints, the number of degrees of freedom the system has is 3N-k. When a system is constrained, the forces of the system are manipulated. Equation 2.25 represents the forces of a system:

$$F_{ckx} = \frac{\lambda_k \partial \sigma_k}{\partial x} \qquad (2.25)$$

λ_k – LaGrange multiplier
x – the constraint

σ_k is the constraint and x is the cartesian coordinate of the atom i or j. If an atom is involved in more than one constraint, then total force is sum of all terms, represented by equation 2.26.

$$\sigma_{ij} = (r_i - r_j)^2 - {d^2}_{ij} = 0 \qquad (2.26)$$

The constraint force is on the bond at all times, and there is an equal and opposite force on the atoms that make up the bond. Overall, the constraint forces do no work on the system. The constraint forces are then factored into the Verlet algorithm as in equation 2.27.[77]

$$r_i(t+\Delta t) = 2r_i(t) - r_i(t-\Delta t) + \frac{\Delta t^2}{m_i}F_i(t) + \sum_k \frac{\lambda_k \Delta t^2}{m_i} r_{ij}(t) \qquad (2.27)$$

In Amber, the SHAKE algorithm is implemented to impose constraints and speed up calculations by reducing the degrees of freedom.

Ensemble Averages

In simulations of molecular systems, macroscopic properties are calculated through microscopic interactions and statistical averaging connects the two. The property of the system depends on the position and momentum of the number of particles (N) inside the system that is given instantaneous value of "A."

An ensemble is a collection of all the possible microscopic states a system can have with the same macroscopic and thermodynamic properties. "A" will fluctuate because of interactions between particles. As time approaches infinity, "A" will approach an average:

$$A_{avg} = \lim_{\tau\to\infty} \frac{1}{\tau} \int_0^\tau A(p^N(t), r^N(t))dt \qquad (2.28)$$

The force on each atom because of interactions with other atoms can be calculated by differentiating the energy function. Boltzmann and Gibbs developed statistical mechanics so that a system involved with time can be replaced by number of replicates which are considered simultaneously.[77]

$$< A >= \iint dp^N dr^N A(p^N, r^N)\rho(p^N, r^N) \quad (2.29)$$

< > = ensemble average value

$\iint$ = should be 6N integral signs for 6N positions and momenta

$\rho(p^N r^N)$=probability density; finding momenta and position of particle N

$A(p^N r^N)$ = property of interest

The importance of ensembles in molecular dynamic simulations is that the simulations should be performed long enough and with the correct parameters to sample enough states. Different variables will have different ensembles with constraints usually involving (N) number of particles, (P) pressure, (V) volume, (T) temperature, (E) energy, and (μ) chemical potential. Types of ensembles include: the microconical ensemble (NVE), the canonical ensemble (NVT), Isobaric-Isothermal ensemble (NPT), and the grand canonical ensemble (μVT). The NVE ensemble shows an isolated system with fixed energy, volume, and particles. The NVT ensemble has its thermodynamics state characterized by constant number of particles, volume, and temperature. The NPT ensemble has fixed number of particles, pressure and temperature and the μVT has set volume, temperature, and chemical potential.

Periodic Boundary Conditions

Molecular dynamic simulations require solvent particles to encase the system; however, a large number of solvent particles can slow down a simulation significantly. Boundary conditions allow for a small number of solvent particles to be used while still being able to calculate the macroscopic properties of the system. An obvious problem with a boundary is that outside of the

boundary is a vacuum; however, periodic boundaries were designed so that the boundary is an array. As a particle exits through one side, it enters again from the other side.

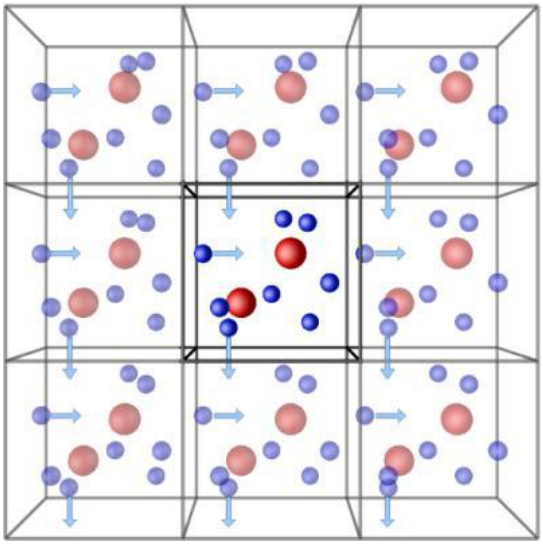


Figure 2.1 Periodic boundary conditions in 2D (Taken from figure 1 of Model Box Periodic Boundary Conditions – P.B.C. at Central Michigan University http://isaacs.sourceforge.net/phys/pbc.html [79])

The above figure shows a 2D cubic box array. The coordinates of each solvent particle are found by adding and subtracting the integral multiples of the box sides.[80] The cubic box is the simplest shape of the types of boundaries, but others include hexagonal prism, truncated octahedron, and elongated dodecahedron. The goal of the periodic boundary is to use as few solvent particles as possible that still simulate the system with accuracy and also make sure that long range interactions are included; however, some simulations are not appropriate for an array such as surface adsorption calculations.[77]

Long Range Interactions

Long range forces are vital to an accurate simulation. Interactions that decay no faster than, r^{-1}, where n is the dimensionality of the system, are problematic because the interactions may be greater than half of the box length. Ewald summation is used to calculate the long range interactions and was first used in 1921 to calculate the energetics of ionic crystals.[81] A particle can interact with all other particles in the simulation and all of their images forming an infinite periodic array.

For simplicity the position of the box is assumed to be a cube of side L and contains N charges as shown in figure 2.2 A vector is then specified of which components are the integral multiple lengths of the box (±iL, ±jL, ±kL; i, j, k = 0,1,2,3…)

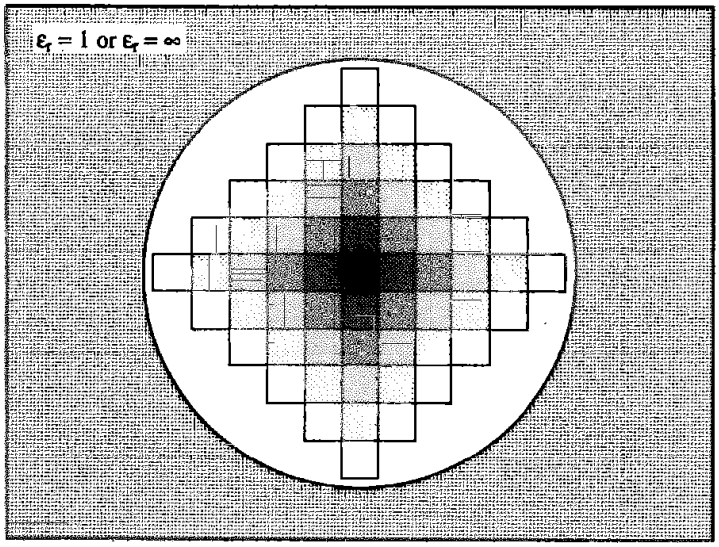


Figure 2.2 Construction of array simulation cells. (Taken from Allen and Tilsedley, 1987, Computer Simulations in Liquids .Oxford,Oxford University Press[82])

The charge-charge contribution to the potential energy is described as:

$$V = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{4\pi\varepsilon_o r_{ij}} \quad (2.30)$$

where r^{ij} is the minimum distance between the charges i and j. Six boxes with a distance L are around the central box with coordinates r_{box} given by (0,0,L), (0,0,-L), (0,L,0), (0,-L,0), (L,0,0), and (-L,0,0). Because figure 2.2 is only in 2D, only four of the boxes are observed. The charge-charge contribution between the central box and the particles in the surrounding box is shown by:

$$V = \frac{1}{2}\sum_{nbox}^{6} = 1\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{4\pi\varepsilon_o(r_0+r_{box})} \quad (2.31)$$

For a cubic box with the lattice point n (=n_x,L, n_yL, n_zL) with n_x, n_y, n_z bein integers, the expression is written as:

$$V = \frac{1}{2}\sum'_{|n|=0}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{4\pi\varepsilon_o(r_{ij}+r_n)} \quad (2.32)$$

In equation 2.32, the interaction does not include interaction i=j for n=0 as indicated by the prime on the summation. The above equation shows interactions with the central box and all images of the surrounding boxes. The disadvantage to the Ewald sum is that it converges extremely slowly and is actually conditionally convergent.[77]

Amber uses a variation of the Ewald method known as Particle Mesh Ewald (PME) and uses a grid based system to replace the point charges.[83] PME modifies the Ewald reciprocal sum to near linear scaling using a three dimensional fast Fourier

transform (3DFFT).[84] PME solves the convergence issue as it converges to zero rapidly.

Accelerated Molecular Dynamics

Standard molecular dynamics may not be efficient enough to cover enough conformational sampling. It allows access to tens to hundreds of nanoseconds which may not be enough time to properly sample biological systems. Accelerated molecular dynamics (aMD) can allow a system to sample beyond the range of standard MD. Two advantages to aMD are that knowledge of the potential energy surface and reaction coordinates do not need to be known. aMD can also be run on GPU's like MD allowing for faster calculations than on CPU versions.[85]

Molecular dynamics samples transitions between potential energy barriers; however, sometimes the simulation gets caught in a well. Accelerated molecular dynamics modifies potential energy landscape by raising the minima that lie below a defined point. The area above that point stays unmodified. It alters the potential energy landscape while leaving the potential energy surface unchanged. This allows enhanced sampling of large biomolecules.

$$V^*(r) = \begin{cases} V(r), V(r) \geq E \\ V(r) + \Delta V(r), V(r) < E \end{cases} \quad (2.33)$$

To do this, a continuous non-negative bias boos potential ($\Delta V(r)$) is applied. When the true potential V(r) is below a certain value of E (boost energy), the simulation is performed on the modified potential $V^*(r) = V(r) + \Delta V(r)$. If V(r) is greater than E, the simulation is performed on the true potential $V^*(r) = V(r)$. This leads to an enhanced escape rate from potential wells as in figure 2.81.

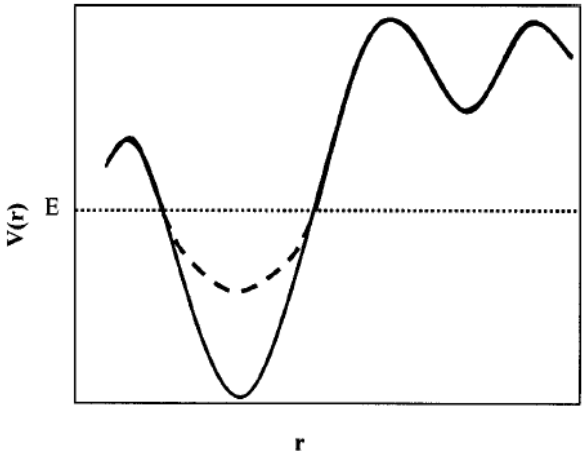


Figure 2.3 Representation of the biased boost potential, normal potential, and the boost energy (Figure taken from figure 1 of Hamelberg et al. J Chem Phys. 2004, 120(24), 11919–11929 [86])

The bias potential helps the system escape from potential wells which increases the rate of evolution of the system with a nonlinear time scale Δt^*.

$$\Delta t^* = \Delta t e^{\beta \Delta V[r(t_i)]} \tag{2.34}$$

Equation 2.34 helps advance the clock based on the potential, and Δt is the time step at the unmodified potential. The total simulation time actually becomes a statistical property resembling the canonical ensemble.

$$t^* = \sum_i^N \Delta t^* = \Delta t \sum_i^N e^{\beta \Delta V[r(t_i)]} \quad (2.35)$$

$$t^* = t e^{\beta \Delta V[r(t_i)]} \quad (2.36)$$

N = number of MD steps carried out during simulation

$e^{\beta \Delta V[r(t_i)]}$ = extent to which simulation has been accelerated

Δt^* is nonlinearly dependent on V(r) and $\Delta t^* = \Delta t$ when system is at the true potential (V(r) = 0). The system must show the canonical averages of an observable A(r) so that thermodynamic and other properties can be determined. The following equation represents the equilibrium ensemble average of observable A(r) at a normal potential V(r):

$$< A > = \frac{\int dr A(r) e^{-\beta V(r)}}{\int dr e^{-\beta V(r)}} \quad (2.37)$$

Equation 2.37 can then be converted to an average on the modified potential:

$$< A > = \frac{\int dr A(r) e^{-\beta V^*(r)}}{\int dr e^{-\beta V^*(r)}} \quad (2.38)$$

V*(r) is then substituted for V(r) + ΔV(r)

$$< A > = \frac{\int dr A(r) e^{-\beta V(r) - \beta \Delta V(r)}}{\int dr e^{-\beta V(r) - \beta V(r)}} \quad (2.39)$$

The phase space is then reweighted by multiplying the configurations by the bias at each position. This yields a corrected ensemble average with equilibrium observable on normal potential.

$$< A^C >= \frac{\int drA(r)e^{-\beta V(r)-\beta\Delta V(r)}e^{\beta\Delta V(r)}}{\int dre^{-\beta V(r)-\beta V(r)}e^{\beta\Delta V(r)}} \quad (2.40)$$

$$< A^C >= \frac{\int drA(r)e^{-\beta V(r)}}{\int dre^{-\beta V(r)}} = < A > \quad (2.41)$$

This shows that aMD does converge on the canonical ensemble and each point is reweighted by the Boltzmann factor of the bias energy:$e^{\beta\Delta V[r(t_i)]}$.

$$\Delta V(r) = \frac{(E-V(r))^2}{\alpha+(E-V(r))^2} \quad (2.42)$$

The selection of the values should follow these basic ideas. E should be greater than the minimum of the potential because if E is less than the minimum of the potential, the simulation would be standard MD. The variable, α, is how deep the modified potential should be, but it is not critical at low values of E. At higher values, $\alpha = E\text{-}V_{min.}$[86]

Amber, the program used for the aMD calculations offers simple equations to help calculate the parameters for the aMD simulation. The manual extends this equation further to include torsion potentials and dihedrals.[87]

$$\Delta V(r) = \frac{(Ep-V(r))^2}{\alpha P+(Ep-V(r))^2} + \frac{(Ed-Vd(r))^2}{\alpha D+(Ed-Vd(r))^2} \quad (2.43)$$

V(r) – normal torsion potential
Vd(r) – normal torsion potential
Ep – average potential
Ed – dihedral energies. Note Ep and Ed serve as reference energy to compare present position of calculation and therefore the amount of boost to be applied

αP and αD – strength which the boost is applied

All values needed for aMD can be found in a short standard MD simulation .out file.

Monte Carlo

Molecular dynamics samples conformations based on time of simulation. Monte Carlo samples based only on predecessor in previous step. It generates the conformations randomly and then uses set of specific criteria that determine whether or not the new conformation will be accepted or rejected.

The probability of finding a given configuration is equal to the Boltzmann factor:

$$e^{-\frac{V(r^N)}{k_B T}} \tag{2.44}$$

$V(r)^N$ is the potential energy function. This equation allows for lower energy configurations to be generated with a greater probability than higher energy functions. If the configuration is accepted, the values of desired properties are calculated and then averaged at the end of the simulation.

$$< A > = \frac{1}{M}\sum_{i=1}^{M} A(r^N), M = number\ of\ values\ calulated \tag{2.45}$$

The above calculation is known as Metropolis Monte Carlo. A new conformation may be generated by moving a single atom or molecule. The energy is calculated through a potential energy function. If the energy is lower, the move is accepted. If the energy is higher, then the Boltzmann factor of energy difference is calculated.

$$\frac{V_{new}(r^N)-V_{old}(r^N)}{k_B T} \quad (2.46)$$

A random number is then generated between 0 and 1. If the calculated number is higher, the move is rejected and if it is lower, the move is accepted. This is known as the metropolis criterion.[77, 88]

Constant pH with discrete protonation state model

The pH of a solvent solution can greatly affect the protein structure and function. Difficulty arises in calculating protonation states at certain pH in proteins because of multiple protonation sites, and the explicit pH models can be very computationally expensive.[89]

The protonation state of a titratable residue is determined by that residues pKa which can vary depending on the position of the amino acid and its neighbors. Standard molecular dynamic simulations use a predetermined protonation states for titratable residues based on the general pKa value. These states are constant throughout the simulation. Constant pH methods allows the pH to be set and titratable residues to be protonated or unprotonated based on their pKa.

There are several CpHMD methods. One method is a continuous protonation state model that uses protonation states as a continuous titration parameter.[90] This advances along with the atomic coordinates of the system. It uses a mean-field approximation [91] and does not consider interactions with nearby residues. This model performs badly for tightly coupled residues. The next type is known as λ-dynamics [92] which uses unphysical fractional protonation states. It has an artificial titration barrier along the coordinate between fully protonated and unprotonated states.[93] Most simulations now use a discrete protonation state models that avoid any nonphysical intermediate charge states.[94] This uses MD for conformational sampling while using Monte Carlo for periodic sampling of the protonation states.

A short MD trajectory employs a time dependent potential energy that interpolates between the old and new protonation state. MC chooses the protonation state and residue at random. The change in free energy is used as the Metropolis Criterion.

$$\Delta G = k_B T\left(pH - pKa_{ref}\right) ln10 + \Delta G_{elec} - \Delta G_{elec,ref} \quad (2.47)$$

k_B – Boltzmann Constant
T – temperature
pH – specific solvent pH
pKa_{ref} – pKa reference
ΔG_{elec} – electrostatic energy of residue
$\Delta G_{elec,ref}$ – electrostatic compenent for the transition energy for reference compound

If the MC move is accepted, the protonation state of the simulation is changed in MD continues, and if the move is rejected, the protonation states remains the same and MD continues.

The biggest limit to constant pH calculations are convergence problems; however, constant pH can also be applied with aMD with the same principals as standard aMD calculations. The CpHaMD will still employ an implicit solvent model so that calculations are not too expensive. The disadvantage to the CpH methods at the time the calculations were performed for either MD or aMD is that it is currently only available on CPU's which makes simulation times much more expensive. [95] Newer versions of Amber have now developed CpH methods for GPU's.

The AMBER Bimolecular Simulation Package

The simulation program used for all MD calculations was AMBER (assisted model building with energy refinement). It was originally developed by the Kollman lab in 1978.[96] It has been cited almost 9000 times and is the collaboration of over 40 scientific researchers.[97] This

program allows for simulations to be carried out on a variety of components such as nucleic acids, proteins, and carbohydrates.

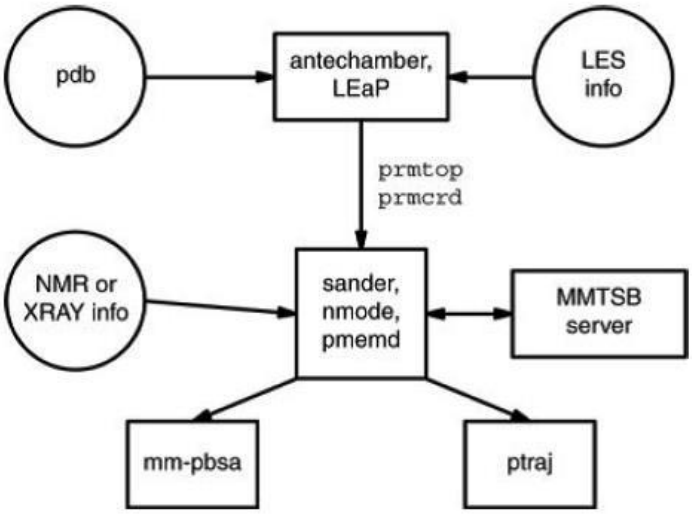


Figure 2.4 Information flow in the Amber program suite (Taken from figure 1 of Case et al. *Journal of computational chemistry*, 2005, 26(16), 1668-1688 [97])

There are three main steps in Amber: sample preparation, simulation, and the trajectory analysis. Amber is not a single program but a collection of programs that are coded and designed to work together. There are three main advantages to having these programs coded separately: 1) individual parts of the Amber package can be upgraded and replaced with minimal impact on other portions of the package, 2) programs can be written in different types of code such as C code or Pearl, 3) the separation of codes allows the porting to new platforms to be done more easily and only principal codes such as sander and pmemd are coded for parallel jobs. Some of the less intensive calculations such as preparation of the unit can be carried out on a desktop while other more intensive processes must be carried out on remote computers. The disadvantage to Amber is it can be difficult to learn and code cannot be easily modified from one section to the next.[84]

Sample preparation starts usually with an input file such as a pdb file. Antechamber constructs the force fields for residues or molecules not part of the Amber library and LEaP

constructs the biopolymers from residues, solvates the system, and prepares the terms for force fields and their related parameters. The prmtop or prmcrd file is generated and includes all information about the coordinates.

Simulation is the second part to the Amber suite. The main program run in Amber is Sander which was written in Fortran 90. Sander offers many variables for the simulation but 32 are most commonly used and can be run in parallel. The way the parallel jobs work are the each processor has its own set of atoms, but all processors know the coordinates. At each step, the potential energy and the gradient is calculated and from this the forces are calculated. Each processor will then get the full force of their specified atoms. The pmemd program is an upgraded version of sander and allows each processor to know only pieces of the coordinates.[84] The biggest advantage to pmemd is it can be run on GPU greatly accelerating the speed of calculations [97]

The third part to the Amber suit is analysis. Ptraj is the analysis portion for the MD trajectories. It was developed to help tackle the large MD trajectories that were produced from the simulation. It can perform cluster, hydrogen bonding, distances, rmsd's, and more.[84]

The General Amber Force Field

Force fields are vital tools in investigating the protein-ligand structure. The Amber force fields were originally developed for proteins and nucleic acids; however, a need for a force field that can easily give parameters to organic molecules was also needed. The general Amber force field (GAFF) can describe most organic molecules.

$$E_{pair} = \sum_{bonds} k_r\left(r - r_{eq}\right)^2 + \sum_{angles} k_\theta\left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} \frac{V_n}{2}\left[1 + \cos(n\phi - \gamma)\right] +$$

$$\sum_{i<j} \frac{A_{ij}}{{R_{ij}}^{12}} - \frac{B_{ij}}{{R_{ij}}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}} \qquad (2.48)$$

r_{eq}, θ_{eq} = equilibrium constants
k_r, k_θ, ν_n = force constants
n = multiplicity
γ=phase angle for torsional parameters
A,B,q – characteristics of non-bonded potentials

The parameters in a force field must replicate experiments. The non-bonded part of the GAFF comes directly from the Amber force fields. Potential charges are assigned using restrained electrostatic potentials (RESP). [98] The first three terms of equation 2.48 are internal terms calculated from bond lengths and bond angles. These terms come from experiment and high level ab initio calculations. Force constants are estimated from hard parameters such as vibrational frequencies which can also be calculated through high level calculations. The torsional angles are coupled to the non-bonded energy and is parametrized last. The GAFF has 35 basic atom types: 5C, 8N, 3O, 5S, 4P, 6H, and one type for each common halogen.

