GENERATION OF MUTANT LIBRARIES FOR DIRECTED EVOLUTION OF A THERMOPHILIC P450 ENZYME

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

MASTER OF SCIENCE

in Biotechnology

by
Emre HAKLI

July 2018
İZMİR

We approve the thesis of **Emre HAKLI**

Examining Committee Members:

Asst. Prof. Nur Başak SÜRMELİ
Department of Bioengineering, İzmir Institute of Technology

Assoc. Prof. Gülşah ŞANLI MOHAMED
Department of Chemistry, İzmir Institute of Technology

Asst. Prof. Hümeyra TAŞKENT SEZGİN
Department of Genetics and Bioengineering, İzmir Economy University

9 July 2018

Asst. Prof. Nur Başak SÜRMELİ
Supervisor,
Department of Bioengineering
İzmir Institute of Technology

Prof. Erdal BEDİR
Co-supervisor,
Department of Bioengineering
İzmir Institute of Technology

Assoc. Prof. Engin ÖZÇİVİCİ
Head of the Programme of Biotechnology
and Bioengineering

Prof. Aysun SOFUOĞLU
Dean of Graduate School of
Engineering and Sciences

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my supervisor Asst. Prof. Nur Başak SÜRMELİ, who has the behavior and the substance of a genius. She continually supported my studies and provided all the knowledge and materials generously. I feel so lucky for being educated by such a precious academician.

I would like to thank from the bottom of my heart to my co-supervisor Prof. Erdal BEDİR who contributed a lot for me to start master's degree in İzmir Institute of Technology. I am grateful for his fatherly approach and supports both mentally and financially. Without his guidance and persistent help, life would be much harder for me.

In addition, I would like to acknowledge 116Z380 project supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK). Also, special thanks to İzmir Institute of Technology, Biotechnology and Bioengineering Applications and Research Centre (BİYOMER) for their contribution.

Besides, a thank you to Asst. Prof. Çağlar KARAYA for opening me his laboratory during my thesis studies.

Finally, I would like to thank to my great family including my mother, father, sister, grandmother, grandfather, aunt, uncles, cousins for supporting and encouraging me to follow my dreams.

ABSTRACT

GENERATION OF MUTANT LIBRARIES FOR DIRECTED EVOLUTION OF A THERMOPHILIC P450 ENZYME

Directed evolution, inspires from natural selection, is a frequently utilized approach in protein engineering for designing enzymes. It allows iterative evolution of existing proteins towards the ones with desired characteristics by the application of random mutagenesis in the laboratory. However, library construction constitutes the most fundamental part of directed evolution. Application of different construction methods affects both the number and diversity of variants created and the screening/selection techniques used. Early procedures including error-prone PCR, mutator strains, chemical mutagens and gene shuffling have been successful in whole gene mutagenesis yet have been required more screening/selection effort by leading larger libraries. On the other hand, recent approaches such as use of degenerate primers and site saturation mutagenesis have decreased the screening/selection effort by allowing random mutagenesis of amino acids located at specific positions in the polypeptide chain. Especially, active site residues of biocatalysts were chosen as targets and the catalytic efficiencies were enhanced.

CYP119, a member of cytochrome P450 protein family, from *Sulfolobus Acidocaldarius* is a thermostable enzyme capable of catalyzing peroxidation, monooxygenation and oxidoreduction reactions. Here, a library of mutants consist of CYP119 variants was created via application of combinatorial active site saturation test (CAST) in amino acid positions 213 – 214 and an effective fluorescence-based method was developed to screen the library for increased peroxidase activity while utilizing hydrogen peroxide as oxidant. After screening of mutant library, a variant with Thr213Arg – Thr214Ile substitutions showed 1.32-fold increased peroxidase activity in the catalysis of Amplex Red compared to wild type CYP119.

ÖZET

TERMOFİLİK BİR P450 ENZİMİN YÖNLENDİRİLMİŞ EVRİMİ İÇİN MUTANT KÜTÜPHANELERİNİN OLUŞTURULMASI

Doğal seçilimden ilham alan yönlendirilmiş evrim, protein mühendisliğinde enzimler tasarlanırken sıklıkla kullanılan bir yöntemdir. Laboratuvar ortamında rastgele mutasyon uygulanarak, mevcut proteinlerin istenen karakterleri taşıyan proteinlere yinelemeli olarak evrilmelerine olanak tanımaktadır. Bununla birlikte, kütüphane oluşturulması, yönlendirilmiş evrimin en temel parçalarından biridir. Farklı inşa metodlarının kullanılması hem oluşturulan varyantların sayısını ve çeşitliliğini hem de kullanılan tarama/seçme tekniklerini etkilemektedir. İlk çıkan prosedürlerden olan hata meyilli PZR, mutasyon yapıcı suşlar, kimyasal mutajenler ve DNA karıştırma tüm genin mutasyonunda başarılı olsa da büyük kütüphaneler oluşturmalarından dolayı daha fazla tarama/seçme zahmeti gerektirdi. Diğer taraftan dejenere primerlerin kullanımı ve bölge doyurma mutasyonu gibi yeni yaklaşımlar, polipeptit zincirinde, belirli bölgelerde bulunan amino asitlerin rastgele mutasyonuna olanak tanıyarak tarama/seçme eforunu azalttı. Özellikle, biyokatalizörlerin aktif bölgeleri hedef olarak seçildi ve katalitik verimlilikleri arttırıldı.

Sulfolobus Acidocaldarius'dan gelen, bir P450 protein ailesi üyesi olan CYP119, peroksidasyon, monooksidasyon ve oksidoredüksiyon tepkimelerini gerçekleştirebilen termostabil bir enzimdir. Burada, 213. ve 214. amino asit pozisyonlarında, kombinatoryal aktif bölge doygunluk testi uygulanarak CYP119 varyantlarından oluşan bir mutant kütüphanesi yaratılmış ve hidrojen peroksiti oksidan olarak kullanırken arttırılmış peroksidaz aktivitesi için florasan temelli, etkili bir tarama yöntemi geliştirilmiştir. Mutant kütüphanenin taranması sonucu, Thr213Arg – Thr214Ile mutasyonlarını taşıyan varyantın doğal CYP119 ile karşılaştırıldığında 1.32 kat daha fazla peroksidaz aktivitesi gösterdiği tespit edilmiştir.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1. Protein Engineering

Protein engineering is the creation of novel enzymes or proteins with new or improved functions. Altering or replacing the amino acids in a protein by utilizing recombinant DNA technology constitute the basis of protein engineering.[1]

In the second half of 20th century, developments occurred in genetic engineering gave rise to the concept of protein engineering. Firstly, the protein of interest has been able to be produced via employing gene cloning. Secondly, chemical synthesis of DNA allowed the application of genetic modifications on a protein. Finally, X-ray crystallography brought a better understanding to the structure-function relationship of proteins by revealing their 3D structure. Through utilizing methodologies improved so far now, either a protein with novel functions or a totally new protein not found in nature can be created.[2]

1.1.1. Enzyme Engineering

Among the proteins, design and improvement of enzymes attract the major interest because of enzyme's diverse utilization in many areas including pharmaceutical, food, environmental, detergent and paper industries. Screening and selection methods have been developed to detect enzyme variants with desired properties to satisfy the needs of industry. Tailor-made mutations of enzymes in general are intended to provide upgraded kinetic characteristics, elimination of allosteric regulation, enhancement of substrate and reaction specificity, increased thermostability, higher stability towards oxidizing agents and heavy metals and resistance to proteolytic degradation. Moreover, optimal pH and temperature of an enzyme can be adjusted according to reaction conditions. Target product yield can be increased by tuning biocatalyst's enantioselectivity, chemoselectivity and regioselectivity through applying enzyme

engineering methods. Conjunction of two or more biocatalyst domains may allow the production of multi-functional enzymes.[3]

1.2. Steps in Protein Engineering

Protein engineering strategies require to be planned considering three basic steps including: determination of the change in the protein (design strategies, such as rational design or random mutagenesis), application of those changes (mutagenesis, library creation) and interpretation of protein variants for upgraded features (screening or selection). Selection of particular strategy in each step may lead facilitated or beclouded operation and preference frequently depends on both existing information about enzyme and desired change in enzyme [4]

The most commonly employed protein engineering approaches are rational design and directed evolution (Figure 1).[5] Before application, both approaches demand the gene(s) encoding the enzyme(s) of interest, an appropriate (generally microbial) expression system and a precise detection system.

Rational design is frequently information-dependent and necessitates both the availability of the structure of the biocatalyst and knowledge about the relevance between sequence, structure and mechanism/function. While structural data of some of the proteins, discovered via X-ray crystallography or Nuclear Magnetic Resonance (NMR) spectroscopy, provided by public data bases, by using molecular modelling and the structure of homologous enzyme it is also possible to predict potential outcomes of mutations such as increase in the selectivity, activity and the stability of corresponding enzyme.

During rational design of an enzyme, frequently site directed mutagenesis is applied. The procedure consists of several steps. Firstly, the enzyme of interest is isolated, later its crystal structure is determined by using X-ray crystallography or NMR. Secondly, the data obtained is examined along with the database of known and putative structural effects of amino acid substitutions on enzyme structure and function. By utilizing all the information collected, possible outcomes of specific amino acid changes are predicted and the ones that will provide the desired improvements in function/structure of enzyme are chosen for further step. The final step involves the construction of the gene that will encode desired amino acid sequence. Predetermined changes in the nucleotide sequence

are created at specific sites of genes and mutant genes are cloned into an expression vector. Afterwards, expression vectors are transformed into appropriate host such as bacteria, yeast, fungi, insect or mammalian cells. The mutant enzymes produced in host organisms are isolated, purified and used for the determination of it structure and properties. Modifications made in the amino acid sequence reveals enzymes with different characteristics and those can be compared with native enzyme characteristics to show improvement(s).[5]

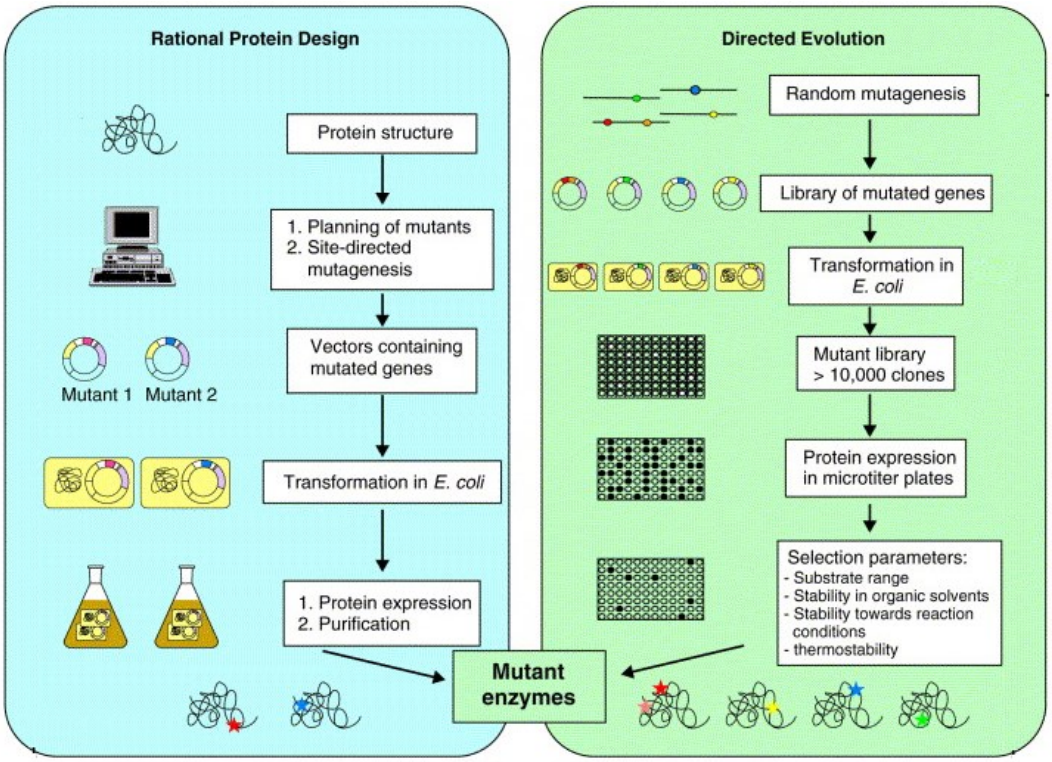


Figure 1. Comparison of rational design and directed evolution.
(Source: Bornscheuer et al., 2001)

Another strategy used in protein engineering is directed evolution, inspired by natural selection, accomplished by either applying random mutagenesis to the target gene encoding the biocatalyst or recombining the fragments of homologous genes. Upon creating library of mutant genes, expression vectors are constituted and transformed into host cells and appropriate variants are selected considering desired characteristics. While library creation, either a specific part of gene or whole gene can be altered randomly depending on applied directed evolution method. In contrast to rational design, random mutagenesis techniques rarely require information about protein structure in advance.[6]

Moreover, application of directed evolution allows discovery of crucial amino acids responsible for particular functions on a protein and provides knowledge for further rational design studies.[7] It leads the creation of large number of variants however only a slight amount satisfies the desired characteristics after screening.

During directed evolution, to provide random mutagenesis, error-prone polymerase chain reaction (epPCR), mutator strains, chemical mutagens, nucleoside analogue incorporation, gene shuffling and degenerate codons can be applied as different methods. Among them, epPCR is the most widely applied method utilizes a thermostable DNA polymerase lack of proof reading provides random mutations on gene of interest entirely. In addition to the type of DNA polymerase used, the concentrations of Mn^{+2}, dNTP and template in the reaction mixture and the number of thermocycles affect the rate of mutations. Although this method has a handicap such as mutation bias, this can be eliminated by using commercial DNA polymerases engineered for balancing transitions and transversions (A→G, T→C and A/G↔C/T respectively).

As an alternative method without PCR employment, chemical mutagens such as nitrous acid, formic acid, hydrazine or ethyl methane sulfonate can be used in the treatment of template DNA to create mutations. Those chemicals affect nucleotide bases and change their hydrogen bonding characteristics eventually increasing their tendency for the formation of noncanonical pairings.[8] By using ethyl methane sulfonate, it is reported that less bias, compared to classical PCR-based methods, was accomplished.[9]

In addition to the methods described above, adding unnatural nucleotide analogs such as 5'-triphosphates of 6-(2-deoxy-beta-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-C] [1,2]oxazin-7-one(dP) and of 8-oxo-2' deoxyguanosine (8-oxodG) leads both transition and transversion mutations as a consequence of displaying different binding characteristics while pairing with canonical nucleotides during PCR reaction.[10]

Mutations can be generated *in vivo* by using mutator strains or UV irradiation on whole cell. Some advantages are provided through elimination of intermediate steps such as template DNA isolation/restriction and ligation. On the other hand, deleterious or malfunctioning mutations in host genome or in essential parts of plasmid DNA such as resistance genes, copy number control sequences or promoter regions are possible undesirable results of this method.

Gene shuffling is another method, used in directed evolution, provides random mutagenesis for whole gene of interest. The procedure begins with the amplification of template DNAs and continue with the restriction of homologue genes by DNase I.

Subsequently, nascent gene fragments in the mixture lack of primers are recombined and amplified by PCR.[11] Recruited genes can be either wild type homolog genes of a gene family or mutated variants of a specific gene. Resulting chimeric gene frequently possesses different sequence fragments from each of the original gene used initially. This method allows to the accumulation of beneficial mutations on a single gene fragment upon iterative application. In addition, compared to standard epPCR, gene shuffling provides more diversity. Since the recombination during PCR is dependent to the reannealing of gene fragments originated from template genes, the divergence between fragments can be a limiting factor thus gene shuffling method requires the consideration of a homology between template genes in advance.

During directed evolution, while whole gene mutagenesis is widely applied, site-targeted mutagenesis can be more practical when the position(s) of mutation(s) will be applied is determined but the amino acids will be employed are unknown. In such cases use of degenerate oligonucleotide primers provide random mutagenesis either for a single site or for multiple sites on a template DNA. The method allows the assessment of all 20 amino acids at a predetermined site or sites. Diverse degenerate codons introduce different amino acid groups with different properties can be used depending on needs (Table 1).[12] The saturation mutagenesis allows the modification of every single position on a gene in an iterative fashion. However, it creates larger libraries and require higher screening effort. On the other hand, site saturation mutagenesis yields smaller libraries as result of confinement of mutagenesis to predetermined positions and provides facilitated screening.

In 2005, Reetz and co-workers brought a new approach to directed evolution by applying Combinatorial Active Site Saturation Test (CAST) thereby substituting amino acid pairs at defined positions in the active site of a lipase. While defining amino acid pairs to be substituted in the active site, secondary structures of proteins aid determination by projecting potential residues located close to each other and point same side which refers to substrate binding site during application (Figure 2).[13] The purpose of the method is to find out amino acid pairs displaying synergistic effect on examined function and the idea arises from if one side chain in the active site has potential to affect catalytic activity then with the association of the closest side chain in same direction, a greater and sophisticated effect can be detected since both side chains are close enough to have interactions both with substrate and between each other. Compared to single site

saturation mutagenesis, CAST provides a more diverse library and higher chance of hit discovery with less screening effort.[13]

Table 1. Degenerate Codons and Corresponding Amino Acids.

