BIOINFORMATICS BASED APPROACH TO DESIGN A THERMOPHILIC P450 FOR INDUSTRIAL BIOCATALYSIS

A Thesis Submitted to
the Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in Bioengineering

by
Ekin KESTEVUR DOĞRU

December 2019
İZMİR

ACKNOWLEDGMENTS

First and foremost, I would like to thank my supervisor Assist. Prof. Dr. Nur Başak SÜRMELİ ERALTUĞ for her help, care and support throughout this project. It would not be possible to come this far without her continuedly and ingeniously guidance.

I would like to express my gratitude to Sürmeli Laboratory members; Muhammet Semih BAŞLAR, Gülce GÜRALP, Fatmanur BOSTAN, Tuğçe SAKALLI, Dilara TİLKİOĞLU and Alper ŞAHİN for their precious friendship and generous help.

I am thankful to my office mates; Eyüp BİLGİ, Gülten KURU, Yiğit Ege ÇÖMLEKÇİ, Öykü SARIGİL and Gamze DOĞAN for their support and friendship.

Finally, I am thankful to my mother and sister; Neriman KESTEVUR and Ezgi KESTEVUR who helped in their special way. The most special thanks go to my husband and my best friend Cengiz DOĞRU for his endless love, support and patience.

ABSTRACT

BIOINFORMATICS BASED APPROACH TO DESIGN A THERMOPHILIC P450 FOR INDUSTRIAL BIOCATALYSIS

Enzyme catalyzed biosynthesis of steroidal drugs is an important process for pharmaceutical manufacturing. Cytochrome P450 (P450) monooxygenases are important for hydroxylation of steroid structures because they can catalyze the oxidation of inactive carbon bonds with high selectivity and efficiency. CYP119 is an acidothermophilic P450 from *Sulfolobus acidocaldarius*, which has the potential to be used as biocatalyst for industrial production since it shows activity at high temperature and low pH conditions. In this work we aim to use CYP119 for selective hydroxylation of progesterone, which is not the original substrate of CYP119, for production of precursor molecules of important hormones like cortisone and aldosterone. Crystal structure of CYP119 (PDB ID: 1F4T) was used for selecting residues that were mutated according to structural alignment with other CYPs that can catalyze progesterone hydroxylation naturally. Progesterone-docking performed with CYP119 to identify residues that create clashes with substrate. Finally selected 12 residues (Leu69, Val151, Phe153, Leu155, Leu205, Ile208, Ala209, Thr213, Thr214, Val254, Thr257, Leu354) were mutated with PyRosetta program to Gly, Glu, Phe, Met, Ala, His, Arg and Ile. Progesterone-docking performed with using DockMCM Protocol of PyRosetta. We used two different starting coordinates of progesterone for docking and results were eliminated according to their energy scores. Best mutants were used for creating double/triple mutants and second round of docking and elimination process were performed with using double/triple mutant enzymes. Final number of 11 mutants with best scores were selected and their possible products were identified.

ÖZET

ENDÜSTRİYEL BİYOKATALİZ İÇİN TERMOFİLİK P450 TASARIMINDA BİYOİNFORMATİK TEMELLİ YAKLAŞIM

Steroid temelli ilaçların enzim katalizli biyosentezi farmasötik üretimi için önemli bir süreçtir. Sitokrom P450 (P450) monooksijenazlar inaktif karbon bağlarının oksidasyonunu yüksek seçicilik ve etkinlikle katalizleyebildikleri için steroid yapıların hidroksilasyonu açısından önemlilerdir. Asidoterfomilik bir arkea olan *Sulfolobus acidocalderius*'tan elde edilen CYP119, yüksek sıcaklık ve düşük pH koşullarında aktivite gösterdiği için endüstriyel üretimde biyokatalist olarak kullanılma potansiyeline sahiptir. Bu çalışmada CYP119 kullanılarak orijinal substratı olmayan progesteron hormonunun seçici hidroksilasyonuyla, aldosteron ve kortizon gibi önemli hormonların öncül moleküllerinin üretimi amaçlanmıştır. Doğal olarak progesteron hidroksilasyonunu katalizleyen CYPlerle yapılacak yapısal hizalanma sonucu mutasyona uğratılacak aminoasitlerin belirlenmesi için CYP119'un kristal yapısı (PDB NO: 1F4T) kullanılmıştır. Substratla çakışma veren amino asitleri belirlemek için CYP119 enzimine progesteron yerleştirme yapılmıştır. Finalde seçilen 12 amino asit (Leu69, Val151, Phe153, Leu155, Leu205, Ile208, Ala209, Thr213, Thr214, Val254, Thr257, Leu354), PyRosetta program kullanılarak Gly, Glu, Phe, Met, Ala, His, Arg ve Ile amino asitleriyle mutasyona uğratılmıştır. Progesteron yerleştirme için PyRosetta programının DockMCM protokolü uygulanmıştır. Yerleştirme için iki farklı progesteron başlangıç koordinatı kullanılmış ve sonuçlar enerji değerlerine göre elenmiştir. En iyi mutantlar ikili/üçlü mutasyonları oluşturmak için kullanılmış ve yerleştirme ve eleme işleminin ikinci etabı ikili/üçlü mutantlar kullanılarak gerçekleştirilmiştir. Finalde en iyi sonuç veren 11 mutant seçilmiş ve olası ürünleri belirlenmiştir.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

Figure **Page**

Figure **Page**

LIST OF TABLES

Table **Page**

Table **Page**

CHAPTER 1

INTRODUCTION

1.1 Protein Design

Proteins are important macromolecules with a wide range of functions as catalysis, transport, cell cycle regulation and receptor signaling. Protein design approaches can be used for production of enzymes with high activity and selectivity, antibodies that can be used for *in vitro* diagnosis or transport molecules for drug delivery (Steiner and Schwab, 2012). Nowadays, number of known protein sequences are increasing rapidly with new improvements using next generation sequencing. Also, a rapid increase in the computational power and decrease in the cost of computers help researchers to develop new algorithms to deal with increasing data and use this data to develop new properties on proteins. Recent advances in the bioengineering field and automation of methods lead to increase in the number of experimentally determined protein structures. Protein Data Bank (RCSB PDB) contains experimentally determined crystal data of approximately 160.000 macromolecular structures (www.rcsb.org).

Enzymes are special protein structures that can catalyze a wide range of chemical reactions in mild conditions by decreasing the activation energy. Enzymes are environmentally friendly alternatives for traditional chemical synthesis procedures in production of a wide range of chemicals from drugs to biofuels (Schmid et al., 2001; Bornscheuer et al., 2012; Schoemaker et al., 2003). Recent advances in molecular biology and chemical synthesis field allow scientist to use enzymes for *in vitro* applications. Even though the improvements in enzyme production area, finding an enzyme with desired properties is still a challenge. Enzymes can be unstable, require additional compounds or proteins to work efficiently or can have narrow substrate range (Hilvert et al., 2013).

Enzyme design process has three main steps; selection of target residues, creating mutations and selection of mutants for desired activity. There are two main approaches to design proteins with desired properties. These are directed evolution approach which

use Darwinian evolution principles and rational design approach that use computational methods during design process.

1.1.1. Directed Evolution of Proteins

Directed evolution mimics Darwinian evolution to get proteins with desired properties (Cobb et al., 2012). Directed evolution creates random mutations on protein sequence and aims to create a library of mutants with high diversity. However, enzymes generally have few numbers of catalytically important residues while directed evolution approach targets larger sequences. As a result of this process directed evolution studies produce lots of unsuccessful mutants for desired activity and this results in waste of money and time. Another negative side of directed evolution is elimination process of library. Best conditions for production must be defined for each enzyme and there must be a specific selection assay for each type of activity. It is relatively easy to select improved mutants when there is a color change occurs during product synthesis. Other kind of reactions require methods like HPLC to detect product amounts, which can be too expensive for screening large libraries.

Advantage of directed evolution is creating variants with an iterative way. Additional mutation rounds can be performed to increase activity after the selection of mutant with desired activity even this activity is so small. Rational design approaches generally have an energy score barrier for elimination process and most of the mutants that have weak interactions with substrate rejected by rational design algorithm.

1.1.2 Rational Design of Proteins

Rational design approach uses computational tools to design and select structures with desired properties. Design procedures use the advantage of known three-dimensional structures of proteins and information about active site residues, substrate recognition sites and functional groups. This information reduces the number of required mutations

during design process and decrease the number of unnecessarily designed mutants that we see in directed evolution-based experiments.

Disadvantage of rational design process is low diversity of designed mutant pool. Design based on the info of target protein generally leads to creation of mutants with same or resembling functions with altered activity or different substrate specificity. Changing activity of protein completely is harder than designing improved mutants. This process often involves the insertion of sequences that are responsible of cofactor binding or using synthetic alternatives of cofactors during production of enzymes that originally have cofactors. There are several heme proteins designed with this approach (Liu et al., 2012; Yeung et al., 2009).

1.1.2.1 *De Novo* Design of Proteins

De novo design process is based on the idea that copying main functional part of enzyme and remodeling only that part for desired purposes. This allows design of peptide-based structures using both rational design and directed evolution methods. Although, computational approaches are mostly preferred for *de novo* design process (Johnnson et al., 1993; Davie et al., 2007), directed evolution methods also used especially in vaccine design (Nelson et al., 2018).

Rational design helps to understand the working mechanism of selected peptide and improve designed structure. There are many peptide-based catalysts that were designed with computational methods that function as kinases (Sculimbrene et al., 2003) and halogenases (Gustafson et al., 2010) with high chemo- and stereoselectivity. Helical bundles are the most common *de novo* designed peptide structures (Hill et al., 2000). They can function similar to catalysts especially when they bind to metal ions like zinc, iron or mercury (Smith et al., 2011). Research shown that 3-helix bundle with Hg(II) and Zn(II) ion can function as carbonic anhydrase enzyme (Zastrow et al., 2012). Another example is a 4-helix bundle with iron centered structure that can catalyze phenol oxidation (Faiella et al., 2009).

1.1.2.2 Ligand Docking

Ligand docking process defines the prediction of a ligand location and orientation inside the user defined binding site which is generally referred as grid box. First aim of docking is modelling the structure with high accuracy and the second aim is prediction of activity.

Docking process always begins with identification of a possible location of ligand in the active site. This process generally referred as creating a pose. Main problem in the posing process is conformational diversity of ligands. Some ligand molecules can have a high number of conformational freedom. However, there are bioinformatic tools that can create ligand conformational library with high accuracy. Ligand docking programs like PyRosetta can use full library of single ligand at the same time during docking process.

Algorithms use several approaches to handle ligand flexibility. These approaches can group in three categories: systematic methods (database search for conformations), random methods (Monte Carlo algorithm) and simulation methods (molecular dynamics) (Brooijman and Kuntz, 2003). Systematic methods use a ligand library which includes all the possible conformations. Random search methods create random changes on the ligand and calculates root mean square deviation (RMSD) values between previous conformation and decides which conformation is more favorable (Liu and Wang, 1999). Simulation methods aims to understand enzyme behavior during ligand binding (Throsset and Sheraga, 1995).

Proteins are very flexible molecules so generating a flexible protein structure during docking process is very important for ligand docking approaches. There are several approaches for creating a flexible target region like molecular dynamics, Monte Carlo algorithm (Morris et al., 1998) or using rotamer libraries (Desmet et al., 1992). Monte Carlo algorithm calculates the energy score of ligand in initial coordinates, than creates a random conformation, translation and rotation. Rescores new configuration of ligand and uses a Metropolis criterion for elimination of results. Metropolis criterion simply explains which configuration should be accepted or rejected with using energy scores. Monte Carlo algorithm repeats this docking cycle until user-defined number of models created. Some ligand docking programs use dead-end-elimination algorithm to

eliminate structurally non-preferred conformations. This algorithm selects conformations that are favorable as energetically (Desmet et al., 1992).

Docking algorithms use scoring functions to evaluate the results of ligand docking process and selecting most suitable ligand for target enzyme. Scoring functions can classified in three groups as force-field-based, empirical and knowledge-based.

Force-field energy scoring function uses ligand-receptor interaction energy and ligand energy for calculation. Force field score functions generally use a single protein conformation while calculating ligand-receptor interaction energy to simplify scoring process. Ligand-receptor interaction energy is explained as a combination of van der Walls and electrostatic energy terms. Lennard-Jones potential function is used for van der Walls interactions while Coulombic function is used to determine electrostatic energy terms. Limitation of force field score function depends on cut – off distance that is used for calculations. It ignores long distance interactions which can be effective for ligand binding.

Empirical score functions use experimental data such as experimentally defined conformations or binding energies. Main idea is calculating binding energy as a function of non-correlated data. Coefficients for experimental data are determined with regression analysis of experimental results and structural information. Regression analysis is the main limitation of empirical score functions because it depends on real data and using regression analysis with different data sets often cause different coefficients for each purpose which makes difficult to compare results.

Knowledge based structures reproduce experimental ligand-protein interaction without using binding energy. Most advantageous part of this scoring function is computational simplicity which provides the screening of large compound libraries efficiently. This approach also has a major disadvantage which can cause false negative results during library screening because of limited background data.

1.2 Cytochrome P450 Monooxygenases (CYPs)

Cytochrome P450 monooxygenase enzymes catalyze the monooxygenation of non-active C-H bonds with a high regio- and stereoselectivity (Urlacher et al., 2004). This

reaction is extremely difficult with chemical synthesis. CYP enzymes can catalyze several types of reactions like hydroxylation, epoxidation and dealkylation of various type of substrates (Cryle et al., 2003). CYP catalyzed reactions are important for drug synthesis and metabolism, steroid biosynthesis, detoxification, fatty acid metabolism (Montelleno, 1995). CYP enzyme family is one of the largest families with more than 300,000 known members that originated from all kingdoms, including viruses (Nelson, 2017).

CYP enzymes are good candidates for industrial applications because they can catalyze various oxygenation reactions and have high substrate diversity. However, CYPs are cofactor and oxygen dependent enzymes and generally have low stability and activity. They need additional proteins for efficient electron transfer from donor molecules like NAD(P)H which are very expensive co-factors (Bernhardt and Urlacher, 2014). H_2O_2 can be used as a redox partner to solve this problem because this system does not require an additional reductase protein (Joo et al., 2009). However, H_2O_2 is not suitable at high concentrations for processes like whole cell catalysis (Gillam et al., 2008). Since cell can provide efficient electron transfer system, currently only whole cell biocatalysis method used for industrial production (Urlacher and Eiben, 2006).

Another limitation is substrates of CYP enzymes which have low solubility in water. These properties of CYP enzymes restrict their applications on industrial production (Bernhardt, 2006). Bacterial CYPs are more advantageous for industrial process because they are soluble while eukaryotic CYPs not soluble and also membrane bound (Munro, 1996).

CYP enzymes generally have low sequence similarity but share a common structure which is highly conserved. Three-dimensional structure of CYP119 has shown in Figure 1.1 as an example of common folding pattern of CYP enzymes. They share only one common sequence motif, EXXR, which locates in the K-helix and important for stabilization of meander region (Ravichandran et al., 1993). Meander region is important for heme binding and tertiary structure of P450s (Seifert et al., 2009). Cysteine residue is one of two conserved residues and essential for heme binding. Mutants of conserved cysteine residue cannot catalyze monooxygenation reaction and generally fail to bind heme group (McIntosh et al., 2015). Second conserved residue is the threonine residue located on I-helix and responsible proton transfer to the heme (Vidakoviç et al., 1998). Secondary structure of CYPs consist of 6 β-sheets and 13 α-helices and 6 substrate recognition sites (SRSs) (Hasemann et al., 1995). SRS residues can interact directly to ligand molecules or can affect binding to active site with forming final structure or

determining flexibility of binding site. Since all CYP enzymes share same three-dimensional structure, it is logical to expect structurally corresponding residues show same function in the enzyme. This approach generally used for determining important residues of novel CYP enzymes.

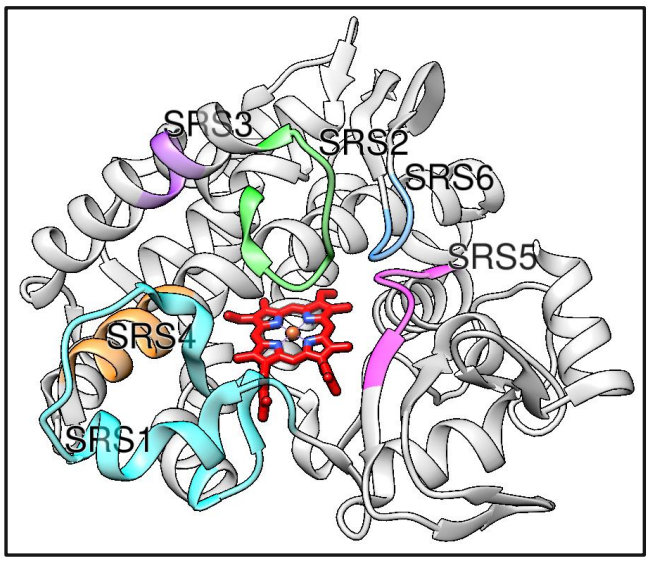

Figure 1.1 Three-dimensional structure of CYP119 with color coded SRS regions. (PDB ID:1F4T)

CYP enzymes have 6 substrate-recognition sites which share same location in the final structure but not have sequence similarity. SRS1 is a part of B-C loop. SRS2 is located on the C terminal side of F-helix and forms the entrance of substrate binding pocket with F-G loop and SRS3 is located on the N terminal side of G-helix. SRS4 is located on the I helix and has the highest residue conservation rate (4 of 18). Most CYPs have conserved AGXXT motif on SRS4 region of I-helix (Mestres, 2005). Conserved glycine residue of SRS4 is structurally important. Following residues generally include a conserved negative amino acid, a threonine and a serine/threonine which are important for proton delivery (Gricman et al., 2014). Conserved Ala and Thr residues of SRS4 are important for heme binding. SRS5 is located on the region starting from the conserved EXXR motif to the β1-4 strand. Several studies have shown that mutations of SRS5 region effect enzyme selectivity, specificity and activity (Urlacher et al., 2006; Liu et al., 2004; Meinhold et al., 2006). Seifert and colleagues shown that 98.4% of all SRS5 regions

contain a hydrophobic residue located at the 5th residue after the EXXR motif and this residue is critical for substrate specificity and regioselectivity. This hydrophobic residue is generally valine (38%) or alanine (18%). 27% of all CYPs contain a second selectivity determining residue located at the 9th position after EXXR motif. Research has shown that 97.7% of all CYPs have a positive charged triad (His-Lys-Arg) 10,11 or 12 residue after EXXR motif (Seifert et al., 2009). SRS6 spans between β strands 4-1 and 4-2.

Substrate binding residues of CYPs locates in the cavity formed by SRS5 region, B-C loop and I-helix (Gotoh, 1992). These residues are important for protein engineering approaches because they are directly affecting selectivity/specificity (Bell et al., 2003; Seifert et al., 2006). B-C loop has high sequence diversity, so it is not possible to use mutations of loop region to identify important residues in CYPs with comparison methods (Seifert et al., 2009).

CYP enzymes classified as Class I and Class II enzymes according to their redox partner type. These groups share some structural similarities among redox partner preferences like the difference in the length of meander region (Gricman et al., 2015). Class I enzymes use ferredoxin/ferredoxin reductase system as electron donor and generally prokaryotic while Class II enzymes accept electrons from NADH/NADPH reductase system and mostly eukaryotic (Gricman et al., 2015). H_2O_2 can be used as electron donor system ($O_2/2e^-/2H^+$ system) which specially named as peroxide shunt pathway. Peroxide shunt pathway generally insufficient for effective electron transport and have a low turnover rate. H_2O_2 can also cause CYP inactivation via heme destruction or oxidative modification of enzyme.

Spin state of CYP enzymes is easy to determine with spectrophotometric analysis, since conversion of low to high spin state cause a shift in Soret peak. CYP enzymes give absorbance maximum at 415 nm in low spin state, upon the loss of distal water molecule enzyme pass high spin state which give absorbance maximum at 390 nm. P450 spin states are depend on environmental factors like temperature, solvent and pH (Koo et al., 2000).

Figure 1.2 shows schematic presentation of CYP catalytic cycle. Iron atom of substrate free enzyme stays in ferric (3^+) state because of distal water molecule (1). Enzyme change spin state from low to high after replacement of water with substrate molecule (2). First reduction forms ferrous (2^+) iron atom (3) and makes O_2 binding possible (4). Second electron transfer to oxygen atom (5a) leads H^+ transfer to O^{2-} atom (5b). Using H_2O_2 as electron donor (shunt pathway) connects step 2 and step 5b. Cleavage of O-O bond produce Compound I (Fe^{IV} state) which is responsible of main activity (6).

Last step is binding substrate and hydroxyl group (7) and replacement of hydroxylated compound with water ligand.

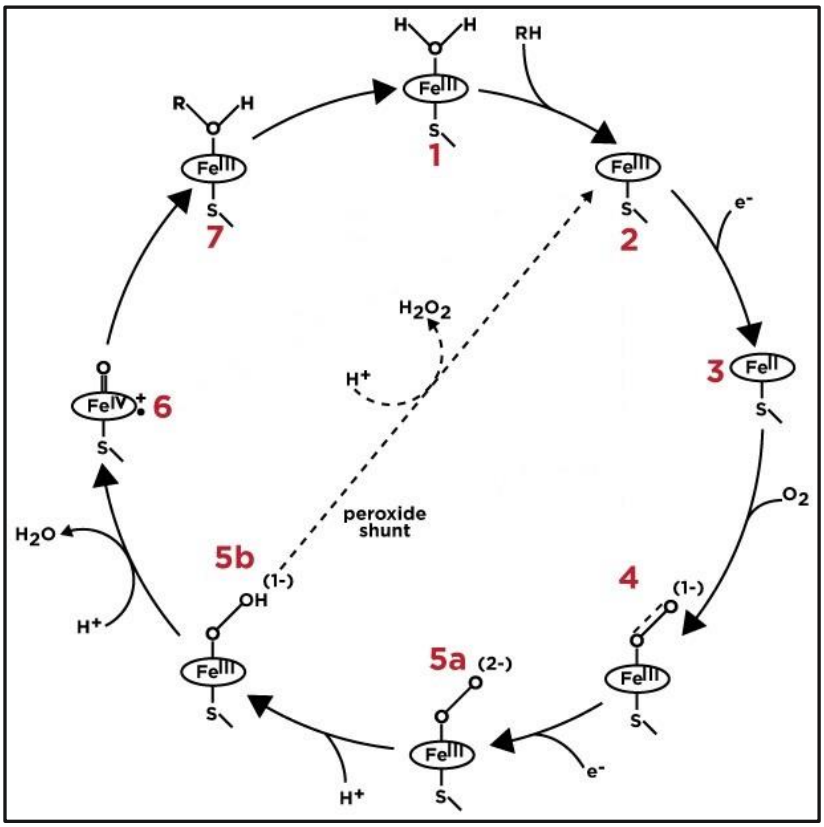

Figure 1.2 Schematic presentation of P450 catalytic cycle.

1.2.1 CYP119

Industrial conditions require enzymes that are stable in high temperatures, organic solvents, extreme pH levels and presence of different type of chemicals (Harris et al., 2017). Thermostable enzymes have lots of benefits for industrial applications. Most important outcome is increase in reaction rates which also means high yield of production with less time and money. Increased solubility of substrates and lower risk of microbial contamination also advantageous properties for large scale production (Chang et al., 2000). There are only four thermophilic P450s that have crystal structure; CYP119 (*Sulfolobus acidocaldarius*) (Park et al., 2002), CYP175A1 (*Thermus thermophilus*)

(Behera and Mazumdar, 2010), CYP231A2 (*Picrophilus torridus*) (Ho et al., 2008), $P450_{st}$ (*Sulfolobus tokadaii*) (Oku et al., 2004).

CYP119 is a thermophilic CYP enzyme isolated from acidothermophilic archaea *Sulfolobus acidocaldarius* with melting temperature of 90°C (McLean et al., 1998). CYP119 has a conserved heme binding region which shows high identity with bacterial and mammalian P450s (Koo et al., 2000). CYP119 is shorter than most of other CYPs with length of 366 amino acids while CYPs generally have 400-450 amino acid length. CYPs generally have a long tail at the N-terminal which is absent in CYP119 (Yano et al., 2000). B5 turn of CYP119 is five residues shorter than other CYPs and this region generally controls the interaction between redox partner protein and CYP enzymes. This region of CYP119 resembles CYP55 (CYP_{nor}) which is a self-sufficient CYP enzyme and uses NADH without the help of any redox partner protein to catalyze nitric oxide reduction (Park et al., 1997).

Melting temperature of CYP119 was determined with differential scanning calorimetry method. CYP119 has approximately 40°C higher melting temperature than other mesophilic CYPs (Park et al., 2000). Thermostability related factors can be listed as; disulfide bonds, aromatic clusters, high number of hydrogen bonds and salt bridges, tightly packed hydrophobic residues and short surface loops (Sterner and Liebl, 2001; Szilagyi and Zabodszky, 2000). Since CYP119 has only one cysteine residue (Cys317), proximal ligand for iron atom, disulfide bridges cannot be the reason of its stability (Wright et al., 1996). Structural alignment with other CYPs reveals the possible parameters for thermostability. CYP119 has some unique salt bridges (i.e. Arg154-Glu212) which may affect enzyme stability towards high temperature (Chang and Loew., 2000). Second important difference is the amount of Ala and Ile residues around active site. CYP119 has an increased number of Ile instead of Ala when compared with other CYPs which probably effect side chain packing. Most important parameter for thermostability of CYP119 is the two unique aromatic clusters which expanding 39 Å on the protein (Yano et al., 2000). Aromatic cluster I includes Tyr2, Trp4, Phe5, Tyr15, Phe24, Trp281 and spans 11.3 Å. Cluster II consist of Phe225, Phe228, Trp231, Tyr250, Phe298, Phe334, Phe338 and spans 24 Å (Nishida et al., 2005). Crystal structure of CYP119 proves that clustering regions are relatively rigid and probably affecting the thermostability of enzyme (Park et al., 2002). Figure 1.3 shows the aromatic cluster I and II of CYP119.

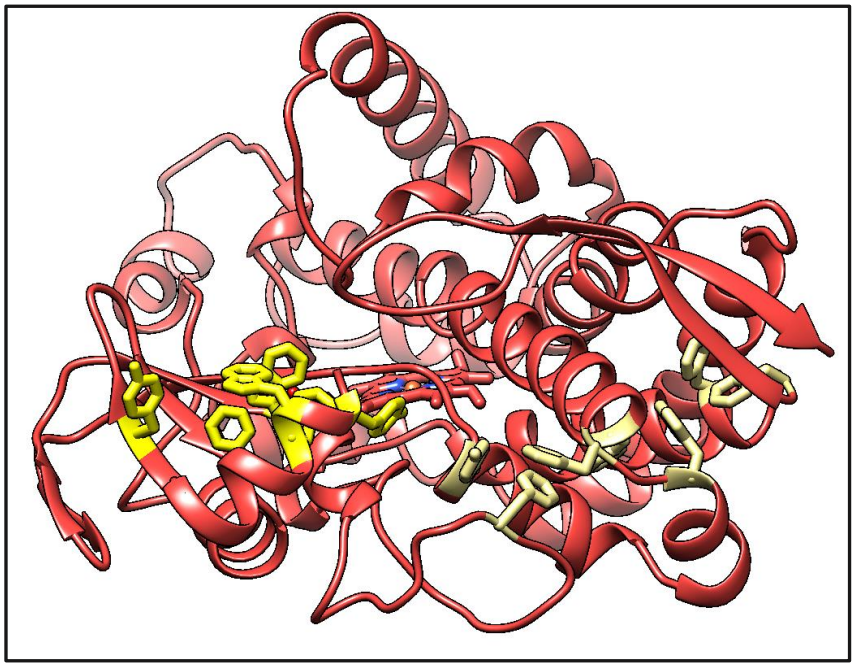

Figure 1.3 Aromatic cluster I (yellow) and II (tan) of CYP119 (1IO7).

Meharenna et al performed a molecular dynamics simulation to compare the behavior of CYP119 and mesophilic P450 CYP101 ($P450_{cam}$). Cys ligand loop of CYP119 stays stable during MD simulations for 20 ns at 500K while it is unfolding in CYP101. The distance between Cys residue and conserved Phe residue (Phe310 in CYP119) is measured as 2.9Å for both enzymes. After 500K molecular dynamics simulations the distance increases to 11 Å for CYP101 while distance between Cys317 and Phe310 shows a small difference (3.1Å) for CYP119. These conserved residues form a hydrogen bond and small movement of CYP119 makes this H bond remain stable (Meharenna et al., 2010).

Comparison of CYP119 amino acid sequence with other CYPs showed percentage of charged residues (32.4%) is higher than mesophilic CYPs (27%). CYPs generally have more negatively charged residues (Asp, Glu) than positively charged ones (Lys, Arg, His), while CYP119 has exact opposite distribution. CYP119 has low content of Pro (6.7% other CYPs, 4.6% CYP119) and Gln (4.2% other CYPs, 1.4% CYP119) residues and Gln residues often replaced by charged residues (Chang et al., 2000).

X-ray structure of CYP119 has shown that it shares same three-dimensional structure with other CYPs. Extreme flexibility of F-G loop is one of the main differences between CYP119 and other CYPs, even though large changes in tertiary structure of

CYPs upon ligand binding is not a rare thing (Li and Poulos, 1997). CYP119 has ten crystal structures in RSCB database. Table 1.1 shows the resolution and PDB ID of these structures. Unfortunately, original substrate of CYP119 is not known thus CYP119 does not have a substrate bound structure in the database which would be the best choice for rational enzyme design.

Table 1.1 PDB ID's and properties of crystal structures of CYP119.

PDB ID	Type	Resolution
1IO7	WT, 100 K	1.5 Å
1IO8	F24L mutant	2.0 Å
1IO9	WT, 297 K	2.05 Å
1F4T	Inhibitor bound, 4-phenylimidazole	1.93 Å
1F4U	Inhibitor bound, imidazole	2.69 Å
4TT5	Inhibitor bound, 4-(4-bromophenyl)-1H imidazole	2.18 Å
4TUV	Inhibitor bound, 4-(4-chlorophenyl)-1H imidazole	2.5 Å
4WPD	Inhibitor bound, 4-(4-flourophenyl)-1H imidazole	2.0 Å
4WQJ	Inhibitor bound, 4-(4-bromophenyl)-1H imidazole, 298 K	2.7 Å
5BV5	T213A/C317H mutant	2.7 Å

Comparison of wild type enzyme structure with inhibitor bound models showed significant difference between the location and structure of F-G loop (Val151-Leu164). All F-G loop residues change their location in a range of 9-18 Å, depending on inhibitor molecule type (Lampe et al., 2010). Phe153 side chain rotates 65° and become more parallel to the heme upon inhibitor binding. The salt bridge between Arg154 and Glu212 of I-helix also disappears upon inhibitor binding (Puckhaev et al., 2002).

Figure 1.4 shows the F-G loop of free and 4-phenylimidazole bound enzyme. Movement of F-G loop might be responsible of binding different size of molecules to the active site (Park et al., 2002). Comparison of imidazole and phenylimidazole bound crystal structures of CYP119 has shown that behavior of F-G loop depends on the molecule size. F-G loop moves closer to the active site in imidazole bound model to increase the contact between small ligand and heme group. Phenylimidazole is bigger

than imidazole thus F-G loop moves slightly to increase active site volume (Puckhaev et al., 2002). This data indicates the effect of F-G loop in ligand binding process.