The default charge method for the GAFF is RESP at HF/631*G. This method is computationally expensive but does allow for fewer torsional terms. The disadvantage of this method is it does not handle many molecules well.

The alternative to the RESP at HF-631*G is AM1-BCC (bond charge correction).[99] This method first carries out a semiempircal AM1 calculation and then corrects the charges with BCC. This method is compatible with RESP and matches the electrostatic potential of HF/6-31G*. The advantage of this method is it can carry out calculations efficiently for large number of molecules.[100]

Principal Component Analysis

Multivariate data analysis is a statistical technique that takes into consideration two or more random variables and sees them as a single entity. It produces a result that takes all of the variables into account and shows observables that may not be shown in other types of statistical methods. Principal component analysis (PCA) describes the structure of the multivariate data.[101]

The main objective of PCA is to minimize the dimensionality of a data set.[102] It consists of linear combinations of original variables with the first PC having the greatest variance, or difference. The idea is that enough principal components are chosen so that the variance is greatly reduced. PCA projects the data, which is defined by the orthogonal components, into space.[103]

PCA is a method based on matrix algebra. A pxp symmetric nonsingular matrix such as covariance matrix S may be reduced to L. This occurs by post-multiplying and pre-multiplying it by an orthonormal matrix U.

$$U'SU = L \tag{2.49}$$

The diagonal elements, $l_1, l_2, \ldots, l_p$, are known as the characteristic roots which are the eigenvalues of S. The columns of U are the eigenvectors of S. The eigenvalues are defined:

$$|S - \ell I| = 0 \tag{2.50}$$

I is the identity matrix and equation 2.50 produces the p^{th} degree polynomial in ℓ where $\ell_1, \ell_2 \ldots . \ell_p$ are obtained. This rotates the original axis of x_1 and x_2.

The starting point for PCA is a covariance matrix S.

$$S = \begin{bmatrix} S_1^2 & S_{1\,2} & S_{1\,p} \\ S_{1\,2} & S_2^2 & S_{2\,p} \\ S_{1\,p} & S_{2\,p} & S_p^2 \end{bmatrix} \quad (2.51)$$

S_1^2 is the variance of the i^{th} variable. x_i and s_{ij} are the covariance between the i^{th} and j^{th} variables. If the covariance is not equal to zero, there is a linear relationship between the variables with the strength represented as

$$r_{ij} = \frac{s_{ij}}{s_i s_j} \quad (2.52)$$

Axis transformation transforms x_1, x_2, …x_p into uncorrelated variables corresponding to z. The axes of these variables which are described as u_i makeup the matrix U.

$$z = U'[x - \bar{x}] \quad (2.53)$$

$x - \bar{x}$ are the pxl vectors of observation of the original variable and their averages. These transformed variables are called principal components, which can be plotted against each other to show data correlation.[101] PCA analysis was used while studying conformational changes in the alkanesulfonate monooxygenase mobile loop. The covariance matrix was calculated through ptraj [104] in Amber14 and the principal components were calculated using Excelstat 2013.

Chapter 3 The Novel L505H Mutation for B-Raf V600E and the Effect on Drug Resistance

Introduction

The purpose of this study was to investigate the conformational changes in B-Raf V600E and B-Raf V600E-L505H when bound with type I and type II inhibitors. Constant pH molecular dynamic simulations were carried out to predict the pKa of His 505 in order to predict its protonation state in the acidic tumor microenvironment.[105] Accelerated molecular dynamic (aMD) simulations were carried out for 500 ns on B-Raf V600E, V600E-L505H (neutral histidine), and B-Raf V600E-L505H(P) (protonated histidine). The aMD method was used to enhance sampling in order to overcome potential energy barriers insurmountable by standard MD. A total of eleven inhibitors were examined in these simulations. Extensive analysis of each system including clustering, atom distances, RMSD, RMSF, and hydrogen bonding were carried out for all inhibitors bound with B-Raf variants.

Experimental Procedures

Modeling: The initial Cartesian coordinates for B-Raf V600E were obtained from the protein databank with pdb ID of 30G7. The crystal structure was missing amino acids 545-547, 597-614, and 627-630. They were modeled in using Modeller 9.10.[106] This program helps determine the sequence based on alignment with known proteins. The protein used to help determine these amino acids was CTR1 kinase (PDB ID: 3PPZ). The V600E-L505H and V600E-L505H(P) were made using Chimera.[107]

Autodock Vina: Inhibitors that did not have a crystal structure for the B-Raf V600E were docked using Autodock Vina, and all inhibitors for B-Raf V600E-L505H and B-Raf V600E-L505H(P) were docked using Autodock Vina.[66] The pdb 30G7 was selected from the protein databank. All waters were removed from the pdb. Using Autodock Tools, all nonpolar hydrogens were merged and Gasteiger charges were added. Flexible residues pertinent to catalysis were selected inside the grid box. The exhaustiveness of the global search was set to 100 and the number of binding modes was set to 20. The maximum energy difference between the best and worst binding modes was set to 5. The binding energies of receptor and enzyme were calculated. The best pose was then used for the molecular dynamics simulations.

Molecular Dynamics: Using Antechamber, charges were added to the inhibitors using AM1-BCC. [87, 108] Using Leap in Amber14, the ff9SB force field was used for standard amino acids and the Generalized Amber force field (GAFF) for nonstandard amino acids. [109,100] Hydrogens were then added to the enzyme-substrate complex, and the complex was solvated using TIP3P water molecules extending 10 Å.[110] Chlorine atoms were added to set the overall system charge to neutral.

Each system was minimized using conjugate gradients (CG). The first step was to use 200 steps of CG to minimize waters only. [111] Some starting structures had bad contacts and would not minimize with CG alone; therefore, 50 steps of steepest decent was also used to remove the bad contacts in the water only minimization. The next minimization step was 10,000 steps of CG to further remove the bad contacts.

The complex then needed to be equilibrated and was gradually heated from 0 K to 300 K for 50 ps of MD using a weak coupling algorithm and a temperature coupling value of 2.8 ps.

Then, the NPT ensemble was used at a constant 300 K and 1atm with a coupling value of 2.0 ps for temperature and pressure. This was run for 500 ps. The system was then switched back to the NVT ensemble and again run for 500 ps. The system was then run for 10ns of production on the GPU-accelerated version of Amber14.[112] Each simulation had a time step of 1 fs and a cutoff distance of 12 Å.

Accelerated Molecular Dynamics: Accelerated molecular dynamics (aMD) was run for all inhibitors and all variants of B-Raf. The previous 10 ns MD simulations were used to obtain potential energies and dihedrals. The parameters calculated for aMD were based on the values recommended for Amber.[86] Once the parameters were obtained, each inhibitor was run for 500 ns of aMD.[85-86] Calculations were carried out on computers located at Auburn University and on the Alabama Supercomputer.

Constant pH at aMD: Accelerated molecular dynamics at constant pH were performed.[89] The apo structure of B-Raf V600E-L505H was loaded into Leap in Amber12. The His505 residue was modified to HIP so it would become the titratable residue. An implicit solvent model was employed. The minimization was carried out for 1000steps and the salt concentration was set to 0.1M. There was no cutoff and the backbone atoms were restrained. The constant pH simulation was turned on; however, protonation states remained constant. The system was then heated gradually to 300 K and the backbone was again restrained. The cutoff was set to 30 Å. The heating of the system was run for 200 ps. The salt concentration was again set to 0.1M and the protonation states remained constant. Next was the equilibration phase. This was run at a pH of 4.5 because the tumor microenvironment is known to be acidic.[105] The protonation state was set to change

every 5 steps; i.e. swapped every 100 fs. The cutoff again was set to 30 Å and the salt concentration was set to 0.1M.

The production phase began next. Simulations at the assigned pH were first carried out at for 1ns at standard molecular dynamics to obtain the potential energy and dihedrals. Each aMD simulation was then run at constant pH with a 0.001 fs time step, and protonation states were changed every 5 steps. The temperature coupling parameter was set to 2 ps and the salt concentration was set to 0.1M. The pH simulations were run from pH 4-9 until the protonation state reached convergence.

Results and Discussion

Docking

Docking calculations were carried out for the V600E, V600E-L505H (neutral and protonated His) for the B-Raf mutants. The first step of this study was to determine if Autodock Vina was an appropriate tool for docking the inhibitors for B-Raf V600E. In order to do this, two of the inhibitors with crystal structures were compared. The first structure docked with B-Raf V600E was PLX 4032. This was overlaid with the pdb file 30G7 from the protein database. (Fig. 3.1) The RMSD between the docked ligand and undocked ligand was 8.01 Å.

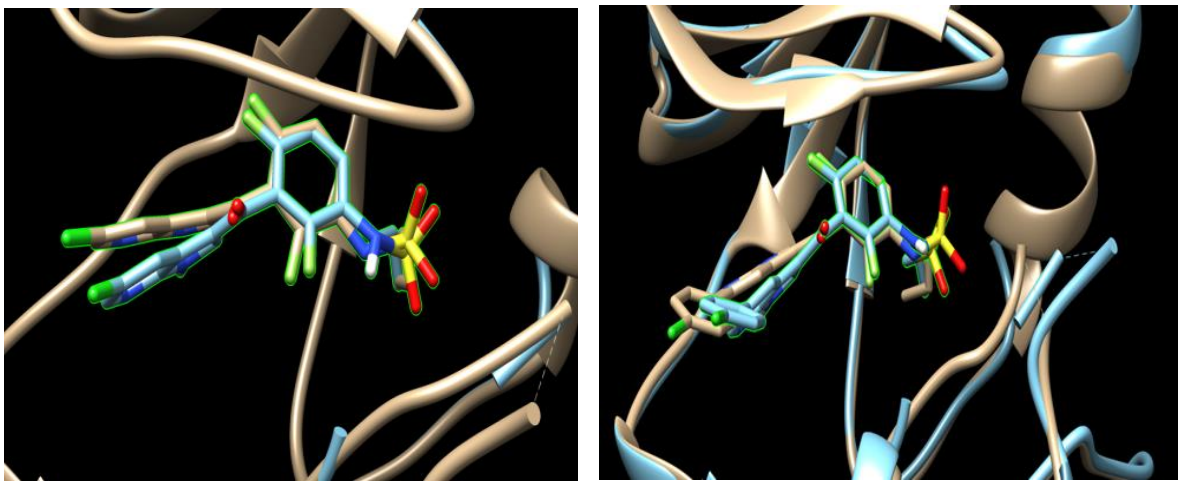
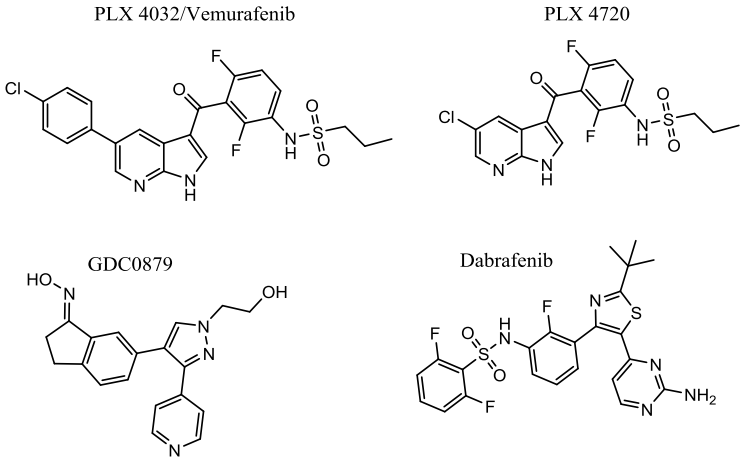


Figure 3.1 Left: Overlay of the crystal structure 3C4C (tan) with the Vina docked structure of the PLX4720 with B-Raf V600E (blue) Right: Overlay of the 30G7 crystal structure (tan) found in the protein database with the Vina docked structure of PLX4032 with B-Raf V600E (blue).

Another inhibitor was tested to determine the accuracy of Autodock Vina. The inhibitor selected was PLX 4720 with a pdb file 3C4C. The RMSD between the two ligands was 7.66 Å. Both docked images show that the inhibitors match the crystal structure well enough to obtain energies and poses for compounds lacking a published crystal structure. These poses also confirmed it was acceptable to continue with aMD simulations based on the generated poses. Figure 3.2 shows all inhibitors and their structure. The binding energies of each inhibitor with B-Raf V600E, B-Raf V600E-L505H (neutral His) and B-Raf V600E-L505H(P) (protonated His) were then calculated and shown in table 3-1.

A. Type I



B. Type II

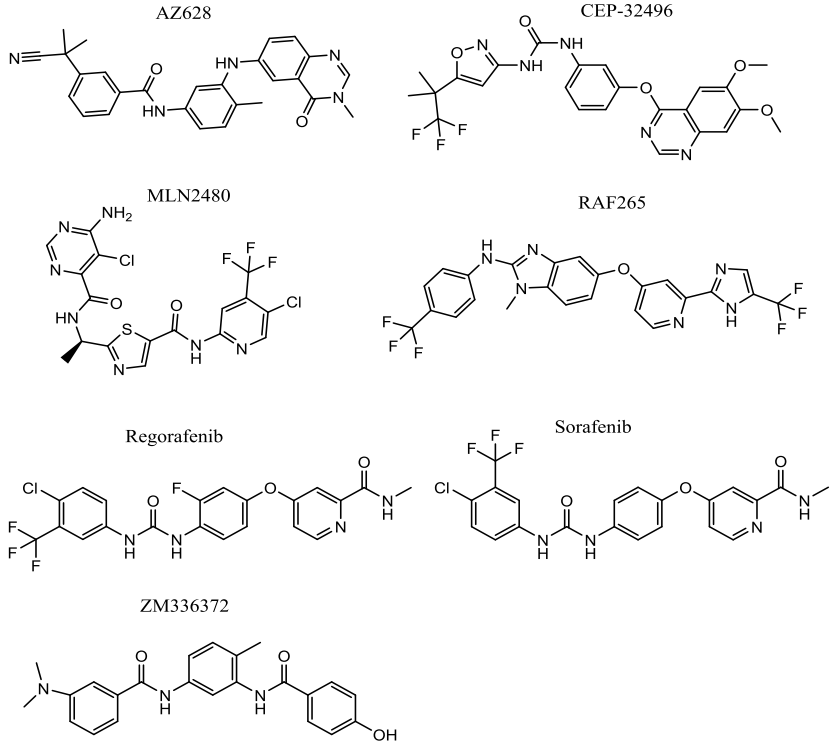


Figure 3.2 Structure of type I and type II inhibitors.

Table 3-1 Relative binding energies (kcal/mol) for B-Raf V600E, B-Raf V600E-L505H, and B-Raf V600E-L505H(P) using Autodock Vina.

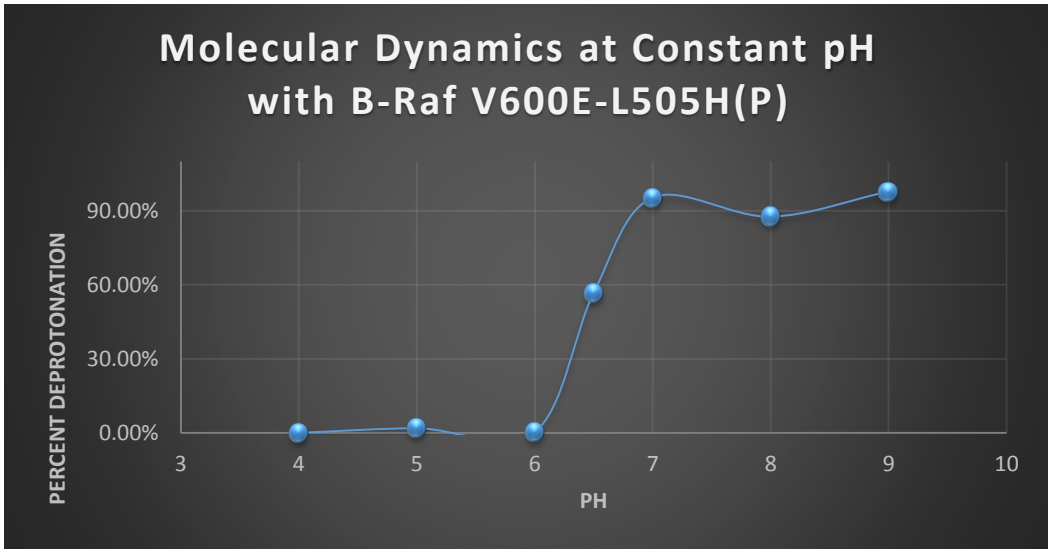
Inhibitor	V600E	V600E-L505H	V600E-L505H(P)
Type 1			
PLX4032 [32, 113]	-11.2	-9.6	-9.9
PLX4720 [114]	-10.5	-10.2	-9.1
Dabrafenib [115]	-10	-9.7	-9.9
GDC-0879 [19]	-9.2	-10	-9.3
Type II			
AZ628 [116]	-12.1	-12.4	-11.4
CEP-32496 [117]	-10.6	-10.5	-9.2
MLN2480 [118]	-8.2	-10.1	-9.1
RAF265 [119]	-11.2	-11	10.7
Regorafenib [120]	-11.1	-10.7	-10.8
Sorafenib [121]	-10.9	-10.5	-11.2
ZM336372 [122]	-11.5	-10	-9.7

PLX 4032 is an FDA approved drug so binding energies for the other potential inhibitors were compared to its binding energies. The energy for the B-Raf V600E bound with PLX 4032 is -11.2 kcal/mol and goes up for the B-Raf V600E-L505H to -9.6 kcal/mol and to -9.9 kcal/mol for the protonated version. This highlights a significant impact to binding in the presence of L505H. Many of the inhibitors follow a similar pattern. A notable exception is AZ628 as it has a lower computed binding energy than PLX 4032 for the V600E and V600E-L505H mutants, which shows it could possibly be a better inhibitor for the mutant. MLN2480 has a poor binding energy relative to the PLX 4032 bound with V600E but shows a significant decrease of -1.9 kcal/mol for the mutant. Sorafenib is interesting in that it was the only inhibitor to have a lower binding energy for the protonated His505 form. The binding energy for V600E-L505H(P) was still comparable to the V600E binding energy for PLX 4032.

The binding energies give a general trend on the potential effectiveness of the inhibitors. The type II inhibitors generally had lower binding energies than the type I inhibitors for the mutant variations. With exception to PLX 4032, they also had better affinities for the V600E variant. Although, these binding affinities are low level calculations, it shows that type II inhibitors have higher affinity than the type I inhibitors.

Constant pH Calculations

Although it is known that the tumor microenvironment is acidic [105], pKa shifts may occur drastically changing the standard pKa, and therefore the protonation state of a particular amino acid.[123] Constant pH simulations were performed to compute the pKa of the mutated His residue in the 505 position.



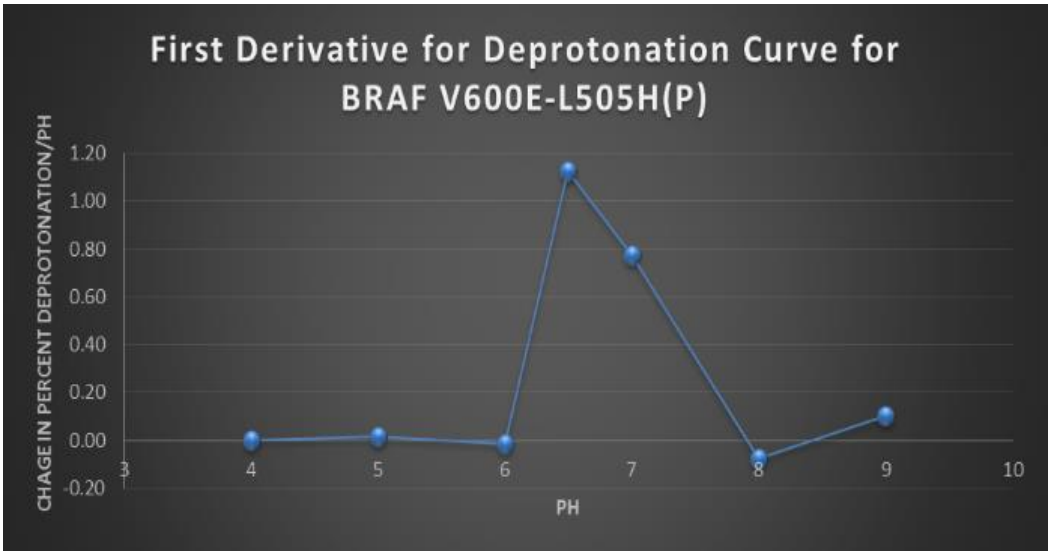


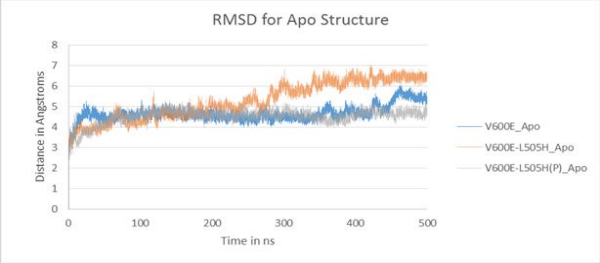
Figure 3.3 Top) The calculated titration curve of the apo V600E-L505H(P) structure. The y-axis represents amount of deprotonation and x-axis shows pH. Bottom) First derivative curve for titration of histidine in apo V600E-L505H(P). Y-axis represents change in deprotonation and x-axis represents pH.

If the pKa shifted away from the standard value of 6.5, it would only be necessary to investigate the His in either the neutral or protonated state. However, as shown in figure 3.2, the pKa was computed to be 6.5, meaning there was not a way to determine if the His was protonated or unprotonated most of the time. As the pH of the system inside the active site is unknown, two different forms of histidine, an unprotonated histidine and protonated histidine, were computed using aMD simulations. All three variations (V600E, V600E-L505H, and V600E-L505H(P)) were evaluated in order to determine the effectiveness of each inhibitor as well as potential differences in conformation of B-Raf.