Degenerate codon	**Corresponding base sequence**	**Encoded codons**	**Stop codons**	**Encoded amino acids**	**Properties**
NNN	(A, T, G, C) (A, T, G, C) (A, T, G, C)	64	TAA, TAG, TGA	All	Fully randomized codon
NNK	(A, T, G, C) (A, T, G, C) (G, T)	32	TAG	All	All 20 amino acids
NNS	(A, T, G, C) (A, T, G, C) (G, C)	32	TAG	All	All 20 amino acids
NDT	(A, T, G, C) (A, T, G) (T)	12	No	Phe, Leu, Ile, Val, Tyr, His, Asn, Asp, Cys, Arg, Ser, Gly	Mixture of polar, nonpoar, positive and negative charged residues
NTN	(A, T, G, C) (T) (A, T, G, C)	16	No	Met, Phe, Leu, Ile, Val	Nonpolar residues
NAN	(A, T, G, C) (A) (A, T, G, C)	16	TAA, TAG	Tyr, His, Gln, Asn, Lys, Asp, Glu	Charged, larger side chains
NCN	(A, T, G, C) (C) (A, T, G, C)	16	No	Ser, Pro, Thr, Ala	Smaller side chains, polar and nonpolar residues
RST	(A, G) (G, C) (T)	4	No	Ala, Gly, Ser, Thr	Small side chains

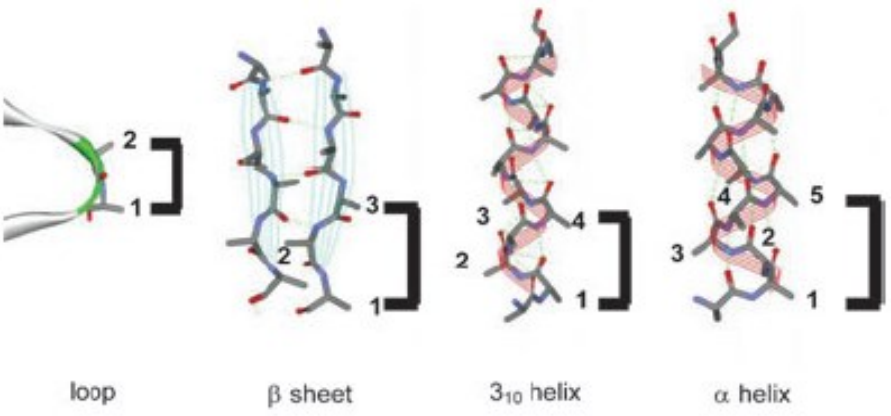


Figure 2. Configuration of amino acids in secondary structure.
(Source: Reetz et al., 2005)

Final step involves the assessment of mutant variants for desired properties. Methods can be categorized as selection-based approach and screening-based approach. In selection-based approach, all mutants created are assayed for a desired biochemical activity. In case of *in vivo* selection, investigated property may provide host cell survivability. For example, β-lactamase expressing cells may be selected by cultivating in β-lactam containing medium. The protein of interest under investigation may be responsible for the synthesis of essential molecules such as amino acids or nucleotides whose deficiency is a limiting factor for cell growth can be considered as a selection parameter. [8] Alternatively, mutant proteins can be introduced on the surface of phages by using phage display system.[14] DNAs of a gene library are fused to the gene of phage encoding phage coat proteins and resulting hybrids are cloned into phagemids and phage particles are transformed with phagemids carrying mutant variants. Phages are reproduced in *E.coli* host cells and a selection based on binding affinity is administered (Figure 3).[15] Binding proteins introduced on the surface of phages are maintained after elution and can be further altered for improved properties. This method allows the representation of diverse libraries with more than 10^{10} unique members as phage pools.

In screening-based methods, biochemical assays are frequently applied via colorimetric or fluorometric measurements to determine substrate turnover rate. Both methods can be applied either *in vitro* or *in vivo*. During *in vitro* application, the colonies carrying protein variants must be grown individually and each colony must be lysed either physically or chemically to obtain enzymes released in the solution. The next step includes the evaluation of enzymes via biochemical assays. Appropriate substrates are

catalyzed by mutant enzymes and specific products are determined via colorimetric or fluorometric measurements either quantitatively or qualitatively.

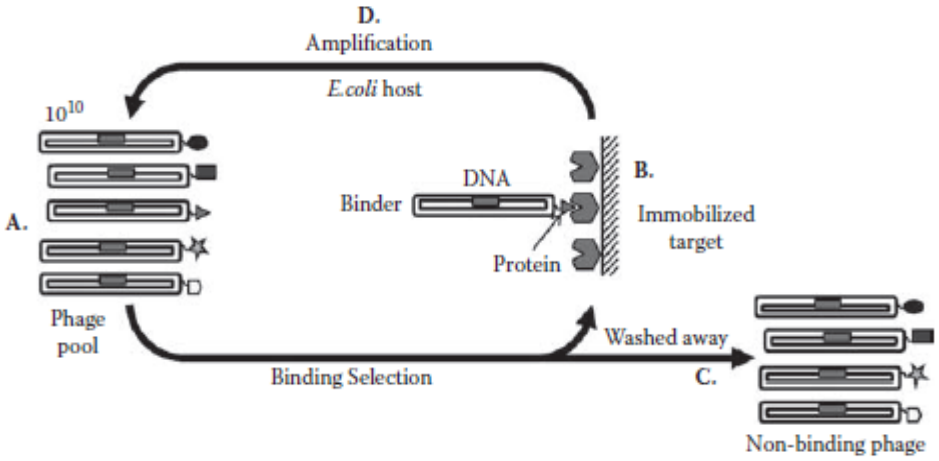


Figure 3. The cycle of phage display for the selection of binding proteins. (A) A library of protein variants fused to phage coat proteins and introduced on phage particles. Each particle contains specifically the gene of protein displayed on. (B) While binding clones are captured by immobilized target, (C) non-binding phages are discarded. (D) Binding clones are further amplified and the proteins are identified through sequence analysis of related DNA. (Source: Sidhu, 2009)

Contrarily in *in vivo* screening-based-methods, the cells are not required to be lysed. However, the cells should be permeable to the substrate but not to the product. Catalysis occurs in the cell and the product formation is examined. Agar plate colony screening, flow cytometric cell sorting and phage display are the methods used in *in vivo* screening. Ligation and transformation efficiency can be limiting factors for *in vivo* screening.[8]

1.3 Cytochrome P450 Proteins (CYPs or P450s)

Cytochrome P450 enzymes are heme *b* containing monooxygenases first characterized by Klingenberg in 1954.[16] CYPs catalyze the site-specific oxidation of non-activated hydrocarbons under mild reaction conditions. In the active site of enzyme, heme group is tied to the apoprotein via a conserved cysteine. The iron ion in the heme is coordinated by four nitrogen atoms of porphyrin and plays a key role during catalysis (Figure 4).

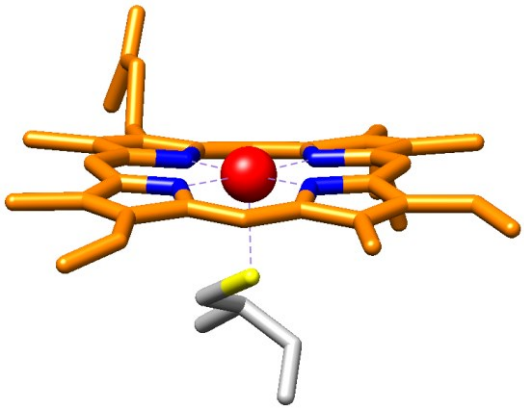


Figure 4. Coordination of iron ion in Protoporphyrin IX. Protoporphyrin IX (orange), nitrogens (blue), iron (red), cysteine (grey) and thiol group (yellow) are shown. The illustration was created on Chimera.

Most P450s catalyze the reductive cleavage of molecular oxygen. Consequently, one oxygen atom is introduced into the substrate, while the second oxygen atom is reduced to water. During catalysis, P450s are reduced via two electrons carried by either nicotinamide adenine dinucleotide or nicotinamide adenine dinucleotide phosphate (NADH or NADPH respectively). However, the electrons are not able to be transferred directly from NAD(P)H to heme center and are mediated by additional proteins called redox partners. Utilization of different redox partner-proteins constitutes several P450-integrated systems (Figure 5).[17] In three-protein systems, a flavoprotein serves as reductase, an iron-sulphur protein serves as electron mediator and a P450 serves as terminal oxidase. P450 and reductase can either be soluble or membrane-bound. In two-protein systems primary protein; cytochrome P450 reductase (CPR) contains both flavin adenine dinucleotide and flavin mononucleotide (FAD and FMN respectively) and transfers electrons accepted from NAD(P)H to P450. Both CPR and P450 present as membrane-bounded. In one-protein systems, CPR and P450 are linked via polypeptide chain and correspond to one-component system with all functional domains. Proteins can be either soluble or membrane-bounded. $P450_{BM-3}$ (CYP102) from Bacillus megaterium represents a good example for one-protein system and is frequently used in P450 studies.[18]

Reactions mediated by P450s initiate via substrate binding (1) followed by 1-electron reduction (2), O_2 binding (3), second 1-electron reduction (4), protonation step (5), homolytic cleavage of O – O bond (6), reaction with the substrate (7, 8) and release of the product (9) respectively (Figure 6, steps 1 to 9).[19] Alternatively, while working with

P450 monooxygenases, addition of H_2O_2 allows direct constitution of hydroperoxo iron from iron in resting state (Fe^{+3} – RH) and the mechanism is known as hydrogen peroxide shunt (Figure 6, step 10). Although H_2O_2-dependent pathway provide less efficiency due to the oxidative inactivation of heme-prosthetic group, elimination of NAD(P)H recruitment introduces cost-effective enzyme systems for industrial applications.[20]

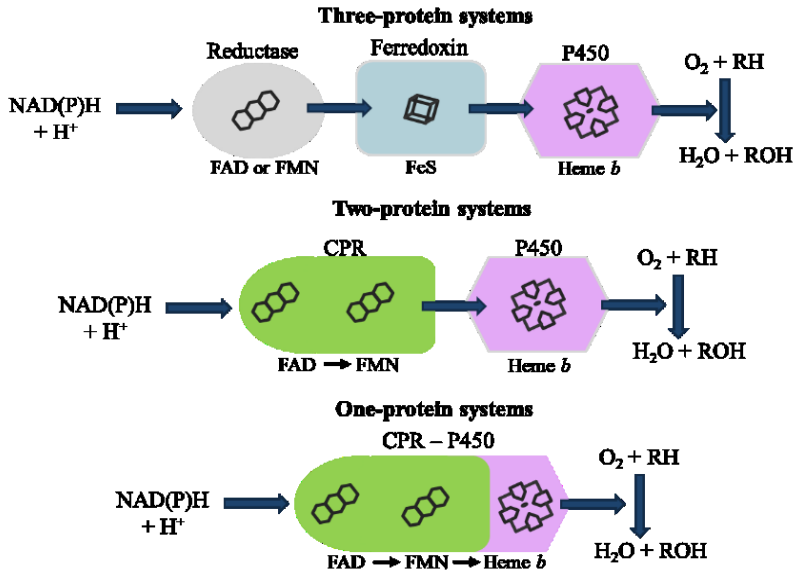


Figure 5. Diverse P450 systems with redox partners.

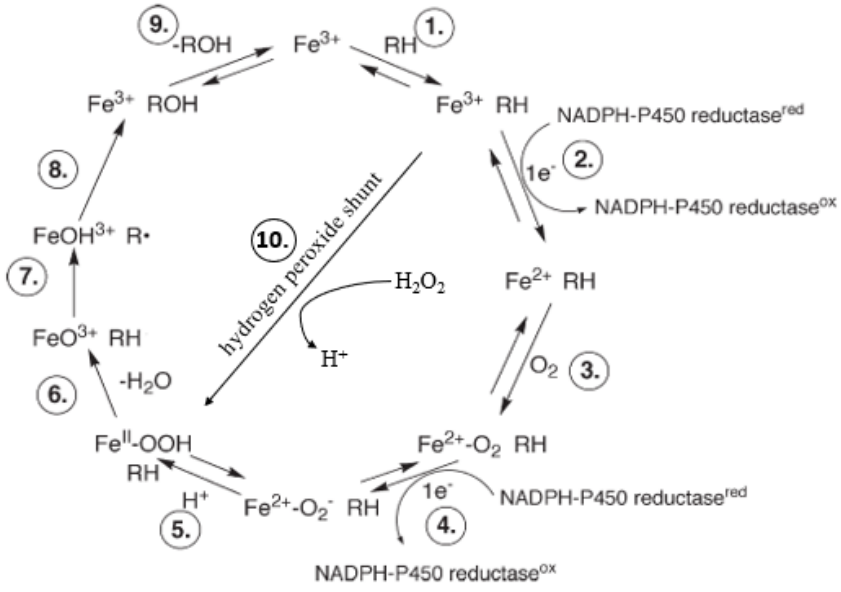


Figure 6. P450 catalytic cycle with NAD(P)H and H_2O_2 utilization are shown via steps from 1 to 9 and step 10 respectively.

P450 enzymes are capable of catalyzing various reactions (Figure 7).[19] Activation of sp^3 hybridized C atoms, epoxidation of C=C double bonds, aromatic hydroxylation, *N*-oxidation, deamination and dehalogenation, as well as *N*-, *O*- and *S*-, dealkylation can be listed as P450 mediated reactions.[17] Under specific circumstances, P450s also catalyze atypical reactions such as C-C and C-O phenol coupling, cleavage of C-C bonds, Baeyer-Villiger oxidation and rearrangement reactions.[17]

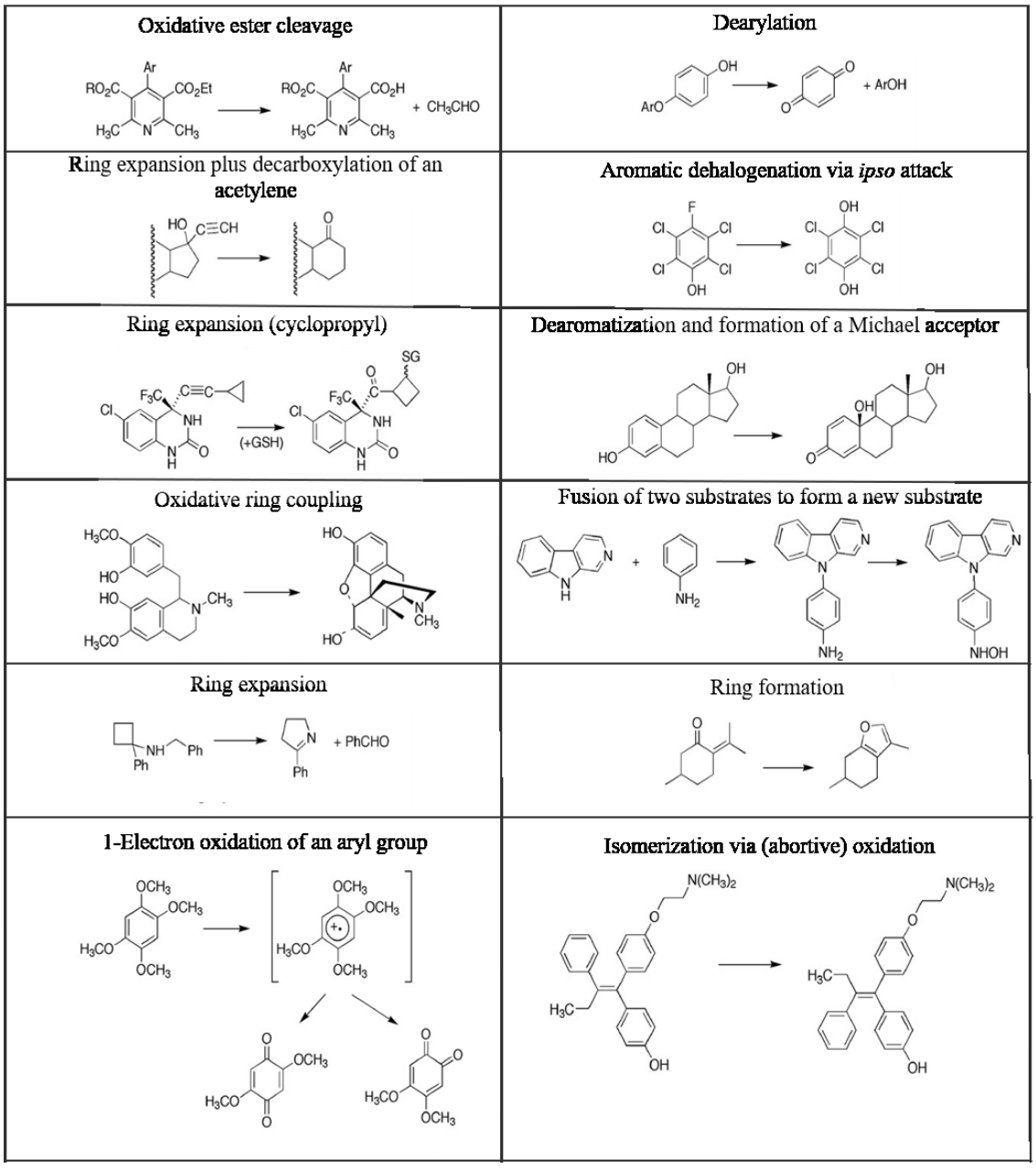


Figure 7. Diverse reactions catalyzed by P450s.