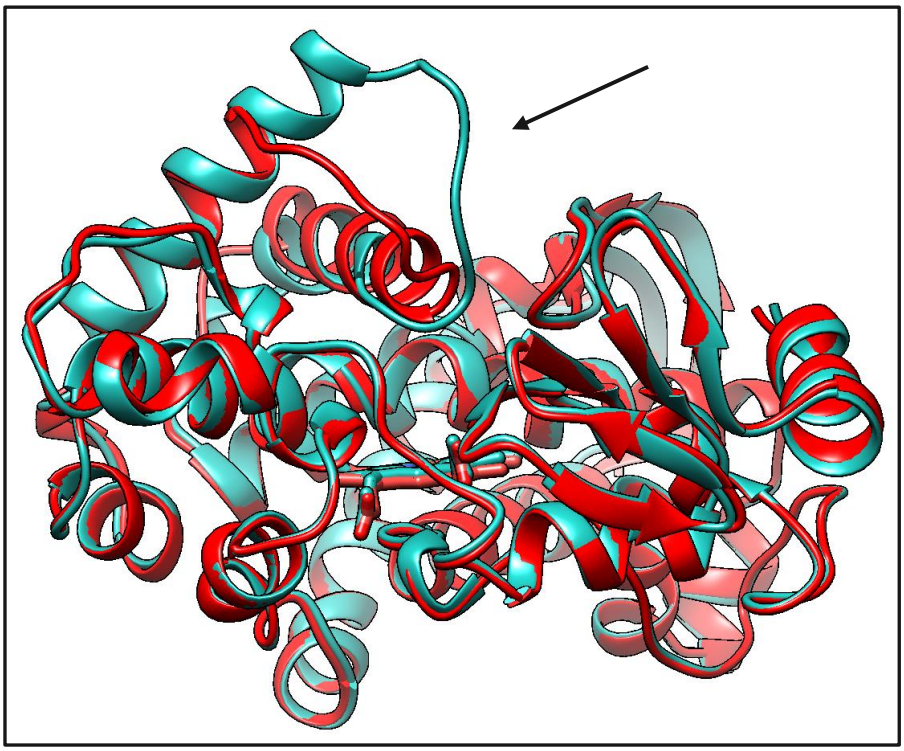

Figure 1.4 Structural alignment of free (1IO7, blue) and inhibitor bound (1F4T, red) crystal structures of CYP119. Arrow shows the F-G loop.

CYP119 enzyme has two additional threonine (Thr214 and Thr215) to conserved Thr213 residue, which is not unique for CYPs but only example for bacteria CYPs (Montelleno, 1995). These threonine residues are important for spin state and catalytic activity of CYP119 but not related to thermostability. Conserved Thr213 residue is directly linked to spin state of other P450s. To understand the effect of this residue on CYP119, it was mutated to Ala, Ser, Val, Phe and Trp. All mutants had a normal spectrum and none of them loss thermal stability. However, mutation of Thr214 residue to Ala and Val resulted increase high spin state enzyme 10-15% (Koo et al., 2002). Both mutations did not affect the folding of enzyme. Most of CYPs have only one Thr residue at conserved position which forms hydrogen bond with Gly210. CYPs that have a second Thr residue like CYP119 use second Thr to make H bonds with conserved Gly residue (Nishida et al., 2005).

Reductase partner protein of CYP119 is not known. Activity of CYP119 measured with using reductase partner groups like putidaredoxin/putidaredoxin reductase, ferredoxin/ferredoxin reductase, human CYP reductase from different sources but none of them work efficiently as electron donor partner (Koo et al., 2000).

1.2.1.1 Secondary Structure of CYP119

CYP119 consist of 366 amino acids and it is clearly shorter than other CYP enzymes. Since all CYP enzymes share a conserved three-dimensional structure which is related to enzyme function, structural properties of the enzyme is an important information source for design process. Even there are some experimentally determined important residues of CYP119, complete list of structural elements can not be found on literature. Secondary structure of CYP119 determined with structural alignment and shown in Figure 1.5. UCSF Chimera program used for structural alignment. Secondary structure elements, SRS regions, structurally important residues and conserved structural elements of CYP119 identified as a part of this project. CYP119 consist of 15 α-helices and 7 β-sheets.

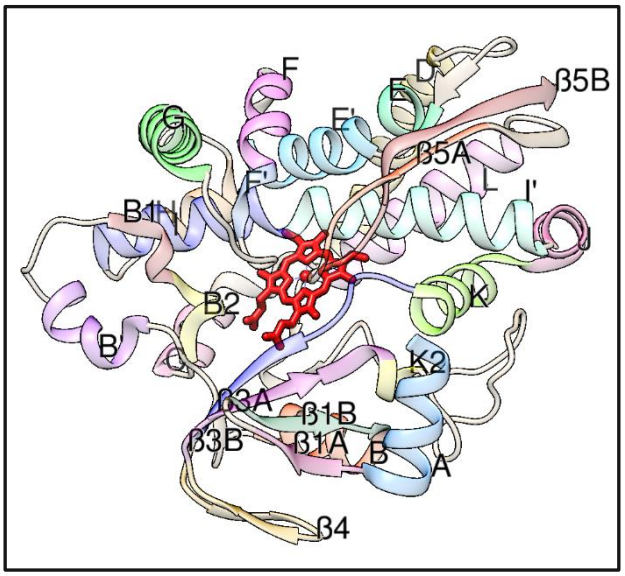

Figure 1.5 Secondary structure of CYP119.

A-helix includes residues 1-12 and followed by β1A-sheet (residues 13-17). Gly18 is a conserved residue that locates between β1A and β1B sheets and changing the rotation of β1 sheet (Figure 1.6).

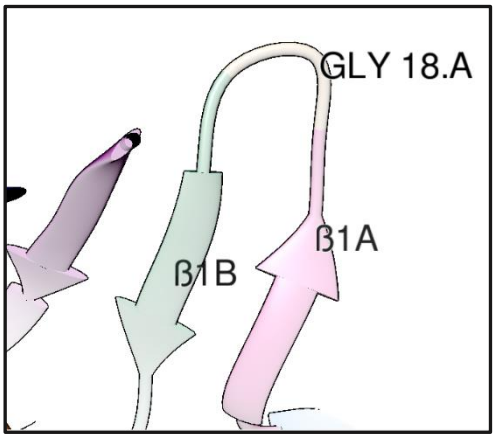

Figure 1.6 Representation of Gly18.

β1B-sheet (residues 19-25) and B-helix (residues 26-35) do not have any conserved residues. B-B' loop (residues 36-44) residue Ser40 forms a conserved H bond with +1 β3 Arg (Arg259) which corresponds to Lys260 in CYP119 (Figure 1.7). These conserved H bonds are important for true folding of enzyme.

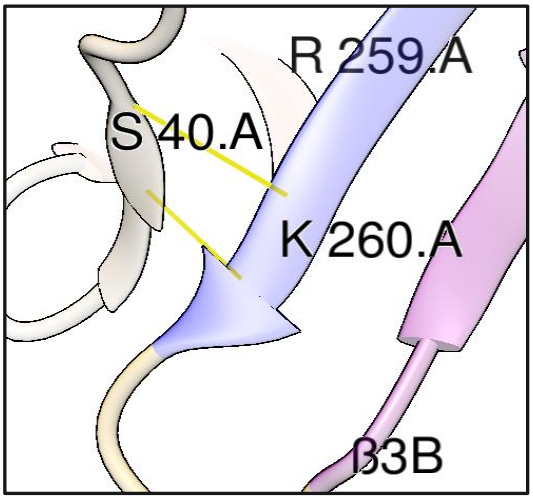

Figure 1.7 Conserved hydrogen bonds (yellow) between Ser40 and Lys260.

B’-helix starts with residue 45 and ends with residue 54. Loop region between B’-helix and C-helix includes two small helices named B1-helix (62-66) and B2-helix (67-71). Leu69 of B2-helix is a critical substrate contact residue. C-helix includes residues 73-81. Arg80 (Figure 1.8) is the most important residue of C-helix, has conserved hydrogen bonds with D-ring propionate and -1 thiolate residue which corresponds to Leu316 in CYP119.

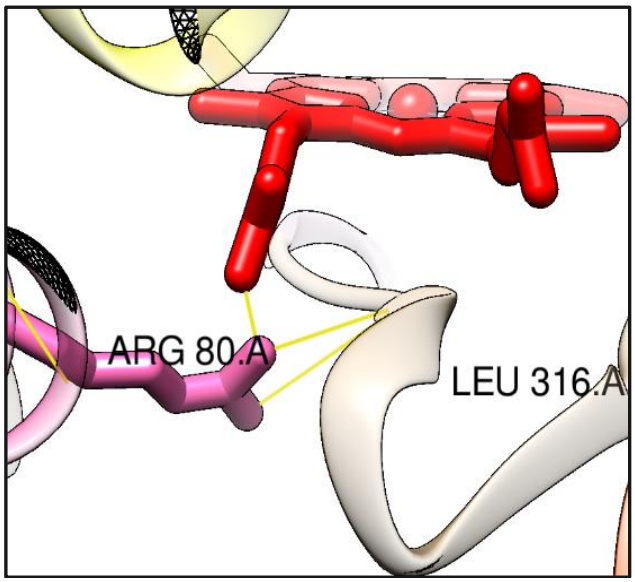

Figure 1.8 Conserved hydrogen bonds (yellow) between Arg80-Leu316 and Arg80-D ring propionate.

C-D loop (residues 82-87), D-helix (residue 88-109), E-helix (residues 117-121) and E’-helix (residues 122-135) do not have any conserved residues. F-helix (residues 140-151) originally consists of 12 residues but its length can change upon ligand binding (9 residues for imidazole, 17 residues for 4-phenylimidazole bound structures). This flexibility is important for acceptance of substrates with various sizes. F-G loop (residues 151-164) is important for substrate binding process and also most flexible part of CYP119 enzyme. Phe153 is a conserved phenylalanine residue of F-G loop and can effect the shape of β5 sheet. Figure 1.9 shows the location of Phe153 in wild type and different inhibitor bound CYP119 models.

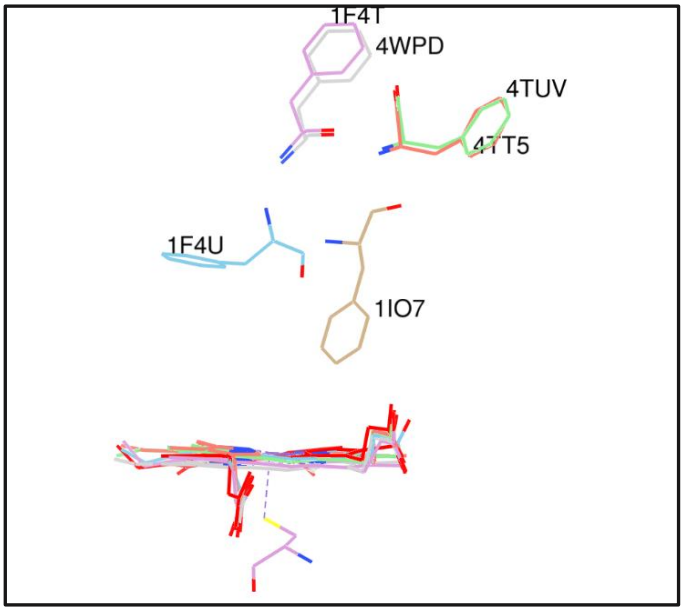

Figure 1.9 Phe153 location in different models of CYP119.

Most flexible residues of F-G loop are Pro158 and Gly159. Both residues displace 13-18 Å in all inhibitor bound models. Figure 1.10 shows the location of Pro158 and Gly159 in wild type and different inhibitor bound CYP119.

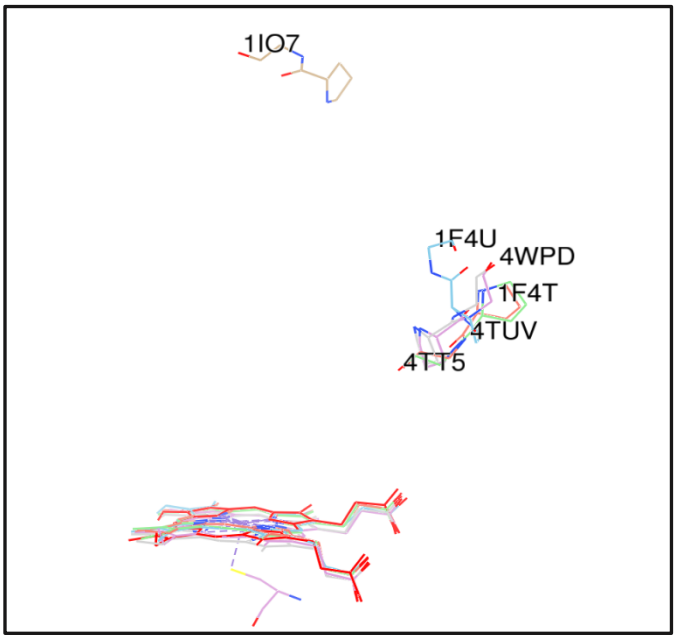

Figure 1.10 Pro158 and Gly159 residues of wild type and inhibitor bound CYP119 structures.

G-helix consists residues 164-179 and gets 4 amino acid shorter in inhibitor bound models. H-helix (residues 183-191) has one structurally important element, Thr183, which forms hydrogen bonds with Leu134 of E'helix (Figure 1.11). This H bond is important for the location of E' and H helices.

Ala209 is the most important residue of I-helix (residues 195-209). It gives flexibility to long I-I'helices with Gly210 of I'-helix. Ala209 moves 1-2 Å upon substrate binding and decrease active site volume with creating a kink between I and I' helices. I'-helix (residues 210-228) is important for enzyme activity, because conserved alcohol-acidic residue pair (Glu212-Thr213) locates on this helix and forms proton delivery pathway. Glu212 also forms a salt bridge with Arg154 of F-G loop, which is important for flexibility of loop (Figure 1.12). Inhibitor binding changes the rotation of Glu212 side chain and disrupts this salt link.

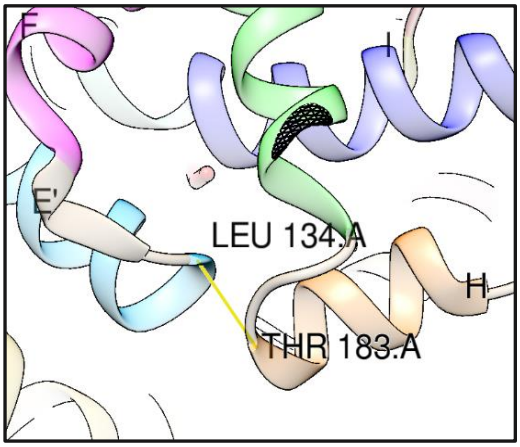

Figure 1.11 Hydrogen bond (yellow) between Leu134-Thr183.

Asn229 creates the turn between I'-helix and J-helix. J-helix consists of residues 230-238 and most important residue is Ile234 which has a hydrogen bond with Leu239 of K-helix. This H bond is important for stabilization of J and K helices and J-K loop. K-helix (residues 239-251) has two members of ERR triad (Figure 1.13) which is a common feature of all CYPs. Glu246 and Arg249 forms ERR triad with Arg302 of meander region. They also forms the only conserved motif of CYPs, EXXR. ERR triad acts as a folding motif, thus effects heme-binding, plays role in redox partner binding and also provides the conserved connection between meander region (Asp288) and B-helix (Tyr26).

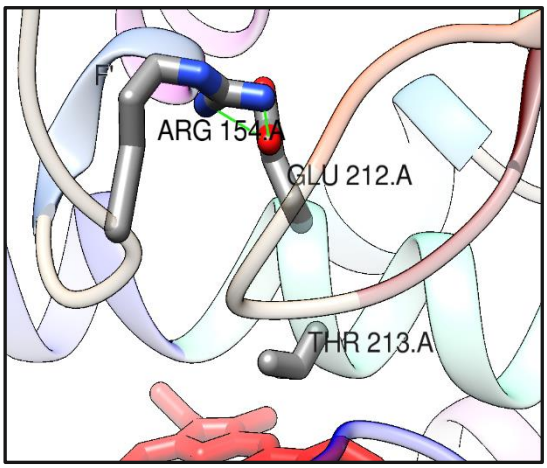

Figure 1.12 Arg154, Glu212 and Thr213 residues. Salt bridges between Arg154 and Glu212 colored green, heme group coloured red.

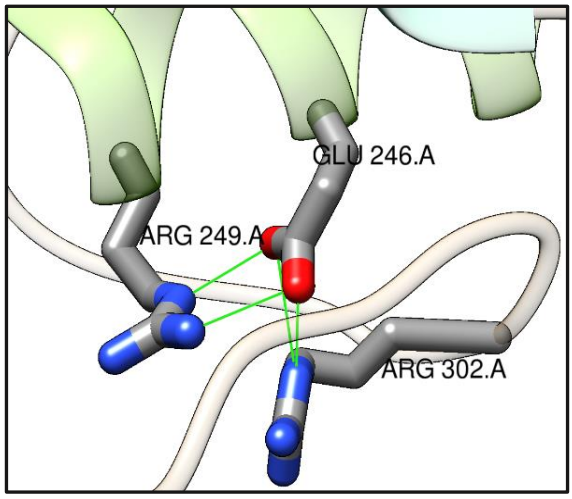

Figure 1.13 ERR triad. Salt bridges colored green.

β3A-sheet connects ERR triad and β5-sheet to heme binding region. β3A length is structurally conserved and potential substrate contact residues are generally specific to CYP subfamily. β3A-sheet (residues 252-261) (Figure 1.14) follows EXXR motif and show some conserved features. PPVM (Pro252, Pro253, Val254, Met255) conserved sequence forms the beginning of β3A-sheet. Pro 253 and Met255 have hydrogen bonds with Asn355 of β5B sheet and Val353 of β5 sheet respectively. These conserved hydrogen bonds are important for folding. Val254 is a substrate contact residue located near the A ring (5.28 Å). Conserved RTV (Arg256, Thr257, Val258) sequence follows

PPVM sequence. RTV sequence is important for substrate recognition and enzymatic activity. Thr257 has structural importance since it has H bond with C-ring propionate of heme group. This H bond is conserved between C-ring propionate and -2 β3 arginine (Arg259). Loss of this hydrogen bond may cause loss of heme group, but Thr residue is not conserved so it can be replaced any other residue as long as hydrogen bond conserved. Val258 has a hydrogen bond with Arg259 which makes it structurally important. Arg259 (known as β3 arginine in literature) is the most important and conserved residue of β3A-sheet. Arg259 has a hydrogen bond with C-ring propionate of heme group also known as heme binding arginine. Mutations of this residue leads complete loss of activity due to loss of heme group.

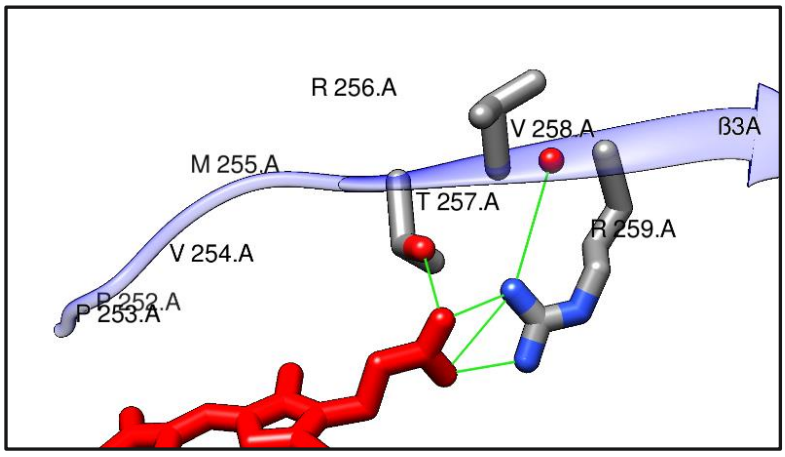

Figure 1.14 β3A-sheet with heme group (red). Conserved hydrogen bonds shown with green.

β4-sheet (residues 262-274) locates between β3A-sheet and β3B-sheet (residues 257-283). Val280 of β3B-sheet corresponds conserved +21 β3 arginine residue and it stabilize -2 β3 arginine residue (Thr257 in CYP119) with hydrogen bond. There is a small K2-helix (residues 284-286) before meander region (287-318). Meander region is the most conserved region of all P450s. Phe310 is a well conserved resdiue and effective on reduction potential of heme. His315 has conserved hydrogen bonds with D-ring propionate of heme group (Figure 1.15) and mutation of this residue disrupted enzyme activity in several P450s like CYP11B1, CYP27A1, CYP17A1, CYP21B1. Cys317 covalently binds iron atom of heme group (2.27 Å) and act as proximal ligand.

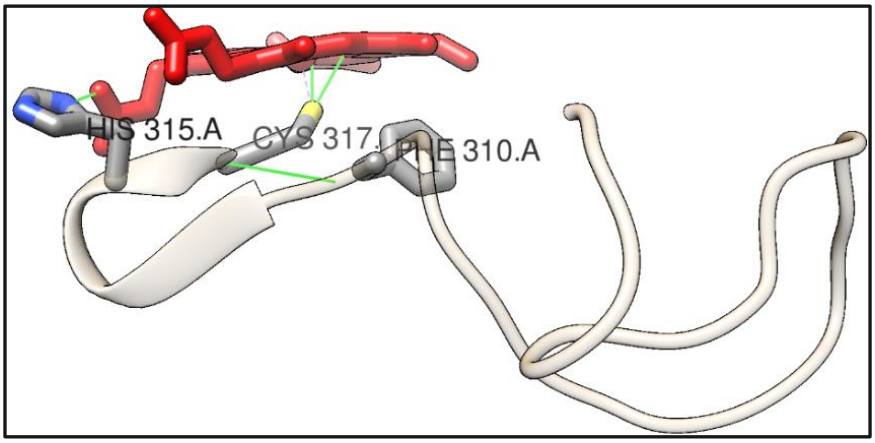

Figure 1.15 Meander region of CYP119.Conserved hydrogen bonds colored with green, covalent bond between Cys317 and iron atom represented with dash line.

L-helix (residues 319-336) has two conserved residues. Gly319 is necessary for stabilization of heme-thiolate loop, it helps to get correct localization for conserved hydrogen bonds. Glu326 is the other conserved residue of L-helix which stabilizes the position of Leu206, Leu207 and Gly210 with hydrogen bonds. β5A-sheet (residues 345-351) and β5B-sheet (residues 355-365) form large and well conserved hydrophobic surface of binding cavity. Turn between β5A and β5B sheets consist of three substrate contact residues; Glu352, Val353 and Leu354 (EVL sequence) (Figure 1.16). Leu354 acts together with Phe153 to shape substrate contact surface and correct orientation of bound substrate molecule. Val353 also has effect on substrate orientation.

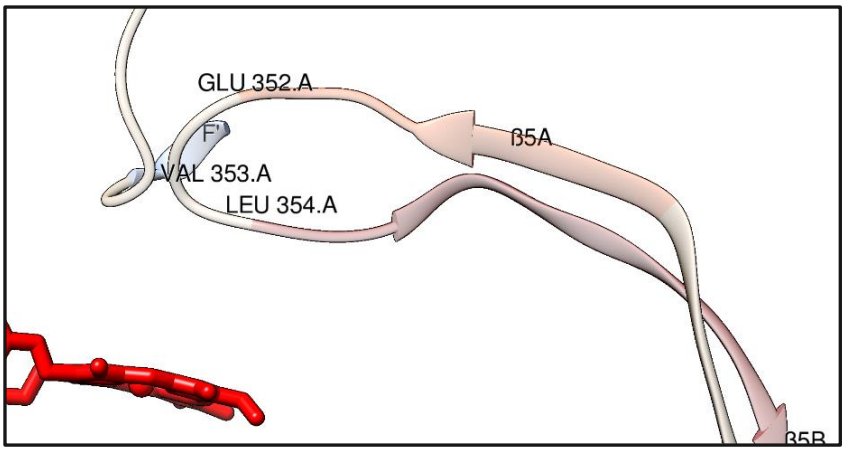

Figure 1.16 EVL region of CYP119 with heme group.

1.2.1.1 Literature About CYP119

Table 1.2 shows all known mutations of CYP119. Original substrate of CYP119 enzyme is not known but several studies shown that it can catalyze lauric acid hydroxylation (Koo et al., 2002) and styrene epoxidation (Koo et al., 2000). CYP119 can also catalyze dehalogenation of CCl_4 to CH_4 (Blair et al., 2004) and reduction of nitrite, nitric oxide and nitrous oxide (Immoos et al., 2004). CYP119 binds imidazole and styrene with weak interactions while binding lauric acid with high affinity as seen in all CYPs. Hydrophobic substrates like lauric acid more effective to removal of distal water molecule thus increase reduction potential of heme (Montelleno, 2005). This effect the oxidation rate of substrate and explains the low activity towards styrene because of non-efficient axial water removal (Koo et al., 2000). CYP119 can bind various length of fatty acids (C8-C20) with different affinity levels (Koo et al., 2001).

CYP119 shows highest hydroxylation activity towards lauric acid at 70°C and activity at 70°C is 10-fold higher than activity at room temperature (25°C) (Puckhaev et al.,2002). Another study shows that CYP119 maintains monooxygenase activity for styrene epoxidation with using H_2O_2 as oxidant up to 80°C (Koo et al., 2000).

Rabe et al., showed that CYP119 can catalyze peroxidation of AmplexRed chemical which was originally used for determining catalytic activity of peroxidase enzymes. AmplexRed is a non-fluorescent molecule which turns to fluorescent resorufin molecule with oxidation and consumption rate of AmplexRed, indirectly activity of oxidation enzyme, can be calculated with measuring absorbance of resorufin at 530nm (excitation maxima) and 584nm (emission maxima) (Zhou et al., 1997). CYP119 showed best activity against AmplexRed at 75°C, pH 8.5 with using 5mM tert-butyl hydroperoxide (TBHP) as redox partner (Rabe et al., 2008).

There are several mutational studies with CYP119 that aims to identify important residues for activity. Arg154-Glu212 salt bridge is important for substrate binding. Puckhaev et al., performed mutations on these residues to understand the effect on lauric acid hydroxylation. E212D mutant completely loss activity while E212Q showed 2-3-fold decrease. On the other hand, R154A and R154Q mutants showed no effect on hydroxylation rate even the loss of salt bridge. This result shows that Glu212 is not only important for F-G loop flexibility but also have a role in activity (Puckhaev et al., 2002).

Table 1.2 Literature about CYP119.

Mutant	Effect on	Reference
	Styrene epoxidation	
T213A	Reduce activity (1.2 fold)	Koo et al., 2000
T213F	Loss of activity	Koo et al., 2000
T213S	Reduce activity (5 fold)	Koo et al., 2000
T213V	Reduce activity (147 fold)	Koo et al., 2000
T213W	Reduce activity (19 fold)	Koo et al., 2000
T214A	Reduce activity (2.7 fold)	Koo et al., 2000
T214V	Reduce activity (3 fold)	Koo et al., 2000
	Lauric acid hydroxylation	
E212D	Loss of activity	Puckhaev et al., 2002
E212Q	No effect on binding	Puckhaev et al., 2002
R154A	Reduce binding affinity (3 fold)	Puckhaev et al., 2002
R154Q	Reduce binding affinity (10 fold)	Puckhaev et al., 2002
D77R	No effect on binding	Koo et al., 2001
T214V	Increase binding affinity (5-fold)	Koo et al., 2001
D77R/T214V	Increase binding affinity (5-fold)	Koo et al., 2001
	Thermostability	
Y2A	Decrease in T_m (10°C)	Puckhaev et al., 2002
W4A	Decrease in T_m (10°C)	Puckhaev et al., 2002
F24S	Decrease in T_m (10°C)	Maves et al., 2000
W231A	Decrease in T_m (10°C)	Puckhaev et al., 2002
Y250A	Decrease in T_m (10°C)	Puckhaev et al., 2002
W281A	Decrease in T_m (10°C)	Puckhaev et al., 2002
W4A/W281A	Decrease in T_m (15°C)	Puckhaev et al., 2003
T213X	No effect on T_m	Puckhaev et al., 2002
T214X	No effect on T_m	Puckhaev et al., 2002
Y26A/L308A	Decrease in T_m (16°C)	Meharenna et al., 2010
	Other activity	
T213A/C317H	Carbene transfer activity	McIntosh et al., 2015

Alanine scanning was performed on residues of aromatic cluster (Tyr2, Trp4, Trp231, Tyr250, Trp281) to understand the effect on thermostability of CYP119. All mutants decreased the T_m value of enzyme approximately 10°C. However double mutants did not decrease thermostability further than 10°C (Puckhaev et al., 2002). W4A/W281A mutant is only exception with 15°C decrease in T_m value (Puckhaev et al., 2003). F24S mutant also showed 10°C decrease in T_m value (Maves et al.,2000). Double mutant Y26A/L308A has T_m value 76°C that is 16°C lower than wild type enzyme (Meharenna et al., 2010). Molecular dynamics simulations of Y26A/L308A mutant unfolds faster than wild type CYP119 and it also make Cys loop of enzyme more unstable (Meharenna et al., 2010). It can be predicted that each mutation from aromatic cluster residues will result at least 10°C decrease in thermostability without depending on residue type.

1.3 Progesterone

Steroids are important pharmaceutical molecules because of various effects of hydroxylated derivatives. Steroid based products follow antibiotics in the biopharmaceutical market (Tong and Dong, 2009). The number of approved steroid-based drugs was more than 300 at 2012 (Donova and Frogova, 2012). CYP-steroid interaction is important because CYP enzymes can catalyze hydroxylation of chemically inaccessible residues (Fernandes et al., 2003). The main problem of steroid catalysis is the transport of steroid molecules into the cell. Whole cell biocatalysis preferred for industrial applications of CYPs because of redox partner problem, but steroid molecules cannot be transported to cell actively, they get into cell by diffusion (Mendel, 1989). There are several ways to overcome this problem and one of them is increasing substrate concentration. Since steroids have low solubility it is not an effective approach. For example; solubility of progesterone in water is 1μM and testosterone solubility is 97μM (Zehentgruber et al., 2010). The second way is increasing membrane permeability of cells with chemicals, but this will obviously affect the large-scale production process. The last method is purification of enzyme and catalyze the hydroxylation *in vitro*, which requires soluble enzymes like CYP119.

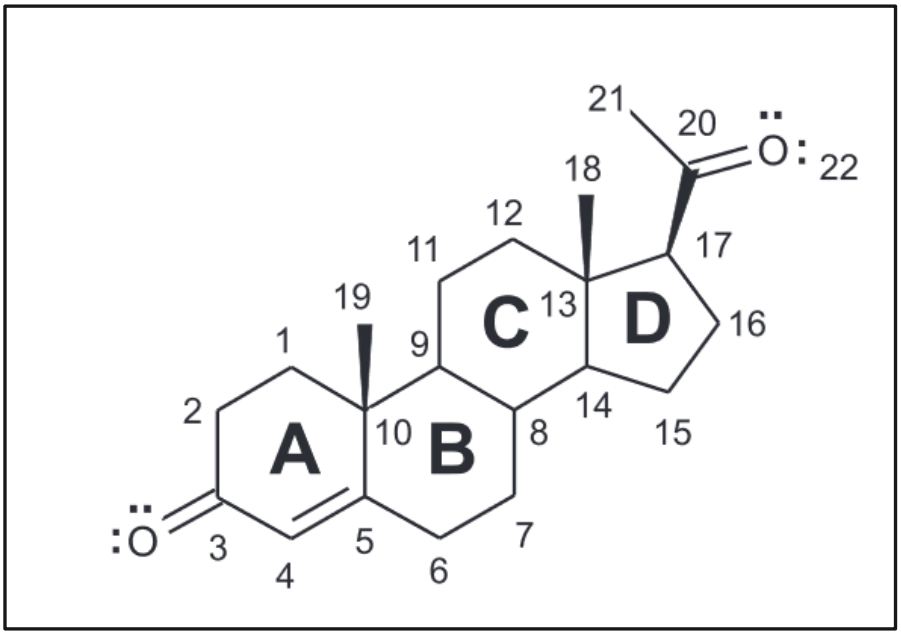

Figure 1.17 Progesterone nomenclature system. (Source: Kiralj and Ferreira, 2003)

Progesterone is a female sex hormone which is naturally produced during the ovulation period. Hydroxylated derivatives of progesterone also important for hormonal system of humans since these molecules used as precursors for synthesis of other steroid based hormones like; cortisone, testosterone and estrogen. Figure 1.17 shows progesterone structure with proper naming. Most positions of progesterone are available for hydroxylation. C2 atom of A ring, C6 atom of B ring, C11 and C12 atoms of C ring and C15, C16 and C17 atoms of D ring are determined as most possible hydroxylation sites of progesterone structure (Khatri et al., 2015). However, there is no data for hydroxylation at positions 3, 5, 10 and 13 in the literature.

1.3.1 Progesterone Derivatives

1α- hydroxyprogesterone does not have any known particular function as a hormone or a signaling molecule. However, hydroxylation of C1 atom of progesterone allows adding different functional groups like methyl to progesterone. 1-methylprogesterone is the most important molecule for synthesis of mesterolone and

methenolone which are anabolic steroids (Schanzer and Donike, 1993). Mesterolone used for treatment of low testosterone levels and methenolone used for treatment of anemia (Camerino and Sciaky, 1975). C1 hydroxylation of progesterone is very complicated and multi-step process (requires nine additional steps) with synthetic catalysis (Mann and Pietrzak, 1989).