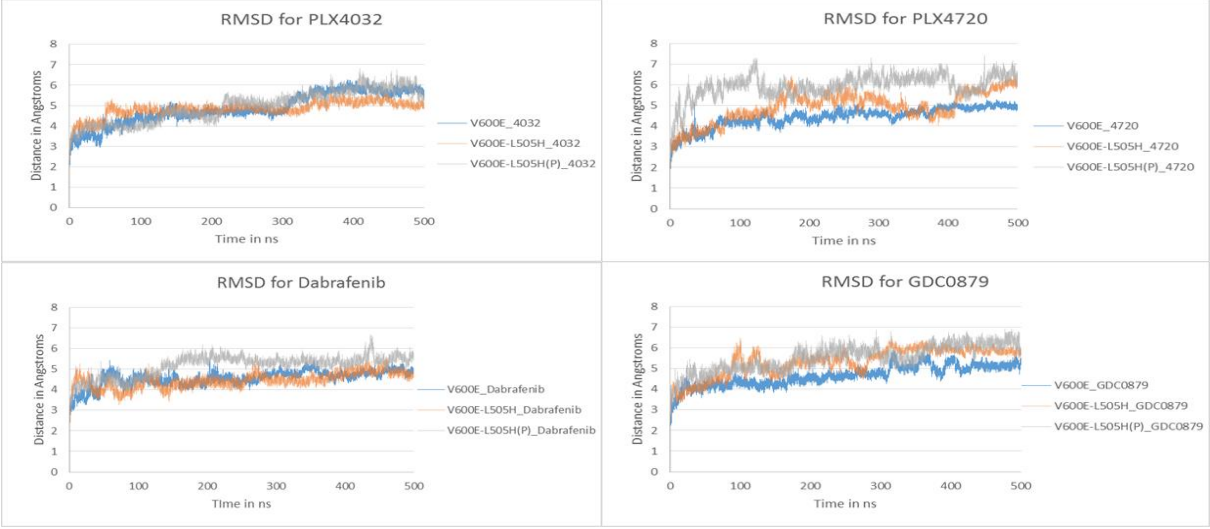
RMSD Calculations

Root mean square deviations (RMSD) were calculated for all the backbone atoms of B-Raf variants bound with the different inhibitors over the 500 ns simulation. The purpose of this was to determine that each simulation was carried out long enough and that no apparent spikes were present in the simulation.

A) Apo Structure



B) Type I Inhibitors



C) Type II Inhibitors

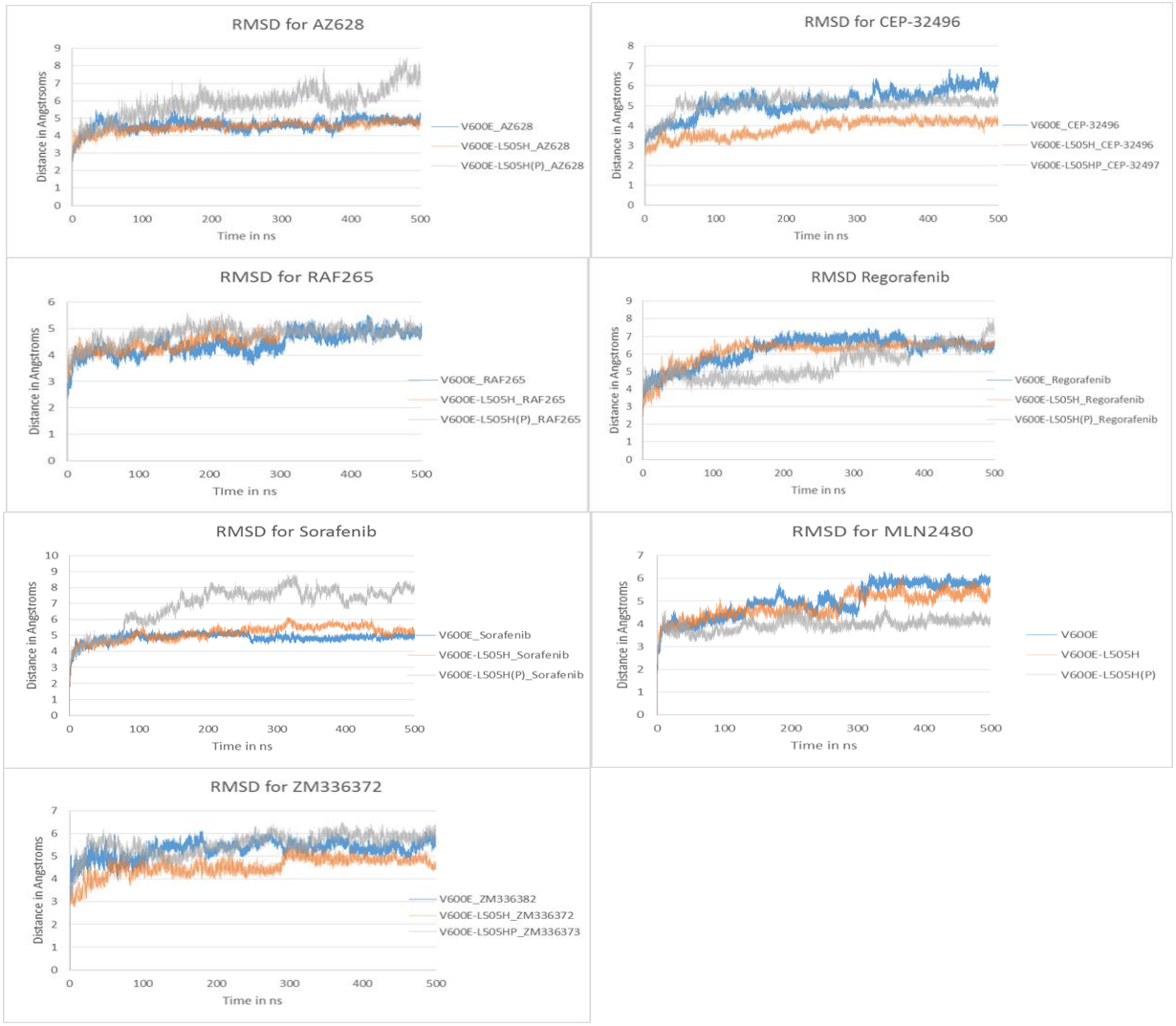


Figure 3.4 RMSD's for each inhibitor and B-Raf mutation. A. RMSD for Apo structure. B. RMSD for Type I inhibitors. C. RMSD for Type II inhibitors.

The RMSD graphs in figure 3.4 show that all calculations were sufficiently run in order to achieve equilibrium with the average RMSD being between 4 and 5 Å. The V600E and V600E-L505H mutations show that they achieve equilibrium quickly into the simulation. Several calculations show a spike at the end of the simulation, such as the apo structure of B-Raf V600E-L505H. This could be due to a large conformational change which will be explained later. The

same spike occurs with PLX 4720, Sorafenib, and AZ628 bound with B-Raf V600E-L505H(P). With the RMSD showing that simulation time was sufficient, further analytical calculations were performed.

Hydrogen Bonding

Inhibitors for B-Raf V600E should have hydrogen bonding, especially with the hinge region.[114] The hinge region involves amino acids 530-536 and is highlighted in orange in Figure 3.5.

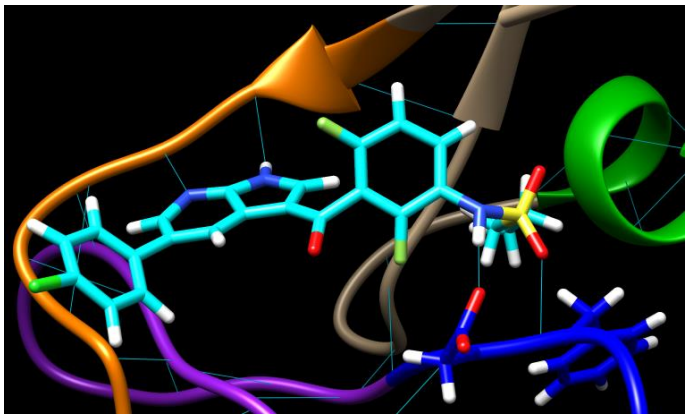


Figure 3.5 PLX 4032 bound with B-Raf V600E. The contacts with the hinge region are highlighted in orange.

Table 3-2 Left: The number of hydrogen bonds between inhibitor and hinge region with bonding occurring greater than 20%. Right: The number of hydrogen bonds between inhibitor and B-Raf active site occurring greater than 20%.

Total Number of Hydrogen Bonds with BRAF Hinge Region			
	V600E	V600E-L505H	V600E-L505H(P)
Type I			
PLX4032	2	1	1
PLX4720	2	2	0
Dabrafenib	2	0	0
GDC0879	1	1	1
Type II			
AZ628	1	0	0
CEP-32496	1	1	0
MLN2480	0	0	0
RAF265	0	0	0
Regorafenib	0	0	2
Sorafenib	1	0	0
ZM336372	0	0	0

Total Number of Hydrogen Bonds with BRAF			
	V600E	V600E-L505H	V600E-L505H(P)
Type I			
PLX4032	5	1	1
PLX4720	5	5	0
Dabrafenib	2	3	4
GDC0879	4	2	3
Type II			
AZ628	3	0	2
CEP-32496	2	1	2
MLN2480	0	1	1
RAF265	0	0	1
Regorafenib	2	1	4
Sorafenib	3	3	1
ZM336372	2	1	3

For a type I inhibitor to be effective it needs to interact with the hinge region in order to compete with ATP.[124] All of the type I inhibitors in table 3-2 have hydrogen bonding with the hinge region in the V600E mutation; however, the 505H mutation shows a decrease in the amount of hydrogen bonding occurring for both the hinge region and the total number of hydrogen bonds in the active site. Type II inhibitors show a different pattern with less boding occurring in the hinge region. All type II inhibitors have hydrogen bonding with the B-Raf mutants except for AZ628 and RAF265 which have zero bonding for the neutral His. While hydrogen bonding is clearly an important factor in inhibitor design, there is no clear pattern that describes the loss of effectiveness in the L505H mutation.

Clustering

Clustering analysis was performed using ptraj.[104] The highest percentage cluster was obtained and observed. Table 3-3 shows each cluster for the inhibitor and the variants. The majority of clusters were less than 40%. Some of the clusters showed a greater than 50% coverage of the simulation. The visuals in figures 3.6, 3.7, and 3.8 will help determine the change in orientation from one variant to the next.

Table 3-3 Shows the top cluster and percentage for each inhibitor and variant.

	V600E	**V600E-L505H**	**V600E-L505H(P)**
Apo	Cluster 5_54.8%	Cluster 9_32.0%	Cluster 8_29.8%
Type I			
PLX4032	Cluster 9_31.4%	Cluster 9_38.2%	Cluster 7_26.9%
PLX4720	Cluster 9_47.8%	Cluster 7_35.5%	Cluster 8_29.8%
Dabrafenib	Cluster 9_24.5%	Cluster 3_32.9%	Cluster 4_30.7%
GDC0879	Cluster 1_37.0%	Cluster 9_43.4%	Cluster 9_27.6%
Type II			
AZ628	Cluster 2_31.3%	Cluster 7_37.3%	Cluster 4_23.7%
CEP-32496	Cluster 5_29.5%	Cluster 9_24.6%	Cluster 9_64.5%
MLN2480	Cluster 9_36.8%	Cluster 4_28.2%	Cluster 9_30.2%
RAF265	Cluster 6_20.4%	Cluster 4_37.0%	Cluster 7_35.4%
Regorafenib	Cluster 9_34.1%	Cluster 9_72.1%	Cluster 2_37.5%
Sorafenib	Cluster 9_46.1%	Cluster 6_35.1%	Cluster 9_31.8%
ZM336372	Cluster 9_38.5%	Cluster 4_40.4%	Cluster 9_56.8%

Apo Structure

V600E V600E-L505H V600E-L505H(P)

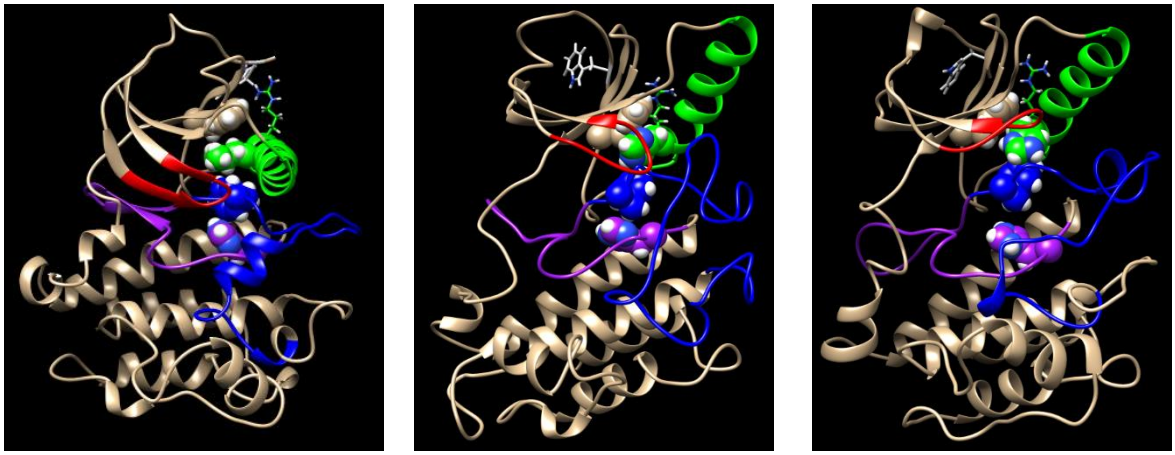


Figure 3.6 Illustrations of the major cluster for each apo B-Raf mutant, i.e. V600E, V600E-L505H, V600E-L505H(P)

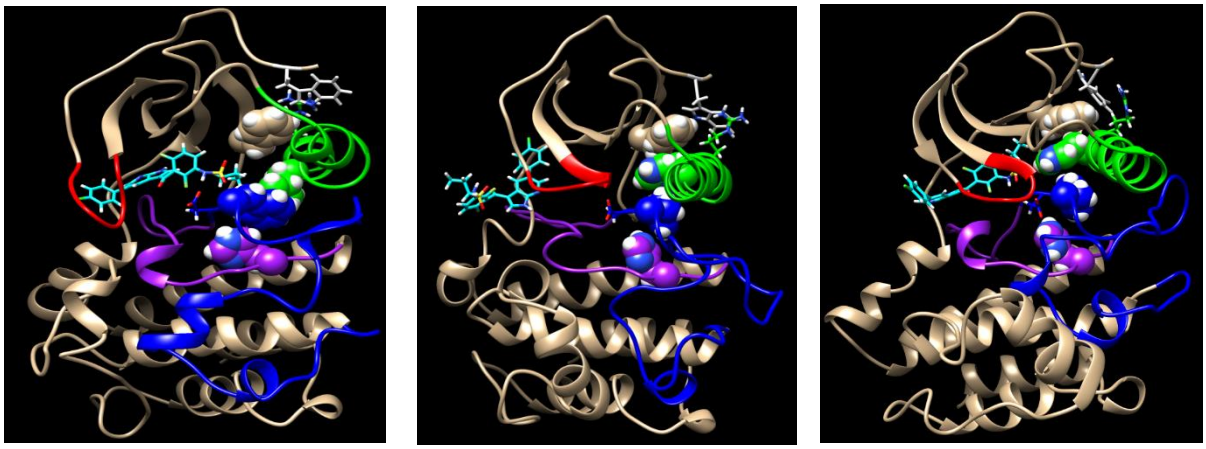
The different conformations of the apo structure show that the L505H mutation has an effect on conformation. The V600E apo structure comprises of 54.8% of the clustering while the His mutants show 32.8% for 505H and 29% for 505H(P). The αC-helix "in" conformation for the V600E shows an active conformation; however, the apo structure for both the mutants show an αC-helix out which would signify an inactive conformation.[14] The R-spine (regulatory spine), consisting of residues L505, F516, F595, and H574, aligns in V600E and 505H B-Raf which represents an active conformation.[125] The R-spine, however, is not aligned in the protonated mutant. Also, the L505H mutant replaces a key residue in the regulatory spine and has a notable effect on the kinase structure.

The net negative charge on the activation segment for the V600E between the key residues Ser599 and Ser602 is what is believed to cause uncontrollable phosphorylating to occur by either acting as a phosomemtic or by stabilizing the region for dimerization.[126] Two key residues are

also believe to play a crucial role in this dimerization process. Trp450 follows the Nta (N-terminalacidic) motif, which consists of SSDD. One of the serines is phosphorylated making the other three acidic. Mutational studies of tryptophan impaired B-Raf as an activator or receiver.[127] The position of this tryptophan dictates the capability of Raf dimerization. In SRC and HKC, the tryptophan is displaced from the hydrophobic pocket causing the αC-helix to conform to an active state.[128, 129] In ITK, tryptophan rotates inward to allow the αC-helix to adopt an active conformation.[130] The exact placement of tryptophan in B-Raf is unknown, however, it should bring the Nta motif in closer to the other key dimerization residue Arg509.[126] Because Arg509 is important to dimerization, helix stabilization is critical. The 505H forms of B-Raf have helix destabilization.

Type I Inhibitors

V600E_PLX 4032 V600E-L505H_PLX 4032 V600E-L505H(P)_PLX 4032



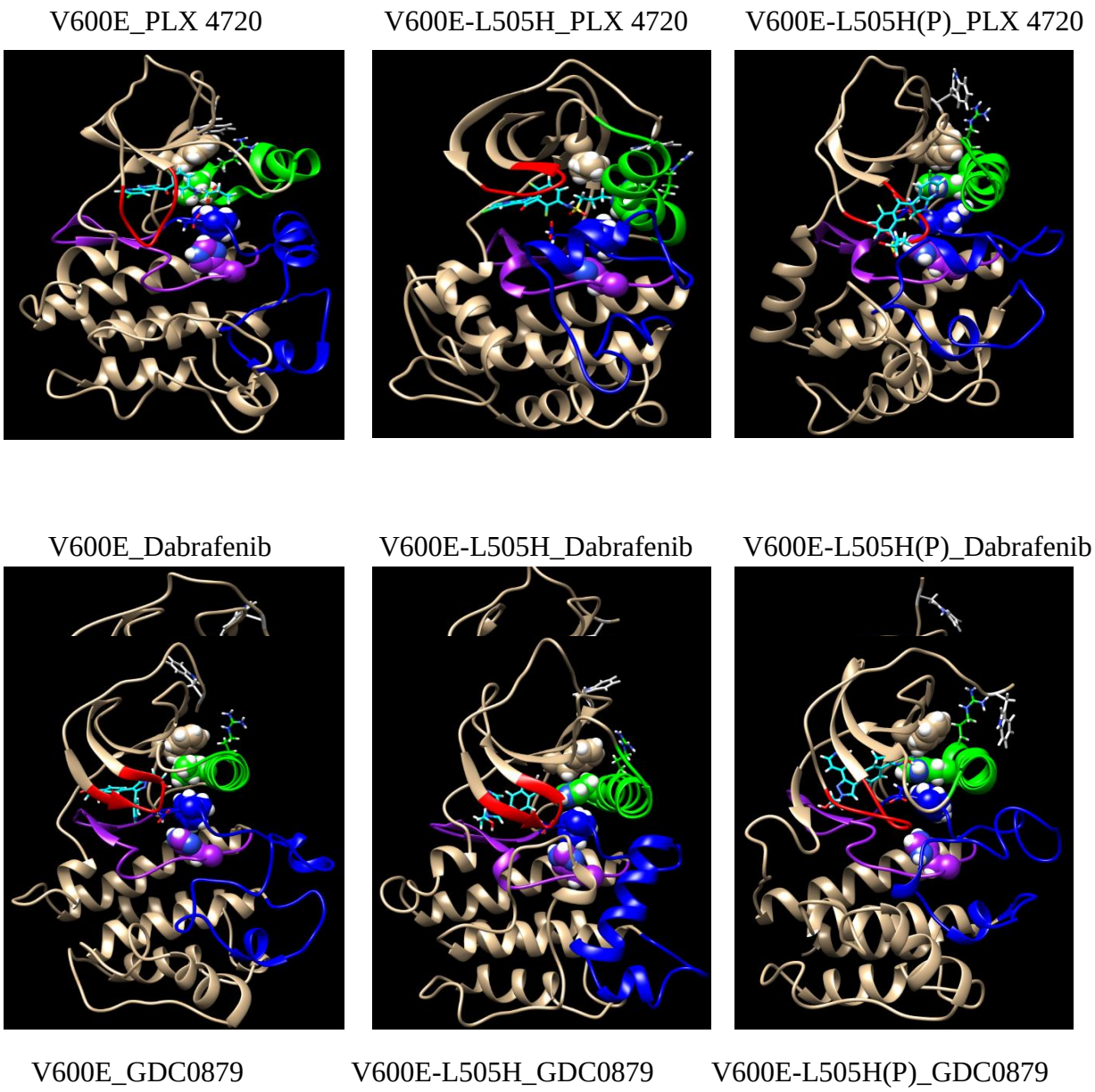


Figure 3.7 Illustrations of the major cluster for each type I inhibitor bound with B-Raf mutant, i.e. V600E, V600E-L505H, V600E-L505H(P)

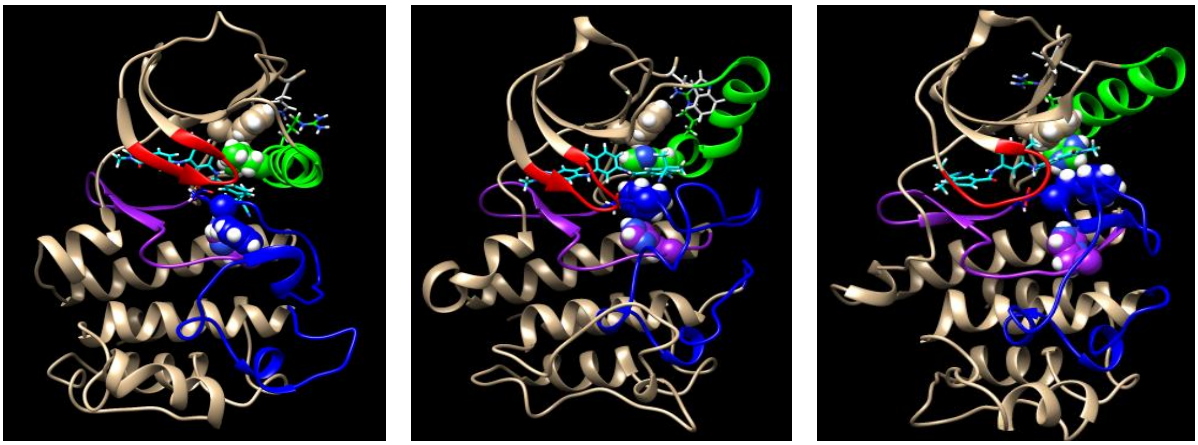
Type I inhibitors bind to a kinase that has a catalytically active binding site orientation and are ATP-competitive.[124] The R-spine as seen from the clustering is aligned in PLX 4032. Not all of the type I inhibitors behaved this way. Dabrafenib has a misaligned R-spine. PLX 4032, PLX

4720, and Dabrafenib all have disruption of the closed αC-helix even though the R-spine is aligned. This disruption prevents downstream phosphorylation and dimerization with other Raf isomers/enzymes.