P450s have a diverse substrate range consisting of fatty acids, terpenes, steroids, prostaglandins, polyaromatic and heteroaromatic compounds, drugs, organic solvents, antibiotics, pesticides, carcinogens and toxins. Among the P450s, P450 BM3 from *Bacillus Megaterium,* is a self-sufficient enzyme with reductase domain, frequently employed in P450 studies.[21] So far now various substrates are subjected to P450 BM3 studies and shown in Figure 8.[17]

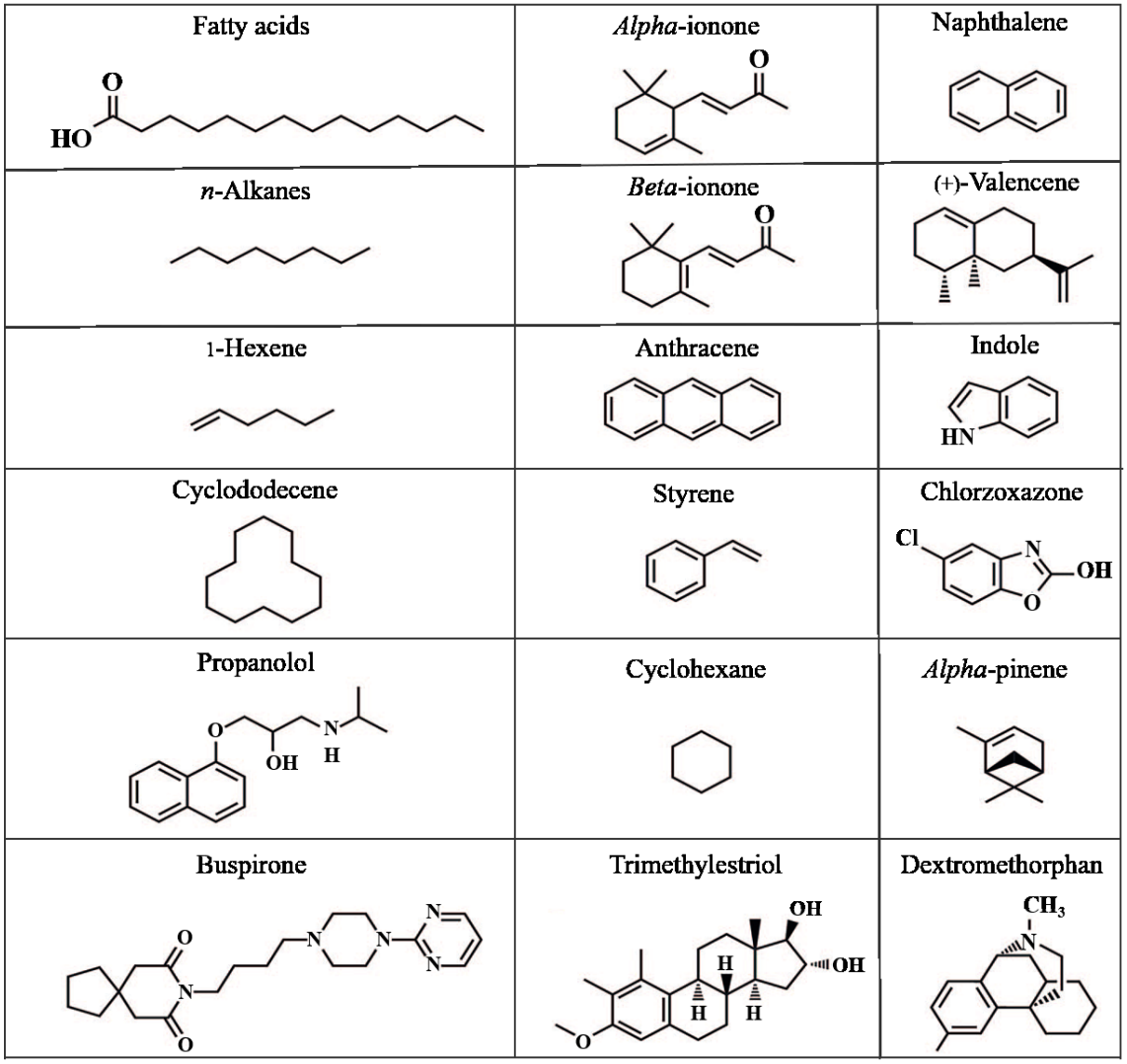


Figure 8. Diverse substrates of P450 BM3.

1.3.1. Cytochrome P450 119 (CYP 119)

CYP119, a member of P450 protein family, identified in *Sulfolobus solfactaricus*, an extreme acidothermophilic, in 1996 by Wright and co-workers.[22] However, in 2008 Rabe and co-workers discovered that original CYP119 originates from *Sulfolobus acidocaldarius* and previous error was due to the contamination of *S.solfactaricus* P1 strain with the *S.acidocaldarius* species. CYP119 is the first known P450 discovered in the domain of archaea and consists of 368 amino acids with a heme in the active site (Figure 9). Exhibiting stability to both high temperature (up to 85 °C) and pressure (up to 2 kbar) are distinguishable features of CYP119 among P450s. The heme cofactor is embedded in the active site of CYP119 ligated to the Cys317, through heme iron.[23] Depending on the substrate, CYP119 can catalyze reactions utilizing either electrons transferred through putidaredoxin/putidaredoxin reductase proteins or H_2O_2 as electron acceptor. While endogenous substrates of CYP119 are unknown, *in vitro* studies revealed that CYP119 is capable of catalyzing epoxidation of styrene, hydroxylation of lauric acid, reduction of nitrite, nitric oxide and nitrous oxide and electrochemical dehalogenation of CCl_4.[24]

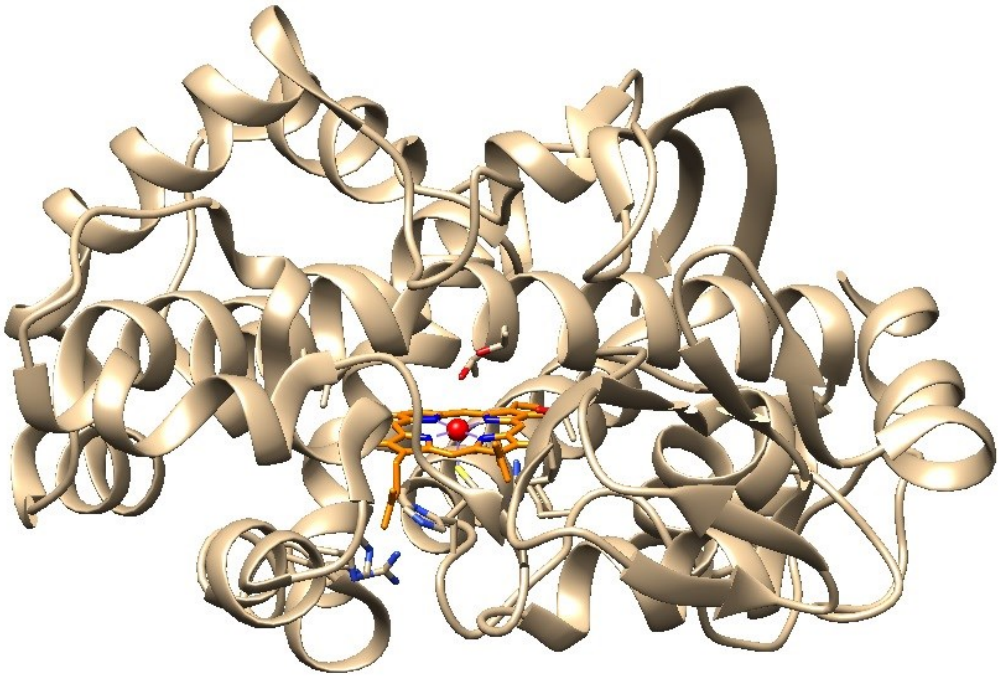


Figure 9. The crystal structure of CYP119. The illustration was created in Chimera.

Besides being water-soluble and satisfactorily thermostable, displaying monooxygenase, oxidoreductase and peroxidase activities attracted attention on CYP119 and led further engineering of the wild type enzyme. First studies related to catalytic properties investigated by Koo and co-workers through mutagenesis of Thr-213 and Thr-214 amino acids positioned close to the heme group in the active site (Figure 10). Significance of Thr-213 in the active site was shown through reactions of CYP119 and the CYP119 mutants with aryldiazene probes. While substitution of Thr-213 with a smaller residue such as serine and alanine increased the rate of formation of the aryl-iron intermediate, substitution with a larger residue caused the opposite. The consequence of Thr-213 mutations is thought due to the steric effects of Thr-213 on aryl-iron intermediate formation. Although the Thr-214 mutations had relatively slight effect on the rate of complex formation, results depicted that the substitutions at this position plays a key role in the determination of spin state of the heme iron due to the participation of Thr-214 in the stabilization of distal water via hydrogen bonds.[25]

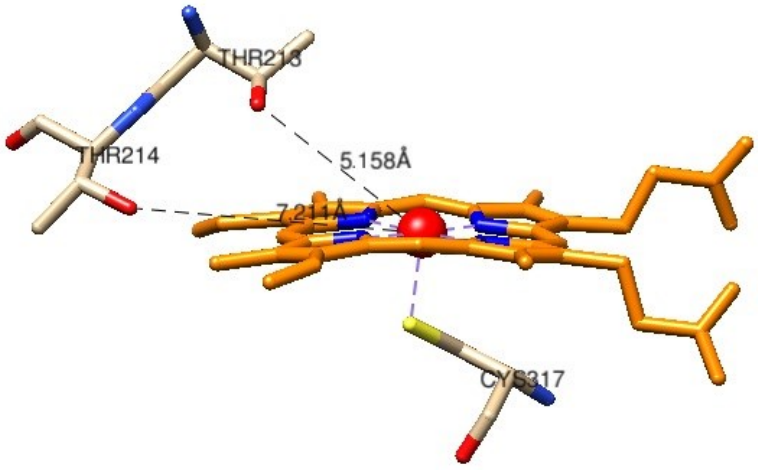


Figure 10. Distance of threonine side chains to heme iron. The illustration was created in Chimera.

Two years later in 2002, another work of Koo and co-workers reported enhanced fatty acid hydroxylation activity via application of T214V and D77R mutations in the wild type CYP119. Unlike previous study focusing on active site residues, Asp-77 is not located in the active site but in the binding site of putidaredoxin, the redox partner of CYP119. D77R mutation increased binding affinity of putidaredoxin to CYP119 and led 5-fold increase in the electron transfer rate from putidaredoxin to mutant CYP119

compared to wild type CYP119. Furthermore, T214V mutation contributed to both improvement of substrate binding and formation of high spin (reducible) state as indicated earlier. Consequently, D77R/T214V double mutant provided 15-fold increase in lauric acid hydroxylation compared to wild type CYP119.[26] Recent efforts to engineer CYP119 included further alterations of amino acids in the active site and uncovered different amino acids as potential candidates for modifications (Figure 11). Dydio and co-workers highlighted Leu-69, Ala-209, Thr-213 and Val-254 as prominent amino acids and showed a quadruple mutant containing C317G, T213G, L69V and V254L mutations led 180 times faster initial turn over frequency in the conversion of diazoester into dihydrobenzofuran compared to wild type CYP119.[27]

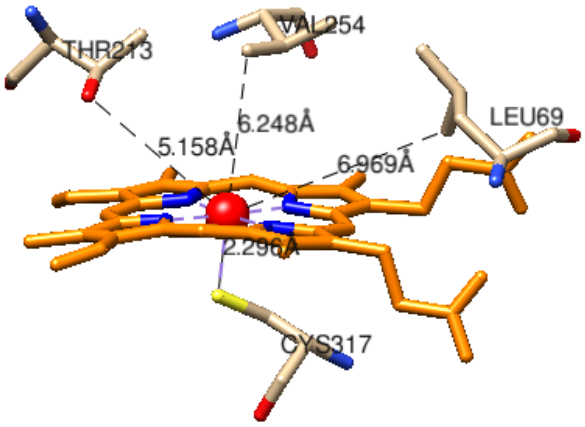


Figure 11. With side chain distances to heme iron, amino acids substituted by Dydio and co-workers are shown. The illustration was created in Chimera.

1.4. Scope of This Study

In this study, a mutant library of CYP119 from *Sulfolobus Acidocaldarius* was created by applying CAST in amino acid positions of 213 and 214. During CAST application, use of NDT containing degenerate primers will provide substitution of T213 and T214 residues with one of the twelve amino acids encoded by NDT. Ultimately, a CYP119 mutant library containing up to 144 different variants will be screened for enhanced peroxidase activity by using hydrogen peroxide as oxygen donor. The expression and isolation of wild type CYP119 will constitute the first step and wild type CYP119 will be tested for its peroxidase activity over Amplex Red with hydrogen peroxide employment. The next step will be the creation of CYP119 mutant library in

PCR and comparison of peroxidase activity of wild type CYP119 and mutant CYP119 variants. Before the screening of the whole library, an effective screening method allowing examination of proteins expressed in small scale will be developed. Selected variants with improved peroxidase activities will be further identified through sequence analysis and help to illuminate the effects of active site residues during hydrogen peroxide mediated oxidation reactions of CYP119. Due to the elimination of NAD(P)H employment, the study has potential to offer cost-effective biocatalysts for industrial applications.

CHAPTER 2

MATERIALS AND METHODS

2.1. Bacterial Transformation and Expression of WT CYP119

The plasmid pET11a+CYP119 which contained WT CYP119 was a gift from Teruyuki Nagamune (Addgene plasmid # 66131).[28] For expression WT CYP119, *E.coli* BL21 DE3 competent cells were transformed with pET11a+CYP119. Likewise, competent *E.coli* DH5α cells were transformed with pET11a+CYP119 for prospective sequence analysis of WT CYP119. In order to be used as control group, *E.coli* BL21 DE3 competent cells were transformed with pET20b plasmid. During transformation heat shock bacterial transformation protocol was applied. Transformants then incubated on LB-agar plates (20 g/L LB and 15 g/L agar) with ampicillin (100 µg/mL) for one night at 37 °C.

For sequence analysis, one colony of *E.coli* DH5α was picked from LB-agar plate and transferred into 5 mL LB broth (20 g/L) with 100 µg/mL ampicillin and incubated at 37 °C for overnight. Growing cells were pelleted by centrifuge and plasmids were isolated with Macherey-Nagel Plasmid DNA purification kit. The concentration of purified plasmids was measured in spectrophotometer and sequence analysis was performed by Biotechnology and Bioengineering Research and Application Center, İzmir Institute of Technology. The results of analysis were interpreted by using GENEIOUS software.

For WT CYP119 expression, one colony of *E.coli* BL21 DE3 was picked from LB-agar plate and transferred into 5 mL LB broth with 100 µg/mL ampicillin and incubated at 37 °C for overnight with 220 RPM shaking. In the morning, 1500 µL of growing cells were transferred into flask containing 500 mL 2xYT medium (16 g/L tryptone, 10g/L yeast extract and 5 g/L NaCl) with 100 µg/mL ampicillin. For four hours cells were grown in shaker incubator at 37 °C with 220 RPM. Upon reaching OD_{600} ~ 0.6 value, protein expression was induced with 1 mM IPTG and cells were incubated for 33 hours at 30 °C with 180 RPM. Before initiation and after termination of induction, 1 mL of samples from cell culture were pelleted and maintained at - 80 °C to be used in SDS-PAGE display.

For control experiments, one colony of *E.coli* BL21 DE3 with pET20b was picked from LB-agar plate and the procedure applied for WT CYP119 expressing cells was followed.

2.2. Isolation of WT CYP119

E.coli BL21 DE3 cells grown in 500 mL LB broth were harvested by centrifuge and pellet were kept in - 80 °C. Frozen pellet was dissolved in lysis buffer (2 mL/g cell) containing 100 mM NaCl, 50 mM potassium phosphate, 0.2 mM phenylmethylsulfonyl fluoride (PMSF) and 1.34 mM benzamidine HCl. Dissolved cells were disrupted by ultrasonicator, powered 69%, for 3 cycles of 15 seconds with 30 seconds resting on ice between each cycle. Disrupted cells were aliquoted into microcentrifuge tubes and subsequently heat-treated at 65 °C for one hour. After heat treatment, cell debris was pelleted by microcentrifuge at room temperature with 12000 RPM for 20 minutes and thermostable WT CYP119 was enriched in the supernatant. Same procedure was also applied to *E.coli* BL21 DE3 cells with pET20b. In order to confirm existence of CYP119, both supernatants from cells with pET11a and pET20b were run on SDS-PAGE. Additionally, absorbance spectrum was created by using spectrophotometer.

2.3. SDS-PAGE Analysis

Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) was utilized to show the success of CYP119 protein expressions by examining both the existence and the density of CYP119 bands at corresponding molecular weight (~ 43 kDa). In order to do this, both cell pellets obtained before and after induction and supernatants of processed CYP119-expressing-cells were separately mixed with SDS loading dye (100 mM Tris-Cl, 4% SDS, 0.2% bromophenol blue and 20% glycerol) and 1 mM dithiothreitol (DTT). After heat treatment at 90 °C for 2 minutes, samples and protein ladder as marker were loaded into the gel and run at 90 Volt. The running was terminated upon observing bands approached to the bottom of the gel. The gel was stained with Coomasie Brilliant Blue dye and destained with a solution containing acetic acid and ethanol for bands to appear.

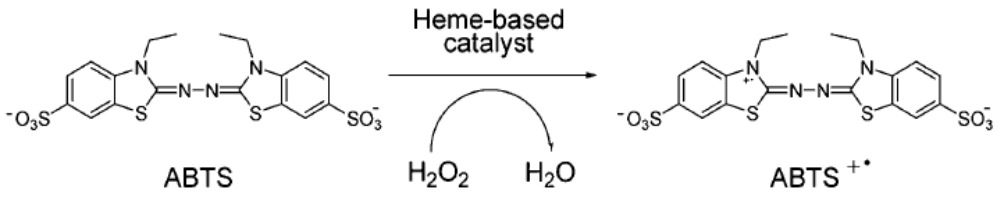
2.4. Detection of Peroxidase Activity of WT CYP119

In order to detect peroxidase activity of wild type CYP119 through hydrogen peroxide shunt, various substrates including ABTS (2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid)), guaiacol and Amplex Red were employed. Before starting experiments with CYP119, the substrates were catalyzed by horse radish peroxidase to confirm the absorbance and fluorescence maxima of the products obtained from literature for guaiacol and Amplex Red oxidation respectively. Whole reactions were investigated through utilizing WT CYP119 enriched in the supernatant. The concentrations of the components were determined by the inspiration of previous studies with same substrates but different peroxidases for ABTS and guaiacol oxidation.[29] For Amplex Red oxidation, a previous study introduced by Rabe and co-workers was followed.[24] While determining CYP119 concentration used, having detectable amount of enzyme in spectrophotometer was considered to follow the change in enzyme absorbance. For control experiments, supernatant of disrupted pET20b-containing-cells were applied the same procedure by employing same volume of supernatant used for CYP119 experiments.

