UNIVERSITY OF CRETE
DEPARTMENT OF CHEMISTRY

MASTER OF CHEMISTRY

ENZYME TECHNOLOGY LABORATORY

Master Thesis

Expression of cytochromes P450 and investigation of their selectivity towards bioactive compounds

THALEIA SAKOLEVA

MASTER THESIS SUPERVISOR: IOANNIS PAVLIDIS

HERAKLION 2020

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΧΗΜΕΙΑΣ

ΓΕΝΙΚΟ ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΧΗΜΕΙΑΣ

ΕΡΓΑΣΤΗΡΙΟ ΕΝΖΥΜΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΔΙΠΛΩΜΑ ΕΙΔΙΚΕΥΣΗΣ

Έκφραση κυτοχρωμάτων P450 και μελέτη της εκλεκτικότητας αυτών προς βιοδραστικές ενώσεις

ΘΑΛΕΙΑ ΣΑΚΟΛΕΒΑ

ΥΠΕΥΘΥΝΟΣ ΚΑΘΗΓΗΤΗΣ: ΙΩΑΝΝΗΣ ΠΑΥΛΙΔΗΣ

ΗΡΑΚΛΕΙΟ 2020

enzyme technology lab

UNIVERSITÄT GREIFSWALD
Wissen lockt. Seit 1456

BEC

To my mother and my father

Examination Committee

Uwe T. Bornscheuer

Professor, Department of Biotechnology and Enzyme Catalaysis, University of Greifswald

Demetrios Ghanotakis

Professor, Department of Chemistry, University of Crete

Ioannis Pavlidis (supervisor)

Assistant Professor, Department of Chemistry, University of Crete

Acknowledgments

Foremost, my sincere gratitude goes to my supervisor, Assist. Prof. Ioannis Pavlidis, for providing this interesting and challenging topic, mentoring, supervising, and advising me during my Master thesis. I would like to thank Prof. Dr. Uwe T. Bornscheuer for providing me the chance to work in his group, “Biotechnology and Enzyme Catalysis”, in the Biochemistry Institute of the University of Greifswald, the cytochromes P450’s constructs and reviewing this thesis.

Furthermore, I would like to thank the Ph.D. candidate Sascha Grobe and Dr. Chris Badenhorst for operational cooperation, great guidance, and useful advice, during my stay in Greifswald. I really enjoyed working alongside them.

Additionally, I would like to thank the group of Enzyme Technology, in the Department of Chemistry, University of Crete, and the group of Biotechnology and Enzyme Catalysis in Biochemistry Institute, University of Greifswald. The warm working atmosphere was really helpful and pleasant.

I would like to thank the Erasmus+ mobility program for the financial support during the mobility, and Mrs. Eugenia Skountaki, as the Erasmus+ student’s coordinator, School of Science, University of Crete, Campus of Heraklion. Also, Prof. Haralambos E. Katerinopoulos, University of Crete, Department of Chemistry for his help regarding the bureaucracy for the Erasmus+ mobility.

Moreover, I would like to thank Prof. Tsiotis and Prof, Ghanotakis, and all the members of their groups for support during the accomplishment of my Master thesis.

At last, I thank my family and my friends for supporting and accompanying me throughout the fulfillment of my studies.

Βιογραφικό Σημείωμα

Προσωπικές Πληροφορίες

Θάλεια Σακολέβα

Βας. Όλγας 24 Α, Θεσσαλονίκη, Ελλάδα
54541

thaliasakoleva@gmail.com

Φύλο Θήλυ | Ημερομηνία γέννησης 07/07/1995 | Εθνικότητα Ελληνική

Εργασιακή εμπειρία

Ιούλιος 2019-Οκτώβριος 2020	Μεταπτυχιακή εργασία, Τμήμα Χημείας, Πανεπιστήμιο Κρήτης, Ελλάδα Επιβλέπων: Επίκουρος Καθηγητής Ιωάννης Παυλίδης «Έκφραση κυτοχρωμάτων P450 και μελέτη της εκλεκτικότητας αυτών προς βιοδραστικές ενώσεις»
Σεπτέμβριος 2019-Δεκέμβριος 2019	Πρακτική Erasmus, Ινστιτούτο Βιοχημείας, Πανεπιστήμιο του Γκράιφσβαλτ, Γερμανία Επιβλέπων: Καθηγητής Uwe T. Bornscheuer «Έτερόλογη έκφραση κυτοχρωμάτων P450, διερεύνηση κατάλληλων οξειδοαναγωγικών συστημάτων και καθορισμός της εκλεκτικότητας αυτών σε οργανικά υποστρώματα για βιοτεχνολογικές εφαρμογές»
Οκτώβριος 2017-Σεπτέμβριος 2018	Διπλωματική εργασία, Τμήμα Επιστήμης και Τεχνολογίας Υλικών, Πανεπιστήμιο Κρήτης, Ελλάδα Επιβλέπων Καθηγητής: Καθηγητής Μιχάλης Κοκκινίδης «Κλωνοποίηση, έκφραση, καθαρισμός και κρυστάλλωση των μεταλλαγμάτων H131A, H135A της απακετυλάσης της πεπτιδογλυκάνης Bc1960»
Ιούλιος 2017-Σεπτέμβριος 2017	Βοηθός έρευνας, Ινστιτούτο Ηλεκτρονικής δομής και Λέιζερ, Κρήτη, Ελλάδα Επιβλέπων: Καθηγητής: Εμμανουήλ Στρατάκης «Κατασκευή φυσικών ικριωμάτων βιοϋλικών με παλμικά λέιζερ»

Εκπαίδευση

Οκτώβριος 2018-Οκτώβριος 2020	Μεταπτυχιακό Δίπλωμα Ειδίκευσης στη Βιοχημεία Βαθμός: 8.42/10 Πανεπιστήμιο Κρήτης, Τμήμα Χημείας, Ελλάδα
Οκτώβριος 2013-Νοέμβριος 2018	Πτυχίο Επιστήμης και Τεχνολογίας Υλικών Βαθμός: 6.97/10 Πανεπιστήμιο Κρήτης, Τμήμα Επιστήμης και Τεχνολογίας Υλικών, Ελλάδα
Σεπτέμβριος 2019-Ιούνιος 2013	Απολυτήριο Λυκείου Πειραματικό Σχολείο Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, Ελλάδα

Προσωπικές Δεξιότητες

Γλώσσες

- Ελληνικά (Μητρική γλώσσα)
- Αγγλικά (C1 επίπεδο)
- Γαλλικά (B2 επίπεδο)

Ψηφιακές δεξιότητες

- MS OFFICE software (Word, Excell, Powerpoint)
- Browsers (Explorer, Firefox, Chrome, Opera)
- λογισμικό Origin (ανάλυση δεδομένων)
- λογισμικό PyMOL (σύστημα μοριακής απεικόνισης)
- λογισμικό YASARA (πρόγραμμα μοντελοποίησης και προσομοίωσης)
- λογισμικό Geneious (ανάλυση αλληλουχιών)
- Fortran '95 (γλώσσα προγραμματισμού)
- Συμμετοχή σε μέσα κοινωνικής δικτύωσης (Linked In, Google+, Skype, Viber, WhatsApp, Facebook)

Εργαστηριακές δεξιότητες

Βιοχημικές

- Καλλιέργεια μικροοργανισμών (βακτηρία, ζυμομύκητες)
- Καλλιέργεια νευρικών κυττάρων
- Ηλεκτροφόρηση πρωτεϊνών και DNA (SDS-PAGE, Western blot, Agarose Gel)
- Πρωτεϊνικός καθαρισμός με χρωματογραφία συγγένειας
- Κρυσταλλογραφία
- Ενζυμικές κινητικές

Αναλυτικές τεχνικές

- Υγρή χρωματογραφία υψηλής απόδοσης (HPLC)
- Χρωματογραφία λεπτής στρώσης (TLC)
- Φασματοσκοπία UV-Vis
- Πυρηνικός Μαγνητικός Συντονισμός (NMR)

Μοριακής Βιολογίας

- Απομόνωση πλασμιδιακού DNA
- Αλυσιδωτή Αντίδραση Πολυμεράσης (PCR)

Πρόσθετες πληροφορίες

Υποτροφίες

-Υποτροφία Erasmus+ (Σεπτέμβριος 2019 – Δεκέμβριος 2019)

Συνέδρια

-Συμμετοχή στο 21ο Συνέδριο Μεταπτυχιακών Φοιτητών στη Χημεία, 2019 Ηράκλειο, Ελλάδα

-Συμμετοχή στο Συνέδριο του Μουσείου Φυσικής Ιστορίας του Πανεπιστημίου Κρήτης, «Natura 2000: Κοινωνία και Οικονομία», Ηράκλειο 2017

Προγράμματα και πιστοποιήσεις

-**Μέλος της Ομάδας Φοιτητικής Εθελοντικής Αιμοδοσίας του Πανεπιστημίου Κρήτης** με παράλληλη συνεργασία με το Τμήμα Αιμοδοσίας του Πανεπιστημιακού Νοσοκομείου Ηρακλείου (ΠΑΓΝΗ) για την οργάνωση εκδηλώσεων και αιμοδοσιών

Ενδιαφέροντα

-Κολύμβηση στον **ΗΡΑΚΛΗ ΘΕΣΣΑΛΟΝΙΚΗΣ**, 1998-2017

-Καταδύσεις, επίπεδο: Advanced (2 αστέρια), PADI, 3 σεζόν εργασιακή εμπειρία

-Πιστοποίηση άδειας ναυαγοσώστη από το 2015 και εργασιακή εμπειρία 3 σεζόν

Βραβεία & διακρίσεις

-3η θέση στον Πανελλήνιο Πρωτάθλημα Κολύμβησης, 2012, Ολυμπιακό Στάδιο Αθηνών (ΟΑΚΑ), Αθήνα, Ελλάδα

Curriculum Vitae

Personal Information

Thaleia Sakoleva

Vas. Olgas 24 A, Thessaloniki, Greece
54541
thaliasakoleva@gmail.com
Sex Woman | Date of birth 1995/07/07 | Nationality Greek

Work experience

July 2019-
October 2020

Master thesis, Department of Chemistry, University of Crete, Greece
Supervisor: Assist. Prof. Ioannis Pavlidis
"Expression of cytochromes P450 and investigation of their selectivity towards bioactive compounds"

September 2019-
December 2019

Erasmus Placement, Institute of Biochemistry, University of Greifswald, Germany
Supervisor: Prof. Uwe T. Bornscheuer
"Heterologous expression of cytochrome P450 enzymes, investigation of suitable redox partner systems and determination of their selectivity towards organic substrates for biotechnological applications"

October 2017-
September 2018

Diploma Thesis, Department of Materials Science and Technology, University of Crete, Greece
Supervisor: Prof. Michael Kokkinidis
"Cloning, expression, purification and crystallization of the mutants H131A , H135A of the bacterial peptidoglycan deacetylase Bc1960"

July 2017-
September 2017

Research assistant, Institute of Electronic Structure and Laser, Crete, Greece
Supervisor: Prof. Emmanuel Stratakis
"Manufacture of natural biomaterial scaffolds using pulse lasers"

Education

October 2018-
October 2020

Master of Science in Biochemistry
Grade: 8.42/10
University of Crete, Dept. of Chemistry, Greece

October 2013-
October 2018

Bachelor's degree in Materials Science and Technology
Grade: 6.97/10
University of Crete, Dept. of Materials Science & Technology, Greece

September 2010-
June 2013

High School Diploma
Experimental High School, Aristotle University of Thessaloniki, Greece

Personal Skills

Languages
- Greek (Mother tongue)
- English (C1 level)
- French (B2 level)

Digital Skills
- MS OFFICE software (Word, Excell, Powerpoint)
- Browsers (Explorer, Firefox, Chrome, Opera)
- Origin software (data analysis)
- PyMOL software (molecular visualization system)
- YASARA software (molecular-graphics, modeling and simulation program)
- Geneious software (sequence analysis analysis)
- Fortran '95 (computer coding)
- Participation in social networks (Linked In, Google+, Skype, Viber, WhatsApp, Facebook)

Laboratory-related skills

Biochemistry
- Cultures of microorganisms (bacteria, yeasts)
- Neural cell cultures
- Protein & DNA Electrophoresis (SDS-PAGE, Western blot, Agarose Gel)
- Protein purification with affinity chromatography
- Crystallography
- Enzyme Kinetics

Analytical Techniques
- High-Performance Liquid Chromatography (HPLC)
- Thin Layer Chromatography (TLC)
- UV-Vis Spectroscopy
- Nuclear Magnetic Resonance (NMR)

Molecular Biology
- Plasmid DNA isolation
- Polymerase Chain Reaction (PCR)

Additional Information

Scholarships

-Erasmus+ Scholarship (September 2019 – December 2019)

Conferences

-Participation in 21st Postgraduates' Conference on Chemistry, 2019, Heraklion, Greece

-Participation in the Natural History's Museum of the University of Crete Conference, "Natura 2000: Society and Economy", Heraklion, 2017

Projects

-**Member of the Blood Donor Association of the University of Crete** and Collaboration with the Blood Donation Department of the University Hospital of Heraklion for the organization of events and donations

Interests

-Swimming in **IRAKLIS THESSALONIKIS,** 1998-2017

-Scuba Diving, level: Advanced (2 stars), PADI and 3 seasons work experience

-Lifeguard Certificate since 2015 and 3 seasons work experience

Awards & Distinctions

-3rd place in Hellenic Swimming Championship, 2012, Olympic Stadium of Athens (OAKA), Athens, Greece

ΠΕΡΙΛΗΨΗ

Τα κυτοχρώματα P450 (CYPs) ή μονοοξυγενάσες P450 ανήκουν στην υπεροικογένεια των μονοοξυγενασών που διαθέτουν αίμη *b* και εμπλέκονται στην οξείδωση μιας ποικιλίας οργανικών ενώσεων, όπως ξενοβιοτικά, αντιβιοτικά, στεροειδή, λιπαρά οξέα και άλλα. Σε μια τυπική αντίδραση που καταλύεται από κυτοχρώματα P450, σχηματίζονται μέσω υδροξυλίωσης, μια αλκοόλη και νερό, με ταυτόχρονη οξείδωση του φωσφορικού νικοτινάμιδο-αδένινο νουκλεοτιδίου (NADPH). Το NADPH δρα ως δότης ηλεκτρονίων μέσω οξειδοαναγωγικών συστημάτων πρωτεινών, για την υδροξυλίωση του στόχευμένου υποστρώματος, αλλά η πολυπλοκότητα αυτών των συστημάτων περιορίζει την ευρύτερη εφαρμογή τους. Σκοπός της παρούσας διατριβής είναι η δημιουργία μιας συλλογής οξειδοαναγωγικών βιοκαταλυτών, ικανών να συνθέτουν βιοδραστικές ενώσεις. Το έργο χωρίζεται σε δύο μέρη.

Στο πρώτο μέρος, μετά από βιοπληροφορική ανάλυση αλληλούχισης προηγούμενων ερευνητών, ένας κατάλογος είκοσι διαφορετικών υποψηφίων CYPs μελετήθηκε αρχικά για τη συν-έκφραση με διαφορετικά οξειδοαναγωγικά συστήματα και έπειτα για την επιλεκτική υδροξυλίωση ενός στερεοειδούς ενδιαφέροντος. Η αντίδραση πραγματοποιήθηκε σε συστήματα ολόκληρων κυττάρων, χρησιμοποιώντας στελέχη *E. coli* ως ξενιστές. Οι μελέτες αποσκοπούσαν στην επιτυχή πρωτεινική έκφραση με σωστή πρωτεινική αναδίπλωση και υδροξυλίωση του υποστρώματος με την εύρεση του κατάλληλου οξειδοαναγωγικού συστήματος. Ο έλεγχος της έκφρασης πραγματοποιήθηκε με ηλεκτροφόρηση πολυακρυλαμιδίου υπό αποδιατακτικες συνθήκες, η αναδίπλωση πρωτεΐνης υποδείχθηκε μετά από μελέτη κορεσμού των κυτοχρωμάτων με CO και ο σχηματισμός προϊόντος μελετήθηκε χρησιμοποιώντας χρωματογραφικές μεθόδους διαχωρισμού. Τα αποτελέσματα έδειξαν ότι παρόλου που αρκετά από τα CYPs ήταν σωστά αναδιπλωμένα, μόνο το OleP υδροξυλίωνε το στερεοειδές ενδιαφέροντος.

Χρησιμοποιώντας το OleP ως την καταλληλότερη μονοοξυγενάση, πραγματοποιήθηκαν αντιδράσεις με άλλα στερεοειδή υποστρώματα, όπως το χολικό οξύ και το δεοξυχολικό οξύ. Από τις βιομετατροπές σε ολόκληρα κύτταρα, αλλά και με καθαρισμένες πρωτεΐνες, μπόρεσε να αποδειχθεί ότι το OleP μπορεί να υδροξυλιώσει το δεόξυχολικό οξύ αλλά όχι το χολικό οξύ, τόσο με το σύστημα PdR/PdX, όσο και με το BMR.

Με βιοπληροφορική ανάλυση μελετήθηκε η εκλεκτικότητα των CYPs ως προς τα υποστρώματα ενδιαφέροντος. Η ανάλυση αυτή αυτή δεν μπόρεσε να αναδείξει το λόγο της εκλεκτικότητας του OleP προς το δεοξυχολικό οξύ, ενώ ήταν ανενεργό με το χολικό

οξύ. Παρόλα αυτά, μπόρεσε να προβλέψει την σωστή πρόσδεση του δεοξυχολικού οξέος, ενώ μέσω της ανάλυσης επιλέχθηκαν κάποιες θέσεις για πειράματα μεταλλαξιγένεσης, που θα μπορούσαν δυνητικά να οδηγήσουν σε διαφορετική τοποεκλεκτικότητα.

Στο δεύτερο μέρος μελετήθηκε η ανάπτυξη και η βελτιστοποίηση πρωτοκόλλων έκφρασης με σκοπό τη μεγιστοποίηση της παραγωγής συντηγμένων κυτοχρωμάτων με μοριακό βάρος ~ 120 kDa σε σωστά αναδιπλωμένη μορφή. Το μεγάλο μοριακό βάρος αποτελεί πρόκληση για την πρωτεινική έκφραση και σταθερότητα σε βακτηριακά συτήματα. Μετά την επιτυχή έκφραση και καθαρισμό των συντηγμένων κυτοχρωμάτων σε σωστά αναδιπλωμένη μορφή, ακολούθησε φωτομετρική ανάλυση αντιδράσεων με τα δύο εναντιομερή του λιμονενίου ως υποστρώματα. Η βιβλιογραφία επιβεβαιώθηκε για το (*R*)-λιμονένιο καθώς παρατηρήθηκε κατανάλωση NADPH, υποδεικνύοντας τη μετατροπή υποστρώματος. Παρόλα αυτά υπήρχαν και αναφορές για ενεργότητα ως προς το (*S*)-λιμονένιο, οι οποίες δεν επιβεβαιώθηκαν. Τα κυτοχρώματα που ταυτοποιήθηκαν στην παρούσα εργασία και εκφράστηκαν σε σωστά αναδιπλωμένη μορφή μπορούν να μελετηθούν περαιτέρω στο μέλλον όσον αφορά το φάσμα των υποστρωμάτων που υποδέχονται, ώστε να οδηγήσουν σε περισσότερα προϊόντα πιθανού βιοτεχνολογικού ενδιαφέροντος.

Λέξεις-κλειδιά: μονοοξυγενάσες P450, οξειδοαναγωγικά συστήματα, NAD(P)H, υδροξυλίωση, στεροειδή

<u>Σημείωση</u>:

Λόγω συμφωνίας εμπιστευτικότητας που υπογράφηκε με το Ινστιτούτο Βιοχημείας, Πανεπιστήμιο του Greifswald, δεν μπορεί να αναφερθεί η δομή του υποστρώματος με το οποίο εργάστηκα κατά την παραμονή μου στη Γερμανία, και για αυτό την αναφέρω ως «στεροειδές ενδιαφέροντος».

Ευχαριστώ για την κατανόηση.

ABSTRACT

Cytochromes P450 (CYPs) or P450 monooxygenases belong to the superfamily of heme *b* containing monooxygenases and they are involved in the oxidation of a variety of organic compounds, such as xenobiotics, antibiotics, steroids, fatty acids, and others. In a typical reaction catalyzed by P450s, an alcohol via hydroxylation and water are formed under the consumption of reduced nicotinamide adenine dinucleotide phosphate (NADPH). NADPH acts as an electron donor via redox partner protein systems for the hydroxylation of the target substrate, but the complexity of these systems limits the wider application of the CYPs. The purpose of the present thesis is the creation of a collection of redox biocatalysts which will be able to synthesize bioactive compounds. The project is separated into two parts.

In the first part, after previous researchers' bioinformatics analysis on sequencing, a list of twenty different P450 candidates was screened firstly for the co-expression with different redox partner systems, and second, for the highly selective hydroxylation of a steroid of interest. The reaction was performed in whole-cell systems, using *E. coli* strains as expression host. The studies aimed for the successful recombinant expression with proper protein folding, and substrate conversion using the proper redox partner. The expression level was tested using SDS-PAGE, the proper folding was indicated with CO titration and the product formation was studied using chromatographic methods. The results indicated that although several of the studied CYPs were properly folded, only OleP was able to hydroxylate the steroid of interest.

Using OleP as the most promising P450 monooxygenase candidate, reactions were performed, using other steroid substrates, such as cholic acid and deoxycholic acid. From the whole-cell biocatalysis, but also using purified proteins, it could be shown that OleP can hydroxylate deoxycholic acid, but not cholic acid, using either PdR/PdX system or BMR reductase.

Using bioinformatic analysis, substrate-selectivity of CYPs was investigated. The analysis could not highlight the underlying reason of the selectivity of OleP towards deoxycholic acid, while it was inactive against cholic acid. Nevertheless, it could predict the right binding conformation of deoxycholic acid, while – via this analysis – several positions were selected for mutagenesis experiments, which could potentially lead to different regioselectivity.

In the second part, the development and optimization of expression protocols to maximize the production of fused CYPs with a molecular weight of ~120 kDa in a properly folded form was investigated. The high molecular weight can be challenging for the protein expression and the protein stability in bacterial systems. The successful expression and purification of these CYPs in proper form was followed from photometric analysis of reactions with limonene as substrate. Literature data was confirmed for (*R*)-limonene, as NADPH was consumed during these reactions, however, despite the reports of activity with (*S*)-limonene as well, no activity could be monitored. CYPs identified in this thesis and were expressed in proper folded conformation can be further studied in the future concerning the substrate scope they accept, in order to get access to further products of potential biotechnological interest.

Keywords: P450-monooxygenases, redox partner, NAD(P)H, hydroxylation, steroids.

Note: Due to confidentiality agreement signed with the Insitute of Biochemistry, University of Greifswald, the structure of the substrate that I worked with in Germany cannot be mentioned, and thus I will refer to it as “steroid of interest”.

Thank you for your understanding.

Table of Contents

Abbreviations

°C	Degrees Celsius
λ	Wavelength
µL	Microliter
µm	Micrometer
AB mix	Acrylamide/bis-acrylamide
Abs	Absorption
APS	Ammonium persulfate
amp	Ampicillin
B.	*Bacillus*
CA	Cholic acid
CO	Carbon monoxide
CPR	Cytochrome P450 reductase from the organism *Candida apicola*
Cys	Cysteine
chl	Chloramphenicol
DCA	Deoxycholic acid
ALA	δ-Aminolevulinic acid
dH_2O	Distilled water
F.	*Fusarium*
FAD	Flavin adenine dinucleotide
FMN	Flavin mononucleotide
g	Gram(s)
h	Hour(s)
HPLC	High-Performance Liquid Chromatography
IPTG	Isopropyl-β-D-1-thiogalacto-pyranoside
kan	Kanamycin
K_{cat}	Specificity catalytic constant
K_D	Dissociation constant
kDa	Kilo Dalton
K_M	Michaelis constant
K_{Pi}	Potassium phosphate (K_2HPO_4/KH_2PO_4) buffer

L	Liter
LB	Lysogenic Broth
mL	Milli Liter
M	Molarity
NAD(P)H	Nicotinamide adenine dinucleotide phosphate
$NAD(P)^+$	Oxidized NAD(P)H
NMR-spectroscopy	Nuclear Magnetic Resonance
M	Meter(s)
Min	Minute(s)
Nm	Nanometer
OD_{600}	Optical density at λ=600 nm
OleP	Cytochrome P450 CYP107D1 from the organism *Streptomyces antibioticus*
PAGE	Polyacrylamide Gel Electrophoresis
P. putida	*Pseudomonas putida*
P450cam	P450 from the organism *Pseudomonas putida*
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PdR	Putidaredoxin reductase from the organism *Pseudomonas putida*
PdX	Putidaredoxin from the organism *Pseudomonas putida*
P450-BM3	P450 from the organism *Bacillus megaterium*
P450-BaLi	P450 from the organism *Bacillus licheniformis*
S	Second(s)
SOC	Super Optimal broth with Catabolite repression
St	Streptomycin
TB medium	Terrific broth
TFA	Trifluoroacetic acid
TLC	Thin Layer Chromatography

CHAPTER 1. Theoretical part

1.1 Biocatalysis in Biotechnology

It is widely known that many chemicals either fine or bulk like pharmaceutical products and biofuels can be chemically synthesized. Simultaneously, biocatalysis provides the synthesis of chemical components using enzymes as nature's reaction catalysts [1-3]. The enzymes can be semipurified, as a lysate, in immobilized form, or even in whole-cell systems [3, 4]. Particularly, the whole-cell system is a reaction method where all the steps included for the protein purification and immobilization can be skipped, and cofactors like NADPH are regenerated from the cells [5].

So far, biocatalysis research in synthetic chemistry and the different technical innovations have generated four milestones during the progress in the biotechnology field, which marked the periods and the development of the application of enzymes as tools in chemistry (Figure 1.1).

The application of natural catalysts started over 100 years ago when the researchers discovered the nowadays called enzymes. Enzymes are capable to perform chemical transformations, as the (*R*)-mandelonitrile from benzaldehyde (first wave) [1, 4].

The second wave (1980s and 1990s) started with protein engineering in the biotechnological field. It was based on the protein structure and site-directed mutagenesis expanded the scope of substrates which could be used from enzymes [1]. With the latest developments, biocatalysis was entered into the pharmaceutical field and the industrial fine chemicals' production [4].

The third wave of biocatalysis (late 1990s) concerns protein engineering tools as DNA shuffling and error-prone PCR combined with high-throughput screening methods. Directed evolution was introduced into the biocatalysis world after Frances Arnold and Pim Stemmer's work [6]. Changes of amino acids randomly, led to the creation of libraries and the screening of them to variants with improved characteristics regarding the substrate selectivity and the enzyme stability [1].

The fourth and approaching wave of biocatalysis integrates bioinformatics tools, molecular genetics, and metagenomics aiming to novel enzymes' discovery [1, 6].

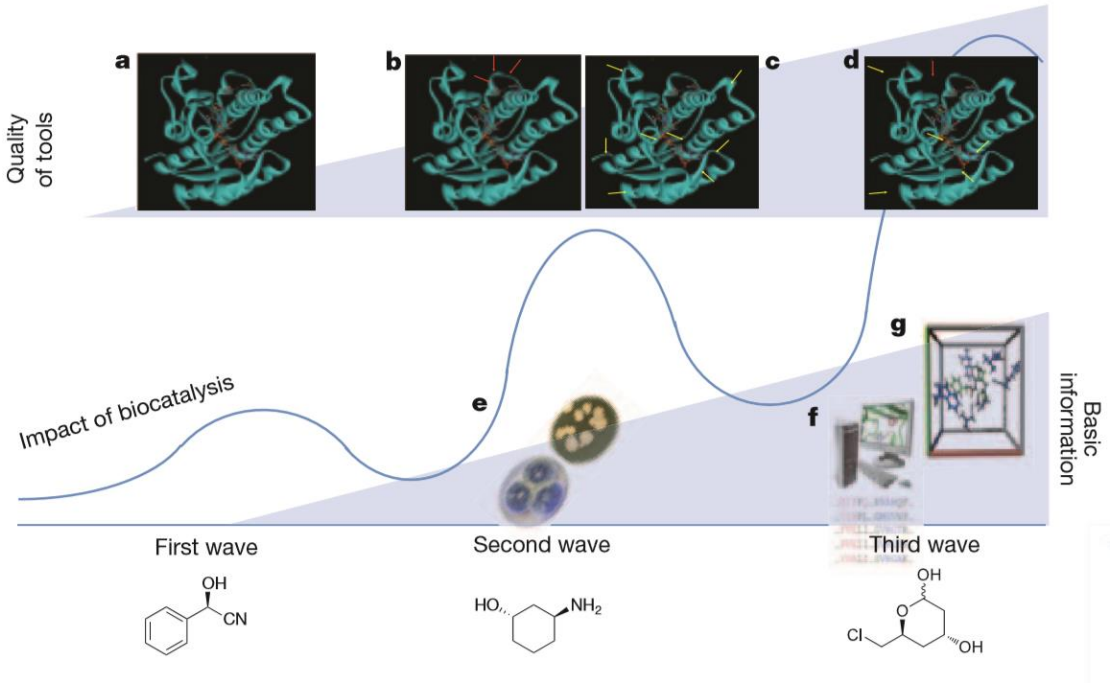


Figure 1.1: The three waves of biocatalysis, from rational design and mutagenesis, based on the protein structure to the directed evolution experiments and the creation of libraries. (a and b) Site directed mutagenesis on proteins with known structure, and (c) identification through rational design, and random mutagenesis combined with screening or selection. (d) A combination of the methods in (c) increases the possibility of the generation of smaller and smarter, libraries. (e) Enzyme screening through culturing. (f) Database search as a more effective for enzyme screening. (g) The computational *de novo* design of enzymes is still in progress. The fine chemicals appeared at the bottom, were used through the different waves of biocatalysis. (*R*)-mandelonitrile (left), (1*S*,3*S*)-3-aminocyclohexanol (centre), and 6-chloro-2,4,6trideoxy-D-erythrohexapyrano-side (right) [1]. Reprinted by permission from Springer Nature: [Journal Publicer: Springer Nature, Journal name: Nature], reference [1], Engineering the third wave of biocatalysis, U. T. Bornscheuer *et al* , [COPYRIGHT] 2012

Each enzyme evolves according to the given environmental conditions and thus has different profile, or optimal conditions. For instance, pH, temperature, pressure, cofactor-dependence, and substrates scope are parameters that have to be considered in biocatalysis (figure 1.2). So far, the biocatalytic process depends mostly on limitations on enzyme kinetics and stability. For example, P450-monooxygenases are challenging enzymes because of stability issues that may rise during the processes. In addition, the catalyst's features and cell's physiology can play a pivotal role during the process. More specifically, cell-growth, induction of overexpression of enzymes, or the use of metabolic pathways for multistep reactions are major factors related to the biocatalytic process and the differentation point from the classic synthetic chemistry financially [7], and environmentally [1]. The biodegradation ability, the non-

toxic feature of biocatalysts, the high selectivity (chemo-, regio- or stereo-selectivity), and the safe work process make the application of enzymes in organic synthesis even more attractive [1, 8].

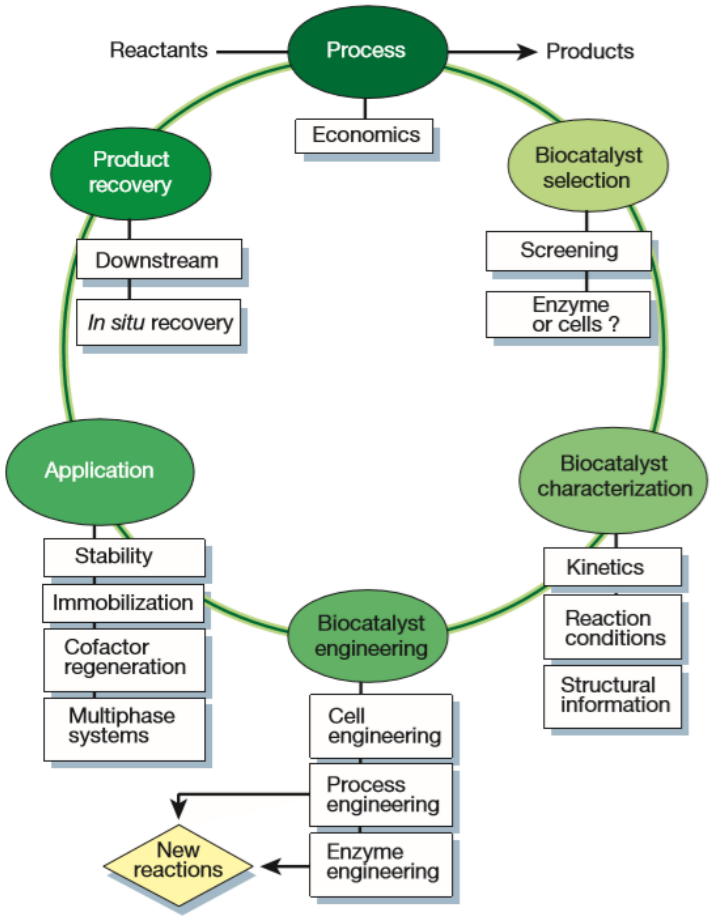


Figure 1.2: The biocatalysis cycle of the development of the biocatalytic process. Several factors are responsible for the economic feasibility of a biocatalytic process. The enzyme selection, reaction conditions and protein structure are needed. Many enzymes are used in the immobilized form as heterogeneous catalysts that can be recovered and reused. Also, some processes are based on homogeneously suspended cells or enzymes without recovery or reuse. Several companies use whole-cell systems for reactions such as specific coenzyme-dependent oxidoreductions [2, 7]. Reprinted by permission from Springer Nature: [Journal Publicer: Springer Nature, Journal name: Nature], reference [7] Industrial biocatalysis today and tomorrow, A. Schmid *et al* , [COPYRIGHT] 2001.