9α-hydroxypregnane molecules show progestational and glucocorticoidal activity. 9α-hydroxyprogesterone derivatives are intermediate molecules for pregnane synthesis. Fluoxymesterone and fludrocortisone are known examples of 9α-hydroxypregnane derivatives and used as drug molecules.

11α-hydroxyprogesterone shows antiandrogenic activity with minimal progestational activity thus can use as a drug for treatment of hair loss in men (Willigen et al., 1987). It also acts as a mineralocorticoid and as effective as aldosterone (Fiet et al., 1989)

16α-hydroxyprogesterone is a side product of 17α-hydroxyprogesterone production reactions of human CYP17A1. Research showed increased levels of 16α-hydroxyprogesterone during pregnancy thus it might be related to successful pregnancy (Stiefel and Ruse, 1969). It is also act as a mineralocorticoid and effective on salt balance of kidneys (Quinkler et al., 2003)

17α-hydroxyprogesterone is the most common product of progesterone hydroxylation which can be used as a drug for the treatment of endometrial cancer and prevention of premature labor (Shindler et al., 2003). C17 hydroxylation allows adding several side chains to progesterone like C1 hydroxylation. C17 derivatives of progesterone used as drugs for prevention of preterm delivery (17α-hydroxyprogesterone caproate) (Romero and Stanczyk, 2013), birth control pills (chlormadinone acetate) (Kuhl et al., 2005), treatment of prostate cancer (cyproterone acetate) (Raudrant and Rabe, 2003) and menopausal hormone therapy (medroxyprogesterone acetate) (Kuhl et al., 2005).

21-hydroxyprogesterone also known as deoxycorticosterone or DOC. It has a mineralocorticoid activity and pre-cursor molecule of aldosterone synthesis (Mattox et al., 1969). 21-hydroxyprogesterone can use as hypertensive drug molecule and metabotoxin in high levels thus it can be used to treat adrenal insufficiency (Deng and Li, 2005). Acetate form of 21-hydroxyprogesterone (deoxycorticosterone (DOCA)) already used for treatment of Addison’s disease.

1.3.2 Progesterone Hydroxylating P450s

CYP106A2 is a steroid hydroxylase from *Bacillus megaterium* which catalyze hydroxylation mainly 15β position (Rauschenbach et al., 1993). Enzyme can catalyze 15β-hydroxylation of progesterone, testosterone and 11-deoxycorticosterone (Berg et al., 1976). CYP106A2 can also produce 9α, 11α and 6β hydroxyprogesterone and -testosterone as side products (Zehentgruber et al., 2010). Amount of side products depend on the redox partner of protein. CYP106A2 normally uses adrenodoxin (Adx) as redox partner but using Etp1(516-618), Adx-like electron transfer protein-1, as electron donor protein significantly reduces the amount of side products (Sagadin et al., 2018).

Corynespora cassicola, *Norcardia canicruria* and *Rhodococcus rhodochrous* are patented organisms for large scale 9α-hydroxyprogesterone production with their own P450 enzyme systems (Rauschenbach et al., 1993, Mersback et al., 1983, Masuda et al., 1989).

CYP17A1 catalyzes production of 17α-hydroxyprogesterone as main product (Gilep et al., 2011). Several mutants of CYP17A1 designed with computational methods to change regioselectivity of enzyme. L206A, V366A and V483A mutants change regioselectivity of enzyme and increase the amount of side product 16α-hydroxyprogesterone up to 40% (Morlock et al., 2018).

CYP260A1 from *Sorangium cellulosum* catalyzes the production of 1α-hydroxyprogesterone. Rational design approach used to change regioselectivity of CYP260A1. S276N mutant catalyzes the 1α-hydroxylation with improved activity, while S276I mutant produce 17α-hydroxyprogesterone as main product (Khatri et al., 2018).

Canine enzyme CYP2B11 produces 21α-hydroxyprogesterone. Mutating residue Leu363 (substrate recognition residue, 5^{th} residue after EXXR motif) to valine changes the regioselectivity of CYP2B11 and produce 16α-hydroxyprogesterone as main product (Born et al., 1995).

CYP154C5 catalyze 16α-hydroxyprogesterone production with high regioselectivity. Crystal structure of CYP154C5 clearly explains regioselectivity of enzyme with the size and polarity distribution of active site (Jozwik et al., 2016).

Human CYP21A2 is responsible to 21-hydroxyprogesterone and 17-hydroxyprogesterone production. These two chemicals used for synthesis of 11-

deoxycorticosterone and 11-deoxycortisol which are important hormones for immune system activity (Miller and Auchus, 2011; Guengerich, 2015).

Human CYP3A4 selectively catalyze 6-hydroxylation of steroid-based substrates (Wang et al., 2004).

1.4 Aim of the Project

This project aims to design the thermophilic CYP119 enzyme using rational design methods for efficient production of hydroxylated progesterone derivatives. Efficient production of progesterone derivatives is a remaining challenge for pharmaceutical industry. CYP enzymes can produce hydroxyprogesterones with high regio- and stereoselectivity. CYP119 as a thermophilic CYP enzyme can increase the yield of hydroxyprogesterone production while decreasing the required amount of money and time for process.

CHAPTER 2

METHODOLOGY

2.1 Structural Alignment of CYP119

CYP enzyme family share a common three-dimensional structure and this can be used to identify structural elements of CYP119 enzyme. UGene alignment program used for structural alignment of CYPs that can bind progesterone naturally. ClustalW alignment method were used for structural alignment process. Sequence alignment of CYP119 were performed with BLAST and phylogenetic trees were constructed with JalView program with using Neighbor-Joining (NJ) method.

2.2 Design of CYP119

Rational design approaches that aim to change the ligand specificity of an enzyme require high-resolution structural data. Crystal structures that have resolution value below 2.0 Å should be used for enzyme design process to make accurate predictions. Enzymes generally show structural differences upon substrate binding like rearranging some secondary elements or having more open active site. Ligand docking simulations should use substrate bound models to get more realistic results.

Since original substrate of CYP119 is not known, CYP119 does not have a substrate bound structural model but have six different inhibitor bound models as shown in Table 1.1. 4-phenylimidazole bound model (PDB ID: 1F4T) has the lowest resolution (1.93 Å) among all inhibitor bound models thus 1F4T used for designing progesterone binding CYP119 enzymes.

2.2.1 PyRosetta Program

PyRosetta is the Python based version of Rosetta molecular modelling suite which is written in C language (Chaudhury et al., 2010). Rosetta is especially effective on protein folding research, but it also has modelling packages for small molecule docking (Lemmon et al., 2012; Davis et al., 2009), protein-protein docking (Weitzner et al., 2017; Marze et al., 2017), DNA/RNA – protein docking (Miao et al., 2017) and antibody design (Lyskov and Gray, 2008). Rosetta online server webpage is free for academic users (Lyskov et al., 2013). PyRosetta works as an interface to Rosetta package and uses same functions with Rosetta.

PyRosetta 3 was used for mutant design and ligand docking procedure in this thesis project. Since PyRosetta 3 is written in Python 2 scripting language, all codes written with Python 2.7.16 package. PyCharm Python IDE used as idle for writing and debugging final script for design. IPython command shell used for running PyRosetta program.

PyRosetta copies structural data of given structure into a "pose". There are three different commands to create a pose;

```
pose = pose_from_pdb("pdb_name" + ".pdb")
# Creates a pose with using the pdb file from computer directory.
pose = pose_from_rcsb("pdb_name")
# Downloads pdb file from RCSB web site and creates the pose.
pose = pose_from_sequence("amino acid sequence", "residue type")
# Creates a pose with using given sequence.
```

PyRosetta applies all user defined protocols on this pose without changing original structure. User can save this altered pose as a new pdb file with following command;

```
pose.dump_pdb("pdb_name" + ".pdb")
```

Copy of a pose can be created with following command;

```
new_pose.assign(pose)
```

PDB files must prepared before starting the design process. That preparation includes removing all water molecules and cofactors from PDB file. Python code for cleaning PDB file;

```
from rosetta.toolbox import cleanATOM
cleanATOM("pdb_name" + ".pdb")
```

This script removes all non-ATOM lines from original pdb file, creates a new pdb file with remaining lines and save new file as pdb_name.clean.pdb. Figure 2.1 shows an example for clean.pdb file.

Atom Type	Atom Number	Atom Type	Residue Name	Chain Name	Residue Number	Atom Coordinates			Occupancy	B-factor
						X	Y	Z		
ATOM	1	N	MET	A	1	37.383	12.642	25.681	1.00	30.52
ATOM	2	CA	MET	A	1	38.655	11.934	25.364	1.00	31.06
ATOM	3	C	MET	A	1	39.471	11.594	26.608	1.00	29.96
ATOM	4	O	MET	A	1	40.077	10.524	26.686	1.00	28.94
ATOM	5	CB	MET	A	1	39.525	12.794	24.434	1.00	32.74
ATOM	6	CG	MET	A	1	38.938	13.058	23.048	1.00	35.28
ATOM	7	SD	MET	A	1	38.794	11.559	22.054	1.00	38.64
ATOM	8	CE	MET	A	1	40.540	11.187	21.762	1.00	38.50
ATOM	9	HA	MET	A	1	38.412	11.005	24.847	1.00	0.00
ATOM	10	HB2	MET	A	1	39.687	13.756	24.920	1.00	0.00
ATOM	11	HB3	MET	A	1	40.490	12.302	24.312	1.00	0.00
ATOM	12	HG2	MET	A	1	39.584	13.762	22.524	1.00	0.00
ATOM	13	HG3	MET	A	1	37.949	13.503	23.162	1.00	0.00
ATOM	14	HE1	MET	A	1	40.623	10.283	21.158	1.00	0.00
ATOM	15	HE2	MET	A	1	41.043	11.034	22.717	1.00	0.00
ATOM	16	HE3	MET	A	1	41.006	12.020	21.235	1.00	0.00
ATOM	17	H1	MET	A	1	36.821	12.067	26.292	1.00	0.00
ATOM	18	H2	MET	A	1	37.591	13.518	26.139	1.00	0.00
ATOM	19	H3	MET	A	1	36.876	12.821	24.826	1.00	0.00
ATOM	20	N	TYR	A	2	39.487	12.496	27.586	1.00	29.67
ATOM	21	CA	TYR	A	2	40.296	12.259	28.778	1.00	29.95
ATOM	22	C	TYR	A	2	39.806	11.204	29.768	1.00	30.02

Figure 2.1 Example of a clean.pdb file.

PyRosetta has a class of scoring functions and it can be used for calculating total energy score of whole structure or energy between two amino acid residues of pose. Energy score has the generic unit of Rosetta Energy Unit (REU). Models that have more negative REU scores generally accepted to be more native like (Misura et al., 2006). Scoring parameters that used for calculating energy scores are given in Figure 2.2. Score function parameters originated from analysis of protein structures from RCSB (Rohl et

al., 2004). Measurements like distance between two polar atoms or angle between hydrogen bonds are converted to energy function with Bayesian statistics (Simons et al., 1997; Dunbrack and Cohen, 1997). Rosetta scoring function is a combination of physic-based and knowledge-based information about structure. Addition of statistical parameters to the score function led improved the correction of Lennard-Jones score terms (Song et al., 2011).

Rosetta Scoring Terms

fa_atr	FA	van der Waals net attractive energy
fa_rep	FA	van der Waals net repulsive energy
hbond_sr_bb, hbond_lr_bb	FA/CEN	Hydrogen-bonding energies, short and long-range, backbone–backbone
hbond_bb_sc, hbond_sc	FA	Hydrogen-bonding energies, backbone–side-chain and side-chain–side-chain
fa_sol	FA	Solvation energies (Lazaridis–Karplus)
fa_dun	FA	Dunbrack rotamer probability
fa_pair	FA	Statistical residue–residue pair potential
fa_intra_rep	FA	Intraresidue repulsive Van der Waals energy
fa_elec	FA	Distance-dependent dielectric electrostatics
pro_close	FA	Proline ring closing energy
dslf_ss_dst, dslf_cs_ang, dslf_ss_dih, dslf_ca_dih	FA	Disulfide statistical energies (S–S distance, *etc.*)
Ref	FA/CEN	Amino acid reference energy of unfolded state
p_aa_pp	FA/CEN	Propensity of amino acid in (ϕ,ψ) bin, P(aa|φ,ψ)
Rama	FA/CEN	Ramachandran propensities
Vdw	CEN	van der Waals "bumps" (repulsive only)
Env	CEN	Residue environment score (statistical)
Pair	CEN	Residue–residue pair score (statistical)
Cbeta	CEN	β-carbon score

Figure 2.2 List of Rosetta scoring terms. (Source: Appendix B - The PyRosetta Interactive Platform for Protein Structure Prediction and Design)

PyRosetta has two main score function classes; centroid (low resolution) and all-atom (high resolution). Centroid docking generally used for *de novo* folding or loop design procedures (Simons,1999). In this method all side chains represented as single atom containing centroid structures around backbone, which still reflects the chemical and physical properties of side chains but limits the free rotation of them. This limitation simply reduces computational time which can be very high number in *de novo* protein design. Users can perform additional steps to get right rotation for each amino acid side chain after creating properly fold protein structure.

All-atom score function includes all side chain atoms and generally used for protein-ligand docking procedures. Scoring function parameters generally depend on knowledge-based terms (Khulman and Baker, 2000; Neria et aal., 1996). Conformation

of each side chain is decided according to Metropolis criterion (Dunbrack and Karplus, 1993).

PyRosetta has multiple score functions with different purposes. Weights of individual scoring parameters cause the difference between score functions. User can also set weights of each score to create own score function. “REF2015” and “ligand” score functions used for this thesis project. Figure 2.3 shows the weights of REF2015 score function. REF2015 was used for calculating energy scores of wild type and mutant enzymes. Relax procedure was also applied according to REF2015 score function.

Scores	Weight
fa_atr	1.000
fa_rep	0.550
fa_sol	1.000
fa_intra_rep	0.005
fa_intra_sol_xover4	1.000
lk_ball_wtd	1.000
fa_elec	1.000
pro_close	1.250
hbond_sr_bb	1.000
hbond_lr_bb	1.000
hbond_bb_sc	1.000
hbond_sc	1.000
dslf_fa13	1.250
omega	0.400
fa_dun	0.700
p_aa_pp	0.600
yhh_planarity	0.625
ref	1.000
rama_prepro	0.450

Figure 2.3 Weights of REF2015 score function.

Ligand docking energy scores are important to understand efficiency of ligand binding. PyRosetta energy score function “ligand” is used for calculating final energy scores of ligand bound mutant enzymes. Figure 2.4 shows the weights of ligand score function.

Scores	Weight
fa_atr	0.800
fa_rep	0.400
fa_sol	0.600
fa_elec	0.250
pro_close	1.000
fa_pair	0.800
hbond_sr_bb	2.000
hbond_lr_bb	2.000
hbond_bb_sc	2.000
hbond_sc	2.000
dslf_ss_dst	0.500
dslf_cs_ang	2.000
dslf_ss_dih	5.000
dslf_ca_dih	5.000
atom_pair_constraint	1.000
coordinate_constraint	1.000
angle_constraint	1.000
dihedral_constraint	1.000
omega	0.500
fa_dun	0.400
p_aa_pp	0.500
ref	1.000
chainbreak	1.000

Figure 2.4 Weights of ligand score function.

Following code explains how to create and set a score function;

```
from rosetta import *
from toolbox import *
init()
scorefxn =create_score_function(score function name)
scorefxn(pose)
protocol_name.set_scorefxn(scorefxn)
```

2.2.1.1 Inclusion of Heme Cofactor

There are several computer programs and online servers for rational design of enzymes. Generally, all docking programs need the removal of water and cofactor

molecules from structural protein data. Removal of cofactors, as in the case heme group, is the main problem while working with P450 enzyme family. Heme group is located at the center of active site and removal of heme will cause false positive ligand docking results.

PyRosetta removes heme group during the cleaning step of pdb file like other design programs. CleanATOM command removes every line starts with HETATM. Every atom belongs to other molecules than amino acids referred as heteroatom and shown as HETATM in pdb files. Although PyRosetta can recognize several cofactors which have required parameter files in the database of program, heme group is not one of them. User should add the necessary files to program database manually. Required files for this process;

HEM.pdb

Create a pdb file with using original coordinates of heme group from target enzyme.

HEM.mol2

Use a .pdb to .mol2 converter to create .mol2 file. (Chimera used for this thesis project.)

Open .mol2 file with a text editor change the pdb file name (pdb_name.pdb → HEM.pdb)

HEM.params

Params file is essential to make heme group recognizable.

molfile2params.py file is necessary for creating params file. User must download it manually. Protocol can be downloaded from PyRosetta web site (http://www.pyrosetta.org/scripts).

Mol file and pdb file of cofactor must be in the PyRosetta\molfile2params directory

Command to create params file with IPython;

molfile_to_params.py <.mol2 filename> -n <Residue name>

Open .params file with a text editor and change the string name in line 2 (IO_STRING PDB_NAME Z →IO_STRING HEM Z)

Move params file to:

C:\\ PyRosetta\database\chemical\residue_type_sets\fa_standard\residue_types

User must add path of params file to the residue database file:

C:\\ PyRosetta\database\chemical\residue_type_sets\fa_standard\residue_types.txt

Original coordinates of heme group must be added to cleaned protein file. PyRosetta should recognize heme group while loading pdb file to a pose, if all steps are completed without any mistake.

2.2.1.2 Mutations

There are two different methods for creation of mutations in PyRosetta. First method is direct mutation of a residue and it is actually not applicable to high resolution design. Following code;

mutate_residue(pose, residue number, “letter code of mutation”)

creates a mutation but does not optimize rotamers. User can use “minmover” function to change side chain rotation to minimize local energy, but this method does not give realistic results and is generally used for protein – protein docking studies when the shape of protein is more important than the side chain rotation of each amino acid.

Second method requires specific file (resfile) to create mutation on the structure. Following code converts the pdb file data of protein structure to resfile;

generate_resfile_from_pose(pose, “mutant name” + “.resfile”)

There are specific abbreviations that can use on resfile to create specific mutations. Figure 2.5 shows resfile codes.

Resfile Codes

NATRO	Use the native amino acid residue and native rotamer (do not repack)
NATAA	Use the native amino acid residue but allow repacking to other rotamers
PIKAA ILV	Pick from amino acid residues Ile, Leu, and Val and allow repacking
ALLAA	Use all amino acid residues and allow repacking

Figure 2.5 Specific mutation codes for resfile. (Source: Appendix B: The PyRosetta Interactive Platform for Protein Structure Prediction and Design)

User must create a specific resfile for each mutation, but it is possible to design multiple mutations on the same resfile to create double/triple etc. mutants. Since resfile does not contain any information about residue names, user must know the chain name and residue number of residues that are selected for mutation. Figure 2.6 shows a resfile example with W21G mutation.

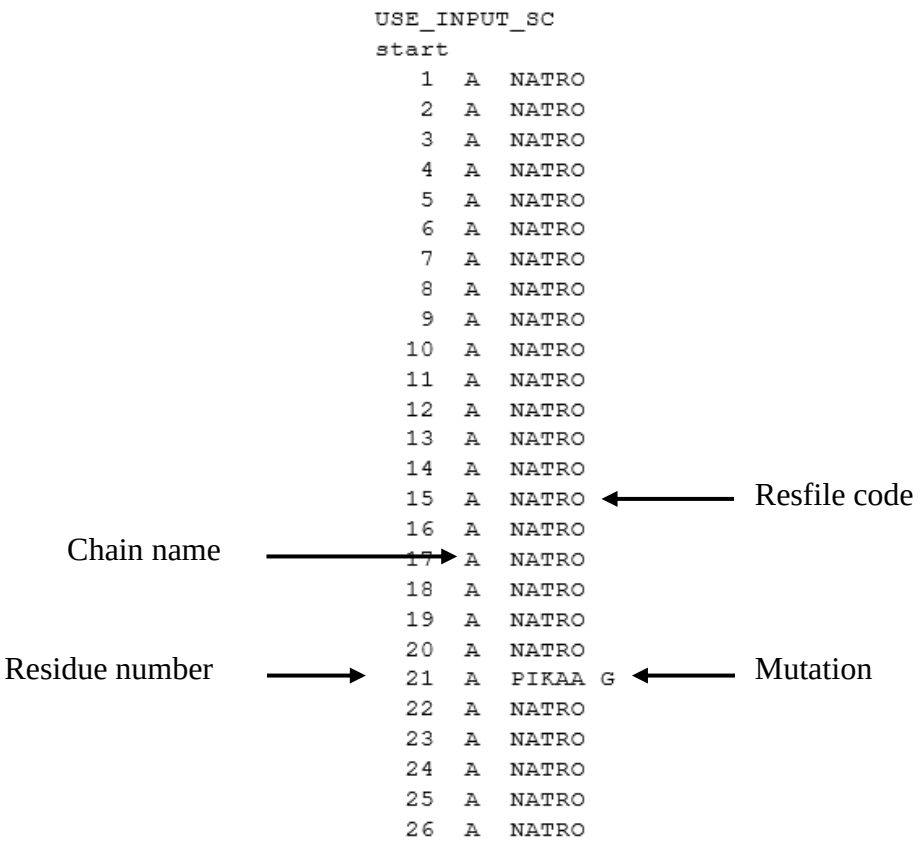

Figure 2.6 W21G.resfile given as an example of resfile.

PyRosetta has TaskFactory module for applying a combination of different commands on the same pose. Following code explains how to create a task to apply mutations and how to select true rotation of new residues' side chain;

```
from rosetta.protocols.rigid import *
from rosetta.core.pack.task import TaskFactory
task_design = standard_task_factory()
task_design.push_back(ReadResfile("mutant name" + ".resfile"))
pack_mover =PackRotamersMover(score_function)
pack_mover = task_factory(task_design)
pack_mover.apply(pose)
```

PackRotamersMover measures the phi and psi angles for mutant residue and takes corresponding chi angles for side chain from Dunbrack Rotamer Library and minimize the side chain according to given score function (REF2015). Dunbrack rotamer library

includes the preferred side chain conformations of amino acids based on experimental results (Leach et al., 1994).

Last step of mutant creation process is energy minimization of the whole structure. PyRosetta has two different relax functions for structure minimization. Relax protocols minimize energy of structure according to given score function (REF2015). ClassicalRelax function is generally used for preparing structures for high resolution docking process because it applies energy minimization to all components of structure including backbone and side chains. This process requires high RAM capacity since it requires huge number of calculations.

FastRelax function is generally used for protein folding process or preparing structures for protein-protein docking (centroid docking) because it applies energy minimization only to backbone structure. FastRelax protocol used for this project because it is 12-fold faster than ClassicalRelax and still effective for structure minimization since side chain of mutant residue already minimized with PackRotamersMover protocol. Code for FastRelax protocol;

```
from rosetta.protocols.relax import *
relax  = FastRelax()
relax.set_scorefxn(scorefunction)
relax.apply(pose)
```

User should calculate the final energy score of mutant protein structure to understand the effect of mutation on the stability of protein. Following code saves the mutated and relaxed structure and calculates energy score of final model;

```
pose.dump_pdb("mutant name" + ".pdb")
print scorefxn(pose)
```

Workflow for mutation creation process can be summarized as;

- load wild type enzyme data to PyRosetta
- generate residue files (.resfile) for each mutation
- create a task to read resfile, apply mutation and select best rotamer for mutant residue(s).
- minimize structure (FastRelax protocol)
- save mutant enzyme and calculate energy score

2.2.1.3 Ligand Docking

Effectiveness of Rosetta ligand protocol was checked with using experimental results and 64% of structures that gave lowest score, had only 2 Å RMSD value when compared with experimental data (Davis and Baker, 2009). Comparison of ligand docking performance of Rosetta and other dock programs like Dock, FlexX, Glide, GOLD, AutoDock showed that Rosetta's performance is equal or better than these programs (Davies et al., 2009).

Rosetta ligand docking protocol accepts both ligand and protein structure as flexible. Minimization of backbone phi and psi angles with a gradient based approach used for modelling protein flexibility (David and Baker, 2009). PyRosetta ligand dock protocol consist of two steps. First step is a low-resolution docking step and checks structures for shape complementarity. Low-resolution docking step is known to be very fast. Second step includes the minimization of side chains and ligand molecule according to Monte Carlo principle. After high resolution step PyRosetta performs an additional gradient minimization on models before calculating final energy score (DeLuca et al., 2015).

Rational design programs require a starting point for ligand docking process as a rule. User should be able to give coordinates of the area that is thought to be active site. Programs like Chimera and AutoDock use a box (Grid box) and perform ligand docking using the coordinates of box as borders of the docking area. PyRosetta uses DockMCM (Dock Monte Carlo Metropolis) protocol for ligand-protein docking process which is specialized for high resolution ligand docking and uses ligand score function to calculate ligand binding scores. Protocol requires coordinates of ligand as a starting point because PyRosetta does not have a special protocol for binding site prediction. This means user must have information about binding site of protein. Before applying docking protocol PyRosetta performs six cycles of side chain rotamer sampling for ligand flexibility. This includes the movement of ligand 0.1 Å in a random direction within 0.05 radians. Ligand torsion angles get minimized according to harmonic constraints (movement within 0.05 radians equals to 1 s.d. of harmonic function) in each cycle of minimization (Combs et al., 2013). Amino acid side chains repack after each minimization step according to Dunbrack rotamer library (Dunbrack and Karplus, 1993).

PyRosetta calculates ligand binding score after every movement and if it is better than the initial score new coordinates accepted as a starting point. If ligand binding score is worse than the original score PyRosetta uses the original coordinates for starting coordinates again. PyRosetta does not accept any docking results with clashes and locates ligand out of structure in these scenarios, continues to use previous coordinates.

Ligand file preparing process is very similar to heme adding protocol. User should prepare a pdb file of ligand. Residue name (three letter code for ligand name - LGN) and chain name (should be X) must be changed with a text editor. Pdb file must be located on PyRosetta/molfile2params directory. Coordinates of ligand pdb file must be added between the coordinates of protein and heme group for each mutant file.

A mol file also must be prepared to create params file for ligand. User should change the string name in line 2 (IO_STRING PDB_NAME Z →IO_STRING LGN Z) and move params file to:

C:\\ PyRosetta\database\chemical\residue_type_sets\fa_standard\residue_types

User should define the parameters for ligand to PyRosetta before opening ligand included pdb file. Code to make ligand recognizable for PyRosetta;

```
params_list = Vector1([“ligand name.params”])
# Input of Vector1 is a list so user can add more ligands into the structure.
res_set = generate_nonstandard_residue_set(params_list)
```

Before starting docking process score function of docking protocol must be determined with the given code;

```
from rosetta.protocols.ligand_docking import *
scorefxn = create_score_function(“ligand”)
docking = DockMCMProtocol()
docking.set_scorefxn(scorefxn)
```

Docking process requires a “job distributor” module since it is a multistep procedure. PyJobDistributor opens the pose, applies docking protocol according to the given score function up to number of docking rounds which determined by user. Number of docking rounds should be at least 1000 to understand the exact position of ligand in the active site. 10 rounds of docking performed in the initial steps of this project since they are informative about the enzyme’s ligand binding capability. 1000 rounds of docking applied the final selected models to identify the possible products of progesterone hydroxylation. Following code explains how to create a job in PyRosetta;

```
pose = Pose()
```

```
pose_from_pdb(pose, res_set, “mutant name.pdb”)
job_output = “mutant name” + “ligand name”
jd = PyJobDistributor(job_output, number of docking rounds, scorefxn)
test_pose = Pose()
counter = 0
while not jd.job_complete:
    test_pose.assign(pose)
    counter +=1
    test_pose.pdb_info().name(job_output  +  “_” + str(counter))
    docking.apply(test_pose)
    test_pose.pdb_info().name(job_output + “_” +str(counter)+ “_fa”)
    jd.output_decoy(test_pose)
```

Last step of ligand docking procedure includes binding score calculation and elimination of bad results that have positive REU score. Same code used for elimination of docked models that have energy score higher than -1180 REU for the final round of docking process. -1180 REU was identified as threshold value because -1180 REU is the highest REU score that we get from ligand docking of natural progesterone hydroxylating CYPs.

Following code used for elimination process;

```
import os
scr = int(scorefxn(pose)
if scr >= 0:
os.remove(r“C:\Users\DockingResults\\”+“docked_file_name.pdb”)
```

Workflow for docking process can be summarized as;

- insert ligand coordinates to mutant .pdb file
- define parameters of ligand and score function to high-resolution docking protocol (DockMCM protocol)
- create a job to load mutant pdb file and ligand parameter file (.params)
- define number of docking rounds and run docking protocol
- eliminate results according to energy scores of docked models

Complete code written for this thesis project given in Appendix A.

2.2.2 UCSF Chimera Program

Chimera is a powerful program for visualization and analysis of protein structures for different purposes like sequence alignment, ligand docking, density maps etc (Petterson et al., 2004). Chimera was used as a visualization tool for this thesis project. We also used ligand docking results of program to control the efficiency of PyRosetta ligand docking scores (Table 2.1). Progesterone docking process of Chimera was applied to randomly selected mutants from first round of mutations. Ligand binding scores of Chimera and PyRosetta plotted (Figure 2.7) to check linear correlation between these two programs binding scores and also check the competence of our PyRosetta docking script.

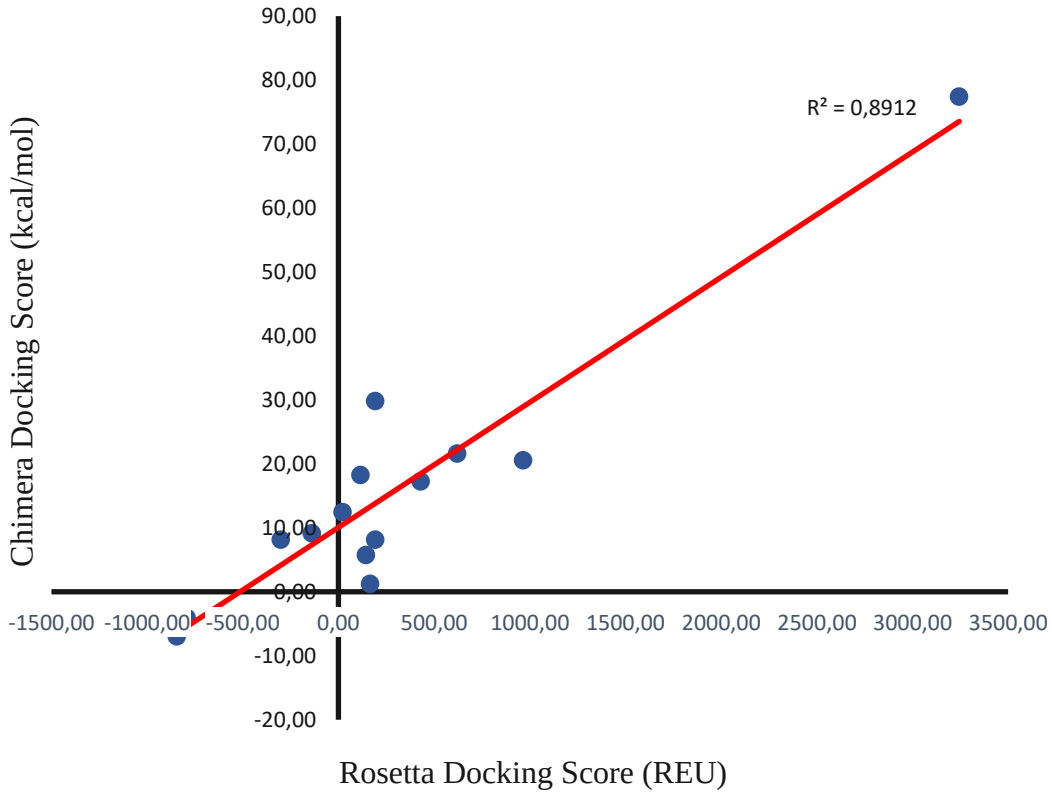

Figure 2.7 PyRosetta vs. Chimera docking results plot.

Table 2.1 Ligand docking scores for PyRosetta and Chimera.