The ABTS oxidation by WT CYP119 was assessed at room temperature in pH 7.4 potassium phosphate buffer. The components of reaction medium and ABTS oxidation are shown in Table 2 and Equation 2.1, respectively.[30] Interpretation was made after 30 minutes of incubation by measuring the absorbance of the product: $ABTS^{\bullet+}$ which has absorbance maxima at 415 nm, 645 nm, 734 nm and 815 nm.[31]

Table 2. The components and concentrations used in ABTS oxidation.

Components	**WT CYP119**	**ABTS**	**H_2O_2**	**Potassium phosphate buffer**
Concentration	4 µM	1 mM	1 mM	50 mM

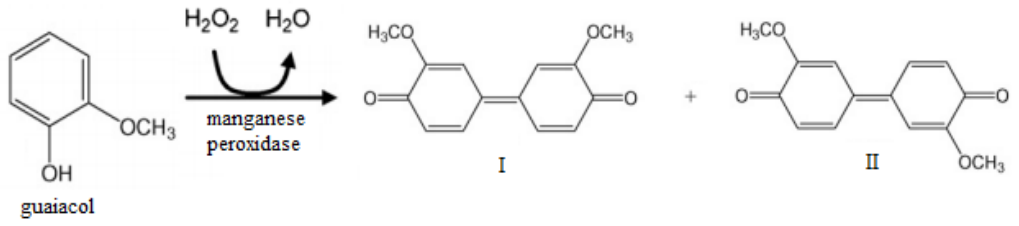


(2.1)

The guaiacol oxidation by WT CYP119 was assessed at room temperature in pH 7.4 potassium phosphate buffer. The components of reaction medium and equation of guaiacol oxidation are shown in Table 3 and Equation 2.2, respectively. Interpretation was made after 2 hours of incubation by measuring the absorbance of the product mixture consisting of (1Z)-3,3'-Dimethoxy-1,1'-bi(2,5-cyclohexadien-1-ylidene)-4,4'-dione (I) and 3,3′-dimethoxy-4,4′-biphenylquinone (II) which have absorbance maxima at 412 nm and 470 nm.[32]

Table 3. The components and concentrations used in guaiacol oxidation.

Components	**WT CYP119**	**Guaiacol**	**H_2O_2**	**Potassium phosphate buffer**
Concentration	4 μM	5 mM	1 mM	50 mM



(2.2)

Further examination of reaction products was assessed by thin layer chromatography (TLC). After 2 hours of incubation at room temperature, the reaction media was run on TLC silica gel 60 F_{254} (Merck). Hexane-ethyl acetate mixture (80:20 v/v) was used as running solvent. The reaction catalyzed by horse radish peroxidase (HRP) was used as positive control group. The components of reaction medium used in TLC were shown in Table 4 with their concentrations. Each lane was loaded equal volume of reaction media and the bands were observed under 312 nm wavelength UV light. Later, silica gel was treated with 5% sulfuric acid - vanillin mixture and heated. After acid and heat treatment the bands were observed under 254 nm and 365 nm wavelengths.

The Amplex Red oxidation by WT CYP119 was assessed at room temperature in pH 7.4 potassium phosphate buffer.[24] The components of reaction medium and equation of Amplex Red oxidation are shown in Table 5 and Equation 2.3, respectively.[33]

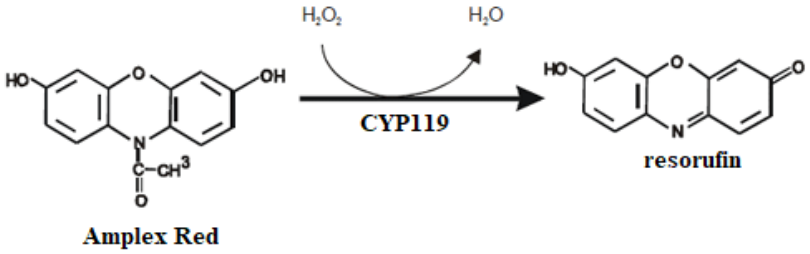
Interpretation was made after 30 minutes of incubation by measuring the fluorescence of the product: resorufin which has excitation and emission maxima of approximately 571 nm and 585 nm, respectively.[34]

Table 4. The components used in TLC.

Components	WT CYP119	Guaiacol	H_2O_2	Potassium phosphate buffer	HRP
Lane 1		5 mM		50 mM	
Lane 2		5 mM	200 µM	50 mM	1 µM
Lane 3	2.8 µM	5 mM	1 mM	50 mM	
Lane 4	4.2 µM	5 mM	1 mM	50 mM	
Lane 5	4.2 µM	5 mM	2 mM	50 mM	
Lane 6	4.2 µM	5 mM	3 mM	50 mM	

Table 5. The components and concentrations used in Amplex Red oxidation.

Components	WT CYP119	Amplex Red	H_2O_2	Buffer
Concentration	2.5 µM	50 µM	1 mM	50 mM sodium phosphate



Amplex Red

(2.3)

2.5. DNA Library Construction of CYP119 by Targeted Random Mutagenesis

The library creation was achieved by applying targeted random mutagenesis by using Q5 Site Directed Mutagenesis Kit (BioLabs) and NDT containing degenerate primers which enable the substitution of one of the 12 amino acids (Table 1) on the amino acid positions Thr213 and Thr214. The degenerate primers were synthesized by Sentegen

(Ankara, Turkey). The sequence of interest was exponentially amplified by Polymerase Chain Reaction (PCR) using the forward and reverse primers shown in Table 6. The PCR conditions are shown in Table 7. In order to examine amplification, PCR product was run on the agarose gel and SafeView-Mini2 (Cleaver Scientific) was used to visualized DNA stained with a fluorophore; SafeView™ Classic (abm).

Table 6. Forward and reverse primer sequences.

Mutant	Forward primer sequence	Reverse primer sequence
T213X-T214X	5’ - GGG TAA TGA GND TND TAC TAA CTT AAT ATC AAA CTC TGT TAT TG - 3’	5’ - GCT ATG AGA AGT AAA ATA ATG TAT C - 3'

X corresponds to the amino acids: Phe, Leu, Ile, Val, Tyr, His, Asn, Asp, Cys, Arg, Ser, Gly encoded by NDT codon.

Table 7. PCR reaction conditions.

Step	Temperature (°C)	Time (seconds)
Initial denaturation	98	45
28 cycles	98	10
	58	30
	72	202
Final extension	72	120
Hold	4	∞

The removal of template DNA and the circularization of nascent DNA was accomplished by the treatment of PCR product with Kinase-Ligase-DpnI (KLD) enzyme mixture for five minutes at room temperature. (Table 8). The plasmids were then transformed into competent DH5α cells with heat shock bacterial transformation protocol. The transformants were incubated on ampicillin-LB-agar plates at 37 °C overnight.

To determine the diversity of CYP119 library, selected DH5α colonies were grown overnight in 5 mL LB medium with 100 µg/mL ampicillin and their plasmids were isolated using Macherey-Nagel kit. Plasmid concentrations were determined after absorbance measurements by using spectrophotometer. The sequence of interest for each plasmid was analyzed by Biotechnology and Bioengineering Research and Application Center, İzmir Institute of Technology and the mutations were examined by using

GENEIOUS software. Sample plasmids with confirmed wild type and mutated sequences were stored in - 20 °C.

Table 8. KLD reaction conditions.

	10 µL Reaction	**Final Concentration**
PCR Product	1 µL	-
2x KLD Reaction Buffer	5 µL	1X
10x KLD Enzyme Mix	1 µL	1X
Nuclease-Free Water	3 µL	-

2.6. High Throughput Screening of CYP119 Mutant Library

2.6.1. Assessment of Screening Methods

In order to develop an effective screening method, *E.coli* BL21 DE3 cells containing WT CYP119, three mutant of CYP119 and pET20b were grown in 5 mL LB medium with 100 µg/mL ampicillin for overnight at 37 °C. 30 µL of each culture were transferred into tubes containing 8 mL 2xYT medium with 100 µg/mL ampicillin. Upon reaching OD_{600} ~ 0.7 cells were induced with 1 mM IPTG and incubated at 30 °C for 32 hours. 8 mL of cell cultures were then divided into three aliquots for sonication (3 mL), Solulyse (Genlantis) treatment (3 mL) and whole-cell activity assays (1 mL). The cells reserved for sonication and Solulyse treatment were centrifuged and the pellets were kept in - 80 °C. The cells reserved for whole-cell activity assays were kept in + 4 °C till catalysis.

To examine proteins harvested by sonicator, cell pellets were dissolved in 300 µL lysis buffer and disrupted by sonicator. Lysed cells were kept in 65 °C for 1 hour and centrifuged. The proteins were collected in the supernatant and used in activity assays with Amplex Red. The reaction media consisted of the followings: 3 µM of enzyme variants, 1 mM H_2O_2, 50 µM Amplex Red and 50 mM sodium phosphate buffer. The supernatant of disrupted pET20b containing cells were used as control group. The reaction occurred in a total volume of 100 µL and interpretation was made by measuring the fluorescence of resorufin at 530 nm excitation and 550 – 700 nm emission values after

30 minutes of incubation. The reactions were run quintuplicate and the graphs were created by calculating the mean values obtained from five wells.

To examine proteins harvested by Solulyse, cell pellets were dissolved in 300 µL Solulyse and centrifuged. Afterwards the proteins were collected in the supernatant and used in activity assays with Amplex Red. The reaction media consisted of the followings: 3 µM of enzyme variants, 1 mM H_2O_2, 50 µM Amplex Red and 50 mM sodium phosphate buffer. The supernatant of disrupted pET20b containing cells were used as control group. The reaction occurred in a total volume of 100 µL and interpretation was made by measuring the fluorescence of resorufin at 530 nm excitation and 550 – 700 nm emission values after 30 minutes of incubation. The reactions were run quintuplicate and the graphs were created by calculating the mean values obtained from five wells.

To investigate whole-cell activity, 100 µL of cell culture was mixed with 50 µL Tris-HCl buffer (100 mM) and 5 µL of polymyxin b sulphate (PB Sulphate 3.6 mM) was used to provide permeability in the cell membrane. [35] After 5 minutes of incubation the cells were added Amplex Red and interpretation was made by measuring the fluorescence of resorufin at 530 nm excitation and 550 – 700 nm emission values after 30 minutes of incubation. The cells containing pET20b plasmid were applied same procedure and used as control group. The reactions were run quintuplicate and the graphs were created by calculating the mean values obtained from five wells.

2.6.2. Rapid Screening of CYP119 Mutant Library

In order to screen mutant CYP119 library, *E.coli* BL21 DE3 cells containing wild type CYP119, pet20b and mutant CYP119 were picked with toothpicks from LB agar plates and grown overnight at 37 °C in 96-well plate containing 140 µL LB and 100 µg/mL ampicillin per well. Twelve hours later 3 µL of each culture was transferred as triplicate into 96-deep-well plates containing 500 µL, 2xYT medium and 100 µg/mL ampicillin per well. About 9 hours later upon observing OD_{578} ~ 0.5, in two replica 96-deep-well plates protein expression was induced with 1 mM IPTG and the cells were incubated for an additional 20 hours at 30 °C with 200 RPM shaking. After completion of induction, plates were centrifuged for 20 minutes with 3500 RPM and cell pellets were maintained at - 80 °C. Later, each well was added 50 µL Solulyse and homogeneous solutions were created by gently pipetting cell pellets. After pipetting, 100 µL of cells

from two replica plates were combined and transferred into micro centrifuge tubes. After transfer, micro centrifuge tubes were shaken for 10 minutes with 220 RPM to provide efficient cell lysis. Upon completion of cell disruption, cell debris and insoluble proteins were removed by centrifugation and soluble proteins were enriched in the supernatant. Later, those supernatants were used in the assessment of peroxidase activity of CYP119 variants. Peroxidase activity of each variant was assessed by mixing 3 μM CYP119, 1 mM H_2O_2 and 10 μM Amplex Red in 50 mM sodium phosphate buffer (pH 7.4) in a total volume of 100 μL. After 30 minutes of incubation at room temperature, the peroxidase activity was examined by measuring the fluorescence of reaction product: resorufin with excitation and emission maxima at 572 nm and 586 nm wavelengths respectively. In total, 66 mutant CYP119 colonies were used for screening and this amount corresponds to the total number of mutant colonies grew in one LB agar plate after transformation. The third replica 96-deep-well plate was used as copy plate and the cells were maintained in - 80 °C after mixing with 500 μL 50% glycerol solution.

After determination of CYP119 variants showing higher peroxidase activity compared to wild type CYP119, colonies in copy plate including those variants were transferred into LB agar plates containing 100 μg/mL ampicillin. Later, one colony from each plate were transferred into tubes containing 5 mL LB media with 100 μg/mL ampicillin. The cells were grown for overnight at 37 °C with 220 RPM shaking and in the morning the plasmids were purified. Plasmids of mutant CYP119 enzymes were sent for sequence analysis (Triogen Biotechnology, İstanbul).

CHAPTER 3

RESULTS AND DISCUSSION

3.1. Expression of Wild Type CYP119

SDS-PAGE analysis showed CYP119 expression in *E.coli* BL21 DE3 cells with 1 mM IPTG induction at 30 °C for 30 hours was successful as shown in Figure 12. Samples were taken before IPTG addition and 30 hours after induction. The cells containing pET20b plasmid were used as negative control group.

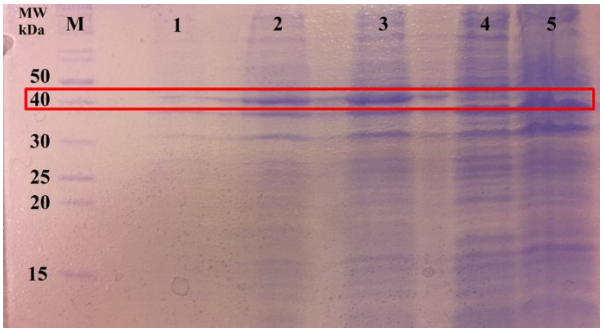


Figure 12. Protein expression bands for WT CYP119 were shown in red frame. M: protein molecular weight marker. Lane 1: wild type CYP119 cell pellet before IPTG (1 mM) induction. Lanes 2 & 3: wild type CYP119 cell pellets, after 30 hours of IPTG (1 mM) induction. Lane 4: negative control, cells transformed with pET20b, before IPTG (1 mM) induction. Lane 5: negative control, cells transformed with pET20b, after 30 hours of IPTG (1 mM) induction.

3.2. Enrichment of Wild Type CYP119

After enrichment of thermostable CYP119 in the supernatant by precipitating irrelevant proteins and cell debris through heat treatment and centrifuge as described in materials and methods 2.2, CYP119 bands were observed more clearly upon applying SDS-PAGE as seen in Figure 13 the supernatant of cells containing pET20b were used as negative control. Further characterization and determination of the concentration of WT CYP119 expressed was accomplished by UV-Vis absorbance spectroscopy as shown

in Figure 14. Existence of CYP119 was confirmed via observing Soret peak at 414 nm and alpha/beta bands at 530 nm and 566 nm respectively. The concentration of CYP119 was calculated by utilizing the extinction coefficient of CYP119 (Ɛ415 = 104 mM^{-1}) and determined to be 33 µM.[23] Higher absorbance at 280 nm compared to 414 nm is thought to be due to the nucleic acids and proteins remained soluble in the supernatant during precipitation which is an expected result of heat-assisted enrichment. Moreover, the bands seen in Figure 13 confirms the existence of other proteins in the supernatant.

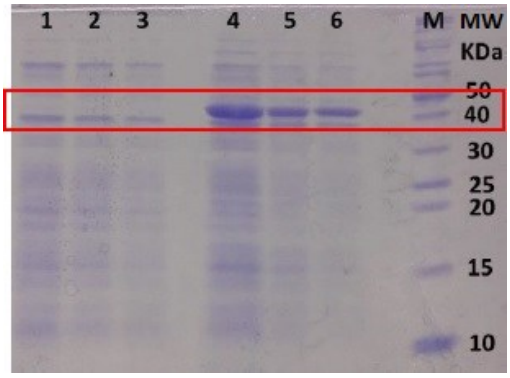


Figure 13. WT CYP119 bands obtained through supernatant loading were shown in red frame. M: protein molecular weight marker. Lane 1: 4 µL supernatant of cells containing pET20b. Lane 2: 2 µL supernatant of cells containing pET20b. Lane 3: 1 µL supernatant of cells containing pET20b. Lane 4: 4 µL supernatant of cells expressing WT CYP119. Lane 5: 2 µL supernatant of cells expressing WT CYP119. Lane 6: 1 µL supernatant of cells expressing WT CYP119.

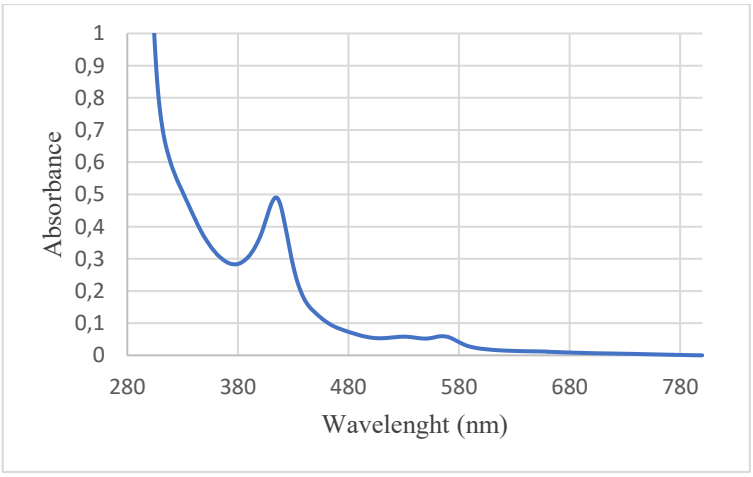


Figure 14. Absorbance spectrum of WT CYP119 obtained from 7x diluted supernatant. Soret band at 414 nm alpha band at 530 nm and beta band at 566 nm are shown.