1.2 Cytochrome P450 monooxygenases

There are seven different classes of enzymes, according to the enzyme commission classification, based on the reaction they catalyse; oxidoreductases, transferases, hydrolases, lyases, isomerases, ligases and translocases are the seven categories and dispose the enzyme families [9]. The P450-monooxygenases (CYPs) belong to the family of oxidoreductases [2].

They are heme-dependent enzymes and their name is derived from their spectral properties, as after saturation with CO on their reduced state [10], the Soret absorption band shifts at 450 nm, while other heme proteins absorb approximately at 420 nm [11].

The CYPs are found in all kingdoms of life [12], and more than 89000 different sequences can be found in P450-related databases (biocatnet CYPED v6.0, Nelson, Dr) [13]. Regarding their classification in families and subfamilies, the criterion used mostly is the sequence identity and not always their substrate-specificity [14]. Particularly, a sequence identity over 40% represents the member of a family while the members of a subfamily represent a percentage of sequence identify over 55% [14].

Although CYPs can catalyze a wide range of reactions, hydroxylation is the most common [12]. For this, activation of O_2 is required, as it is introduced into a non-activated C-H bond [12]. Also, the reactions catalyzed by CYPs can be highly selective [12, 15]. On the other hand, several limitations seem to affect the application of P450-monooxygenases in biotechnology. First, the dependence on the cofactor NAD(P)H [12] for the enzymatic reactions and the need for a redox partner system for the electron transfer can be partially transcended using whole-cell systems, but still, the low water solubility of the substrates can be highly challenging [12] (paragraphs 1.2.2 and 1.3).

Although the obstacles faced in their application, CYPs are claimed to be highly perspective. Protein engineering tools and recombinant whole-cell systems can enhance their activity in selected substrates, applied in synthetic biology, and multi-enzyme cascades [12].

1.2.1 History

The P450 protein system was discovered in 1955 in the endoplasmic reticulum in rat liver [16, 17]. Two years later the CYP-catalyzed hydroxylation of squalene leading to cholesterol's synthesis and cholesterol conversion to several steroid hormones were studied in detail. The need for O_2, and the fact that the cofactor NAD(P)H acts as an electron donor centered the research interest [18, 19]. A study of the same group

concerned progesterone's hydroxylation at C-21 [18]. In 1958 the ability of CO binding to the enzyme and in 1962 the special spectral property with absorption at 450 nm was indicated *in vitro*. The same year, it was proved that the staining named "Pigment-450" or "P-450" concerns a protein with heme in the catalytic center and is closely related to the photometric property indicated [11]. The first PDB number for a P450 crystal structure was released in 1985 and concerned the P450cam from the organism *P. putida* (PDB ID: 2CCP) [20, 21]. The reaction mechanism of the cytochrome P450 enzymes' hydroxylation was first proposed in 1971, with the cyclic mechanism model [19, 22]. Of course, since then, further studies regarding the catalytic cycle were conducted and there are many alternative versions published [23].

The ability to metabolize xenobiotic compounds like drugs in the liver brought them under the spotlight of several research groups. However, as CYPs can generate toxic chemically-reactive intermediates, their study was challenging [24]. The question was if the overproduction of these enzymes is "beneficial" or "lethal" for the host-cell, but this can be answered case by case [24].

P450s and their properties are until today a field of research thanks to their unique properties and capabilities in C-H activation.

1.2.2 Reactions catalyzed by cytochrome P450s

The heme-containing enzymes are considered to be the most versatile biocatalysts in nature [25]. They can accept various organic substrates found in nature, as terpenoids, polyketides, fatty acids, alkaloids, and polypeptides [26-28]. In a living organism, for example, CYPs have been noticed to degrade xenobiotics [29], metabolize steroids, carcinogens, and drugs [10] and biologically synthesize several endogenous compounds like vitamins and fatty acids in reactions that require regio- and stereo-specificity [30].

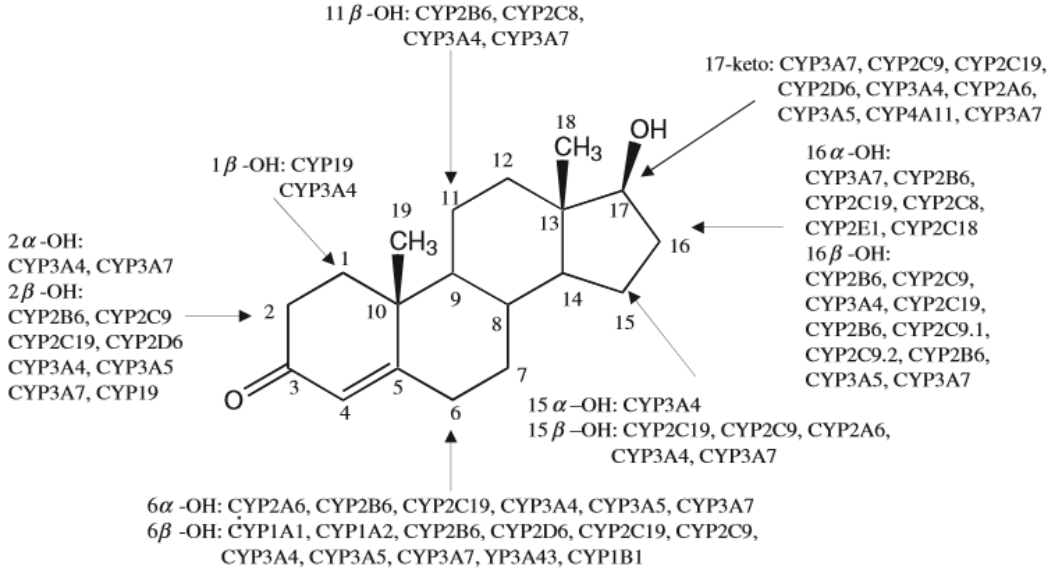
The reactions in which CYPs are involved are more than 20, including hydroxylation of non-activated C-H bonds and epoxidation, aromatic coupling, double bond epoxidation, and dealkylation on C-, N- and S-atoms [29] and C–C bond coupling or cleavage [31, 32].

The hydroxylation, forms alcohol and water under the consumption of NAD(P)H, which provides two electrons via redox partner protein systems (Scheme 1.1) [12, 33].

$$NAD(P)H + RH + O_2 + H^+ \rightarrow R\text{-}OH + NAD(P)^+ + H_2O$$

Scheme 1.1: The general reaction scheme catalyzed by cytochrome P450s. The NAD(P)H works as the electron donor and is oxidized to $NAD(P)^+$. Alcohol (R-OH) and H_2O are formed, while O_2 is needed [12].

A characteristic reaction catalyzed by P450s is the testosterone's hydroxylation, producing hydroxylation derivatives at various positions (Scheme 1.2.). The resulting products can be multiple times hydroxylated [34].



Scheme 1.2: The hydroxylation of testosterone by P450 and the camA/B gene (PdR/PdX from *P. putida*) [34]. Reprinted by permission from: Bioscience, Biotechnology and Biochemistry, Journal Publicer: Taylor and Francis, Journal name: Bioscience, Biotechnology and Biochemistry], reference [34], Hydroxylation of Testosterone by Bacterial Cytochromes P450 Using the *Escherichia coli* Expression System, H. Agematu *et al* , [COPYRIGHT] 2005 .

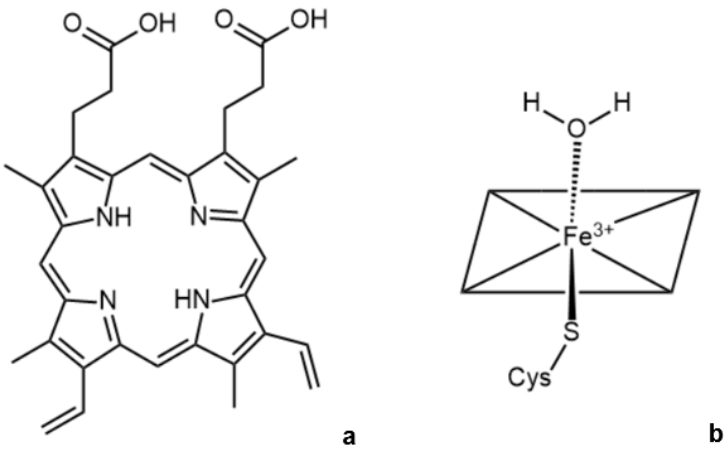
1.2.3 The active site

P450-monooxygenases are heme-containing enzymes and crystal or computational structures for CYPs or for their mutants or even in cases of metal substitution of the heme have been extensively studied, as for instance, the P450-BM3 mutant F87V [35-37].

The protoporphyrin IX or heme *b* is in the catalytic center of the cytochrome P450 and consists of a porphyrin ring coordinated to ferric iron and is a highly-conjugated and symmetrical prosthetic group [38]. The ferric iron has six ligands, four of which are nitrogen atoms that belong to the planar of the protoporphyrin ring, one axial ligand is the thiolate anion from the cysteine residue of the C-terminal region of the polypeptide

chain [39, 40] and the final distal ligand is the H_2O molecule which is coordinated to the iron when there is no substrate bounded (Scheme 1.3) [38].

The heme is located in a hydrophobic pocket. The hydrophobic residues of the pocket are responsible for the substrate-binding to the enzyme, using either ionic, H bonding, Van der Waals, or π–π bond stacking forces [38].



Scheme 1.3: (a) The protoporphyrin IX structure with the nitrogen atoms in the planar of the protoporphyrin ring and (b) the structure of the heme *b*, where the coordinated Fe^{3+} is displayed with all the six ligands, including nitrogen atoms and the axial thiolate anion from cysteine residue, and H_2O molecule as distal ligand [41].

The protein folding in the catalytic area has been maintained during the years, although the percentage of the amino acid sequence similarity is low. As shown in Figure 1.3, there are 12 α-helical regions (labeled A to L) and 5-beta (β) sheets (labeled β1–β5). The heme group is shown in the center of the structure, between the α-helix domain and the β-sheet domain [20, 38, 42].

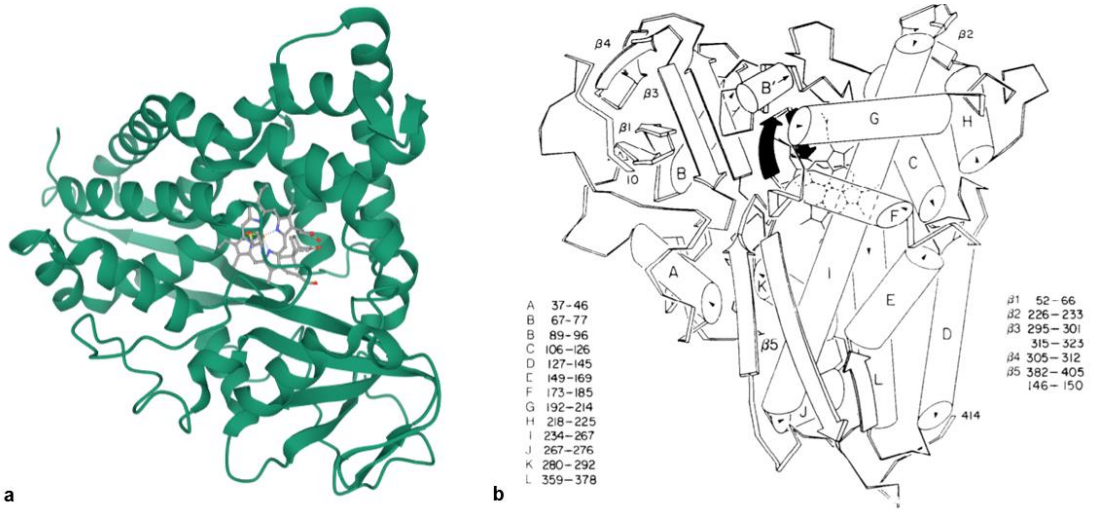


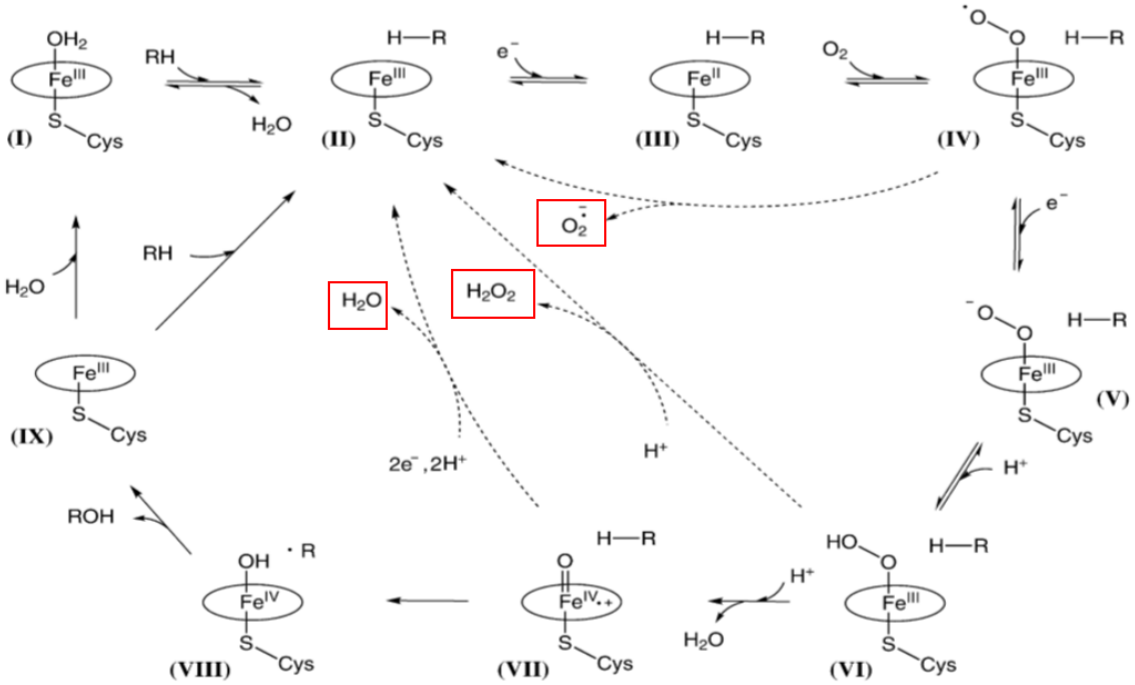
Figure 1.3: (a) Cartoon structure of the P450cam with the heme group in grey, together with the camphor, from the PDB [43]. (b) All the secondary structure regions labeled are displayed, corresponded to α-helix, or β-sheets. The lists mention the residues matching to each one of the regions of the structure. Reprinted by permission from The Journal of Biological Chemistry: [Journal Publicer: The Journal of Biological Chemistry, Journal name: The Journal of Biological Chemistry], reference [20] The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450, T. L. Poulos *et al* , [COPYRIGHT] 1985 .

1.2.4 Catalytic cycle of cytochrome P450s

The iron, as a chemical element, as described in the catalytic cycle (Scheme 1.4), can be found into two different redox forms, the reduced as the ferrous ion (Fe^{2+}), and the oxidized as the ferric ion (Fe^{3+}).

In the resting state, the water is bound to the iron of the heme (I). The cycle starts with the substrate binding in the active site, leading to the conformational change of the CYP and the removal of the H_2O [23] (II). In a next step, because of the conformational transition, the ferric spin state changes from low spin to high spin and the electron transferred from the NAD(P)H oxidation to $NAD(P)^+$, reduces the iron to the ferrous (from Fe^{3+} to Fe^{2+}) [44]. As a consequence, the iron is in the high spin form and the complex becomes a good electron acceptor, as the d-orbitals interaction is weakened (III) [23] [44]. The high-spin complex binds molecular oxygen and forms the dioxygen-Fe^{2+} complex (IV) [23, 44]. The singulet spin state of the latest complex is a good electron acceptor and a second electron transfer from the NAD(P)H forms the peroxo complex [23, 45] (V). The anion (peroxo-Fe^{3+}) is protonated and the hydroperoxy (Compound 0) intermediate is formed, because the peroxo is a good Lewis base [23, 44] (VI). As Compound 0 is a good Lewis base, the cycle is continued with a second protonation and the O-O bond is cleaved. As a result, a water molecule is formed and

the porphyrin complex is oxidized to Fe^{4+}, creating the reactive species (Compound 1) (VII) [23]. After the hydroxylation, the intermediate (VIII) is formed from Compound 1. The hydroxylated product (ROH) leaves the active site, and the heme returns to the pentacoordinate ferric state (IX) [44]. The cycle starts again with the addition of a water molecule in the complex, resulting in the resting state (I), and the cycle can start again [23].



Scheme 1.4: Scheme of the catalytic cycle of cytochrome P450 enzymes [44]. Reprinted by permission from The Chemical Society reviews: [Journal Publicer: Royal Society of Chemistry, Journal name: The Chemical Society reviews], reference [44] P450(BM3) (CYP102A1): connecting the dots, J. C. Whitehouse *et al* , [COPYRIGHT] 2012.

The catalytic cycle described in Scheme 1.4 is the general process CYPs follow for the substrate conversion. However, the enzyme can be reduced even without substrate binding [46]. This phenomenon indicates that the substrate is not necessarily converted into a product and hydrogen peroxide or superoxide can be formed. This is called “uncoupling” and can take place in two different states of the catalytic cycle, which are displayed in red in Scheme 1.4 [23, 38, 47, 48]. However, in the shunt pathway, where Compound 1 is directly formed, H_2O_2 use is limited because of the low-efficiency most CYPs have [31].

1.2.5 Spectral properties of cytochrome P450s

The name of the CYPs is based on the spectral properties these heme-dependent enzymes have and they are attributed to the coordination of the heme iron to the thiolate ligand [38].

The oxidized state of the ferrous (Fe^{3+}) presents a maximum at 418-424 nm, similar to other hemoproteins, like hemoglobin, which absorbs at the same range of wavelengths [10]. The reduced state of the hemic iron (Fe^{2+}) has a strong affinity to O_2 but it can bind with even higher affinity to CO [10, 38], which absorbs at 450 nm [10] and determines the CYP's proper folding [11, 49]. This spectral property is CYPs' signature and generates their term, P450s. On the other hand, if the signal of the Soret band remains at 420 nm, after the saturation with CO at the enzyme's reduced state, the protein is not properly folded around the heme, and especially the axial ligand [50]. In Figure 1.4 there are two spectra of CYP51B1 in its oxidezed and reduced form.

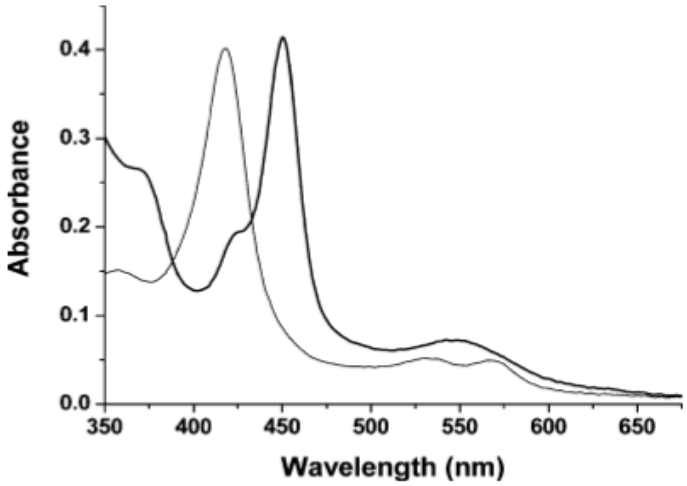


Figure 1.4: Spectra of P450 CYP51B1 from *Mycobacterium tuberculosis* [51], where A_{max} = 418 nm in its oxidized form (grey) and A_{max} = 450 nm Fe^{2+}- CO complex (black). The shoulder at 420 nm indicates a misfolded fraction of the enzyme. Reprinted by permission from Royal Society of Chemistry: [Journal Publicer: Royal Society of Chemistry, Journal name: The Natural Product Reports], reference [51] Variations on a (t)heme--novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily, A. W. Munro *et al* , [COPYRIGHT] 2007 .

1.2.6 Applications of P450s

Although CYPs are known since the middle of the 20th century [52], after the nomination of the Nobel prize in Chemistry to Prof. Frances H. Arnold Nobel about directed evolution [53], they are highlighted as promising catalysts, as they can be evolved to catalyze non-natural reactions [12, 53]. Thus, in biotechnology and synthetic biology, CYPs are very attractive catalysts because of their broad substrate

selectivity, which can be expanded using protein engineering. The reactions catalyzed by these CYPs could be applied in the synthesis of pharmaceuticals, natural products and fine chemicals [54].

Simultaneously, the limitations of CYPs, such as low catalytic efficiency, low stability, dependence on redox partners, high cost of cofactors, and electron uncoupling [12], make them unsuitable for the production of bulk chemicals and limit their industrial and commercial application [31]. On the other hand, some CYPs with high catalytic activity, as some fused proteins containing both the P450 and the reductase in the same polypeptide chain, can find application in bioconversion process for pravastatin production, conversion of vitamin D_3 to 1α,25 dihydroxy vitamin D3 (1α,25(OH)2D3), steroid production and in other compounds other with pharmaceutical interest, but only in lab-scale so far [55].

Another application of CYPs is in the field of biosensors [55]. For instance, Joseph *et al.* proved that cytochrome CYP3A4 can be used by monitoring the electron transfer to the heme [56]. Substrate's conversion causes a concentration-dependent increase of reduction current, based on cyclic voltametry. As a result, it is suggested that this biosensor can be used for drug detection [55, 56].

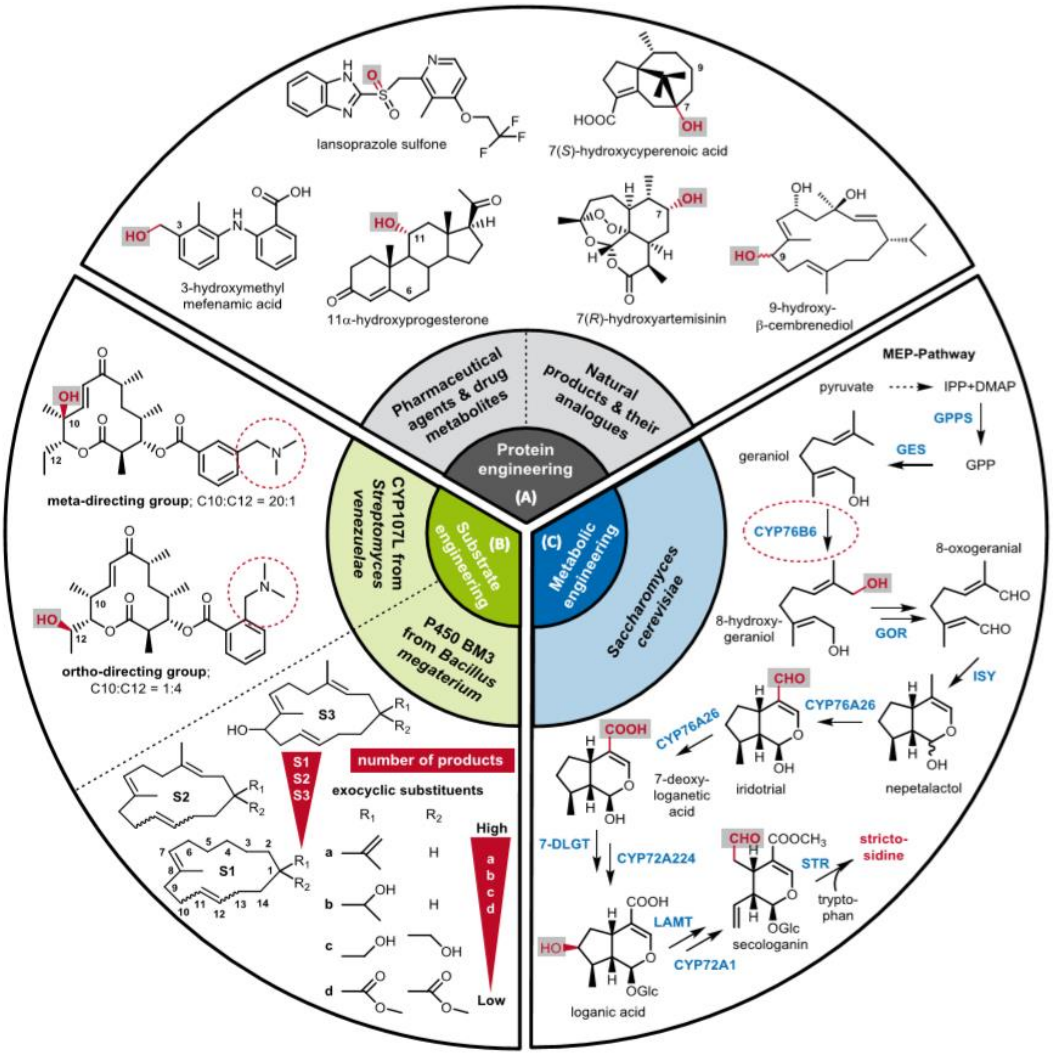


Figure 1.5: Examples of regio- and stereo-selective oxidations catalyzed by P450s after (A) protein engineering, (B) substrate engineering, and (C) metabolic engineering [54]. Reprinted by Elsevier and permission was not required. [Journal Publicer: Science Direct, Journal name: Current opinion in Chemical Biology], reference [54], Applications of microbial cytochrome P450 enzymes in biotechnology and synthetic biology, H.M Girvan, [COPYRIGHT] 2016

1.3 Redox partner systems

CYPs catalysis is based on an electron transfer system (paragraph 1.2.4). The cofactor NAD(P)H is converted a reductase of the redox partner system and constitutes the electron donor. The electrons are provided to the active site of the CYP after the oxidation of NAD(P)H, the catalytic cycle starts and the reaction to be accomplished [57].

Based on the reductase and the electron transfer system, the CYPSs are classified in ten different groups [58].

Class I are three-component systems consisted of FAD containing reductase that oxidizes NADH to NAD+, an iron-sulfur protein (FdR/FdX) that plays the role of the electron carrier and the P450 (Table 1.1, Figure 1.6). In this class CYPs from mitochondria and bacteria can be found. The P450cam from *P. putida* is the best-studied bacterial P450 of class I [58].

Class II are membrane-bound CYPs and receive their electrons from NAD(P)H through the CPR reductase. The CPR is a FAD and FMN containing reductase and membrane-bound (Table 1.1, Figure 1.6) [58].

Class III are self-sufficient monooxygenases and have similarities to the Class I, but instead of the iron-sulfur, clusters contain a flavodoxin (Fld) and an FMN group (Table 1.1) [58].

The CYP119 from the organism *Sulfolobus solfataricus* is found in Class IV and has been reported to be thermostable (up to 91°C). It is the first enzyme that does not obtain electrons from NAD(P)H-flavoprotein but a 2-oxoacid as pyruvic acid (Table 1.1) [57, 59].

In Class V there is an FdR domain and the ferredoxin (Fdx) is fused to the P450 at the C-terminus with a flexible alanine rich linker (Table 1.1) [57].

Class VI is an NAD(P)H dependent system and is consisted of an FdR domain and an Fld domain, instead of an Fdx, fused to the P450 [57]. This class was first described in a study from 2006 focussing on the use of P450 XplA as a remediation strategy in areas contaminated from the explosive hexahydro1,3,5-trinitro-1,3,5-triazine (Table 1.1,) [60].

The class VII of P450 enzymes is fused at its C-terminal side of the polypeptide chain with a phthalate dioxygenase reductase domain and the first P450 reported in this class was the CYP116B2 from *Rhodococcus* sp (Table 1.1, Figure 1.6) [57, 61].

Class VIII is a naturally fused system consisted from the components of the class II system but self-sufficient (CPR fused). Microsomal P450 enzymes are included in this class, as the human hepatic CYPs (Table 1.1, Figure 1.6) [55, 57].

Class IX is a self-sufficient P450 monooxygenase fused to a nitric oxide reductase, using NADH as cofactor. An example is the CYP55 which protects the fungus *Fusarium oxysporum* from the nitric oxide inhibition in the mitochondria (Table 1.1) [57].

Class X consists of cytochrome P450 independent enzymes as fatty acid hydroperoxide lyase, thromboxane synthases, allene oxide synthase, divinyl ether synthases and prostacyclin synthases (Table 1.1) [57]

Table 1.1: Classification of the different CYP systems, based on the topology of the electron transfer chain [58].

P450 Class	Electron transport chain
I	NAD(P)H → [FdR] → [Fdx] → [P450]
II	NAD(P)H → [CPR] → [P450]
III	NAD(P)H → [FdR] → [Fld] → [P450]
IV	Pyruvat, CoA → [OFOR] → [Fdx] → [P450]
V	NADH → [FdR] → [Fdx-P450]
VI	NAD(P)H → [FdR] → [Fld-P450]
VII	NADH → [PFOR-P450]
VIII	NADPH → [CPR-P450]
IX	NADH → [P450]
X	[P450]

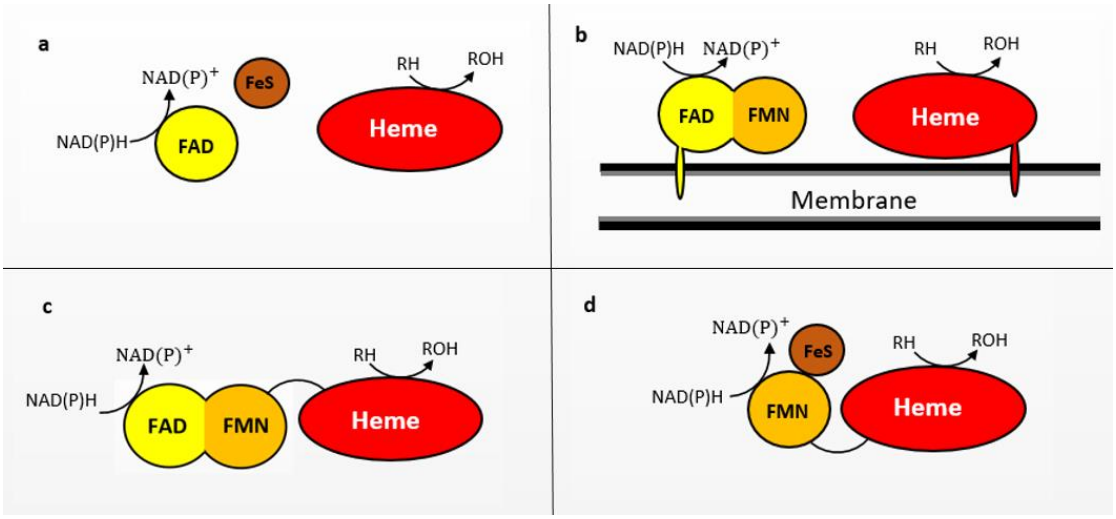


Figure 1.6: The four most common P450 classes with the CYP and the different reductases as electron carriers. (a) Class I, with the FAD-containing reductase, the ferredoxin [Fe-S] and the P450. (b) Class II, with the membrane-bound P450 and reductase. (c) Class VIII, with the self-sufficient system, fused with the FAD and FMN containing CPR reductase, via a linker. (d) Class VII, with the natural fused via linker P450 with an FMN and [Fe-S] containing reductase (phthalate dioxygenase) [33]. The figure is redrawn from Büchsenschütz *et al*.

Pdx/PdR is a redox partner system from *P. putida* and consists of two different components: the FAD-containing oxidoreductase which is a putidaredoxin reductase

(PdR) and the iron-sulfur cluster containing [Fe-S] ferrodoxin, called putidaredoxin (PdX). The electron transfer in this system proceeds from the oxidation of NAD(P)H by the reductase, and is transferred to the iron-sulfur cluster containing, PdX. PdX transports one electron at a time towards P450cam (Figure 1.6) [62, 63]. The binding of PdX towards P450cam is favored in the open conformation, with Asp251 being the key-residue for proton relay [64]. PdR/PdX is classified as class I redox system and applied together with P450cam in the stereospecific 5-exo hydroxylation of the terpene camphor [20, 65].