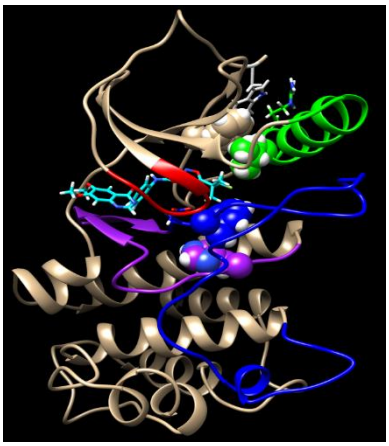
The mutation significantly changes the conformation of B-Raf for both the protonated and unprotonated version. GDC0879 has stabilization of the αC-helix in all three variations; however, the other three inhibitors show αC-helix stabilization in the 505H mutants. The protonated version shows the greatest affinity for helix stabilization. Note in the PLX 4720 simulation, the inhibitor was ejected out of the binding pocket. Trp450 also shows a pattern for the type I simulation and swings toward the αC-helix potentially setting B-Raf up for side to side dimerization. The key dimerization residue, Arg509, also points outward in the mutants. Stability of key components such as the helix and activation segment as well as orientation of the Nta motif allow the dimerization of B-Raf.[126] The type I inhibitors show a pattern that they may paradoxically activate B-Raf because of the 505 mutant.

Type II Inhibitors

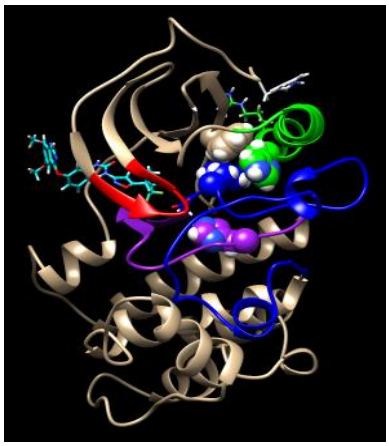
V600E_AZ628 V600E-L505H_AZ628 V600E-L505H(P)_AZ628



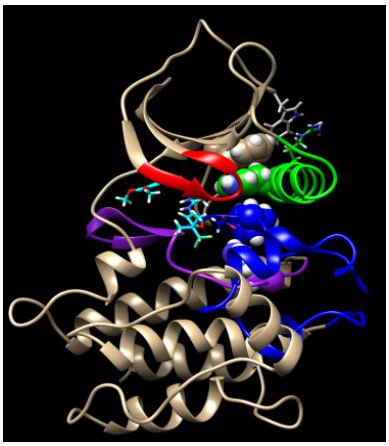
V600E_CEP-32496



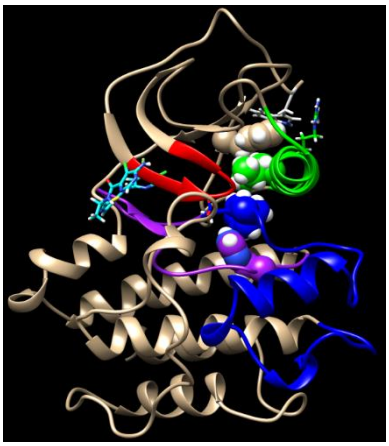
V600E-L505H_CEP-32496



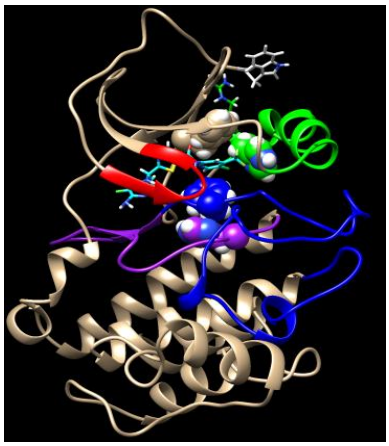
V600E-L505H(P)_CEP-32496



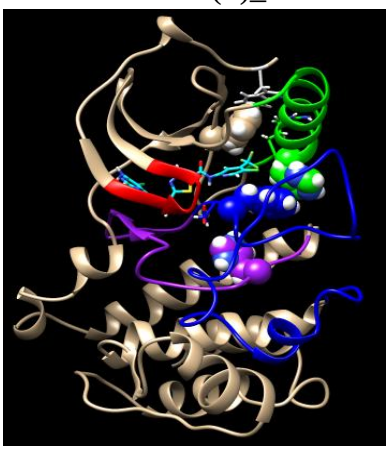
V600E_MLN2480



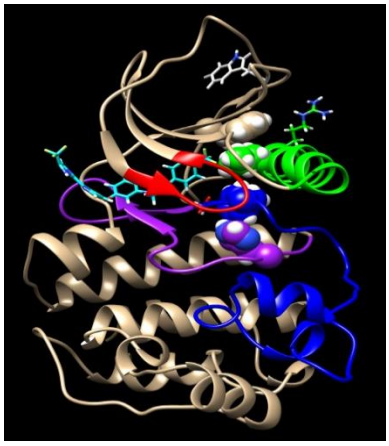
V600E-L505H_MLN2480



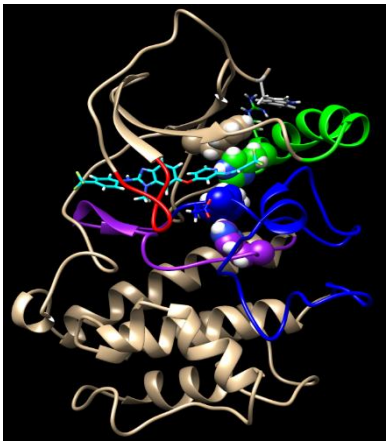
V600E-L505H(P)_MLN2480



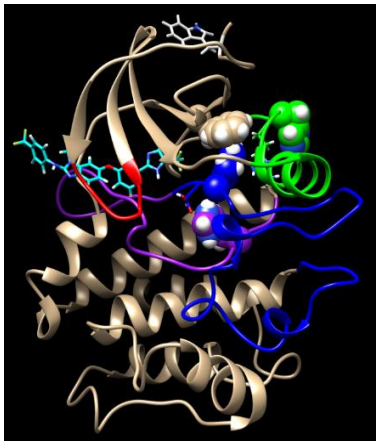
V600E_RAF265



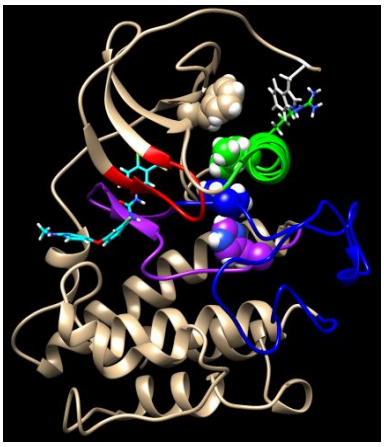
V600E-L505H_RAF265



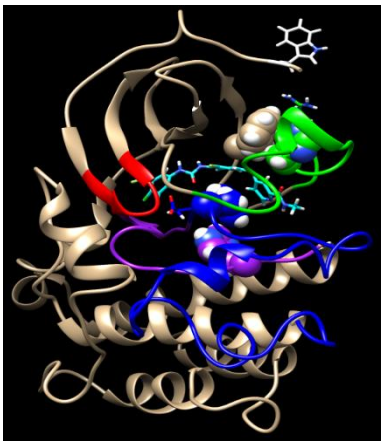
V600E-L505H(P)_RAF265



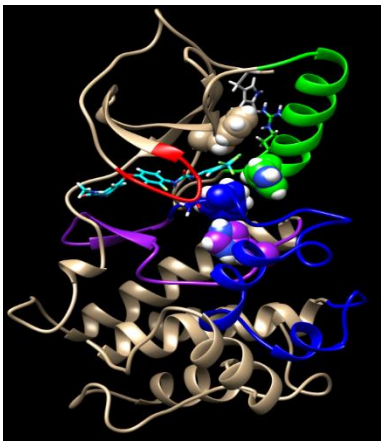
V600E_Regorafenib



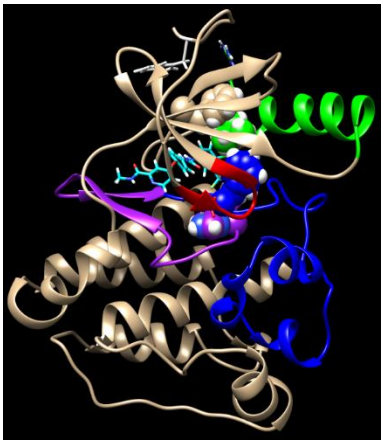
V600E-L505H_Regorafenib



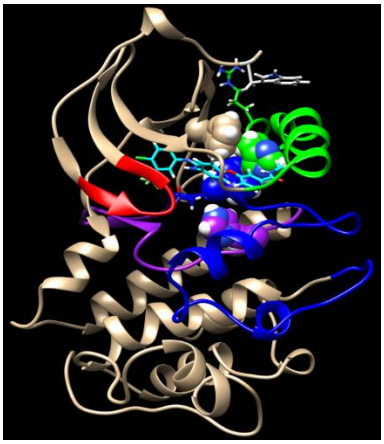
V600EL505H(P)_Regorafenib



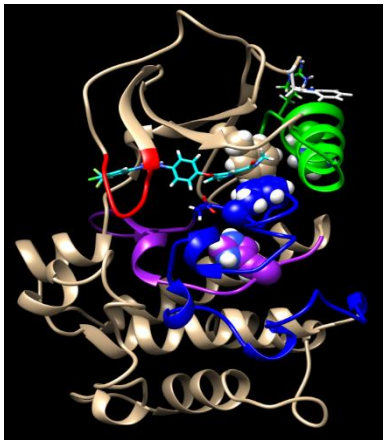
V600E_Sorafenib



V600E-L505H_Sorafenib



V600E-L505H(P)_Sorafenib



V600E_ZM336372 V600E-L505H_ZM336372 V600E-L505H(P)_ZM336372

Figure 3.8 Illustrations of the major cluster for each type II inhibitor bound with each B-Raf mutant, i.e. V600E, V600E-L505H, V600E-L505H(P)

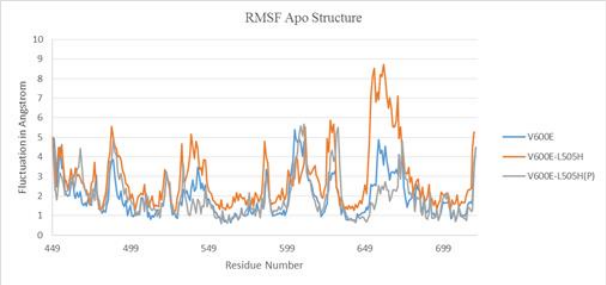
The type II inhibitors can bind to the binding pocket or an allosteric site, may or may not be ATP competitive, and may form an interaction with the R-spine displacing the F595 residue of the DFG motif. [124] Figure 3.8 shows clustering for all type II inhibitors bound the V600E, V600E-L505H, and V600E-L505H(P). Some of the type II inhibitors are active in at least one mutant form. AZ628, MLN2480, and Regorafenib all have a stable αC-helix as well as the Trp450 brought into the dimer interface. The type II C-Raf inhibitor, ZM336372, stabilizes the αC-helix in the V600E and 505HP. CEP-32496 exhibited a similar patter as ZM336372.

The approved type II inhibitor for metastatic melanoma is Sorafenib. In all three B-Raf mutants, Sorafenib induces a conformational change of the αC-helix. The V600E mutant has distortion of the αC-helix as well as the Trp450 residue pointing away from the dimer interface. The inhibitor flips in the active site for the 505H mutants and does bring the tryptophan closer to the dimer interface; however, the stability of the αC-helix is distorted making dimerization unlikely.

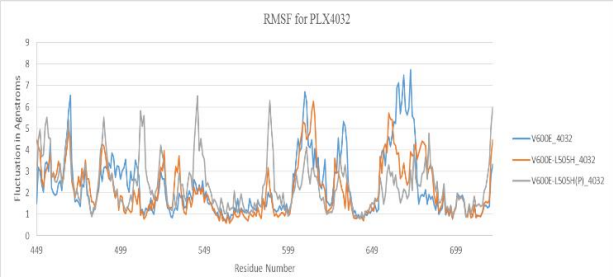
RMSF Calculations

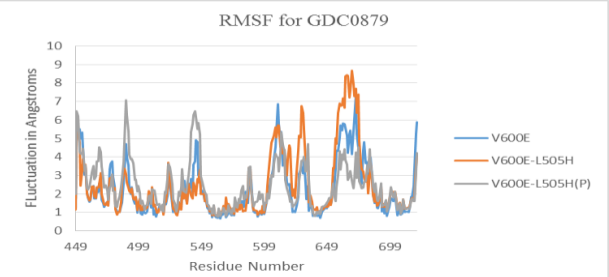
Root mean squared calculations (RMSF) were performed for all inhibitors on all three variants of B-Raf. RMSF calculations show the fluctuations of backbone atoms in the protein. Figure 3.9 illustrates the RMSF for each inhibitor with all B-Raf mutants.

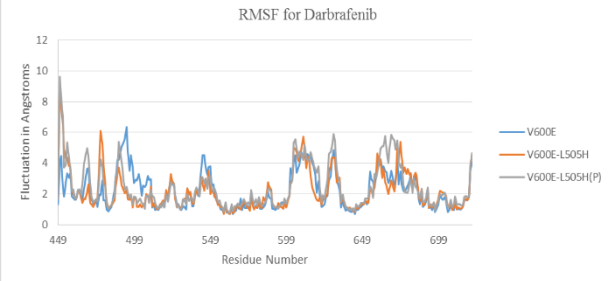
I. Apo Structure

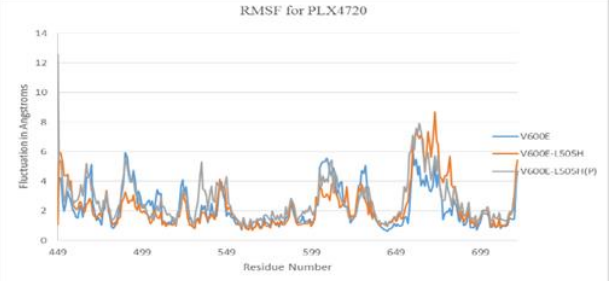


II. Type I Inhibitors









III. Type II Inhibitors

Figure 3.9 RMSF calculations for inhibitors bound with B-Raf mutants.

The fluctuation of each residue was calculated with respect to an average structure. The intensity of the fluctuations show the dynamic nature of B-Raf. Intense fluctuations may also be occurring because the simulations are crossing the high energy barriers because aMD raises the minima potential to allow for more sampling. The fluctuations show several dynamic regions in B-Raf. The αC-helix region, ranging from region 491-510, spikes in all calculations. The activation

segment is also another mobile region ranging from residues 593-617. These regions are vital to the activity of B-Raf as well as its ability to dimerize. The other highly mobile region is the helical C-lobe from residue 645-721.

Distance for Salt Bridge LYS 507 and GLU 600

A salt bridge was believed to form between Lys507 on the αC-helix and the Glu600 mutation on the activation segment.[126] Lys507 is part of the RKRT motif on the αC-helix, a conserved region on kinases that allows for dimerization.[131] This interaction forces the activation segment to fold over as well as the αC-helix to adopt an "in" orientation, mimicking the active dimer interface.[132]

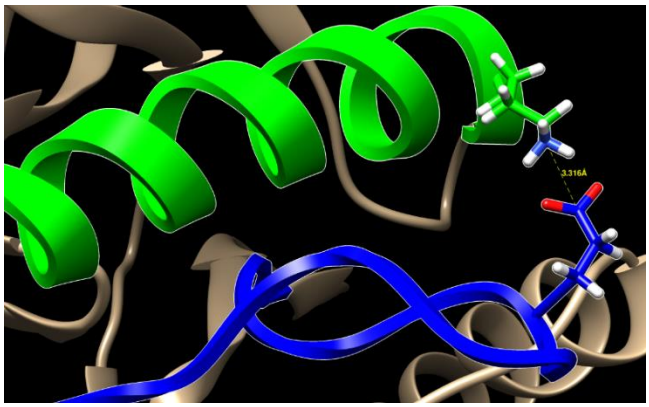


Figure 3.10 Illustration of the Glu600-Lys507 salt bridge for the apo B-Raf V600E which was taken at 500 ns.

The average distance calculated for the Lys507-Glu600 salt bridge for the apo V600E was 5.90Å. This distance shows that the interaction between these two residues may induce a potential dimer interface. The salt bridge distance for the 505H mutants was 9.52Å for the neutral His and

7.04Å for the protonated His. This further elucidates that the His mutation in both the neutral and protonated version disrupts the potential for a dimer interface by moving the αC-helix out.

PLX 4032, PLX 4720, and Dabrafenib all show a broken salt bridge for V600E; however, they all reform the salt bridge for the 505H mutants. PLX 4720 showed the salt bridge reconnecting when the inhibitor was ejected out of the active site. GDC0879 has the salt bridge connected in all three mutants of B-Raf indicating it is potentially ineffective against the V600E mutant as well. This further elucidates the type I inhibitor may paradoxically reactivate the B-Raf 505 mutants through stabilizing the dimer interface.

Type II inhibitors with the exception of CEP-32496 and ZM336372 show disruption of this salt bridge. Several of the type II inhibitors shown in table 3-4 activate the salt bridge for the V600E variant. Sorafenib was the only inhibitor out of all the type I and type II inhibitors to deactivate the salt bridge for each mutant.

Table 3-4 Average distances shown in angstroms for Glu600 and Lys507 salt bridge. Distances were calculated every .05 ns and averages computed for the 500 ns simulation.

	V600E	V600E-L505H	V600E-L505H(P)
Apo	5.90	9.52	7.04
Type I			
PLX4032	10.87	5.44	5.08
PLX4720	11.17	4.03	11.81
Dabrafenib	11.19	4.53	4.65
GDC0879	4.82	4.56	4.62
Type II			
AZ628	5.13	5.68	20.38
CEP-32496	14.88	10.42	3.96
MLN2480	4.67	19.65	10.09
RAF265	4.78	12.20	13.15
Regorafenib	4.10	12.33	16.63
Sorafenib	12.88	16.34	29.52
ZM336372	12.36	10.68	4.06

Distance for Salt Bridge LYS 483 and GLU 501

A salt bridge exists between lysine 483 and glutamate 501 that also dictates whether or not the protein is in the active or inactive conformation.[6] For the B-Raf V600E, it is expected that the salt bridge would remain intact as continuous phosphorylation would occur with the phosomemtic 600 glutamate mutation.

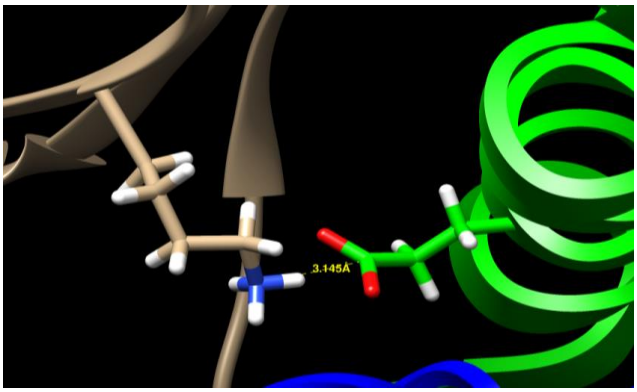
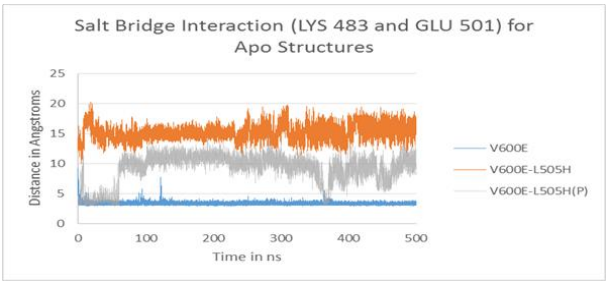
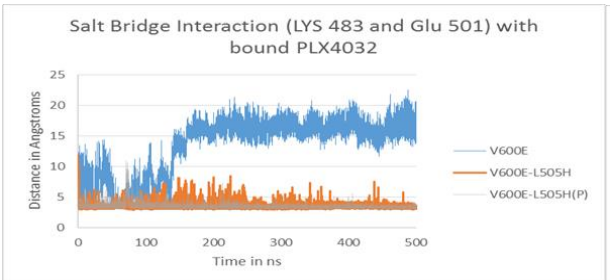


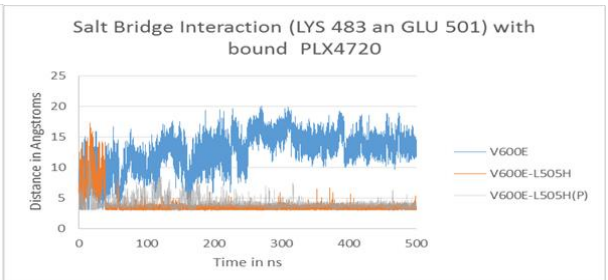
Figure 3.11 Illustration of the Glu501 and Lys483 salt bridge for the apo B-Raf V600E which was taken at 500 ns.

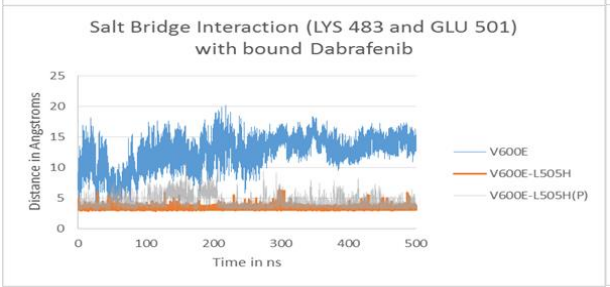
I. Apo

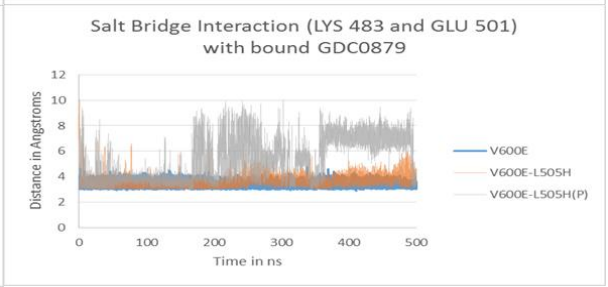


II. Type I









III. Type II

Figure 3.12 Distances for the Glu501 and Lys 483 salt bridge shown over the 500 ns simulation for inhibitors bound with B-Raf V600E, B-Raf V600E-L505H, and B-Raf V600E-L505H(P)

The apo salt bridge shows that B-Raf adopts an active state for V600E; however, the 505H mutations break this salt bridge. This relates to the clustering calculations as the 505H structures destabilize the αC-helix and adopt an "out" configuration. The mutation causes bad interactions

within B-Raf to cause it to become inactive. However when a type I inhibitor is bound to the protein, the salt bridge breaks for V600E but forms for the 505H mutations.

Each type I inhibitor except for GDC0879, which had the highest binding energy according to the docking calculations, broke the salt bridge for the V600E mutation. However, they each caused the salt bridge to form again for both the 505H mutations. The type I inhibitors may be causing a paradoxical activation because of the 505H mutation as the apo structure showed that 505H mutations put B-Raf in an inactive form. This further hints that the type I inhibitors may be forming a favorable dimerization environment.