3.3. Qualification of Peroxidase Activity of Wild Type CYP119

The peroxidase activity of wild type CYP119 was examined via using ABTS, guaiacol and Amplex Red as substrates and H_2O_2 as oxidant. The absorbance maxima of the products of guaiacol oxidation and fluorescence excitation/emission maxima of the product of Amplex Red oxidation were determined by spectrophotometer and fluorometer, respectively, after catalysis of substrates by horse radish peroxidase.

Oxidation of ABTS by CYP119 was achieved at room temperature in 30 minutes by using 1mM ABTS, 1 mM H_2O_2, 4 µM CYP119 and 50 mM potassium phosphate buffer (pH 7.4) in a total volume of 700 µL. The equation of ABTS oxidation and absorbance spectrum of the product ($ABTS^{\bullet+}$ radical cation) obtained at t_0 and t_{30} were shown in Equation 2.1 and Figure 15, respectively.

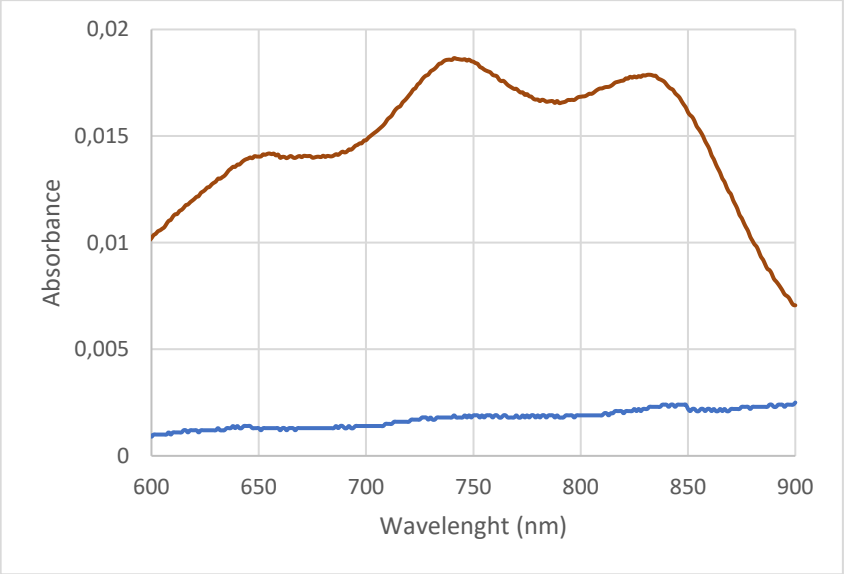


Figure 15. Absorbance of ABTS•+ radical cation produced via oxidation of ABTS by CYP119. The absorbance at t0 and t30 were indicated by blue and brown lines respectively.

The control experiment was accomplished by mixing the supernatant of cells with pET20b, 1 mM H_2O_2, 1 mM ABTS and 50 mM potassium phosphate buffer. The reaction media was incubated at room temperature and at incubation times of t_0 and t_{30} (in minutes) absorbance spectrums were taken. The result was shown in Figure 16. As expressed in the literature, absorbance peaks of $ABTS^{\bullet+}$ were obtained at 645 nm, 734 nm and 815 nm

in the employment of CYP119 and as expected absorbance of product in control group was not observed.[31] However, the magnitude of the peaks obtained via employment of CYP119 was not high enough to develop a sensitive screening method prevented utilization of ABTS in library screening.

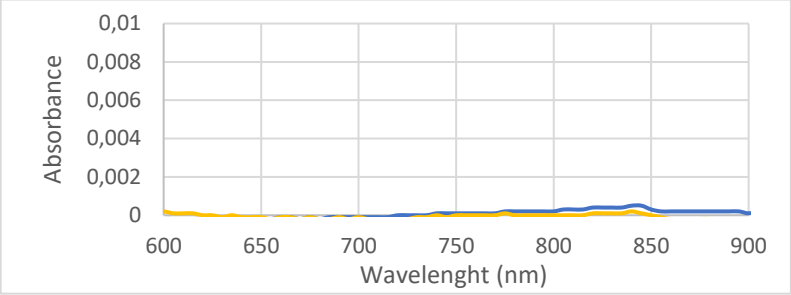


Figure 16. Absorbance spectrum of control group consists of pET20b, ABTS, H_2O_2 and potassium phosphate. The absorbance at t0 and t30 were indicated by blue and yellow lines respectively.

Confirmation of absorbance maxima of guaiacol oxidation product was accomplished by catalyzing guaiacol with horse radish peroxidase (HRP). In the reaction media 1 µM HRP, 5 mM guaiacol, 200 µM H_2O_2 and 50 mM potassium phosphate buffer (pH 7.4) were used and absorbance spectrum was taken after 30 seconds of incubation at room temperature. The absorbance peaks of the products produced via guaiacol oxidation with HRP were observed at 412 nm and 470 nm as expressed in previous studies.[32] Guaiacol oxidation and the absorbance of products were shown in Equation 2.2 and Figure 17, respectively.

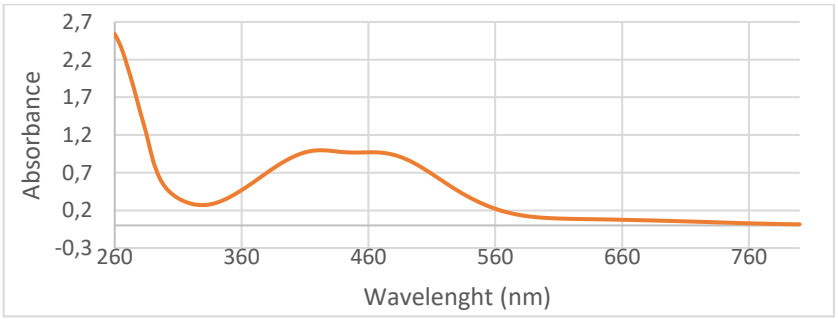


Figure 17. Absorbance spectrum of guaiacol oxidation products with the employment of HRP. The absorbance spectrum after 30 seconds of incubation was shown with orange line.

Oxidation of guaiacol by CYP119 was achieved at room temperature in 2 hours by using 5 mM guaiacol, 1 mM H_2O_2, 4 µM CYP119 and 50 mM potassium phosphate buffer (pH 7.4) in a total volume of 700 µL. The absorbance spectrums were taken at incubation times of t_0 and t_{120} (in minutes) and the result was shown in Figure 18.

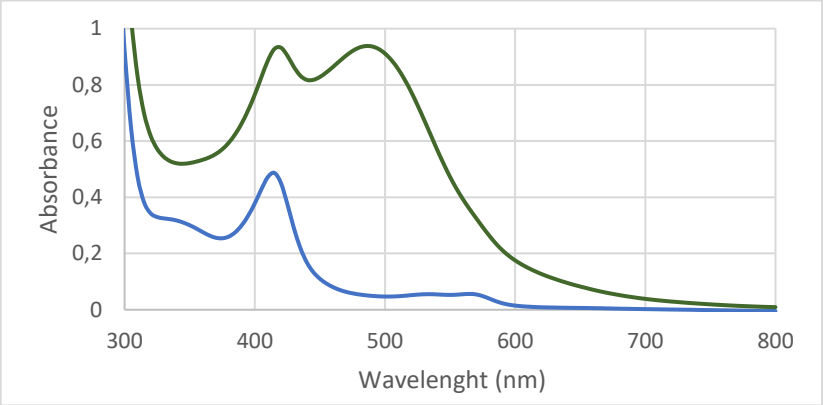


Figure 18. Absorbance spectrum of guaiacol oxidation products obtained with the employment of CYP119. The absorbance at t0 and t120 were indicated by blue and green lines respectively.

The control experiment was accomplished by mixing the supernatant of cells with pET20b, 1 mM H_2O_2, 5 mM guaiacol and 50 mM potassium phosphate buffer. The reaction media was incubated at room temperature and at incubation times of t_0 and t_{120} (in minutes) absorbance spectrums were taken. The result was shown in Figure 19.

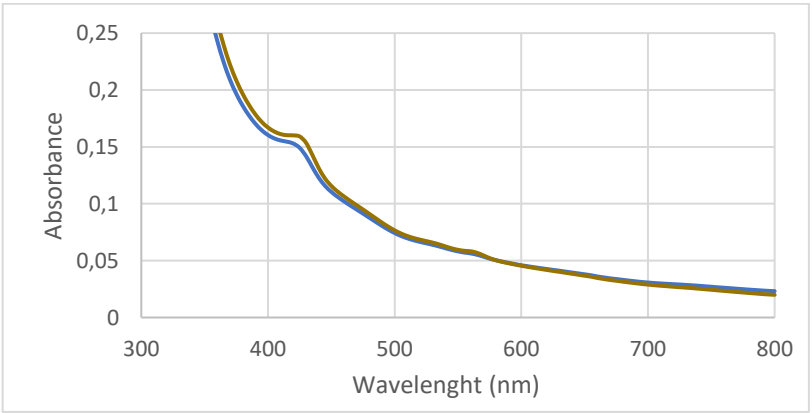


Figure 19. Absorbance spectrum of control group consist of pET20b, guaiacol, H_2O_2 and potassium phosphate. The absorbance at t0 and t120 were indicated by blue and brown lines respectively.

The absorbance peaks observed via guaicol oxidation by CYP119 were obtained at 420 nm and 490 nm. Although the absorbance peaks obtained via guaiacol oxidation with the by CYP119 does not overlap with the absorbance peaks obtained via guaiacol oxidation by HRP, the increase in the absorbance confirms the existence of reaction. Moreover, work of Wang and co-workers showed the absorbance of guaiacol oxidation product at 490 nm while employing guaiacol oxidase (EC 1.11.1.7.) from culture filtrates of soil fungi.[36] During control experiments, as expected no absorbance peaks that belong to guaiacol oxidation product was observed. The magnitudes of absorbance peaks with the employment of CYP119 were satisfying and further studies of guaiacol oxidation by CYP119 continued by repeating the same reaction to confirm the consistency of previous results obtained. Surprisingly, guaiacol oxidation was not successful in every trial with the employment of CYP119. Moreover, it was observed that CYP119 stocks previously succeed in guaiacol catalysis could not catalyze the reaction when employed second time. This contradiction has prevented the use of guaiacol as substrate in the screening of CYP119 library. However, first time oxidation of guaiacol by CYP119 was achieved. Further confirmation of product formation was accomplished by applying TLC. Reaction media with different concentrations of wt CYP119 and H_2O_2 were run on silica gel and horse radish peroxidase was used in the catalysis of guaiacol to determine the band of product. After completion of mobile phase, the bands were observed under 312 nm, 254 nm and 365 nm wavelengths and shown in Figure 20, Figure 21 and Figure 22, respectively. TLC results confirmed the occurrence of reaction and existence of the product. In the lanes (3, 4, 5 and 6) wt CYP119 used, bright bands, thought to be products, were observed in the same linage with a bright tick band that belongs to the lane 2 where horse radish peroxidase used. In the lane 1 only guaiacol was run and no product was observed. In the lane 7 water was run as blank and no component was observed.

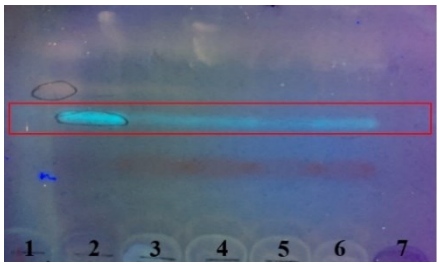


Figure 20. TLC of guaiacol oxidation products at 312 nm wavelength. Lane 1 guaiacol, lane 2 horse radish peroxidase, lanes 3, 4, 5 and 6 wt CYP119 and lane 7 water.

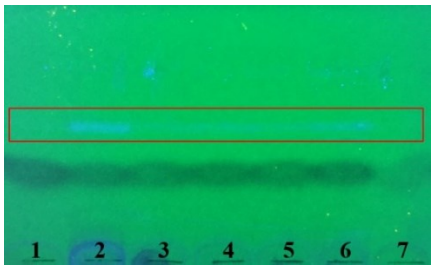


Figure 21. TLC of guaiacol oxidation products at 254 nm wavelength. Lane 1 guaiacol, lane 2 horse radish peroxidase, lanes 3, 4, 5 and 6 wt CYP119 and lane 7 water.

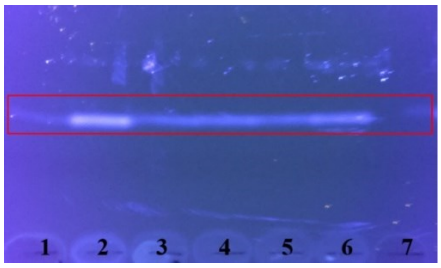


Figure 22. TLC of guaiacol oxidation products at 365 nm wavelength. Lane 1 guaiacol, lane 2 horse radish peroxidase, lanes 3, 4, 5 and 6 wt CYP119 and lane 7 water.

Confirmation of fluorescence excitation/emission maxima of Amplex Red oxidation product; resorufin was accomplished by catalyzing Amplex Red with HRP. In the reaction media 0,05 U/mL HRP, 50 µM Amplex Red, 1 mM H_2O_2 and 50 mM sodium phosphate buffer (pH 7.4) were used and catalysis occurred in 30 seconds at room temperature. The excitation and emission maxima of resorufin were detected at 572 nm and 586 nm respectively as expressed in kit manual.[34] Oxidation of Amplex Red and the fluorescence of resorufin was shown in Equation 2.3 and Figure 23, respectively.

Oxidation of Amplex Red by CYP119 was achieved at room temperature in 30 minutes by using 50 µM Amplex Red, 1 mM H_2O_2, 2.5 µM CYP119 and 50 mM sodium phosphate buffer (pH 7.4) in a total volume of 100 µL. The fluorescence of the resorufin produced by WT CYP119 was measured by fluorometer and confirmed the peroxidase activity of the CYP119. The result was shown in Figure 24.

The control experiment was accomplished by mixing 1 mM H_2O_2, 50 µM Amplex Red and 50 mM sodium phosphate buffer. The reaction media was incubated at room temperature for 30 minutes and the fluorescence was measured by fluorometer. The result was shown in Figure 25.

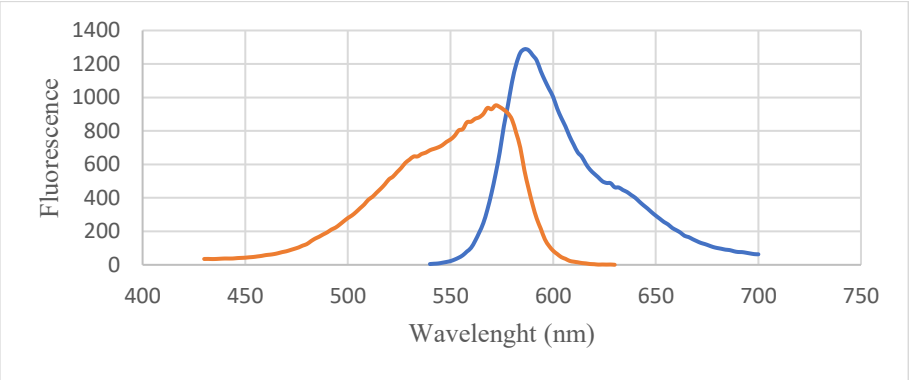


Figure 23. Excitation and emission maxima of resorufin produced by 0,05 U/mL HRP after catalysis of 50 µM Amplex Red in 50 mM sodium phosphate buffer with the aid of 1 mM H_2O_2. The peaks are seen at wavelengths 572 nm and 586 nm for excitation (red line) and emission (blue line), respectively.

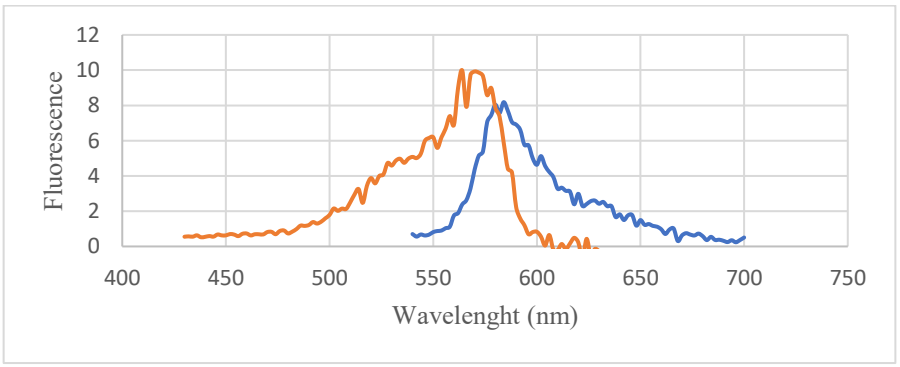


Figure 24. The fluorescence of the resorufin produced by 2.5 µM WT CYP119. 50 µM Amplex Red, 1 mM H_2O_2 and 50 mM sodium phosphate buffer were used in the reaction medium. The peaks are seen at wavelenghts 572 nm and 580 nm for excitation (red line) and emission (blue line) respectively.