CPR (cytochrome P450 reductase) from *C. apicola*, is an N-terminally membrane-anchored protein of the endoplasmic reticulum, found in eukaryotic CYPs, especially in plants, and is classified as class II redox system (figure 1.6, b) [66]. CPR is NAD(P)H-dependent reductase, contains two prosthetic groups, FAD and FMN and can transfer the reducing equivalents from NAD(P)H to the FAD cofactor, and afterwards to the FMN. The electrons are transported to the heme in the active site [57]. Regarding the range of substrate efficiency CYPs can convert together with the reductase CPR from *C. apicola*, varies, as they can hydroxylate fatty acids, as long-chain (ω)- or (ω-1)-hydroxy fatty acids [66, 67]. For instance, CPR was tested together with the monooxygenase domain of CYP102A1 from *B. megaterium*, for the conversion of myristic acid, and proved 100 % conversion within 2 hours [66].

The reductase BMR (BM3 reductase) originates from *B. megaterium* (see also paragraph 1.4) and constitutes the electron transport system of the P450-BM3. The FAD/FMN-containing BMR consists of only one protein for the electron transfer. The monooxygenase of BM3 together with the reductase BMR catalyzes the ω-2-hydroxylation of saturated fatty acids and pharmaceuticals, drug-related biomolecules, aromatic compounds, and even surfactants as sodium dodecyl sulfate (SDS) have been investigated [44]. Additionally, Dodhia *et al.* fused the BMR as a redox-partner domain, to different P450 monooxygenases. The tests proved that these “molecular legos” can improve the studies regarding drug metabolism from CYPs and simplify them.

1.4 CYP102A1 (BM3) from *Bacillus megaterium*

The cytochrome CYP102A1 or P450-BM3 from the organism *B. megaterium* is a widely used and well-investigated monooxygenase of 120 kD [68] [69]. It is assigned in class VIII according to the P450-classification, based on the electron transfer system, with the heme-group in the active site and the FAD/FMN containing reductase, BMR for the electron transfer (paragraph 1.3) [44]. The fused protein is soluble and fairly stable

compared to other monooxygenases during processes [70]. In Figure 1.7, the heme-domain of the cytochrome P450-BM3 is monitored.

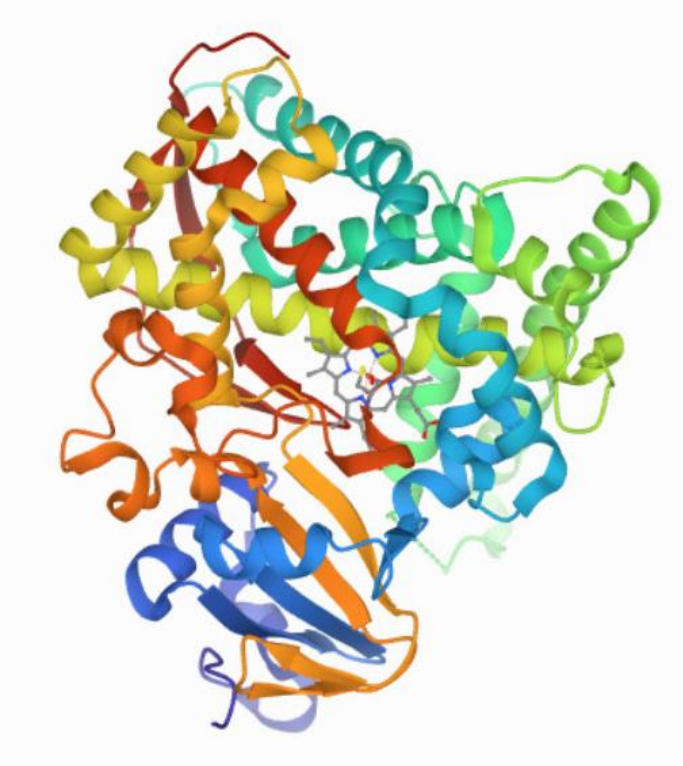


Figure 1.7: The structure of the heme domain of the cytochrome P450-BM3, obtained from the Protein Data Bank (PDB) [71].

P450-BM3 can catalyze the hydroxylation of saturated fatty acids, including the epoxidation of non-saturated, where the arachidonic acid as a substrate is a well-studied hydroxylation reaction of BM3, although it can convert a wide range of substrates' [72-75]. For the fatty acids, the hydroxylation is noticed mostly at ω-1, ω-2, and ω-3 position [75]. More specifically, high turnover rates, of 91.4 ± 8.9 s^{-1} were observed for the wild type BM3 converting palmitic acid [76].

Furthermore, the CYP102A1 monooxygenase, apart from the ability to convert fatty acids has been indicated for 4-(*R*)-limonene into various oxidative products [70]. Particularly, the creation of mutant libraries can improve or change the properties of the protein, as in the case of the mutations at positions 87 and 328, which were considered to be hotspots [70]. Seifert *et al*. introduced five hydrophobic amino acids (alanine, valine, phenylalanine, leucine, and isoleucine) at these positions. Six different products were revealed, where 4-(*R*)-limonene-8,9 epoxide was identified to be the major product [70].

Additionally, the residue Thr268, at the distal side of the heme is considered to be responsible for several roles, as oxygen activation, substrate recognition and electron transfer [77]. The mutagenesis Thr268Ala confirmed the threonine's importance to the enzymatic reaction, as the rate constant for NADPH consumption was decreased and the hydrogen peroxide production instead of the product formation was increased [77].

According to Li *et al.* in P450-BM3, the hypothesis of Phe87 to be responsible for oxidation is maintained [75]. Among other residues, the Phe87 has a strong impact on the substrate specificity and regio-selectivity [68]. In general, this residue seems to be responsible for blocking the substrate to closely approach the heme, by rotating and provoking a small change in the catalytic area [78]. Furthermore, the mutation of Phe87 to valine (F87V) has been proved to increase significantly the turnover rate [79]. It influences the activity properties of the P450-BM3 by giving 100% 14-(*S*)-15-(*R*)-epoxidation of arachidonic acid, compared to a percentage of 20% for the wild type variant, which also gave 80% 18-(*R*)-hydroxylation [78, 80].

Although P450-BM3 as a wild type is mostly known as a fatty acid hydroxylase, another mutant, the F87A has been indicated to be able to catalyze testosterone and generate an 18:82 mixture of two hydroxylated products, the 2β- and 16β-product, respectively [79].

1.5 CYP102A7 (BaLi) from *Bacillus licheniformis*

The P450 monooxygenase CYP102A7 (BaLi) from the organism *B. licheniformis*, likely to other proteins of the CYP101A subfamily, is a fatty acid hydroxylase, with preference to the medium length chains [29]. It also appeared to have moderate activity for unsaturated fatty acids and the highest activities towards saturated fatty acids [29]. Additionally, it catalyzes the oxidation of cyclic and acyclic terpenes, as (*R*)-(+)-limonene and (*S*)-(+)-limonene were converted into two main products (32% of cis-1,2-limonene epoxide and 28% of cis-carveol) and four minor products for the (*R*)-(+)-limonene, and respectively, for the (*S*)-(+)-limonene into the main product of (*S*)-(−)-limonene oxidation was trans-1,2 limonene epoxide (43%) [29].

As a member of the CYP102A family, BaLi is a self-sufficient monooxygenase, with molecular weight ~120 kDa. The reductase domain, is FAD/FMN containing [29, 81]. Regarding the structure of CYP102A7 has not been identified yet, but CYP102A7 is stable in DMSO, in contrast to other CYPs of the CYP102A family [29].

1.6 CYP107D1 (OleP) from *Streptomyces antibioticus*

The P450 monooxygenase CYP107D1 (OleP) is a bacterial epoxidase from the organism *S. antibioticus* and part of the biosynthesis of the 14-membered polyketide antibiotic oleandomycin [30, 82]. Particularly, OleP introduces an epoxide in the macrolide in a selective manner [15, 83]. Among the know macrolide-epoxygenase, OleP is the only one capable to introduce the epoxide function on a non-activated C-C bond, but the mechanism remains unknown [83].

CYP107D1 which has been crystallized together with 6 deoxyerythronoloide B and chlortrimazole has also been reported to hydroxylate steroids [30]. Studies have proven the unspecific hydroxylation of testosterone [34] and a recent study from Grobe *et al.* indicated the highly selective hydroxylation of lithocholic acid and deoxycholic acid, forming only 6β- hydroxylated products using the PdR/PdX redox partner system [15].

The crystallographic structure of OleP in complex with clotrimazole was obtained by Montemiglio *et al* (Figure 1.8) [83].

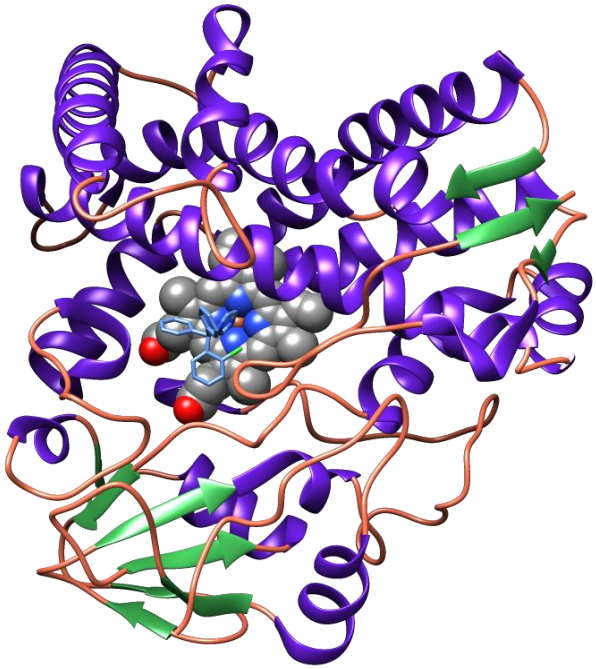


Figure 1.8: Crystal structure of the OleP-clotrimazole (CTZ) complex, where the azole moiety of the substrate is coordinated to the iron of the heme-group. CTZ with cornflower blue stick bound to OleP structure [83]. The structure was obtained from PDB [84] and edited in Chimera 1.13.1.

1.7 Substrates

The CYPs have a broad range of substrates. Endogenous compounds as steroids, fatty acids, lipid-soluble vitamins, natural products as terpenoids, alkaloids and xenobiotic chemical as carcinogens and drugs can be converted by them [28, 79, 85]. In the present thesis, steroids and terpenes have been used as CYPs' substrates and will be discussed.

1.7.1 Steroids

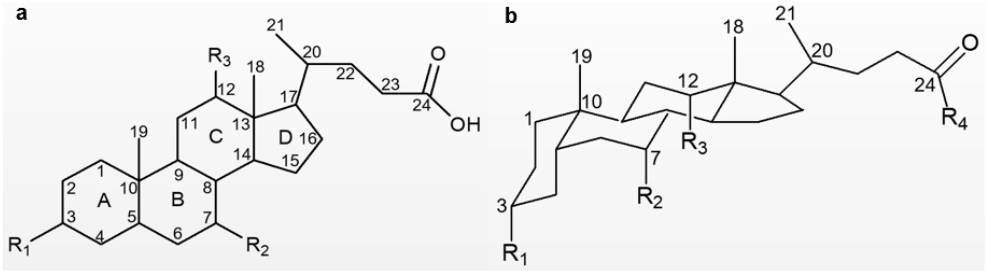
Bile acids are synthesized from cholesterol in liver through several pathways regulated by the levels of different bile acid species [86]. For example, they can regulate, apart from their synthesis and enterohepatic circulation, the homeostasis of triglyceride, cholesterol, glucose, and energy through signaling pathways [87]. These organic molecules and their derivatives are steroid compounds and play a pivotal role in the metabolism of fatty acids and vitamins and act as biomarkers for early diagnosis of several diseases. For instance, high levels of intestinal bile acids, and in particular, high levels of deoxycholic acid is an indication of colon carcinogenesis [15, 86].

Biocatalysis of steroid compounds in whole-cell systems represent a well-studied research area in white biotechnology [88], and interestingly, the microbial steroid transformation seems to be a powerful technique for generating novel steroidal drugs [88]. Although bile acids are involved in the pathogenesis and etiology of several diseases, their physical, chemical and biological properties made them an interesting scaffold for drug development [86]. Steroid-based drugs are widely applied for anti-tumor, anti-inflammatory, anti-microbial, anti-viral, anti-fungal, anti-estrogenic, and anti-convulsant activity and as anti-allergy agents [88]. Especially the hydroxylated derivatives of some bile acids, as chenodeoxycholic acid, are estimated to be target molecules and murideoxycholic acid's properties are under study [15].

As steroidal compounds, bile acids share several chemical, physical and biological characteristics [86]. The chemical structure is one of them, with the 24 carbon atoms (Scheme 1.5) and the core structure with the four rings being the main carbon skeleton of several compound-families, as triterpenoids, hormones, hopanoids, and others [86]. In general, their natural activity is structure-dependent. For example, the type, number, and the regio- and stereo- position of the functional groups are crucial for their activities [88]. Bile acids have a hydrophilic side (α-face) and a hydrophobic side (β-face), where the hydroxyl groups and the carboxylic side chain are oriented towards the α-side to give the hydrophilic character. The hydrophobic methyl groups (at C-18 and C-19) are oriented towards the β-side (Scheme 1.5) [86, 89]. Biologically, steroids are considered

to be digestive molecules and they metabolize fats and vitamins [86]. For example, biliary lipids are secreted because of bile acids action and combined to the chemical and physical properties these compounds have, mixed micelles together with biliary phosphospholipids are generated, and as a result, cholesterol is soluble in bile [86]. Additionally, bile acids are strong antimicrobial agents, and the growth of microorganisms in the intestine, is not affordable. Similarly, many other biological properties of bile acids regarding the liver's function, their role as signaling molecules, and pancreatic enzyme secretion have been further studied [86].

Steroid skeleton



Name	R_1	R_2	R_3	R_4
Cholanic acid	H	H	H	OH
Cholic acid	OH	OH	OH	OH
Chenodeoxycholic acid	OH (α)	OH (α)	H	OH
Deoxycholic acid	OH	OH	OH	OH
Ursodeoxycholic acid	OH (α)	OH (β)	H	OH
Lithocholic acid	OH	OH	H	OH
Glicocholate	OH	OH	OH	$NHCH_2COO^-$
Taurocholate	OH	OH	OH	$NHCH_2CH_2SO_3^-$

Scheme 1.5: The structures of the most common bile acids in humans and their glycine (glycocholate) and taurine (taurocholate) conjugates [86]. The figure is redrawn from Monte *et al* [86].

For steroid functionalization, hydroxylation is one of the most effective reactions for increasing their biological activity by turning them into more polar compounds and as a result, affect their toxicity [88]. Several publications and reviews are describing the conversion of different bile acids from CYPs [15, 18, 79, 85]. Around the white biotechnology framework, this kind of hydroxylations using microorganisms have been reported for the generation of intermediates able to provide access to inaccessible sites of the steroid molecule and provide therapeutic applications, as anti-inflammatory action by the presence of an oxygen group at C-11β position [88].

Cholic acid (3α-7α-12α-trihydroxy 5β-cholan-24-oic acid, CA), is a primary bile acid produced in liver, synthesized by cholesterol and deoxycholic acid (3α-6β-12α-trihydroxy 5β-cholan-24-oic acid, DCA) is a secondary bile acid, CA's derivative and is

considered to be an impurity of CA's bulk drugs (Scheme 1.5) [90]. OleP with PdR/PdX system was reported in a study with several bile acids that it can produce the 6β-hydroxylated product of DCA [15]. Interestingly, CA is not accepted as a substrate despite the similarity of the structure [15].

1.7.2 Terpenes

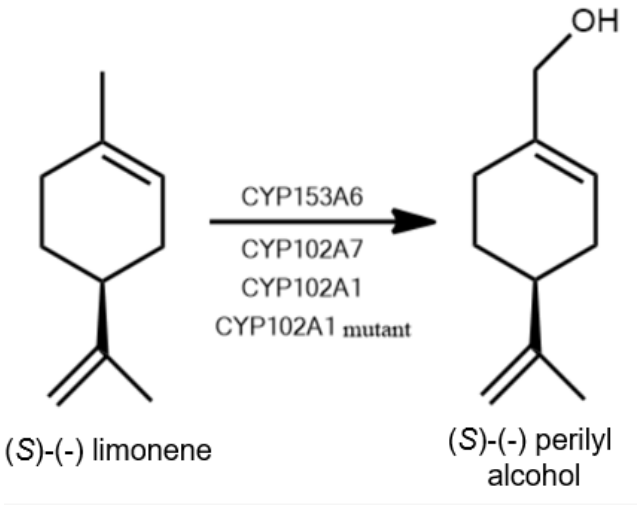
Natural products from plants have been extensively used by humans to improve health and nutrition throughout history. In parallel, many synthetic pharmaceuticals are based on natural products and are widely used in industrial area. They are attractive to manufacturers of cosmetics, perfumes, and fine chemicals and they represent a large family of plant natural products with over 20000 compounds.

Terpenes constitute characteristic examples of natural bioactive molecules and they meet two basic criteria associated with compounds that are considered to be good pharmaceuticals [14, 91]. Firstly, the C_{15} and C_{20} scaffolds of sesqui-, di-, and triterpenes carry an exceptionally high fraction of carbons in tetrahedral sp^3—compared to planar sp^2—and linear sp-hybridization (Fsp^3). Secondly, they contain many chiral centers. Chiral terpene building blocks have been the starting point of chemical synthesis of many natural products and as a consequence, pharmaceuticals, as well [92]. Metal catalysts as Ti/SiO_2 and Cu/SiO_2 have been used for the epoxidation, isomerization, and hydrogenation of terpenes [93]. However, the hydrophobic nature of terpenes can affect the catalytic process in aqueous media and CYPs were introduced as biocatalysts with non-polar substrates [93]. For biocatalysis in whole-cell systems, engineered *E. coli* with ferric hydroxamate was used to overcome the problem of cell diffusion [92]. The membrane proteins acted as passive diffusion channels and the hydrophobic substrate could enter the cell [93].

Terpenes can be found in plants and extracted from them. On the other hand, after recent research, terpenes can be produced in alternative ways for industrial purposes, using biotechnological methodology, with enzymes involved in their biosynthetic pathways [14]. Usually, terpene synthases are used in the first step to form the scaffolds and CYPs catalyze the highly-specific oxidation of them [14].

Limonene is a cyclic monoterpene, and its enantiomers are found to be the main in several citrus peel oils, like lemons and oranges. Limonene, can find applications in different industrial products such as cosmetics, perfumery, food flavoring, and cleaning products [14]. As a compound, it works as a starting material for natural products, and disposes biological properties, like antimicrobial, antioxidant, antinociceptive and anticancer activity.

Perillyl alcohol enantiomers, are limonene metabolites found in human body and are derived from the 7-hydroxylation catalyzed by CYP2C9 and CYP2C19 in the liver (Scheme 1.6) [14]. The (*S*)-(−)-perillyl alcohol is an organic compound with pharmaceutical interest because of its anticarcinogenic properties [14, 94, 95]. After studies, the P450-BaLi converts both (*R*)- and (*S*)- stereoisomers of limonene and form different products [29]. As a consequence, it is hypothesized that protein engineering in P450-BaLi can change the regioselectivity [14].



Scheme 1.6: Hydroxylation reaction of (*S*)-(-)-limonene to (*S*)-(-)-perillyl alcohol catalyzed by CYPs [14].

CHAPTER 2. Aim and methodology

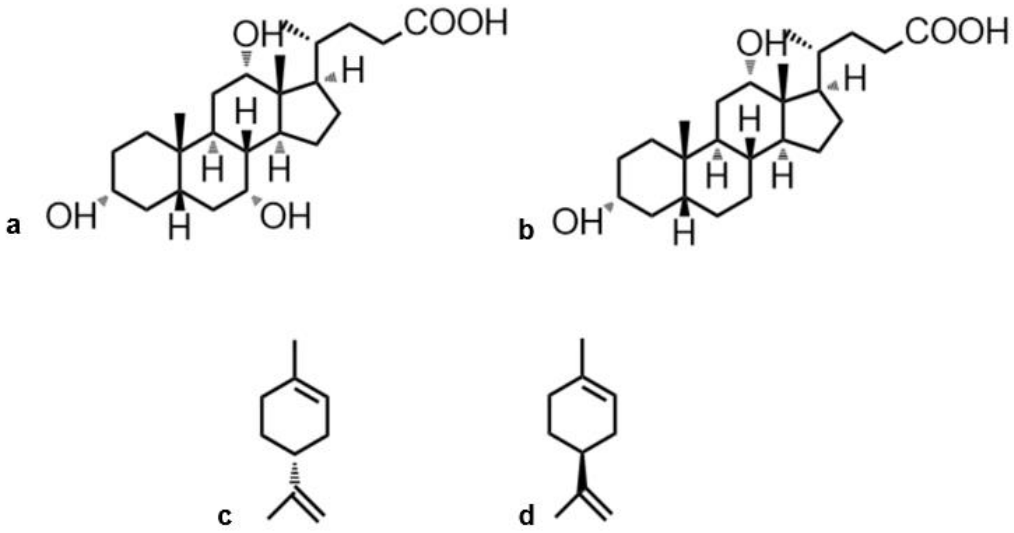
2.1 Aim of the thesis

The goal of the present thesis is the creation of a library of recombinantly expressed redox biocatalysts which will be able to synthesize bioactive compounds.

First of all, the thesis focuses on the development and optimization of recombinant expression protocols of CYPs, to maximize the production of self-sufficient CYPs in *E. coli* with a molecular weight of ~120 kDa in a properly folded form. NADPH consumption assays aimed to the establishment of positive controls, using (*R*)- and (*S*)-limonene as substrates.

In the second part of the thesis, the regio- and stereo-selectivity of CYPs in hydroxylation of bile acids was investigated. The project aimed at the screening of 20 different CYPs for (a) optimal expression in *E. coli*, (b) determination of the suitable electron transfer system, testing different redox partners and (c) identifying suitable whole-cell systems for the hydroxylation of a specific steroid of interest. Furthermore, more substrates were tested for hydroxylation using the suitable P450 and the redox system. The substrates tested are indicated in Scheme 2.1. This part of the research was mainly contacted at the Department of Biotechnology and Enzyme Catalysis, Institute of Biochemistry, University of Greifswald, after cooperation with Prof. Dr. Uwe T. Bornscheuer, via an Erasmus Placement Internship, and the experiments were finalized back in the Department of Chemistry, University of Crete.

In this framework, bioinformatics analysis focused on understanding the differences between CA and DCA.



Scheme 2.1: The substrates investigated during the thesis (steroid of interest not included). (a) Cholic acid, (b) Deoxycholic acid, (c) (*R*)-limonene, (d) (*S*)-limonene.

2.2 Materials and methods

2.2.1 Candidate enzymes

The genes tested during the present Master thesis are displayed in table 2.1:

Table 2.1: Candidate enzymes tested.

Candidate gene	CYP / Redox partner	Vector	Organism	Selection	His-tag
A0A1A9DXM0 (sgdx)	SGDX	pCDFDuet	*S.griseus.*	st	-
A0A1V2N1E9 (sce9)	SCE9	pCDFDuet	*S. cyaneofuscatus*	st	-
B1VU44 (sgg4)	SGG4	pCDFDuet	*S. griseus.*	st	-
bmr	BMR	pet28α	*B. megaterium*	kan	6x
camA/B	PdR/PdX	paCYC	*P. putida*	chl	-
CPR_{His}	CPR	pet22	*Candida apicola*	amp	6x
bm3	BM3	pet22	*B. megaterium*	amp	6x
bali	BaLi	pet22	*B. lichemiformis*	amp	6x
cyp3A4	CYP3A4	pCDFDuet	*Homo sapiens*	st	-
FG0101	FG0101	pet28α	*F. graminearum*	kan	-
FG022	FG022	pet28α	*F. graminearum*	kan	-
FG048	FG048	pet28α	*F. graminearum*	kan	-
FG081	FG081	pet28α	*F. graminearum*	kan	-
FG094	FG094	pet28α	*F. graminearum*	kan	-
FG110	FG110	pet28α	*F. graminearum*	kan	-
oleP	OleP	pet28α	*S.antibioticus*	kan	6x
WP_012377925	WP_012377925	pet28α	*S. griseus.*	kan	-
WP_012378285	WP_012378285	pet28α	*S.griseus.*	kan	-
WP_030573948	WP_030573948	pet28α	*S. cyaneofuscatus*	kan	-
WP_034313748	WP_034313748	pet28α	*A. lurida*	kan	-
WP_084493705 (ne05)	NE05	pCDFDuet	*Nocardia elegans*	st	-
WP_084494404	WP_084494404	pet28α	*Nocardia elegans*	kan	-
WP_091596541	WP_091596541	pet28α	*A. lurida*	kan	-

x1XP_020043035 (cc35)	cc35	pCDFDuet	*Castor canadensis*	st	-
XP_020043033 (cc33)	cc33	pCDFDuet	*Castor canadensis*	st	-

2.2.2 Growth medium preparation

The LB medium contaied 0.5% Yeast Extract (Condalab), 1.0% Tryptone (AcumediaLab) and 1.0% NaCl (Riedel-de Haën). For the preparation of LB-agar plates, the mixture additionally contained 1.5% Agar (Sigma Aldrich). dH_2O is added up to the corresponding volume and the solution was autoclaved. The TB medium was prepared using TB mixture (Carl Roth) and 0.4% glycerol (Sigma Aldrich). dH_2O was added up to the corresponding volume and the solution was autoclaved.

2.2.3 Single-transformation and co-transformation in *E. coli* with chemically competent cells

The reductases BMR and the CYP-candidates were tested, using OleP as positive control by transformation in *E. coli* BL21 (DE3), C41 (DE3), C43 (DE3) and Shuffle T7 Express. For the transformation, 1 µL of plasmid DNA (50–100 $ng \cdot \mu L^{-1}$) was added to each aliquot and the cells were incubated on ice for 1 h. For co-transformation of a CYP and redox system PdR/PdX, 1 µL of plasmid DNA was added also from the second plasmid, camA/B. Heat-shock followed, at 42°C for 40 s and the cells were incubated on ice for 2 min. Afterwards, 1 mL of pre-warmed LB medium (Carl Roth) was added and the cells were shaken vigorously at 37°C for 1 h. From the resulting cell culture, 0.1 mL were spread on LB agar plate with the corresponding antibiotic. The agar plates were incubated at 37°C for 24 h.

2.2.4 Single-transformation and co-transformation in *E. coli* with electro-competent cells

The CYPss were tested using OleP as a positive control (single and with the PdR/PdX redox system) by transformation in *E. coli* C41 (DE3) and C43 (DE3). For the electro transformation, 1 µL of plasmid DNA(50–100 $ng \cdot \mu L^{-1}$) was added to each aliquot and the cells were incubated on ice for 1 h. The electroporator device (Bio-Rad micropulser) was used for the transformation, where 0.1 mL of the cells were transferred in cuvettes and the program selected was the EC2. After using the device, 1 mL of pre-warmed LB medium was added immediately and the cells were shaken vigorously at 37°C for 1 h. From the resulting cell culture, 0.1 mL were spread on LB agar plate with the corresponding antibiotic and the plates were incubated at 37°C for 24 h.

2.2.5 Glycerol stocks

After the transformations, overnight cultures (Paragraph 2.2.6) were prepared in LB medium from a single colony and glycerol stocks were prepared the next day by adding 1 mL of the cell culture and 1 mL of glycerol 60% (glycerol in water (v/v)) under sterile conditions. The glycerol stocks were stored at -80°C.

2.2.6 Recombinant protein expression

A single colony from the single transformations in *E. coli* or 10 µL from the glycerol stock were cultured in a 5 mL overnight culture in LB using the corresponded antibiotic (with final concentrations of 50 µg/mL streptomycin , 50 µg/mL kanamycin, 100 µg/mL ampicillin, see table 2.1). After 16-18 h of cultivation, 50 mL cultures were prepared in TB medium (Carl Roth, in Greifswald) or LB medium (in Heraklion) (paragraph 2.2.2) at flasks with a ratio liquid: air, 1:4, with the corresponding antibiotic, at 160 rpm, 37°C, until OD_{600} = 0.6-0.8 was reached. Prior to induction, the cultures were cooled down to 20°C and inducer, as well as supplements, were added to the following final concentrations: 0.5mM IPTG, 0.3 mM $FeSO_4$, 84 µg/mL δ-aminolevulinic acid (ALA). The cultures were incubated at 20°C, 160 rpm, for 16-18 h (Infors HT Minitron/Multitron).

2.2.7 Harvest and cell lysis

The cultures were harvested at 6000 rpm (Thermo Scientific Labofuge 400R Heraeus), for 30 min, at 4°C and resuspended in sodium phosphate buffer (50 mM, pH =7.4). Cell lysis was accomplished by sonication, 40% power, 3 min of lysis (30 s lysis followed by 30 s rest on ice, Sonotrode Bandelein sonoplus). The soluble fraction was separated from cell debris 30 min centrifugation at 4000 rpm, 4°C. A 50 µL sample was collected from the supernatant for analysis by SDS-PAGE (Paragraph 2.2.11).

2.2.8 Protein purification with a Ni-NTA column

For purification using a Ni-NTA column with gravity flow, a 3 mL column was washed with dH_2O and Buffer A (50 mM K_{Pi} buffer (K_2HPO_4/KH_2PO_4), 300 mM NaCl, 10 mM imidazole, pH=7.4) and the supernatant after filtration with 0.2 µm filters was loaded on the column for purification. Afterwards, a flow-through sample was collected and the washing step was performed using Buffer A (50 mM K_{Pi} buffer, 300 mM NaCl, 10 mM imidazole, pH=7.4). A sample of the washing step was collected. The elution step followed with Buffer B (50 mM K_{Pi} buffer, 300 mM NaCl, 500 mM imidazole, pH =7.4) and 3 elution fractions of each 1.5 mL were collected (E1, E2, E3). All the samples mentioned during the purification procedure were tested for expression (Paragraph 2.2.11).

For purification using a Ni-NTA column and an ÄKTA device (Synthetic Biomaterials Lab, Assistant Prof. Kelly Velonia), three K_{Pi} buffers were prepared: Buffer A, containing 50 mM K_{Pi} and 500 mM NaCl pH =7.4, Buffer B with 50 mM K_{Pi}, 500 mM NaCl and 300 mM imidazole pH=7.4 and Buffer C containing 50 mM K_{Pi} pH =7.4. Buffers B and C were used to prepare an additional buffer of 15 mM imidazole, to be used as a lysis buffer and for the first washing steps. A 5 mL Ni-NTA column was primed by using 3 column volumes. The column equilibration was performed with 2 column volumes, the wash was performed in 5 column volumes (10% buffer B) and the elution with 100% buffer B in 4 column volumes. The flow of the column was determined at 1.5 mL/min.

After the protein purification, the protein solution was concentrated in 30 kDa concentrators and desalinized in the same concentrators by 3 washings with Buffer C. The procedure was performed at 6000 rpm, 4 °C (Labofuge 400R Heraeus, Thermo Scientific).

2.2.9 Cultivations with auto-induction protocol

The single transformations and the co-transformations in *E. coli* BL21(DE3), C41(DE3), C43(DE3) and Shuffle T7 express, were cultivated in 50 mL TB medium (Paragraph 2.2.2) with glucose 0.05% and lactose 0.2%, at 300 mL flasks, at 160 rpm, at 37°C for BL21(DE3), C41(DE3), C43(DE3) and 30°C for Shuffle T7 express. Supplements were added before the inoculation, to an end concentration of 0.3 mM $FeSO_4$ and 84 μg/mL ALA. After 3-5 h, the cultures were cooled down to 20°C and further incubated for 24 h.

2.2.10 SDS-Polyacrylamide Gel Electrophoresis (SDS PAGE)

The samples were prepared using 20 μL of sample, mixed with 20 μL loading buffer, incubated at 95°C for 15 min (Eppendorf thermomixer comfort) and centrifuged for 1 min, 13000 rpm (Thermo Scientific 17R, Heraeus). All samples and the protein marker (Unstained Molecular Weight Protein Marker, Thermo Scientific) were loaded on to the 12.5 % gel and run in the vertical electrophoresis apparatus from Amersham Pharmacia Biotech, with 25 mA per gel for 1 h. Afterwards, the gels were stained with SDS-PAGE staining (α-cyclodextrin, H_2O, phosphate acid, 5x Bradford solution) for 20 min.

The gels of the electrophoresis consists of 2 parts, the resolving gel and the stacking gel. The composition of both gels is shown in table 2.2. Rgarding the loading buffer, is consisted of 12% SDS, 6% β-mercaptoethanol, 30% glycerol, 0.05% servablue-G and 150 mM Tris/HCl pH =7.5.

Table 2.2: Components of one gel for SDS-PAGE

Components	Resolving gel (12.5 %)	Stacking gel (4 %)
Tris/HCl buffer, pH=7.4	2 mL	-
Tris/HCl buffer, pH=6.8	-	1 mL
AB mix 30%	3.33 mL	0.53 mL
dH_2O	2.67 mL	2.67 mL
APS (10%)	40 µL	40 µL
TEMED	4 µL	4 µL

2.2.11 Protein concentration

To measure CYPs' concentration, two different methods were used.