The type 2 inhibitors had a different pattern for salt bridge activation. AZ628, MLN2480, RAF265, and Regorafenib all activated the salt bridge for the V600E form, but deactivated the salt bridge for the neutral 505H mutant. Others such as CEP-32496 and ZM336372 activated the salt bridge for the protonated form of 505H. Sorafenib was the only inhibitor to deactivate the salt bridge for all three forms of B-Raf.

The type I and type II inhibitors show distinct differences in the structure and differences in the conformation of B-Raf variants. Based on these distance calculations, type I inhibitors would be ineffective for this mutation because they activate B-Raf. The type II inhibitor does not show enough consistency to make a determination, but Sorafenib was overall the best inhibitor.

Conclusion

B-Raf V600E has effective inhibitors; however, the L505H mutation causes conformational changes within B-Raf to reject current treatments. The apo structure of B-Raf V600E-L505H/L505H(P) shows an inactive conformation in both clustering calculations and

distance calculations. The salt bridge between Lys507 and Glu600 also shows that the mutation causes unfavorable interactions between the activation segment and αC-helix.

Type I inhibitors potentially have an activating effect on the B-Raf L505H mutants by stabilizing the dimerization region. Experiments have shown that some of the inhibitors are strong inducers of B-Raf dimerization.[133] Many of the type II inhibitors show activation for the V600E based on both salt bridge calculations, but they do show deactivation for the 505H mutations. Sorafenib was the only inhibitor to show deactivation for all three mutations. In conclusion, the L505H mutation deactivates B-Raf; however in the presence of an inhibitor either type I or type II, B-Raf can be stabilized for dimerization and activation. The most potent inhibitor for this mutation was determined to be Sorafenib.

Chapter 4 Mechanistic Role of the Mobile Loop for Alkanesulfonate Monooxygenase

Introduction

This study focuses on the mobile loop of SsuD to determine the types of closure as well as the interaction of Arg297 with reduced flavin, FMN, and oxidized flavin, FMNOO-, and octanesulfonate. The purpose of this study was to examine the mobile loop region of SsuD, and the conformational changes that occur depending on the substrate. Accelerated molecular dynamics (aMD) was performed on the SsuD for 1) apo 2) FMN and octanesulfonate bound, and 3) FMNOO- and octanesulfonate bound. These aMD simulations were carried out for 1000 ns to ensure ample conformational sampling. Extensive analysis such as RMSD, RMSF, clustering, hydrogen bonding, principal component, and atom distance were performed on each system to determine the motions of the mobile loop as well as its effects. Hence forth, the reduced flavin will be known as FMR and the oxidized flavin will be known as FMP.

Methods

Enzyme Preparation

The coordinates of SsuD were obtained from a 2.3 Å resolution crystal structure of 1M41 from the Protein Data Bank.[47] This structure was missing the part of the dynamic region with residues ranging from 250-282 and was also missing the C-terminal residues 362-380. MODELLER 9.10 was used to insert the residues based on proteins with similar templates.[106] FMN substrate was inserted based on the coordinates from the LadA enzyme with bound FMN (PDB 3B90).

AutoDock Vina

Octanesulfonate was inserted into SsuD through Autodock Vina.[66] A grid box was set over the active site to encompass all pertinent residues and 10-20 binding predictions were performed to find the lowest energy conformer.[55,44,49,47] All active site residues were set to be rotatable. The iterated local search global optimizer algorithm was used to evaluate the binding affinities of the enzyme and substrate. Calculations were conducted with an exhaustiveness set to 100, the number of binding modes set to 20, and the maximal energy difference between best and worst binding mode set to 5. The lowest energy pose was then used for the molecular dynamic simulations.

Ligand Optimization

Ligands for SsuD were optimized using Monte Carlo conformational search using BOSS program [134] and the OPLS force field. [135] The lowest energy structures were optimized using MP2/6-31G(d) theory level using Gaussian 09 software.[136] The lowest energy structure was then used to obtain restrained electrostatic potential (RESP) atomic charges using HF/6-31G(d) and the AmberTools antechamber module.[108]

For further information about enzyme preparation, docking, and ligand optimization for SsuD reference "Exploring the Catalytic Mechanism of Alkanesulfonate Monooxygenase Using Molecular Dynamics." [55]

Accelerated Molecular Dynamics

Molecular dynamic simulations were carried out on WT SsuD using the following combinations: Apo, FMR (reduced flavin) and octanesulfonate, and FMP (oxidized flavin) and octanesulfonate. Using Amber14 [87], the ff99SB force field [109] was used to create topology for standard amino acids, and the general Amber force field (GAFF) [100] was used to create the topology for the ligands. Leap was used to add hydrogens to the structure. The systems were solvated using the TIP3P water model [110] extending to 10 Å, and ions were added to obtain neutrality.

Each system was minimized using conjugate gradients (CG) [111] for 200 steps with water molecules only. The system was then minimized for 10,000 steps using CG to remove bad contacts. The next step was equilibration. The system was heated from 0 K to 300 K for 50 ps of MD using a weak coupling algorithm and a temperature coupling value of 2.8 ps. It was then equilibrated using the NPT ensemble at 300 K and 1atm with a coupling value of 2 ps for temperature and pressure. This was run for 500 ps. Finally, the system was switched to NVT and equilibrated for another 500 ps. The system was then run for 10ns of production on the GPU-accelerated version of Amber14.[112] Each simulation had a time step of 1fs and a cutoff distance of 12 Å.

Accelerated molecular dynamics (aMD) was then implemented. The parameters calculated for aMD were based on the recommended values for Amber.[86] The previous production phase was used to obtain dihedrals and potential energies. Each system was run for 1000 ns to obtain ample conformational sampling. All simulations were carried out using particle mesh Ewald, a non-bonded cutoff of 12 Å, and a time step of 1 fs.[137,83] These calculations were carried out on computers located at the Alabama Super Computer Center and Auburn University.

Results and Discussion

RMSD Calculations

Root mean square deviations (RMSD) were calculated for all the backbone atoms of the monooxygenase system for the Apo, FMR_OCS bound, and FMP_OCS bound variants over the 1000ns simulation. The purpose of this was to determine that each simulation was properly equilibrated with apparent spikes present in the simulation. Each system showed convergence for the simulation. This allows for further analysis of the system to continue.

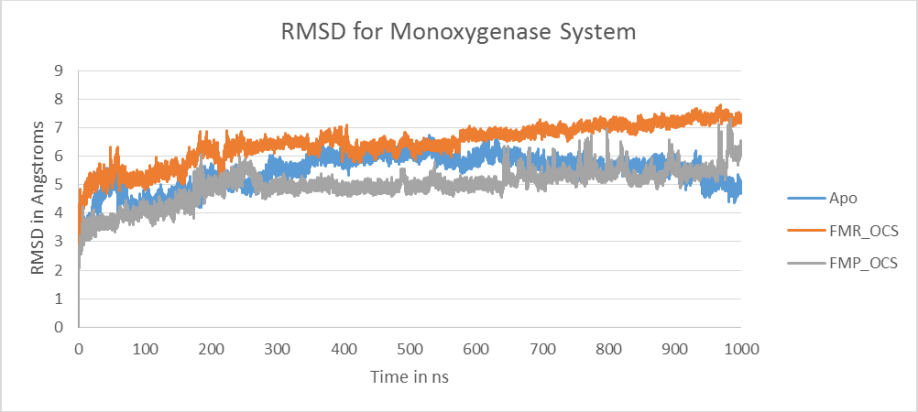


Figure 4.1 Root mean squared deviation for accelerated molecular dynamic simulation for monooxygenase variants.

Clustering

For each run of accelerated molecular dynamics, clustering calculations using ptraj were performed to determine the most common configuration of monooxygenase, and the location of the mobile loop. [104] Ten clusters were calculated for each simulation and table 4-1 shows the occupancy of each cluster.

Table 4-1 Percentage occupancy of each cluster. All values represent a percentage number.

Simulation	1	2	3	4	5	6	7	8	9	10
Apo	0.20	1.80	1.00	0.40	1.50	21.40	0.40	1.00	6.90	65.40
FMR_OCS	0.20	0.50	3.60	11.10	2.20	4.50	0.20	1.90	14.40	61.40
FMP_OCS	0.90	0.30	8.00	8.20	0.10	1.30	1.90	6.60	69.50	3.00

Cluster 10 showed the highest percentage for both the Apo and FMR_OCS simulation, and cluster 8 showed the highest occupancy for the FMP_OCS simulation. Because these numbers represent a large portion of the simulation, these clusters were analyzed to determine the most likely orientation of the mobile loop.

The Apo and FMP_OCS cluster showed an open configuration over the entire simulation and the top clusters consist of >65% of the simulation. The FMR_OCS simulation showed the closed conformation more than 60% of the time as shown in figure 4.2.

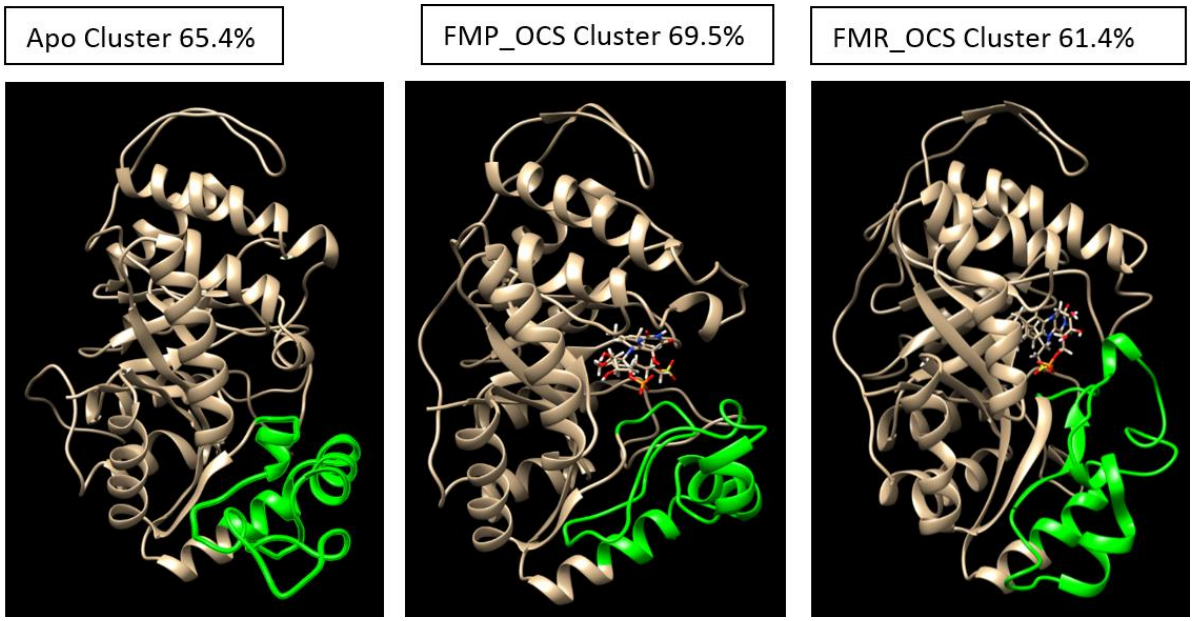


Figure 4.2 Top cluster for each simulation. The top Apo cluster was shown 65.4% of the time. The top FMP_OCS cluster was shown 69.5% of the time, and the top FMR_OCS cluster was shown 61.4% of the time. The mobile loop is highlighted in green.

The purpose of performing the clustering calculations was to determine the favored mobile loop conformation over the 1000 ns simulation. The Apo structure shows an open conformation. FMP_OCS also shows an open conformation, but it is different than the Apo loop possibly showing more movement which will be further analyzed in a later section. FMR_OCS shows a closed mobile loop which was anticipated to help the oxidation reaction facilitate for the conversion of FMR to FMP. The mobile loop is potentially used to block out bulk solvent and also to pertinent residues available for catalysis.[138] This loop also contains the important residue of Arg297 to potentially help anchor either FMR or OCS in the active site.

RMSF Calculations

Root mean squared fluctuation (RMSF) calculations were computed for each aMD simulation to determine the fluctuation of the backbone atoms for each residue.

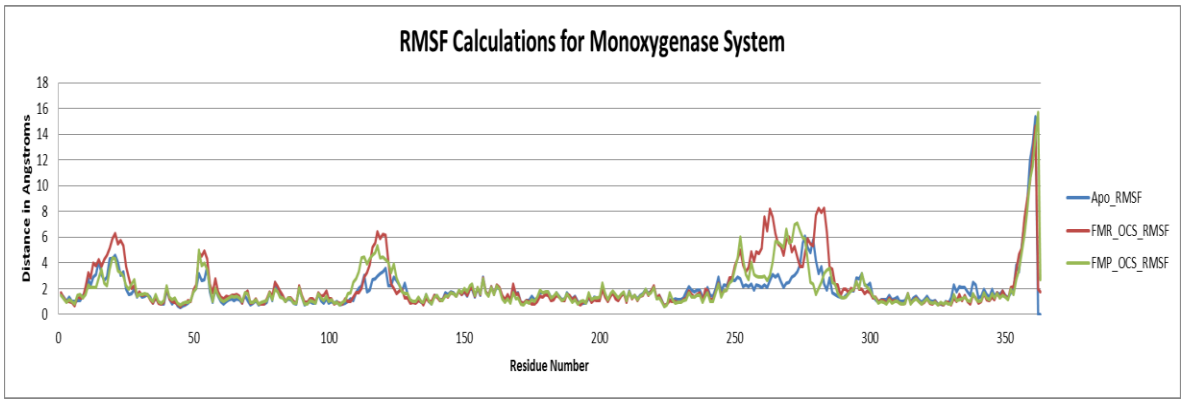


Figure 4.3 RMSF for all residues. Each simulation was overlaid to determine differences in fluctuations.

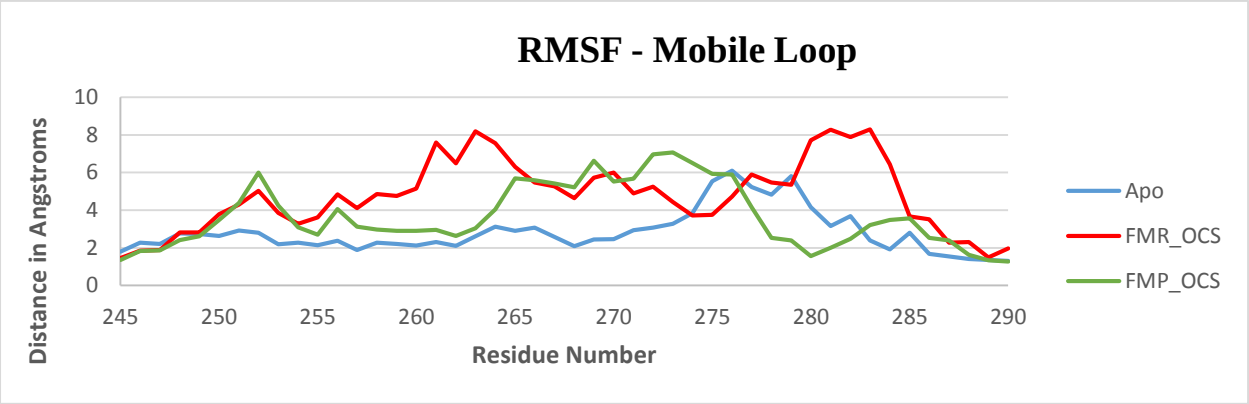


Figure 4.4 RMSF for mobile loop residues

The RMSF of the system (Figure 4.3) shows that the majority of residues have similar fluctuations with the exception of the mobile loop region. The mobile loop, i.e. residues 245-290, shows large fluctuations for FMR_OCS and FMP_OCS (Figure 4.4) The RMSF of the mobile loop shows that FMR_OCS has the greatest fluctuation which coincides with the clustering analysis showing a closed conformation. The Apo structure does not show as much movement with most of the fluctuation in the end of the mobile region. This is represented as an open conformation. FMP_OCS shows some fluctuations and at certain points it is higher than FMR_OCS; however, FMR_OCS has more movement over the entire loop showing a greater displacement of the mobile loop. Residues 1-20 are close to the active site and the large spike may be due to orientation of key residues during catalysis. The other peaks at residues 50-60 and 100-130 fluctuate the most for FMR_OCS and FMP_OCS because they undergoing orientation to allow for loop closure. These calculations show that the mobile loop has more motility with FMR_OCS confirming theories that this loop is dynamic and may play a role in catalysis.

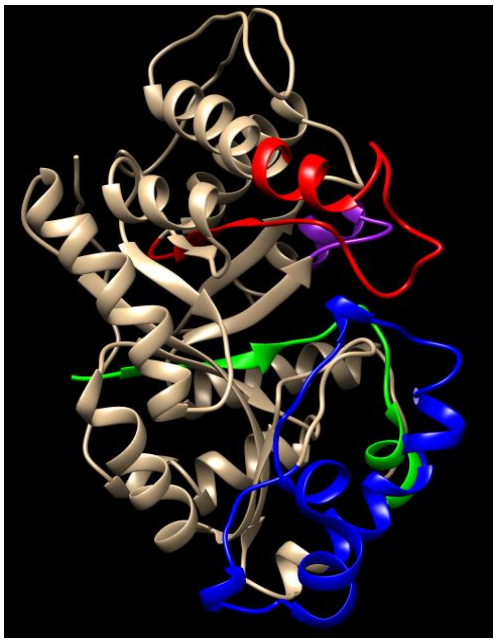


Figure 4.5 Represents monomeric SsuD system. Residues 1-20 are highlighted in green. Residues 50-60 are highlighted in purple. Residues 100-130 are highlighted in red. Residues 245-290 are the mobile loop region and highlighted in blue.

Principal Component Analysis

A covariance matrix was generated for each simulation to obtain the eigen values and principal components. Table 4-2 shows the eigen values and the cumulative percent of each component. Two simulations show a higher than 50% cumulative percent for the first two principal components. The FMP_OCS shows only a 46% cumulative for the first two principal components. A higher percent would be desired; however, principal component graph was analyzed even with the lower percentage for the entire system. Figure 4.5 shows the principal component graph for each simulation.

Table 4-2 Table shows the Eigenvalues and cumulative percent for the first three principal components for each simulation.

FMR_OCS			
	PC1	PC2	PC3
Eigenvalue	1033.28	315.73	184.37
Cumulative %	41.51	54.20	61.60

Apo			
	PC1	PC2	PC3
Eigenvalue	757.23	112.93	72.88
Cumulative %	47.05	54.07	58.60

FMP_OCS			
	PC1	PC2	PC3
Eigenvalue	626.902	182.987	132.47
Cumulative %	35.87	46.34	53.92

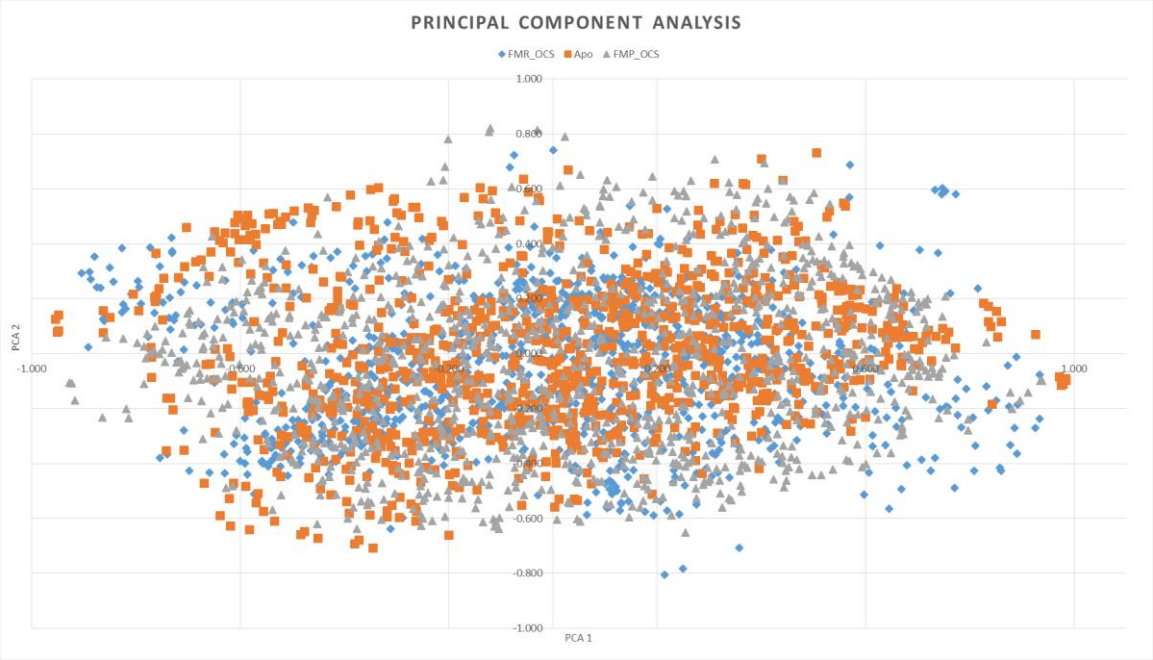


Figure 4.6 Principal component graph for all residues over 1000 ns.

Each simulation overlays in the graph with the exception of one section. The FMP_OCS and FMR_OCS are the only two present in the bottom right quadrant. This may represent the different conformations of the monooxygenase mobile loop. The FMR_OCS also has a small section of its own possibly showing an even bigger difference in conformation and representing the closed conformation of the mobile loop. A further breakdown of only the mobile loop, amino acids 241-300, would show the difference in the projections. Table 4-3 shows the cumulative percentage for each variant. Both the Apo and FMP_OCS cover less than 50% of the space in the two projections. The FMR_OCS has a percentage of 72% for the first two projections. In figure 4.6, as the mobile loop becomes more dynamic, Apo< FMP_OCS < FMR_OCS, the projection occupies more space. Overall, this graph shows a resemblance to the PCA graph of the entire system. It shows that as the loop becomes more dynamic, the projections cover a greater amount of space

Table 4-3 Table shows Eigenvalues and cumulative percentage for the first projection of the mobile loop only.

FMR_OCS			
	PC1	PC2	PC3
Eigenvalue	7989.00	934.89	694.85
Cumulative %	63.67	71.12	76.66
Apo			
	PC1	PC2	PC3
Eigenvalue	1366.10	538.55	463.56
Cumulative %	31.46	43.86	54.54
FMP_OCS			
	PC1	PC2	PC3
Eigenvalue	1795.93	1548.38	652.64
Cumulative %	25.40	47.29	56.52

.