Mutants	Ligand Binding Scores	
	PyRosetta (REU)	Chimera (kcal/mol)
V151R	-809,00	-4,50
A209R	966,00	20,50
A209G	116,00	18,20
L69G	-792,00	-4,10
L69G	430,00	17,20
L205G	23,00	12,40
A209F	3245,00	77,40
V151R	-844,00	-7,00
L354M	-138,00	9,10
T213A	193,00	29,80
T214E	-301,00	8,10
L354M	145,00	5,70
I208E	167,00	1,20
F153A	622,00	21,60
T214A	193,00	8,10

CHAPTER 3

RESULTS AND DISCUSSION

3.1 Sequence Alignment of CYP119

Enzymes with high sequence similarity generally share same substrate type selectivity. Since CYP119's original substrate is not known, we decided to perform a sequence alignment to create a phylogenetic tree for CYP119 and try to get an idea about enzyme's substrate preferences with using the data of similar CYPs. JalView program used for sequence alignment and tree created with NJ method (Figure 3.1) (Waterhouse et al., 2009). According to sequence alignment results; the most related CYPs are originated from *Bacillus sp.* however, original substrates of these enzymes are also not known.

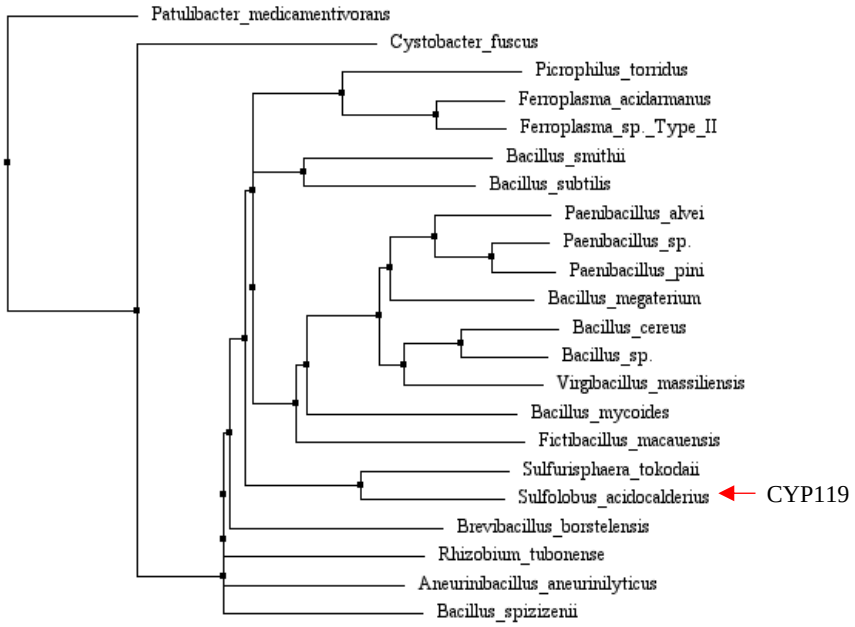

Figure 3.1 Phylogenetic tree of CYP119 gene created with Jalview program.

3.2 Structural Alignment of CYP119

CYP enzymes share a common three-dimensional structure even though they generally have low sequence similarity. This feature can be used to identify important parts of a CYP enzyme, as in the case CYP119. Park et al., already identified some important heme binding residues (Arg259, Thr257, His315, Cys317, Glu212) of CYP119 (Park et al., 2002). Thermostability related residues were also defined as the aromatic cluster I and II (Nishida et al., 2005). Prosser and colleagues used structural alignment to identify important residues of CYP27A1 (Prosser et al., 2006). In this work we aligned CYP119 to the structural alignment by Prosser and colleagues to identify all secondary structure elements, SRS regions, structurally important residues and conserved structural elements of CYP119. Figure 3.2 shows secondary structure of CYP119 named according to nomenclature.

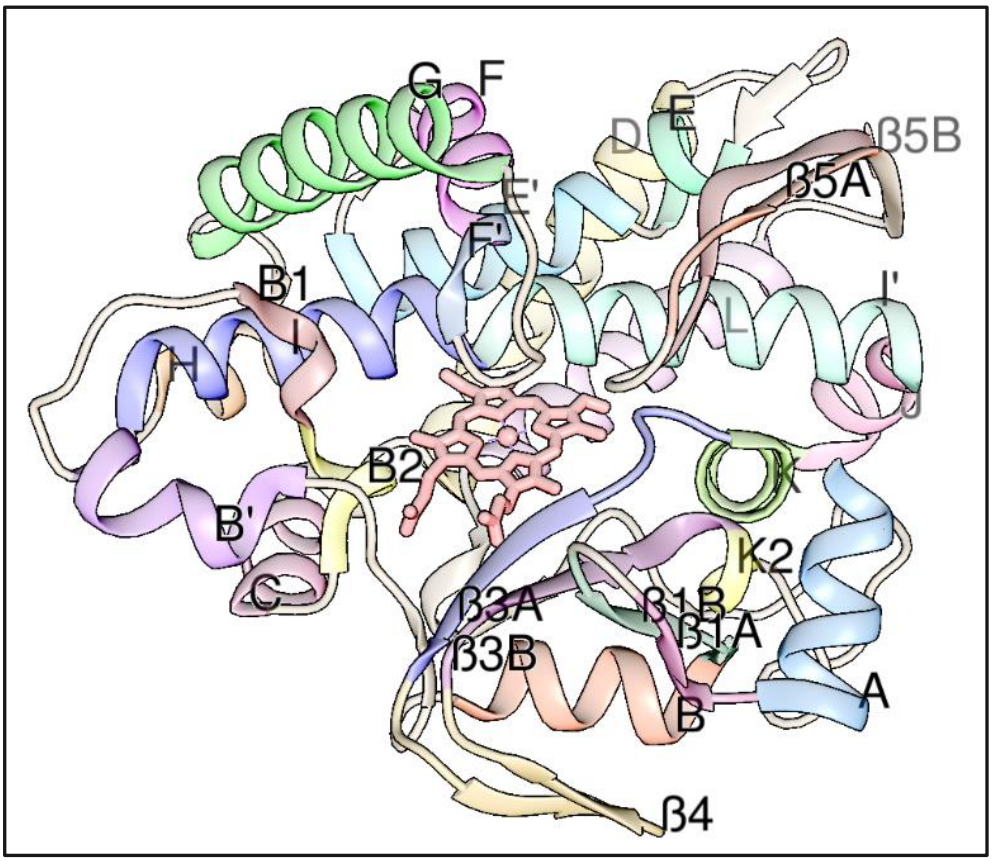

Figure 3.2 Crystal structure of CYP119 with annotated secondary structure elements.

Important residues of CYP119 were identified according to secondary structure information of enzyme. These important residues were classified as; thermostability related regions, heme binding regions, substrate binding residues, ERR triad, EXXR motif and residues forming conserved H bonds. Figure 3.3 shows important residues of CYP119.

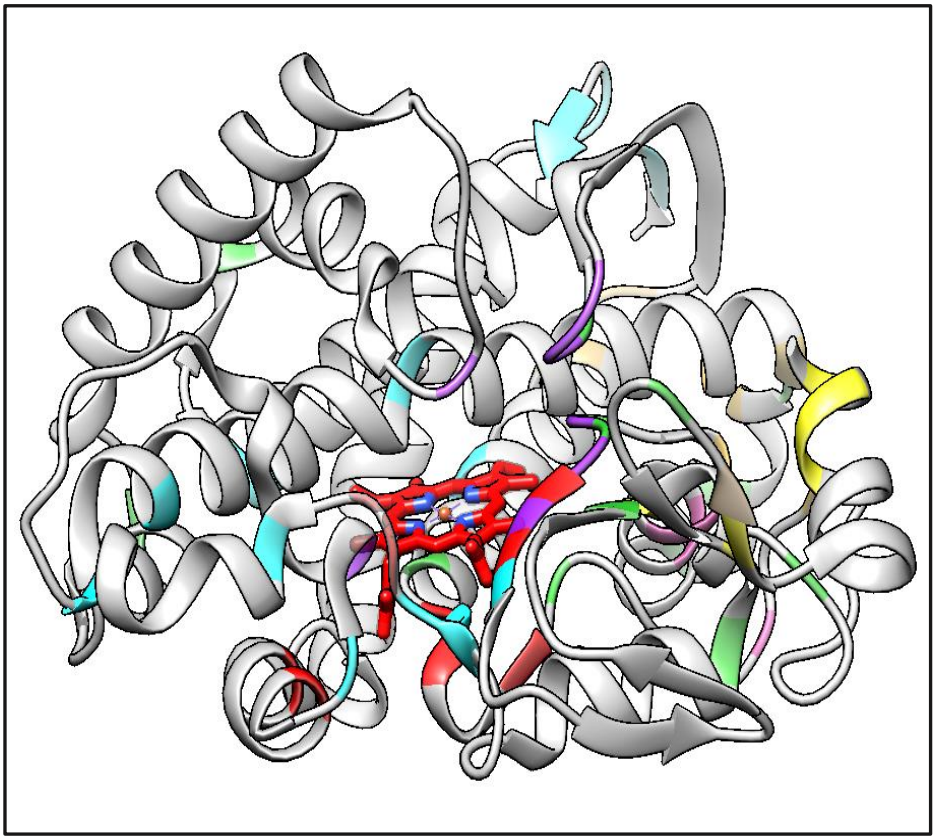

Figure 3.3 CYP119 structure with color coded important residues. (Yellow: thermostability, blue: conserved H bond, purple: substrate binding residues, green: structurally important residues, red: heme binding regions)

Structural alignment was performed with the sequences of natural progesterone binding CYPs to find a common feature among them that can be important for progesterone binding activity. CYP119 amino acid sequence was also aligned with progesterone binding group to identify residues that are completely different so they may be candidates for mutations. Figure 3.4 shows sequence alignment of natural progesterone binding CYPs and CYP119 and color codes for important residues.

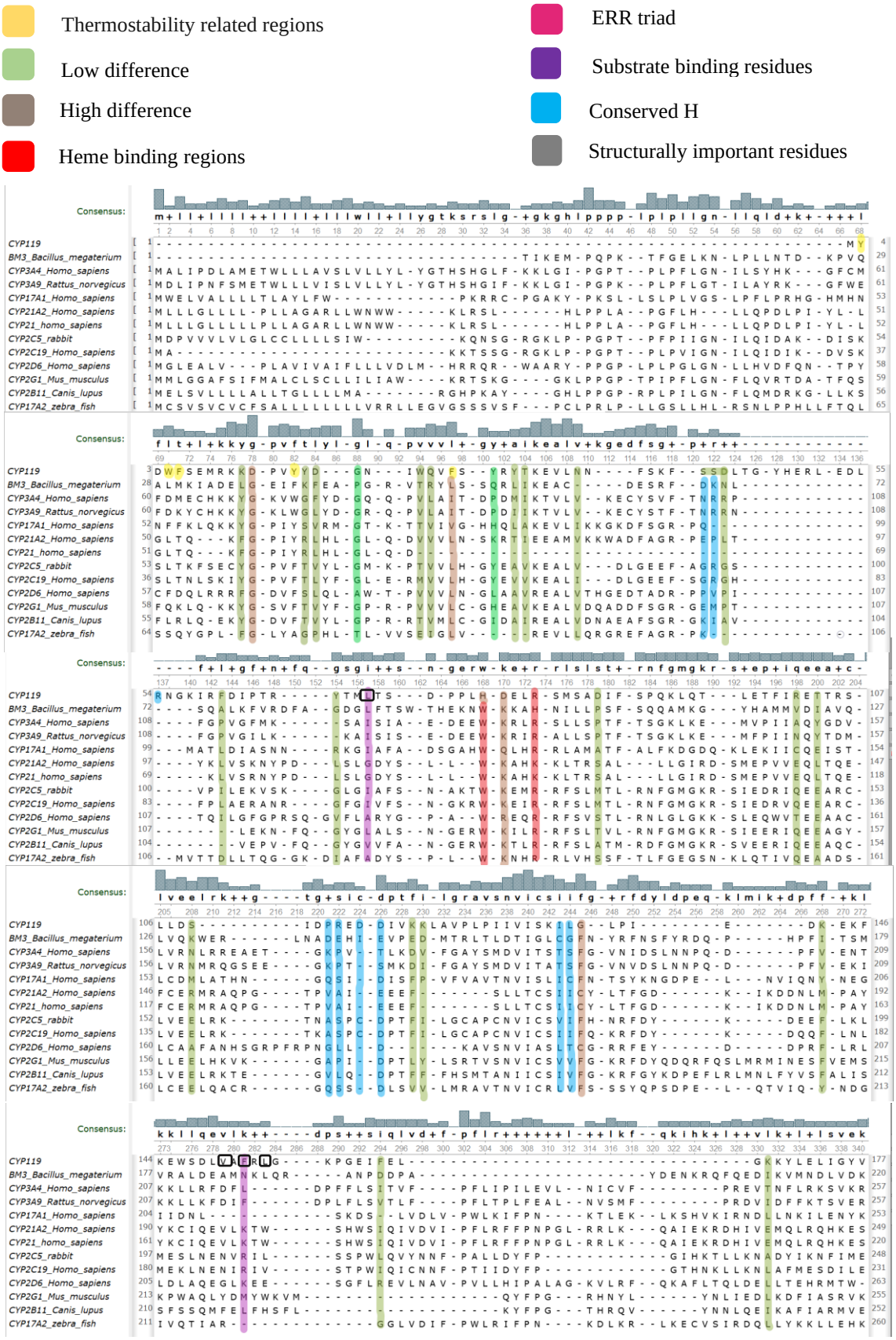

Figure 3.4 Structural alignment of natural progesterone binding P450s and CYP119. Black boxes indicate selected residues for mutational design. (cont. in next page).

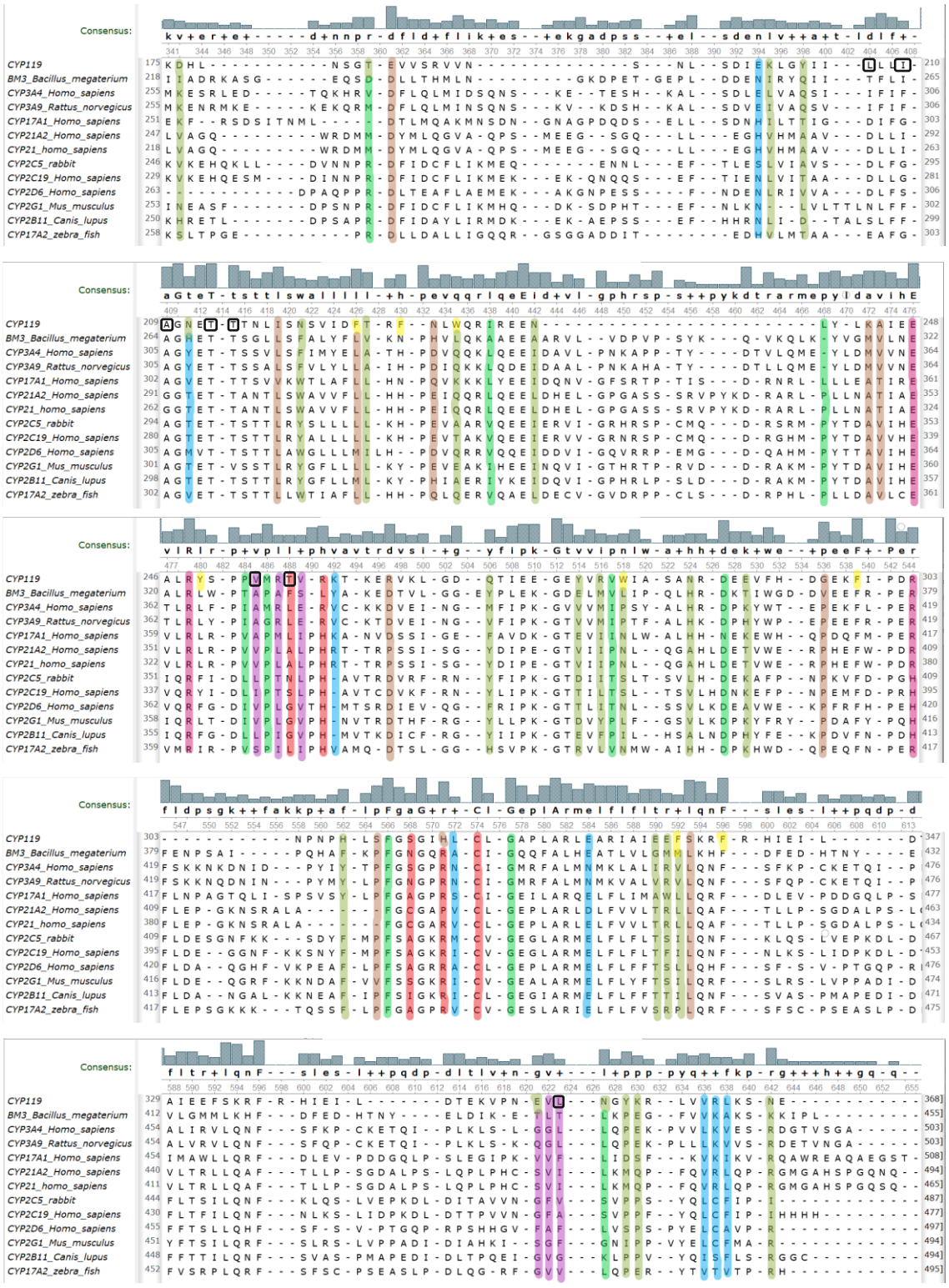

Figure 3.4 (cont.)

3.3 Mutations of CYP119

A total of 12 residues were selected for mutational scanning according to structural alignment results. Selected residues are Leu69, Val151, Phe153, Leu155, Leu205, Ile208, Ala209, Thr213, Thr214, Val254, Thr257, Leu354 shown in the Figure 3.5 with heme.

Leu69 is located on B2 helix and is an important substrate contact residue. Val151, Phe153 and Leu155 are located on flexible F-G loop of enzyme, which controls the substrate binding and helps binding to substrates with different sizes by rearranging the distance of the loop from active site. Structural alignment of all progesterone hydroxylating CYPs show that Leu205 residue of CYP119 is often replaced with phenylalanine and Ile208 is often replaced with phenylalanine or glycine. Ala209 affect the flexibility of I-helix with Gly210 and mutation of this residue can help to bind substrates of different sizes. Thr213 and Thr214 form conserved acid-alcohol pair with Glu212 which is important for proton delivery. There are already several mutational studies that investigated these residues to understand their effect on the activity, folding and thermostability. Val254 corresponds to a highly conserved hydrophobic residue which is located 5 residues after the EXXR motif and is also an important substrate contact residue. Thr257 is close to highly conserved β3-arginine (Arg259) and also forms a conserved hydrogen bond with heme group. Leu354 is part of the substrate recognition sequence “EVL” and located at the beginning of substrate access channel.

All residues were mutated to 7 different amino acids which are Alanine (small), Arginine (positive), Glutamic acid (negative), Glycine (small), Histidine (positive), Methionine (Sulphur-containing), Phenylalanine (aromatic) for the first round of selection. Figure 3.5 shows selected amino acids for mutant design. Energy score of all mutants calculated with PyRosetta program according to score function REF2015.Total energy score of selected mutants compared with the energy score of wild type protein to eliminate mutations that will affect stability of the enzyme.

Mutations were applied with TaskDesign protocol of PyRosetta program that read instructions for mutations from resfile files. Energy minimization of rotamers performed with PackRotamerMover protocol according to score function REF2015. FastRelax

protocol was applied to all mutants before calculating final energy of the enzyme and performing ligand docking.

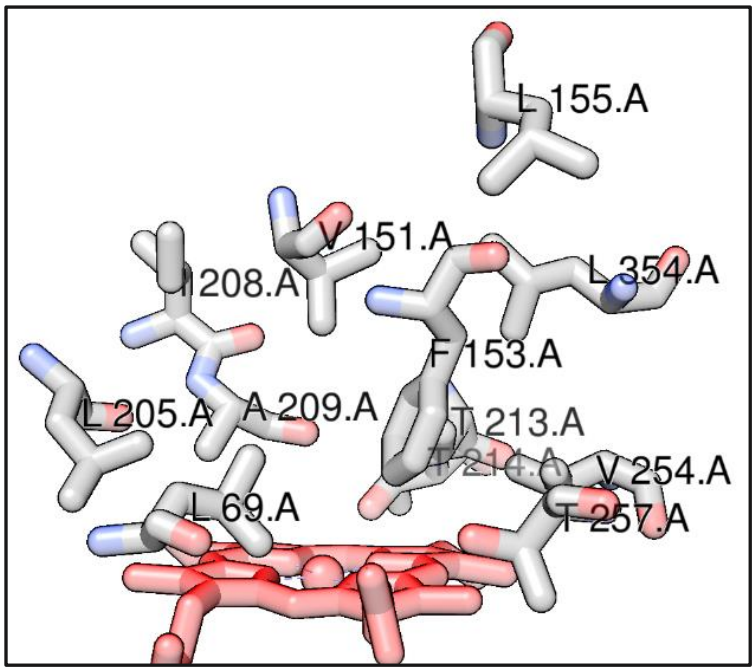

Figure 3.5 Selected residues for designing progesterone hydroxylating CYP119 (PDB ID: 1F4T). Heme group colored red, amino acids colored by element. Image created with Chimera.

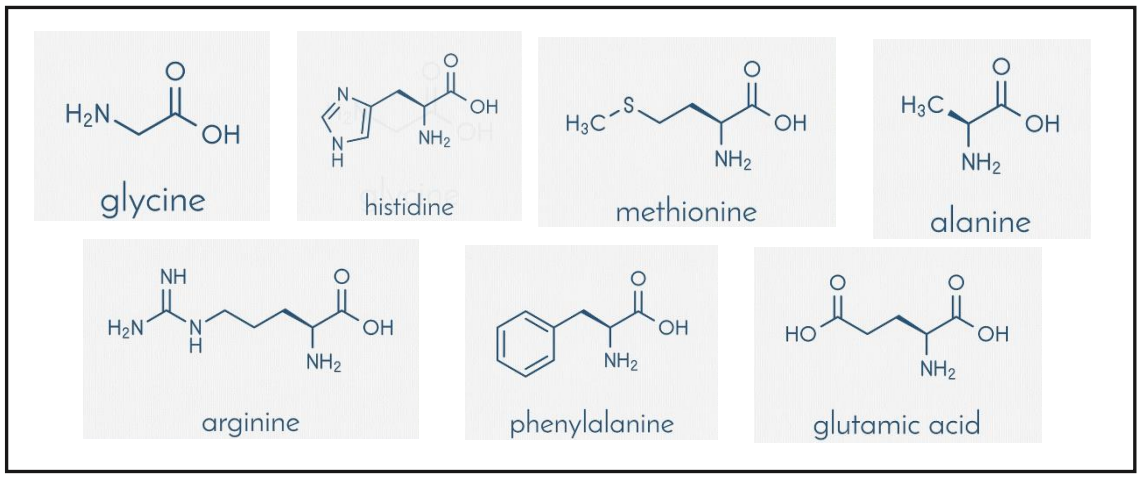

Figure 3.6 Seven amino acids that used for creating mutations.

Single Mutant (SM) group contains mutants of 12 selected residues. Ligand docking was performed with all mutants with two different coordinates of progesterone (PRG-1 and PRG-2) and mutants were eliminated according to binding scores. Combinations of selected single mutants were used for creating double mutants.

Double/triple mutants were designed by using two different approaches. For the first approach; combination of selected mutants simultaneously created on wild type enzyme and then double/triple mutant energy minimized. Double Mutant (DM), Double + Single Mutant (D+SM) and Triple Mutant (TM) groups created with the first method. In the second approach; single/double mutants that can bind progesterone were used as a template for second/third mutation. Mutations were performed iteratively. Double Single Mutant (2SM) and Triple Single Mutant (3SM) groups created with the second method. Figure 3.6 shows a schematic presentation of mutant design process.

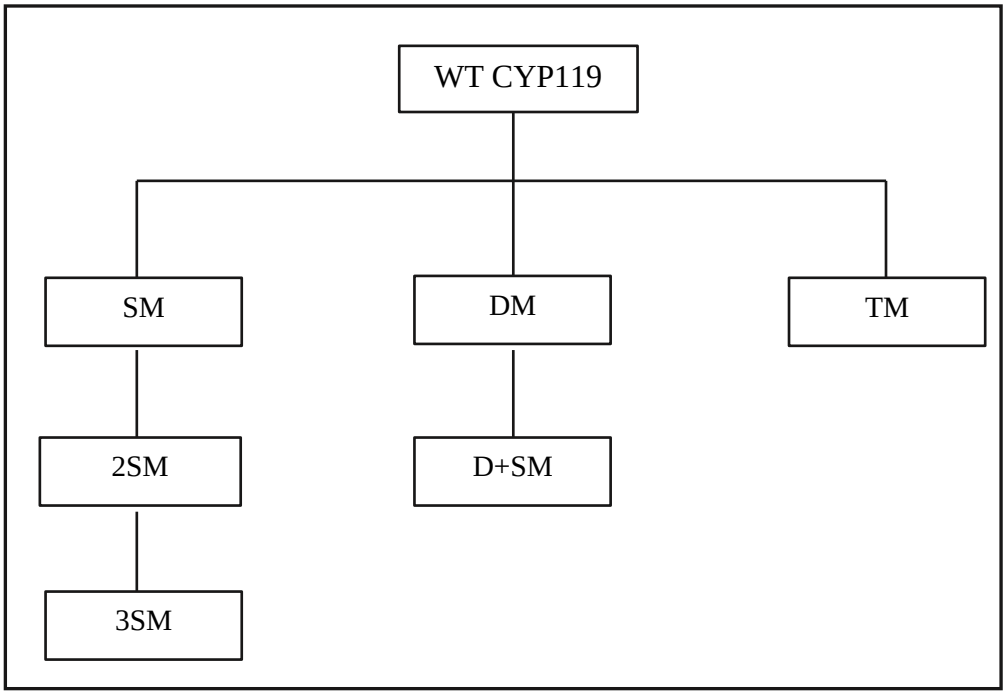

Figure 3.7 Schematic presentation of mutant design process. SM: Single Mutant, DM: Double Mutant, TM: Triple Mutant, 2SM: 2-Single Mutant with wild type CYP119, 3SM: 3-Single Mutant with wild type CYP119, D+SM: Single mutant with double mutant.

Ligand docking was performed with designed mutants after relax protocol was applied. Mutants were eliminated according to their progesterone binding scores

calculated with ligand score function. Table 3.1 shows the numbers of designed mutants and docking rounds performed for each group.

Table 3.1 Total numbers of designed mutants and performed docking rounds

	Number of Mutants	Number of Docking Rounds
SM	96	1920
2SM	350	7000
DM	118	2360
D+SM	60	1200
3SM	29	580
TM	20	400

Table 3.2 Total energy scores of final selected mutants and WT CYP119.

Name of Protein	Energy Score (REU)	Name of Protein	Energy Score (REU)
WT CYP119	-690.670	L69G-L205E-L354R	-692.285
L69G-T257G	-682.594	T214M-I208R-F153A	-706.183
L69G-F153A	-687.187	L69G-F153F-L354R	-692.481
L69G-F153F	-692.639	L69G-F153F-L205E	-689.294
L69E-T257E	-712.162	F153A-T214M-L354R	-716.439
L69E-T214M	-709.546	L205G-L69G-T257F	-686.111
L69G-L205G	-690.530	F153A-T214M-L69G	-708.454
L69G-T257E	-680.426	L205G-L69G-V254E	-687.605
L69G-L354R	-688.209	L354R-L69G-T257F	-695.466
F153A-T214M	-716.267	L354R-L69G-L205G	-688.661
V151G-L155R	-684.376	L354R-T213F-L69G	-702.840
		L69G-V151A-L205G	-687.517
		L69G-V254E-L354R	-695.648

Double/triple mutants were designed according to docking results of first/second round and total energy score was calculated for each mutant. Total number of 694 mutants were designed for this project and 24 of these mutants can bind to progesterone according to docking results. Table 3.2 shows total energy scores of 24 selected mutant and wild type CYP119. Results indicated none of these mutations show a negative effect on the stability of the enzyme.

3.4 Ligand Docking of Progesterone to CYP119 Mutants

PyRosetta program was used for selecting mutants according to their binding energy scores. Energy minimization process were performed with FastRelax protocol to every mutant and DockMCM protocol was used for high resolution ligand docking with progesterone. CYPs that can naturally catalyze hydroxylation of progesterone used for determining a standard REU value for successful docking process. Table 3.3 shows REU scores of natural progesterone hydroxylating CYPs' after 100 rounds of docking with PyRosetta.

Table 3.3 Properties of naturally progesterone hydroxylating P450s and their progesterone binding scores calculated with DockMCM protocol of PyRosetta.

Name	Organism	PDB ID	Resolution	REU
CYP17A2	*Danio rerio*	4R21	2.7 Å	-1198
CYP3A4	*Homo sapiens*	5A1R	2.45 Å	-1380
CYP17A1 (A105L)	*Homo sapiens*	4NKX	2.794 Å	-1421
CYP21A2	*Homo sapiens*	4Y8W	2.64 Å	-1287
CYP3A4	*Homo sapiens*	5A1P	2.5 Å	-1362
CYP260A1 (S276N)	*Sorangium cellulosum*	6F88	1.75 Å	-1242
CYP260A1 (S276I)	*Sorangium cellulosum*	6F8C	1.9 Å	-1254
CYP3A4	*Homo sapiens*	1W0F	2.65 Å	-1302
CYP154C5	*Nocardia farcinica*	4J6C	1.9 Å	-1233

PyRosetta requires starting coordinates of ligand for ligand docking process. Natural progesterone hydroxylating CYPs were grouped according to their ligand locations in the active site. Coordinates of progesterone were taken from group 2 (PRG-2) and group 3 (PRG-1) used as starting coordinates for ligand docking.

Group 1 includes models 1W0F, 5A1P and 5A1R which belong to the same human enzyme, CYP3A4. Figure 3.7 shows progesterone bound models and distances between heme group and progesterone for each model. Group 1 coordinates were not used for the design because all mutated amino acids are located around heme and using Group 1 coordinates would be meaningless since progesterone is located very far from the heme group in these structures.

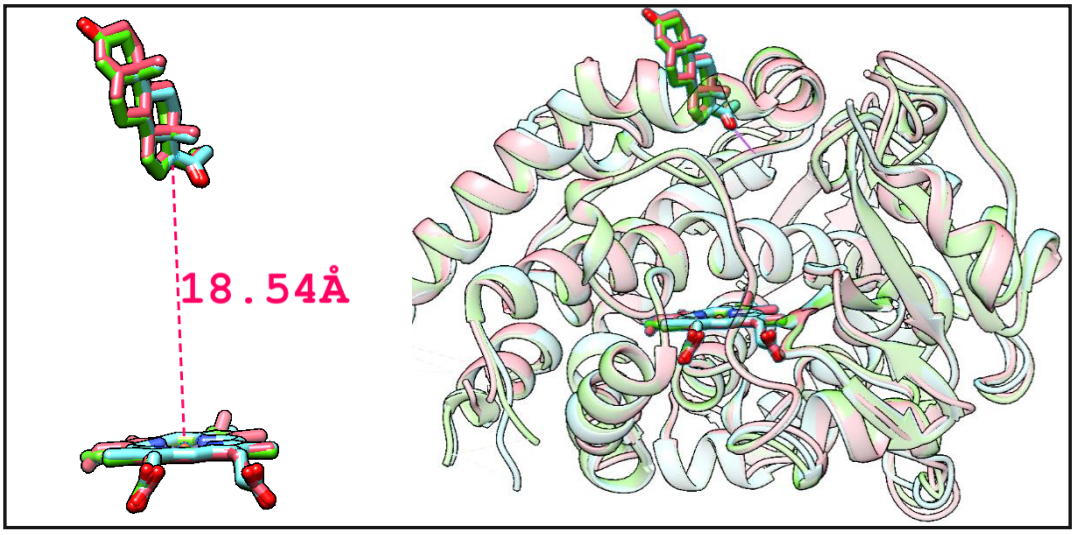

Figure 3.8 Presentation of heme group and bound progesterone of Group-1 P450s and aligned structures of 1W0F (pink), 5A1P (blue) and 5A1R (green). Distance between progesterone and heme group shown in yellow.

Models 6F88 and 6F8C from bacteria *Sorangium cellulosum* were classified as Group 2. Figure 3.8 shows progesterone bound models and distance between heme group and progesterone for each model. Original coordinates of model 6F88 were used as PRG-2 and coordinates of both models identified to program as rotamer library of progesterone.

Group 3 included models 4Y8W (*Homo sapiens*), 4J6C (*Nocardia farcinicia*), 4NKX (*Homo sapiens*), 4R21 (*Danio rerio*) that have different origins but same binding type for progesterone. Figure 3.9 shows progesterone bound models and distance between

heme group and progesterone for each model. Original coordinates of 4Y8W were used as PRG-1 and coordinates of all models used for creating rotamer library of progesterone.