Based on the spectral properties, the spectrum which came up of the protein saturation with CO in the reduced state was used. A spatula tip of sodium dithionite was added to reduce the protein, from Fe^{3+} to Fe^{2+} (of the heme-iron) to a 2 mL sample of lysate or purified protein. One mL of the solution was transferred to a cuvette to be used as a reference, while the second mL of the sample was saturated with CO (1 bubble/s). Both samples' spectra were recorded at the range of 400-500 nm. Properly folded CYPs' concentration was determined with the following formula [10]:

Equation 2.1: Formula for proper folded protein concentration

$$[(A_{450} - A_{490})_{sample} - (A_{450} - A_{490})_{blank}] /0.091 = nmol_{CYP}/mL$$

For the misfolded CYP, the formula is the following:

Equation 2.2: Formula for misfolded protein concentration

$$[(A_{420} - A_{490})_{sample} - (A_{420} - A_{490})_{blank}] /0.091 = nmol_{CYP}/mL \text{ [10]}$$

The protein concentration was measured also using the Bradford method. Bovine serum alvumin (BSA) solutions with concentrations 0.1 mg/mL to 1mg/mL were used for the standard curve. The Bradford staining solution was prepared using 0.1 g Brilliant Blue G250 dissolved in 50 mL ethanol, 100 mL 85 % phosphoric acid and 100 mL dH_2O. Afterwards, the staining was dissolved with dH_2O in a volume ratio 5 mL Bradford solution: 10 mL H_2O and filtered with 0.2 µm filters. The samples for the standard curve and the unknown concentration sample were prepared by mixing 200 µL of Bradford staining and 15 µL sample in triplicates. The unknown samples

were diluted based on the standart curve and the absorption was measured at 595 nm using photometer (Thermo Scientific, Multiskan Sky).

2.2.12 DNA extraction

The GeneJET Plasmid Miniprep Kit #K0502 was used. An overnight culture (paragraph 2.2.6) of 5 mL volume was prepared with the the *E. coli* containing the plasmid of interest. The biomass was harvested and resuspended in 250 µL of resuspension buffer. The solution was transferred in a 1.5 mL tube, gently shaked and 350 µL of the equilibration buffer were added. After a 5 min centrifugation at 13000 rpm the supernatant was transferred by pipetting in the GeneJET column and washed with 1 mL of the washing buffer. The spin GeneJET column was transferred in a 1.5 mL tube and 50 µL of the elution buffer were added, incubated in room temperature for 2 min, and centrifuged for 2 min, at 13000rpm. The column was discarded afterwards and the plasmid concentration was measured at 260 nm using 2 µL of the plasmid solution on the µDrop plate, using the photometer Multiskan Sky (Thermo Scientific). The extracted plasmid was stored at -20°C.

2.2.13 Biotransformation in whole-cell catalysis

After the co-expression of the P450-redox partner system and the P450-BM3, P450-BaLi, the cells were harvested by centrifugation at 6000 rpm, for 30 min, at 4°C. The pellets were resuspended with biotransformation buffer (200 mM K_{Pi} buffer pH =7.4, 20 mM NaCl, 1% glucose, 0.4% glycerol) under sterile conditions after normalization up to OD_{600} =30. The cultures with the resting cell medium were transferred in sterile flasks (liquid:air ratio 1:5). One mL of culture was taken as a sample reference. Afterwards, the steroid substrate was added, to an end concentration 4 mg/mL to initiate the reactions. The mixtures were incubated for 5 min and 1 mL sample was collected from each flask as time zero* (0*). The flasks were incubated for 24 h at 20°C, 160 rpm (Infors HT Unitron) and 1 mL sample was collected from each flask after 24 h and 48 h.

The 1 mL samples taken during the whole-cell catalysis, were extracted with 1 mL of ethyl acetate added to each sample, followed by mixing for 15 s. The samples were centrifuged for 3 min at 13000 rpm and the two phases were separated. The upper, organic phase was transferred to a new reaction tube. The steps were repeated by adding ethyl acetate again and the organic phases were joined. Then, 1 spatula tip of anhydrous Na_2SO_4 was added to the tube with the organic phase, to dehydrate the sample. The samples were centrifuged at 13000 rpm for 1 min and the liquid was

transferred to a new reaction tube. After evaporation, the dried samples were stored at -20°C until further use.

2.2.14 Thin Layer Chromatography (TLC)

For TLC analysis, 200 µL of ethanol were added at each sample, and the samples were mixed by vortexing and centrifuged at 13000 rpm, for 1 min (Thermo Scientific 17R, Heraeus). The TLC chamber was prepared (mobile phase) in 5 mL total volume with the ratio chloroform: acetone: acetic acid being 90:15:1. Then, 2 µL of each sample were loaded onto the TLC plates and the plates were put into the chamber. After the development of the plates within the chamber, substrate and product were stained with Cerium (IV) molybdate stain (3.65 g $Ce(SO_4)_2 \cdot 4H_2O$, 7.5 g $H_3[P(Mo_3O_{10})_4] \cdot xH_2O$, 22.5 mL H_2SO_4). The plates were dried and heated until substrates and reference compounds were visible.

2.2.15 High-Performance Liquid Chromatography (HPLC)

In the Department of Biotechnology and Enzyme Catalysis, in the University of Greifswald, the dried samples collected after the extraction and the evaporation (paragraph 2.13), were dissolved in 200 µL ethanol 100%. Then 20 µL of the samples were analyzed using a Hitachi LaChrom Elite HPLC System (Hitachi High-Technologies, Krefeld, Germany) with a Luna® Omega 5 µm/100 Å PS C18 column (Phenomenex, Aschaffenburg, Germany). The mobile phase was acetonitrile:water (50:50) % (v/v) containing 0.1% TFA. An isocratic method (1 mL/min) was employed at room temperature. The bile acids were detected using a LaChrom Elite L-2490 RI Detector (Hitachi High-Technologies, Krefeld, Germany).

In the Department of Chemistry, University of Crete, analysis was performed in a Shimadzu Prominence HPLC system equipped with SPD-M20A diode array detector and LICHOSPHERE RP18-5 column (Supelco), using as mobile phase acetonitrile:water (50:50) % (v/v) containing 0.1% acetic acid.

DCA was detected at 208 nm and CA at 212 nm.

2.2.16 Photometric analysis

Greiner 96-well plates were used and all the reactions were tested with purified enzymes in triplicates, at 340 nm. Using steroid as substrates, the well contained K_{Pi} buffer (50 mM, pH =7.4), 0.5 mM NADPH, 1 µM of protein and 0.2 mg/mL of substrate (stock dissolved in dH_2O) in a final reaction volume of 200 µL. The reaction was performed in photometer (Multiskan Sky, Thermo Scientific) at 20°C, for 30 min, with 10 s measurement intervals.

Using (*R*) or (*S*)-limonene as substrates, the well contained K_{Pi} buffer (50 mM, pH =7.4) 0.5 mM NADPH, 0.5 µM of protein and 1.3 mM of substrate (stock dissolved in DMSO), in a final reaction volume 200 µL. The "blank" measurement was performed under the same conditions without the enzyme, in a final reaction volume of 200 µL.

Enzyme activity was calculated using the Equation 2.3.

Equation 2.3: Formula for the calculation of the enzyme activity in Units (U) per mL.

$$\frac{\mathbf{U}}{\mathbf{mL}} = \frac{\text{Slope of the enzymes reaction } (\frac{\Delta A}{s})}{\text{Slope of the standart curve } (\frac{\Delta A}{mM})} \cdot \frac{60\ s}{1\ \text{min}} \cdot \text{Df}$$

Where D_f is the dilution factor before the measurement multiplied by the volumetric dilution in the microtiter plate.The NADPH consumption was calculated using the NADPH standard curve. From the linear equation the NADPH concentration can come up in t=0 and in t=final and then the NADPH consumption equals to (Equation 2.4):

Equation 2.4: Formula for the calculation of the NADPH consumed in µM/min.

$$\frac{C_{NADPH,t=0} - C_{NADPH,t=final}}{\text{Reaction time } [min]} = \text{NADPH consumed } (\frac{\mu M}{min})$$

2.2.17 Bioinformatic analysis

The 3D structures of the BM3 (PDB ID: 3EKF) and OleP (PDB ID: 5MNS) were retrieved from PDB [96]. The MolProbity web service was used for the evaluation of the protein structures studied [97]. The substrates were initially drawn in 2D with the ChemDraw molecular editor and modified using YASARA (v.20.4.24) to get their 3D structures [98].

YASARA was also used for energy minimization, structure refinement, and docking experiments were performed using it.

2.2.18 Mutagenesis Polymerase Chain Reaction (PCR)

The bioinformatic analysis provided some hints for rational mutagenesis of OleP. OleP gene was cloned in vector pET-28a(+), with an N-terminal 6xHistag.

The primers were designed in GENEious software (v 9.1) and ordered from Eurofins genomics.

Table 2.3: Name and sequence of the mutagenesis's primers.

Primer name	Sequence
OleP_P88H_forward	5'-CGA CCC CGC ATG AAC C-3'
OleP_P88H_reverse	5'-GCT GGG GCG TAC TTG G-3'
OleP_R82F_forward	5'-CCG GCA ACC CCG TTT ATG TTC C-3'

OleP_R82F_reverse	5'-GGC CGT TGG GGC AAA TAC AAG G-3'
OleP_F296R_forward	5'-G CTC GCG TGT TCG TGT CG-3'
OleP_F296R_reverse	5'-CGA GCG CAC AAG CAC AGC-3'
OleP_R82V_forward	5'-GCA ACC CCG GTG ATG TTC C-3'
OleP_R82V_reverse	5'-CGT TGG GGC CAC TAC AAG G-3'
OleP_V291K_forward	5'-CCG CTG AAA TCA GCG GGC TCG-3'
OleP_V291K_reverse	5'-GGC GAC TTT AGT CGC CCG AGC-3'

For the mutagenesis PCR, the plasmid pET-28a(+) with OleP gene (wild type) was used as template for the polymerase chain reaction (PCR). The reaction mixture was as shown in table 2.4 and the program used for the PCR was shown in table 2.5:

Table 2.4: Composition of each polymerase chain reaction.

Component	Final concentration	1 reaction volumes (µL)
Sterile MQ water		17.5
5x HF buffer (Minotech)	1x	5
Primer mix (10 µM each)	0.2 µM each	0.5
Template (10 ng/µl)	0.2 ng/µL	0.5
DMSO	2% (v/v)	0.5
dNTPs (10 mM)	0.2 mM	0.5
DNA polymerase 2u/µL, 0.5 KU (Minotech)		0.5
Total volume		~25 µL

Table 2.5: Polymerase chain reaction program used mutagenesis of OleP.

Step	Temperature (°C)	Time (min:sec)
	95	05:00
25 cycles:		
Denaturation:	95	00:30
Annealing:	60	00:30
Elongation:	72	03:35
2x elongation duration	72	07:00
Cooling	10	∞

After detecting product with DNA electrophoresis (paragraph 2.2.19), 1 µL DpnI was added to the PCR product to completely digest the methylated DNA.

2.2.19 Agarose gel electrophoresis

Agarose gel electrophoresis was used for the separation of DNA molecules after PCR reactions. A 0.8% (w/v) agarose gel was prepared (70 mL gel volume) by dissolving 0.56 g of agarose in 1x TE buffer (from a 10x stock, 100 mM Trismabase, pH =8,

100 mM EDTA). The mixture was completely dissolved by heating in the microwave and afterwards cooled down to about 60 ^{o}C. One drop of ethidium bromide (EtBr) was added in the mixture before pouring in the electrophoresis apparatus for the gel to solidify. Then, 100 mL of running buffer were added in the device (1x TE buffer, pH =8), to cover the gel.

The samples were prepared by transferring in a tube 5 μL of the sample and 2 μL of the loading solution. In each well 6 μL of this mixture were loaded and in another well were loaded 5 μL DNA ladder. The gel run at 90 V for 30 min and the gel was observed on a UV table.

CHAPTER 3. CYPs expression, purification and folding

3.1 Introduction

CYPs have been widely studied as they can convert a broad range of substrates and perform a variety of reactions. Their catalytic system has centered the scientific interest and their expression using *E. coli* as expression hosts was applied in several studies and for different applications [12, 13, 15, 29, 45, 57, 83].

In the present master thesis, specific plasmid vectors, targeted genes, *E. coli* strains and several chemicals were used for the transformation experiments and the proper folding of the proteins, during the heterologous expression. Several challenges were faced and overcame, as CYPs are unstable, and additionally, the necessity of redox partner systems can be a major obstacle during the expression.

3.1.1 Plasmid vectors and gene targets

The genes of the CYP candidates used during this master thesis, except for CYP102A1 (BM3) from *B. megaterium*, and CYP102A7 (BaLi) from *B. licheniformis,* which are self-sufficient, were selected after bioinformatics analysis.

The bioinformatic analysis performed was first and second-generation sequence similarity network (2^G SSN) from Dr. Emil Hamnevik and Dr. Thomas Bayer, from the Department of Biotechnology and Enzyme Catalysis at the Institute of Biochemistry, University of Greifswald and focused on the selection of specific cytochrome genes, for the specific hydroxylation of steroids of biotechnological interest. The cytochrome genes chosen are presented in table 2.1. The plasmid vectors used for each gene are also displayed in table 2.1, together with the corresponding antibiotic resistance.

3.1.2 Bacterial expression systems and *E. coli* strains

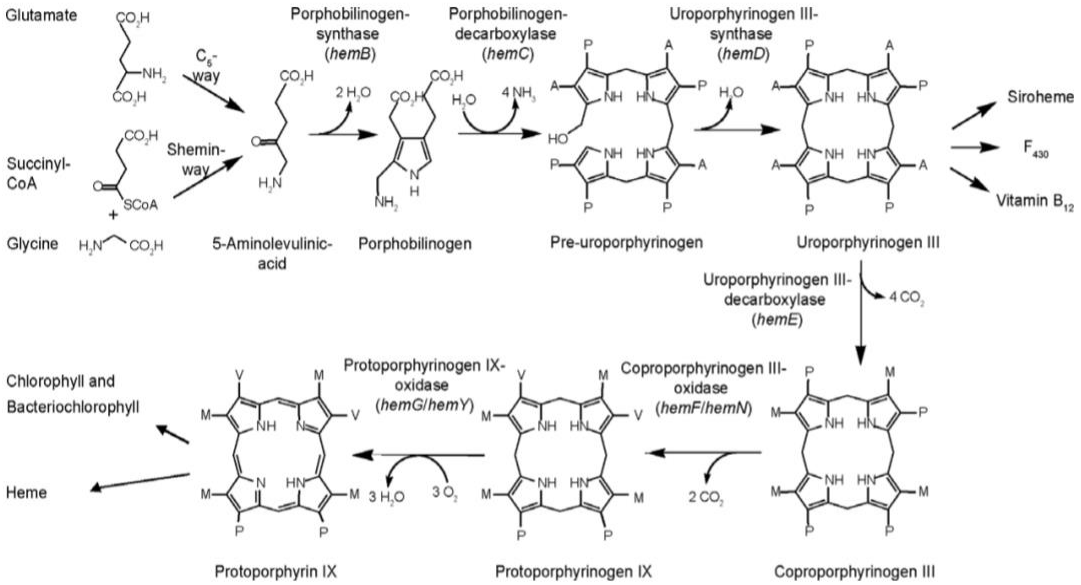
During this thesis, the expression of the CYPs was optimized by testing different *E. coli* strains, and different growth media. These enzymes are considered to be toxic proteins for the bacterial *E. coli* hosts, and thus, the C41(DE3) and C43(DE3) *E. coli* strains, derived from the BL21(DE3), were used, because of their ability to overcome the toxicity [99].

The reason of the CYPs' toxicity remains unclear but is hypothesized that it might be due to their activity and the effects at the cell-host metabolism [100].

3.1.3 Challenges in CYP expression and application

Several CYPs are not stable as proteins and a suitable buffer at appropriate pH is important to be used. For CYPs, usually K_{Pi} buffer is used with pH=7.4-7.5 [15, 36, 54, 66]. The expression conditions are specific and require additional chemicals such as aminolevulinic acid (ALA) as precursor for the heme synthesis. Particularly, in *E. coli* heme can be synthesized as there are heme-containing enzymes expressed [19, 101, 102], but the addition of ALA and $FeSO_4$ can enhance the heme synthesis in the active site and work as heme precursors, which facilitates the proper folding [19, 103].

Heme is built and degraded continually, and its synthesis is depended and balanced by the rate of its disposal. This is crucial, because heme, in free form and not bound to some heme-dependent protein, is a pro-oxidant agent and toxic. Therefore, its synthesis and degradation are regulated. Simultaneously, the term “free heme” indicates that it is loosely bound to intracellular proteins and not committed to specifically to a particular hemoprotein [104]. The heme biosynthetic pathway is displayed in Scheme 3.1



Scheme 3.1: The biosynthetic pathway of heme. The synthesis of heme requires ALA and several enzymes which participate in the process of the biosynthetic pathway. The enzymes are labeled above the arrows. Reprinted by permission from Elsevier. [Journal Publicer: Elsevier, Journal name: Archives of Biochemistry and Biophysics], reference [105], The biochemistry of heme biosynthesis, I. U. Heinemann, [COPYRIGHTS] 2008.

Regarding the redox partner system, it does not enhance the expression and the application of CYPs because an additional protein complex is used [103]. As a

consequence, either co-expression should be performed, or the redox system is at the same polypeptide chain with the CYP as fused protein.

On the other hand, sometimes proteins cannot be expressed in the soluble fraction and are found as inclusion bodies. In case of high toxicity levels while expressed, other bacterial strains are used, as C41(DE3) and C43(DE3) which were developed to express toxic proteins [99].

Lastly, during enzymatic reactions, co-factors as NAD(P)H are necessary [2] and the substrates used are often characterized by their high toxicity and limited solubility [12, 45].

3.2 Results and Discussion

In the presented thesis, twenty different CYP genes were investigated (olep, cc33, cc35, cyp3A4, sgdx, sce9, sgg4, ne05, WP_091596541, WP_034313748, WP_084494404, WP_030573948, WP_012378285, WP_012377925, FG081, FG022, FG048, FG094, FG0101, FG110, Table 2.1) together with 3 different redox partners systems (PdR/PdX from *P. putida*, CPR_{His} from *Candida apicola* and BMR from *B. megaterium*, Table 2.1) for expression and protein proper folding. Additionally, the fused CYP genes, *bali* and *bm3* were investigated, as well.

Only the enzymes OleP, BaLi and BM3 were purified, because of the additional 6xHis-tag. The absence of the His-tag at the rest of the enzymes did not allow a purification.

3.2.1 Expression, purification and folding tests of the fused CYPs BM3 (CYP102A1) and BaLi (CYP102A7)

The CYPs BM3 (CYP102A1) from *B. megaterium* and BaLi (CYP102A7) from *B. licheniformis,* were studied in Enzyme Technology Laboratory, Department of Chemistry, University of Crete, Greece.

The CYP OleP has always been used as the positive control, as all the protocols have been established by Sascha Grobe, Biotechnology and Enzyme Catalysis group, Institute of Biochemistry, University of Greifswald, Germany [15].

First, BaLi and BM3 were expressed using *E. coli*, C41(DE3) and C43(DE3) strains, respectively, as expression hosts. As displayed in figure 3.1 (a), the signal for BaLi in *E. coli*, C41(DE3) can be visualized in the after induction sample, in comparison to the before induction sample (compare lanes 2, 3, figure 3.1 a) at the height of 121 kDa, indicating the successful expression. Also, the protein can be noticed in the pure protein fractions, before and after desalting however there is a signal in the pellet

fraction. P450-BM3 is expressed in *E. coli*, C43(DE3), as shown in figure 3.1 b where it can be seen by the signal in the after induction fraction at the height of 116 kDa. Also, the expressed BM3 can be visualized in the pure protein fractions, before and after desalting and in the pellet fraction, indicating that either *E. coli* were not completely lysed with the sonication, or a fraction of the enzyme is expressed in insoluble form.

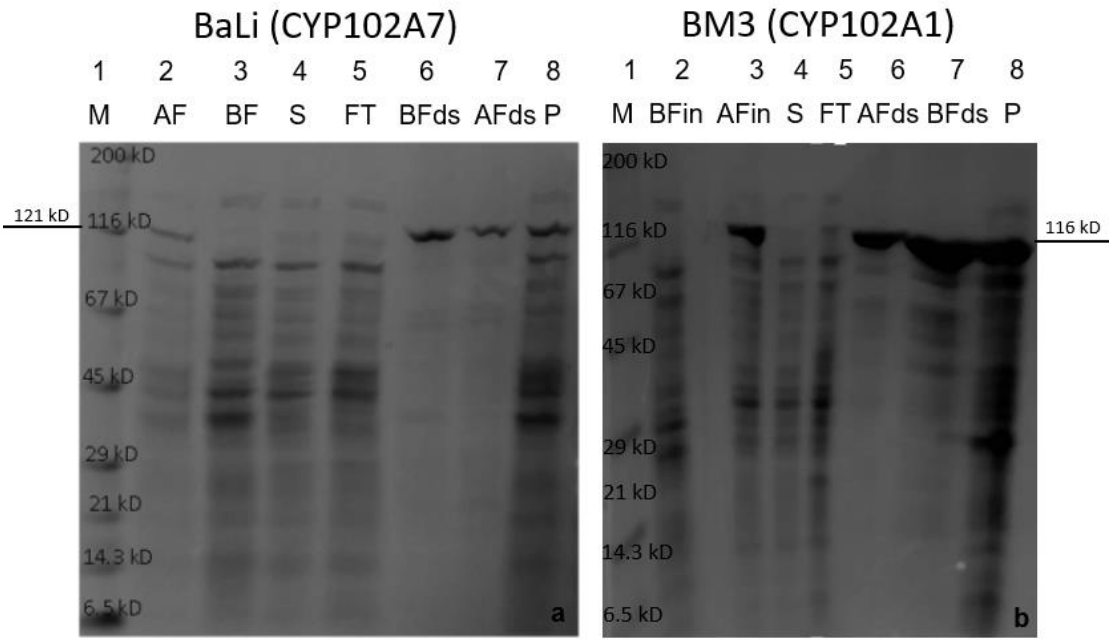


Figure 3.1: 10 % SDS-PAGE of (a) BaLi (CYP102A7) and (b) BM3 (CYP102A1) in *E.coli*, C41(DE3) and C43(DE3), respectively. (a): The samples are denoted as follows: marker (M), after induction (AF), before induction (BF), supernatant after cell lysis using sonification (S), flow-through (FT), pure protein before desalting (BFds), pure protein after desalting (AFds) and pellet after sonication (P).

To confirm the proper folding of the enzymes, the CO-assay was performed. As displayed in Figure 3.2, BM3 is expressed in properly folded form (figure 3.2 a). On the other side, BaLi, has a small signal at 420 nm, indicating a fraction of the purified protein is misfolded.

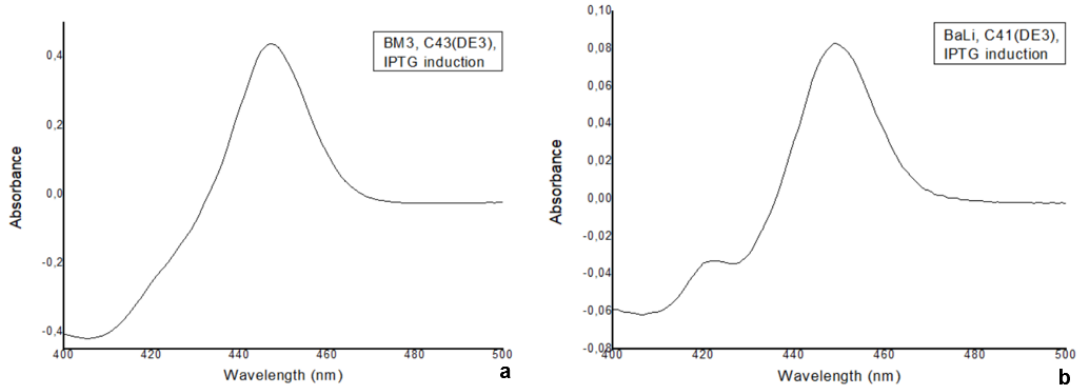


Figure 3.2: Spectra of the CO assay for the CYPs BM3 and BaLi in *E.coli* BL21(DE3), C41(DE3) and C43(DE3), respectively. On the x-axis the wavelength is displayed from 400 nm to 500 nm and on the y-axis the absorbance is displayed ranging from 0.1 to 0.6 (a), 0.04 to 0.12 (b).

To calculate BM3's concentration, equation 2.1 was used, and the concentration for properly folded BM3 was calculated to be 4.74 µM.

Regarding BaLi, the enzyme was not as much concentrated compared to BM3, as the folded form was calculated to be 0.92 µM and the misfolded 0.02 µM.

3.2.2 Expression, purification and folding tests of the CYPs

Regarding the positive control OleP, it was expressed using IPTG as inducer (paragraph 2.2.6), as well as auto-induction media (paragraph 2.2.9), in different *E.coli* strains (BL21(DE3), C41(DE3), C43(DE3) and Shuffle T7 Express). The expression was performed after the single and co-transformation of the OleP with the redox partner systems PdR/PdX and CPR_{His}.

As displayed in figure 3.3, the expression of OleP in *E. coli* BL21(DE3), using IPTG as inducer, and its purification were successful. This is indicated by the band at 46 kDa in the after induction sample when compared to the pre-induction sample. The cell lysis by sonication seems to be a suitable for OleP, as a clear band is visible in the soluble fraction corresponding to the size of OleP, although a signal of similar intensity is visible in the pellet fraction. In the flow-through fraction only a small band is visible at 46 kDa, indicating that not all targeted protein binds to the column. In the wash fraction, no OleP was found and clear bands are obtained in the elution fractions E1-E3. With this, the expression and purification of OleP in *E. coli* BL21(DE3), using IPTG as inducer, was successful. The expression of OleP's using auto-induction medium (paragraph 2.2.9), as shown in figure 3.3b, was also succesful.

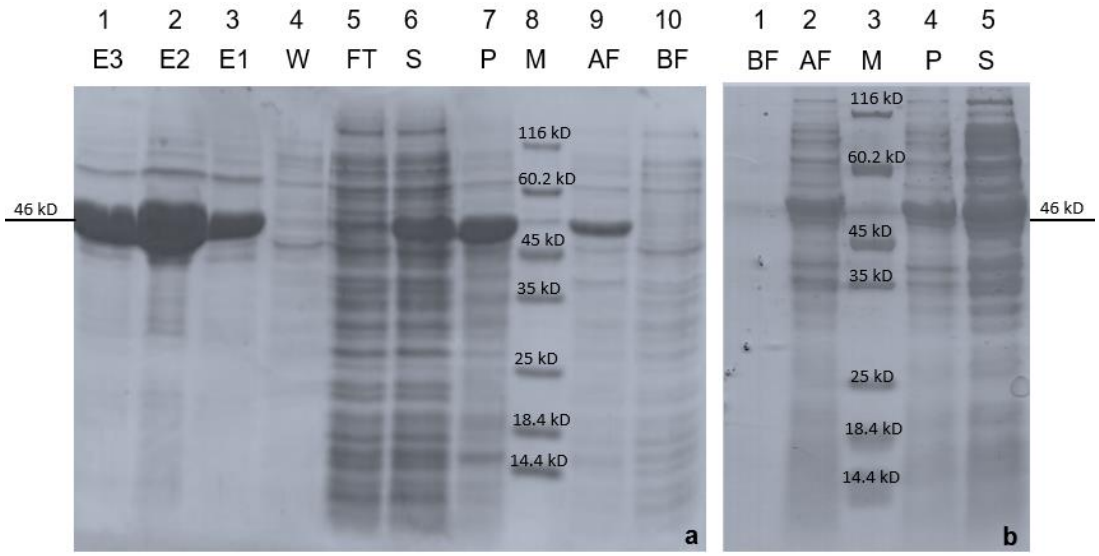


Figure 3.3: 12.5 % SDS-PAGE of OleP in *E. coli*, BL21 (DE3), using IPTG as inducer (a) and auto-induction medium (b). (a): The samples are denoted as follows: elution 3 (E3), elution 2 (E2), elution 1 (E1), wash (W), flowthrough (FT), supernatant after sonification (S), pellet after sonification (P) marker (M), after induction (AF), before induction (BF).

Similarly, OleP's expression as single and co-expressed with the PdR/PdX system and also with CPR_{His} redox-partner system was investigated in BL21(DE3), C41(DE3), C43(DE3) and Shuffle T7 Express, using IPTG or auto-induction and confirmed to be successful.

The folding of OleP was verified using the CO-assay. As displayed in figure 3.4, from the CO spectra, OleP and OleP-camA/B can be expressed and are folded properly in the strains BL21(DE3) (figure 3.4 (a) and (b)), C41(DE3) (figure 3.4, (c) and (d)) and C43(DE3) (figure 3.4, (e) and (f)) using IPTG as inducer but also with auto-induction (C41(DE3), figure 3.4, (g)). The folding results were determined from the Soret band observed at 450 nm, at the spectra in figure 3.4.

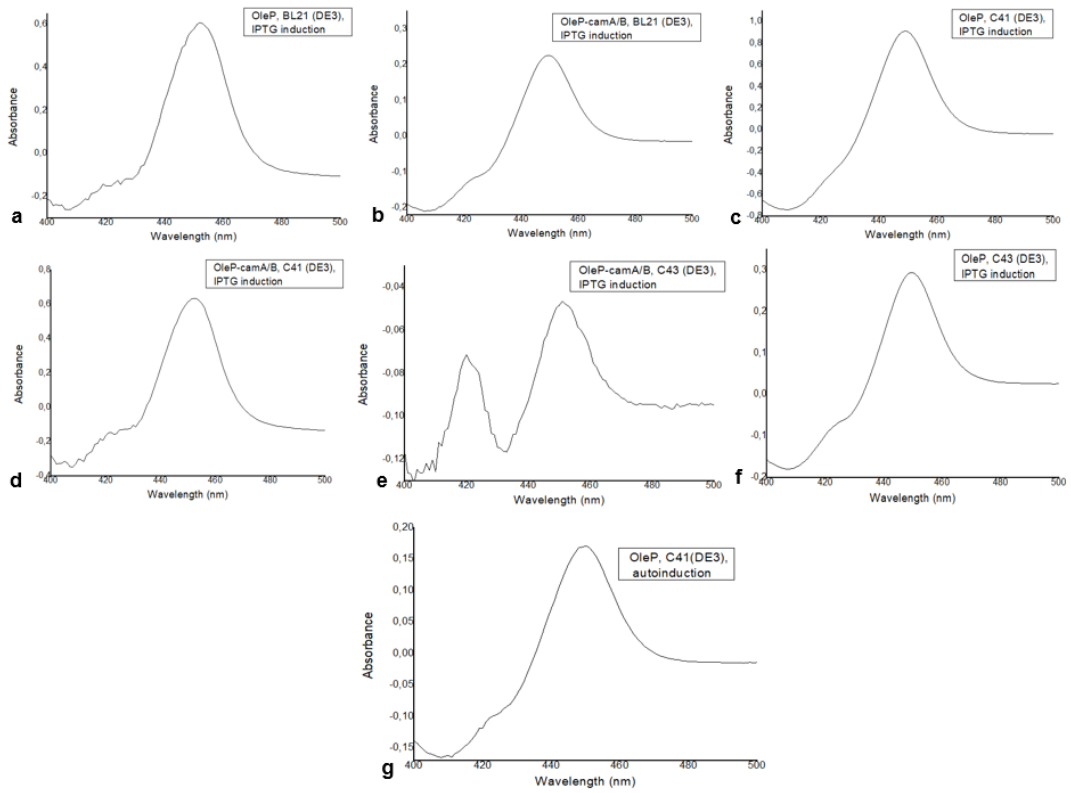


Figure 3.4: Spectra of the CO assay for OleP and OleP-camA/B using IPTG induction, in *E. coli* BL21(DE3) ((a) and (b)), C41(DE3) ((c) and (d)), C43(DE3) (e) and (f) and using auto-inductionin C41(DE3) (g). On the x-axis the wavelength is displayed from 400 nm to 500 nm and on the y-axis the absorbance is displayed ranging from -0.2 to 0.6 (a), -0.2 to 0.3 (b), -0.8 to 1.0 (c), -0.4 to 0.8 (d), -0.12 to -0.04 (e), -0.3 to 0.3 (f) and -0.15 to 0.20.

In table 3.1 the concentration of properly folded OleP samples are displayed.

Table 3.1: Concentrations of the OleP and OleP-camA/B samples in the *E. coli* strains BL21(DE3), C41(DE3), C43(DE3).

#	Sample	Concentration (µM)
a	OleP in BL21(DE3), IPTG induction	7.57
b	OleP-camA/B in BL21(DE3), IPTG induction	2.58
c	OleP in C41(DE3), IPTG induction	9.57
d	OleP-camA/B in C41(DE3), IPTG induction	8.35
e	OleP-camA/B in C43(DE3), IPTG induction	2.88
f	OleP in C43(DE3), IPTG induction	1.56
g	OleP in C41(DE3), auto-induction	2.02

Using the successful experiments of OleP as a positive control, the P450 genes cyp3A4, cc33, cc35, sgdx, sce9, sgg4, ne05, WP_091596541, WP_034313748, WP_084494404, WP_030573948, WP_012378285, WP_012377925, FG081, FG022, FG048, FG094, FG0101 and FG110 (Table 2.1) with and without the redox partner systems PdR/PdX and CPR_{His}, were also investigated for expression and further analyzed for proper folding via their CO difference spectra.