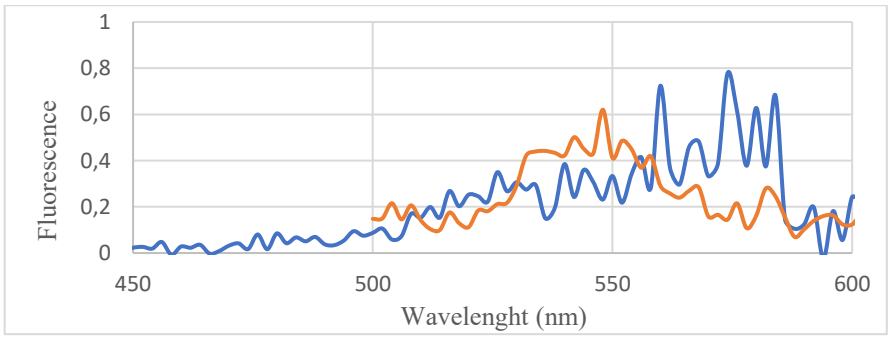


Figure 25. Fluorescence created in reaction medium containing 1 mM H_2O_2, 50 µM Amplex Red and 50 mM sodium phosphate buffer. The orange and blue lines respectively refer to excitation and emission spectrum of the reaction medium.

As expected no significant amount of fluorescence was detected in the control group. Consequently, oxidation of Amplex Red by CYP119 provided fluorescence peaks with satisfying magnitudes and allowed the utilization of Amplex Red in CYP119 library screening.

3.4. DNA Library Construction of CYP119

Mutant CYP119 library was created by applying CAST to the amino acids positioned at T213 and T214. The mutagenesis was accomplished by PCR and the success of amplification was confirmed by running PCR product on agarose gel as shown in Figure 26.

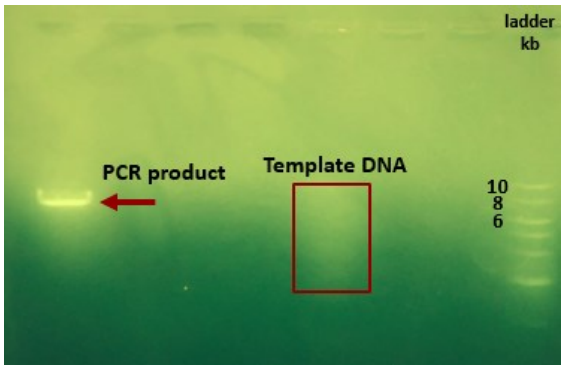


Figure 26. PCR product of amplified CYP119 gene variants shown with the arrow. The area in frame represents the lane template CYP119 run in. (pET11a + CYP119 6748bp).

Mutant gene containing PCR product were transformed into *E.coli* DH5α cells and individual plasmids possessing CYP119 variants were purified from selected *E.coli* colonies. The efficiency of transformation was proved by observation of numerous colonies on agar plate as shown in Figure 27.

In order to determine the frequency of amino acids substituted as well as mutational biases, DNA sequence of plasmids obtained from 15 colonies was analyzed and no significant mutational bias was observed as shown in Figure 28. The sequence analysis also revealed that a diverse library was created as shown in Figure 29 and Figure 30. Ten different amino acids and seven different amino acids out of possible 12 amino

acids with the use of NDT degenerate codons, substituted at the positions 213 and 214, respectively.

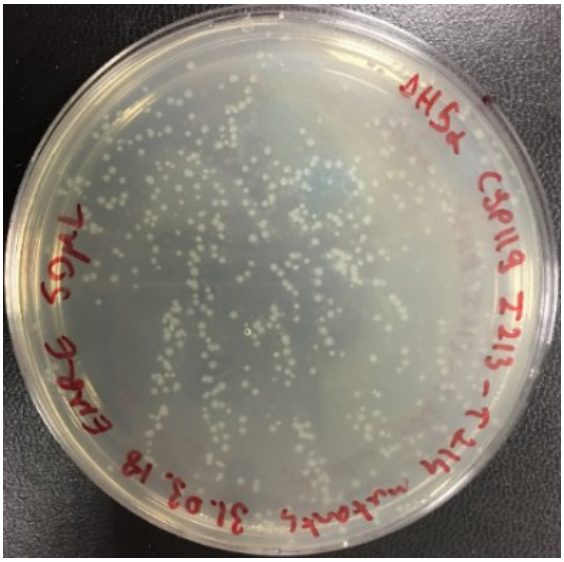


Figure 27. *E.coli* DH5α colonies transformed with pET11a containing CYP119 gene variants.

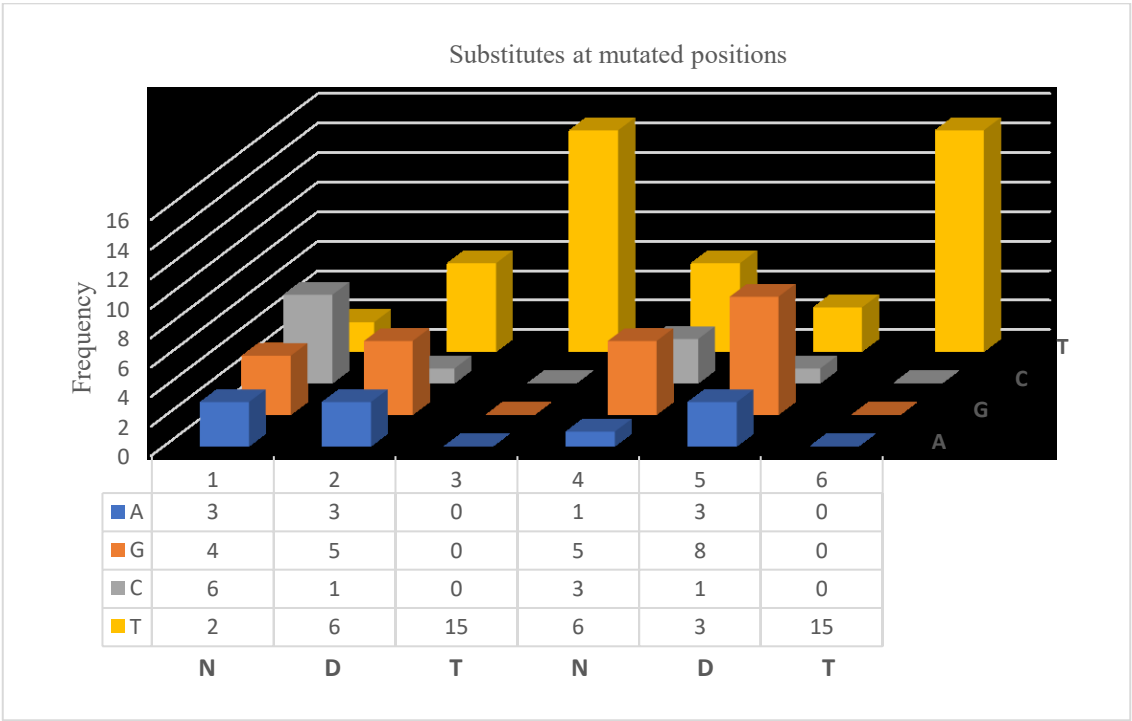


Figure 28. Substituted bases encoding amino acids at positions 213 and 214 with their frequencies at corresponding positions.

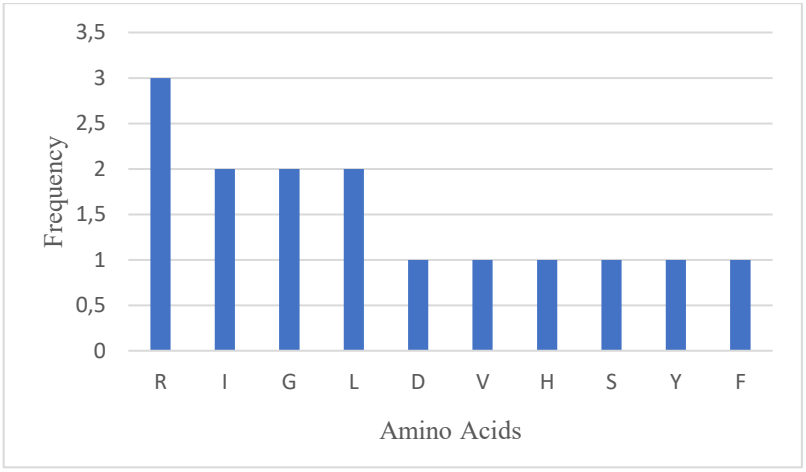


Figure 29. Amino acids employed at the position 213 after mutagenesis.

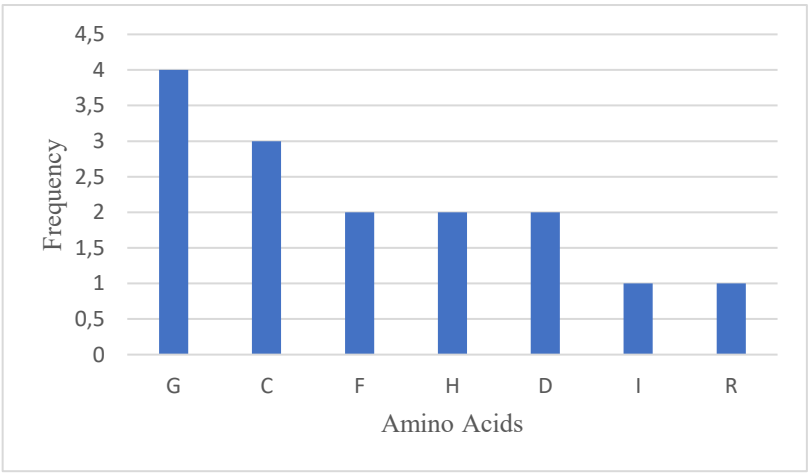


Figure 30. Amino acids employed at the position 214 after mutagenesis.

3.5. High Throughput Screening of CYP119 Mutant Library

3.5.1. Development of the Screening Method

In order to develop an appropriate screening method, peroxidase activity of CYP119 harvested after sonication and heat treatment and Solulyse treatment was investigated with whole-cell catalysts. The reaction media consisted of 10 µM Amplex Red, 1 mM H_2O_2, 50 mM sodium phosphate and 3 µM of enzymes and all the reactions were run quintuplicate. WT CYP119, and 2 CYP119 mutants were expressed in *E.coli*

BL21 DE3 cells. The cells containing pET20b plasmid were used as control group. The fluorescence produced with the employment of sonicated cells were shown in Figure 31. Samples were compared to each other by comparing fluorescence emission signals measured at 584 nm and the results were shown in Figure 32. The fluorescence produced with the employment of Solulyse-treated cells were shown in Figure 33. Samples were compared to each other by comparing fluorescence emission signals measured at 584 nm and the results were shown in Figure 34.

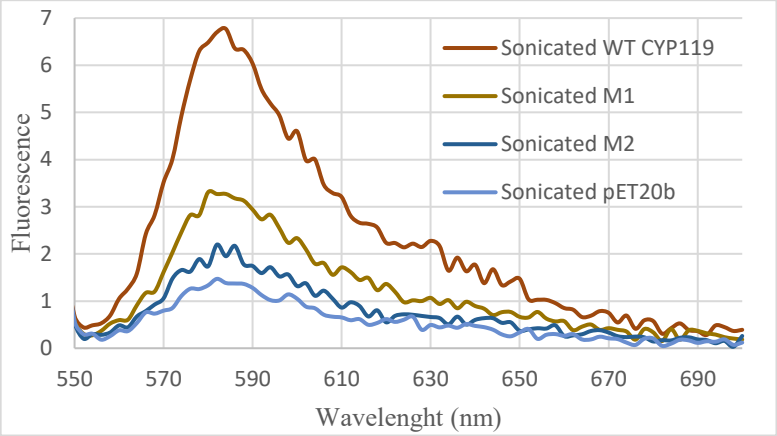


Figure 31. Fluorometric spectrum of sonicated samples in the catalysis of Amplex Red. Fluorescence created by resorufin at 530 nm excitation and 550 – 700 nm emission values. The catalysis was accomplished by CYP119 variants isolated via sonication. The result depicts the mean value of five identical reaction media for each sample.

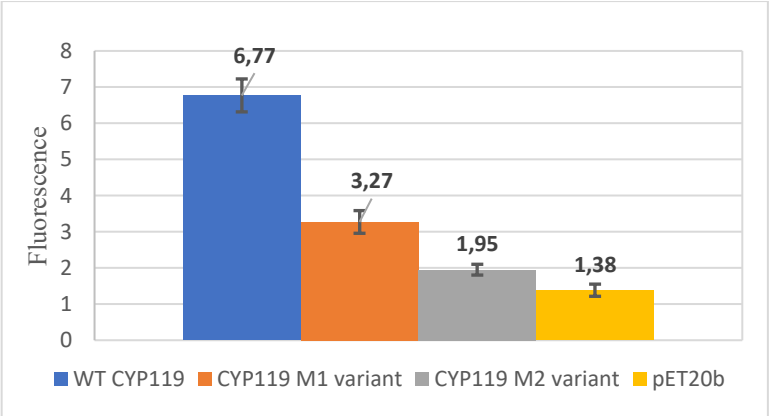


Figure 32. Comparison of fluorescence amounts detected at 584 nm emission produced via employment of sonicated samples in the catalysis of Amplex Red. The result depicts the mean value of five identical reaction media for each sample.

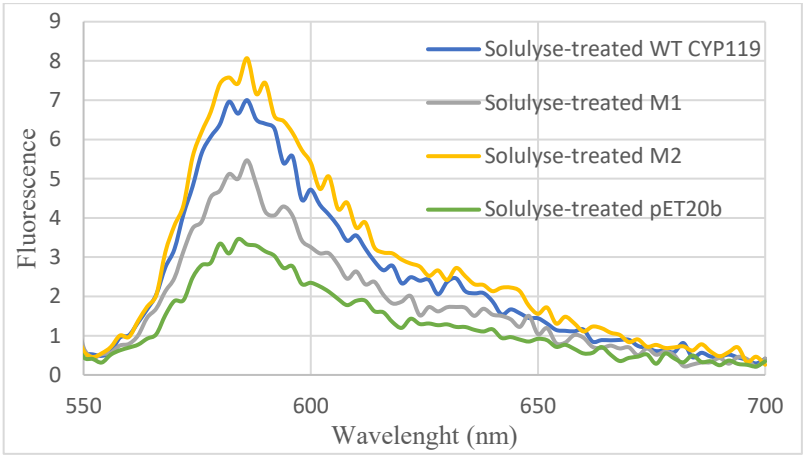


Figure 33. Fluorometric spectrum of Solulyse-treated samples in the catalysis of Amplex Red. Fluorescence created by resorufin was measured at 530 nm excitation and 550 – 700 nm emission values. The catalysis was accomplished by CYP119 variants extracted by Solulyse. The result depicts the mean value of five identical reaction media for each sample.

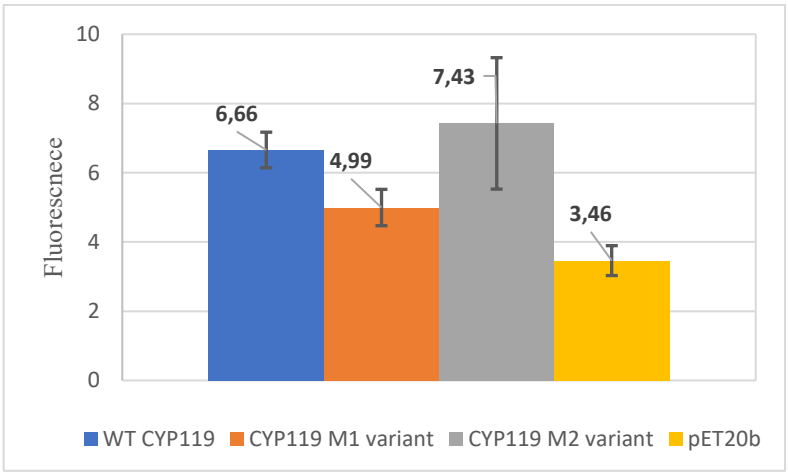


Figure 34. Comparison of fluorescence amounts detected at 584 nm emission produced via employment of Solulyse-treated samples in the catalysis of Amplex Red. The result depicts the mean value of five identical reaction media for each sample.

The difference between fluorescence values obtained through two different extraction methods are thought to be due to the lower efficiency of sonication in intact protein isolation.[37] This difference for individual samples of WT CYP119, M1, M2 and pET20b are shown in Figure 35. Among the fluorescence values obtained with the employment of sonicated and Solulyse-treated cells, only for WT CYP119 both extraction

method provided similar efficiency. Utilization of Solulyse in the extraction of M1 and M2 enzyme variants led higher fluorescence yields during oxidation of Amplex Red. Furthermore, more than 2-fold higher fluorescence detected through employment of Solulyse-treated WT CYP119 containing cells over Solulyse-treated pET20b containing control cells allows the utilization of Solulyse as an efficient protein extraction detergent during high throughput screening.

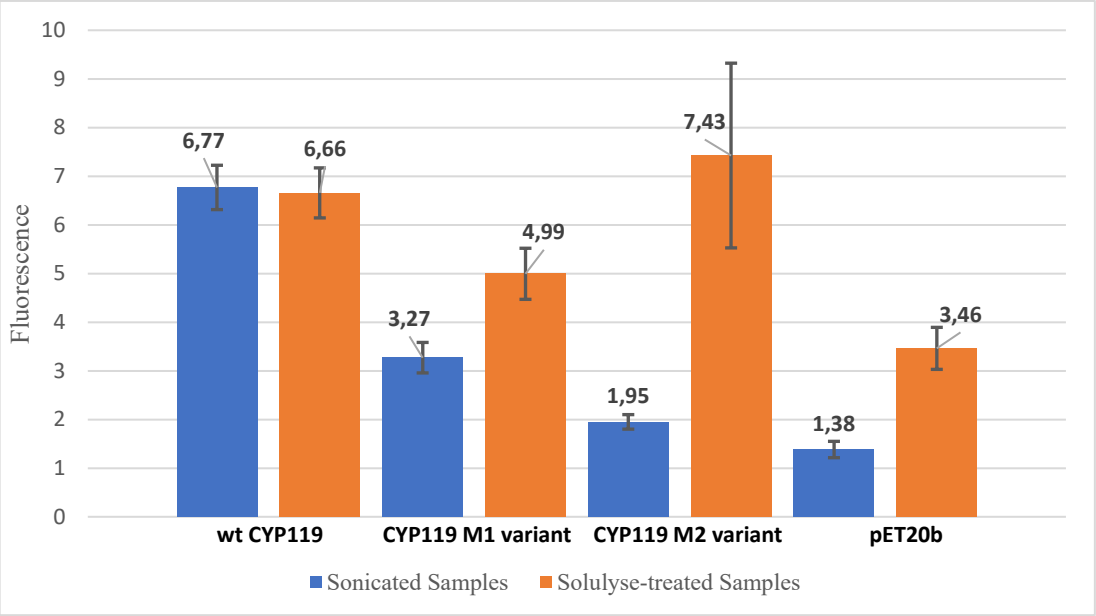


Figure 35. Comparison fluorescence amounts detected at 584 nm with the employment of samples obtained via sonication and Solulyse treatment. The result depicts the mean value of five identical reaction media for each sample.