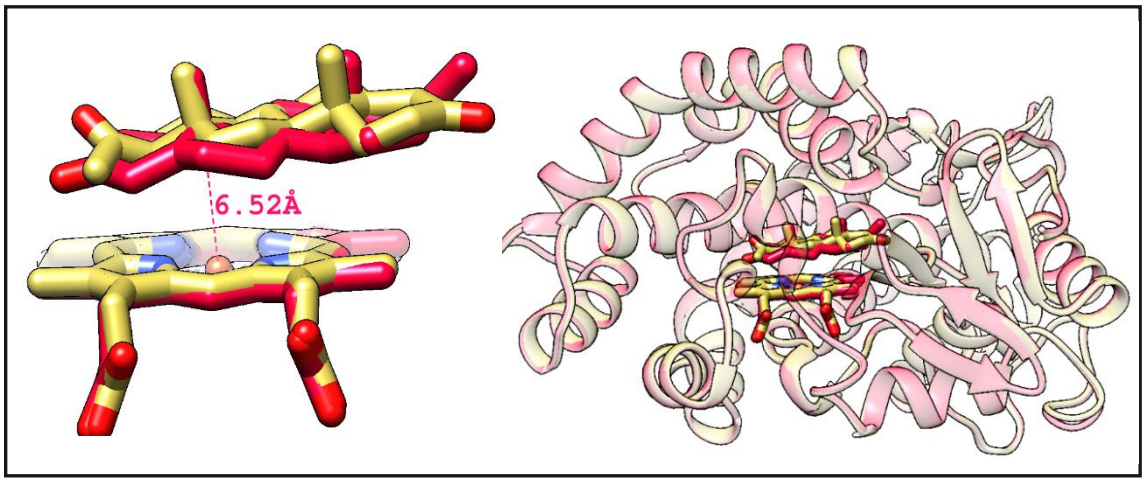

Figure 3.9 Presentation of heme group and bound progesterone of Group-2 P450s and aligned structures of 6F88 (red) and 6F8C (gold). Distance between progesterone and heme group shown in yellow.

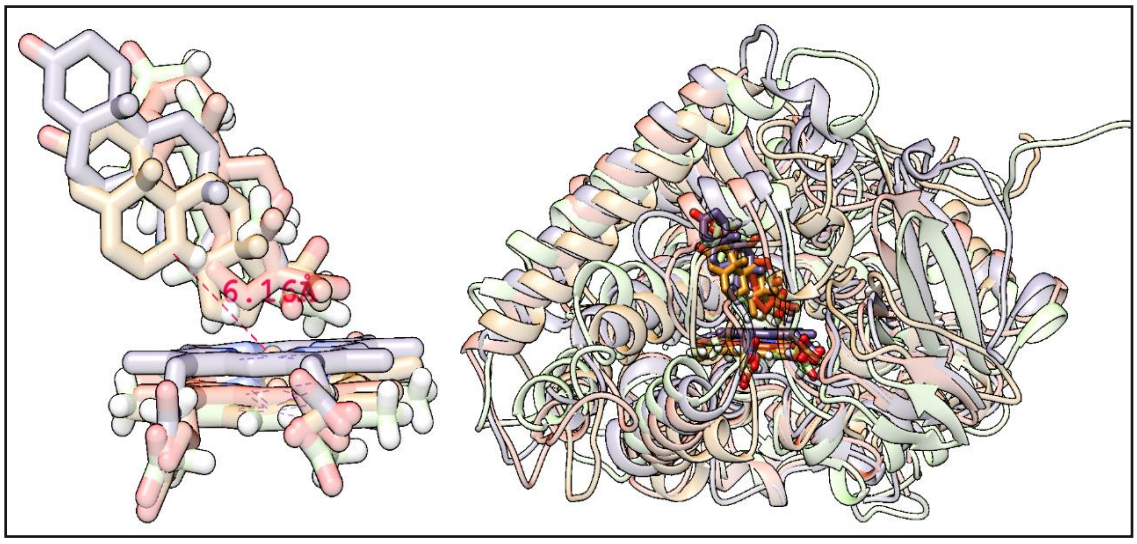

Figure 3.10 Presentation of heme group and bound progesterone of Group-3 P450s and aligned structures of 4Y8W (red), 4J6C (green), 4NKX (purple), 4R21 (orange). Distance between progesterone and heme group shown in yellow.

PRG-1 and PRG-2 coordinates were used as starting coordinates and docking was performed with two different ligand coordinate groups. Best scoring mutants of each group was selected for next round of mutations and docking was performed again with two different starting coordinates. Figure 3.10 shows location of PRG-1 and PRG-2 in proximal pocket of WT CYP119.

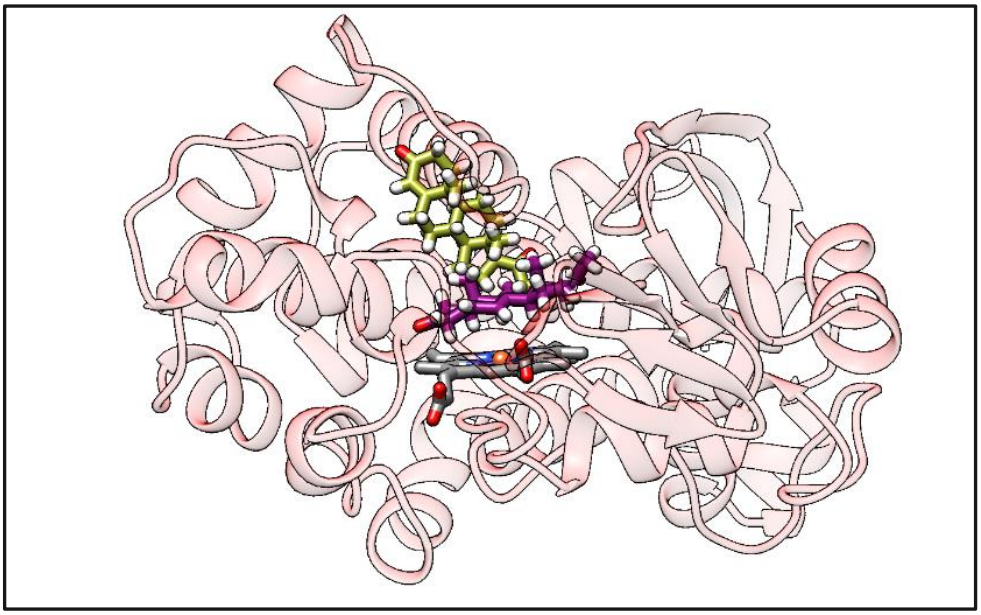

Figure 3.11 Presentation of PRG-1 (yellow) and PRG-2 (purple) in the active site of wild type CYP119.

3.4.1 Single Mutant (SM) Group of CYP119

96 mutants were created for SM group and 10 high-resolution docking rounds performed with PRG-1 and PRG-2 coordinates, separately. Positively scored progesterone-enzyme structures, which indicated progesterone binding to these structures is not favorable, were automatically deleted by program. Remaining models were analyzed with visualization program Chimera. Models that did not bind progesterone at the active site, were selected and deleted manually. Table 3.4 and Table 3.5 shows remaining mutants after first round of selection and their docking energy scores for PRG-1 and PRG-2.

3.4.2 Double Mutant (DM) Group of CYP119

DM group was designed with using the first approach, so all mutations applied at the same time to the wild type enzyme. Two different DM group created based on the origin of single mutant. 41 double mutants designed (DM-PRG1) with using PRG-1 binding mutant data. All two pair combinations were designed and mutated at the same time on wild type enzyme. Double mutant energies were minimized with FastRelax protocol and ligand docking was performed with PRG-1 and PRG-2 as ligands again. Table 3.6 and Table 3.7 shows the docking results for DM-PRG1 mutants.

77 double mutants were designed (DM-PRG2) according to PRG-2 binding results of SM group. In Table 3.5, yellow and green colored residues were used for creating mutations in DM-PRG2 group. At least one mutant was selected for each amino acid to create a more diverse in double mutant group, even though some mutations had low binding scores. Yellow color indicates that these mutations were only selected for DM-PRG2 group and green colored means these mutations were also used for creation of 2SM group. Table 3.8 and Table 3.9 show high resolution docking results of DM-PRG2 group for both PRG-1 and PRG-2.

Table 3.4 Selected progesterone binding SM Group mutants according to PRG-1 binding scores calculated with PyRosetta.

PRG-1	
Mutant	REU
L69G	-792
V151G	-567
V151R	-844
V151F	-654
L155R	-484
L155E	-551
L205G	-320
T214E	-302
V254M	-297
L354M	-138

3.4.3 Single-Single Mutant (2SM) Group of CYP119

350 double mutants were designed according to PRG-2 binding results of SM group using the second approach. All two pairs of mutations were designed using a single mutant as template and two double mutants were designed for each pair. Names of mutants coded as template and mutation order. For example; L69G-F153A double mutant was designed with using L69G as template, while F153A-L69G double mutant was created with F153A as template. Mutants with REU score under -800 were selected for creating a more stable library. Blue (only for 2SM group) and green (both 2SM and DM groups) colored boxes in Table 3.5 show the single mutants that were used for creating 2SM group.. FastRelax protocol of PyRosetta was used for energy minimization of all mutants before ligand docking was performed with PRG-1 and PRG-2. Table 3.10 and Table 3.11 show the ligand binding scores of 2SM group.

Table 3.5 Selected progesterone binding SM Group mutants according to PRG-2 binding scores. Yellow (DM-PRG 2 Group), blue (2SM Group) and green (both groups).

PRG-2					
Mutant	REU	Mutant	REU	Mutant	REU
L69M	-838	L205F	-297	T213R	-472
L69E	-846	L205G	-820	T214M	-798
L69R	-763	I208A	-324	T214F	-894
L69F	-486	I208E	-700	T214G	-702
L69G	-877	I208G	-551	V254R	-550
V151A	-590	I208R	-825	V254M	-496
V151E	-487	I208M	-220	V254F	-423
V151M	-551	A209A	-460	V254E	-793
V151G	-282	A209G	-673	T257R	-707
F153A	-869	A209E	-523	T257F	-886
F153M	-886	A209M	-804	T257G	-805
F153R	-776	A209R	-414	T257M	-877
F153F	-886	A209F	-185	T257E	-809
L155A	-634	T213E	-607	L354R	-895
L155R	-463	T213A	-643	L354F	-481
L205E	-793	T213M	-788	L354E	-577
L205M	-666	T213F	-871	L354M	-882

Table 3.6 PRG-1 ligand docking results of DM-PRG 1 Group. X indicates such mutation is not possible. Structures that cannot bind progesterone shown with “ 0” as a score. All REU values colored in scale of green-yellow-red (highest-medium-lowest).

	V254M	V151R	V151G	V151F	T214E	L354M	L205G	L155R	L155E	L69G
L69G	-1104	-1167	-869	0	0	0	0	0	-94	X
L155E	0	-1107	0	0	0	-1090	-1108	0	X	-94
L155R	0	-1093	-1116	-690	-596	0	0	X	0	0
L205G	0	0	-917	-471	-586	0	X	0	-1108	0
L354M	0	0	0	0	0	X	0	0	-1090	0
T214E	0	0	0	0	X	0	-586	-596	0	0
V151F	-887	0	0	X	0	0	-471	-690	0	0
V151G	0	0	X	0	0	0	-917	-1116	0	-869
V151R	-1133	X	0	0	0	0	0	-1093	-1107	-1167
V254M	X	-1133	0	-887	0	0	0	0	0	-1104

Table 3.7 PRG-2 ligand docking results of DM-PRG 1 Group. X indicates such mutation is not possible. Structures that cannot bind progesterone shown with “ 0” as a score. All REU values colored in scale of green-yellow-red (highest-medium-lowest).

	V254M	V151R	V151G	V151F	T214E	L354M	L205G	L155R	L155E	L69G
L69G	-667	-891	-1069	-800	-1151	-1111	-1119	-1083	-702	X
L155E	-24	0	-786	-961	-977	-573	-1085	0	X	-702
L155R	-1139	-477	-1180	-752	-381	-914	-1135	X	0	-1083
L205G	-1007	-1109	-988	-950	-901	-670	X	-1135	-1085	-1119
L354M	-24	-1109	-1025	0	-959	X	-670	-914	-573	-1111
T214E	-271	-544	-522	-1127	X	-959	-901	-381	-977	-1151
V151F	-467	0	0	X	-1127	0	-950	-752	-961	-800
V151G	-956	0	X	0	-522	-1025	-988	-1180	-786	-1069
V151R	-265	X	0	0	-544	-1109	-1109	-477	0	-891
V254M	X	-265	-956	-467	-271	-24	-1007	-1139	-24	-667

Table 3.8 PRG-1 ligand docking results of DM-PRG 2 Group. X indicates such mutation is not possible. Structures that cannot bind progesterone shown with “ 0” as a score. All REU values colored in scale of green-yellow-red (highest-medium-lowest).

	T257F	T213F	I208R	V254E	V151A	T214F	L69G	L354R	L205G	L155A	F153M	F153F	A209M
A209M	0	0	0	0	0	0	0	-1144	0	-1128	0	0	X
F153F	-1175	0	0	-1143	0	0	-815	0	-587	-1052	0	X	0
F153M	0	0	0	0	-876	0	0	0	0	-1096	X	0	0
L155A	0	0	0	-1141	-1150	-1065	-1100	-1047	0	X	-1096	-1052	-1128
L205G	-806	0	0	0	0	0	0	-374	X	0	0	-587	0
L354R	0	0	0	0	0	-673	0	X	-374	-1047	0	0	-1144
L69G	0	0	0	0	0	0	X	0	0	-1100	0	-815	0
T214F	-883	0	0	-1006	0	X	0	-673	0	-1065	0	0	0
V151A	0	0	0	0	X	0	0	0	0	-1150	-876	0	0
V254E	0	0	0	X	0	-1006	0	0	0	-1141	0	-1143	0
I208R	0	0	X	0	0	0	0	0	0	0	0	0	0
T213F	0	X	0	0	0	0	0	0	0	0	0	0	0
T257F	X	0	0	0	0	-883	0	0	-806	0	0	-1175	0

Table 3.9 PRG-2 ligand docking results of DM-PRG 2 Group. X indicates such mutation is not possible. Structures that cannot bind progesterone shown with “ 0” as a score. All REU values colored in scale of green-yellow-red (highest-medium-lowest).

	V254E	V151A	T257F	T214F	T213F	L69G	L354R	L205G	L155A	I208R	F153M	F153F	A209M
A209M	-800	-799	-1135	-793	-475	-510	-366	-692	-1129	-1059	-559	-747	X
F153F	-1159	-446	-1178	-1128	-615	-1040	-1050	-840	-867	-887	0	X	-747
F153M	-1099	-975	-770	-1179	-173	-852	-639	-890	-599	-941	X	0	-559
I208R	-962	-660	-1046	-356	-401	-887	-768	-540	-460	X	-941	-887	-1059
L155A	-1065	-1072	-613	-929	-599	-999	-1046	-1145	X	-460	-599	-867	-1129
L205G	-970	-1035	-1138	-591	-851	-1182	-791	X	-1145	-540	-890	-840	-692
L354R	-1042	-953	-1117	-550	-1100	-1046	X	-791	-1046	-768	-639	-1050	-366
L69G	-1164	-1109	-1161	-876	-842	X	-1046	-1182	-999	-887	-852	-1040	-510
T213F	0	-721	-743	-20	X	-842	-1100	-851	-599	-401	-173	-615	-475
T214F	-851	-993	-856	X	-20	-876	-550	-591	-929	-356	-1179	-1128	-793
T257F	-1075	-982	X	-856	-743	-1161	-1117	-1138	-613	-1046	-770	-1178	-1135
V151A	-1053	X	-982	-993	-721	-1109	-953	-1035	-1072	-660	-975	-446	-799
V254E	X	-1053	-1075	-851	0	-1164	-1042	-970	-1065	-962	-1099	-1159	-800

Table 3.10 PRG-1 ligand docking results of 2SM Group. X indicates such mutation is not possible. Structures that cannot bind progesterone shown with “ 0” as a score. All REU values colored in scale of green-yellow-red (highest-medium-lowest).

	L354M	L354R	T257E	T257M	T257G	T257F	V254E	T214F	T214M	T213F	A209M	I208R	L205G	L205E	F153F	F153M	F153A	L69M	L69E	L69G
L69M	-835	0	0	0	0	0	0	0	0	0	0	0	0	-821	0	-657	-559	X	X	X
L69E	0	0	0	0	-1054	-17	-929	0	0	0	0	0	0	0	0	0	0	X	X	X
L69G	-1097	0	-1158	0	0	0	0	0	0	0	0	0	-1072	0	0	0	0	X	X	X
F153A	0	0	0	0	-1067	0	0	0	-246	0	0	0	0	0	X	X	X	0	-1143	-1185
F153M	-966	-287	0	0	0	0	-871	0	0	0	0	0	0	0	X	X	X	0	-1159	0
F153F	-12	-603	0	0	0	0	0	0	0	0	0	0	0	0	X	X	X	0	-1116	-1181
L205E	0	-217	0	0	0	0	-973	0	-849	0	0	0	X	X	0	-1024	-1082	0	-1079	0
L205G	-1060	-952	0	0	0	0	0	0	-928	0	0	0	X	X	0	0	0	0	-1157	0
I208R	0	0	0	0	0	0	0	0	-1172	0	0	X	0	0	0	-208	0	0	-1139	0
A209M	-389	0	0	0	0	0	0	-818	0	0	X	0	-750	0	0	0	0	0	-874	-54
T213F	-996	0	0	0	0	0	0	-1064	0	X	0	0	0	0	0	0	0	0	-58	0
T214M	-312	0	0	0	-899	0	-686	X	X	0	0	0	0	0	0	0	0	0	-1183	0
T214F	0	0	0	0	0	0	0	X	X	0	0	0	0	0	0	0	0	0	-1165	0
V254E	0	0	0	0	0	0	X	0	-1144	0	0	0	0	0	0	0	0	0	-934	-949
T257F	0	0	X	X	X	X	0	0	0	0	0	0	-1012	-866	0	0	0	0	-1172	-801
T257G	-23	0	X	X	X	X	0	0	0	0	0	0	-86	-1132	0	0	0	0	-1171	-1191
T257M	0	-880	X	X	X	X	0	0	0	0	0	0	-46	0	0	0	0	0	-1094	-1164
T257E	0	0	X	X	X	X	-969	0	0	0	0	0	0	0	0	0	0	0	-1183	0
L354R	X	X	0	0	0	0	-423	0	0	0	0	0	-131	0	0	0	0	0	-1125	0
L354M	X	X	0	0	0	0	0	0	0	0	0	0	0	-1004	0	0	0	0	-1132	-1031

Table 3.11 PRG-2 ligand docking results of 2SM Group. X indicates such mutation is not possible. Structures that cannot bind progesterone shown with “ 0” as a score. All REU values colored in scale of green-yellow-red (highest-medium-lowest).

	L354M	L354R	T257E	T257M	T257G	T257F	V254E	T214F	T214M	T213F	A209M	I208R	L205G	L205E	F153F	F153M	F153A	L69M	L69E	L69G
L69M	-803	-1152	-1111	-1114	-1121	-690	-974	-908	-1154	-968	-911	0	0	-1029	-776	-762	0	X	X	X
L69E	-680	-1159	-1135	-1099	-1109	-980	-1064	-705	-1127	-561	-860	-540	-629	-696	-504	-1143	-1155	X	X	X
L69G	-1071	-1163	-840	-560	-1074	-481	-1075	-805	-619	-1022	-1081	-502	-1016	-1172	-1168	-538	-570	X	X	X
F153A	0	-933	-947	-1055	-1015	-249	-1030	-956	-1123	-872	-656	-869	-1084	-880	X	X	X	-964	-593	-1070
F153M	-177	-868	-782	-961	-869	-517	-948	-956	-788	-352	-857	-865	-571	-1020	X	X	X	-777	-1076	-1176
F153F	-248	-1068	-1146	-667	-1087	-515	-1015	-717	-302	-882	-858	-845	-655	-272	X	X	X	-793	-715	-1180
L205E	-970	-994	-1119	-321	-1038	-566	-920	-880	-995	-875	-911	-786	X	X	-686	0	-993	-800	-919	-1179
L205G	-1044	-815	-763	-667	-930	-903	-1044	-993	0	-833	-750	-866	X	X	-752	-1054	-987	-1111	-889	-1186
I208R	-695	-1025	-1118	-654	-1051	-135	-765	-579	-1155	-796	-255	X	0	-829	-758	-657	0	-756	-378	-1175
A209M	-123	-802	-972	0	-997	-383	-1069	-789	-181	-706	X	-637	-867	-907	-671	-928	-304	-778	-979	-1149
T213F	-873	-1100	-1060	-366	-954	-446	-911	-549	-955	X	-619	-667	-125	-882	-254	-651	0	-932	-579	-1060
T214M	-788	-1015	-1096	-366	-878	-810	-1040	X	X	-956	-937	-908	-981	-949	-680	-1131	-1208	-754	-1043	-495
T214F	-613	-995	-860	-712	-676	-142	0	X	X	-567	-321	-918	-562	-1071	-895	-1054	-862	-719	-977	-1038
V254E	-996	-1027	-1023	-652	-1071	-1035	X	-865	-1111	-789	-904	-1059	-851	-1014	-771	-1073	-920	-1091	-1013	-1089
T257F	-396	-946	X	X	X	X	-1050	-250	-95	-915	-822	-860	-810	-1098	-660	-213	-422	-1056	-758	-882
T257G	-476	-1074	X	X	X	X	-838	-870	-1110	-602	-882	-819	-966	-822	0	-995	-808	-659	-710	-1095
T257M	-238	-1063	X	X	X	X	-926	-843	-851	-943	-666	-682	-394	-695	-307	-776	-510	-1051	-1114	-934
T257E	-373	-1075	X	X	X	X	-720	-857	-1137	-955	-879	-411	0	-911	-494	-967	-958	-1082	-573	-1181
L354R	X	X	-1092	-1044	-943	-987	-931	-796	-703	-878	-911	-266	-937	-895	-873	-506	-1103	-1109	-1097	-1203
L354M	X	X	-736	-1083	-791	-823	-1039	-1015	-1172	-471	-325	-913	-800	-912	-805	-502	-773	-1019	-1085	-1164

3.4.4 Triple Mutant (TM) Group of CYP119

TM group of triple mutants were designed according to the first approach. All mutations were performed at the same time with using wild type enzyme as template. 6 mutations (L69G, T257F, V151A, V254E, L205G, L354R) with low docking score from DM group selected for creating TM group. Every possible three-pair combinations of these 6 mutations were created with using PyRosetta program. Triple mutant energy was minimized with FastRelax protocol and ligand docking was performed with PRG-1 and PRG-2 separately. Table 3.12 and 3.13 show ligand docking scores of TM group with PRG-1 and PRG-2 coordinates, respectively.

Table 3.12 PRG-1 ligand docking results of TM Group. X indicates such mutation is not possible. Structures that cannot bind progesterone shown with "0" as a score. All REU values colored in scale of green-yellow-red (highest-medium-lowest).

	L69G	T257F	V151A	V254E	L205G	L354R	
L69G	X	X	0	0	-735	0	T257F
	X	0	X	0	0	0	V151A
	X	0	0	X	-1152	0	V254E
	X	-735	0	-1152	X	0	L205G
	X	0	0	0	0	X	L354R
T257F	X	X	X	0	0	0	V151A
	X	X	0	X	0	0	V254E
	X	X	0	0	X	-69	L205G
	X	X	0	0	-69	X	L354R
V151A	X	X	X	X	0	-977	V254E
	X	X	X	0	X	0	L205G
	X	X	X	-977	0	X	L354R
V254E	X	X	X	X	X	-183	L205G
	X	X	X	X	-183	X	L354R

Table 3.13 PRG-2 ligand docking results of TM Group. X indicates such mutation is not possible. Structures that cannot bind progesterone shown with "0" as a score. All REU values colored in scale of green-yellow-red (highest-medium-lowest)

	L69G	T257F	V151A	V254E	L205G	L354R	
L69G	X	X	-761	-933	-1154	-1058	T257F
	X	-761	X	-1043	-1193	-1098	V151A
	X	-933	-1043	X	-1126	-1206	V254E
	X	-1154	-1193	-1126	X	-1133	L205G
	X	-1058	-1098	-1206	-1133	X	L354R
T257F	X	X	X	-401	-1150	-727	V151A
	X	X	-401	X	-1162	-799	V254E
	X	X	-1150	-1162	X	-1035	L205G
	X	X	-727	-799	-1035	X	L354R
V151A	X	X	X	X	-1022	-958	V254E
	X	X	X	-1022	X	-955	L205G
	X	X	X	-958	-955	X	L354R
V254E	X	X	X	X	X	-911	L205G
	X	X	X	X	-911	X	L354R

3.4.5 Double + Single Mutant (D+SM) Group of CYP119

Total number of 60 triple mutants designed for D+SM group. Double mutants with lowest score selected from DM group combined with single mutations that seem to affect ligand binding according to DM group docking results. All triple mutant's energy minimized before high resolution docking performed. 10 rounds of docking performed for both PRG-1 and PRG-2. Table 3.14 and Table 3.15 show ligand binding scores of D+SM group mutants.

Table 3.14 PRG-1 ligand docking results of D+SM Group. X indicates such mutation is not possible. Structures that cannot bind progesterone shown with “0” as a score. All REU values colored in scale of green-yellow-red (highest-medium-lowest).

	L69G	T257F	V151A	V254E	L354R	L205G	T213F
L205G-L69G	X	-1181	0	0	0	X	0
V254E-L69G	X	0	0	X	0	0	0
L205G-T257F	-1029	X	-1024	-722	0	X	0
L205G-V151A	0	-986	X	0	0	X	0
V254E-V151A	-146	0	X	X	0	0	-22
V254E-L354R	0	-694	0	X	X	-1082	0
V254E-T257F	0	X	0	X	-1159	0	0
L354R-L69G	X	0	-1146	-1104	X	-1140	0
L354R-T213F	0	0	0	-914	X	-609	X
L354R-T257F	-1177	X	0	0	X	-933	0
L69G-T257F	X	X	0	-1178	0	0	0
L69G-V151A	X	0	X	0	0	0	0

Table 3.15 PRG-2 ligand docking results of D+SM Group. X indicates such mutation is not possible. Structures that cannot bind progesterone shown with “0” as a score. All REU values colored in scale of green-yellow-red (highest-medium-lowest).

	L69G	T257F	V151A	V254E	L354R	L205G	T213F
L205G-L69G	X	-1166	-1102	-1196	-1080	X	x
V254E-L69G	X	-187	-1082	X	-1083	-1109	x
L205G-T257F	-1143	X	-805	-917	-1068	X	0
L205G-V151A	-1078	-684	X	-1156	-909	X	-643
V254E-V151A	-870	-853	X	X	-1020	-366	-527
V254E-L354R	-1117	-898	-812	X	X	-1059	0
V254E-T257F	-1163	X	-788	X	-934	-1014	-1076
L354R-L69G	X	-1192	-1169	-1161	X	-1180	-1149
L354R-T213F	-1185	-1172	0	-1160	X	-1137	X
L354R-T257F	-1170	X	-1041	-1004	X	-1160	-1058
L69G-T257F	X	X	-1080	-1106	-1156	-1174	-1160
L69G-V151A	X	-1054	X	-1070	-1022	-1146	-781

3.4.6 Single-Single-Single Mutant (3SM) Group of CYP119

29 triple mutants were created with the second approach and grouped as 3SM. Every mutant that scored below -1100 REU listed from 2SM group. Table 3.16 and Table 3.17 show REU values of PRG-1 and PRG-2 docked 3SM group of mutants.

Table 3.16 PRG-1 ligand docking results of 3SM Group. X indicates such mutation is not possible. Structures that cannot bind progesterone shown with "0" as a score. All REU values colored in scale of green-yellow-red (highest-medium-lowest).

	L205E	L354R	T214M	F153A	F153F	I208R	L69G	L69M	L69E	T257E
L69G-F153F	-1182	-1186	-655	X	X	X	X	X	X	X
L69G-L205E	X	-1194	0	0	0	X	X	X	X	X
L69G-L354R	-274	X	0	0	0	X	X	X	X	X
L69M-L354R	0	X	0	0	0	X	X	X	X	X
F153A-T214M	0	-1181	X	X	X	X	0	0	X	X
T214M-I208R	X	X	X	-1191	-1155	X	X	X	0	-1149
L69E-F153A	X	X	0	X	X	-1141	X	X	X	-1153
L69E-F153F	X	X	-1131	X	X	-1153	X	X	X	-968

Table 3.17 PRG-2 ligand docking results of 3SM Group. X indicates such mutation is not possible. Structures that cannot bind progesterone shown with "0" as a score. All REU values colored in scale of green–yellow–red (highest–medium-lowest).

	L205E	L354R	T214M	F153A	F153F	I208R	L69G	L69M	L69E	T257E
L69G-F153F	-1099	-1097	-1087	X	X	X	X	X	X	X
L69G-L205E	X	-1114	-1011	-1069	-1064	X	X	X	X	X
L69G-L354R	-1106	X	-963	-1082	-1116	X	X	X	X	X
L69M-L354R	-1092	X	-989	-993	-962	X	X	X	X	X
F153A-T214M	-849	-802	X	X	X	X	-1184	-1082	X	X
T214M-I208R	X	X	X	-1169	-1123	X	X	X	-1144	-1135
L69E-F153A	X	X	-1012	X	X	-1029	X	X	X	-582
L69E-F153F	X	X	-421	X	X	-1021	X	X	X	-416

Total number of 674 mutants were designed and eliminated due to their ligand binding scores. Structures that have REU values below -1180 were selected from all designed groups and additional 1000 rounds of docking performed with selected 24 mutants to get most stable bound form of progesterone to enzyme (Table 3.18). REU score distribution graphics of 1000 rounds of docking for all models shown in Appendix B. Possible hydroxylation locations were determined according to most common ligand-protein complexes of 1000 rounds of docking. Docking scores calculated for each round and every structure that get a score higher than -1100 REU deleted by program automatically.

Table 3.18 Lowest REU scores of selected mutants and wild type CYP119 after 1000 rounds of docking with PRG-1 and/or PRG-2. Mutants in green (Thr257 mutants) and blue (Leu354 mutants) boxes eliminated after structural analysis.

PRG-1		PRG-2	
WT CYP119	+1456	WT CYP119	-567
L69G-T257G	-1192	L69G-L205G	-1198
L69G-F153A	-1189	L69G-T257E	-1181
L69G-F153F	-1200	L69G-F153F	-1194
L69E-T257E	-1191	L69G-L354R	-1203
L69E-T214M	-1197	F153A-T214M	-1211
L205G-L69G-T257F	-1196	V151G-L155R	-1180
L69G-L205E-L354R	-1201	F153A-T214M-L69G	-1184
T214M-I208R-F153A	-1210	L205G-L69G-V254E	-1198
L69G-F153F-L354R	-1189	L354R-L69G-T257F	-1192
L69G-F153F-L205E	-1183	L354R-L69G-L205G	-1181
F153A-T214M-L354R	-1196	L69G-V151A-L205G	-1194
		L354R-T213F-L69G	-1185
		L69G-V254E-L354R	-1206

3.5 Docking Results of CYP119 Mutants

A total of 24 mutant CYP119 enzymes were selected at the end of elimination process (Table 3.18). However, this elimination process was performed only depending on energy score levels. We further examined our models with using Chimera as visualization tool and eliminated 12 more mutants according to structural analysis results. Parameters like loss of conserved hydrogen bonds and changes in the substrate access channels were used as a cause of elimination. Remaining 12 models were examined structurally and possible hydroxylation sites of progesterone determined.

3.5.1 L69G – L69E Mutations

Leu69 residue was located B'- C loop and Leu69 is important as a substrate contact residue. L69G is the most common mutation for both models. L69G mutation was observed in 7 out of 11 of PRG-1 docking results and 11 out of 13 of PRG-2 docking results.. Figure 3.12 shows L69G mutation in the active site of the wild type enzyme. In the original structure side chain of Leu69 stands above heme group and reduce the volume of binding pocket. L69G mutation increase the volume thus ligands can bind the active site of mutant enzymes. This mutation seems necessary for both rotations of progesterone therefore this mutation is thought to improve progesterone binding.