Particularly for the CYP genes cyp3A4 (*Homo sapiens*), cc33 (*Castor canadensis*) and cc35 (*Castor canadensis*), the expression and folding were challenging, and as a consequence, different *E. coli* strains and different induction methods were used (IPTG induction (paragraph 2.2.6) and auto-induction (paragraph 2.2.9)). As described in table 3.2, the CYPs CYP3A4, cc33 and cc35 were successfully expressed in the *E. coli* strains BL21(DE3), C41(DE3), C43(DE3) using IPTG but in a denatured form. On the other hand, using the auto-induction method, only CYP3A4 could be expressed, but again, misfolded. These results were determined using SDS-PAGE 12.5%, comparing the signals before and after induction at the corresponding molecular weights (57.3 kDa for CYP3A4, 56.6 kDa for cc33 and 56.3 kDa for cc35). The denatured form of the enzymes was determined by the Soret band at 420 nm instead of the 450 nm in the CO difference spectra, and it was noticed that the redox partner PdR/PdX did not affect the results, as the enzymes CYP3A4, cc33 and cc35 had the same folding profile as single expressed proteins, using the same *E. coli* hosts and induction methods, compared to the co-expression with the redox system .

Table 3.2: Expression of the CYPs in differents *E. coli* strains, with the PdR/PdX system.

P450	**Organism**	***E. coli* strain**			
		BL21 (DE3)	**C41 (DE3)**	**C43 (DE3)**	**Shuffle T7 Express**
CYP3A4	*Homo sapiens*	+	+	+	_
cc33	*C. canadensis*	+	+	+	_
cc35	*C. canadensis*	_	+	+	_

The expression of the CYPs SGG4, SCE9, SGDX and NE05 was tested, similarly to the positive control enzyme, OleP, together with the PdR/PdX and CPR_{His} redox partner systems, using IPTG for induction. After the evaluation of the data obtained, the expression method was determined to be sutitable as an expression method. The expression and folding results are displayed in table 3.3:

Table 3.3: Expression and folding results of the CYPs, as co-transformed in *E. coli*, with the PdR/PdX and the CPR_{His} redox-partner system.

P450	**Organism**	**Expressed with:**		***E. coli* strain**	**Folded (co-expressed with)**		**Denatured (co-expressed with)**	
		PdR/PdX	**CPR_{His}**		**PdR/ PdX**	**CPR_{His}**	**PdR/PdX**	**CPR_{His}**
SGG4	*S. griseus*	+	+	BL21(DE3)	_	_	0.15 µM	0.21 µM
SGDX	*S. griseus*	+	+	BL21(DE3)	_	_	0.73 µM	0.61 µM
SCE9	*S. cyaneofuscatus*	+	+	BL21(DE3)	0.24 µM	0.23 µM	0.06 µM	0.002 µM
NE05	*N. elegans*	+	+	C43(DE3)	_	1.50 µM	0.74 µM	-

As described in table 3.3, the CYPs SGG4, SGDX and SCE9 were successfully expressed in the *E. coli*, BL21(DE3), and NE05 in C43(DE3) using IPTG, but in a denatured form. On the other hand, using the auto-induction method, only SCE9 together with the PdR/PdX system and SCE9 and NE05 together with the CPR_{His} system were in the properly folded form. These results were obtained again by the SDS-PAGE 12.5%, by the comparison of the signal at before and after induction fractions at the corresponding molecular weights (46.3 kDa for SGDX, 57.1 kDa for SGG4, 44.8 kDa for NE05 and 44.9 kDa for SCE9). The CO-folding tests, with, the signal at 450 nm proved the proper folding of the active site for SCE9 and NE05, respectively.

In parallel though, the CYPs SGG4, SGDX, SCE9 and NE05 presented a similar folding profile when expressed as single proteins to the one as expressed with the CPR_{His} system (Tables 3.3, 3.4). Supplementary, twelve more enzymes (WP_091596541, WP_034313748, WP_084494404, WP_030573948, WP_012378285, WP_012377925, FG081, FG022, FG048, FG094, FG0101, FG110, Table 2.1) were tested for expression and folding, after single-transformation in *E. coli* BL21(DE3). The analysis proved that four of them seemed to be successfully expressed and properly folded. The rest of the candidates were either insoluble or not expressed at all. The results of the folding tests are displayed in table 3.4.

According to the data obtain from the CO-difference spactra (tables 3.3 and 3.4) the co-expression of the PdR/PdX system does not favor the proper folding of the active site of SGG4, SGDX and NE05, compared to the CPR_{His} system, but the reason remains unknown. The rest of the candidates were not tested together with a targeted redox-partner system but after single-expression, they were indicated to be partially

properly folded and partially denatured. The spectra of these results can be found in Chapter 8. Appendix, figure 8.1.

Table 3.4: Expression and folding results of the CYPs, in *E. coli*.

P450	Organism	Folded P450 (μM)	Misfolded P450 (μM)
SGG4	*S. griseus*	0.11	0.45
NE05	*N. elegans*	0.53	-
SCE9	*S. cyaneofuscatus*	1.77	0.04
SGDX	*S. griseus*	-	1.52
WP_091596541	*A. lurida*	0.208	0.090
WP_084494404	*N. elegans*	0.516	0.549
WP_030573948	*S. cyaneofuscatus*	0.334	0.711
FG110	*F. graminearum*	0.030	0.681

3.3 Conclusions

As presented in Chapter 3, 20 CYP constructs were tested for expression and folding, using different *E. coli* strains and induction methods.

The fused self-sufficient cytochromes BM3 and BaLi were successfully expressed and folded (figures 3.1 and 3.2), and they can be further used for enzymatic reactions with target substrates.

OleP was used as a positive control, as it is already established from previous work of Sascha Grobe, Biotechnology and Enzyme Catalysis group, Biochemistry Institute, University of Greifswald, Germany. As a consequence, all the targeted CYP genes, which were chosen after 2nd generation sequencing analysis, from Dr. Emil Hamnevik and Dr. Thomas Bayer, were handled similarly to OleP. According to the expression and folding tests performed, only the SGG4 (*S. griseus*), SCE9 (*S. cyaneofuscatus*), NE05 (*N. elegans*), WP_091596541 (*A. lurida*), WP_084494404 (*N. elegans*), WP_030573948 (*S. cyaneofuscatus*) and FG110 (*F. graminearum*) were expressed in a proper folded form and further investigated for reactions with the substrates of interest.

CHAPTER 4. Whole-cell biocatalysis with steroids substrates

4.1 Introduction

CYPs tested for expression and folding in Chapter 3 were used in whole-cell biocatalysis with a steroid substrate, for evaluating their potential activity and selectivity.

Acording to Chapter 3, 8 out of 20 CYPs were expressed in a properly folded form. As the reason for the denaturation of the rest of the candiates has not been determined, and as it can be that the cell lysis method could be a reason for their inactive form, all biotransformations with the candidates were performed in whole-cell systems.

4.1.1. Biocatalysis in whole-cell systems

The advantage of the whole-cell catalysis as method, is that cell lysis and protein purification are not required and, at the same time, the host cell provides a stabilizing environment to the overexpressed enzyme [5].

In addition, the NADPH is necessary for reaction processes with oxidoreductases as natural catalysts, as these enzymes are nicotinamide cofactors-depended. Using whole cells, the cofactors can be regenerated by the host cell metabolism [5, 106].

4.1.2 Oxidizing agents

It is known that CYPs use a redox partner system for electron transfer from the nicotinamide cofactor to a reductase, and then, to the heme.

However, there are chemicals acting as oxidizing agents. Sodium periodate, diacetoxyiodo benzene and 3-chlorobenzoic acid are able to act as electron donors, in the place of NADPH [107]. As a result, there is no requirement for redox partner system, and the experimental procedure during the biocatalysis can be easier and less challenging [107, 108]. Gustaffson *et al.*, could prove with partially purified cytochromes from liver microsomes and from *B. megaterium* that steroid hydroxylation was feasible in the absence of NADPH and molecular oxygen, only with the presence of $NaIO_4$ [107].

Scheme 4.1: The oxidizing agents used for the reaction of the P450 candidates with the steroid substrate. The molecular structures are presented of (a) sodium periodate, (b) 3-chlorobenzoic acid, (c) (diacetoxyiodo) benzene.

4.2 Results and discussion

All the reactions performed with the steroid of interest in whole-cell biocatalysis (paragraph 2.2.13) were compared to the reactions performed with the positive control, which was OleP co-expressed together with the PdR/PdX redox partner system (OleP-camA/B). Reactants were extracted as described in paragraph 2.2.13 and the analysis was performed with TLC according to paragraph 2.2.14 and HPLC, according to paragraph 2.2.15.

4.2.1 Chromatographic analysis of the OleP with the steroid of interest

The reactions were evaluated by analyzing samples before adding the substrate into the reaction (background), 5 min of incubation after adding the substrate (time zero), and after 24 h and 48 h of reaction. As shown in figure 4.1, from 0 h time to 24 h time only the band of the substrate is visible. However, at 48 h, there is a small band of product, while the band of the substrate was still at a similar size. This is the first indication that OleP with the PdR/PdX system in BL21(DE3) host cells can accept the steroid of interest. The same results were obtained also when OleP was co-expressed with the PdR/PdX system in C41(DE3) and C43(DE3).

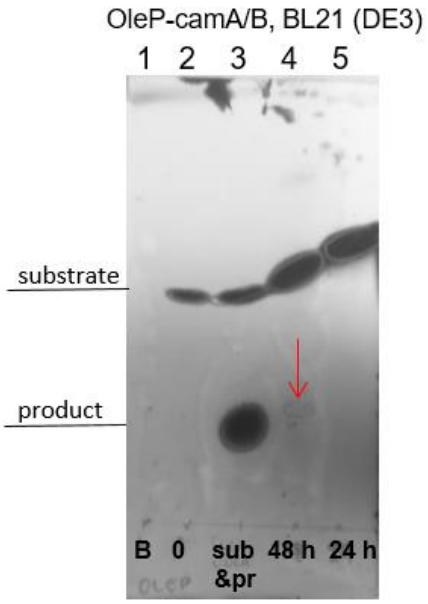


Figure 4.1: TLC plate for the biotransformation of the steroid of interest as substrate, using OleP with the PdR/PdX system (OleP-camA/B). The samples were loaded from left to right with the order: background (B), time zero (0), standards of substrate/product, sample after 48 hours of reaction (48 h) and sample after 24 h of reaction (24 h).

To confirm the results, all samples tested by TLC, were analyzed and evaluated by HPLC, as well.

With this method the substrate's retention time was 30.7 min with minor peaks of impurities at 6.1 min, 7.7 min and 9.4 min.

The analysis for the positive control OleP-camA/B in BL21(DE3) is shown in figure 4.2. As seen, the substrate is eluted at 30.7 min as expected, and 6-β hydroxylated product was eluted at 4.59 min. With this, OleP-camA/B expressed in BL21(DE3) is active with the steroid of interest, as indicated from the product formation observed by HPLC analysis. These results support the TLC analysis (figure 4.1). Production of 6-β product was also verified with enzymes expressed in *E. coli* C41(DE3) strain (Chapter 8, figure 8.2) and in *E. coli* C43(DE3) strain (Chapter 8, figure 8.3).

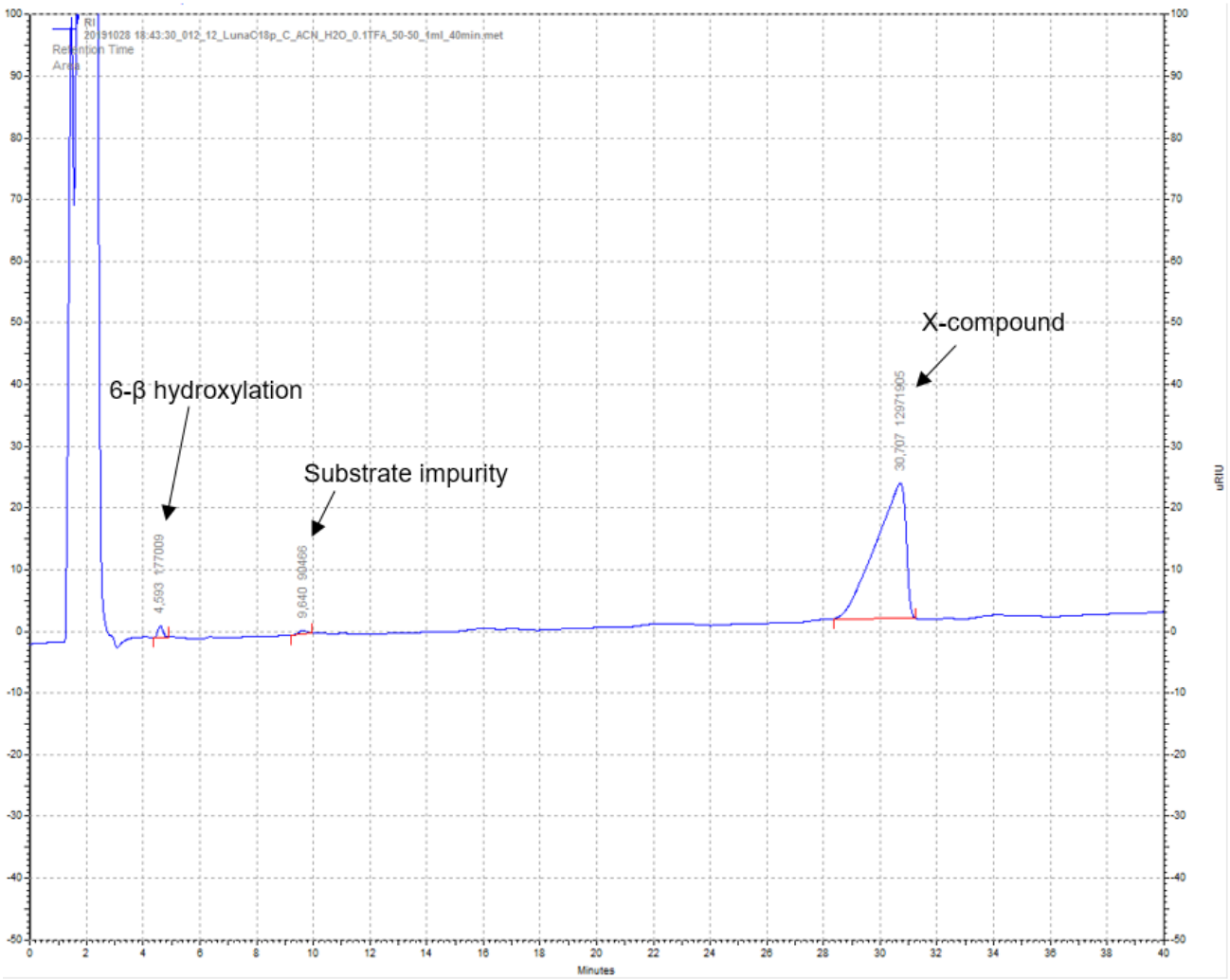


Figure 4.2: Chromatogram of OleP-camA/B in *E. coli* BL21(DE3) after HPLC analysis. On the x-axis the time is displayed from 0 min to 40 min and on the y-axis the absorbance (mAU) is displayed ranging from -50 to 100 mAU.

4.2.2 Investigation of CYPs library in the desired reaction

Using the reaction extracts of OleP-camA/B as a positive control, all CYPs under investigation were interrogated in whole-cells biocatalysis with the steroid of interest, and they were subsequently analyzed via TLC and HPLC. The chromatographic analysis determined that none of the targeted enzymes is capable to convert the substrate, as there was no signal indicating the existence of any product. It is worth to mention that the reactions were not only tested with the CYPs co-transformed with the PdR/PdX redox-partner system but also with the CPR_{His} system, as well, in case that the PdR/PdX system was not compatible with these CYPs.

In table 4.1 the results of the reactions analyzed are displayed. The properly folded enzymes (tables 3.3, 3.4) do not seem to have activity towards this specific substrate (Chapter 8, figure 8.4). Even the enzymes which could not be expressed in properly

folded form, were investigated for substrate conversion in whole-cell biotransformation, but none of them exhibited any activity. On the other hand, the positive control, OleP biocatalysis presented activity in all cases studied. The best conversion and best folding were attained with induction in in *E. coli* BL21(DE3) host cells, used with the PdR/PdX system from *P. putida* (table 4.1).

Table 4.1: The CYPs tested with biocatalysis in whole-cell systems for substrate conversion.

CYP	Organism	Best folding conditions	Product
OleP	*S. antibioticus*	IPTG, BL21(DE3), PdR/PdX	+
CYP3A4	*Homo sapiens*	-	-
cc33	*C. canadensis*	-	-
cc35	*C. canadensis*	-	-
SGG4	*S. griseus*	-	-
NE05	*N. elegans*	IPTG, C43(DE3), CPR_{His}	-
SCE9	*S. cyaneofuscatus*	IPTG, BL21(DE3), PdR/PdX	-
SGDX	*S. griseus*	-	-

4.2.3 Biocatalysis using oxidizing agents with crude lysates

Sodium periodate, diacetoxyiodo benzene and 3-chlorobenzoic acid were used to replace the redox-partner systems and act as electron donors instead of the NADPH. Reactions with the properly folded CYPs were performed and are presented in table 4.2. As seen, the properly folded CYPs after cell lysis were interrogated in the desired reaction with oxidizing agents as electron donors. The reactions were analyzed with TLC and showed that the enzymes are not able to convert the substrate under these conditions. The same results also were obtained by the positive control, OleP. This is not strange, as no NADPH was provided and it could not use the typical mechanism for the observed activity, and, as seen from these experiments, it cannot use the suggested electron donors. With this, it is hypothesized that the chemicals sodium periodate, diacetoxyiodo benzene and 3-chlorobenzoic acid are not suitable electron donors for the hydroxylation of the substrate of interest with any of the investigated CYPs.

Table 4.2: Results of reactions of CYPs with the steroid of interest, by replacing the redox systems with the oxidizing agents sodium periodate, diacetoxyiodo benzene and 3-chlorobenzoic acid.

CYPs	Organism	Best folding conditions	Product
OleP	*S. antibioticus*	IPTG, BL21(DE3)	-
NE05	*N. elegans*	IPTG, C43(DE3)	-
SCE9	*S. cyaneofuscatus*	IPTG, BL21(DE3)	-
WP_091596541	*A. lurida*	IPTG, BL21(DE3)	-
WP_084494404	*N. elegans*	IPTG, BL21(DE3)	-
WP_030573948	*S. cyaneofuscatus*	IPTG, BL21(DE3)	-
FG110	*F. graminearum*	IPTG, BL21(DE3)	-

4.3 Conclusions

The hydroxylation of the steroid of interest was investigated with the CYPs found from the bioinformatics analysis, by using whole-cell biocatalysis. The reactions were analyzed and evaluated by TLC and HPLC, using OleP-camA/B as positive control.

All reactions with OleP and PdR/PdX were successful, as there was a band at the corresponding height in the TLC experiments and a peak at the determined retention time in HPLC. On the other hand, all reactions with this specific substrate with CYPs chosen (CYP3A4, cc33, cc35, SGDX, SCE9, NE05, SGG4) did not lead to any product, although they were tested in different expression conditions, *E. coli* strains and with different redox partner systems (PdR/PdX from *P. putida* and CPR_{His} from *C. apicola*).

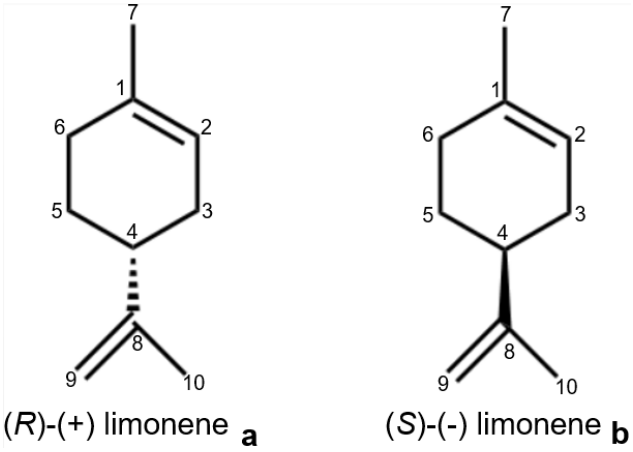
Consequently, the properly folded CYPs were further tested for conversion in crude lysate by using chemicals-oxidizing agents as electron donors. These agents were sodium periodate, (diacetoxyiodo) benzene and 3-chlorobenzoic acid. The TLC analysis proved that there was no product formation after the reaction and these chemical electron donors cannot be accepted from the selected CYPs, not even from the positive control (OleP).

An outlook of this experimental process is to test the proper expressed and folded CYPs which were not co-expressed, together with different redox partners (ex. BMR, CPR_{His}) for product formation with the targeted substrate. However, there are many combinations and for this reason this work was not possible to be finalized in the framework of this thesis.

CHAPTER 5. Establishment of reaction-protocols with limonene

5.1 Introduction

The fused self-sufficient CYPs BM3 (CYP102A1) from the organism *B. megaterium*, and BaLi (CYP102A7), from the organism *B. licheniformis*, have been reported in literature to be active with the terpenes (*R*)-(+)-limonene and (*S*)-(-)-limonene (Scheme 5.1) [29, 36, 70]. After the successful expression and folding tests performed (figure 3.1), the activity of these enzymes against these substrates was investigated monitoring the NADPH consumption, in order to establish the activity protocol as control reaction in future.



Scheme 5.1: The structures of the terpenes (*R*)-(+) limonene (a) and (*S*)-(-) limonene (b).

In redox reactions, the NAD(P)H consumption does not necessarily lead to substrate conversion to a product; the NAD(P)H oxidation can happen also due to uncoupling phenomenon, and Reactive Oxygen Species (ROS) such as hydrogen peroxide can be produced [109, 110]. ROS derived from the uncoupling of the P450 cycle can lead to protein degradation and modify cell macromolecules, provoking toxic effects for the cell and leading to its death (figure 5.1). However, in some cases, the high level of ROS is normal, as in the case of mitochondria [110].

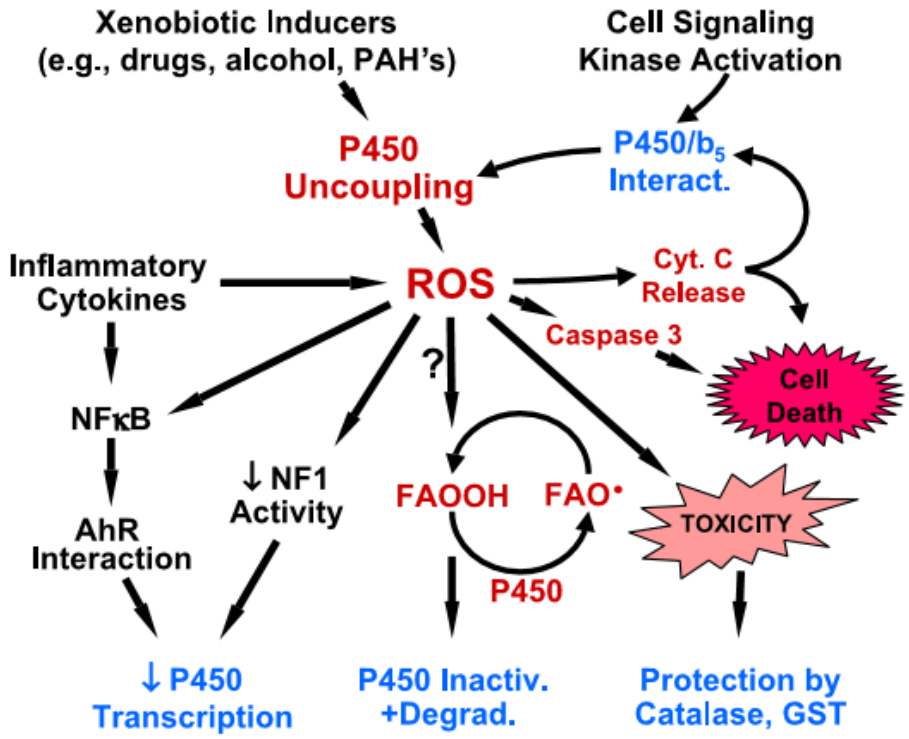


Figure 5.1: Representation of putative ROS regulation by microsomal CYPs. In red are toxicity-related steps and in blue the antioxidant processes. NF1, NFκB and AhR represent regulatory factors. Reprinted by permission from Elsevier. [Journal Publicer: Elsevier, Journal name: Toxicology and applied Pharmacology], reference [110], Mechanisms that regulate production of reactive oxygen species by cytochrome P450, R. C. Zangar *et al*, [COPYRIGHT] 2004

5.2 Results and Discussion

5.2.1 NADPH consumption enzymatic assay

The NADPH consumption was measured at 340 nm. First, the standard curve of NADPH in K_{Pi} buffer (50 mM, pH=7.5) was plotted, concentration of NADPH ranging from 0.50 mM to 0.45 mM, corresponding up to 10% conversion under the experimental conditions (figure 5.2):

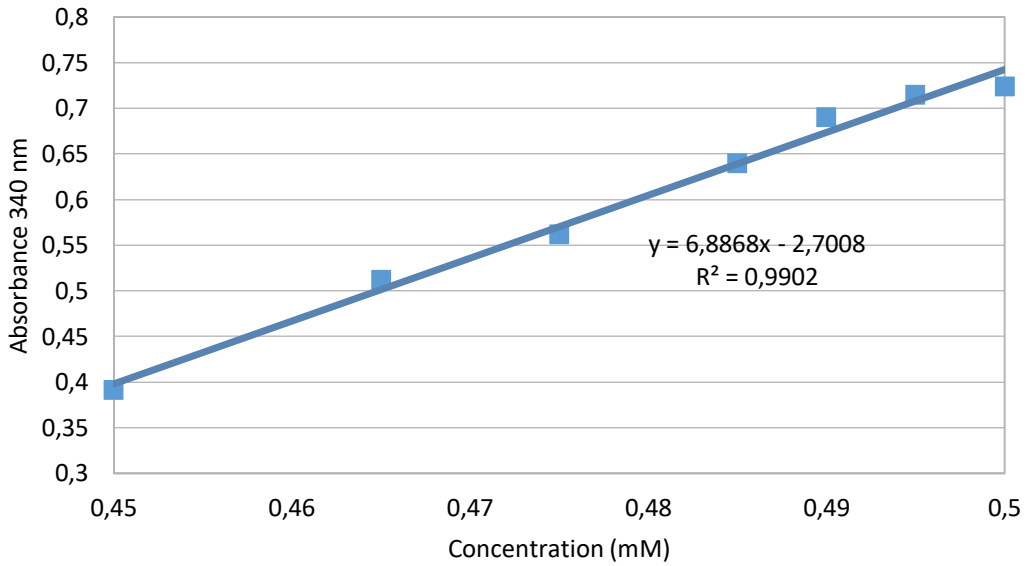


Figure 5.2: Standard curve of the NADPH. The measurements were performed in a 96-well plate, 200 µL volume, and route's length 0.627 nm.

After photometric analysis of BaLi and BM3 CYPs, the specific activity was calculated as the conversion of 1 µmol substrate in 1 min per mg of enzyme, where the enzyme concentration was determined after CO titration (U/mg, equation 2.1). As seen in table 5.1, both self-sufficient CYPs are active against (*R*)-limonene, however, no NADPH consumption was observed when (*S*)-limonene was used as substrate. This is not in accordance to the literature, where activity against both enantiomers has been reported for BaLi [29], however, this can be due to the low activity that may be under the detection limit.

Table 5.1: The enzyme activities of BaLi with (*R*)-limonene.

CYP	U/mL	U/mg	NADPH consumed (µM/min)
BaLi	$2.5 \pm 0.8 \cdot 10^{-5}$	$7.4 \pm 0.8 \cdot 10^{-4}$	0.0007
BM3	$1.6 \pm 0.1 \cdot 10^{-4}$	$3.7 \pm 0.1 \cdot 10^{-4}$	0.00325

5.3 Conclusions

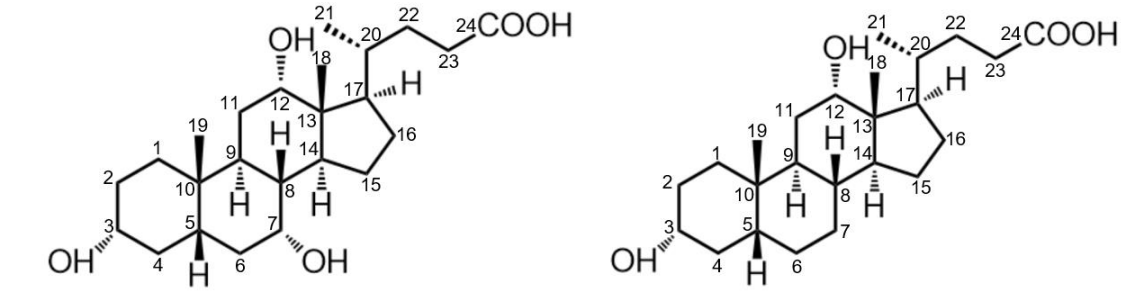
The enzymatic assays performed for BM3 and BaLi with (*R*)-limonene as substrate confirmed the reaction-profile from literature [29, 68, 70]. However, no activity against (*S*)-limonene was observed [29]. Thus, the activity test with (*R*)-limonene could act as a positive control for these self-sufficient CYPs. As it is reported that CYPs convert limonene, an outlook of this work could be to characterize the product and engineer the CYPs to produce a different one of biotechnological interest.

CHAPTER 6. Understanding OleP substrate scope

6.1 Introduction

Whole-cell biotransformations were performed with OleP and the PdR/PdX redox partner systems using the steroids cholic acid and deoxycholic acid. The enzyme has been determined to be able to hydroxylate deoxycholic acid, but not cholic acid [15]. As these compounds have only one hydroxyl group difference (Scheme 6.1), it is striking that the enzyme is so specific and for this reason, we decided to further investigate its substrate scope.

In the Enzyme Technology Laboratory, Department of Chemistry, University of Crete, reactions of OleP and the redox partners BMR and PdR/PdX with the substrates cholic acid and deoxycholic acid were performed, aiming to test the NADPH consumption in the photometric assay using purified proteins (OleP and BMR include the 6xHis-tag (see table 2.1)), but also to check the activity by HPLC (OleP and PdR/PdX in whole-cells). In parallel, bioinformatic analysis was performed using YASARA, to determine the amino acids that are responsible for the observed selectivity.



Scheme 6.1: The steroids used as substrates in this chapter, cholic acid (a) and deoxycholic acid (b).

6.2 Results and Discussion

6.2.1 Activity of OleP against steroid substrates

The results obtained by the photometric analysis of the NADPH consumption of OleP-BMR with cholic acid and deoxycholic acid are displayed in table 6.1. As seen, no NADPH consumption was observed with OleP-BMR system when using cholic acid as substrate, confirming the reported results in literature [15]. On the other hand, in the reactions with deoxycholic acid, the NADPH consumption confirmed that the enzyme is active towards the steroid substrate. The activity of OleP in U/mg was calculated after the saturation with CO.

Table 6.1: Enzyme activities of OleP-BMR with cholic acid and deoxycholic acid

Substrate	U/mL	U/mg	NADPH consumed (μM/min)
Cholic acid	-	-	-
Deoxycholic acid	$1.52 \pm 0.89 \cdot 10^{-4}$	$1.52 \pm 0.84 \cdot 10^{-3}$	0.0011

These reactions were investigated with whole-cell systems, where OleP and PdR/PdX were co-expressed in one BL21(DE3) host and the samples were analyzed via HPLC. As the HPLC in the Department of Chemistry, University of Crete, is equipped with diode array detector, the first step was to identify the optimal wavelength to monitor the reactants. As seen in figure 6.1, the maximum absorbance of cholic acid is at 212 nm and deoxycholic acid at 208 nm. As the concentrations of CA and DCA used for the spectra are equal (4 mg/mL), the absorbance at 208 nm is equal for the two compounds.