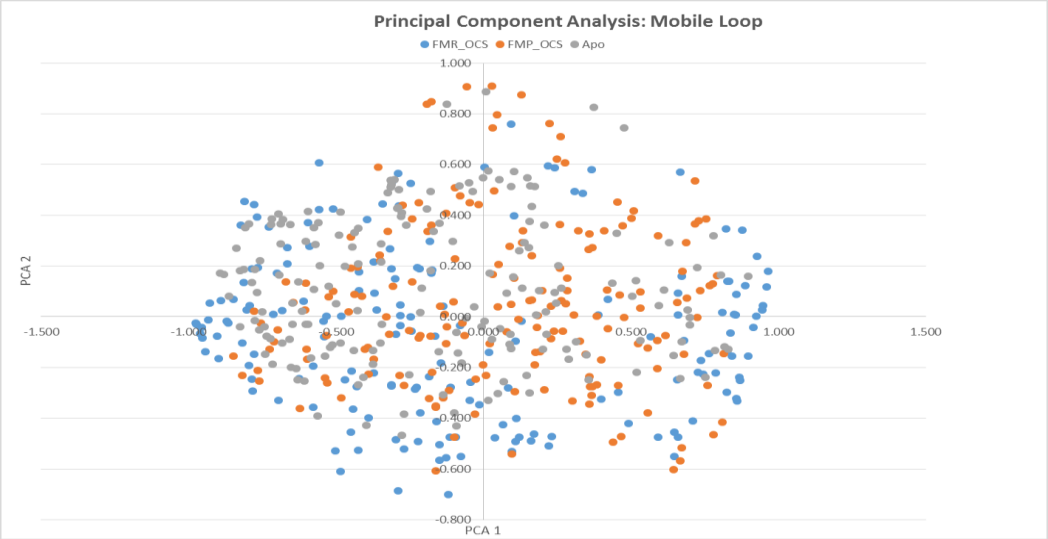


Figure 4.7 Principal component analysis of mobile loop

Distance Calculations

Distance between Phe261 and Pro112 was also calculated. The purpose of this calculation was to see evaluate the movement of the mobile loop. The graph in figure 4.7 shows the distance for all three variants over the 1000 ns simulation. The Apo and FMP_OCS simulations showed a distance around 20 Å which is depicted in the images taken at 900 ns. For the FMP_OCS simulation, the image shows the active pocket is wide open allowing for the potential exit of substrates. The FMR_OCS shows that the mobile loop has a big shift around 300 ns. The Phe261 and Pro112 interaction was thought to potentially help stabilize the mobile look through intramolecular inter-residue interactions in the FMR_OCS simulation.[139] Mutational studies were performed to see if these residues may have any pertinence to catalysis; however, the activity of the enzyme was unchanged in the mutant (unpublished data). These residues, however, may play

a role in keeping solvent out of the active site as they are both hydrophobic in nature. These residues close off the pocket so that the oxidation of the reduced flavin can take place.

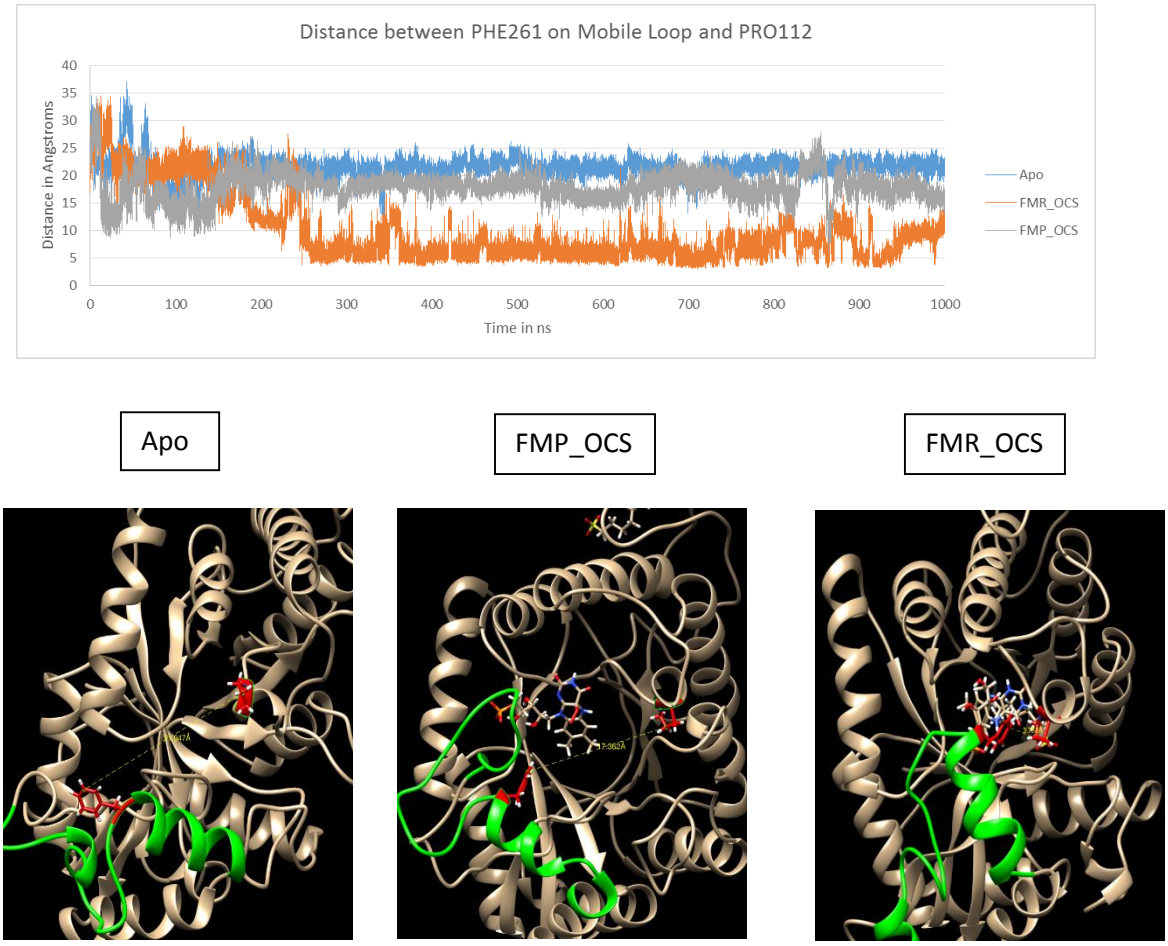
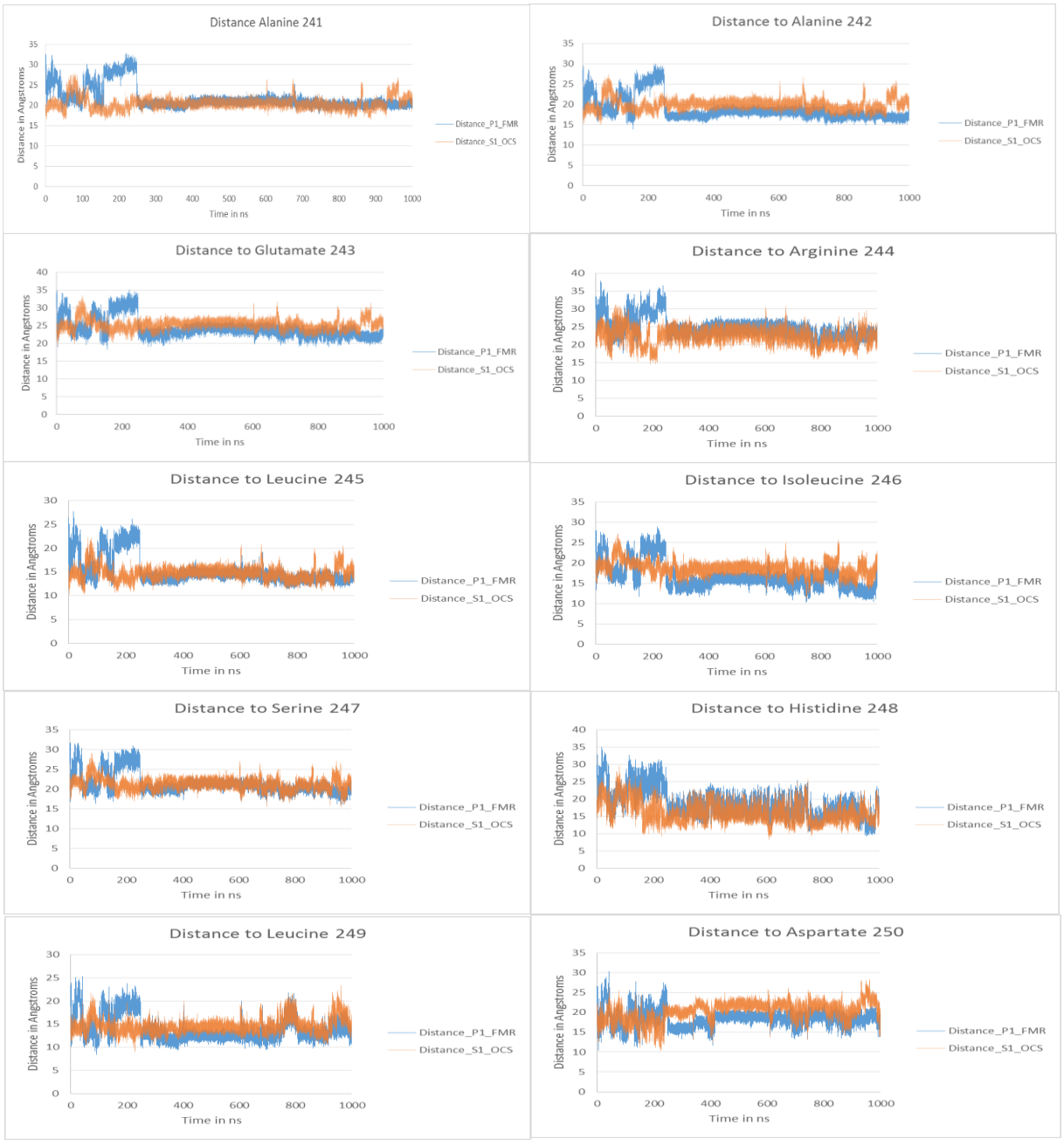
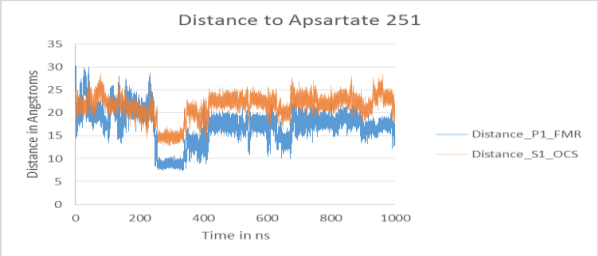


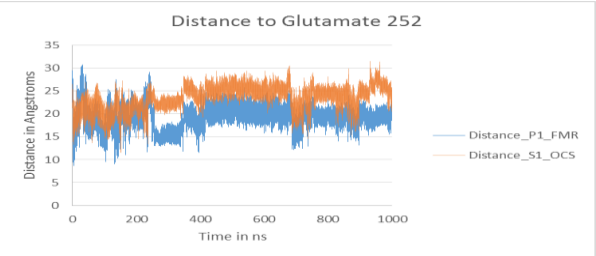
Figure 4.8 Distance between Phe261 (mobile loop) and Pro112. The graph shows the distance over the 1000 ns simulation. Each image was taken at 900 ns and shows the movement of the entire mobile loop.

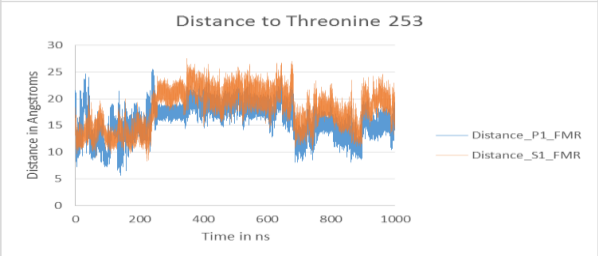
With the mobile loop found to be very dynamic, distance calculations were performed for every amino acid on the mobile loop to the phosphate group on FMR and sulfur group on OCS. The goal was to determine if there were any other residues that may be important to catalysis. It was found that the only residue that seems important to catalysis was the Arg297. All other

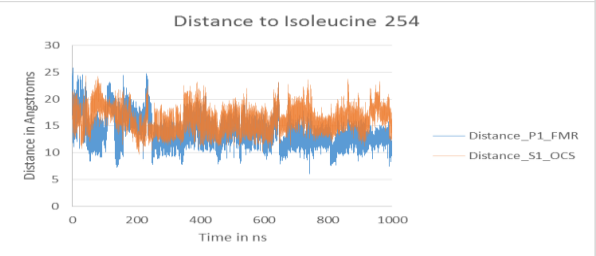
residues had a distance too great to be part in any catalytic reaction. Figure 4.8 shows all the distance calculations.

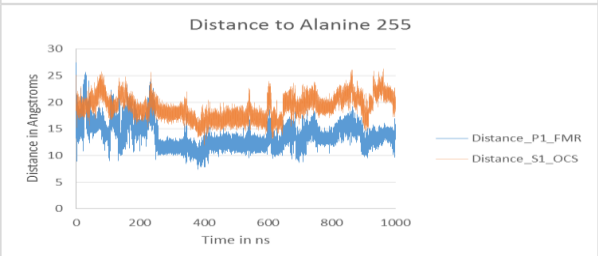


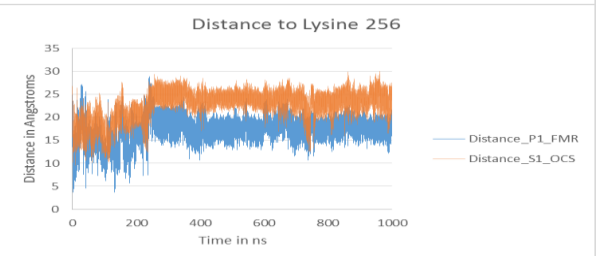

Distance to Apsartate 251
Distance in Angstroms
Time in ns
Distance_P1_FMR
Distance_S1_OCS

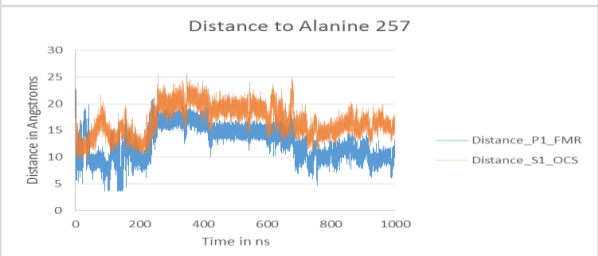

Distance to Glutamate 252
Distance in Angstroms
Time in ns
Distance_P1_FMR
Distance_S1_OCS

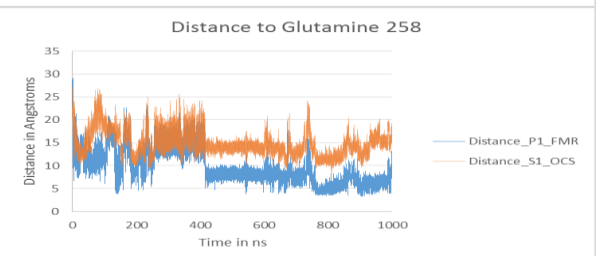

Distance to Threonine 253
Distance in Angstroms
Time in ns
Distance_P1_FMR
Distance_S1_FMR

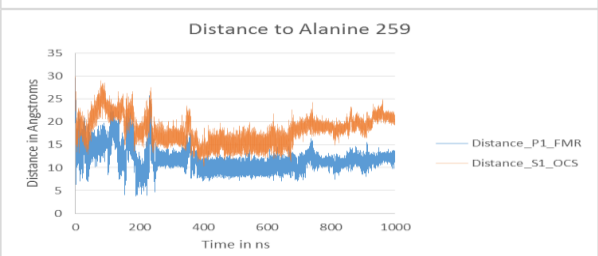

Distance to Isoleucine 254
Distance in Angstroms
Time in ns
Distance_P1_FMR
Distance_S1_OCS

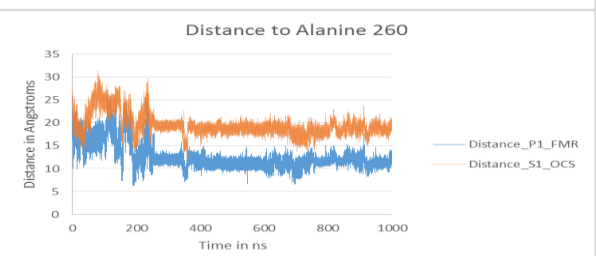

Distance to Alanine 255
Distance in Angstroms
Time in ns
Distance_P1_FMR
Distance_S1_OCS

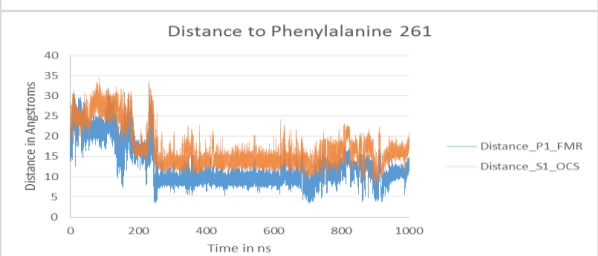

Distance to Lysine 256
Distance in Angstroms
Time in ns
Distance_P1_FMR
Distance_S1_OCS

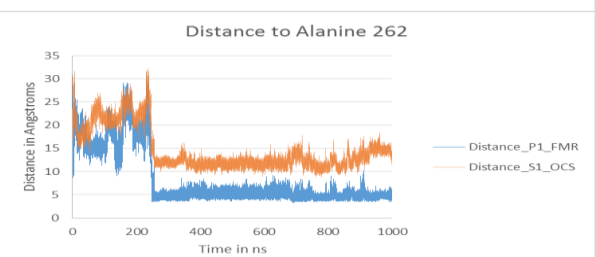

Distance to Alanine 257
Distance in Angstroms
Time in ns
Distance_P1_FMR
Distance_S1_OCS


Distance to Glutamine 258
Distance in Angstroms
Time in ns
Distance_P1_FMR
Distance_S1_OCS

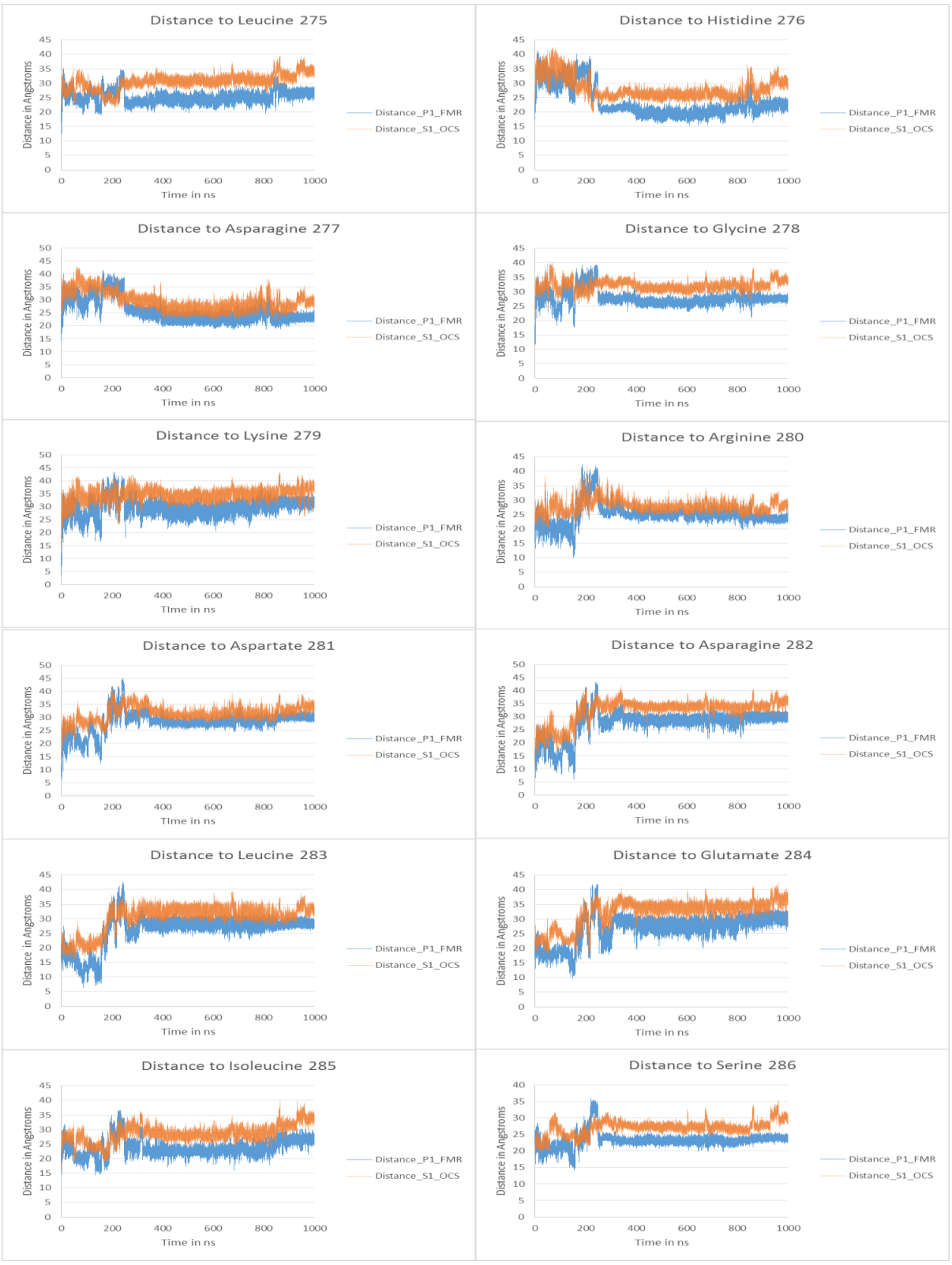

Distance to Alanine 259
Distance in Angstroms
Time in ns
Distance_P1_FMR
Distance_S1_OCS