R-R distance maps were calculated with Chimera program. These maps are important to identify residues that are most effected from a mutation. Chimera calculates Cα-Cα distances within a single chain. Chimera can also calculate the difference of overall distance values of two different chains (Chen et al., 2015). Mutant and wild type CYP119 structures were loaded to Chimera program to understand the effect of each mutation on the location of all residues. R-R distance maps have a color code to show relocation of each residue. While comparing two different chains; if the Cα-Cα distance of first structure is longer than the second one the difference between two values will be positive and shown with blue color. If the Cα-Cα distance of first structure is shorter than

the second one the difference between two values will be negative and shown with yellow color (Chen et al., 2015).

Residues that moved significantly further investigated on the structure to understand the cause of this relocations. Some of the significant relocations are result of relax protocol of PyRosetta and can be shown on each mutant. Some of the relocations are mutant specific while others are specific to certain type of progesterone location in the active site. Four residues of L69G mutant have relocated significantly (Figure 3.13). Replacement of residues 56 and 78 was observed in all structures and this occurred a result of relax procedure of PyRosetta program. Residue 339 is located on the surface of enzyme and does not effect substrate binding. Residue 354 (Leu354) is one of our target residues and known as substrate binding residue. Relocation of this residue (2.09 Å) has increased the active site volume and allowed progesterone binding.

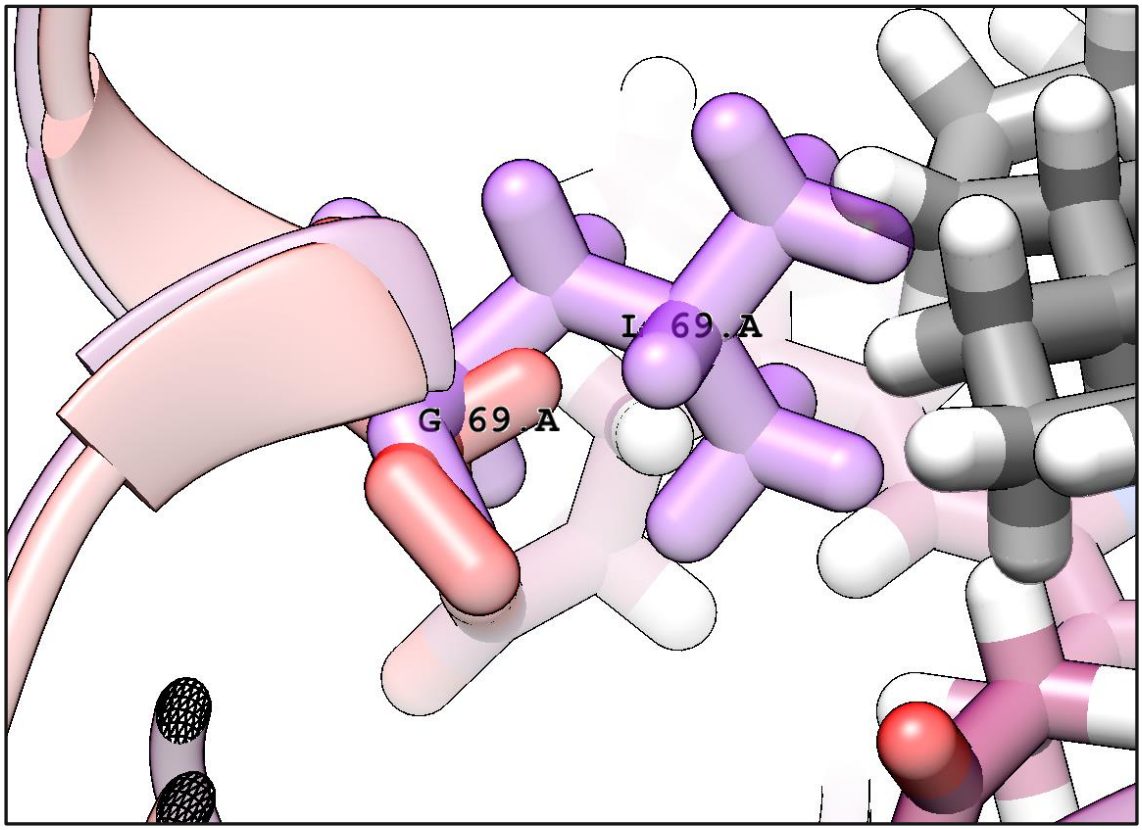

Figure 3.12 Presentation of wild type structure (purple), Gly69 mutation (red) and progesterone (grey).

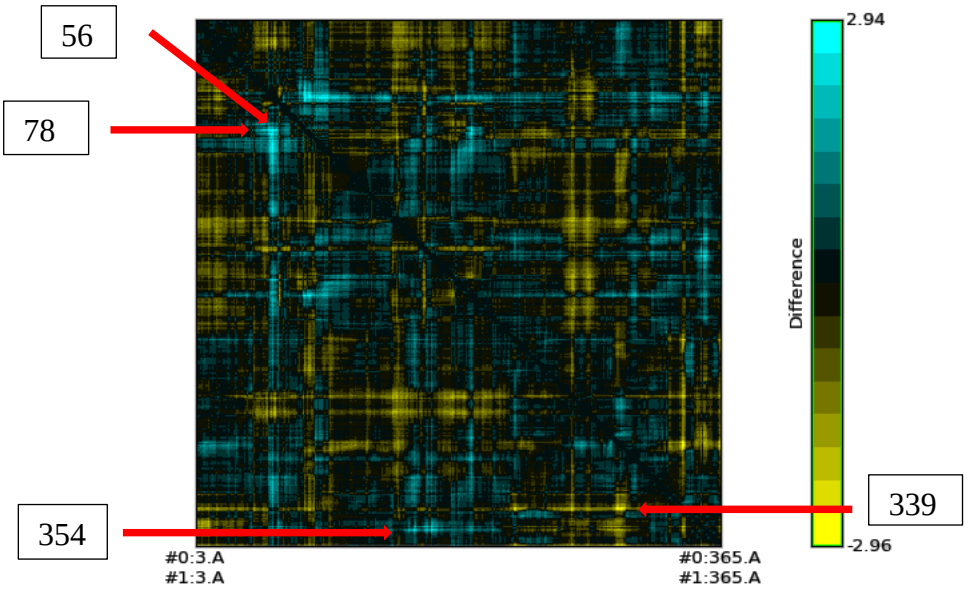

Figure 3.13 R-R distance map of L69G vs. wild type CYP119. Numbers of most flexible residues indicated in boxes.

L69E mutation (Figure 3.14) is only acceptable for PRG-1 coordinates of progesterone. Glutamate side chain positioned above heme and does not allow a parallel binding of progesterone thus none of PRG-2 mutants include L69E mutation.

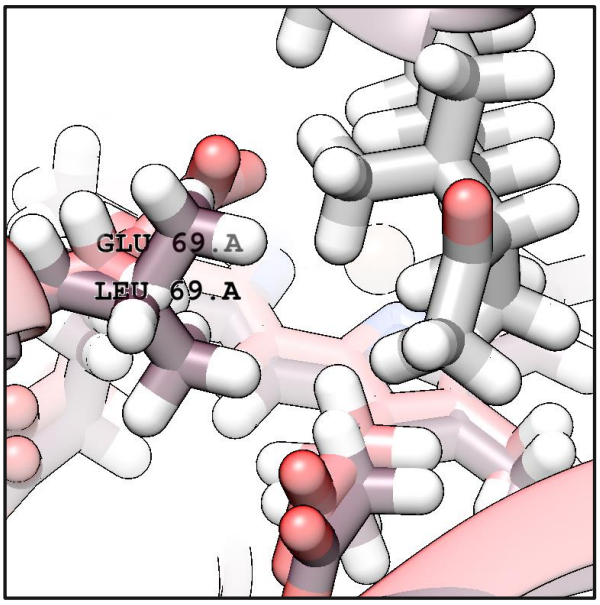

Figure 3.14 Presentation of wild type structure (brown), Glu69 mutation (red) and progesterone (grey).

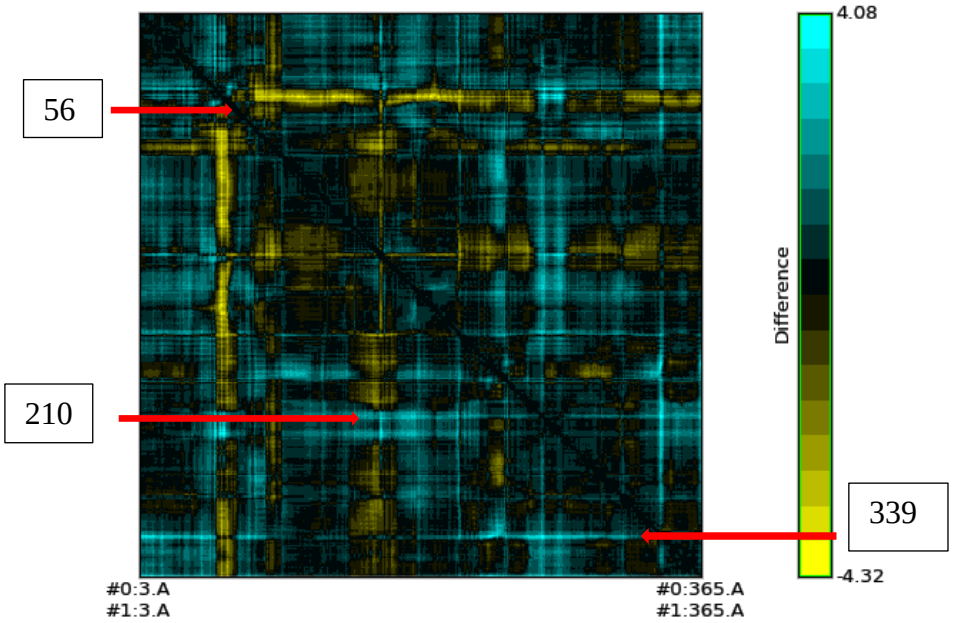

Figure 3.15 R-R distance map of L69E vs. wild type CYP119. Numbers of most flexible residues indicated in boxes.

R-R distance map of L69E mutation (Figure 3.15) did not show any significant replacement. Only Gly210 residue moves slightly, this residue is important because it gives flexibility to I-helix with Ala209 and their movement is related to substrate binding. We cannot see replacement of Leu354 in L69E mutant, because Glu and Leu residues have approximately the same size and this replacement did not give structural flexibility given by Gly69. Lack of Leu354 replacement also explains why L69E mutants are only applicable for PRG-1 substrates. PRG-2 binding needs extra spaces created by Leu354 replacement.

3.5.2 F153F – F153A Mutations

Phe153 is a well conserved residue of F-G loop and it is important for substrate binding. Phe153 can also affect the shape of the β5 sheet which includes substrate contact residues Glu352, Val353 and Leu354 (EVL sequence).

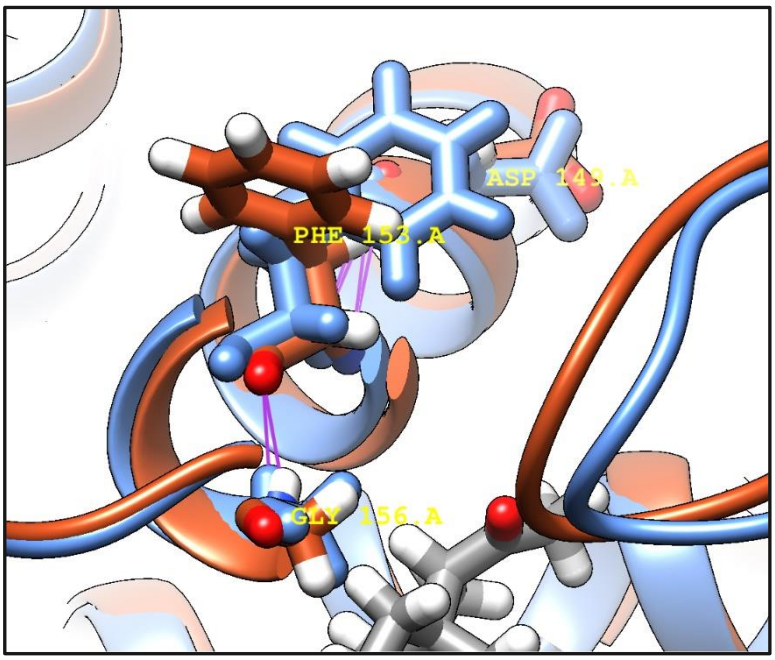

Figure 3.16 Comparison of F153F mutation (blue) with wild type enzyme (red). Residues Phe153, Asp149 and Gly156 labeled with yellow color. Hydrogen bonds (purple) and progesterone (grey) also shown.

F153F is not actually referring a real mutation at this site, it is just changing the rotation of phenylalanine side chain as a result of energy minimization done by FastRelax protocol. Most obvious difference between wild type crystal structure and relaxed enzyme is lack of the hydrogen bond between Phe153 and Gly156 backbone atoms. Figure 3.16 shows the rotation of phenylalanine side chain before and after energy minimization. This movement leads to relocation of backbone and disrupts the hydrogen bond between Phe153 and Gly156. F153F mutation creates new hydrogen bonds with Asp149. This rotational change cannot be expected in the wild type structure since F153F mutation is not real, so mutants that include F153F cannot be selected as a final model. This situation is unfortunate since 4 of total 24 mutants include F153F mutation. Even though F153F is not a real mutation, analysis of F153F still shows the importance of this residue for progesterone binding.

R-R distance map of F153F mutation (Figure 3.17) only show changes that depend on the relax procedure. Leu354 residue is only important residue that changes its location. Relocation of F-G loop changes the location of EVL sequence. As a result of Leu354 replacement F153F mutants bind progesterone both PRG-1 and PRG-2 coordinates.

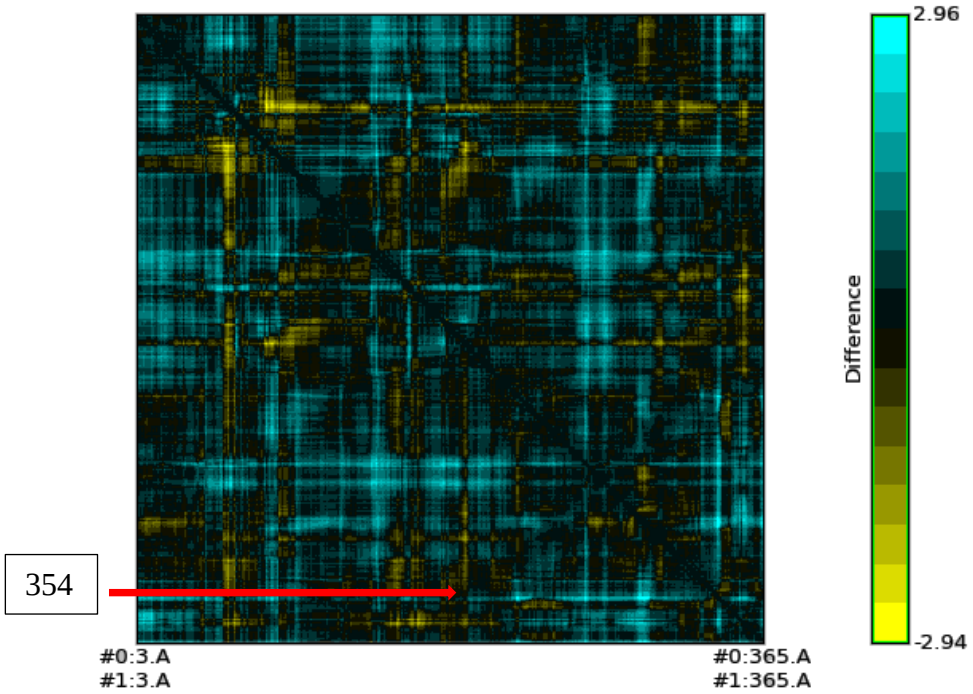

Figure 3.17 R-R distance map of F153F vs. wild type CYP119. Numbers of most flexible residues indicated in boxes.

Figure 3.18 shows F153A mutation and its effects on the location of F-G loop. F-G loop is important because it is the most flexible part of wild type CYP119 enzyme. Inhibitor bound crystal structures show that F-G loop can move up to 12 Å and can rotate more than 45°. This relocation is important for substrate binding as well as inhibitor binding. F153A mutation changes both location and structure of the F-G loop. As shown in Figure 3.18 F153A mutation moves Gly156 residue 2.14 Å and due to the loss of conserved hydrogen bond between Gly156 and Phe153. Loss of this hydrogen bond effect structure of F-G loop, moves it away from the B-helix.

R-R distance map of F153A mutation (Figure 3.19) resembles F153F mutation as expected. The only difference is that a replacement occurs in Val353 instead of Leu354. Both residues are part of EVL sequence which is responsible of substrate binding in CYP119. Interesting fact is F153A mutation often works with T214M mutation (4 out of 5) while T214M stays exactly at the same location and also does not form any kind of hydrogen bond, salt bridge etc. which can effect substrate binding.

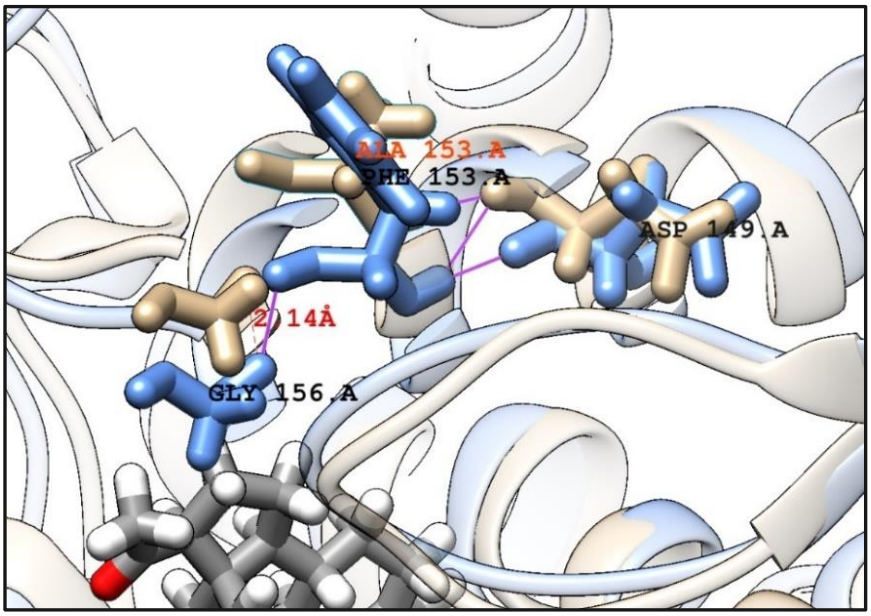

Figure 3.18 Comparison of F153A mutation (tan) with wild type enzyme (blue). Residues Phe153, Asp149 and Gly156 labeled with black and Ala153 with orange color. Distance between Gly156 residues of mutant and wild type enzyme shown with red. Hydrogen bonds (purple)

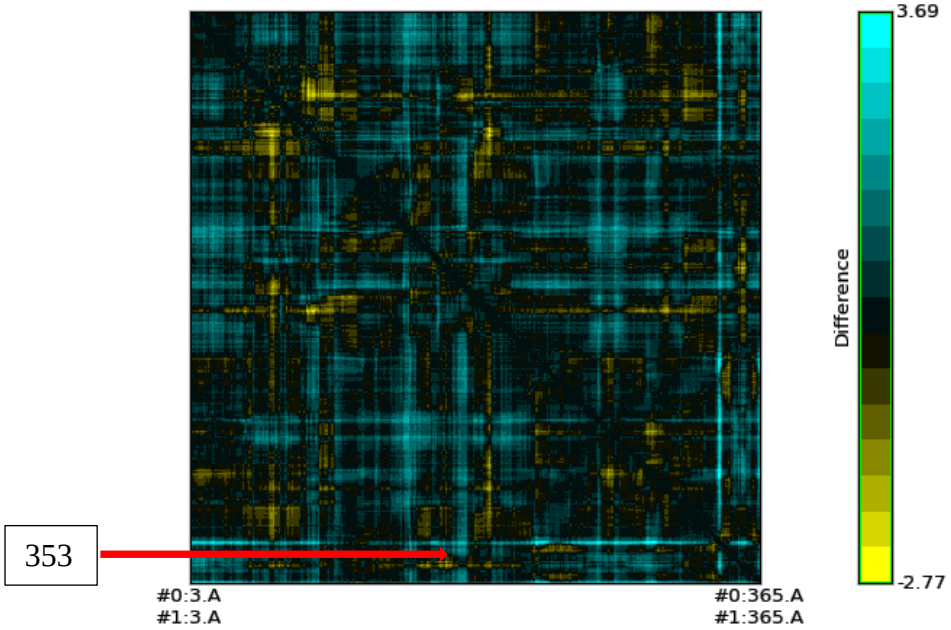

Figure 3.19 R-R distance map of F153A vs. wild type CYP119. Numbers of most flexible residues indicated in boxes.

3.5.3 T214M Mutation

T214M mutation (Figure 3.20) always pairs with F153A mutation. This connection is interesting because Thr214 residue is not pointed towards active site. It is located behind the porphyrin ring. Thr214 has no interaction with heme group. T214M mutation does not seem to be effective, especially since it's not creating new hydrogen bonds or disrupting conserved ones. It is also not changing location of any nearby residues (Figure 3.19) and not causing any clashes. The effect of this mutation is still not understood when it is combined with F153A, double mutants is binding progesterone efficiently according to docking scores.

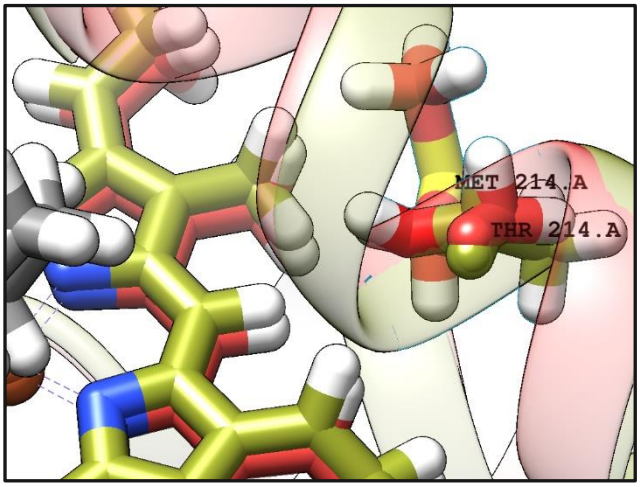

Figure 3.20 Presentation of wild type structure (red), Met214 mutation (green) and progesterone (grey).

3.5.4 L205G – L205E Mutations

Leu205 mutations have same effect of Leu69 mutations. L205G (Figure 3.21) increases the binding volume of CYP119 and helps to bind progesterone, especially in PRG-2 coordinates. Since PRG-2 is located parallel to the heme group, binding progesterone with these coordinates requires a large active site area.

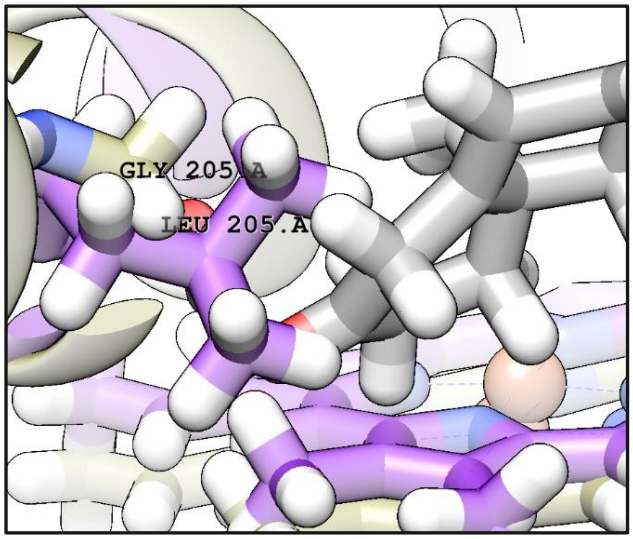

Figure 3.21 Presentation of wild type structure (purple), Glu69 mutation (tan) and progesterone (grey).

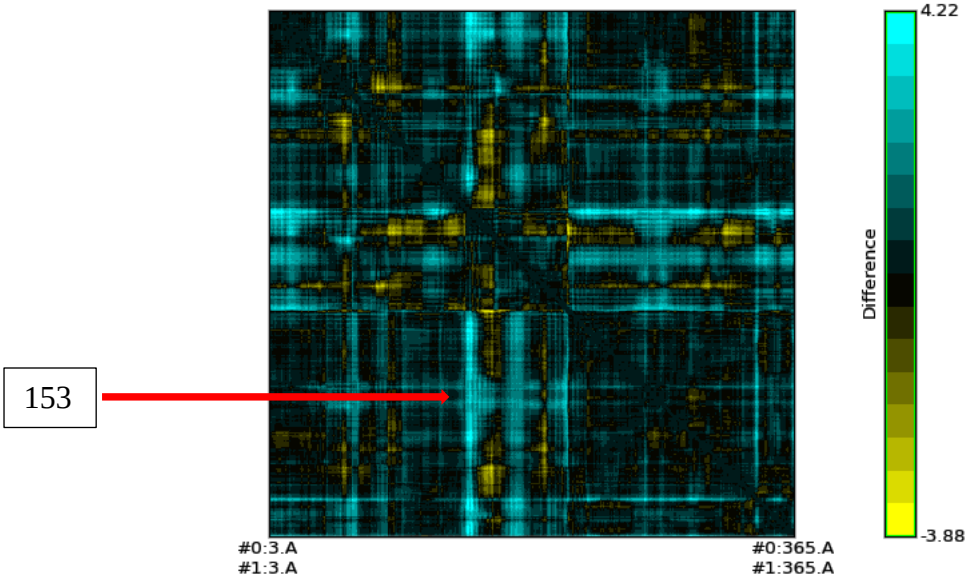

Figure 3.22 R-R distance map of L205G vs. wild type CYP119. Numbers of most flexible residues indicated in boxes.

L205E has the same effect of L69E and it helps to bind progesterone in PRG-1 coordinates since it located above of heme (Figure 3.23). Glutamate side chain along above the heme and helps to bind progesterone in PRG-1 coordinates. There is not any

significant difference in R-R map of L205E mutation that can be related to substrate binding.

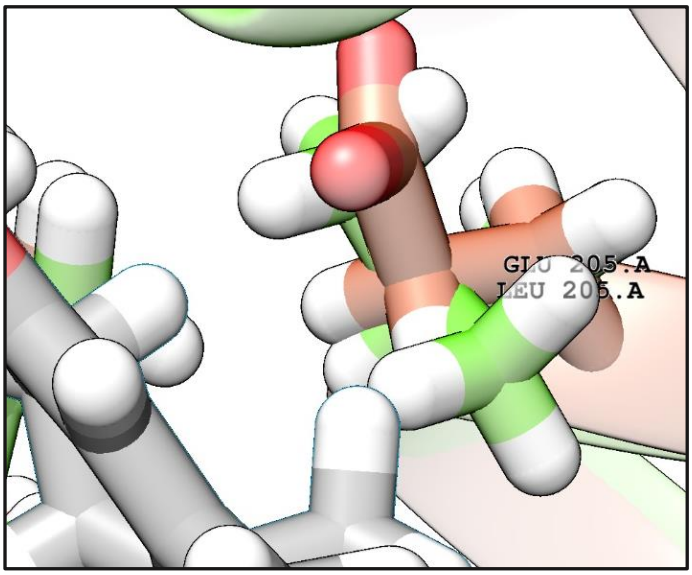

Figure 3.23 Presentation of wild type structure (green), Glu69 mutation (red) and progesterone (grey).

3.5.5 V254E Mutation

Val254 residue is a member of PPVM motif which is common for all P450s and responsible for connection conserved ERR triad to β5 sheet. Val254 acts as a substrate contact residue. V254E mutation (Figure 3.24) is only acceptable for PRG-2 coordinates of progesterone and it is not actually affecting the binding capability of mutant enzyme. Only 2 of 13 selected PRG-2 mutants have V254E mutation and there is not a significant difference between these triple mutants and their double mutant counterparts as ligand binding score. L69G-L205G and L69G-L205G-V254E have exactly same ligand binding score (-1198 REU) and difference between L69G-L354R (-1203 REU) and L69G-L354R-V254E (-1206 REU) is only three. Even though V254E mutation did not affect the binding of progesterone directly, it changes the final location of the ligand in the active site by increasing the distance from the heme group.

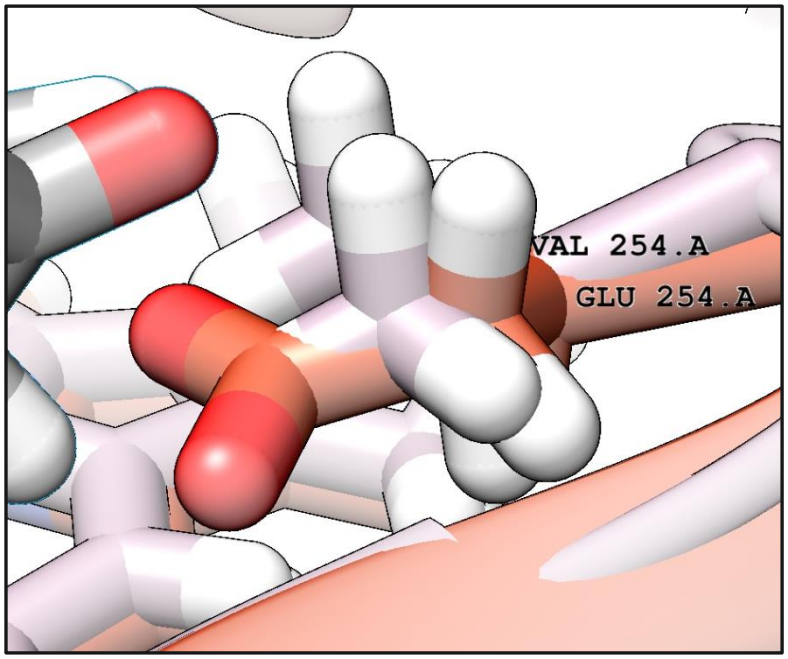

Figure 3.24 Presentation of wild type structure (pink), Glu254 mutation (red) and progesterone (grey).

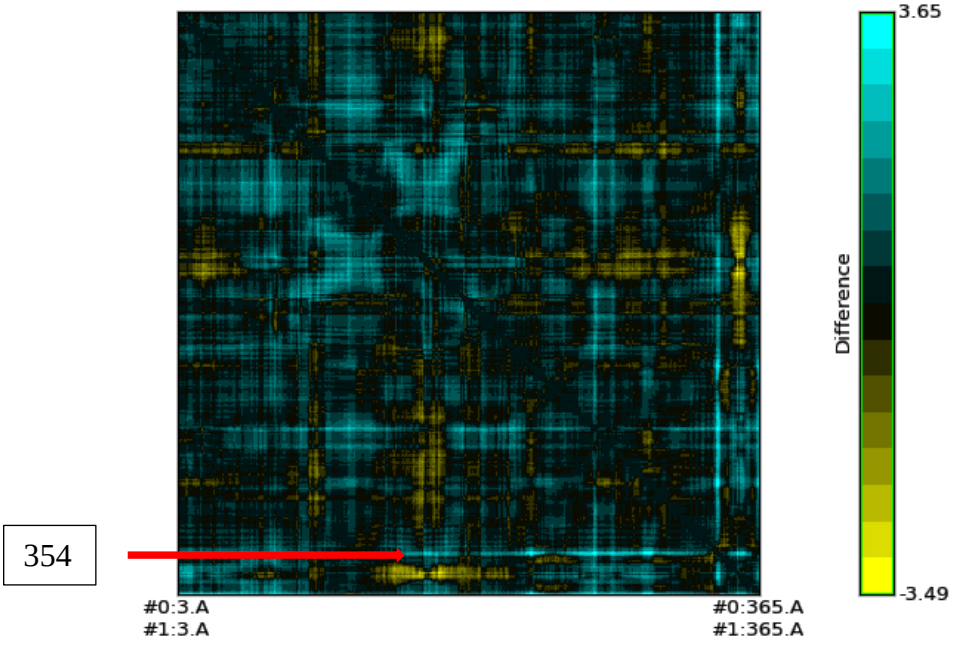

Figure 3.25 R-R distance map of V254E vs. wild type CYP119. Numbers of most flexible residues indicated in boxes.