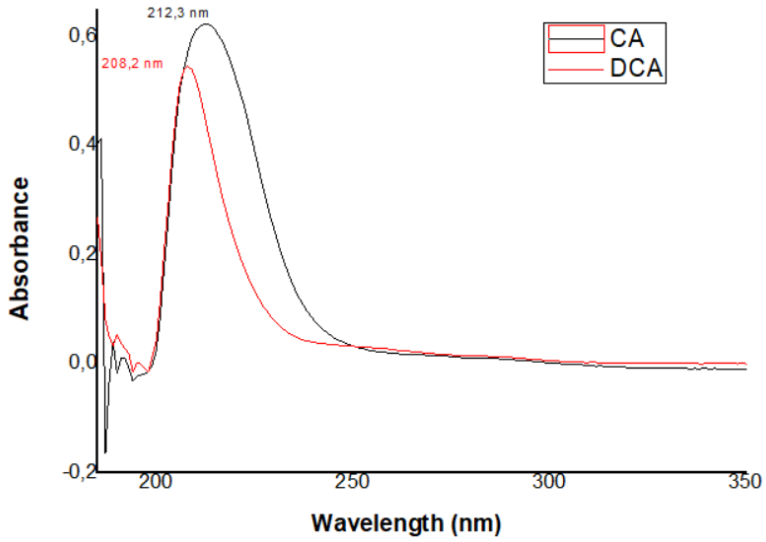


Figure 6.1.: Absorbance spectra of cholic acid (CA, black) and deoxycholic acid (DCA, red),

Analyzing the standards of CA and DCA with the described HPLC method (paragraph 2.2.15), the retention time for DCA was 18.5 min and for CA was 7.2 min. After 48 h of whole cell biocatalysis of the OleP-PdR/PdX system with DCA as substrate, a more hydrophilic product peak appears at 14.9 min (figure 6.2). As this is the only peak observed, we conclude that this is the 6-β hydroxylated DCA. As we do not have standards to calculate the extinction coefficient, or at least compare the signal with that of DCA, we cannot quantify the amount of product produced. If we base the analysis according to the reduction of the substrate peak, a conversion of about 20% is calculated, however without internal standard, this is just an estimation and further experiments should be performed to be able to quantify the conversion.

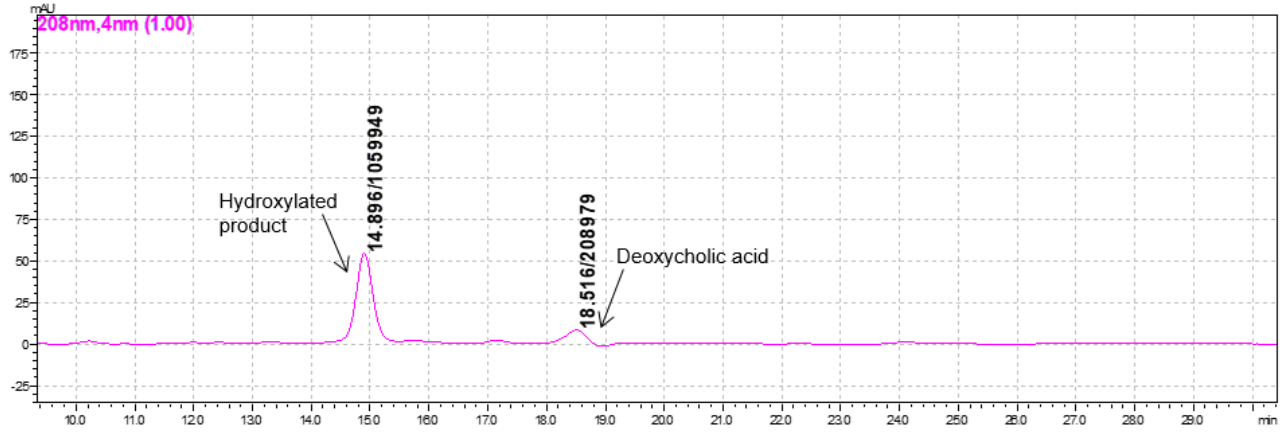


Figure 6.2: Chromatogram of OleP-camA/B in *E. coli* BL21(DE3) after HPLC analysis.

6.2.2 Bioinformatic analysis

The 3D structure of OleP was retrieved from PDB (PDB ID: 5MNS) [30]. The structure was evaluated using the web service MolProbity and analyzing sterics and geometry. As 5MNS consists of six identical monomeric chains, the MolProbity data were used to identify the chain with the better quality (chain A), for the following experiments. The chain A was isolated in a single PDB file and this was used as the model structure of OleP. After the refinement of the structure of the selected chain, the docking experiments of OleP were performed with cholic acid and deoxycholic acid as substrates. The hydroxylation positions were determined based on the distance between the substrate's nearest carbon atom and the oxygen of the Fe=O bond of the heme. In figure 6.3, deoxycholic acid is docked in OleP active site. The docking revealed the proper orientation of DCA for 6-β-hydroxylation, which is the position certified from wet lab experiments.

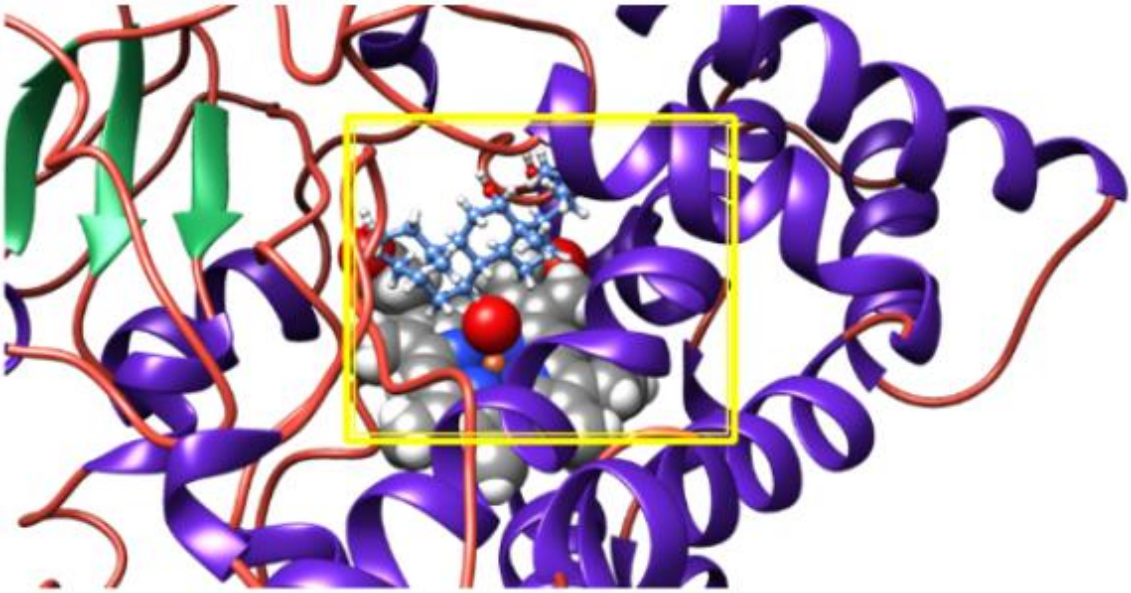


Figure 6.3: The docked complex of the OleP ferryl-oxo intermediate with deoxycholic acid. The heme is depicted as gray spheres and the substrate is shown in ball and stick and colored light blue. Helices are colored purple, strands green and coils coral. The image was created using Chimera.

After reviewing the docked complexes of OleP with cholic acid and deoxycholic acid, some positions of interest have been identified, for the selectivity towards the substrate, namely Arg82, Pro88, Val 291 and Phe296. These positions seem to be able to turn the substrate in a slightly different orientation and thus alter the selectivity.

The point mutation Phe296Arg was chosen after bioinformatics analysis, aiming to change the hydroxylation position to 4-β. Phe296 is a hydrophobic amino acid in the moiety of the carboxylic function of the substrate and by adding a positive charged amino acid, hydrogen bond can be formed between the Arg296 and the substrate and thus differentiate the docking of the substrate in the active site. In figure 6.4 the docking was performed with the mutant Phe296Arg and it seems that the variant would be able to hydroxylate position 4-β.

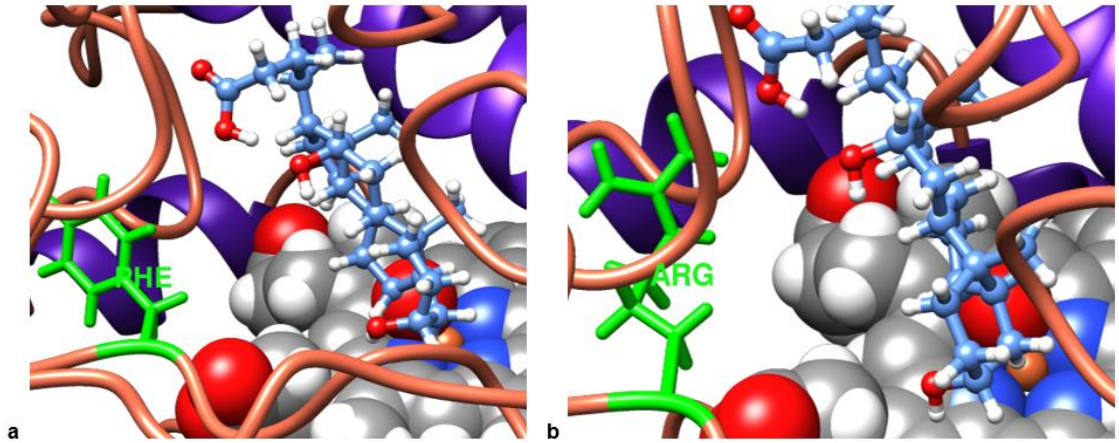


Figure 6.4: The docked complex of the wild type OleP ferryl-oxo intermediate (a) and the mutant F286R (b) with deoxycholic acid. The heme is depicted as gray spheres and the substrate is shown in ball and stick and colored light blue. Helices are colored purple, strands green and coils coral. The image was created using Chimera.

Proline is a nonpolar aliphatic imino acid that has significant restrains on the conformations it can take. Pro88 is also close to the carboxylic acid of the substrate and it was hypothesized, that Pro88His, would provide more mobility to the turn that the proline is located, while the positive charge of histidine can attract the substrate by interacting with the acidic group. The docking revealed an alternative binding and also suggests that this mutant would provide the 4-β-hydroxylated product (figure 6.5).

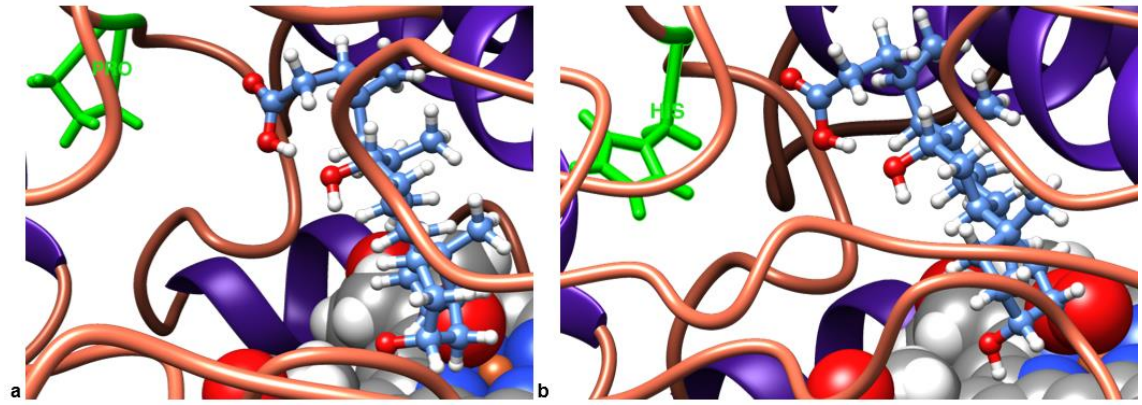


Figure 6.5: The docked complex of the wild type OleP ferryl-oxo intermediate (a) and the mutant P88H (b) with deoxycholic acid. The heme is depicted as gray spheres and the substrate is shown in ball and stick and colored light blue. Helices are colored purple, strands green and coils coral. The image was created using Chimera.

In parallel, the mutant Val291Lys was considered to be promising as well for the formation of the hydroxylated 4-β product, by keeping the same motif of analysis. The positively charged Lys is interesting to be tested for attracking the hydroxyl group in

position 3 of the steroidal skeleton of DCA, when the wild type protein with the hydrophobic valine hydroxylates the steroid substrate in position 6-β (Figure 6.6).

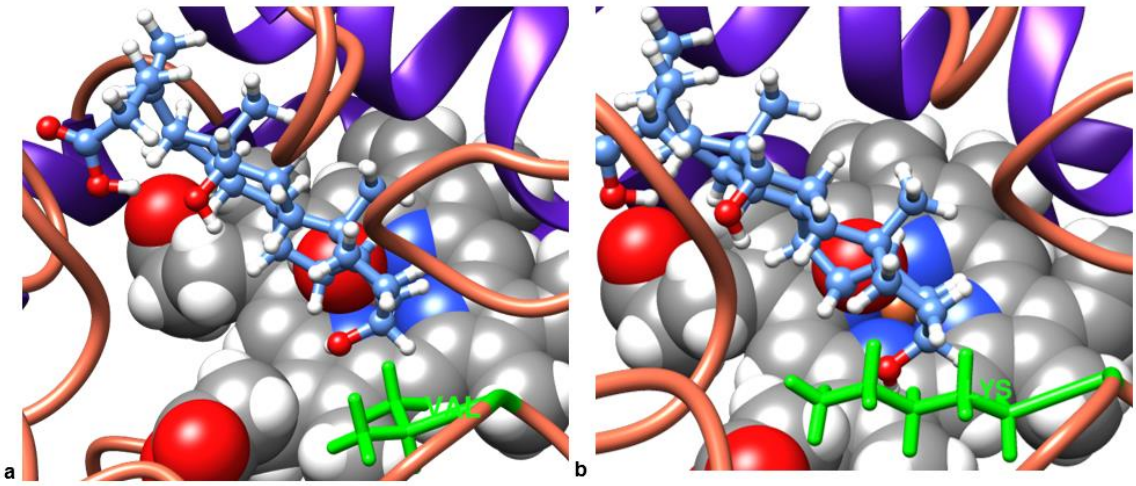


Figure 6.6: The docked complex of the wild type OleP ferryl-oxo intermediate (a) and the mutant V291K (b) with deoxycholic acid. The heme is depicted as gray spheres and the substrate is shown in ball and stick and colored light blue. Helices are colored purple, strands green and coils coral. The image was created using Chimera.

Lastly, the position 82 is located further of the binding site and it interacts with the carboxyl function. The idea of mutating this position is to destabilize this position by providing a more hydrophobic character in this area. Thus, the mutants Arg82Phe and Arg82Val were selected. Both mutants' analysis shows that the hypothetical product would be 4-β hydroxylated (figure 6.7).

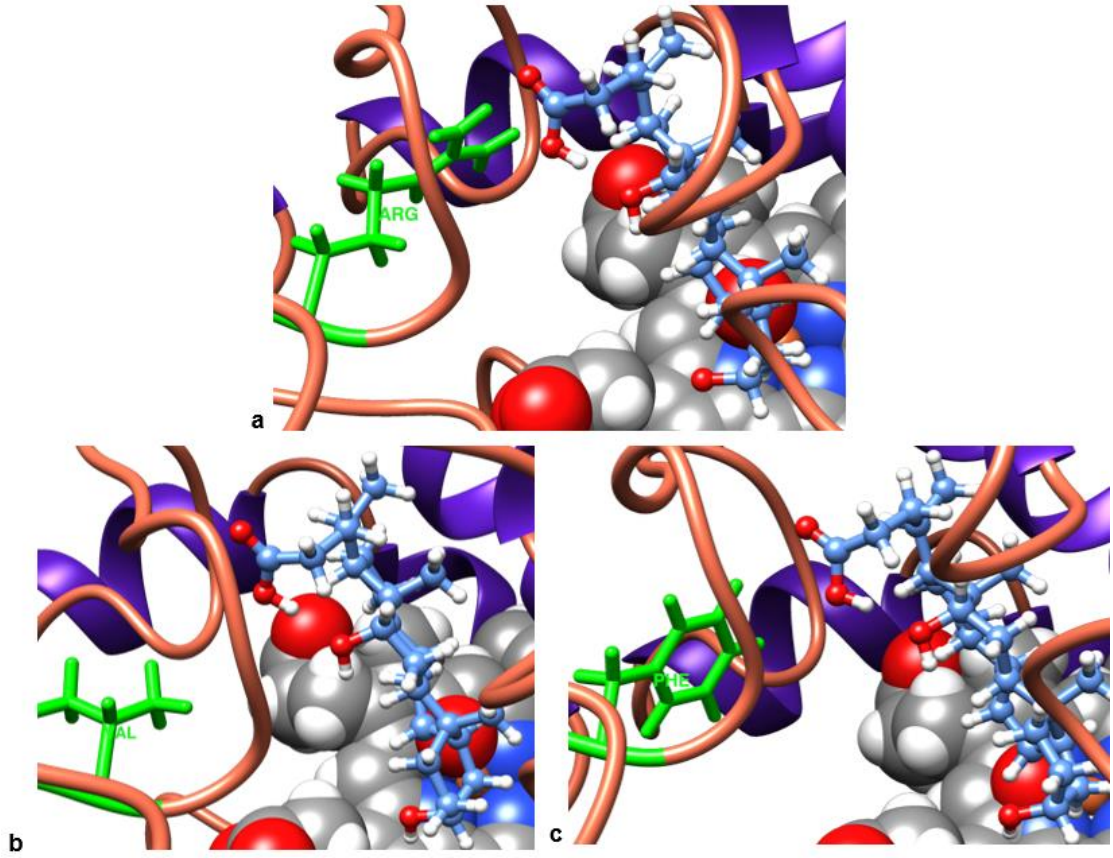


Figure 6.7: The docked complex of the wild type OleP ferryl-oxo intermediate (a) and the mutants R82F (b) and R82V (c), with deoxycholic acid. The heme is depicted as gray spheres and the substrate is shown in ball and stick and colored light blue. Helices are colored purple, strands green and coils coral. The image was created using Chimera.convert

6.3 Conclusions

The enzymatic assays performed for OleP with the reductase BMR and the steroid substrates proved that the enzyme in pure form can convert DCA but not CA, based on NADPH consumption. Similarly, OleP with the PdR/PdX redox partner system, after biocatalysis in whole-cells, can hydroxylate DCA but not CA, based on HPLC analysis. As the 6-β hydroxylated product was not available for preparing a standard curve, we cannot quantify the reactions, however only one peak is observed in HPLC, agreeing with literature that OleP is only hydroxylating in the 6-β position [15].

Bioinformatic analysis was performed for the study of the CYP OleP. Using the YASARA software the aim was to determine point mutants that will understand the reasons that OleP accepts DCA but not CA and change the regioselectivity. Concerning the first part, there was no obvious reason why the enzyme would not accept CA as substrate. However, several position were identified for mutagenesis, to provide alternative hydroxylation profile. The bioinformatics analysis indicated that the

shifting of the hydroxylation position is possible with the mutants and as an outlook, wet-lab experiments will determine if the mutants of OleP can be expressed in the properly folded form and if product conversion with biocatalysis in whole-cell systems can be accomplished. The molecular biology part of this work was performed as the last part of this thesis, but unfortunately we were not able to successfully introduce the mutations in the limited time we had and thus the mutants remain to be prepared and characterized in the future.

CHAPTER 7. Conclusions

This Master thesis aimed at the creation of a collection of redox biocatalysts that will be able to synthesize bioactive compounds. The project consists of two parts. In the first part, twenty different CYP enzymes were selected, after 2nd generation sequencing bioinformatics analysis, and they were tested firstly for expression and folding, single and together with different redox partner systems, using different *E. coli* strains and different expression conditions. Afterwards, the highly selective hydroxylation of a steroid of interest was studied, by investigating the suitable redox partner system catalyzing the reaction of interest, cooperatively with the selected monooxygenases. The experiments of the first part were performed in the Department of Biotechnology and Enzyme Catalysis, Institute of Biochemistry, University of Greifswald and the CYP OleP (*S. antibioticus*) was used as positive control in all these experiments.

The expression experiments determined that OleP can be expressed in *E. coli* BL21(DE3), C41(DE3) and C43(DE3) using the inducer IPTG, and that this CYP can be also successfully co-expressed with the PdR/PdX redox partner. According to the TLC and HPLC analysis, after biotransformations in whole-cells, it was determined that a product is formed after the reaction of the steroid of interest with OleP and PdR/PdX in *E. coli* BL21(DE3), C41(DE3) and C43(DE3). Additionally, to enable the industrial applicability of the process, the protein expression in auto-induction media was feasible in the *E. coli* strain C41(DE3).

Based on the results obtained from OleP, the selected CYPs were evaluated for expression, proper folding and for the conversion of the steroid of interest. The cytochrome CYP3A4 from *Homo sapiens*, could be expressed alone and together with PdR/PdX, but in a denatured form, using IPTG. The tests were performed in BL21(DE3), C41(DE3) or C43(DE3), with IPTG as inducer and with auto-induction media, and did not form a product in whole-cell biotransformation. Alike to CYP3A4, the candidates cc33 and cc35 from *C. canadensis* were studied and they were expressed only in C41(DE3) and C43(DE3) with IPTG boosting the process, but misfolded and they did not convert the steroid of interest together with PdR/PdX. In parallel, NE05 from *N. elegans* was investigated for expression with IPTG, in different *E. coli* strains and with the redox system PdR/PdX and the reductase CPR_{His} from *C. apicola*. In strain C43(DE3), NE05 was obtained properly folded when expressed stand-alone, partially properly folded when co-expressed with the CPR_{His} and misfolded with PdR/PdX. SGG4 from *S. griseus* was indicated as inappropriate CYP to work, as it was partially properly folded when single-expressed, but denatured together with the redox partners, something that does not allow the application in redox

reactions in whole cell biocatalysis. Likely to SGG4, the CYP SGDX from *S. griseus* was expressed in BL21(DE3), but the protein did not have the proper folded form at the active site. On the other hand, SCE9 from *S. cyaneofuscatus* obtained the partially folded, as single and co-expressed in BL21(DE3) with the IPTG inducer. NE05, SCE9, SGG4 and SGDX, after biotransformations in whole-cells did not convert the steroid of interest, although the CO-difference spectra for NE05 and SCE9 proved that they could be promising candidates.

Additionally, twelve more genes were studied in *E. coli* BL21(DE3) and four of them (WP_091596541 (*A. lurida*), WP_084494404 (*N. elegans*), WP_030573948 (*S. cyaneofuscatus*) and FG110 (*F.graminearum*)) were stable. Further tests for product formation were performed by using the steroid of interest as a substrate and chemical oxidizing agents. Sodium periodate, (diacetoxyiodo)benzene and 3-chlorobenzoic acid were studied as electron donors, however TLC analysis proved that there was no product formation after the reaction and thus the enzymes, even OleP, require the redox partner proteins in order to be active.

Thus, the next step on this project would be to test the proper expressed and folded CYPs together with different redox partner systems (ex. BMR, CPR_{His}) for the conversion of the substrate of interest and further test them with steroid substrates.

In the second part of the thesis, in Department of Chemistry, University of Crete, OleP (*S. antibioticus*) was selected as the most suitable monooxygenase for studying the conversion of the steroid substrates CA and DCA, cooperatively with the redox partner system PdR/PdX. After repeating the results from the literature, for OleP catalyzing the 6-β hydroxylation of DCA and inability to convert CA, the enzyme was further tested for the reactions photometrically with the reductase BMR (*B. megaterium*). The consumption of NADPH was observed at 340 nm, in the reaction with DCA, but there was no consumption when studied with the CA. With this, the reactions were confirmed with a different electron transfer system.

OleP was analyzed with bioinformatics, aiming at the protein engineering of the enzyme to change the hydroxylation position when reacting with DCA. Additionally, the study also aimed at the understanding of why OleP does not accept CA as substrate. There is no clear reason why this is happening, as the structural differences between CA and DCA are minimal and there are no steric hintrances or disfavored interactions with the hydroxyl group of CA. Concerning the regioselectivity alteration, the analysis mainly focused on introducing or removing charges to residues close to the active site and the substrate, in order to change the docking of the substrate. The mutants

selected to be most suitable after the bioinformatic analysis were Phe296Arg, Arg82Phe, Arg82Val, Pro88His and Val291Lys. As an outlook, the protein engineering of OleP can be performed and the mutants should be tested for NADPH consumption with the steroid substrates in whole-cell biotransformations.

Lastly, two fused CYPs were additionally studied for expression and folding and reactions analyzed photometrically with the terpenes (*R*)- and (*S*)-limonene as substrates. The CYPs BM3 (*B. megaterium*) and BaLi (*B. licheniformis*) were expressed in properly folded form and NADPH consumption was observed only with the substrate (*R*)-limonene. This study was performed to establish the protocols with the fused proteins and in future experiments, more terpenes as substrates for several reactions. In addition, bioinformatics analysis can suggest mutants for the CYPs to be studied with substrates of interest.

CHAPTER 8. Appendix

8.1 DNA and amino-acid sequences of the cytochromes P450

Below are the sequences of each tested gene, and the corresponding amino acid sequence.

BMR Wild type 1728 bp DNA with the C-terminal 6xHis-tag

5′–ATGCCGCTGCTTGTGCTATACGGTTCAAATATGGGAACAGCTGAAGGAACGGCG
CGTGATTTAGCAGATATTGCAATGAGCAAAGGATTTGCACCGCAGGTCGCAACGCTT
GATTCACACGCCGGAAATCTTCCGCGCGAAGGAGCTGTATTAATTGTAACGGCGTCT
TATAACGGTCATCCGCCTGATAACGCAAAGCAATTTGTCGACTGGTTAGACCAAGCG
TCTGCTGATGAAGTAAAAGGCGTTCGCTACTCCGTATTTGGATGCGGCGATAAAAAC
TGGGCTACTACGTATCAAAAAGTGCCTGCTTTTATCGATGAAACGCTTGCCGCTAAA
GGGGCAGAAAACATCGCTGACCGCGGTGAAGCAGATGCAAGCGACGACTTTGAAGGC
ACATATGAAGAATGGCGTGAACATATGTGGAGTGACGTAGCAGCCTACTTTAACCTC
GACATTGAAAACAGTGAAGATAATAAATCTACTCTTTCACTTCAATTTGTCGACAGC
GCCGCGGATATGCCGCTTGCGAAAATGCACGGTGCGTTTTCAACGAACGTCGTAGCA
AGCAAAGAACTTCAACAGCCAGGCAGTGCACGAAGCACGCGACATCTTGAAATTGAA
CTTCCAAAAGAAGCTTCTTATCAAGAAGGAGATCATTTAGGTGTTATTCCTCGCAAC
TATGAAGGAATAGTAAACCGTGTAACAGCAAGGTTCGGCCTAGATGCATCACAGCAA
ATCCGTCTGGAAGCAGAAGAAGAAAAATTAGCTCATTTGCCACTCGCTAAAACAGTA
TCCGTAGAAGAGCTTCTGCAATACGTGGAGCTTCAAGATCCTGTTACGCGCACGCAG
CTTCGCGCAATGGCTGCTAAAACGGTCTGCCCGCCGCATAAAGTAGAGCTTGAAGCC
TTGCTTGAAAAGCAAGCCTACAAAGAACAAGTGCTGGCAAAACGTTTAACAATGCTT
GAACTGCTTGAAAAATACCCGGCGTGTGAAATGAAATTCAGCGAATTTATCGCCCTT
CTGCCAAGCATACGCCCGCGCTATTACTCGATTTCTTCATCACCTCGTGTCGATGAA
AAACAAGCAAGCATCACGGTCAGCGTTGTCTCAGGAGAAGCGTGGAGCGGATATGGA
GAATATAAAGGAATTGCGTCGAACTATCTTGCCGAGCTGCAAGAAGGAGATACGATT
ACGTGCTTTATTTCCACACCGCAGTCAGAATTTACGCTGCCAAAAGACCCTGAAACG
CCGCTTATCATGGTCGGACCGGGAACAGGCGTCGCGCCGTTTAGAGGCTTTGTGCAG
GCGCGCAAACAGCTAAAAGAACAAGGACAGTCACTTGGAGAAGCACATTTATACTTC
GGCTGCCGTTCACCTCATGAAGACTATCTGTATCAAGAAGAGCTTGAAAACGCCCAA
AGCGAAGGCATCATTACGCTTCATACCGCTTTTTCTCGCATGCCAAATCAGCCGAAA
ACATACGTTCAGCACGTAATGGAACAAGACGGCAAGAAATTGATTGAACTTCTTGAT
CAAGGAGCGCACTTCTATATTTGCGGAGACGGAAGCCAAATGGCACCTGCCGTTGAA
GCAACGCTTATGAAAAGCTATGCTGACGTTCACCAAGTGAGTGAAGCAGACGCTCGC
TTATGGCTGCAGCAGCTAGAAGAAAAAGGCCGATACGCAAAAGACGTGTGGGCTGGG
CATCATCATCATCATCATTAA–3′

BMR Wild type translation 576 aa with the C-terminal 6xHis-tag

MPLLVLYGSNMGTAEGTARDLADIAMSKGFAPQVATLDSHAGNLPREGAVLIVTASY
NGHPPDNAKQFVDWLDQASADEVKGVRYSVFGCGDKNWATTYQKVPAFIDETLAAKG
AENIADRGEADASDDFEGTYEEWREHMWSDVAAYFNLDIENSEDNKSTLSLQFVDSA
ADMPLAKMHGAFSTNVVASKELQQPGSARSTRHLEIELPKEASYQEGDHLGVIPRNY
EGIVNRVTARFGLDASQQIRLEAEEEKLAHLPLAKTVSVEELLQYVELQDPVTRTQL
RAMAAKTVCPPHKVELEALLEKQAYKEQVLAKRLTMLELLEKYPACEMKFSEFIALL
PSIRPRYYSISSSPRVDEKQASITVSVVSGEAWSGYGEYKGIASNYLAELQEGDTIT
CFISTPQSEFTLPKDPETPLIMVGPGTGVAPFRGFVQARKQLKEQGQSLGEAHLYFG

CRSPHEDYLYQEELENAQSEGIITLHTAFSRMPNQPKTYVQHVMEQDGKKLIELLDQ
GAHFYICGDGSQMAPAVEATLMKSYADVHQVSEADARLWLQQLEEKGRYAKDVWAGH
HHHHH

PdR/PdX Wild type DNA

camA 1269 bp DNA

5′—ATGAACGCAAACGACAACGTGGTCATCGTCGGTACCGGACTGGCTGGCGTTGAG
GTCGCCTTCGGCCTGCGCGCCAGCGGCTGGGAAGGCAATATCCGGTTGGTGGGGGAT
GCGACGGTAATTCCCCATCACCTACCACCGCTATCCAAAGCTTACTTGGCCGGCAAA
GCCACAGCGGAAAGCCTGTACCTGAGAACCCCAGATGCCTATGCAGCGCAGAACATC
CAACTACTCGGAGGCACACAGGTAACGGCTATCAACCGCGACCGACAGCAAGTAATC
CTATCGGATGGCCGGGCACTGGATTACGACCGGCTGGTATTGGCTACCGGAGGGCGT
CCAAGACCCCTACCGGTGGCCAGTGGCGCAGTTGGAAAGGCGAACAACTTTCGATAC
CTGCGCACACTCGAGGACGCCGAGTGCATTCGCCGGCAGCTGATTGCGGATAACCGT
CTGGTGGTGATTGGTGGCGGCTACATTGGCCTTGAAGTGGCTGCCACCGCCATCAAG
GCGAACATGCACGTCACCCTGCTTGATACGGCAGCCCGGGTTCTGGAGCGGGTTACC
GCCCCGCCGGTATCGGCCTTTTACGAGCACCTACACCGCGAAGCCGGCGTTGACATA
CGAACCGGCACGCAGGTGTGCGGGTTCGAGATGTCGACCGACCAACAGAAGGTTACT
GCCGTCCTCTGCGAGGACGGCACAAGGCTGCCAGCGGATCTGGTAATCGCCGGGATT
GGCCTGATACCAAACTGCGAGTTGGCCAGTGCGGCCGGCCTGCAGGTTGATAACGGC
ATCGTGATCAACGAACACATGCAGACCTCTGATCCCTTGATCATGGCCGTCGGCGAC
TGTGCCCGATTTCACAGTCAGCTCTATGACCGCTGGGTGCGTATCGAATCGGTGCCC
AATGCCTTGGAGCAGGCACGAAAGATCGCCGCCATCCTCTGTGGCAAGGTGCCACGC
GATGAGGCGGCGCCCTGGTTCTGGTCCGATCAGTATGAGATCGGATTGAAGATGGTC
GGACTGTCCGAAGGGTACGACCGGATCATTGTCCGCGGCTCTTTGGCGCAACCCGAC
TTCAGCGTTTTCTACCTGCAGGGAGACCGGGTATTGGCGGTCGATACAGTGAACCGT
CCAGTGGAGTTCAACCAGTCAAAACAAATAATCACGGATCGTTTGCCGGTTGAACCA
AACCTACTCGGTGACGAAAGCGTGCCGTTAAAGGAAATCATCGCCGCCGCCAAAGCT
GAACTGAGTAGTGCCTGA—3′

camB 324 bp DNA

5′—ATGTCTAAAGTAGTGTATGTGTCACATGATGGAACGCGTCGCGAACTGGATGTG
GCGGATGGCGTCAGCCTGATGCAGGCTGCAGTCTCCAATGGTATCTACGATATTGTC
GGTGATTGTGGCGGCAGCGCCAGCTGTGCCACCTGCCATGTCTATGTGAACGAAGCG
TTCACGGACAAGGTGCCCGCCGCCAACGAGCGGGAAATCGGCATGCTGGAGTGCGTC
ACGGCCGAACTGAAGCCGAACAGCAGGCTCTGCTGCCAGATCATCATGACGCCCGAG
CTGGATGGCATCGTGGTCGATGTTCCCGATAGGCAATGGTAA—3′