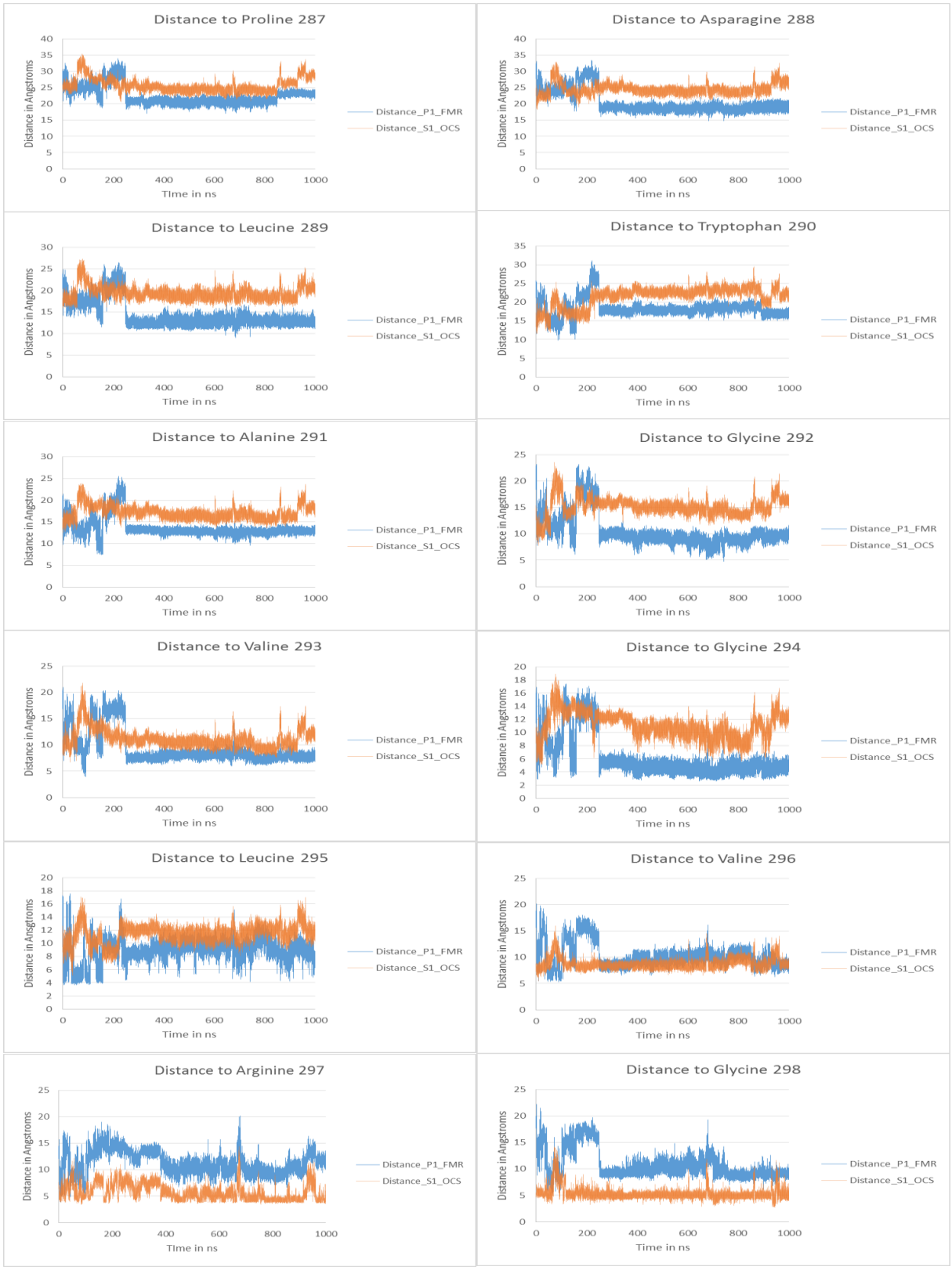

Distance to Alanine 260
Distance in Angstroms
Time in ns
Distance_P1_FMR
Distance_S1_OCS


Distance to Phenylalanine 261
Distance in Angstroms
Time in ns
Distance_P1_FMR
Distance_S1_OCS


Distance to Alanine 262
Distance in Angstroms
Time in ns
Distance_P1_FMR
Distance_S1_OCS


Distance to Arginine 263
Distance to Threonine 264
Distance to Aspartate 265
Distanec to Serine 266
Distance to Valine 267
Distance to Glycine 268
Distance to Glutamine 269
Distance to Glutamine 270
Distance to Arginine 271
Distance to Methionine 272
Distance to Alanine 273
Distance to Alanine 274
Distance in Angstroms
Distanec in Angstroms
Time in ns
TIme in ns
Distance_P1_FMR
Distance_S1_OCS


Distance to Leucine 275
Distance to Histidine 276
Distance to Asparagine 277
Distance to Glycine 278
Distance to Lysine 279
Distance to Arginine 280
Distance to Aspartate 281
Distance to Asparagine 282
Distance to Leucine 283
Distance to Glutamate 284
Distance to Isoleucine 285
Distance to Serine 286
Distance in Angstroms
Time in ns
Distance_P1_FMR
Distance_S1_OCS


Distance to Proline 287
Distance to Asparagine 288
Distance to Leucine 289
Distance to Tryptophan 290
Distance to Alanine 291
Distance to Glycine 292
Distance to Valine 293
Distance to Glycine 294
Distance to Leucine 295
Distance to Valine 296
Distance to Arginine 297
Distance to Glycine 298
Distance in Angstroms
Time in ns
Distance_P1_FMR
Distance_S1_OCS

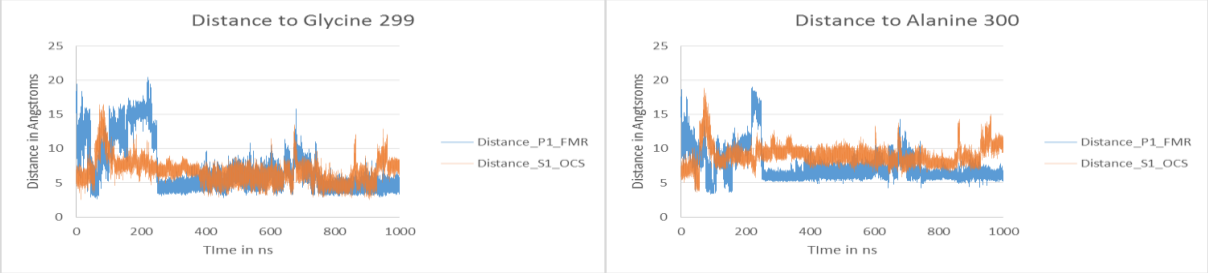


Figure 4.9 Distance calculations for all amino acids on the mobile loops to the phosphate group on FMR and the sulfur group on OCS.

Arg297 is believed to play a substantial role in the desulfonation mechanism.[44] This arginine residue is conserved in all bacterial SsuD enzymes and is located on the mobile loop. In the inactive form (mobile loop open), this arginine is far away from the catalytic site. However, once SsuD is ready for catalysis with both FMR and OCS bound, the mobile loop closes over the active site bringing arginine in range to help with catalysis. Mutational studies have shown that no sulfite product is produced without this residue.[44] Figure 4.9 shows the distance over the simulation for Arg297 for FMR_OCS.

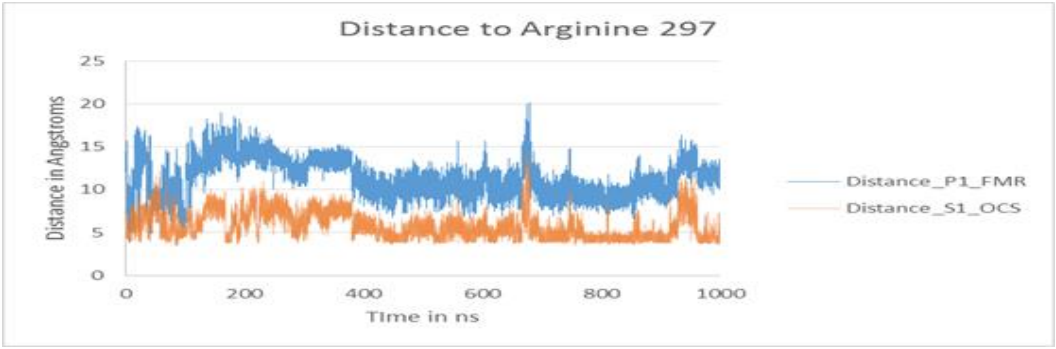


Figure 4.10 Distance calculations between Arg297 for FMR_OCS.

It was originally thought that the Arg297 was interacting with the phosphate group on FMR; however, the distance calculations show that it may be stabilizing the sulfur residue in OCS. The distance calculation with FMP_OCS shows that the OCS is eventually ejected out of the active site.

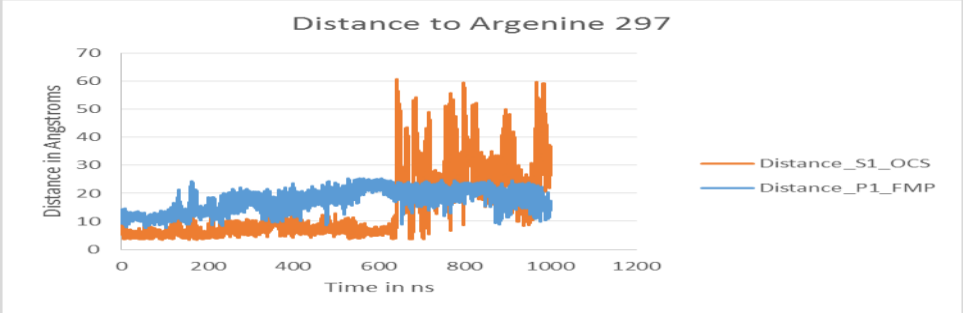


Figure 4.11 Distance calculations between Arg297 for FMP_OCS

The distance for FMR_OCS over the 1000 ns stays around 5 Å for the sulfur group on OCS and closer to 10 Å for the phosphate group on FMR. The FMP_OCS shows a distance of around 15 Å for FMP and eventually between 30-50 Å for OCS. The octanesulfonate is ejected out of the pocket. This may suggest an ordered mechanism on substrate disassociation from SsuD. Once the octanesulfonate leaves, SsuD is then ready to transfer the oxidized flavin back to SsuE. A longer simulation would be needed as FMP was shown to stay in the pocket during the 1000 ns. Figure 4.11 shows a visual representation of the distance between Arg297 and OCS.

A. FMR_OCS B. FMP_OCS

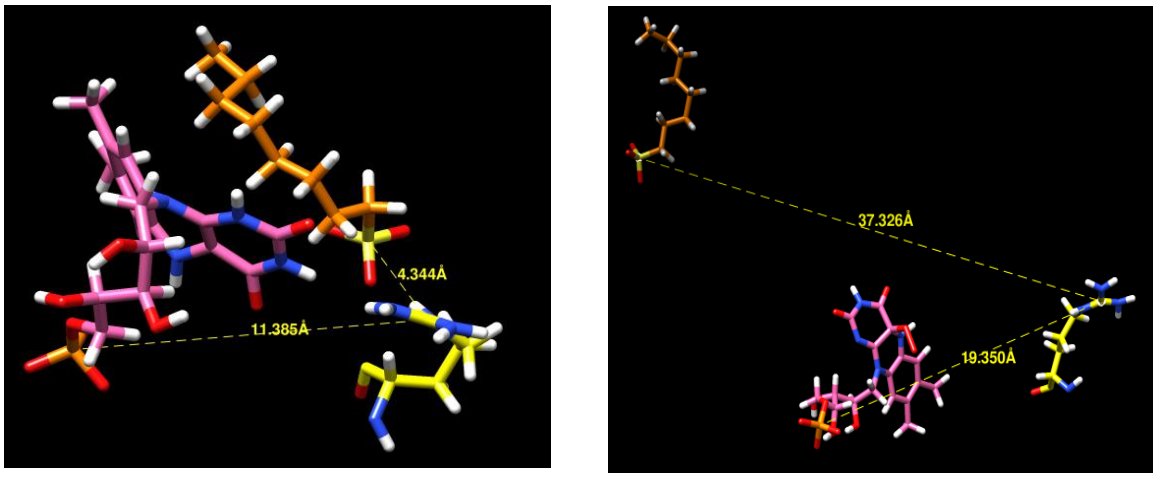


Figure 4.12 A) shows the distance between Arg297 and OCS. The distance was taken at 900 ns and shows a distance of 4.3 Å between Arg297 and OCS. B) Shows the distance of OCS with Arg297. This distance was taken at 900 ns and shows a distance of 37 Å.

Conclusion

This study using accelerated molecular dynamics explored the importance of the mobile loop in SsuD for catalytic activity and in potential flavin transfer. Clustering analysis, PCA analysis, and RMSF analysis examined the different conformations available for the Apo, FMR-OCS bound, and FMP-OCS bound structures. The Apo and FMP_OCS bound have open conformations. The Apo structure is able to accept the substrate with its open conformation while the FMP_OCS is able to eject the substrate. FMR_OCS has a closed conformation in order for catalysis to take place.

Deletion studies of the mobile loop on SsuD showed that the enzyme could still bind with flavin; however, it lost all catalytic activity.[56] Further mutational studies were done on the

conserve Arg297 residue which showed that activity was lost when mutated. It was also postulated that this residue plays an important role in catalysis by anchoring the phosphate group of FMR.[44] From the distance calculations performed in this experiment, the arginine residue appears to be the only residue on the mobile loop that plays a role in catalytic activity. Also, this residue seems to lock in and stabilize the sulfate on octanesulfonate instead of the phosphate group in FMR as originally thought. Once the oxidation reaction is completed, the OCS is ejected out of the active site through an open conformation on SsuD. These simulations also elucidate the mechanism of potential flavin transfer back to SsuE.

References

1. Hunter, T., A thousand and one protein kinases. *Cell* **1987,** *50* (6), 823-829.
2. Hanks, S. K., Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. *Genome Biology* **2003,** *4* (5), 111-111.
3. Lewis, T. S.; Shapiro, P. S.; Ahn, N. G., Signal Transduction through MAP Kinase Cascades. In *Advances in Cancer Research*, George, F. V. W.; George, K., Eds. Academic Press: 1998; Vol. Volume 74, pp 49-139.
4. Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S., The Protein Kinase Complement of the Human Genome. *Science* **2002,** *298* (5600), 1912-1934.
5. Hanks, S. K.; Hunter, T., Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *The FASEB Journal* **1995,** *9* (8), 576-96.
6. Taylor, S. S.; Keshwani, M. M.; Steichen, J. M.; Kornev, A. P., *Evolution of the eukaryotic protein kinases as dynamic molecular switches*. 2012; Vol. 367, p 2517-2528.
7. Steichen, J. M.; Kuchinskas, M.; Keshwani, M. M.; Yang, J.; Adams, J. A.; Taylor, S. S., Structural Basis for the Regulation of Protein Kinase A by Activation Loop Phosphorylation. *The Journal of Biological Chemistry* **2012,** *287* (18), 14672-14680.
8. Coulombe, P.; Meloche, S., Atypical mitogen-activated protein kinases: Structure, regulation and functions. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **2007,** *1773* (8), 1376-1387.
9. Cargnello, M.; Roux, P. P., Activation and Function of the MAPKs and Their Substrates, the MAPK-Activated Protein Kinases. *Microbiology and Molecular Biology Reviews : MMBR* **2011,** *75* (1), 50-83.
10. Davis, R. J., The Mitogen-Activated Protein-Kinase Signal-Transduction Pathway. *J Biol Chem* **1993,** *268* (20), 14553-14556.
11. McCubrey, J. A.; Steelman, L. S.; Chappell, W. H.; Abrams, S. L.; Wong, E. W. T.; Chang, F.; Lehmann, B.; Terrian, D. M.; Milella, M.; Tafuri, A.; Stivala, F.; Libra, M.; Basecke, J.; Evangelisti, C.; Martelli, A. M.; Franklin, R. A., Roles of the Raf/MEK/ERK pathway in cell growth, malignant transformation and drug resistance. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **2007,** *1773* (8), 1263-1284.
12. Yan, J.; Roy, S.; Apolloni, A.; Lane, A.; Hancock, J. F., Ras Isoforms Vary in Their Ability to Activate Raf-1 and Phosphoinositide 3-Kinase. *J Biol Chem* **1998,** *273* (37), 24052-24056.
13. Chong, H.; Vikis, H. G.; Guan, K.-L., Mechanisms of regulating the Raf kinase family. *Cellular Signalling* **2003,** *15* (5), 463-469.
14. Leicht, D. T.; Balan, V.; Kaplun, A.; Singh-Gupta, V.; Kaplun, L.; Dobson, M.; Tzivion, G., Raf Kinases: Function, Regulation and Role in Human Cancer. *Biochimica et biophysica acta* **2007,** *1773* (8), 1196-1212.
15. Mercer, K. E.; Pritchard, C. A., Raf proteins and cancer: B-Raf is identified as a mutational target. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **2003,** *1653* (1), 25-40.
16. Freeman, A. K.; Ritt, D. A.; Morrison, D. K., The importance of Raf dimerization in cell signaling. *Small GTPases* **2013,** *4* (3), 180-185.
17. Roskoski, R., Jr., RAF protein-serine/threonine kinases: structure and regulation. *Biochemical and biophysical research communications* **2010,** *399* (3), 313-7.

18. El-Nassan, H. B., Recent progress in the identification of BRAF inhibitors as anti-cancer agents. *European Journal of Medicinal Chemistry* **2014,** *72* (0), 170-205.
19. Hoeflich, K. P.; Herter, S.; Tien, J.; Wong, L.; Berry, L.; Chan, J.; O'Brien, C.; Modrusan, Z.; Seshagiri, S.; Lackner, M.; Stern, H.; Choo, E.; Murray, L.; Friedman, L. S.; Belvin, M., Antitumor Efficacy of the Novel RAF Inhibitor GDC-0879 Is Predicted by BRAF(V600E) Mutational Status and Sustained Extracellular Signal-Regulated Kinase/Mitogen-Activated Protein Kinase Pathway Suppression. *Cancer Res* **2009,** *69* (7), 3042-3051.
20. Wan, P. T. C.; Garnett, M. J.; Roe, S. M.; Lee, S.; Niculescu-Duvaz, D.; Good, V. M.; Project, C. G.; Jones, C. M.; Marshall, C. J.; Springer, C. J.; Barford, D.; Marais, R., Mechanism of Activation of the RAF-ERK Signaling Pathway by Oncogenic Mutations of B-RAF. *Cell 116* (6), 855-867.
21. Melanoma Treatment (PDQ). http://www.cancer.gov/types/skin/patient/melanoma-treatment-pdq#section/all.
22. Guide to Staging - Melanoma. http://www.skincancer.org/skin-cancer-information/melanoma/the-stages-of-melanoma/guide-to-staging-melanoma.
23. Diseases and Conditions: Melanoma. http://www.mayoclinic.org/diseases-conditions/melanoma/basics/causes/con-20026009.
24. Bollag, G.; Hirth, P.; Tsai, J.; Zhang, J.; Ibrahim, P. N.; Cho, H.; Spevak, W.; Zhang, C.; Zhang, Y.; Habets, G.; Burton, E. A.; Wong, B.; Tsang, G.; West, B. L.; Powell, B.; Shellooe, R.; Marimuthu, A.; Nguyen, H.; Zhang, K. Y.; Artis, D. R.; Schlessinger, J.; Su, F.; Higgins, B.; Iyer, R.; D'Andrea, K.; Koehler, A.; Stumm, M.; Lin, P. S.; Lee, R. J.; Grippo, J.; Puzanov, I.; Kim, K. B.; Ribas, A.; McArthur, G. A.; Sosman, J. A.; Chapman, P. B.; Flaherty, K. T.; Xu, X.; Nathanson, K. L.; Nolop, K., Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature* **2010,** *467* (7315), 596-9.
25. Wilson, T. R.; Fridlyand, J.; Yan, Y.; Penuel, E.; Burton, L.; Chan, E.; Peng, J.; Lin, E.; Wang, Y.; Sosman, J.; Ribas, A.; Li, J.; Moffat, J.; Sutherlin, D. P.; Koeppen, H.; Merchant, M.; Neve, R.; Settleman, J., Widespread potential for growth-factor-driven resistance to anticancer kinase inhibitors. *Nature* **2012,** *487* (7408), 505-509.
26. Liu, Y.; Gray, N. S., Rational design of inhibitors that bind to inactive kinase conformations. *Nat Chem Biol* **2006,** *2* (7), 358-364.
27. Zhao, Z.; Wu, H.; Wang, L.; Liu, Y.; Knapp, S.; Liu, Q.; Gray, N. S., Exploration of Type II Binding Mode: A Privileged Approach for Kinase Inhibitor Focused Drug Discovery? *ACS Chemical Biology* **2014,** *9* (6), 1230-1241.
28. Janne, P. A.; Gray, N.; Settleman, J., Factors underlying sensitivity of cancers to small-molecule kinase inhibitors. *Nat Rev Drug Discov* **2009,** *8* (9), 709-723.
29. Jambrina, P. G.; Bohuszewicz, O.; Buchete, N. V.; Kolch, W.; Rosta, E., Molecular mechanisms of asymmetric RAF dimer activation. *Biochem Soc Trans* **2014,** *42* (4), 784-90.
30. Zuccotto, F.; Ardini, E.; Casale, E.; Angiolini, M., Through the "gatekeeper door": exploiting the active kinase conformation. *J Med Chem* **2010,** *53* (7), 2681-94.
31. Nazarian, R.; Shi, H.; Wang, Q.; Kong, X.; Koya, R. C.; Lee, H.; Chen, Z.; Lee, M.-K.; Attar, N.; Sazegar, H.; Chodon, T.; Nelson, S. F.; McArthur, G.; Sosman, J. A.; Ribas, A.; Lo, R. S., Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature* **2010,** *468* (7326), 973-977.
32. Choi, J.; Landrette, S. F.; Wang, T.; Evans, P.; Bacchiocchi, A.; Bjornson, R.; Cheng, E.; Stiegler, A. L.; Gathiaka, S.; Acevedo, O.; Boggon, T. J.; Krauthammer, M.; Halaban, R.; Xu, T.,

Identification of PLX4032-resistance mechanisms and implications for novel RAF inhibitors. *Pigment cell & melanoma research* **2014,** *27* (2), 253-62.
33. Torres Pazmiño, D. E.; Winkler, M.; Glieder, A.; Fraaije, M. W., Monooxygenases as biocatalysts: Classification, mechanistic aspects and biotechnological applications. *Journal of Biotechnology* **2010,** *146* (1–2), 9-24.
34. Ellis, H. R., The FMN-dependent two-component monooxygenase systems. *Archives of Biochemistry and Biophysics* **2010,** *497* (1–2), 1-12.
35. Becvar, J. E.; Hastings, J. W., Bacterial luciferase requires one reduced flavin for light emission. *Proceedings of the National Academy of Sciences* **1975,** *72* (9), 3374-3376.
36. Li, L.; Liu, X.; Yang, W.; Xu, F.; Wang, W.; Feng, L.; Bartlam, M.; Wang, L.; Rao, Z., Crystal Structure of Long-Chain Alkane Monooxygenase (LadA) in Complex with Coenzyme FMN: Unveiling the Long-Chain Alkane Hydroxylase. *Journal of Molecular Biology* **2008,** *376* (2), 453-465.
37. Bohuslavek, J.; Payne, J. W.; Liu, Y.; Bolton, H.; Xun, L., Cloning, Sequencing, and Characterization of a Gene Cluster Involved in EDTA Degradation from the Bacterium BNC1. *Applied and Environmental Microbiology* **2001,** *67* (2), 688-695.
38. Uetz, T.; Schneider, R.; Snozzi, M.; Egli, T., Purification and characterization of a two-component monooxygenase that hydroxylates nitrilotriacetate from "Chelatobacter" strain ATCC 29600. *Journal of Bacteriology* **1992,** *174* (4), 1179-1188.
39. Sciara, G.; Kendrew, S. G.; Miele, A. E.; Marsh, N. G.; Federici, L.; Malatesta, F.; Schimperna, G.; Savino, C.; Vallone, B., The structure of ActVA-Orf6, a novel type of monooxygenase involved in actinorhodin biosynthesis. *The EMBO Journal* **2003,** *22* (2), 205-215.
40. Blanc, V.; Lagneaux, D.; Didier, P.; Gil, P.; Lacroix, P.; Crouzet, J., Cloning and analysis of structural genes from Streptomyces pristinaespiralis encoding enzymes involved in the conversion of pristinamycin IIB to pristinamycin IIA (PIIA): PIIA synthase and NADH:riboflavin 5'-phosphate oxidoreductase. *Journal of Bacteriology* **1995,** *177* (18), 5206-5214.
41. Galán, B.; Díaz, E.; García, J. L., Enhancing desulphurization by engineering a flavin reductase-encoding gene cassette in recombinant biocatalysts. *Environmental Microbiology* **2000,** *2* (6), 687-694.
42. Gao, B.; Ellis, H. R., Altered mechanism of the alkanesulfonate FMN reductase with the monooxygenase enzyme. *Biochemical and biophysical research communications* **2005,** *331* (4), 1137-1145.
43. Ellis, H. R., Mechanism for sulfur acquisition by the alkanesulfonate monooxygenase system. *Bioorganic Chemistry* **2011,** *39* (5–6), 178-184.
44. Carpenter, R. A.; Xiong, J.; Robbins, J. M.; Ellis, H. R., Functional Role of a Conserved Arginine Residue Located on a Mobile Loop of Alkanesulfonate Monooxygenase. *Biochemistry* **2011,** *50* (29), 6469-6477.
45. Gao, B.; Ellis, H. R., Mechanism of flavin reduction in the alkanesulfonate monooxygenase system. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2007,** *1774* (3), 359-367.
46. Zhan, X.; Carpenter, R. A.; Ellis, H. R., Catalytic Importance of the Substrate Binding Order for the FMNH2-Dependent Alkanesulfonate Monooxygenase Enzyme†. *Biochemistry* **2008,** *47* (7), 2221-2230.