3.5.6 L354R Mutation

Leu354 is one of the three substrate contact residues located on β5 sheet of CYP119. Leu354 acts together with Phe153 to shape substrate contact surface and plays role in the orientation of bound substrate. 7 of 24 selected mutants have L354R mutation. Arg354 makes hydrogen bonds with Asp149, Glu212 and Asn351. Hydrogen bonds with Asp149 and Glu212 make these mutants not suitable for enzyme activity. This hydrogen bonds completely close the entrance of substrate access channel thus we decided to not produce these mutants for progesterone hydroxylation. Figure 3.26 shows L354R mutation and H bonds.

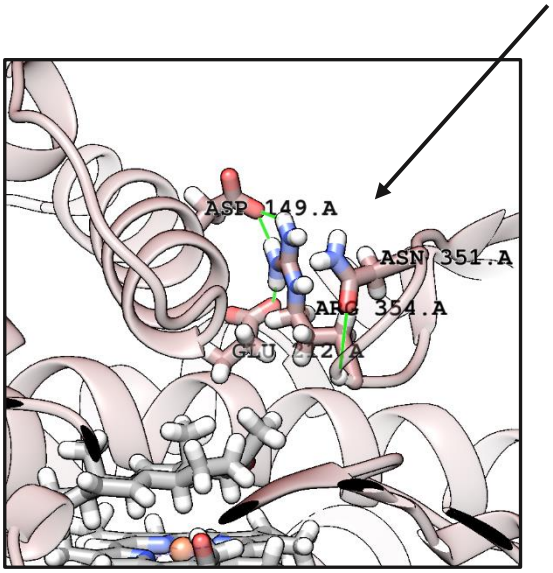

Figure 3.26 Residues forming hydrogen bond with Arg254. All hydrogen bonds colored with green. Arrow indicates substrate access channel of CYP119.

3.5.7 T257E – T257F – T257G Mutations

Thr257 forms a conserved H bond with the C ring propionate of heme group. This H bond always form between C ring and -2β arginine (Arg259 in CYP119). Mutation of Thr257 residue is acceptable as long as H bond conserved. Among to 7 amino acids only three of these mutations had acceptable ligand binding scores. Unfortunately, none of

these mutations (T257E, T257F, T257G) formed a H bond with C ring propionate. Loss of this H bond probably will lead loss of activity due to problems in heme binding. 5 of 24 final mutants have Thr257 mutation (L69E-T257E, L69G-L205G-T257F, L69G-L354R-T257F, L69G-T257E, L69G-T257G) and we decided to not produce these mutants for progesterone hydroxylation.

3.6 Structural Analysis of Docked Models

Hydroxylation site of substrate generally depends on the distance between the target atom and the heme group of CYPs. The distance between iron atom and target carbon atom should be below 5Å and the C-H-Fe angle should be between 135°-180° (Szklarz and Halpert, 1994). C17 is generally an exception of this rule because C17 is energetically more favorable than C16 even though C16 stands closer to the heme group. Molecular Dynamics simulations of human CYP17A1 enzyme shows that C17 residue of progesterone is approximately 4-fold more preferred because required energy to form tertiary radical of C17 is lower than less stable C16 secondary radical (Auchus and Miller, 1999). Table 3.19 gives the information about hydroxylation type of selected natural progesterone hydroxylating P450s, distances between Fe atom of porphyrin group and selected carbon atoms and C-H-Fe angles of corresponding residues. C-H-Fe angles and C-Fe distances measured with Avogadro program. Data clearly shows that distance and angle values are in the range of given values in the literature and can be use as an elimination criteria.

Final 11 mutants analyzed with these two perspectives (distance and angle) and possible hydroxylation sites were determined for each of them. This approach was previously used by Pinzon et al to determine all possible hydroxylation sites of steroidal drugs by CYP260B1 (Pinzon et al., 2016). Hydroxylation at C3, C5, C10 and C13 has never been observed in literature. If one of these residues are determined as the closest residue to heme group, second closest residue annotated as possible hydroxylation site. Table 3.20 – Table 3.21 and Table 3.22 – Table 3.23 show the distance between heme group and each carbon atom for best models of PRG-1 and PRG-2, respectively.

Table 3.19 Distances between targeted carbon atom and Fe atom of heme group and the angle between C-H-Fe atoms for naturally progesterone hydroxylating P450s.

	PDB ID	Product Name	Distance	Angle (Fe-C-H)
PRG-1	4Y8W	21-hydroxyprogesterone	4.869 Å	147.9°
	4J6C	16-hydroxyprogesterone	3.438 Å	138.6°
	4NXK	17-hydroxyprogesterone	4.323 Å	116.2°
	4R21	17-hydroxyprogesterone	4.192 Å	126.3°
PRG-2	6F88	1-hydroxyprogesterone	3.367 Å	155.1°
	6F8C	17-hydroxyprogesterone	4.051 Å	118.7°

Table 3.20 Double mutants of PRG-1 binding group and distances between each carbon atom and Fe atom of heme group. Predicted hydroxylation sites shown in red boxes.

Mutant Name	L69G-F153A	L69G-F153F	L69E-T214M
Possible Hydroxylation Site			
C1	9.80 Å	10.03 Å	8.72 Å
C2	11.18 Å	11.41 Å	8.75 Å
C3	11.74 Å	11.89 Å	7.65 Å
C4	10.95 Å	11.05 Å	6.31 Å
C5	9.61 Å	9.70 Å	6.18 Å
C6	9.14 Å	9.15 Å	4.95 Å
C7	8.18 Å	8.16 Å	5.63 Å
C8	6.99 Å	7.04 Å	6.89 Å
C9	7.71 Å	7.87 Å	7.87 Å
C10	8.78 Å	8.95 Å	7.50 Å
C11	7.00 Å	7.25 Å	9.30 Å
C12	6.02 Å	6.26 Å	9.92 Å
C13	5.03 Å	5.16 Å	9.16 Å
C14	6.08 Å	6.10 Å	7.82 Å
C15	5.73 Å	5.62 Å	7.51 Å
C16	4.92 Å	4.80 Å	8.97 Å
C17	4.71 Å	4.76 Å	9.88 Å
C18	4.00 Å	4.19 Å	9.26 Å
C19	8.36 Å	8.55 Å	7.61 Å
C20	3.79 Å	3.91 Å	11.20 Å
C21	3.04 Å	3.00 Å	11.59 Å

Table 3.21 Triple mutants of PRG-1 binding group and distances between each carbon atom and Fe atom of heme group. Predicted hydroxylation sites shown in red boxes.

Mutant Name	T214M-I208R-F153A	L69G-F153F-L205E
Possible Hydroxylation Site		
C1	3.80 Å	5.19 Å
C2	2.80 Å	6.04 Å
C3	3.78 Å	7.18 Å
C4	4.83 Å	7.79 Å
C5	5.15 Å	7.45 Å
C6	6.51 Å	8.58 Å
C7	7.50 Å	8.22 Å
C8	7.18 Å	7.29 Å
C9	6.03 Å	5.89 Å
C10	4.64 Å	6.25 Å
C11	6.31 Å	4.95 Å
C12	7.83 Å	5.32 Å
C13	8.65 Å	6.84 Å
C14	8.49 Å	7.38 Å
C15	9.80 Å	8.82 Å
C16	10.80 Å	8.88 Å
C17	10.14 Å	7.57 Å
C18	8.62 Å	7.60 Å
C19	4.29 Å	6.98 Å
C20	11.02 Å	7.98 Å
C21	12.25 Å	9.46 Å

Table 3.22 Double mutants of PRG-2 binding group and distances between each carbon atom and Fe atom of heme group. Predicted hydroxylation sites shown in red boxes.

Mutant Name	L69G-F153F	L69G-L205G	F153A-T214M	V151G-L155R
Possible Hydroxylation Site				
C1	8.23 Å	5.77 Å	9.08 Å	4.44 Å
C2	9.43 Å	5.88 Å	10.60 Å	3.40 Å
C3	10.41 Å	7.26 Å	11.48 Å	2.58 Å
C4	10.47 Å	8.30 Å	11.03 Å	3.40 Å
C5	9.70 Å	8.18 Å	9.69 Å	4.79 Å
C6	10.27 Å	9.44 Å	9.69 Å	5.85 Å
C7	9.37 Å	9.72 Å	8.31 Å	6.98 Å
C8	8.50 Å	8.93 Å	7.29 Å	7.31 Å
C9	7.64 Å	7.44 Å	7.25 Å	6.37 Å
C10	8.73 Å	7.12 Å	8.75 Å	5.22 Å
C11	6.73 Å	6.66 Å	6.16 Å	7.39 Å
C12	6.12 Å	7.31 Å	4.94 Å	8.45 Å
C13	6.89 Å	8.62 Å	4.75 Å	9.18 Å
C14	7.76 Å	9.39 Å	5.97 Å	8.50 Å
C15	8.85 Å	10.98 Å	6.17 Å	9.92 Å
C16	8.27 Å	11.17 Å	4.90 Å	11.00 Å
C17	6.75 Å	9.86 Å	3.61 Å	10.58 Å
C18	8.09 Å	9.03 Å	5.42 Å	9.81 Å
C19	9.56 Å	7.15 Å	9.17 Å	6.05 Å
C20	6.60 Å	9.93 Å	2.63 Å	11.58 Å
C21	7.97 Å	11.14 Å	3.35 Å	12.78 Å

Table 3.23 Triple mutants of PRG-2 binding group and distances between each carbon atom and Fe atom of heme group. Predicted hydroxylation sites shown in red boxes.

Mutant Name	L205G-L69G-V254E	L69G-V151A-L205G	F153A-T214M-L69G
Possible Hydroxylation Site			
C1	4.75 Å	4.93 Å	4.59 Å
C2	4.07 Å	5.59 Å	3.74 Å
C3	3.09 Å	6.81 Å	2.99 Å
C4	3.31 Å	7.47 Å	3.63 Å
C5	4.63 Å	7.18 Å	4.94 Å
C6	5.45 Å	8.28 Å	5.95 Å
C7	6.46 Å	8.19 Å	6.96 Å
C8	6.94 Å	7.41 Å	7.30 Å
C9	6.19 Å	5.96 Å	6.38 Å
C10	5.26 Å	6.14 Å	5.37 Å
C11	7.39 Å	5.25 Å	7.41 Å
C12	8.34 Å	5.64 Å	8.40 Å
C13	8.93 Å	6.93 Å	9.13 Å
C14	8.07 Å	7.58 Å	8.42 Å
C15	9.39 Å	9.19 Å	9.84 Å
C16	10.51 Å	9.30 Å	10.90 Å
C17	10.21 Å	7.92 Å	10.46 Å
C18	9.67 Å	7.78 Å	9.85 Å
C19	6.31 Å	6.73 Å	6.33 Å
C20	11.31 Å	8.14 Å	11.48 Å
C21	12.51 Å	9.54 Å	12.72 Å

Table 3.24 shows Fe-C distances and C-H-Fe angles of predicted hydroxylation sites of final 11 mutants. All measurements made with Avogadro program. Structures that

have C-H-Fe angles below 135 ° eliminated. Selected mutants indicated with red boxes in the table.

Table 3.24 C-Fe distances and C-H-Fe angles of selected mutants. Red boxes indicate selected mutants for laboratory production step.

Mutant	Possible Product	Distance	Angle (C-H-Fe)
L69G-F153F (PRG-1)	21-hydroxyprogesterone	3.04	129.4
L69G-F153F (PRG-2)	12-hydroxyprogesterone	6.12	151,3
L69G-L205G	1-hydroxyprogesterone	5.77	146.5
F153A-T214M	20-hydroxyprogesterone	2.63	120.9
V151G-L155R	4-hydroxyprogesterone	3.40	97.5
L69G-F153A	21-hydroxyprogesterone	3.03	100.1
L69E-T214M	6-hydroxyprogesterone	4.94	124.0
L205G-L69G-V254E	4-hydroxyprogesterone	3.31	123.2
L69G-V151A-L205G	1-hydroxyprogesterone	4.93	133.2
F153A-T214M-L69G	4-hydroxyprogesterone	3.63	109.4
L69G-F153F-L205E	11-hydroxyprogesterone	4.95	162.1
T214M-I208R-F153A	2-hydroxyprogesterone	2.80	142.5

Final 5 mutants; L69G-F153F, L69G-L205G, L69G-V151A-L205G, L69G-F153F-L205E and T214M-I208R-F153A, will be produced in our laboratory to check their progesterone hydroxylation activity experimentally. Table 3.25 shows a summary of final selected models.

Table 3.25 Summary of selected mutants for laboratory production step.

	Mutant Name	REU Score	Possible Hydroxylation Site	Distance	Angle
PRG-1	L69G-F153F-L205E	-1183	C11	4.95	162.1
	T214M-I208R-F153A	-1210	C2	2.80	142.5
PRG-2	L69G-F153F	-1194	C12	6.12	151.3
	L69G-L205G	-1198	C1	5.77	146.5
	L69G-V151A-L205G	-1184	C1	4.93	133.2

CHAPTER 4

CONCLUSION

Production of hydroxylated progesterone derivatives with high regio- and stereoselectivity is important for pharmaceutical industry. This study aims to design novel catalysts for progesterone derivative production. For this purpose, thermophilic P450 enzyme CYP119 was designed with rational design methods.

PyRosetta program was used for creating mutations and selecting mutants based on their progesterone binding scores. Total number of 674 mutants were designed and approximately 50000 rounds of docking were performed. First elimination step was applied according to REU score of each docking step. Mutants with lower docking scores were selected for following rounds of mutation creation process. 24 mutants with best docking scores were selected for structural analysis. Distances between the iron atom of heme group and each carbon atom of progesterone were measured, closest carbon atom was defined as possible hydroxylation site for each mutant. According to literature distance between iron atom and targeted carbon atom (Fe-C) should be below 5 Å and the angle between iron atom and targeted carbon atom (C-H-Fe) should be between 135°-180° (Szklarz and Halpert, 1994). Fe-C and C-H-Fe values were measured for each mutant and data used for second elimination step.

Total number of 5 mutants were finally selected at the end of process. Selected mutants are T214M-I208R-F153A and L69G-F153F-L205E for PRG-1 docking group and their expected products are 2-hydroxyprogesterone and 11-hydroxyprogesterone, respectively. L69G-F153F, L69G-L205G and L69G-V151A-L205G mutants were selected from PRG-2 docking group. Expected products of these mutants are; 1-hydroxyprogesterone for both L69G-L205G and L69G-V151A-L205G mutants and 12-hydroxyprogesterone for L69G-F153F mutant. All these mutants will be produced in our laboratories, their activity against progesterone will be measured and products will be characterized to compare experimental data with theoretical prediction. Each progesterone derivative has a unique function and importance as a pharmaceutical product. Production of these derivatives are normally catalyzed by membrane bound

mesophilic P450 enzymes. Our mutant CYP119 enzymes, which are also soluble, will be first progesterone binding thermophilic enzymes in the literature. This project provides novel catalysts for production of hydroxyprogesterones in high temperature conditions which is an important development for industrial production.

REFERENCES

Kieslich, K. (1979). *US4237220A*. United States Patent.

Marsheck, W. J., Jiu, J., & Wang, P. T. (1881). *US4397947*. United States Patent.

Masuda, Y., Nishimura, N., Tanaka, K., Takakura, I., & Shiozaki, S. (1989). *1171496*. Japanese Patent.

Albertolle, M. E., & Peter Guengerich, F. (2018). The relationships between cytochromes P450 and H_2O_2 : Production, reaction, and inhibition. *Journal of Inorganic Biochemistry*, *186*, 228–234.

Auchus, R. J., & Miller, W. L. (1999). Molecular Modeling of Human P450c17 (17α-Hydroxylase/ 17,20-Lyase): Insights into Reaction Mechanisms and Effects of Mutations. *Molecular Endocrinology*, *13*(7), 1169–1182.

Barradell, L. B., & Faulds, D. (1994). Cyproterone. *Drugs & Aging*, *5*(1), 59–80.

Behera, R. K., & Mazumdar, S. (2010). Thermodynamic basis of the thermostability of CYP175A1 from Thermus thermophilus. *International Journal of Biological Macromolecules*, *46*(4), 412–418.

Bell, S. G., Chen, X., Sowden, R. J., Xu, F., Williams, J. N., Wong, L.-L., & Rao, Z. (2003). Molecular Recognition in (+)-α-Pinene Oxidation by Cytochrome P450cam. *Journal of the American Chemical Society*, *125*(3), 705–714.

Bell, S. G., Chen, X., Xu, F., Rao, Z., & Wong, L.-L. (2003). Engineering substrate recognition in catalysis by cytochrome P450cam. *Biochemical Society Transactions*, *31*(3), 558–562.

Berg, A., Gustafsson, J. A., & Ingelman-Sundberg, M. (1976). Characterization of a cytochrome P-450-dependent steroid hydroxylase system present in Bacillus megaterium. *The Journal of Biological Chemistry*, *251*, 2831–2838.

Bernhardt, R. (2006). Cytochromes P450 as versatile biocatalysts. *Journal of Biotechnology*, *124*(1), 128–145.

Blair, E., Greaves, J., & Farmer, P. J. (2004). High-Temperature Electrocatalysis Using Thermophilic P450 CYP119: Dehalogenation of CCl4to CH4. *Journal of the American Chemical Society*, *126*(28), 8632–8633.

Brooijmans, N., & Kuntz, I. D. (2003). Molecular Recognition and Docking Algorithms. *Annual Review of Biophysics and Biomolecular Structure*, *32*(1), 335–373.

Camerino, B., & Sciaky, R. (1975). Structure and effects of anabolic steroids. *Pharmacology & Therapeutics. Part B: General and Systematic Pharmacology*, *1*(2), 233–275.

Chang, Y.-T., & Loew, G. (2000). Homology Modeling, Molecular Dynamics Simulations, and Analysis of CYP119, a P450 Enzyme from Extreme Acidothermophilic Archaeon Sulfolobus solfataricus†. *Biochemistry*, *39*(10), 2484–2498.

Chang, Y.-T., & Loew, G. H. (1999). Molecular Dynamics Simulations of P450 BM3; Examination of Substrate-Induced Conformational Change. *Journal of Biomolecular Structure and Dynamics*, *16*(6), 1189–1203.

Combs, S. A., DeLuca, S. L., DeLuca, S. H., Lemmon, G. H., Nannemann, D. P., Nguyen, E. D., Meiler, J. (2013). Small-molecule ligand docking into comparative models with Rosetta. *Nature Protocols*, *8*(7), 1277–1298.

Cryle, M. J., Stok, J. E., & De Voss, J. J. (2003). Reactions Catalyzed by Bacterial Cytochromes P450. *Australian Journal of Chemistry*, *56*(8), 749.

Cupp-Vickery, J., Anderson, R., & Hatziris, Z. (2000). Crystal structures of ligand complexes of P450eryF exhibiting homotropic cooperativity. *Proceedings of the National Academy of Sciences*, *97*(7), 3050–3055.

Dai, Z.-R., Ge, G.-B., Feng, L., Ning, J., Hu, L.-H., Jin, Q., Yang, L. (2015). A Highly Selective Ratiometric Two-Photon Fluorescent Probe for Human Cytochrome P450 1A. *Journal of the American Chemical Society*, *137*(45), 14488–14495.

Davis, I. W., & Baker, D. (2009). RosettaLigand Docking with Full Ligand and Receptor Flexibility. *Journal of Molecular Biology*, *385*(2), 381–392.

Davis, I. W., Raha, K., Head, M. S., & Baker, D. (2009). Blind docking of pharmaceutically relevant compounds using RosettaLigand. *Protein Science*, *18*(9), 1998–2002.

DeLuca, S., Khar, K., & Meiler, J. (2015). Fully Flexible Docking of Medium Sized Ligand Libraries with RosettaLigand. *PLOS ONE*, *10*(7), e0132508.

Desmet, J., Maeyer, M. D., Hazes, B., & Lasters, I. (1992). The dead-end elimination theorem and its use in protein side-chain positioning. *Nature*, *356*(6369), 539–542.

Dietrich, J. A., Yoshikuni, Y., Fisher, K. J., Woolard, F. X., Ockey, D., McPhee, D. J., Keasling, J. D. (2009). A Novel Semi-Biosynthetic Route for Artemisinin Production Using Engineered Substrate-Promiscuous P450BM3. *ACS Chemical Biology*, *4*(4), 261–267.

Di Nardo, G., & Gilardi, G. (2012). Optimization of the Bacterial Cytochrome P450 BM3 System for the Production of Human Drug Metabolites. *International Journal of Molecular Sciences*, *13*(12), 15901–15924.

Donova, M. V., & Egorova, O. V. (2012). Microbial steroid transformations: current state and prospects. *Applied Microbiology and Biotechnology*, *94*(6), 1423–1447.

Dunbrack, R. L., & Karplus, M. (1993). Backbone-dependent Rotamer Library for Proteins Application to Side-chain Prediction. *Journal of Molecular Biology*, *230*(2), 543–574.

Dunbrack, R. L., & Cohen, F. E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Science*, *6*(8), 1661–1681.

Fernandes, P., Cruz, A., Angelova, B., Pinheiro, H. M., & Cabral, J. M. S. (2003). Microbial conversion of steroid compounds: recent developments. *Enzyme and Microbial Technology*, *32*(6), 688–705.

Girvan, H. M., Waltham, T. N., Neeli, R., Collins, H. F., McLean, K. J., Scrutton, N. S., Munro, A. W. (2006). Flavocytochrome P450 BM3 and the origin of CYP102 fusion species. *Biochemical Society Transactions*, *34*(6), 1173–1177.

Gonzalez, E., & Guengerich, F. P. (2017). Kinetic processivity of the two-step oxidations of progesterone and pregnenolone to androgens by human cytochrome P450 17A1. *Journal of Biological Chemistry*, *292*(32), 13168–13185.

Gotoh, O. (1992). Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *The Journal of Biological Chemistry*, *267*, 83–90.

Gottardi, M., Tyzack, J. D., Bender, A., & Cedergreen, N. (2018). Can the inhibition of cytochrome P450 in aquatic invertebrates due to azole fungicides be estimated with in silico and in vitro models and extrapolated between species? *Aquatic Toxicology*, *201*, 11–20.

Guengerich, F. P. (2015). Human Cytochrome P450 Enzymes. *Cytochrome P450*, 523–785.

Guengerich, F. P., Wilkey, C. J., & Phan, T. T. N. (2019). Human cytochrome P450 enzymes bind drugs and other substrates mainly through conformational-selection modes. *Journal of Biological Chemistry*, *294*(28), 10928–10941.

Harris, K. L., Thomson, R. E. S., Strohmaier, S. J., Gumulya, Y., & Gillam, E. M. J. (2018). Determinants of thermostability in the cytochrome P450 fold. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1866*(1), 97–115.

Hasemann, C. A., Kurumbail, R. G., Boddupalli, S. S., Peterson, J. A., & Deisenhofer, J. (1995). Structure and function of cytochromes P450:a comparative analysis of three crystal structures. *Structure*, *3*(1), 41–62.

Herzog, K., Bracco, P., Onoda, A., Hayashi, T., Hoffmann, K., & Schallmey, A. (2014). Enzyme–substrate complex structures of CYP154C5 shed light on its mode of highly selective steroid hydroxylation. *Acta Crystallographica Section D Biological Crystallography*, *70*(11), 2875–2889.

Ho, W. W., Li, H., Nishida, C. R., Ortiz de Montellano, P. R., & Poulos, T. L. (2008). Crystal Structure and Properties of CYP231A2 from the Thermoacidophilic Archaeon Picrophilus torridus†,‡. *Biochemistry*, *47*(7), 2071–2079.

Immoos, C. E., Chou, J., Bayachou, M., Blair, E., Greaves, J., & Farmer, P. J. (2004). Electrocatalytic Reductions of Nitrite, Nitric Oxide, and Nitrous Oxide by Thermophilic Cytochrome P450 CYP119 in Film-Modified Electrodes and an Analytical Comparison of Its Catalytic Activities with Myoglobin. *Journal of the American Chemical Society*, *126*(15), 4934–4942.

Jaenicke, R., & Böhm, G. (1998). The stability of proteins in extreme environments. *Current Opinion in Structural Biology*, *8*(6), 738–748.

Jennewein, S., Rithner, C. D., Williams, R. M., & Croteau, R. B. (2001). Taxol biosynthesis: Taxane 13 -hydroxylase is a cytochrome P450-dependent monooxygenase. *Proceedings of the National Academy of Sciences*, *98*(24), 13595–13600.

Joo, H., Lin, Z., & Arnold, F. H. (1999). Laboratory evolution of peroxide-mediated cytochrome P450 hydroxylation. *Nature*, *399*(6737), 670–673.

Jóźwik, I. K., Kiss, F. M., Gricman, Ł., Abdulmughni, A., Brill, E., Zapp, J., … Thunnissen, A.-M. W. H. (2016). Structural basis of steroid binding and oxidation by the cytochrome P450 CYP109E1 fromBacillus megaterium. *The FEBS Journal*, *283*(22), 4128–4148.

Jung, C., Vries, S. de, & Schünemann, V. (2011). Spectroscopic characterization of cytochrome P450 Compound I. *Archives of Biochemistry and Biophysics*, *507*(1), 44–55.

Kellner, D. G., Hung, S.-C., Weiss, K. E., & Sligar, S. G. (2002). Kinetic Characterization of Compound I Formation in the Thermostable Cytochrome P450 CYP119. *Journal of Biological Chemistry*, *277*(12), 9641–9644.

Khatri, Y., Ringle., M., Lisurek, M., von Kries, J. P., Zapp, J., & Bernhardt, R. (2015). Substrate Hunting for the Myxobacterial CYP260A1 Revealed New 1α-Hydroxylated Products from C-19 Steroids. *ChemBioChem*, *17*(1), 90–101.

Khatri, Y., Jóźwik, I. K., Ringle, M., Ionescu, I. A., Litzenburger, M., Hutter, M. C., Bernhardt, R. (2018). Structure-Based Engineering of Steroidogenic CYP260A1 for Stereo- and Regioselective Hydroxylation of Progesterone. *ACS Chemical Biology*, *13*(4), 1021–1028.

Kingsley, L. J., & Lill, M. A. (2014). Including ligand-induced protein flexibility into protein tunnel prediction. *Journal of Computational Chemistry*, *35*(24), 1748–1756.

Kirton, S. B., Kemp, C. A., Tomkinson, N. P., St.-Gallay, S., & Sutcliffe, M. J. (2002). Impact of incorporating the 2C5 crystal structure into comparative models of cytochrome P450 2D6. *Proteins: Structure, Function, and Genetics*, *49*(2), 216–231.

Koo, L. S., Tschirret-Guth, R. A., Straub, W. E., Moënne-Loccoz, P., Loehr, T. M., & Ortiz de Montellano, P. R. (2000). The Active Site of the Thermophilic CYP119 from Sulfolobus solfataricus. *Journal of Biological Chemistry*, *275*(19), 14112–14123.

Koo, L. S., Immoos, C. E., Cohen, M. S., Farmer, P. J., & Ortiz de Montellano, P. R. (2002). Enhanced Electron Transfer and Lauric Acid Hydroxylation by Site-Directed Mutagenesis of CYP119. *Journal of the American Chemical Society*, *124*(20), 5684–5691.

Kuhl, H. (2005). Pharmacology of estrogens and progestogens: influence of different routes of administration. *Climacteric*, *8*(sup1), 3–63.

Kuhlman, B., & Baker, D. (2000). Native protein sequences are close to optimal for their structures. *Proceedings of the National Academy of Sciences*, *97*(19), 10383–10388.

Kumar, S., Chen, C. S., Waxman, D. J., & Halpert, J. R. (2005). Directed Evolution of Mammalian Cytochrome P450 2B1. *Journal of Biological Chemistry*, *280*(20), 19569–19575.

Lampe, J. N., Floor, S. N., Gross, J. D., Nishida, C. R., Jiang, Y., Trnka, M. J., & Ortiz de Montellano, P. R. (2008). Ligand-Induced Conformational Heterogeneity of Cytochrome P450 CYP119 Identified by 2D NMR Spectroscopy with the Unnatural Amino Acid13C-p-Methoxyphenylalanine. *Journal of the American Chemical Society*, *130*(48), 16168–16169.

Leach, A. R. (1994). Ligand docking to proteins with discrete side-chain flexibility. *Journal of Molecular Biology*, *235*(1), 345–356.

Leach, A. R., Shoichet, B. K., & Peishoff, C. E. (2006). Prediction of Protein−Ligand Interactions. Docking and Scoring: Successes and Gaps. *Journal of Medicinal Chemistry*, *49*(20), 5851–5855.

Lentz, O., Feenstra, A., Habicher, T., Hauer, B., Schmid, R. D., & Urlacher, V. B. (2005). Altering the Regioselectivity of Cytochrome P450 CYP102A3 of Bacillus subtilis by Using a New Versatile Assay System. *ChemBioChem*, *7*(2), 345–350.

Leys, D., Mowat, C. G., McLean, K. J., Richmond, A., Chapman, S. K., Walkinshaw, M. D., & Munro, A. W. (2002). Atomic Structure of Mycobacterium tuberculosis CYP121 to 1.06 Å Reveals Novel Features of Cytochrome P450. *Journal of Biological Chemistry*, *278*(7), 5141–5147.

Li, H., & Poulos, T. L. (1997). The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. *Nature Structural Biology*, *4*(2), 140–146.

Lisewski, A. M., & Lichtarge, O. (2006). Rapid detection of similarity in protein structure and function through contact metric distances. *Nucleic Acids Research*, *34*(22), e152.

Liu, J., Ericksen, S. S., Sivaneri, M., Besspiata, D., Fisher, C. W., & Szklarz, G. D. (2004). The effect of reciprocal active site mutations in human cytochromes P450 1A1 and 1A2 on alkoxyresorufin metabolism. *Archives of Biochemistry and Biophysics*, *424*(1), 33–43.

Liu, M., & Wang, S. (1999). MCDOCK: a Monte Carlo simulation approach to the molecular docking problem. *Journal of Computer-Aided Molecular Design*, *13*(5), 435–451.

Liu, Z., Lemmonds, S., Huang, J., Tyagi, M., Hong, L., & Jain, N. (2018). Entropic contribution to enhanced thermal stability in the thermostable P450 CYP119. *Proceedings of the National Academy of Sciences*, *115*(43), E10049–E10058.

Mann, J., & Pietrzak, B. (1989). Synthesis of 1-α-hydroxytestosterone. *Tetrahedron*, *45*(5), 1549–1552.

Mast, N., Norcross, R., Andersson, U., Shou, M., Nakayama, K., Bjorkhem, I., & Pikuleva, I. A. (2003). Broad Substrate Specificity of Human Cytochrome P450 46A1 Which Initiates Cholesterol Degradation in the Brain†. *Biochemistry*, *42*(48), 14284–14292.

Maves, S. A. (2001). Understanding thermostability in cytochrome P450 by combinatorial mutagenesis. *Protein Science*, *10*(1), 161–168.

McLean, M. A., Maves, S. A., Weiss, K. E., Krepich, S., & Sligar, S. G. (1998). Characterization of a Cytochrome P450 from the Acidothermophilic Archaea Sulfolobus solfataricus. *Biochemical and Biophysical Research Communications*, *252*(1), 166–172.

McIntosh, J. A., Heel, T., Buller, A. R., Chio, L., & Arnold, F. H. (2015). Structural Adaptability Facilitates Histidine Heme Ligation in a Cytochrome P450. *Journal of the American Chemical Society*, *137*(43), 13861–13865.

Meharenna, Y. T., & Poulos, T. L. (2010). Using Molecular Dynamics to Probe the Structural Basis for Enhanced Stability in Thermal Stable Cytochromes P450. *Biochemistry*, *49*(31), 6680–6686.

Meinhold, P., Peters, M. W., Hartwick, A., Hernandez, A. R., & Arnold, F. H. (2006). Engineering Cytochrome P450 BM3 for Terminal Alkane Hydroxylation. *Advanced Synthesis & Catalysis*, *348*(6), 763–772.

Mendel, C. M. (1989). The Free Hormone Hypothesis: A Physiologically Based Mathematical Model. *Endocrine Reviews*, *10*(3), 232–274.

Mestres, J. (2004). Structure conservation in cytochromes P450. *Proteins: Structure, Function, and Bioinformatics*, *58*(3), 596–609.

Miller, W. L., & Auchus, R. J. (2011). The Molecular Biology, Biochemistry, and Physiology of Human Steroidogenesis and Its Disorders. *Endocrine Reviews*, *32*(1), 81–151.

Misura, K. M. S., Chivian, D., Rohl, C. A., Kim, D. E., & Baker, D. (2006). Physically realistic homology models built with ROSETTA can be more accurate than their templates. *Proceedings of the National Academy of Sciences*, *103*(14), 5361–5366.