PdR/PdX Wild type translation

camA 422 aa

MNANDNVVIVGTGLAGVEVAFGLRASGWEGNIRLVGDATVIPHHLPPLSKAYLAGKA
TAESLYLRTPDAYAAQNIQLLGGTQVTAINRDRQQVILSDGRALDYDRLVLATGGRP
RPLPVASGAVGKANNFRYLRTLEDAECIRRQLIADNRLVVIGGGYIGLEVAATAIKA
NMHVTLLDTAARVLERVTAPPVSAFYEHLHREAGVDIRTGTQVCGFEMSTDQQKVTA
VLCEDGTRLPADLVIAGIGLIPNCELASAAGLQVDNGIVINEHMQTSDPLIMAVGDC
ARFHSQLYDRWVRIESVPNALEQARKIAAILCGKVPRDEAAPWFWSDQYEIGLKMVG

LSEGYDRIIVRGSLAQPDFSVFYLQGDRVLAVDTVNRPVEFNQSKQIITDRLPVEPN
LLGDESVPLKEIIAAAKAELSSA

camB 107 aa

MSKVVYVSHDGTRRELDVADGVSLMQAAVSNGIYDIVGDCGGSASCATCHVYVNEAF
TDKVPAANEREIGMLECVTAELKPNSRLCCQIIMTPELDGIVVDVPDRQWC

BaLi (CYP102A7) Wild type 3227 bp DNA with the C-terminal 6xHis-tag through a ligand

5′—ATGAACAAACTGGACGGTATCCCGATCCCGAAAACCTACGGTCCGCTGGGTAAC
CTGCCGCTGCTGGACAAAAACCGTGTTTCTCAGTCTCTGTGGAAAATCGCTGACGAA
ATGGGTCCGATCTTCCAGTTCAAATTCGCTGACGCTATCGGTGTTTTCGTTTCTTCT
CACGAACTGGTTAAAGAAGTTTCTGAAGAATCTCGTTTCGACAAAAACATGGGTAAA
GGTCTGCTGAAAGTTCGTGAATTCTCTGGTGACGGTCTGTTCACCTCTTGGACCGAA
GAACCGAACTGGCGTAAAGCTCACAACATCCTGCTGCCGTCTTTCTCTCAGAAAGCT
ATGAAAGGTTACCACCCGATGATGCAGGACATCGCTGTTCAGCTGATCCAGAAATGG
TCTCGTCTGAACCAGGACGAATCTATCGACGTTCCGGACGACATGACCCGTCTGACC
CTGGACACCATCGGTCTGTGCGGTTTCAACTACCGTTTCAACTCTTTCTACCGTGAA
GGTCAGCACCCGTTCATCGAATCTATGGTTCGTGGTCTGTCTGAAGCTATGCGTCAG
ACCAAACGTTTCCCGCTGCAGGACAAACTGATGATCCAGACCAAACGTCGTTTCAAC
TCTGACGTTGAATCTATGTTCTCTCTGGTTGACCGTATCATCGCTGACCGTAAACAG
GCTGAATCTGAATCTGGTAACGACCTGCTGTCTCTGATGCTGCACGCTAAAGACCCG
GAAACCGGTGAAAAACTGGACGACGAAAACATCCGTTACCAGATCATCACCTTCCTG
ATCGCTGGTCACGAAACCACCTCTGGTCTGCTGTCTTTCGCTATCTACCTGCTGCTG
AAACACCCGGACAAACTGAAAAAAGCTTACGAAGAAGCTGACCGTGTTCTGACCGAC
CCGGTTCCGTCTTACAAACAGGTTCAGCAGCTGAAATACATCCGTATGATCCTGAAC
GAATCTATCCGTCTGTGGCCGACCGCTCCGGCTTTCTCTCTGTACGCTAAAGAAGAA
ACCGTTATCGGTGGTAAATACCTGATCCCGAAAGGTCAGTCTGTTACCGTTCTGATC
CCGAAACTGCACCGTGACCAGTCTGTTTGGGGTGAAGACGCTGAAGCTTTCCGTCCG
GAACGTTTCGAACAGATGGACTCTATCCCGGCTCACGCTTACAAACCGTTCGGTAAC
GGTCAGCGTGCTTGCATCGGTATGCAGTTCGCTCTGCACGAAGCTACCCTGGTTCTG
GGTATGATCCTGCAGTACTTCGACCTGGAAGACCACGCTAACTACCAGCTGAAAATC
AAAGAATCTCTGACCCTGAAACCGGACGGTTTCACCATCCGTGTTCGTCCGCGTAAA
AAAGAAGCTATGACCGCTATGCCGGGTGCTCAGCCGGAAGAAAACGGTCGTCAGGAA
GAACGTCCGTCTGCTCCGGCTGCTGAAAACACCCACGGTACCCCGCTGCTGGTTCTG
TACGGTTCTAACCTGGGTACCGCTGAAGAAATCGCTAAAGAACTGGCTGAAGAAGCT
CGTGAACAGGGTTTCCACTCTCGTACCGCTGAACTGGACCAGTACGCTGGTGCTATC
CCGGCTGAAGGTGCTGTTATCATCGTTACCGCTTCTTACAACGGTAACCCGCCGGAC
TGCGCTAAAGAATTCGTTAACTGGCTGGAACACGACCAGACCGACGACCTGCGTGGT
GTTAAATACGCTGTTTTCGGTTGCGGTAACCGTTCTTGGGCTTCTACCTACCAGCGT
ATCCCGCGTCTGATCGACTCTGTTCTGGAAAAAAAAGGTGCTCAGCGTCTGCACAAA
CTGGGTGAAGGTGACGCTGGTGACGACTTCGAAGGTCAGTTCGAATCTTGGAAATAC
GACCTGTGGCCGCTGCTGCGTACCGAATTCTCTCTGGCTGAACCGGAACCGAACCAG
ACCGAAACCGACCGTCAGGCTCTGTCTGTTGAATTCGTTAACGCTCCGGCTGCTTCT
CCGCTGGCTAAAGCTTACCAGGTTTTCACCGCTAAAATCTCTGCTAACCGTGAACTG
CAGTGCGAAAAATCTGGTCGTTCTACCCGTCACATCGAAATCTCTCTGCCGGAAGGT
GCTGCTTACCAGGAAGGTGACCACCTGGGTGTTCTGCCGCAGAACTCTGAAGTTCTG

ATCGGTCGTGTTTTCCAGCGTTTCGGTCTGAACGGTAACGAACAGATCCTGATCTCT
GGTCGTAACCAGGCTTCTCACCTGCCGCTGGAACGTCCGGTTCACGTTAAAGACCTG
TTCCAGCACTGCGTTGAACTGCAGGAACCGGCTACCCGTGCTCAGATCCGTGAACTG
GCTGCTCACACCGTTTGCCCGCCGCACCAGCGTGAACTGGAAGACCTGCTGAAAGAC
GACGTTTACAAAGACCAGGTTCTGAACAAACGTCTGACCATGCTGGACCTGCTGGAA
CAGTACCCGGCTTGCGAACTGCCGTTCGCTCGTTTCCTGGCTCTGCTGCCGCCGCTG
AAACCGCGTTACTACTCTATCTCTTCTTCTCCGCAGCTGAACCCGCGTCAGACCTCT
ATCACCGTTTCTGTTGTTTCTGGTCCGGCTCTGTCTGGTCGTGGTCACTACAAAGGT
GTTGCTTCTAACTACCTGGCTGGTCTGGAACCGGGTGACGCTATCTCTTGCTTCATC
CGTGAACCGCAGTCTGGTTTCCGTCTGCCGGAAGACCCGGAAACCCCGGTTATCATG
GTTGGTCCGGGTACCGGTATCGCTCCGTACCGTGGTTTCCTGCAGGCTCGTCGTATC
CAGCGTGACGCTGGTGTTAAACTGGGTGAAGCTCACCTGTACTTCGGTTGCCGTCGT
CCGAACGAAGACTTCCTGTACCGTGACGAACTGGAACAGGCTGAAAAAGACGGTATC
GTTCACCTGCACACCGCTTTCTCTCGTCTGGAAGGTCGTCCGAAAACCTACGTTCAG
GACCTGCTGCGTGAAGACGCTGCTCTGCTGATCCACCTGCTGAACGAAGGTGGTCGT
CTGTACGTTTGCGGTGACGGTTCTCGTATGGCTCCGGCTGTTGAACAGGCTCTGTGC
GAAGCTTACCGTATCGTTCAGGGTGCTTCTCGTGAAGAATCTCAGTCTTGGCTGTCT
GCTCTGCTGGAAGAAGGTCGTTACGCTAAAGACGTTTGGGACGGTGGTGTTTCTCAG
CACAACGTTAAAGCTGACTGCATCGCTCGTACCTAAGC—3′

BaLi (CYP102A7) Wild type translation 1074 aa with the C-terminal 6xHis-tag through a ligand

MNKLDGIPIPKTYGPLGNLPLLDKNRVSQSLWKIADEMGPIFQFKFADAIGVFVSSH
ELVKEVSEESRFDKNMGKGLLKVREFSGDGLFTSWTEEPNWRKAHNILLPSFSQKAM
KGYHPMMQDIAVQLIQKWSRLNQDESIDVPDDMTRLTLDTIGLCGFNYRFNSFYREG
QHPFIESMVRGLSEAMRQTKRFPLQDKLMIQTKRRFNSDVESMFSLVDRIIADRKQA
ESESGNDLLSLMLHAKDPETGEKLDDENIRYQIITFLIAGHETTSGLLSFAIYLLLK
HPDKLKKAYEEADRVLTDPVPSYKQVQQLKYIRMILNESIRLWPTAPAFSLYAKEET
VIGGKYLIPKGQSVTVLIPKLHRDQSVWGEDAEAFRPERFEQMDSIPAHAYKPFGNG
QRACIGMQFALHEATLVLGMILQYFDLEDHANYQLKIKESLTLKPDGFTIRVRPRKK
EAMTAMPGAQPEENGRQEERPSAPAAENTHGTPLLVLYGSNLGTAEEIAKELAEEAR
EQGFHSRTAELDQYAGAIPAEGAVIIVTASYNGNPPDCAKEFVNWLEHDQTDDLRGV
KYAVFGCGNRSWASTYQRIPRLIDSVLEKKGAQRLHKLGEGDAGDDFEGQFESWKYD
LWPLLRTEFSLAEPEPNQTETDRQALSVEFVNAPAASPLAKAYQVFTAKISANRELQ
CEKSGRSTRHIEISLPEGAAYQEGDHLGVLPQNSEVLIGRVFQRFGLNGNEQILISG
RNQASHLPLERPVHVKDLFQHCVELQEPATRAQIRELAAHTVCPPHQRELEDLLKDD
VYKDQVLNKRLTMLDLLEQYPACELPFARFLALLPPLKPRYYSISSSPQLNPRQTSI
TVSVVSGPALSGRGHYKGVASNYLAGLEPGDAISCFIREPQSGFRLPEDPETPVIMV
GPGTGIAPYRGFLQARRIQRDAGVKLGEAHLYFGCRRPNEDFLYRDELEQAEKDGIV
HLHTAFSRLEGRPKTYVQDLLREDAALLIHLLNEGGRLYVCGDGSRMAPAVEQALCE
AYRIVQGASREESQSWLSALLEEGRYAKDVWDGGVSQHNVKADCIARTAALEHHHHH
H

BM3 (CYP102A1) F87V 3168 bp DNA with a C-terminal 6xHis-tag

5′—ATGACAATTAAAGAAATGCCTCAGCCAAAAACGTTTGGAGAGCTTAAAAATTTA
CCGTTATTAAACACAGATAAACCGGTTCAAGCTTTGATGAAAATTGCGGATGAATTA
GGAGAAATCTTTAAATTCGAGGCGCCTGGTCGTGTAACGCGCTACTTATCAAGTCAG

CGTCTAATTAAAGAAGCATGCGATGAATCACGCTTTGATAAAAACTTAAGTCAAGCG
CTTAAATTTGTACGTGATTTTGCAGGAGACGGGTTATTTACAAGCTGGACGCATGAA
AAAAATTGGAAAAAAGCGCATAATATCTTACTTCCAAGCTTCAGTCAGCAGGCAATG
AAAGGCTATCATGCGATGATGGTCGATATCGCCGTGCAGCTTGTTCAAAAGTGGGAG
CGTCTAAATGCAGATGAGCATATTGAAGTACCGGAAGACATGACACGTTTAACGCTT
GATACAATTGGTCTTTGCGGCTTTAACTATCGCTTTAACAGCTTTTACCGAGATCAG
CCTCATCCATTTATTACAAGTATGGTCCGTGCACTGGATGAAGCAATGAACAAGCTG
CAGCGAGCAAATCCAGACGACCCAGCTTATGATGAAAACAAGCGCCAGTTTCAAGAA
GATATCAAGGTGATGAACGACCTAGTAGATAAAATTATTGCAGATCGCAAAGCAAGC
GGTGAACAAAGCGATGATTTATTAACGCATATGCTAAACGGAAAAGATCCAGAAACG
GGTGAGCCGCTTGATGACGAGAACATTCGCTATCAAATTATTACATTCTTAATTGCG
GGACACGAAACAACAAGTGGTCTTTTATCATTTGCGCTGTATTTCTTAGTGAAAAAT
CCACATGTATTACAAAAAGCAGCAGAAGAAGCAGCACGAGTTCTAGTAGATCCTGTT
CCAAGCTACAAACAAGTCAAACAGCTTAAATATGTCGGCATGGTCTTAAACGAAGCG
CTGCGCTTATGGCCAACTGCTCCTGCGTTTTCCCTATATGCAAAAGAAGATACGGTG
CTTGGAGGAGAATATCCTTTAGAAAAAGGCGACGAACTAATGGTTCTGATTCCTCAG
CTTCACCGTGATAAAACAATTTGGGGAGACGATGTGGAAGAGTTCCGTCCAGAGCGT
TTTGAAAATCCAAGTGCGATTCCGCAGCATGCGTTTAAACCGTTTGGAAACGGTCAG
CGTGCGTGTATCGGTCAGCAGTTCGCTCTTCATGAAGCAACGCTGGTACTTGGTATG
ATGCTAAAACACTTTGACTTTGAAGATCATACAAACTACGAGCTGGATATTAAAGAA
ACTTTAACGTTAAAACCTGAAGGCTTTGTGGTAAAAGCAAAATCGAAAAAAATTCCG
CTTGGCGGTATTCCTTCACCTAGCACTGAACAGTCTGCTAAAAAAGTACGCAAAAAG
GCAGAAAACGCTCATAATACGCCGCTGCTTGTGCTATACGGTTCAAATATGGGAACA
GCTGAAGGAACGGCGCGTGATTTAGCAGATATTGCAATGAGCAAAGGATTTGCACCG
CAGGTCGCAACGCTTGATTCACACGCCGGAAATCTTCCGCGCGAAGGAGCTGTATTA
ATTGTAACGGCGTCTTATAACGGTCATCCGCCTGATAACGCAAAGCAATTTGTCGAC
TGGTTAGACCAAGCGTCTGCTGATGAAGTAAAAGGCGTTCGCTACTCCGTATTTGGA
TGCGGCGATAAAAACTGGGCTACTACGTATCAAAAAGTGCCTGCTTTTATCGATGAA
ACGCTTGCCGCTAAAGGGGCAGAAAACATCGCTGACCGCGGTGAAGCAGATGCAAGC
GACGACTTTGAAGGCACATATGAAGAATGGCGTGAACATATGTGGAGTGACGTAGCA
GCCTACTTTAACCTCGACATTGAAAACAGTGAAGATAATAAATCTACTCTTTCACTT
CAATTTGTCGACAGCGCCGCGGATATGCCGCTTGCGAAAATGCACGGTGCGTTTTCA
ACGAACGTCGTAGCAAGCAAAGAACTTCAACAGCCAGGCAGTGCACGAAGCACGCGA
CATCTTGAAATTGAACTTCCAAAAGAAGCTTCTTATCAAGAAGGAGATCATTTAGGT
GTTATTCCTCGCAACTATGAAGGAATAGTAAACCGTGTAACAGCAAGGTTCGGCCTA
GATGCATCACAGCAAATCCGTCTGGAAGCAGAAGAAGAAAAATTAGCTCATTTGCCA
CTCGCTAAAACAGTATCCGTAGAAGAGCTTCTGCAATACGTGGAGCTTCAAGATCCT
GTTACGCGCACGCAGCTTCGCGCAATGGCTGCTAAAACGGTCTGCCCGCCGCATAAA
GTAGAGCTTGAAGCCTTGCTTGAAAAGCAAGCCTACAAAGAACAAGTGCTGGCAAAA
CGTTTAACAATGCTTGAACTGCTTGAAAAATACCCGGCGTGTGAAATGAAATTCAGC
GAATTTATCGCCCTTCTGCCAAGCATACGCCCGCGCTATTACTCGATTTCTTCATCA
CCTCGTGTCGATGAAAAACAAGCAAGCATCACGGTCAGCGTTGTCTCAGGAGAAGCG
TGGAGCGGATATGGAGAATATAAAGGAATTGCGTCGAACTATCTTGCCGAGCTGCAA
GAAGGAGATACGATTACGTGCTTTATTTCCACACCGCAGTCAGAATTTACGCTGCCA
AAAGACCCTGAAACGCCGCTTATCATGGTCGGACCGGGAACAGGCGTCGCGCCGTTT
AGAGGCTTTGTGCAGGCGCGCAAACAGCTAAAAGAACAAGGACAGTCACTTGGAGAA
GCACATTTATACTTCGGCTGCCGTTCACCTCATGAAGACTATCTGTATCAAGAAGAG

CTTGAAAACGCCCAAAGCGAAGGCATCATTACGCTTCATACCGCTTTTTCTCGCATG
CCAAATCAGCCGAAAACATACGTTCAGCACGTAATGGAACAAGACGGCAAGAAATTG
ATTGAACTTCTTGATCAAGGAGCGCACTTCTATATTTGCGGAGACGGAAGCCAAATG
GCACCTGCCGTTGAAGCAACGCTTATGAAAAGCTATGCTGACGTTCACCAAGTGAGT
GAAGCAGACGCTCGCTTATGGCTGCAGCAGCTAGAAGAAAAAGGCCGATACGCAAAA
GACGTGTGGGCTGGGCATCATCATCATCATTAA–3′

BM3 (CYP102A1)_F87V translation 1055 aa with a C-terminal 6xHis-tag

MTIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADELGEIFKFEAPGRVTRYLSSQR
LIKEACDESRFDKNLSQALKFVRDFAGDGLFTSWTHEKNWKKAHNILLPSFSQQAMK
GYHAMMVDIAVQLVQKWERLNADEHIEVPEDMTRLTLDTIGLCGFNYRFNSFYRDQP
HPFITSMVRALDEAMNKLQRANPDDPAYDENKRQFQEDIKVMNDLVDKIIADRKASG
EQSDDLLTHMLNGKDPETGEPLDDENIRYQIITFLIAGHETTSGLLSFALYFLVKNP
HVLQKAAEEAARVLVDPVPSYKQVKQLKYVGMVLNEALRLWPTAPAFSLYAKEDTVL
GGEYPLEKGDELMVLIPQLHRDKTIWGDDVEEFRPERFENPSAIPQHAFKPFGNGQR
ACIGQQFALHEATLVLGMMLKHFDFEDHTNYELDIKETLTLKPEGFVVKAKSKKIPL
GGIPSPSTEQSAKKVRKKAENAHNTPLLVLYGSNMGTAEGTARDLADIAMSKGFAPQ
VATLDSHAGNLPREGAVLIVTASYNGHPPDNAKQFVDWLDQASADEVKGVRYSVFGC
GDKNWATTYQKVPAFIDETLAAKGAENIADRGEADASDDFEGTYEEWREHMWSDVAA
YFNLDIENSEDNKSTLSLQFVDSAADMPLAKMHGAFSTNVVASKELQQPGSARSTRH
LEIELPKEASYQEGDHLGVIPRNYEGIVNRVTARFGLDASQQIRLEAEEEKLAHLPL
AKTVSVEELLQYVELQDPVTRTQLRAMAAKTVCPPHKVELEALLEKQAYKEQVLAKR
LTMLELLEKYPACEMKFSEFIALLPSIRPRYYSISSSPRVDEKQASITVSVVSGEAW
SGYGEYKGIASNYLAELQEGDTITCFISTPQSEFTLPKDPETPLIMVGPGTGVAPFR
GFVQARKQLKEQGQSLGEAHLYFGCRSPHEDYLYQEELENAQSEGIITLHTAFSRMP
NQPKTYVQHVMEQDGKKLIELLDQGAHFYICGDGSQMAPAVEATLMKSYADVHQVSE
ADARLWLQQLEEKGRYAKDVWAGHHHHHH

OleP Wild type 1284 bp DNA with an N-terminal 6xHis-tag

5′–ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGC
AGCCATATGACCGATACGCATACCGGCCCGACCCCGGCAGACGCAGTCCCGGCATAC
CCGTTTAGCCTGCCGCACGCCCTGGACCTGGACCCGCATTATGCGGAACTGCGTCGC
GATGAACCGGTCAGCCGTGTGCGTCTGCCGTACGGCGAGGGTACGGCATGGCTGGTC
ACCCGCATGTCTGACGCTCGTATTGTGCTGGGCGATAGTCGCTTCTCCACCGCCGCG
GCAACCGATCCGGCAACCCCGCGTATGTTCCCGACCCCGCCGGAACCGGATGGTGTT
CTGGCCCAGGATCCGCCGGACCATACCCGTCTGCGTCGCCTGGTTGGTAAAGCGTTT
ACCGCCCGTCGCGTCGAAGAAATGCGTCCGCGCGTTCGTAGCCTGGTCGACTCTCTG
CTGGATGACATGGTTGCGCACGGCAGTCCGGCCGATCTGGTTGAATTTCTGGCAGTC
CCGTTCCCGGTCGCTGTGATTTGCGAACTGCTGGGTGTGCCGCTGGAAGATCGCGAC
CTGTTTCGTACGTTCAGCGATGCGATGCTGAGCTCTACCCGCCTGACGGCTGCGGAA
ATCCAGCGTGTGCAGCAAGACTTTATGGTTTATATGGATGGCCTGGTCGCACAACGT
CGCGATGCTCCGACCGAAGATCTGCTGGGTGCTCTGGCGCTGGCCACGGATAACGAT
GACCATCTGACCAAAGGCGAAATTGTGAATATGGGTGTTAGTCTGCTGATCGCCGGT
CATGAAACGTCCGTGAACCAGATCACCAATCTGGTTCACCTGCTGCTGACCGAACGC
AAACGTTATGAATCTCTGGTTGCAGATCCGGCTCTGGTCCCGGCAGCAGTGGAAGAA
ATGCTGCGCTACACGCCGCTGGTTTCAGCGGGCTCGTTTGTTCGTGTCGCCACCGAA
GACGTGGAACTGTCAACCGTGACCGTTCGTGCGGGTGAACCGTGCGTGGTTCACTTC

GCATCGGCTAACCGTGATGAAGAAGTGTTTGATCATGCAGACGAACTGGATTTCCAC
CGCGAACGTAATCCGCATATTGCGTTTGGCCACGGTGCCCATCACTGTATCGGCGCC
CAGCTGGGTCGTCTGGAACTGCAAGAAGCGCTGAGCGCACTGGTTCGTCGCTTCCCG
ACCCTGGATCTGGCAGAACCGGTGGCTGGTCTGAAATGGAAACAGGGTATGCTGATT
CGCGGTCTGGAACGTCAAATCGTGAGCTGGTAA–3′

OleP Wild type translation 427 aa with an N-terminal 6xHis-tag

MGSSHHHHHHSSGLVPRGSHMTDTHTGPTPADAVPAYPFSLPHALDLDPHYAELRRD
EPVSRVRLPYGEGTAWLVTRMSDARIVLGDSRFSTAAATDPATPRMFPTPPEPDGVL
AQDPPDHTRLRRLVGKAFTARRVEEMRPRVRSLVDSLLDDMVAHGSPADLVEFLAVP
FPVAVICELLGVPLEDRDLFRTFSDAMLSSTRLTAAEIQRVQQDFMVYMDGLVAQRR
DAPTEDLLGALALATDNDDHLTKGEIVNMGVSLLIAGHETSVNQITNLVHLLLTERK
RYESLVADPALVPAAVEEMLRYTPLVSAGSFVRVATEDVELSTVTVRAGEPCVVHFA
SANRDEEVFDHADELDFHRERNPHIAFGHGAHHCIGAQLGRLELQEALSALVRRFPT
LDLAEPVAGLKWKQGMLIRGLERQIVSW

cc33 Wild type 1496 bp DNA

5′–ATGGCCCTGCTGCTGGCAGTGTTTCTGGTGCTGCTGTATCTGTATGGTACCCAT
AGCCACGGTCTGTTTAAAAAACTGGGTATTCCGGGCCCGAAACCGCTGCCGTTTCTG
GGCACCATTCTGGGTTATCGTAAAGTTAGCGGTTTTGCAGATTTTGATATTGAATGC
TATAAGAAGTACGGTAAAATGTGGGGTCTGTATGATGGTCGCCAGCCGGTTCTGGCA
ATTACCGATACCGATATGATTAAGACCGTGTTTGTGAAAGAATGTTATAGCGTGTTT
ACCAATCGCCGTTTTCTGGGTCCGATTGGCTTTATGAAAAAAGCAGTGAGTATTAGC
GAAGATGAAGAATGGAAACGTATTCGCACCCTGCTGAGTCCGACCTTTACCAGTGGC
AAACTGAAAGAAATGCTGCCGATTATTACCCAGTATGCAGATATGCTGGTTAAAAAT
CTGCGCATGAAAGCAGAAAAAGATAATGCCATTACCATGAAAGATATCTTTGGCGCC
TATAGTATGGATGTGATTACCGGCACCAGTTTTGGTGTGAATGTGGATAGCCTGAAT
AATCCGCAGGATCCGTTTGTGGAAAAAACCAAAAAACTGCTGAAATTTGACTTCTTT
GACCCGATTCTGTTTAGTGTGGTGCTGTTTCCGTTTCTGACCAGTTTTTATGAAGCC
CTGAATATTAGCATGTTTCCGAAAGATGTGACCGATTTTCTGAAAAGTAGTGTTGAA
CGTATGAAAAAAGACCGCCTGCAGGATAAAACCAAACATCGTGTGGATCTGCTGCAG
CTGATGATTAATAGCCAGAATAGCAAAGATACCGAAAGCCATAAAGTGCTGAGTGAT
CTGGAAATTGTTGCACAGAGCATTATTTTTATCTTCGCAGGTTATGAAACCACCAGC
AGCGCCCTGAGCTTTGTGACCTATCTGCTGGCAACCCATCCGGATGTTCAGAAAAAA
CTGCAGGAAGAAATTGATGTTACCTTTCCGAATAAGGCCCCGGCCACCTATGATGCA
CTGGTGCAGATGGAATATCTGGATATGGTTGTGAATGAAACCCTGCGCCTGTATCCG
ATTGCCGGTCGTCTGGAACGTCTGTGCAAAACCGATGTTGAAATTAATGGTGTTTTC
ATTCCGAAAGGCACCGTTGTGATGGTTCCGAGCTATGCCCTGCATCGCGATCCGAAA
CTGTGGCAGGAACCGGATGAATTTCATCCGGAACGTTTTAGCAAACAGAATAAGGAT
AGTATTGACCCGTATATCTATATGCCGTTTGGCAATGGCCCGCGTAATTGCATTGGC
ATGCGTTTTGCACTGATGAATATGAAAGTTGCCCTGGTGCGCGTGATTCAGAATTTT
ACCTTTAAAACCTGCAAGGAAACCCAGATTCCGGTGAAACTGGGTAAACAGGGTCTG
CTGCAGGCCGAAAAACCGATTGTTCTGAAAGTTGTTCCGCGTGATGGTATTATTAGC
GGTGCCTAAGAATTC–3′

cc33 Wild type translation 498 aa

MAELNYIPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVVSM
VERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVPALFLDVSDQTP
INSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSARLRLAGWHKYLTRNQIQP
IAEREGDWSAMSGFQQTMQMLNEGIVPTAMLVANDQMALGAMRAITESGLRVGADIS
VVGYDDTEDSSCYIPPLTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKR
KTTLAPNTQTASPRALADSLMQLARQVSRLESGQ

cc35 Wild type 1494 bp DNA

5′—ATGGCACTGCTGCTGGCAGTGTTTCTGGTTCTGCTGTATCTGTATGGCACCCAT
AGTCATGGTCTGTTTAAAAAACTGGGCATTCCGGGCCCGAAACCGCTGCCGTTTCTG
GGTACCATTCTGGGTTATCGCAAAGTTAGCGGCGTTGCCGATTTTGATATTGAATGC
TATAAAAAGTACGGCAAAATGTGGGGTCTGTATGATGGTCGCCAGCCGGTTCTGGTT
ATTACCGATACCGATATGATTAAGACCGTTTTTGTTAAAGAGTGCTATAGTGTGTTT
ACCAATCGCCGTTTTCTGGGTCCGGTTGGTTTTATGAAAAAAGCCGTGAGTATTAGC
AAAGATGAAGAATGGAAACGTATTCGTACCCTGCTGAGTCCGACCTTTACCAGTGGT
AAACTGAAAGAAATGTTTCCGATTATTACCCAGTATACCGATGTGCTGGTGAAAAAT
CTGCGCATGAAAGTGGAAAATGATAATACCGTTAACATGAAAGACATCTTTGGTGCC
TATAGTATGGATGTGATTACCGGCACCAGCTTTGGCGTGAATGTGGATAGTCTGAAT
AATCCGCAGGATCCGTTTGTTGAAAAAAGTAAAAAACTGCTGAAGCTGGATCTGTTT
GATCCGATTCTGTTTAGTGTTGTGCTGTTTCCGTTTCTGACCAGCTTTTATGAAGCC
CTGAATATTAGTATTTTCCCGAAAGTTAGTACCAATTTCTTTAAAAGTAGCGTGGAA
AAAATCAAGAAGGATCGTCTGCAGGATAAAACCAAACATCGTGTGGATCTGCTGCAG
CTGATGATTAATAGCCAGAATAGCAAAGATACCGAAAGCCATAAAGCACTGAGTGAT
CTGGAAATTATTGCACAGAGTATTATTTTCATCTTCGCAGGTTATGAAACCACCAGC
AGTGCCCTGAGCTTTGTGACCTATCTGCTGGCAACCCATCCGGATGTGCAGAAAAAA
CTGCAGGAAGAAATTGATGTGACCTTTCCGAATAAGGCACCGGCAACCTATGATAGC
CTGGTTCAGATGGAATATCTGGATATGGTGGTTAATGAAACCCTGCGTCTGTATCCG
ATTGCAAGTCGCCTGGGTCGTGTTTGTAAAACCGATATTGAAATTAAGGGCGTTTTT
ATTCCGAAAGGTACAATGGTTATGGCCCCGATCTATGCACTGCATCGTGATCCGAAA
CTGTGGCAGGAACCGGAAGAATTTCGTCCGGAACGCTTTAGCAAACAGAATAAGGAT
AGTATTGACCCGTATATTTTCATGCCGTTTGGCAATGGTCCGCGTAATTGCATTGGT
ATGCGCTTTGCACTGATGAATATGAAAGTGGCCCTGGTTCGTGTTCTGCAGAATTTT
ACCTTTAAACCGTGCAAAGAAACCCAGATTCCGATGAAACTGAATACCCAGGGTCTG
CTGCAGGCAGAAAAACCGATTGTTCTGAAAGTTGTTCCGCGTGATGGTATTATTAGC
GGCGCATAAGAATTC-3′

cc35 Wild type translation 495 aa

MALLLAVFLVLLYLYGTHSHGLFKKLGIPGPKPLPFLGTILGYRKVSGVADFDIECY
KKYGKMWGLYDGRQPVLVITDTDMIKTVFVKECYSVFTNRRFLGPVGFMKKAVSISK
DEEWKRIRTLLSPTFTSGKLKEMFPIITQYTDVLVKNLRMKVENDNTVNMKDIFGAY
SMDVITGTSFGVNVDSLNNPQDPFVEKSKKLLKLDLFDPILFSVVLFPFLTSFYEAL
NISIFPKVSTNFFKSSVEKIKKDRLQDKTKHRVDLLQLMINSQNSKDTESHKALSDL
EIIAQSIIFIFAGYETTSSALSFVTYLLATHPDVQKKLQEEIDVTFPNKAPATYDSL
VQMEYLDMVVNETLRLYPIASRLGRVCKTDIEIKGVFIPKGTMVMAPIYALHRDPKL
WQEPEEFRPERFSKQNKDSIDPYIFMPFGNGPRNCIGMRFALMNMKVALVRVLQNFT
FKPCKETQIPMKLNTQGLLQAEKPIVLKVVPRDGIISGA