47. Eichhorn, E.; Davey, C. A.; Sargent, D. F.; Leisinger, T.; Richmond, T. J., Crystal Structure of Escherichia coli Alkanesulfonate Monooxygenase SsuD. *Journal of Molecular Biology* **2002,** *324* (3), 457-468.
48. Massey, V., The Chemical and Biological Versatility of Riboflavin. *Biochemical Society Transactions* **2000,** *28* (4), 283-296.
49. Carpenter, R. A.; Zhan, X.; Ellis, H. R., Catalytic role of a conserved cysteine residue in the desulfonation reaction by the alkanesulfonate monooxygenase enzyme. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2010,** *1804* (1), 97-105.
50. Robbins, J. M.; Ellis, H. R., Identification of Critical Steps Governing the Two-Component Alkanesulfonate Monooxygenase Catalytic Mechanism. *Biochemistry* **2012,** *51* (32), 6378-6387.
51. Xin, X.; Xi, L.; Tu, S. C., Functional consequences of site-directed mutation of conserved histidyl residues of the bacterial luciferase .alpha. subunit. *Biochemistry* **1991,** *30* (47), 11255-11262.
52. Robbins, J. M.; Ellis, H. R., Steady-State Kinetic Isotope Effects Support a Complex Role of Arg226 in the Proposed Desulfonation Mechanism of Alkanesulfonate Monooxygenase. *Biochemistry* **2014,** *53* (1), 161-168.
53. Malabanan, M. M.; Amyes, T. L.; Richard, J. P., A Role for Flexible Loops in Enzyme Catalysis. *Current opinion in structural biology* **2010,** *20* (6), 702-710.
54. Ferrario, V.; Braiuca, P.; Tessaro, P.; Knapic, L.; Gruber, C.; Pleiss, J.; Ebert, C.; Eichhorn, E.; Gardossi, L., Elucidating the structural and conformational factors responsible for the activity and substrate specificity of alkanesulfonate monooxygenase. *Journal of Biomolecular Structure and Dynamics* **2012,** *30* (1), 74-88.
55. Armacost, K.; Musila, J.; Gathiaka, S.; Ellis, H. R.; Acevedo, O., Exploring the Catalytic Mechanism of Alkanesulfonate Monooxygenase Using Molecular Dynamics. *Biochemistry* **2014,** *53* (20), 3308-3317.
56. Xiong, J.; Ellis, H. R., Deletional studies to investigate the functional role of a dynamic loop region of alkanesulfonate monooxygenase. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2012,** *1824* (7), 898-906.
57. Brandsdal, B. O.; Österberg, F.; Almlöf, M.; Feierberg, I.; Luzhkov, V. B.; Åqvist, J., Free Energy Calculations and Ligand Binding. In *Advances in Protein Chemistry*, Academic Press: 2003; Vol. Volume 66, pp 123-158.
58. Singh, N.; Warshel, A., Absolute Binding Free Energy Calculations: On the Accuracy of Computational Scoring of Protein-ligand Interactions. *Proteins* **2010,** *78* (7), 1705-1723.
59. Fogolari, F.; Brigo, A.; Molinari, H., The Poisson–Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *Journal of Molecular Recognition* **2002,** *15* (6), 377-392.
60. Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *Journal of Computational Chemistry* **2011,** *32* (5), 866-877.
61. Gohlke, H.; Case, D. A., Converging free energy estimates: MM-PB(GB)SA studies on the protein–protein complex Ras–Raf. *Journal of Computational Chemistry* **2004,** *25* (2), 238-250.
62. Guvench, O.; MacKerell, A. D., Computational evaluation of protein – small molecule binding. *Current opinion in structural biology* **2009,** *19* (1), 56-61.

63. Spitzer, R.; Jain, A. N., Surflex-Dock: Docking Benchmarks and Real-World Application. *Journal of Computer-Aided Molecular Design* **2012,** *26* (6), 687-699.
64. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *Journal of computational chemistry* **2009,** *30* (16), 2785-2791.
65. Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D., Improved protein–ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics* **2003,** *52* (4), 609-623.
66. Trott, O.; Olson, A. J., Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *Journal of Computational Chemistry* **2010,** *31* (2), 455-461.
67. Rastelli, G.; Del Rio, A.; Degliesposti, G.; Sgobba, M., Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J Comput Chem* **2010,** *31* (4), 797-810.
68. Wang, R.; Lu, Y.; Wang, S., Comparative Evaluation of 11 Scoring Functions for Molecular Docking. *Journal of Medicinal Chemistry* **2003,** *46* (12), 2287-2303.
69. Alder, B. J.; Wainwright, T. E., Phase Transition for a Hard Sphere System. *The Journal of Chemical Physics* **1957,** *27* (5), 1208-1209.
70. Rahman, A., Correlations in the Motion of Atoms in Liquid Argon. *Physical Review* **1964,** *136* (2A), A405-A411.
71. McCammon, J. A.; Gelin, B. R.; Karplus, M., Dynamics of folded proteins. *Nature* **1977,** *267* (5612), 585-590.
72. Petrenko, R.; Meller, J., Molecular Dynamics. In *eLS*, John Wiley & Sons, Ltd: 2001.
73. Paquet, E.; Viktor, H. L., Molecular Dynamics, Monte Carlo Simulations, and Langevin Dynamics: A Computational Review. *BioMed Research International* **2015,** *2015*, 18.
74. Verlet, L., Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **1967,** *159* (1), 98-103.
75. Grubmüller, H.; Heller, H.; Windemuth, A.; Schulten, K., Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-range Interactions. *Molecular Simulation* **1991,** *6* (1-3), 121-142.
76. Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R., A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics* **1982,** *76* (1), 637-649.
77. Leach, A., *Molecular Modelling: Principles and Applications (2nd Edition)*. Prentice Hall: 2001.
78. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., 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.
79. LeRoux, S., Petkov, V. Model Box Periodic Boundary Conditions - P.B.C. isaacs.sourceforge.net/phys/pbc.html.
80. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E., Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* **1953,** *21* (6), 1087-1092.
81. Ewald, P. P., Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* **1921,** *369* (3), 253-287.

82. Allen, M. P.; Tildesley, D. J., *Computer Simulation of Liquids (Oxford Science Publications)*. Oxford University Press: 1989.
83. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993,** *98* (12), 10089-10092.
84. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **2005,** *26* (16), 1668-1688.
85. Pierce, L. C.; Salomon-Ferrer, R.; Augusto, F. d. O. C.; McCammon, J. A.; Walker, R. C., Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *Journal of chemical theory and computation* **2012,** *8* (9), 2997-3002.
86. Hamelberg, D.; Mongan, J.; McCammon, J. A., Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *The Journal of Chemical Physics* **2004,** *120* (24), 11919-11929.
87. D.A. Case, V. B., J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu and P.A. Kollman *AMBER 14*, University California, San Francisco, 2014.
88. Cramer, C. J., *Essentials of Computational Chemistry: Theories and Models (2nd Edition)*. John Wiley & Sons: 2004.
89. Mongan, J.; Case, D. A., Biomolecular simulations at constant pH. *Current Opinion in Structural Biology* **2005,** *15* (2), 157-163.
90. Mongan, J.; Case, D. A.; McCammon, J. A., Constant pH molecular dynamics in generalized Born implicit solvent. *Journal of Computational Chemistry* **2004,** *25* (16), 2038-2048.
91. Hotta, T., Mean-Field Approximation. In *Nanoscale Phase Separation and Colossal Magnetoresistance*, Springer Berlin Heidelberg: 2003; Vol. 136, pp 157-167.
92. Kong, X.; Brooks, C. L., λ-dynamics: A new approach to free energy calculations. *The Journal of Chemical Physics* **1996,** *105* (6), 2414-2423.
93. Chen, J.; Brooks, C. L.; Khandogin, J., Recent advances in implicit solvent based methods for biomolecular simulations. *Current opinion in structural biology* **2008,** *18* (2), 140-148.
94. Baptista, A. M.; Teixeira, V. H.; Soares, C. M., Constant-pH molecular dynamics using stochastic titration. *The Journal of Chemical Physics* **2002,** *117* (9), 4184-4200.
95. Williams, S. L.; de Oliveira, C. A. F.; McCammon, J. A., Coupling Constant pH Molecular Dynamics with Accelerated Molecular Dynamics. *Journal of chemical theory and computation* **2010,** *6* (2), 560-568.
96. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham Iii, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **1995,** *91* (1–3), 1-41.
97. Salomon-Ferrer, R.; Case, D. A.; Walker, R. C., An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2013,** *3* (2), 198-210.

98. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry* **1993,** *97* (40), 10269-10280.
99. Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of Computational Chemistry* **2002,** *23* (16), 1623-1641.
100. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004,** *25* (9), 1157-1174.
101. Jackson, J. E., Getting Started. In *A User's Guide to Principal Components*, John Wiley & Sons, Inc.: 2004; pp 4-25.
102. Cordella, C. B. Y., *PCA: The Basic Building Block of Chemometrics*. 2012.
103. Ballabio, D., A MATLAB toolbox for Principal Component Analysis and unsupervised exploration of data structure. *Chemometrics and Intelligent Laboratory Systems* **2015,** *149, Part B*, 1-9.
104. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of chemical theory and computation* **2013,** *9* (7), 3084-3095.
105. Trédan, O.; Galmarini, C. M.; Patel, K.; Tannock, I. F., Drug Resistance and the Solid Tumor Microenvironment. *Journal of the National Cancer Institute* **2007,** *99* (19), 1441-1454.
106. Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M.-y.; Pieper, U.; Sali, A., Comparative Protein Structure Modeling Using MODELLER. In *Current Protocols in Protein Science*, John Wiley & Sons, Inc.: 2001.
107. Pettersen, E.; Goddard, T.; Huang, C.; Couch, G.; Greenblatt, D.; Meng, E.; Ferrin, T., UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004,** *25* (13), 1605-1612.
108. Salomon-Ferrer, R.; Case, D. A.; Walker, R. C., An overview of the Amber biomolecular simulation package. *Wires Comput Mol Sci* **2013,** *3* (2), 198-210.
109. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **2006,** *65* (3), 712-725.
110. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983,** *79* (2), 926-935.
111. Watowich, S. J.; Meyer, E. S.; Hagstrom, R.; Josephs, R., A stable, rapidly converging conjugate gradient method for energy minimization. *Journal of Computational Chemistry* **1988,** *9* (6), 650-661.
112. Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *Journal of chemical theory and computation* **2012,** *8* (5), 1542-1555.
113. (a) Poulikakos, P. I.; Persaud, Y.; Janakiraman, M.; Kong, X. J.; Ng, C.; Moriceau, G.; Shi, H. B.; Atefi, M.; Titz, B.; Gabay, M. T.; Salton, M.; Dahlman, K. B.; Tadi, M.; Wargo, J. A.; Flaherty, K. T.; Kelley, M. C.; Misteli, T.; Chapman, P. B.; Sosman, J. A.; Graeber, T. G.; Ribas, A.; Lo, R. S.; Rosen, N.; Solit, D. B., RAF inhibitor resistance is mediated by dimerization of aberrantly spliced BRAF(V600E). *Nature* **2011,** *480* (7377), 387-U144; (b) Sosman, J. A.; Kim, K. B.; Schuchter, L.; Gonzalez, R.; Pavlick, A. C.; Weber, J. S.; McArthur,

G. A.; Hutson, T. E.; Moschos, S. J.; Flaherty, K. T.; Hersey, P.; Kefford, R.; Lawrence, D.; Puzanov, I.; Lewis, K. D.; Amaravadi, R. K.; Chmielowski, B.; Lawrence, H. J.; Shyr, Y.; Ye, F.; Li, J.; Nolop, K. B.; Lee, R. J.; Joe, A. K.; Ribas, A., Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. *The New England journal of medicine* **2012,** *366* (8), 707-14.
114. Tsai, J.; Lee, J. T.; Wang, W.; Zhang, J.; Cho, H.; Mamo, S.; Bremer, R.; Gillette, S.; Kong, J.; Haass, N. K.; Sproesser, K.; Li, L.; Smalley, K. S.; Fong, D.; Zhu, Y. L.; Marimuthu, A.; Nguyen, H.; Lam, B.; Liu, J.; Cheung, I.; Rice, J.; Suzuki, Y.; Luu, C.; Settachatgul, C.; Shellooe, R.; Cantwell, J.; Kim, S. H.; Schlessinger, J.; Zhang, K. Y.; West, B. L.; Powell, B.; Habets, G.; Zhang, C.; Ibrahim, P. N.; Hirth, P.; Artis, D. R.; Herlyn, M.; Bollag, G., Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proceedings of the National Academy of Sciences of the United States of America* **2008,** *105* (8), 3041-6.
115. Robert, C.; Karaszewska, B.; Schachter, J.; Rutkowski, P.; Mackiewicz, A.; Stroiakovski, D.; Lichinitser, M.; Dummer, R.; Grange, F.; Mortier, L.; Chiarion-Sileni, V.; Drucis, K.; Krajsova, I.; Hauschild, A.; Lorigan, P.; Wolter, P.; Long, G. V.; Flaherty, K.; Nathan, P.; Ribas, A.; Martin, A.-M.; Sun, P.; Crist, W.; Legos, J.; Rubin, S. D.; Little, S. M.; Schadendorf, D., Improved Overall Survival in Melanoma with Combined Dabrafenib and Trametinib. *New England Journal of Medicine* **2015,** *372* (1), 30-39.
116. Montagut, C.; Sharma, S. V.; Shioda, T.; McDermott, U.; Ulman, M.; Ulkus, L. E.; Dias-Santagata, D.; Stubbs, H.; Lee, D. Y.; Singh, A.; Drew, L.; Haber, D. A.; Settleman, J., Elevated CRAF as a potential mechanism of acquired resistance to BRAF inhibition in melanoma. *Cancer Res* **2008,** *68* (12), 4853-4861.
117. James, J.; Ruggeri, B.; Armstrong, R. C.; Rowbottom, M. W.; Jones-Bolin, S.; Gunawardane, R. N.; Dobrzanski, P.; Gardner, M. F.; Zhao, H.; Cramer, M. D.; Hunter, K.; Nepomuceno, R. R.; Cheng, M. G.; Gitnick, D.; Yazdanian, M.; Insko, D. E.; Ator, M. A.; Apuy, J. L.; Faraoni, R.; Dorsey, B. D.; Williams, M.; Bhagwat, S. S.; Holladay, M. W., CEP-32496: A Novel Orally Active BRAF(V600E) Inhibitor with Selective Cellular and In Vivo Antitumor Activity. *Mol Cancer Ther* **2012,** *11* (4), 930-941.
118. Cunniff, E. G. C.; Zhang, J.; Chouitar, J.; Mettetal, J.; Nakamura, K.; Arita, T.; Nakamura, A.; Okaniwa, M.; Ishikawa, T.; Yoshida, S.; Fabrey, R.; Vincent, P.; Eng, K.; Garcia, K.; Borelli, D.; Vagrhese, T.; Stroud, S.; Menon, S.; Kuranda, M.; Galvin, K., Abstract C146: Combination treatment with the investigational RAF kinase inhibitor MLN2480 and the investigational MEK kinase inhibitor TAK-733 inhibits the growth of BRAF mutant and RAS mutant preclinical models of melanoma and CRC. *Mol Cancer Ther* **2013,** *12* (11 Supplement), C146.
119. Su, Y.; Vilgelm, A. E.; Kelley, M. C.; Hawkins, O. E.; Liu, Y.; Boyd, K. L.; Kantrow, S.; Splittgerber, R. C.; Short, S. P.; Sobolik, T.; Zaja-Milatovic, S.; Dahlman, K. B.; Amiri, K. I.; Jiang, A.; Lu, P.; Shyr, Y.; Stuart, D. D.; Levy, S.; Sosman, J. A.; Richmond, A., RAF265 Inhibits the Growth of Advanced Human Melanoma Tumors. *Clinical Cancer Research* **2012,** *18* (8), 2184-2198.
120. Wilhelm, S. M.; Dumas, J.; Adnane, L.; Lynch, M.; Carter, C. A.; Schütz, G.; Thierauch, K.-H.; Zopf, D., Regorafenib (BAY 73-4506): A new oral multikinase inhibitor of angiogenic, stromal and oncogenic receptor tyrosine kinases with potent preclinical antitumor activity. *International Journal of Cancer* **2011,** *129* (1), 245-255.

121. Mangana, J.; Levesque, M. P.; Karpova, M. B.; Dummer, R., Sorafenib in melanoma. *Expert Opinion on Investigational Drugs* **2012,** *21* (4), 557-568.
122. Holderfield, M.; Nagel, T. E.; Stuart, D. D., Mechanism and consequences of RAF kinase activation by small-molecule inhibitors. *Br J Cancer* **2014,** *111* (4), 640-645.
123. Edgcomb, S. P.; Murphy, K. P., Variability in the pKa of histidine side-chains correlates with burial within proteins. *Proteins: Structure, Function, and Bioinformatics* **2002,** *49* (1), 1-6.
124. Blanc, J.; Geney, R.; Menet, C., Type II kinase inhibitors: an opportunity in cancer for rational design. *Anticancer Agents Med Chem* **2013,** *13* (5), 731-47.
125. Kornev, A. P.; Taylor, S. S.; Ten Eyck, L. F., A helix scaffold for the assembly of active protein kinases. *Proceedings of the National Academy of Sciences* **2008,** *105* (38), 14377-14382.
126. Lavoie, H.; Therrien, M., Regulation of RAF protein kinases in ERK signalling. *Nat Rev Mol Cell Biol* **2015,** *16* (5), 281-298.
127. Hu, J.; Stites, Edward C.; Yu, H.; Germino, Elizabeth A.; Meharena, Hiruy S.; Stork, Philip J. S.; Kornev, Alexandr P.; Taylor, Susan S.; Shaw, Andrey S., Allosteric Activation of Functionally Asymmetric RAF Kinase Dimers. *Cell* **2013,** *154* (5), 1036-1046.
128. Cowan-Jacob, S. W.; Fendrich, G.; Manley, P. W.; Jahnke, W.; Fabbro, D.; Liebetanz, J.; Meyer, T., The Crystal Structure of a c-Src Complex in an Active Conformation Suggests Possible Steps in c-Src Activation. *Structure* **2005,** *13* (6), 861-871.
129. LaFevre-Bernt, M.; Sicheri, F.; Pico, A.; Porter, M.; Kuriyan, J.; Miller, W. T., Intramolecular Regulatory Interactions in the Src Family Kinase Hck Probed by Mutagenesis of a Conserved Tryptophan Residue. *J Biol Chem* **1998,** *273* (48), 32129-32134.
130. Joseph, R. E.; Min, L.; Andreotti, A. H., The Linker between SH2 and Kinase Domains Positively Regulates Catalysis of the Tec Family Kinases†. *Biochemistry* **2007,** *46* (18), 5455-5462.
131. Baljuls, A.; Mahr, R.; Schwarzenau, I.; Müller, T.; Polzien, L.; Hekman, M.; Rapp, U. R., Single Substitution within the RKTR Motif Impairs Kinase Activity but Promotes Dimerization of RAF Kinase. *The Journal of Biological Chemistry* **2011,** *286* (18), 16491-16503.
132. Roring, M.; Herr, R.; Fiala, G. J.; Heilmann, K.; Braun, S.; Eisenhardt, A. E.; Halbach, S.; Capper, D.; von Deimling, A.; Schamel, W. W.; Saunders, D. N.; Brummer, T., Distinct requirement for an intact dimer interface in wild-type, V600E and kinase-dead B-Raf signalling. *EMBO J* **2012,** *31* (11), 2629-47.
133. Hugo, L.; Neroshan, T.; Gwenaëlle, G.; John, J. L.; Abbas, P.; Sébastien, G.; Jean, D.; Daniel, Y. L. M.; Michel, B.; Frank, S.; Marc, T., Inhibitors that stabilize a closed RAF kinase domain conformation induce dimerization. *Nature Chemical Biology* **2013,** *9* (7), 428-436.
134. Jorgensen, W. L.; Tirado–Rives, J., Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *Journal of Computational Chemistry* **2005,** *26* (16), 1689-1700.
135. Jorgensen, W. L., OPLS Force Fields. In *Encyclopedia of Computational Chemistry*, John Wiley & Sons, Ltd: 2002.
136. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene,

M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., Gaussian 09, Revision B.01. Wallingford CT, 2009.
137. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of chemical theory and computation* **2013,** *9* (9), 3878-3888.
138. Campbell, Z. T.; Baldwin, T. O., Two Lysine Residues in the Bacterial Luciferase Mobile Loop Stabilize Reaction Intermediates. *J Biol Chem* **2009,** *284* (47), 32827-32834.
139. Biedermannova, L.; E. Riley, K.; Berka, K.; Hobza, P.; Vondrasek, J., Another role of proline: stabilization interactions in proteins and protein complexes concerning proline and tryptophane. *Physical Chemistry Chemical Physics* **2008,** *10* (42), 6350-6359.