Munro, A. W., & Lindsay, J. G. (1996). Bacterial cytochromes P-450. *Molecular Microbiology*, *20*(6), 1115–1125.

Murtazina, D., Puchkaev, A. V., Schein, C. H., Oezguen, N., Braun, W., Nanavati, A., & Pikuleva, I. A. (2002). Membrane-Protein Interactions Contribute to Efficient 27-Hydroxylation of Cholesterol by Mitochondrial Cytochrome P450 27A1. *Journal of Biological Chemistry*, *277*(40), 37582–37589.

Nelson, D. R. (2018). Cytochrome P450 diversity in the tree of life. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1866*(1), 141–154.

Neria, E., Fischer, S., & Karplus, M. (1996). Simulation of activation free energies in molecular systems. *The Journal of Chemical Physics*, *105*(5), 1902–1921.

Oku, Y., Ohtaki, A., Kamitori, S., Nakamura, N., Yohda, M., Ohno, H., & Kawarabayasi, Y. (2004). Structure and direct electrochemistry of cytochrome P450 from the thermoacidophilic crenarchaeon, Sulfolobus tokodaii strain 7. *Journal of Inorganic Biochemistry*, *98*(7), 1194–1199.

Ortiz de Montellano, P. R. (1995). *Cytochrome P450: Structure, Mechanism, and Biochemistry*. (P. R. Ortiz de Montellano, Ed.) (2nd ed.). New York: Plenum Press.

Pallan, P. S., Wang, C., Lei, L., Yoshimoto, F. K., Auchus, R. J., Waterman, M. R., Egli, M. (2015). Human Cytochrome P450 21A2, the Major Steroid 21-Hydroxylase. *Journal of Biological Chemistry*, *290*(21), 13128–13143.

Park, S.-Y., Shimizu, H., Adachi, S., Nakagawa, A., Tanaka, I., Nakahara, K., Shiro, Y. (1997). Crystal structure of nitric oxide reductase from denitrifying fungus Fusarium oxysporum. *Nature Structural Biology*, *4*(10), 827–832.

Park, S.-Y., Yamane, K., Adachi, S., Shiro, Y., Weiss, K. E., & Sligar, S. G. (2000). Crystallization and preliminary X-ray diffraction analysis of a cytochrome P450 (CYP119) from Sulfolobus solfataricus. *Acta Crystallographica Section D Biological Crystallography*, *56*(9), 1173–1175.

Park, S.-Y., Yamane, K., Adachi, S., Shiro, Y., Weiss, K. E., Maves, S. A., & Sligar, S. G. (2002). Thermophilic cytochrome P450 (CYP119) from Sulfolobus solfataricus: high resolution structure and functional properties. *Journal of Inorganic Biochemistry*, *91*(4), 491–501.

Puchkaev, A. V., Wakagi, T., & Ortiz de Montellano, P. R. (2002). CYP119 Plus a Sulfolobus tokodaii Strain 7 Ferredoxin and 2-Oxoacid:Ferredoxin Oxidoreductase Constitute a High-Temperature Cytochrome P450 Catalytic System. *Journal of the American Chemical Society*, *124*(43), 12682–12683.

Puchkaev, A. V., Koo, L. S., & Ortiz de Montellano, P. R. (2003). Aromatic stacking as a determinant of the thermal stability of CYP119 from Sulfolobus solfataricus. *Archives of Biochemistry and Biophysics*, *409*(1), 52–58.

Puchkaev, A. V., & Ortiz de Montellano, P. R. (2005). The Sulfolobus solfataricus electron donor partners of thermophilic CYP119: an unusual non-NAD(P)H-dependent cytochrome P450 system. *Archives of Biochemistry and Biophysics*, *434*(1), 169–177.

Raudrant, D., & Rabe, T. (2003). Progestogens with Antiandrogenic Properties. *Drugs*, *63*(5), 463–492.

Rauschenbach, R., Isernhagen, M., Noeske-Jungblut, C., Boidol, W., & Siewert, G. (1993). Cloning sequencing and expression of the gene for cytochrome P450meg, the steroid-15?-monooxygenase from Bacillus megaterium ATCC 13368. *MGG Molecular & General Genetics*, *241–241*(1–2), 170–176.

Ravichandran, K., Boddupalli, S., Hasermann, C., Peterson, J., & Deisenhofer, J. (1993). Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. *Science*, *261*(5122), 731–736.

Rohl, C. A., Strauss, C. E. M., Misura, K. M. S., & Baker, D. (2004). Protein Structure Prediction Using Rosetta. *Methods in Enzymology*, 66–93.

Romero, R., & Stanczyk, F. Z. (2013). Progesterone is not the same as 17α-hydroxyprogesterone caproate: implications for obstetrical practice. *American Journal of Obstetrics and Gynecology*, *208*(6), 421–426.

Sagadin, T., Riehm, J. L., Milhim, M., Hutter, M. C., & Bernhardt, R. (2018). Binding modes of CYP106A2 redox partners determine differences in progesterone hydroxylation product patterns. *Communications Biology*, *1*(1).

Sagadin, T., Riehm, J., Putkaradze, N., Hutter, M. C., & Bernhardt, R. (2019). Novel approach to improve progesterone hydroxylation selectivity by CYP 106A2 via rational design of adrenodoxin binding. *The FEBS Journal*, *286*(6), 1240–1249.

Salamanca-Pinzón, S. G., & Guengerich, F. P. (2011). A tricistronic human adrenodoxin reductase-adrenodoxin–cytochrome P450 27A1 vector system for substrate hydroxylation in Escherichia coli. *Protein Expression and Purification*, *79*(2), 231–236.

Sawada, N., Sakaki, T., Yoneda, S., Kusudo, T., Shinkyo, R., Ohta, M., & Inouye, K. (2004). Conversion of vitamin D3 to 1α,25-dihydroxyvitamin D3 by Streptomyces griseolus cytochrome P450SU-1. *Biochemical and Biophysical Research Communications*, *320*(1), 156–164.

Schänzer, W., & Donike, M. (1993). Metabolism of anabolic steroids in man: synthesis and use of reference substances for identification of anabolic steroid metabolites. *Analytica Chimica Acta*, *275*(1–2), 23–48.

Schindler, A. E., Campagnoli, C., Druckmann, R., Huber, J., Pasqualini, J. R., Schweppe, K. W., & Thijssen, J. H. H. (2008). Reprint of Classification and pharmacology of progestins. *Maturitas*, *61*(1–2), 171–180.

Schmitz, D., Janocha, S., Kiss, F. M., & Bernhardt, R. (2018). CYP106A2—A versatile biocatalyst with high potential for biotechnological production of selectively hydroxylated steroid and terpenoid compounds. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1866*(1), 11–22.

Seifert, A., Tatzel, S., Schmid, R. D., & Pleiss, J. (2006). Multiple molecular dynamics simulations of human p450 monooxygenase CYP2C9: The molecular basis of substrate binding and regioselectivity toward warfarin. *Proteins: Structure, Function, and Bioinformatics*, *64*(1), 147–155.

Seifert, A., & Pleiss, J. (2009). Identification of selectivity-determining residues in cytochrome P450 monooxygenases: A systematic analysis of the substrate recognition site 5. *Proteins: Structure, Function, and Bioinformatics*, *74*(4), 1028–1035.

Sheng, X., Horner, J. H., & Newcomb, M. (2008). Spectra and Kinetic Studies of the Compound I Derivative of Cytochrome P450 119. *Journal of the American Chemical Society*, *130*(40), 13310–13320.

Simons, K. T., Kooperberg, C., Huang, E., & Baker, D. (1997). Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. *Journal of Molecular Biology*, *268*(1), 209–225.

Sivaramakrishnan, S., Ouellet, H., Du, J., McLean, K. J., Medzihradszky, K. F., Dawson, J. H., Ortiz de Montellano, P. R. (2011). A Novel Intermediate in the Reaction of Seleno CYP119 withm-Chloroperbenzoic Acid. *Biochemistry*, *50*(14), 3014–3024.

Song, Y., Tyka, M., Leaver-Fay, A., Thompson, J., & Baker, D. (2011). Structure-guided forcefield optimization. *Proteins: Structure, Function, and Bioinformatics*, *79*(6), 1898–1909.

Sterner, R. hard, & Liebl, W. (2001). Thermophilic Adaptation of Proteins. *Critical Reviews in Biochemistry and Molecular Biology*, *36*(1), 39–106.

Suzuki, R., Hirakawa, H., & Nagamune, T. (2014). Electron donation to an archaeal cytochrome P450 is enhanced by PCNA-mediated selective complex formation with foreign redox proteins. *Biotechnology Journal*, *9*(12), 1573–1581.

Szilágyi, A., & Závodszky, P. (2000). Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure*, *8*(5), 493–504.

Szklarz, G. D., & Halpert, J. R. (1997). Use of Homology Modeling in Conjunction with Site-Directed Mutagenesis for Analysis of Structure-Function Relationships of Mammalian Cytochromes P450. *Life Sciences*, *61*(26), 2507–2520.

Tong, W.-Y., & Dong, X. (2009). Microbial Biotransformation: Recent Developments on Steroid Drugs. *Recent Patents on Biotechnology*, *3*(2), 141–153.

Trosset, J.-Y., & Scheraga, H. A. (1998). Reaching the global minimum in docking simulations: A Monte Carlo energy minimization approach using Bezier splines. *Proceedings of the National Academy of Sciences*, *95*(14), 8011–8015.

Urlacher, V. B., Lutz-Wahl, S., & Schmid, R. D. (2004). Microbial P450 enzymes in biotechnology. *Applied Microbiology and Biotechnology*, *64*(3), 317–325.

Urlacher, V. B., Makhsumkhanov, A., & Schmid, R. D. (2005). Biotransformation of β-ionone by engineered cytochrome P450 BM-3. *Applied Microbiology and Biotechnology*, *70*(1), 53–59.

Urlacher, V. B., & Eiben, S. (2006). Cytochrome P450 monooxygenases: perspectives for synthetic application. *Trends in Biotechnology*, *24*(7), 324–330.

van der Willigen , A. H., Peereboom-Wynia, J. D., van Joost, T. , & Stolz, E. . (1987). A preliminary study of the effect of 11a-hydroxyprogesterone on the hair growth in men suffering from androgenetic alopecia. *Acta Dermato-Venereologica*, *67*(1), 82–85.

Vidakovic, M., Sligar, S. G., Li, H., & Poulos, T. L. (1998). Understanding the Role of the Essential Asp251 in Cytochrome P450cam Using Site-Directed Mutagenesis, Crystallography, and Kinetic Solvent Isotope Effect†. *Biochemistry*, *37*(26), 9211–9219.

Wang, Y., Han, K.-L., Yang, S.-L., & Yang, L. (2004). Structural determinants of steroids for cytochrome P450 3A4-mediated metabolism. *Journal of Molecular Structure: Theochem*, *710*(1–3), 215–221.

Wright, R. L., Harris, K., Solow, B., White, R. H., & Kennelly, P. J. (1996). Cloning of a potential cytochrome P450 from the Archaeon Sulfolobus solfataricus. *FEBS Letters*, *384*(3), 235–239.

Xiong, Y., Qiao, Y., Kihara, D., Zhang, H.-Y., Zhu, X., & Wei, D.-Q. (2019). Survey of Machine Learning Techniques for Prediction of the Isoform Specificity of Cytochrome P450 Substrates. *Current Drug Metabolism*, *20*(3), 229–235.

Yano, J. K., Koo, L. S., Schuller, D. J., Li, H., Ortiz de Montellano, P. R., & Poulos, T. L. (2000). Crystal Structure of a Thermophilic Cytochrome P450 from the Archaeon Sulfolobus solfataricus. *Journal of Biological Chemistry*, *275*(40), 31086–31092.

Yano, J. K., Blasco, F., Li, H., Schmid, R. D., Henne, A., & Poulos, T. L. (2002). Preliminary Characterization and Crystal Structure of a Thermostable Cytochrome P450 from Thermus thermophilus. *Journal of Biological Chemistry*, *278*(1), 608–616.

Yasuda, K., Ikushiro, S., Kamakura, M., Ohta, M., & Sakaki, T. (2010). Metabolism of Sesamin by Cytochrome P450 in Human Liver Microsomes. *Drug Metabolism and Disposition*, *38*(12), 2117–2123.

Zhou, M., Diwu, Z., Panchuk-Voloshina, N., & Haugland, R. P. (1997). A Stable Nonfluorescent Derivative of Resorufin for the Fluorometric Determination of Trace Hydrogen Peroxide: Applications in Detecting the Activity of Phagocyte NADPH Oxidase and Other Oxidases. *Analytical Biochemistry*, *253*(2), 162–168.

APPENDIX A

```
from rosetta import *
from toolbox import *
from rosetta.protocols.rigid import *
from rosetta.protocols.ligand_docking import *
init()

import os

pose = pose_from_pdb("1f4t.clean.pdb")

scorefxn = create_score_function("ref2015")

outfile = open(r"C:\Users\Ekin Kestevur\Desktop\Dock_results.txt","a")

from rosetta.protocols.relax import *
relax = FastRelax()
relax.set_scorefxn(scorefxn)

generate_resfile_from_pose(pose, "1f4t.resfile")

kT = 1
outfile.write("Native energy of wild type protein = %f\r\n" %
scorefxn(pose))

from rosetta.core.pack.task import TaskFactory

pose_list = ["L69E-F153A-T214M", "L69E-F153A-I208R", "L69E-F153A-
T257E"]
for i in pose_list:
    v = i
    i = Pose()
    i.assign(pose)
    print i
    task_design = standard_task_factory()
    task_design.push_back(ReadResfile(str(v) + ".resfile"))
    pack_mover = PackRotamersMover(scorefxn)
    pack_mover.task_factory(task_design)
```

```
    pack_mover.apply(i)
    relax.apply(i)
    i.dump_pdb(str(v) + ".pdb")
    outfile.write("Total energy of " + str(v) + " mutant = %f\r\n" %
scorefxn(i))
    outfile.write("*" * 12 + "\n")

outfile.close()


# DON'T FORGET TO COPY MUTANT FILES BEFORE THIS STEP!!!!
# WE WILL ADD DIFFERENT PROGESTERONE COORDINATES!

for v in pose_list:
    with open(r"C:\Users\Ekin Kestevur\\" + str(v) + ".pdb", "r+") as
f:
        new_f = f.readlines()
        f.seek(0)
        for line in new_f:
            if "HETATM" not in line:
                f.write(line)
        f.truncate()
    output = open(r"C:\Users\Ekin Kestevur\\" + str(v) + ".pdb", "a")
    output.write("""
HETATM    1  C1  STR X   1      48.921  18.092   5.026  1.00 30.82
HETATM    2  C2  STR X   1      48.630  17.921   3.530  1.00 30.71
HETATM    3  C3  STR X   1      49.868  18.235   2.719  1.00 30.75
HETATM    4  O1  STR X   1      49.772  18.671   1.585  1.00 30.37
HETATM    5  C4  STR X   1      51.206  18.006   3.303  1.00 30.72
HETATM    6  C5  STR X   1      51.379  17.607   4.578  1.00 30.50
HETATM    7  C6  STR X   1      52.814  17.352   5.014  1.00 30.58
HETATM    8  C7  STR X   1      53.084  18.061   6.342  1.00 30.37
HETATM    9  C8  STR X   1      52.030  17.679   7.386  1.00 30.72
HETATM   10  C9  STR X   1      50.606  18.088   6.905  1.00 30.67
HETATM   11  C10 STR X   1      50.210  17.426   5.554  1.00 30.76
HETATM   12  C11 STR X   1      49.518  18.005   8.012  1.00 30.63
HETATM   13  C12 STR X   1      49.931  18.474   9.420  1.00 30.40
HETATM   14  C13 STR X   1      51.286  17.918   9.864  1.00 30.93
HETATM   15  C14 STR X   1      52.313  18.268   8.780  1.00 30.87
HETATM   16  C15 STR X   1      53.695  18.095   9.405  1.00 30.98
HETATM   17  C16 STR X   1      53.458  18.478  10.862  1.00 31.53
```

```
HETATM   18  C17 STR X   1      51.935  18.652  11.042  1.00 31.38
HETATM   19  C18 STR X   1      51.139  16.406  10.169  1.00 30.97
HETATM   20  C19 STR X   1      49.957  15.913   5.706  1.00 30.84
HETATM   21  C20 STR X   1      51.422  18.309  12.432  1.00 31.63
HETATM   22  O2  STR X   1      50.238  18.485  12.697  1.00 32.11
HETATM   23  C21 STR X   1      52.343  17.792  13.495  1.00 31.67
HETATM   24  H1  STR X   1      48.080  17.671   5.578  1.00  0.00
HETATM   25  H2  STR X   1      48.973  19.159   5.244  1.00  0.00
HETATM   26  H3  STR X   1      48.325  16.892   3.338  1.00  0.00
HETATM   27  H4  STR X   1      47.826  18.597   3.239  1.00  0.00
HETATM   28  H5  STR X   1      52.077  18.164   2.685  1.00  0.00
HETATM   29  H6  STR X   1      52.970  16.280   5.138  1.00  0.00
HETATM   30  H7  STR X   1      53.497  17.733   4.254  1.00  0.00
HETATM   31  H8  STR X   1      53.056  19.139   6.184  1.00  0.00
HETATM   32  H9  STR X   1      54.071  17.777   6.707  1.00  0.00
HETATM   33  H10 STR X   1      52.043  16.593   7.481  1.00  0.00
HETATM   34  H11 STR X   1      50.687  19.151   6.678  1.00  0.00
HETATM   35  H12 STR X   1      48.675  18.618   7.693  1.00  0.00
HETATM   36  H13 STR X   1      49.181  16.971   8.082  1.00  0.00
HETATM   37  H14 STR X   1      49.172  18.147  10.131  1.00  0.00
HETATM   38  H15 STR X   1      49.977  19.563   9.429  1.00  0.00
HETATM   39  H16 STR X   1      52.209  19.345   8.646  1.00  0.00
HETATM   40  H17 STR X   1      54.037  17.064   9.321  1.00  0.00
HETATM   41  H18 STR X   1      54.411  18.774   8.943  1.00  0.00
HETATM   42  H19 STR X   1      53.971  19.412  11.091  1.00  0.00
HETATM   43  H20 STR X   1      53.820  17.686  11.517  1.00  0.00
HETATM   44  H21 STR X   1      51.729  19.712  10.893  1.00  0.00
HETATM   45  H22 STR X   1      52.101  16.004  10.486  1.00  0.00
HETATM   46  H23 STR X   1      50.407  16.266  10.964  1.00  0.00
HETATM   47  H24 STR X   1      50.805  15.885   9.272  1.00  0.00
HETATM   48  H25 STR X   1      50.860  15.429   6.079  1.00  0.00
HETATM   49  H26 STR X   1      49.141  15.750   6.410  1.00  0.00
HETATM   50  H27 STR X   1      49.691  15.490   4.737  1.00  0.00
HETATM   51  H28 STR X   1      51.774  17.588  14.402  1.00  0.00
HETATM   52  H29 STR X   1      52.817  16.873  13.149  1.00  0.00
HETATM   53  H30 STR X   1      53.109  18.538  13.707  1.00  0.00
TER
HETATM    1  C1  HEM X   1      58.435  17.806  15.041  1.00 20.00
HETATM    2  C2  HEM X   1      54.889  19.461  17.930  1.00 20.00
HETATM    3  C3  HEM X   1      55.661  23.971  16.283  1.00 20.00
HETATM    4  C4  HEM X   1      58.792  22.213  13.083  1.00 20.00
```

```
HETATM    5  C5  HEM X   1      57.463  17.884  16.047  1.00 20.00
HETATM    6  C6  HEM X   1      57.005  16.748  16.809  1.00 20.00
HETATM    7  C7  HEM X   1      55.982  17.210  17.603  1.00 20.00
HETATM    8  C8  HEM X   1      55.831  18.610  17.315  1.00 20.00
HETATM    9  C9  HEM X   1      55.183  16.402  18.633  1.00 20.00
HETATM   10  C10 HEM X   1      57.558  15.324  16.699  1.00 20.00
HETATM   11  C11 HEM X   1      56.545  14.355  16.053  1.00 20.00
HETATM   12  C12 HEM X   1      57.070  12.932  15.923  1.00 20.00
HETATM   13  O1  HEM X   1      57.980  12.555  16.692  1.00 20.00
HETATM   14  O2  HEM X   1      56.550  12.176  15.065  1.00 20.00
HETATM   15  C13 HEM X   1      54.768  20.833  17.694  1.00 20.00
HETATM   16  C14 HEM X   1      53.964  21.714  18.504  1.00 20.00
HETATM   17  C15 HEM X   1      54.157  22.977  18.032  1.00 20.00
HETATM   18  C16 HEM X   1      55.141  22.862  16.975  1.00 20.00
HETATM   19  C17 HEM X   1      53.152  21.317  19.734  1.00 20.00
HETATM   20  C18 HEM X   1      53.491  24.150  18.374  1.00 20.00
HETATM   21  C19 HEM X   1      53.246  24.596  19.742  1.00 20.00
HETATM   22  C20 HEM X   1      56.608  23.887  15.273  1.00 20.00
HETATM   23  C21 HEM X   1      57.162  25.026  14.538  1.00 20.00
HETATM   24  C22 HEM X   1      57.977  24.502  13.559  1.00 20.00
HETATM   25  C23 HEM X   1      57.984  23.085  13.783  1.00 20.00
HETATM   26  C24 HEM X   1      56.947  26.504  14.897  1.00 20.00
HETATM   27  C25 HEM X   1      58.678  25.123  12.512  1.00 20.00
HETATM   28  C26 HEM X   1      58.502  26.463  12.049  1.00 20.00
HETATM   29  C27 HEM X   1      58.935  20.861  13.343  1.00 20.00
HETATM   30  C28 HEM X   1      59.786  20.007  12.550  1.00 20.00
HETATM   31  C29 HEM X   1      59.764  18.802  13.149  1.00 20.00
HETATM   32  C30 HEM X   1      58.832  18.890  14.252  1.00 20.00
HETATM   33  C31 HEM X   1      60.572  20.378  11.286  1.00 20.00
HETATM   34  C32 HEM X   1      60.594  17.625  12.699  1.00 20.00
HETATM   35  C33 HEM X   1      61.724  17.254  13.643  1.00 20.00
HETATM   36  C34 HEM X   1      62.412  15.976  13.233  1.00 20.00
HETATM   37  O3  HEM X   1      61.745  15.109  12.615  1.00 20.00
HETATM   38  O4  HEM X   1      63.615  15.826  13.536  1.00 20.00
HETATM   39  N1  HEM X   1      56.737  19.024  16.350  1.00 20.00
HETATM   40  N2  HEM X   1      55.505  21.543  16.761  1.00 20.00
HETATM   41  N3  HEM X   1      57.129  22.691  14.808  1.00 20.00
HETATM   42  N4  HEM X   1      58.298  20.177  14.349  1.00 20.00
HETATM   43 FE1  HEM X   1      56.853  20.817  15.518  1.00 20.00
HETATM   44  H1  HEM X   1      58.906  16.850  14.864  1.00 20.00
HETATM   45  H2  HEM X   1      54.207  19.017  18.640  1.00 20.00
```

```
HETATM   46  H3  HEM X   1      55.299  24.951  16.556  1.00 20.00
HETATM   47  H4  HEM X   1      59.362  22.620  12.261  1.00 20.00
HETATM   48  H5  HEM X   1      58.462  15.344  16.090  1.00 20.00
HETATM   49  H6  HEM X   1      57.809  14.962  17.696  1.00 20.00
HETATM   50  H7  HEM X   1      56.299  14.725  15.058  1.00 20.00
HETATM   51  H8  HEM X   1      55.638  14.341  16.657  1.00 20.00
HETATM   52  H9  HEM X   1      53.130  24.775  17.570  1.00 20.00
HETATM   53  H10 HEM X   1      59.424  24.528  12.007  1.00 20.00
HETATM   54  H11 HEM X   1      59.936  16.762  12.601  1.00 20.00
HETATM   55  H12 HEM X   1      61.018  17.852  11.721  1.00 20.00
HETATM   56  H13 HEM X   1      61.316  17.127  14.646  1.00 20.00
HETATM   57  H14 HEM X   1      62.455  18.062  13.657  1.00 20.00
HETATM   58  H15 HEM X   1      53.271  25.685  19.780  1.00 20.00
HETATM   59  H16 HEM X   1      54.017  24.193  20.399  1.00 20.00
HETATM   60  H17 HEM X   1      52.268  24.243  20.070  1.00 20.00
HETATM   61  H18 HEM X   1      54.214  16.134  18.213  1.00 20.00
HETATM   62  H19 HEM X   1      55.036  17.003  19.531  1.00 20.00
HETATM   63  H20 HEM X   1      55.732  15.495  18.888  1.00 20.00
HETATM   64  H21 HEM X   1      56.425  26.574  15.851  1.00 20.00
HETATM   65  H22 HEM X   1      56.350  26.983  14.121  1.00 20.00
HETATM   66  H23 HEM X   1      57.913  27.004  14.973  1.00 20.00
HETATM   67  H24 HEM X   1      53.083  20.231  19.789  1.00 20.00
HETATM   68  H25 HEM X   1      52.151  21.741  19.660  1.00 20.00
HETATM   69  H26 HEM X   1      53.642  21.696  20.631  1.00 20.00
HETATM   70  H27 HEM X   1      61.309  21.144  11.528  1.00 20.00
HETATM   71  H28 HEM X   1      59.885  20.760  10.531  1.00 20.00
HETATM   72  H29 HEM X   1      61.080  19.494  10.901  1.00 20.00
HETATM   73  H30 HEM X   1      58.989  26.580  11.081  1.00 20.00
HETATM   74  H31 HEM X   1      58.945  27.156  12.765  1.00 20.00
HETATM   75  H32 HEM X   1      57.438  26.676  11.949  1.00 20.00
TER """)
    output.close()


output = open(r"C:\Users\Ekin Kestevur\Desktop\Dock_results.txt", "a")
output.write("DOCKING RESULTS FOR PRG-1 \n")
scorefxn = create_score_function("ligand")
docking = DockMCMProtocol()
docking.set_scorefxn(scorefxn)

params_list = Vector1(["STR.params"])
```

```
res_set = generate_nonstandard_residue_set(params_list)

for i in pose_list:
    v = str(i) + ".pdb"
    pose = Pose()
    pose_from_pdb(pose,res_set, str(v))
    job_output = str(i) + "_prg_1"
    jd = PyJobDistributor(job_output, 10, scorefxn)
    test_pose = Pose()
    counter = 0
    while not jd.job_complete:
        test_pose.assign(pose)
        counter += 1
        test_pose.pdb_info().name(job_output + '_' + str(counter))
        docking.apply(test_pose)
        test_pose.pdb_info().name(job_output  +  '_'  +  str(counter)  +
'_fa')
        jd.output_decoy(test_pose)

for i in pose_list:
    x = 1
    name = str(i) + ".pdb"
    output.write("*" * 12 + "\n")
    output.write("Docking results of model " + str(name) + "\n")
    for x in range(1,11):
        v = str(i) + "_prg_1_" + str(x) + ".pdb"
        pose = Pose()
        pose_from_pdb(pose, res_set, str(v))
        scr = int(scorefxn(pose))
        if scr <= 0:
            output.write("Docking score " + str(x) + " = %f\r\n" % scr)
        if scr >= 0:
            os.remove(r"C:\Users\Ekin Kestevur\\" + str(v))

output.close()
```

APPENDIX B

Figure B.1 REU score distribution graphics of PRG-1 docked mutants and WT CYP119 enzyme. 1000 rounds of docking performed for each mutant. (cont. in the next page)

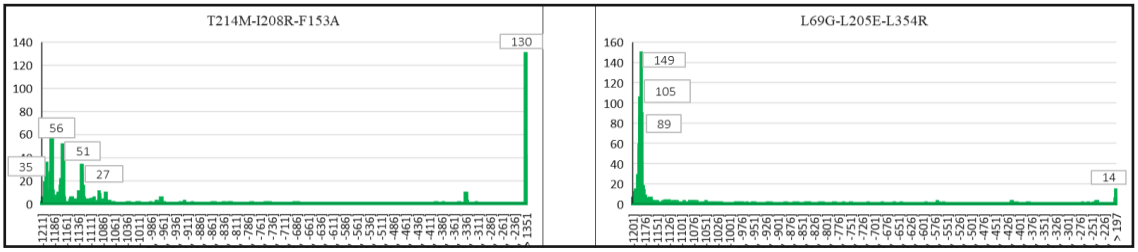

Figure B.1 (cont.)

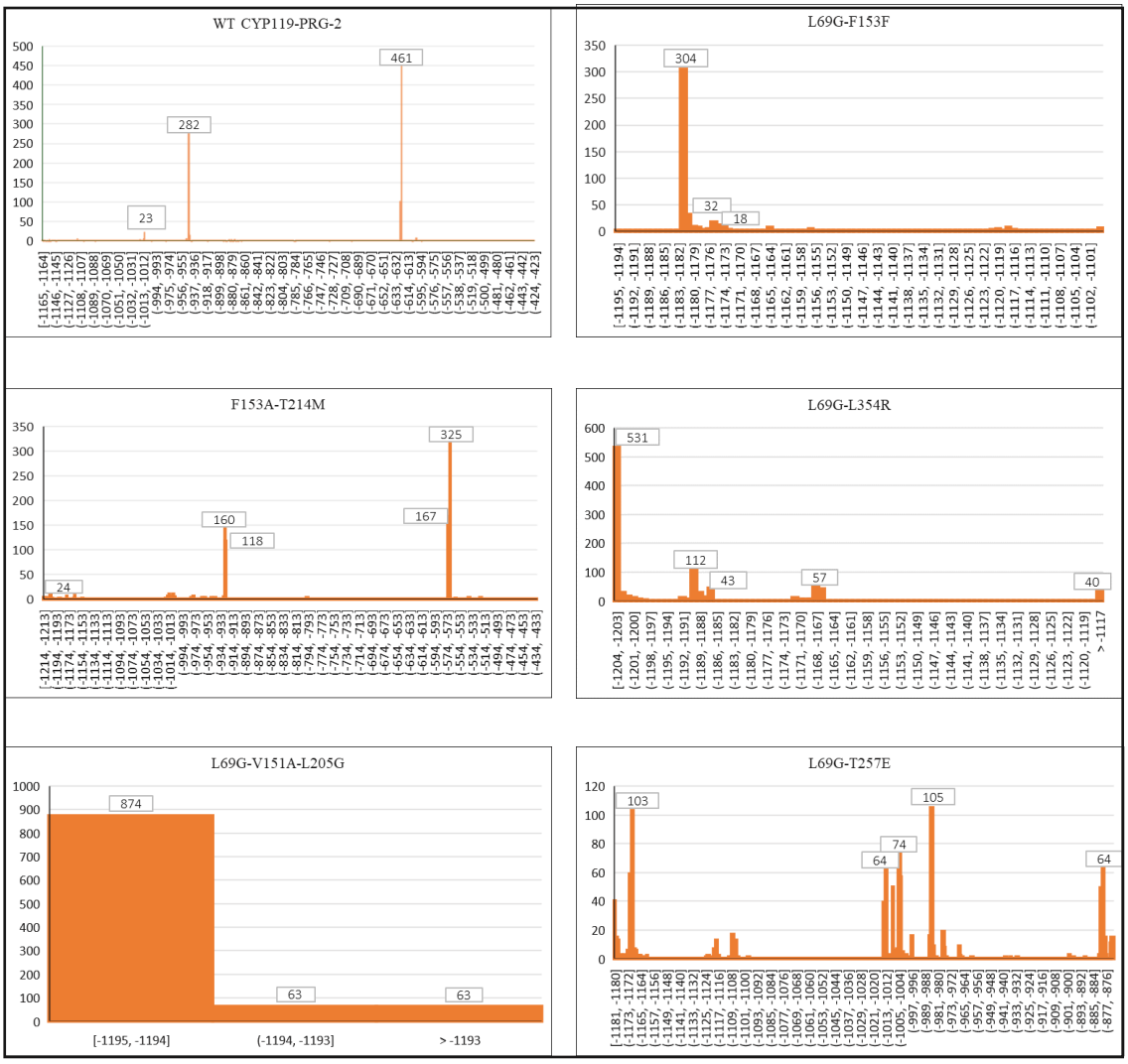

Figure B.2 REU score distribution graphics of PRG-2 docked mutants and WT CYP119 enzyme. 1000 rounds of docking performed for each mutant. (cont. in the next page)

Figure B.2 (cont.)