In order to develop a facile alternative screening method with less steps, peroxidase activity of samples was investigated as whole-cell catalysts. After 32 hours of induction, 1 mL of samples were transferred into micro centrifuge tubes and maintained at +4 °C until reaction. In the reaction medium 50 mM Tris-HCl, 3.6 mM PB Sulphate, 50 µM Amplex Red and 1 mM H_2O_2 used with employment of 100 µL of cells. Fluorescence created by resorufin was measured at 500 nm excitation and 540 – 700 nm emission values. Although highest fluorescence value was obtained from CYP119-M2 variant, other samples could not display reliable results as shown in Figure 36.

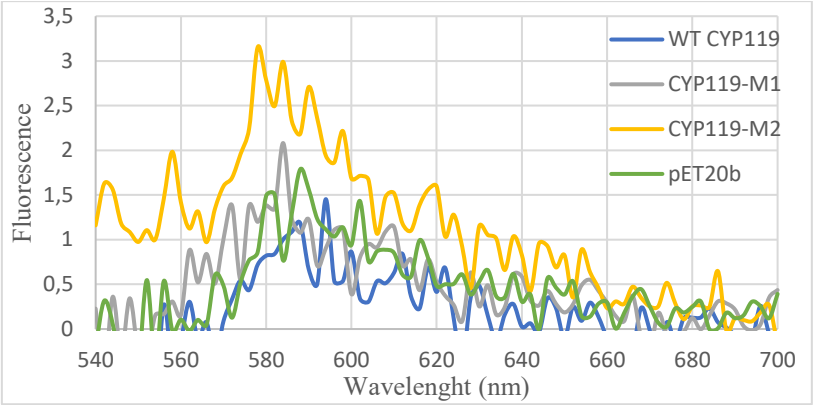


Figure 36. Comparison of the fluorescence detected with the employment of WT CYP119, CYP119-M1, CYP119-M2 and pET20b containing cells. The result depicts the mean value of five identical reaction media for each sample.

3.5.2. Rapid Screening of CYP119 Mutant Library

After observing peroxidase activity from CYP119 enriched by Solulyse treatment, 66 mutant CYP119 variants were screened for increased peroxidase activity. The fluorescence emission of resorufin created by improved variant was measured at 584 nm wavelenght and compared to fluorescence of resorufin created by wild type CYP119 as shown in Figure 37. Later, sequence analysis of improved mutant was determined. During activity assays supernatant of cells containing pET20b was used as negative control group. A scheme representing rapid screening method in steps was shown in Figure 38. Among 66 mutant colonies, a variant with Thr213Arg and Thr214Ile mutations showed 1.32-fold higher activity compared to wild type CYP119. Further confirmation of existence of improved variant was achieved by expressing wt CYP119 and CYP119 Thr213Arg – Thr214Ile mutant in 10 mL 2xYT broth. After protein extraction with Solulyse, Amplex Red was catalyzed with both wt CYP119 and CYP119 Thr213Arg – Thr214Ile mutant with the aid of H_2O_2. The reactions were catalyzed as duplicate in two different wells for each sample. Fluorescence emission values at 584 nm detected after 30 minutes later were shown in Figure 39. In CYP119 Thr213Arg – Thr214Ile double mutant, amino acid configurations were shown in Figure 40. Contribution of Arg213 to activity might be arise from nitrogen groups on Arg213 residue, which may be facilitating either formation of compound I by making hydrogen bonds with oxygen or release of oxygen from heme iron by attracting electrons on oxygen through positively charged

groups. Contribution of Ile214 to activity was an expected result and thought to be due to destabilizing distal water which was previously stabilized by hydrogen bonds where hydroxyl group of Thr214 included. Since the substitution of Arg in the position of Thr213 is a mutation first time reported for CYP119, the effects of using charged groups in the position 213 requires further analysis in order to determine the role of amino acids at this position.

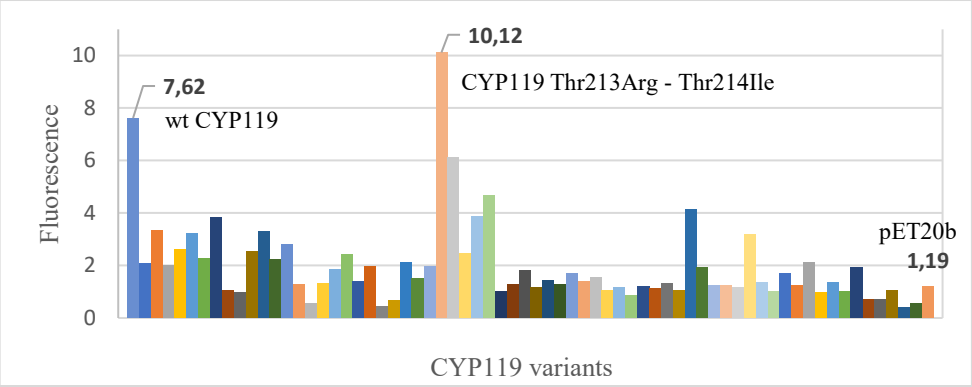


Figure 37. Comparison of the fluorescence detected with the employment of WT CYP119, CYP119 variants and pET20b samples obtained through Solulyse treatment in 96-well plates. The graph displays fluorescence amounts obtained from reactions occurred in a single well for each sample.

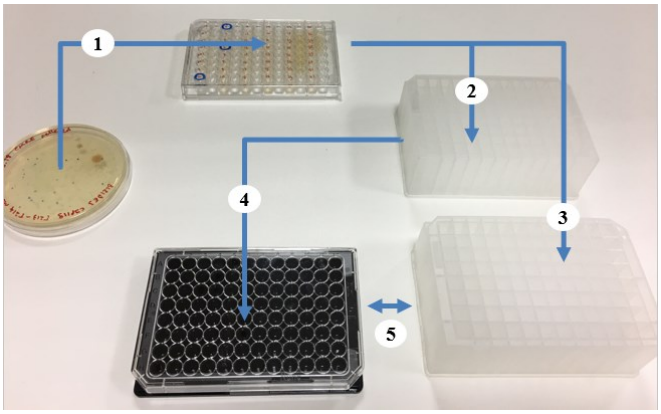


Figure 38. Steps of high-throughput screening method. Step 1: mutant colonies grown on LB agar plates are transferred into a 96-well plate containing LB broth. Step 2 and 3: cultures grown in 96-well plate are transferred into 96-deep-well plates containing 2xYT broth. Step 4: proteins in one of the 96-deep-well plate are extracted via Solulyse treatment and used in the catalysis of Amplex Red. Step 5: Enzymes displaying desired activity are determined and cultures possessing those enzymes are further cultured from the second 96-deep-well plate corresponds to copy plate.

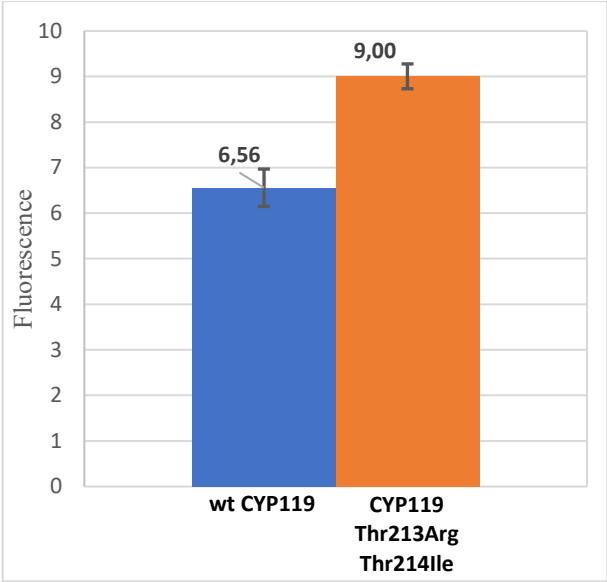


Figure 39. Comparison of the fluorescence detected with the employment of WT CYP119 and CYP119 Thr213Arg – Thr214Ile mutant. The graph displays fluorescence mean values obtained from reactions occurred in two different wells for each sample.

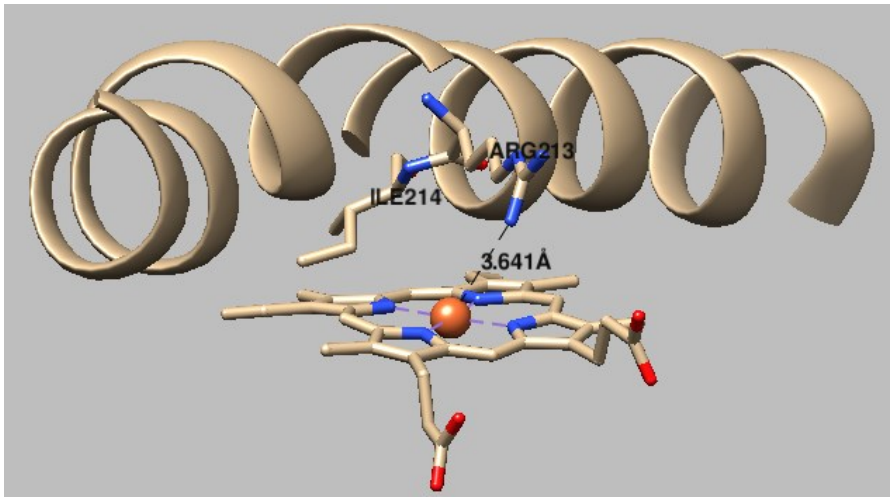


Figure 40. Configuration of amino acids at positions 213 and 214 after substitution. The illustration was created in Chimera.

CHAPTER 4

CONCLUSION

Improvements in recombinant DNA technology allowed massive protein expression and pave the way for employment of biocatalysts in industrial applications instead of chemical catalysts. The advantages such as having substrate specificity, enantioselectivity, chemoselectivity and stability to both temperature and pH provided by enzymes made them valuable tools. Moreover, it is possible to obtain desired properties or further improve through protein engineering approaches. In this study, a mutant library of a thermophilic enzyme; CYP119 was created via directed evolution and an effective screening method was developed to determine variants displaying higher peroxidase activity while employing hydrogen peroxide as oxidant. Three different substrates including ABTS, guaiacol and Amplex Red were used to evaluate peroxidase activity of CYP119. All substrates were catalyzed by CYP119 and Amplex Red was found suitable for the screening of mutant library due to higher fluorescence yield compared to control group. The method developed allows the rapid screening of soluble enzymes in 96-well plates. In the screening of 66 CYP119 variants, CYP119 Thr213Arg – Thr214Ile double mutant showed 1.32-fold higher peroxidase activity compared to wild type CYP119.

Consequently, in this study first time combinatorial active site saturation test was applied to CYP119 and an improved screening method was developed to screen enzymes enriched in the supernatant rather than purified. In addition, first time oxidation of guaiacol by CYP119 was reported here.

REFERENCES

(1) Turanli-Yildiz, B.; Alkim, C.; Petek Cakar, Z. Protein Engineering Methods and Applications. In *Protein Engineering*; Kaumaya, P., Ed.; InTech, **2012**; p 344.

(2) Ulmer, K. M. Protein Engineering. *Science* **1983**, *219* (4585), 666–671.

(3) Basheer, S. M.; Chellappan, S. Enzyme Engineering. In *Bioresources and Bioprocess in Biotechnology*; Springer Singapore: Singapore, **2017**; pp 151–168.

(4) Kazlauskas, R. J.; Bornscheuer, U. T. Finding Better Protein Engineering Strategies. *Nat. Chem. Biol.* **2009**, *5* (8), 526–529.

(5) Bornscheuer, U. T.; Pohl, M. Improved Biocatalysts by Directed Evolution and Rational Protein Design. *Curr. Opin. Chem. Biol.* **2001**, *5* (2), 137–143.

(6) Cobb, R. E.; Chao, R.; Zhao, H. Directed Evolution: Past, Present and Future. *AIChE J.* **2013**, *59* (5), 1432–1440.

(7) Kaur, J.; Sharma, R. Directed Evolution: An Approach to Engineer Enzymes. *Crit. Rev. Biotechnol.* **2006**, *26* (3), 165–199.

(8) Nannemann, D. P.; Birmingham, W. R.; Scism, R. A.; Bachmann, B. O. Assessing Directed Evolution Methods for the Generation of Biosynthetic Enzymes with Potential in Drug Biosynthesis. *Future Med. Chem.* **2011**, *3* (7), 809–819.

(9) Lai, Y.-P.; Huang, J.; Wang, L.-F.; Li, J.; Wu, Z.-R. A New Approach to Random Mutagenesis in Vitro. *Biotechnol. Bioeng.* **2004**, *86* (6), 622–627.

(10) Zaccolo, M.; Williams, D. M.; Brown, D. M.; Gherardi, E. An Approach to Random Mutagenesis of DNA Using Mixtures of Triphosphate Derivatives of Nucleoside Analogues. *J. Mol. Biol.* **1996**, *255* (4), 589–603.

(11) Stemmer, W. P. DNA Shuffling by Random Fragmentation and Reassembly: In Vitro Recombination for Molecular Evolution. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (22), 10747–10751.

(12) Currin, A.; Swainston, N.; Day, P. J.; Kell, D. B. Synthetic Biology for the Directed Evolution of Protein Biocatalysts: Navigating Sequence Space Intelligently. *Chem. Soc. Rev.* **2015**, *44* (5), 1172–1239.

(13) Reetz, M. T.; Bocola, M.; Carballeira, J. D.; Zha, D.; Vogel, A. Expanding the Range of Substrate Acceptance of Enzymes: Combinatorial Active-Site Saturation Test. *Angew. Chemie Int. Ed.* **2005**, *44* (27), 4192–4196.

(14) Smith, G. P. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science* **1985**, *228* (4705), 1315–1317.

(15) Sidhu, S.; Ernst, A. Phage Display Systems For Protein Engineering. In *Protein Engineering and Design*; CRC Press, **2009**.

(16) Klingenberg, M. The Dragging Emergence of the P450 Cytochrome. *Arch. Biochem. Biophys.* **2003**, *412* (1), 2.

(17) Urlacher, V. B.; Girhard, M. Cytochrome P450 Monooxygenases: An Update on Perspectives for Synthetic Application. *Trends Biotechnol.* **2012**, *30* (1), 26–36.

(18) Degtyarenko, K. N. Structural Domains of P450-Containing Monooxygenase Systems. *Protein Eng.* **1995**, *8* (8), 737–747.

(19) Isin, E. M.; Guengerich, F. P. Complex Reactions Catalyzed by Cytochrome P450 Enzymes. *Biochim. Biophys. Acta - Gen. Subj.* **2007**, *1770* (3), 314–329.

(20) Holtmann, D.; Fraaije, M. W.; Arends, I. W. C. E.; Opperman, D. J.; Hollmann, F. The Taming of Oxygen: Biocatalytic Oxyfunctionalisations. *Chem. Commun.* **2014**, *50* (87), 13180–13200.

(21) Sevrioukova, I. F.; Li, H.; Zhang, H.; Peterson, J. A.; Poulos, T. L. Structure of a Cytochrome P450-Redox Partner Electron-Transfer Complex. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (5), 1863–1868.

(22) Wright, R. L.; Harris, K.; Solow, B.; White, R. H.; Kennelly, P. J. Cloning of a Potential Cytochrome P450 from the Archaeon Sulfolobus Solfataricus. *FEBS Lett.* **1996**, *384* (3), 235–239.

(23) McLean, M. A.; Maves, S. A.; Weiss, K. E.; Krepich, S.; Sligar, S. G. Characterization of a Cytochrome P450 from the Acidothermophilic ArchaeaSulfolobus Solfataricus. *Biochem. Biophys. Res. Commun.* **1998**, *252* (1), 166–172.

(24) Rabe, K. S.; Kiko, K.; Niemeyer, C. M. Characterization of the Peroxidase Activity of CYP119, a Thermostable P450 FromSulfolobus Acidocaldarius. *ChemBioChem* **2008**, *9* (3), 420–425.

(25) Koo, L. S.; Tschirret-Guth, R. A.; Straub, W. E.; Moënne-Loccoz, P.; Loehr, T. M.; Ortiz de Montellano, P. R. The Active Site of the Thermophilic CYP119 from Sulfolobus Solfataricus. *J. Biol. Chem.* **2000**, *275* (19), 14112–14123.

(26) Koo, L. S.; Immoos, C. E.; Cohen, M. S.; Farmer, P. J.; Ortiz de Montellano, P. R. Enhanced Electron Transfer and Lauric Acid Hydroxylation by Site-Directed Mutagenesis of CYP119. *J. Am. Chem. Soc.* **2002**, *124* (20), 5684–5691.

(27) Dydio, P.; Key, H. M.; Nazarenko, A.; Rha, J. Y.-E.; Seyedkazemi, V.; Clark, D. S.; Hartwig, J. F. An Artificial Metalloenzyme with the Kinetics of Native Enzymes. *Science (80-.).* **2016**, *354* (6308), 102–106.

(28) Suzuki, R.; Hirakawa, H.; Nagamune, T. Electron Donation to an Archaeal

Cytochrome P450 Is Enhanced by PCNA-Mediated Selective Complex Formation with Foreign Redox Proteins. *Biotechnol. J.* **2014**, *9* (12), 1573–1581.

(29) Doerge, D. R.; Divi, R. L.; Churchwell, M. I. Identification of the Colored Guaiacol Oxidation Product Produced by Peroxidases. *Anal. Biochem.* **1997**, *250* (1), 10–17.

(30) Gao, W.-Y.; Chrzanowski, M.; Ma, S. Metal–metalloporphyrin Frameworks: A Resurging Class of Functional Materials. *Chem. Soc. Rev.* **2014**, *43* (16), 5841–5866.

(31) Re, R.; Pellegrini, N.; Proteggente, A.; Pannala, A.; Yang, M.; Rice-Evans, C. Antioxidant Activity Applying an Improved ABTS Radical Cation Decolorization Assay. *Free Radic. Biol. Med.* **1999**, *26* (9–10), 1231–1237.

(32) Hwang, S.; Lee, C.-H.; Ahn, I.-S. Product Identification of Guaiacol Oxidation Catalyzed by Manganese Peroxidase. *J. Ind. Eng. Chem.* **2008**, *14* (4), 487–492.

(33) Rabe, K. S.; Erkelenz, M.; Kiko, K.; Niemeyer, C. M. Peroxidase Activity of Bacterial Cytochrome P450 Enzymes: Modulation by Fatty Acids and Organic Solvents. *Biotechnol. J.* **2010**, *5* (8), 891–899.

(34) Amplex Red Hydrogen Peroxide/Peroxidase Assay Kit - Thermo Fisher Scientific https://www.thermofisher.com/order/catalog/product/A22188 (accessed Jun 3, **2018**).

(35) Schwaneberg, U.; Otey, C.; Cirino, P. C.; Farinas, E.; Arnold, F. H. Cost-Effective Whole-Cell Assay for Laboratory Evolution of Hydroxylases in *Escherichia Coli*. *J. Biomol. Screen.* **2001**, *6* (2), 111–117.

(36) Wang, W.; Li, Y.; Wang, H.; Zu, Y. Differences in the Activities of Eight Enzymes from Ten Soil Fungi and Their Possible Influences on the Surface Structure, Functional Groups, and Element Composition of Soil Colloids. *PLoS One* **2014**, *9* (11), e111740.

(37) Listwan, P.; Pédelacq, J.-D.; Lockard, M.; Bell, C.; Terwilliger, T. C.; Waldo, G. S. The Optimization of in Vitro High-Throughput Chemical Lysis of *Escherichia Coli*. Application to ACP Domain of the Polyketide Synthase ppsC from *Mycobacterium Tuberculosis*. *J. Struct. Funct. Genomics* **2010**, *11* (1), 41–49.

APPENDIX A

AMINO ACID SEQUENCES

- Wild type CYP119

MYDWFSEMRKKDPVYYDGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLED
LRNGKIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLETFIRETTRSLLD
SIDPREDDIVKKLAVPLPIIVISKILGLPIEDKEKFKEWSDLVAFRLGKPGEIFELG
KKYLELIGYVKDHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAGNETTTNLISNS
VIDFTRFNLWQRIREENLYLKAIEEALRYSPPVMRTVRKTKERVKLGDQTIEEG
EYVRVWIASANRDEEVFHDGEKFIPDRNPNPHLSFGSGIHLCLGAPLARLEARIA
IEEFSKRFRHIEILDTEKVPNEVLNGYKRLVVRLKSNE

- CYP119 Thr213Arg – Thr214Ile (Improved variant)

MYDWFSEMRKKDPVYYDGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLED
LRNGKIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLETFIRETTRSLLD
SIDPREDDIVKKLAVPLPIIVISKILGLPIEDKEKFKEWSDLVAFRLGKPGEIFELG
KKYLELIGYVKDHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAGNERITNLISNS
VIDFTRFNLWQRIREENLYLKAIEEALRYSPPVMRTVRKTKERVKLGDQTIEEG
EYVRVWIASANRDEEVFHDGEKFIPDRNPNPHLSFGSGIHLCLGAPLARLEARIA
IEEFSKRFRHIEILDTEKVPNEVLNGYKRLVVRLKSNE

- CYP119 Variants Used in the Determination of the Diversity of Library

Variant 1:

MYDWFSEMRKKDPVYYDGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLED
LRNGKIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLETFIRETTRSLLD
SIDPREDDIVKKLAVPLPIIVISKILGLPIEDKEKFKEWSDLVAFRLGKPGEIFELG
KKYLELIGYVKDHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAGNEIGTNLISNS
VIDFTRFNLWQRIREENLYLKAIEEALRYSPPVMRTVRKTKERVKLGDQTIEEG
EYVRVWIASANRDEEVFHDGEKFIPDRNPNPHLSFGSGIHLCLGAPLARLEARIA
IEEFSKRFRHIEILDTEKVPNEVLNGYKRLVVRLKSNE

Variant 2:

MYDWFSEMRKKDPVYYDGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLED
LRNGKIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLETFIRETTRSLLD
SIDPREDDIVKKLAVPLPIIVISKILGLPIEDKEKFKEWSDLVAFRLGKPGEIFELG
KKYLELIGYVKDHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAGNEIRTNLISNS
VIDFTRFNLWQRIREENLYLKAIEEALRYSPPVMRTVRKTKERVKLGDQTIEEG
EYVRVWIASANRDEEVFHDGEKFIPDRNPNPHLSFGSGIHLCLGAPLARLEARIA
IEEFSKRFRHIEILDTEKVPNEVLNGYKRLVVRLKSNE

Variant 3:

MYDWFSEMRKKDPVYYDGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLED
LRNGKIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLETFIRETTRSLLD
SIDPREDDIVKKLAVPLPIIVISKILGLPIEDKEKFKEWSDLVAFRLGKPGEIFELG
KKYLELIGYVKDHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAGNERGTNLISNS
VIDFTRFNLWQRIREENLYLKAIEEALRYSPPVMRTVRKTKERVKLGDQTIEEG
EYVRVWIASANRDEEVFHDGEKFIPDRNPNPHLSFGSGIHLCLGAPLARLEARIA
IEEFSKRFRHIEILDTEKVPNEVLNGYKRLVVRLKSNE

Variant 4:

MYDWFSEMRKKDPVYYDGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLED
LRNGKIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLETFIRETTRSLLD
SIDPREDDIVKKLAVPLPIIVISKILGLPIEDKEKFKEWSDLVAFRLGKPGEIFELG
KKYLELIGYVKDHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAGNEDITNLISNS
VIDFTRFNLWQRIREENLYLKAIEEALRYSPPVMRTVRKTKERVKLGDQTIEEG
EYVRVWIASANRDEEVFHDGEKFIPDRNPNPHLSFGSGIHLCLGAPLARLEARIA
IEEFSKRFRHIEILDTEKVPNEVLNGYKRLVVRLKSNE

Variant 5:

MYDWFSEMRKKDPVYYDGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLED
LRNGKIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLETFIRETTRSLLD
SIDPREDDIVKKLAVPLPIIVISKILGLPIEDKEKFKEWSDLVAFRLGKPGEIFELG
KKYLELIGYVKDHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAGNERGTNLISNS
VIDFTRFNLWQRIREENLYLKAIEEALRYSPPVMRTVRKTKERVKLGDQTIEEG
EYVRVWIASANRDEEVFHDGEKFIPDRNPNPHLSFGSGIHLCLGAPLARLEARIA
IEEFSKRFRHIEILDTEKVPNEVLNGYKRLVVRLKSNE

Variant 6:

MYDWFSEMRKKDPVYYDGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLED
LRNGKIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLETFIRETTRSLLD
SIDPREDDIVKKLAVPLPIIVISKILGLPIEDKEKFKEWSDLVAFRLGKPGEIFELG
KKYLELIGYVKDHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAGNERFTNLISNS
VIDFTRFNLWQRIREENLYLKAIEEALRYSPPVMRTVRKTKERVKLGDQTIEEG
EYVRVWIASANRDEEVFHDGEKFIPDRNPNPHLSFGSGIHLCLGAPLARLEARIA
IEEFSKRFRHIEILDTEKVPNEVLNGYKRLVVRLKSNE

Variant 7:

MYDWFSEMRKKDPVYYDGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLED
LRNGKIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLETFIRETTRSLLD
SIDPREDDIVKKLAVPLPIIVISKILGLPIEDKEKFKEWSDLVAFRLGKPGEIFELG
KKYLELIGYVKDHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAGNEGDTNLISNS
VIDFTRFNLWQRIREENLYLKAIEEALRYSPPVMRTVRKTKERVKLGDQTIEEG
EYVRVWIASANRDEEVFHDGEKFIPDRNPNPHLSFGSGIHLCLGAPLARLEARIA
IEEFSKRFRHIEILDTEKVPNEVLNGYKRLVVRLKSNE

Variant 8:

MYDWFSEMRKKDPVYYDGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLED
LRNGKIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLETFIRETTRSLLD
SIDPREDDIVKKLAVPLPIIVISKILGLPIEDKEKFKEWSDLVAFRLGKPGEIFELG
KKYLELIGYVKDHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAGNEGCTNLISNS
VIDFTRFNLWQRIREENLYLKAIEEALRYSPPVMRTVRKTKERVKLGDQTIEEG
EYVRVWIASANRDEEVFHDGEKFIPDRNPNPHLSFGSGIHLCLGAPLARLEARIA
IEEFSKRFRHIEILDTEKVPNEVLNGYKRLVVRLKSNE

Variant 9:

MYDWFSEMRKKDPVYYDGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLED
LRNGKIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLETFIRETTRSLLD
SIDPREDDIVKKLAVPLPIIVISKILGLPIEDKEKFKEWSDLVAFRLGKPGEIFELG
KKYLELIGYVKDHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAGNEVGTNLISNS
VIDFTRFNLWQRIREENLYLKAIEEALRYSPPVMRTVRKTKERVKLGDQTIEEG
EYVRVWIASANRDEEVFHDGEKFIPDRNPNPHLSFGSGIHLCLGAPLARLEARIA
IEEFSKRFRHIEILDTEKVPNEVLNGYKRLVVRLKSNE

Variant 10:

MYDWFSEMRKKDPVYYDGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLED
LRNGKIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLETFIRETTRSLLD
SIDPREDDIVKKLAVPLPIIVISKILGLPIEDKEKFKEWSDLVAFRLGKPGEIFELG
KKYLELIGYVKDHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAGNELCTNLISNS
VIDFTRFNLWQRIREENLYLKAIEEALRYSPPVMRTVRKTKERVKLGDQTIEEG
EYVRVWIASANRDEEVFHDGEKFIPDRNPNPHLSFGSGIHLCLGAPLARLEARIA
IEEFSKRFRHIEILDTEKVPNEVLNGYKRLVVRLKSNE

Variant 11:

MYDWFSEMRKKDPVYYDGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLED
LRNGKIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLETFIRETTRSLLD
SIDPREDDIVKKLAVPLPIIVISKILGLPIEDKEKFKEWSDLVAFRLGKPGEIFELG
KKYLELIGYVKDHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAGNEHHTNLISNS
VIDFTRFNLWQRIREENLYLKAIEEALRYSPPVMRTVRKTKERVKLGDQTIEEG
EYVRVWIASANRDEEVFHDGEKFIPDRNPNPHLSFGSGIHLCLGAPLARLEARIA
IEEFSKRFRHIEILDTEKVPNEVLNGYKRLVVRLKSNE

Variant 12:

MYDWFSEMRKKDPVYYDGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLED
LRNGKIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLETFIRETTRSLLD
SIDPREDDIVKKLAVPLPIIVISKILGLPIEDKEKFKEWSDLVAFRLGKPGEIFELG
KKYLELIGYVKDHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAGNESFTNLISNS
VIDFTRFNLWQRIREENLYLKAIEEALRYSPPVMRTVRKTKERVKLGDQTIEEG
EYVRVWIASANRDEEVFHDGEKFIPDRNPNPHLSFGSGIHLCLGAPLARLEARIA
IEEFSKRFRHIEILDTEKVPNEVLNGYKRLVVRLKSNE

Variant 13:

MYDWFSEMRKKDPVYYDGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLED
LRNGKIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLETFIRETTRSLLD
SIDPREDDIVKKLAVPLPIIVISKILGLPIEDKEKFKEWSDLVAFRLGKPGEIFELG
KKYLELIGYVKDHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAGNELCTNLISNS
VIDFTRFNLWQRIREENLYLKAIEEALRYSPPVMRTVRKTKERVKLGDQTIEEG
EYVRVWIASANRDEEVFHDGEKFIPDRNPNPHLSFGSGIHLCLGAPLARLEARIA
IEEFSKRFRHIEILDTEKVPNEVLNGYKRLVVRLKSNE

Variant 14:

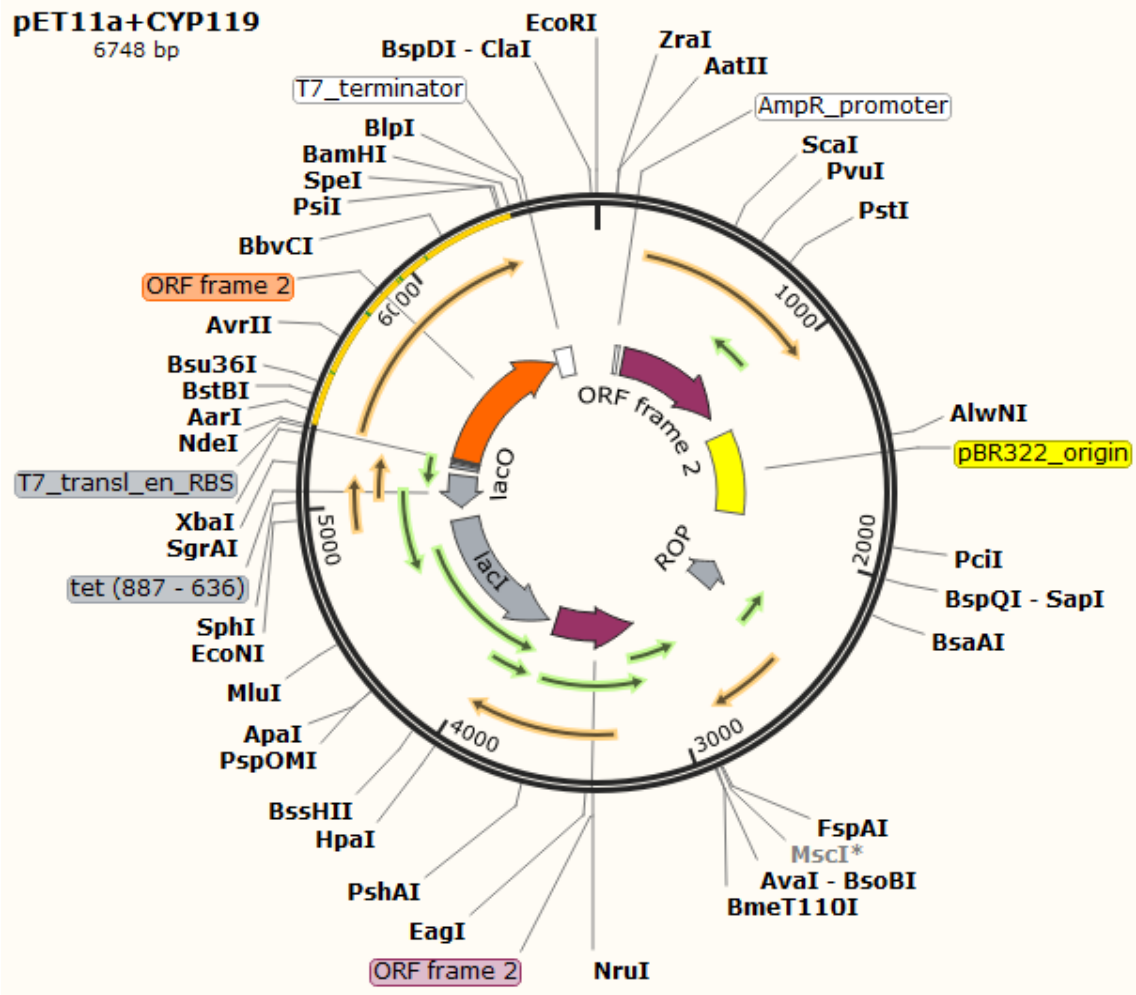
MYDWFSEMRKKDPVYYDGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLED
LRNGKIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLETFIRETTRSLLD
SIDPREDDIVKKLAVPLPIIVISKILGLPIEDKEKFKEWSDLVAFRLGKPGEIFELG
KKYLELIGYVKDHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAGNEYCTNLISNS
VIDFTRFNLWQRIREENLYLKAIEEALRYSPPVMRTVRKTKERVKLGDQTIEEG
EYVRVWIASANRDEEVFHDGEKFIPDRNPNPHLSFGSGIHLCLGAPLARLEARIA
IEEFSKRFRHIEILDTEKVPNEVLNGYKRLVVRLKSNE

Variant 15:

MYDWFSEMRKKDPVYYDGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLED
LRNGKIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLETFIRETTRSLLD
SIDPREDDIVKKLAVPLPIIVISKILGLPIEDKEKFKEWSDLVAFRLGKPGEIFELG
KKYLELIGYVKDHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAGNEFHTNLISNS
VIDFTRFNLWQRIREENLYLKAIEEALRYSPPVMRTVRKTKERVKLGDQTIEEG
EYVRVWIASANRDEEVFHDGEKFIPDRNPNPHLSFGSGIHLCLGAPLARLEARIA
IEEFSKRFRHIEILDTEKVPNEVLNGYKRLVVRLKSNE

APPENDIX B

VECTOR MAP



The illustration was created in SnapGene Viewer.