CYP3A4 Wild type 1488 bp DNA

5′—ATGGCCCTGCTGCTGGCCGTTTTTCTGGTTCTGCTGTATCTGTATGGTACCCAT
AGTCATGGCCTGTTTAAAAAACTGGGTATTCCGGGCCCGACCCCGCTGCCTTTTCTG
GGTAATATTCTGAGTTATCATAAGGGTTTTTGCATGTTTGATATGGAATGTCATAAG
AAATACGGTAAAGTGTGGGGCTTTTATGATGGTCAGCAGCCGGTTCTGGCCATTACC
GATCCGGATATGATTAAGACCGTTCTGGTTAAAGAATGCTATAGTGTTTTTACCAAC
CGTCGCCCGTTTGGCCCGGTGGGCTTTATGAAAAGCGCAATTAGTATTGCAGAAGAT
GAAGAATGGAAACGTCTGCGTAGCCTGCTGAGTCCGACCTTTACCAGTGGTAAACTG
AAAGAAATGGTGCCGATTATTGCCCAGTATGGTGACGTTCTGGTTCGCAATCTGCGT
CGCGAAGCCGAAACCGGCAAACCGGTGACCCTGAAAGATGTTTTTGGTGCATATAGC
ATGGATGTTATTACCAGTACCAGCTTTGGTGTTAATATTGATAGCCTGAATAACCCG
CAGGATCCGTTTGTGGAAAATACCAAAAAACTGCTGCGTTTTGATTTTCTGGATCCG
TTTTTCCTGAGCATTACCGTTTTTCCGTTTCTGATTCCGATTCTGGAAGTGCTGAAT
ATTTGCGTGTTTCCGCGCGAAGTTACCAATTTTCTGCGCAAAAGCGTGAAACGTATG
AAAGAAAGTCGTCTGGAAGATACCCAGAAACATCGTGTGGATTTTCTGCAGCTGATG
ATTGATAGCCAGAATAGCAAAGAAACCGAAAGTCATAAAGCCCTGAGCGATCTGGAA
CTGGTGGCACAGAGTATTATTTTTATTTTTGCCGGTTACGAGACCACCAGCAGCGTT
CTGAGTTTTATTATGTATGAACTGGCAACCCATCCGGATGTGCAGCAGAAACTGCAG
GAAGAAATTGATGCCGTTCTGCCGAATAAGGCACCGCCGACCTATGATACCGTTCTG
CAGATGGAATATCTGGATATGGTTGTTAATGAAACCCTGCGCCTGTTTCCGATTGCA
ATGCGCCTGGAACGTGTGTGCAAAAAAGATGTGGAAATTAATGGTATGTTCATCCCG
AAAGGTGTTGTTGTGATGATTCCGAGTTATGCACTGCATCGTGATCCGAAATATTGG
ACCGAACCGGAAAAATTTCTGCCGGAACGCTTTAGCAAAAAGAATAAGGATAATATC
GACCCGTATATCTATACCCCGTTTGGTAGTGGCCCGCGTAATTGCATTGGTATGCGC
TTTGCACTGATGAATATGAAACTGGCCCTGATTCGCGTGCTGCAGAATTTTAGTTTT
AAACCGTGTAAAGAGACCCAGATTCCGCTGAAACTGAGTCTGGGTGGTCTGCTGCAG
CCGGAAAAACCGGTTGTGCTGAAAGTGGAAAGCCGTGATGGTACCGTTAGTGGTGCC
TAAGAATTC-3′

CYP3A4 Wild type translation 496 aa

MALLLAVFLVLLYLYGTHSHGLFKKLGIPGPTPLPFLGNILSYHKGFCMFDMECHKK
YGKVWGFYDGQQPVLAITDPDMIKTVLVKECYSVFTNRRPFGPVGFMKSAISIAEDE
EWKRLRSLLSPTFTSGKLKEMVPIIAQYGDVLVRNLRREAETGKPVTLKDVFGAYSM
DVITSTSFGVNIDSLNNPQDPFVENTKKLLRFDFLDPFFLSITVFPFLIPILEVLNI
CVFPREVTNFLRKSVKRMKESRLEDTQKHRVDFLQLMIDSQNSKETESHKALSDLEL
VAQSIIFIFAGYETTSSVLSFIMYELATHPDVQQKLQEEIDAVLPNKAPPTYDTVLQ
MEYLDMVVNETLRLFPIAMRLERVCKKDVEINGMFIPKGVVVMIPSYALHRDPKYWT
EPEKFLPERFSKKNKDNIDPYIYTPFGSGPRNCIGMRFALMNMKLALIRVLQNFSFK
PCKETQIPLKLSLGGLLQPEKPVVLKVESRDGTVSGA

SGG4 Wild type 1587 bp DNA

5′—ATGGGCACCACCCCGTTTCATCATGAACCGGGCACCGTGCCGCCGCCGCAATGT
CCTGCTCATAATCTGGATATTGGCCCGGGCGGCCTGCGCCGTCTGCATGGTCCTGAA
GCAGAAAATAATCCGGCAGGCCTGTATGATAAACTGCGTGCCGAACATGGCACCGTG
GCCCCGATTCTGCTGCATGGTGACGTTCCGGCATGGCTGGTGCTGGGTCATAGTGAA
AATCTGCATGTGACCCGCACCCCGAGTCAGTTTAGTCGTGATAGCCGTCGTTGGCGC
GCACTGCAGGATGGTAGCGTGGCCCCGGATCATCCGCTGGCACCGATTTTTACCTGG

CAGCCGATTTGTGTTTTTGCCGATGGCGCAAAACATGAACGCCAGCGTGGTGCCGTG
ACCGATAGCATGGAACGTATTGATACCCGTGGTGTTCGCCGTCATATTAATCGCTTT
AGTAATCGCCTGGTTAATGATTTTTGCGAAAAAGGTACCGCCGATCTGGTGGGCCAG
TTTGCCGAACATCTGCCGATGATGGTTGTGTGCGCCATTTTTGGCATGCCGGAAGAA
TATGATGAACGCCTGGTTCAGGCCGCCCGTGATATGACCCGTGGTACCGAAACCGCA
GTTGCAAGCAATGCCCATATTGTTAGTGTTCTGACCCGTCTGGTTGAACGTCGCCGT
GCCGAACCGAGCCCGGATCTGGCAAGCTGGCTGGTTGAACATCCGGCAACCATGACC
GATACCGAAGTGATTGAACATCTGCGTCTGATTATGATTGCCGCATATGAAAGTACC
GCAAATCTGATTGCAAATGTGCTGCGTATGGTGCTGATTGATCCGCGCTTTCGTGCC
CGCCTGAGTGGTGGTCATATGACCGTGCCGGAAGCAGTTGAACAGACCCTGTGGGAT
GAACCGCCGTTTACCGCCGTGTTTGGTCGTTGGGCCGTTGGCGATACCGAACTGGGC
GGTCAGCAGATTAAGGCAGGCGATGCCCTGCTGGTGGGCATTGCACCGGCCAATACC
GATCCGACCGTTCGTCCGGATCTGGGTGCAGATATGGGTGGTAATCGTGCCCATCTG
GCATTTTCTGGTGGTCCGCATGAATGTCCGGGCCAGGATATTGGTCGCGCAATTGCC
GATGTTGGCGTTGATGCCCTGTTAATGCGCCTGCCGGATCTGGAACTGGGCGTTGGC
GAAAGTGAACTGCATTGGGTGGGCAATATTATGAGTCGTCATCTGGTGGAACTGCCG
GTGAAATTTGCCCCGGGCCCGCAGCAGAAACTGGATGCCGATCCGCTGACCGTTATG
GCCCGCGCCCCGAGACCGGCAGATGCTTGGGAAATTAGTAGTCCGGCCCGTCAGGTG
CCGGAACCGCGTCATGAAGCCGTGGTGGCACAGCCGGCACATGCCCCTGGTGCAGCA
CCGACCGCAGAACCGGATCCGGCACCGGCAGCCCCTCCTGCACCTGAACCGGCAGCC
GCACCGGAACCGGCTCCTGTTGCCACCATTCCGCAGCAGCGCCGTCCGGCTGCACCG
GCAAGATTTTGGCAGGCCGTTACCCGCTGGTGGAGTGGTTATTAAGAATTC—3′

SGG4 Wild type translation 529 aa

MGTTPFHHEPGTVPPPQCPAHNLDIGPGGLRRLHGPEAENNPAGLYDKLRAEHGTVA
PILLHGDVPAWLVLGHSENLHVTRTPSQFSRDSRRWRALQDGSVAPDHPLAPIFTWQ
PICVFADGAKHERQRGAVTDSMERIDTRGVRRHINRFSNRLVNDFCEKGTADLVGQF
AEHLPMMVVCAIFGMPEEYDERLVQAARDMTRGTETAVASNAHIVSVLTRLVERRRA
EPSPDLASWLVEHPATMTDTEVIEHLRLIMIAAYESTANLIANVLRMVLIDPRFRAR
LSGGHMTVPEAVEQTLWDEPPFTAVFGRWAVGDTELGGQQIKAGDALLVGIAPANTD
PTVRPDLGADMGGNRAHLAFSGGPHECPGQDIGRAIADVGVDALLMRLPDLELGVGE
SELHWVGNIMSRHLVELPVKFAPGPQQKLDADPLTVMARAPRPADAWEISSPARQVP
EPRHEAVVAQPAHAPGAAPTAEPDPAPAAPPAPEPAAAPEPAPVATIPQQRRPAAPA
RFWQAVTRWWSGY

SCE9 Wild type 1239 bp DNA

5′—ATGGGCACCCGCATTGCCCTGGATCCGTTTGTGCGTGATCTGGATGGTGAAAGT
GCAGCACTGCGTGCAGCCGGCCCGCTGGCAGAAGTTGAACTGCCGGGTGGCGTTCAT
GTGTATGCAGTGACCCGCCATGCCGAAGCACGCGCCCTGCTGACCGATAGTCGCGTT
GTGAAAGATATTAATGTTTGGGGTGCCTGGCAGCGCGGTGAAATTCCGATGGATTGG
CCGCTGATTGGTCTGGCCAATCCGGGTCGCAGTATGCTGACCGTTGATGGCGCAGAT
CATCGTCGTCTGCGCACCCTGGTTGCACAGGCCCTGACCGTGAAACGTGTGGAACGC
CTGCGCGCCGGTATTGAAGCACTGACCAATGCCAGTCTGGAAAAACTGGCAGCACTG
CCGGCAGGTCAGCCGGTTGATCTGAAAGCAGAATTTGCCTATCCGCTGCCGATGAAT
GTTATTAGCGAACTGATGGGCGTTGATGCCGCAGATCATCCGCGCCTGAAAGAACTG
TTTGAAAAATTTTTCAGCACCCAGACCCCGCCGGAAGAAGTTCCGCAGATGATGGCC
GATCTGGGTACCCTGTTTACCAAAATTGTTGATGATAAACGCGCCAATCCGGGCGAT

GATCTGACCAGCGCACTGATTGCAGCAAGCGAAGATGGTGACCATCTGACCGATGAA
GAAATTGTGAATACCCTGCAGCTGATTATTGCCGCCGGTCATGAAACCACCATTAGT
CTGATTGTTAATGTTGTTGAAGCACTGCAGACCCATCCGGAACAGCGCAAAAAAGTG
CTGAATGGTGAAATTGGCTGGGATGGTGTGATTGAAGAAACCCTGCGTTGGAATACC
CCGACCAGTCATGTTCTGATTCGCTTTGCCACCGAAGATATTGAAGTTGGTGACAAA
ATTCTGCCGAAAGGCGAAGGCCTGATTATTAGCTTTGGTGCACTGGGCCGCGATGAA
GAACAGTATGGTCCGACCGCAGGTGAATTTGATGCCACCCGTACCCCGAATCGTCAT
ATTGCCTTTGGTCATGGCCCGCATGTTTGTCCGGGCGCAGCACTGAGTCGCCTGGAA
GCAGGTATTGCCCTGCCGGCACTGTATGAACGTTTTCCGGAACTGGAACTGGCCGTT
CCGGCAACCGAACTGCGTAATAAGCCGATTGTTACCCAGAATGATCTGTATGAACTG
CCGGTTGAACTGGGTTGTCCGTTTGGCCATGCCAGTTAAGAATTC—3′

SCE9 Wild type translation 413 aa

MGTRIALDPFVRDLDGESAALRAAGPLAEVELPGGVHVYAVTRHAEARALLTDSRVV
KDINVWGAWQRGEIPMDWPLIGLANPGRSMLTVDGADHRRLRTLVAQALTVKRVERL
RAGIEALTNASLEKLAALPAGQPVDLKAEFAYPLPMNVISELMGVDAADHPRLKELF
EKFFSTQTPPEEVPQMMADLGTLFTKIVDDKRANPGDDLTSALIAASEDGDHLTDEE
IVNTLQLIIAAGHETTISLIVNVVEALQTHPEQRKKVLNGEIGWDGVIEETLRWNTP
TSHVLIRFATEDIEVGDKILPKGEGLIISFGALGRDEEQYGPTAGEFDATRTPNRHI
AFGHGPHVCPGAALSRLEAGIALPALYERFPELELAVPATELRNKPIVTQNDLYELP
VELGCPFGHAS

SGDX Wild type 1278 bp DNA

5′—ATGGGCAGTACCAGCAGCCCGAGCTTTGGTCCGCAGGCCCCGGCAAGTTGCCCG
GTTGGTGCAGGCGCCGGTGCCGTTCGTCTGAGCGGTGCAAGTTATCAGCAGACCCCG
ACCCAGCTGTATCGCAGTCTGCGCCGTGATCATGGTGCAGTTGCACCGGTGCTGCTG
GATGGCGATGTGCCGGCATGGCTGGTTCTGGGCTATGCAGAACTGAGCTATGTTCTG
ACCCATGATGAACTGTTTGCACGTGATAGCCGCCGTTGGAATCAGTGGGAAACCATT
CCGCCGGATTGGCCGCTGATGCCGTTTGTGGGCTATCAGCCGAGTGTGCTGTTTACC
GAAGGCGATGAACATCGTCGCCGTGCCGGTGTTATTACCGAAGCCCTGGAAGGTATT
GATCAGTTTGAACTGGCCCGCGATTGCCGTCGCATTGCAGATCGCCTGATTGCCGAT
TTTGCCGGTAGCGGCCGTACCGAACTGATGAGTAGTTATGTGCATGCCCTGCCGATG
CGTGCAGTTGTGGAAATGTGCGGCATGCCGGTGAGTGGTAGCGATACCCAGCAGCTG
GTTGATGATCTGCGCATTAGTCTGGATGCCGGTGAAGGCGATGATCCGGTGGCCGCC
TATGGCCGTGTTGGTGACCGTCTGCGCCAGCTGGTTGAAGATAAACGCGCCGCACCG
GGTGCAGATATTACCAGTCGTATGGTGACCCACGGTGCCGGCCTGACCGATGAAGAA
ATTGTTCAGGATCTGATTAGCGTTATTGCAGCAGCCCAGCAGCCGACCGCAAATTGG
ATTTGCAATACCCTGCGCCTGCTGCTGACCGATGAGCGCTTTGCCCTGAATGTGAGT
GGCGGCCGTCTGAGCGTGGGTGAAGCCCTGAATGAAGTTCTGTGGCTGGATACCCCG
ACCCAAAATTTTATTGGTCGCTGGGCAGTTCGCGATACCCAGCTGGGTGGTCGTCAT
ATTCGCGCAGGCGATTGCCTGGTTCTGGGTCTGGCAGCAGCCAATACCGATCCGGAA
ATTTGGCCGGAAAGCTATGTGGGTGCCGAAAATAGTGCCCATCTGAGTTTTAGCGGC
GGCGAACATCGTTGCCCGTATCCGGCACCGCTGCTGGCCGATGTTATGGCCCGCACC
GCAGTTGAAACCCTGCTGGAACAGCTGCCGGATCTGATGCTGGCAGTGGATCCGACC
GAACTGAGTTGGCGTCCGAGTATTTGGATGCGCGGTCTGAGCACCCTGCCGGTTCAG
TTTAGTCCGATGGCCCAGTAAGAATTC—3′

SGDX Wild type translation 426 aa

MGSTSSPSFGPQAPASCPVGAGAGAVRLSGASYQQTPTQLYRSLRRDHGAVAPVLLD
GDVPAWLVLGYAELSYVLTHDELFARDSRRWNQWETIPPDWPLMPFVGYQPSVLFTE
GDEHRRRAGVITEALEGIDQFELARDCRRIADRLIADFAGSGRTELMSSYVHALPMR
AVVEMCGMPVSGSDTQQLVDDLRISLDAGEGDDPVAAYGRVGDRLRQLVEDKRAAPG
ADITSRMVTHGAGLTDEEIVQDLISVIAAAQQPTANWICNTLRLLLTDERFALNVSG
GRLSVGEALNEVLWLDTPTQNFIGRWAVRDTQLGGRHIRAGDCLVLGLAAANTDPEI
WPESYVGAENSAHLSFSGGEHRCPYPAPLLADVMARTAVETLLEQLPDLMLAVDPTE
LSWRPSIWMRGLSTLPVQFSPMAQ

NE05 Wild type 1227 bp DNA

5′—ATGGGCACCAATCGTCCGGCATTTCCGGATGCCAATGAACGCCCGCTGGATACC
TGTCTGCGTCTGCGCAAAACCGCCCCGGTTGTGGAAGTGGATTTTCCGGGCGGCGTG
CCGGCATATCTGGCACTGAGCCATCGTGCCGTGGAAGAAATTCTGGCAGGTGACAAT
AAGACCTTTGCCCGTGATCCGAAACATTGGCCGGCCCTGTATGATGGCAGCATTCCG
GAAGATTGGCCGTTTCGTGCCATTGTGCAGGGCGATCATCTGAGTACCAAAGATGGC
GCCGATCATCGTCGTCTGCGCGGTCTGGTTGGTAAAGGTTTTACCCCGGCCCGCGTT
CGCGATCTGGAACCGCGTATTCAGGCAATTATTGATGGCCTGCTGGATGGTGTTGCC
GCAGCAGGTGACACCGTGGATCTGGTTCCGGCATTTTGCGATGCACTGCCGATGGCC
GTTATTAGCGAACTGTTTGGTGTGCCGGAAAGCGAACGCGCACAGCTGCGCCAGTGG
ACCCTGACCAGCCTGAGTCAGGAAATTAGCGCCGAAGAAGCCTTTAGCACCCAGGTG
GCCCTGCGTGGCTATCTGTATGAACTGGTGGAACGTAAACAGCGTGAACCGGGTGAC
GATCTGACCAGTGATCTGGTGCGCGCCCGTGATGAAGGTGACCGCCTGACCACCGAT
GAACTGGTTGCAATTCTGTGGCTGATGCTGATTGCAGGCCATGAAACCACCGTGTAT
CTGCTGGGCAATGCAGTGGTTGCCCTGGGTCGTCATCCGGATCAGCTGGCCCTGGTG
AAAGCAGAAGATCGTTGGGCAGATGTGGTGGAAGAAACCCTGCGTTATCGTCATAGC
GTTATGATGACCAGCTTTCGCTTTACCCTGCAGGATGTTACCATTGATGGCGTGCCG
ATTCCGAAAGGTAATGCCGTTGGTGTGGTTTATCAGGCAACCGGTGTGGATACCGCA
CAGCATGGTGAAACCGCAGATGAATTTGATATTACCCGCCCGCCGCGCCCGCATCTG
GGTTTTGGTCATGGTCCGCGCTATTGTATTGGCGCACCGCTGGCACGCCTGGAAGGC
AGACTGGGTCTGACCAGCTTATATCGTCGCTTTCCGGATCTGACCCTGGCAGTTGAT
CCGGCCGAAATTCCGTATACCCCGAGCTTTTTCACCGTGGGTCCGGTTAGCATTCCG
GTGCGCCCGGGCGCTGATCGCGGTTAAGAATTC-3′

NE05 Wild type translation 409 aa

MGTNRPAFPDANERPLDTCLRLRKTAPVVEVDFPGGVPAYLALSHRAVEEILAGDNK
TFARDPKHWPALYDGSIPEDWPFRAIVQGDHLSTKDGADHRRLRGLVGKGFTPARVR
DLEPRIQAIIDGLLDGVAAAGDTVDLVPAFCDALPMAVISELFGVPESERAQLRQWT
LTSLSQEISAEEAFSTQVALRGYLYELVERKQREPGDDLTSDLVRARDEGDRLTTDE
LVAILWLMLIAGHETTVYLLGNAVVALGRHPDQLALVKAEDRWADVVEETLRYRHSV
MMTSFRFTLQDVTIDGVPIPKGNAVGVVYQATGVDTAQHGETADEFDITRPPRPHLG
FGHGPRYCIGAPLARLEGRLGLTSLYRRFPDLTLAVDPAEIPYTPSFFTVGPVSIPV
RPGADRG

WP_091596541 Wild type 1383 bp DNA

5′–ATGGCCACCGCCCTGCCTCATCCGCCGCGTAGAATCCCTCTGATTGGTGATGTG
CTGGGCGTTAGTCCGAAAACACCTGTGCAGGATAGCATGCGCCATGCCGCAGAACTG
GGTCCGGTGTTCGAACGCAAAGTGTTCGGCCGTAGCATTGTGTTCGTTAATGGTGCA
GATATGGTTGCAGATCTGAGCGATGAAAAACGCTTCGCAAAACATATTACCCCGGCA
ATTAGCAATCTGCGTCCGCTGGGCGGTGATGGCCTGTTCACCGCCCATAATGAAGAA
CCGAATTGGCGCCGTGCCCATGAAATTCTGGCACCGGCATTCAGCCGTAATGCAATG
CAGCGCTATCATCCGACCATGCTGGCAATTACCCGTGAACTGCTGGATACCTGGGAT
CGTGGCGGTGAAGTTGATGTGGCCGATGATATGACCAAACTGACCCTGGAAACCATT
GGTCGCACCGGCTTCGGCTATAGCTTCAGTAGCTTCGCACGTGCCGAACCGCATCCG
TTCGTTGCAGCAATGGTGCGCACCCTGCGTCATGCCCAGCGTCGCGCAATTCAGCCG
CCGGTTATTGGTCCGCTGCTGAGTCGCAAAGCAAATGGTCGTAATGAAGCAAATCTG
GCATTCCTGCATAGTGTGGTGGCAGATGTTATTGAAGCCCGCCGCAATGATAGTAGT
ACCGAAGATCTGCTGGGTCTGATGCTGAATACCACCCAGCCGAGTACCGGTGAAGCA
CTGGATGAAGATAATATTCGTCATCAGATTATCACCTTCCTGGTTGCCGGCCATGAA
ACCACCAGCGGTGCCCTGAGCTTCGCACTGTATTATCTGGCAGGCAATCCGGATGTT
CTGGCCCGTGCCCAGAGCGAAGTTGATAAAGTGTGGGGTGATAATCCGGATCCGAGT
TATGAAGAAGTTGCAAAACTGCGTTATGTTCGTCGTGTGCTGGATGAAGCACTGCGC
CTGTGGCCGACCGCACCGGCATTCGCTCGTCAGGCACAGGTTGATACCGTGGTGGGT
GGTGAATATCCGATGCGTAAAGGCCAGTGGGCACTGGTTCTGATTCCGGCACTGCAT
CGCGATCCGGTGTGGGGCGATGATCCGGAAAGCTTCGATCCGGATCGCTTCAGCCCG
GAACGTAATCGTGCCCGTCCGGCACATGTGTATAAACCGTTCGGTACCGGCGAACGT
GCCTGTATTGGCCGTCAGTTCGCACTGCATGAAGCAACCCTGGTGCTGGGCATGGTG
CTGAGCCGTTATGAACTGCGCGGCGATCCGAGTTACCGCCTGAAAGTTCAGGAACTG
CTGACCCTGAAACCGGAAGGCTTCAAACTGCAGGTGAGCCTGCGTGATCGCGCAACC
GTTAGCGCATAACTCGAG-3′

WP_091596541 Wild type translation 461 aa

MATALPHPPRRIPLIGDVLGVSPKTPVQDSMRHAAELGPVFERKVFGRSIVFVNGAD
MVADLSDEKRFAKHITPAISNLRPLGGDGLFTAHNEEPNWRRAHEILAPAFSRNAMQ
RYHPTMLAITRELLDTWDRGGEVDVADDMTKLTLETIGRTGFGYSFSSFARAEPHPF
VAAMVRTLRHAQRRAIQPPVIGPLLSRKANGRNEANLAFLHSVVADVIEARRNDSST
EDLLGLMLNTTQPSTGEALDEDNIRHQIITFLVAGHETTSGALSFALYYLAGNPDVL
ARAQSEVDKVWGDNPDPSYEEVAKLRYVRRVLDEALRLWPTAPAFARQAQVDTVVGG
EYPMRKGQWALVLIPALHRDPVWGDDPESFDPDRFSPERNRARPAHVYKPFGTGERA
CIGRQFALHEATLVLGMVLSRYELRGDPSYRLKVQELLTLKPEGFKLQVSLRDRATV
SA

WP_034313748 Wild type 1359 bp DNA

5′–ATGACCCACGGTGCCGAAACCAGCCGTGCAGCACGTCTGGCACCGCCGGGTCCT
CCTCGTCGTGCTACCTTCGGCCTGCTGAAAAAACTGTTCACCGATCGCCTGGCCCTG
ATGAGCGATAGTGCCGAAGCACATGGTGATGTGGTTCGTATTGCAATTGGCCCGAAA
ACCATGTATCTGGTGAATCATCCGGATCTGGCAAAACATGTGCTGGCAGATAATGCC
GCCAATTATCATAAAGGTATTGGCCTGCAGGAAGCACGCCGTGCACTGGGCGATGGT
CTGCTGACCAGTGATGGCGATGTGTGGAAAAAACAGCGCCGTACCATTCAGCCGGTG
TTCCAGCCGAAACGTATTGCCCGTCAGGCAAGTGTGGTTGCAAATGAAGTTGAAGGC
CTGGTTAAACGTCTGCGCGATACCGAAGGTCCGGTTGAAATTCTGCATGAAATGACC
GGCCTGACCCTGGGCGTTCTGGGCAAAACCTTACTGGATGCCGATCTGGGTGGCTTC

ACCAGCCTGGGTCATAGCTTCGAAGCAGTTCAGGATCAGGCCATGTTCGAAGCCGTG
ACCCTGAGTATGGTGCCGCAGTGGGCACCGCTGAAAAAACAGCTGCGCTTCCGCGAA
AGCCGTGCAGATCTGCGCCGTATTGCCGATGAACTGGTGGAACAGCGCCTGGCAAAT
CCGGTGGAAAATGGTGAAGATGTGCTGAGTCGTCTGATTGCCACCGATGGTACCCGC
GAACAGATGCGCGATGAACTGATTACCCTGCTGCTGGCAGGCCATGAAACCACCGCC
AGCACCCTGGGCTGGGCATTCCATCTGCTGGATGAACATCCGGAAATTGCAGCCAAA
CTGCGCGCAGAAGCCGATGCCGTTCTGGGCGATCAGCTGCCGACCCATGATGATCTG
CATCGTCTGCCGTATACCGCACGCGTGGTTGAAGAAGTTATGCGCCTGTATCCGCCG
GTGTGGCTGCTGCCGCGTGTGGCACAGGCCGATGATGAAATTGGCGGCTATCATATT
CCGGCAGGCAGCGATGTGGTTGTTGTTCCGTATACCCTGCATCGCCATCCGGAGTTC
TGGCCGGAACCGGAAAAATTCGATCCGGATCGCTTCGATCCGGACCGTCCGAGCGGC
CGTCCGAGATATGCATATATTCCGTTCGGTGCCGGTCCGCGCTTCTGCATTGGCAAT
AGCCTGGGCGTGATGGAAGCAGTGTTCGTGCTGACAATGGTTCTGCGTGATCTGGAA
CTGCGTAAACTGCCGGGTTATGATGTTGTGCCGGAAGCAATGCTGAGCCTGCGTGTT
CGTGGTGGCCTGCCGATGACCGTTCATACCCGTAATCGCCGCTAACTCGAG—3′

WP_034313748 Wild type translation 453 aa

MTHGAETSRAARLAPPGPPRRATFGLLKKLFTDRLALMSDSAEAHGDVVRIAIGPKT
MYLVNHPDLAKHVLADNAANYHKGIGLQEARRALGDGLLTSDGDVWKKQRRTIQPVF
QPKRIARQASVVANEVEGLVKRLRDTEGPVEILHEMTGLTLGVLGKTLLDADLGGFT
SLGHSFEAVQDQAMFEAVTLSMVPQWAPLKKQLRFRESRADLRRIADELVEQRLANP
VENGEDVLSRLIATDGTREQMRDELITLLLAGHETTASTLGWAFHLLDEHPEIAAKL
RAEADAVLGDQLPTHDDLHRLPYTARVVEEVMRLYPPVWLLPRVAQADDEIGGYHIP
AGSDVVVVPYTLHRHPEFWPEPEKFDPDRFDPDRPSGRPRYAYIPFGAGPRFCIGNS
LGVMEAVFVLTMVLRDLELRKLPGYDVVPEAMLSLRVRGGLPMTVHTRNRR*LE

WP_084494404 Wild type 2577 bp DNA

5′—ATGCATATCGCAATGTTCATGGATTATCATCCGGCAACCCTGGGTGGCGTTCAG
ACCGCCGTGGCAGCACTGCGTCGCGGTCTGGAACGCGCCGGTCATCGTGTTACCCTG
TTCGTTGCCCCGCTGAGCGGTACCCCGAGTCCGGCAACCGAAGCCGATGATGGTGTG
ATTGAACTGGCCCCGCTGGCCGGTGCCGTGGTTAATGGCTTCCCGATGGTGCTGCCG
ACCGCACGCAATGCCGCCCTGATTGATGCCGCCTTCGCAGCACGCGGTCCGGTTGAT
CTGGTGCATACCCATACCACCTATGGTGTGGCAATTAGTGGTATGAAAGCAGCACGT
CGTCATCGTATTCCGCTGGTTCATACCGCACAGAGCCGCGATGATGCCTTCATTGAA
CATACCGCACCGGCCCCGTATCTGAGCGCACTGGCCCTGCGTATTCTGCATGGTCGC
TTCGTGAATCATCCGGCCCGTATGCCGCGCGGTGCACAGAGTCGTGCAGCACGTCAT
GCATGGCATACCATTGTTGGTCAGGGCCAGGCAGCCGATCGCGTTATTGCCCCGACC
CGCCACTTCGCCGATCTGCTGCTGGATCATGGCCTGACCCGCCCGATTCGTGTGGTG
AGTAATGGCGTTGATGATGCACTGGTTGATATTCGTTGTGCCGTTGAACGTGATGCA
GATGGCCCGCTGCGCCTGATCTGGTGCGGCAGACTGAGCGCAGAAAAACGTCTGCTG
GAAAGCATTGATGCAGTTCGTCGCGTGCCGGATTGTACCCTGGATGTGTATGGCGAA
GGTGAACTGTTCGATCGCGCAGCCGCAGTGGTTGCAGAACATGGCCTGGCCGATCGT
ATTACCCTGCATGGCCGTGTGAGTCAGGAACAGTGCCTGGATGCAATGGCAGCCGCA
GATGCACTGCTGTTCCCGAGTTGGGGCTTCGATACCCAGGGTATGGTGCTGCTGGAA
GCCATTGCCACCGGTCTGCCGGTGCTGTATTGTGATCCGGATCTGGCAGAAAGTATT
CCGGCCGGCGGTGGCCTGCGTGCCGATGACGCAAGTCCGGCAGCCCTGGCCGATGTT
ATTGCACGTCTGGCAACCGATCGTGCAGAACTGCTGGCAATGCGCGCCGAAATGGCC

GGTTATCGTGATAGCGTGCGCCAGAGCGGCCGTATTGATGCAATTGTGGAAGTGTAT
CGCCAGGCACGCGCAGATACCGAACCGGCCGCACAGCCGCCGGTTCCTCGTAGACTG
AGTGATGTGCCGACCGCACCGGGCGCACTGCCTCTTATTGGTCATAGCCTGCGCGCC
CTGCGTGATGCACCGGGCTTCGTTACCAGCCTGGCAGATCTGGGCCCGATTGTTCGC
ATTCGCTTCGGCCGCAAAACCGGTTATGTTCTGACCACCCCGGAACTGGTTCGCGAA
GTGGGTCTGGGCGATGCCGAACTGAATCGCGATGATCTGCGTGAAGCCATTGCAGAT
GTGGCCGGCGGTAGCGTGAATGTGCTGCGTGGTGCAGAACATAAACTGCGTCGTCGT
ATGATTGCACCGGCCCTGCGCCAGAGCCGCCTGGCAGAGTATACCGTTACCGCAGCC
GGTCTGGCCGATACCTGGAGCGCAGCCCTGCCGGCTGGTGGTCGTGTGGATCTGATG
GATGAAGCCCACGGTCTGATTCTGGATACCGTTAGTAGTACCCTGTTCACCGCCGAG
TTCAGCGAAACCGCACGTCGTCGTATTCGCGATAATGTTCCGTGGCTGCTGAGTCAG
GTGATTCTGCGCACCGCCCTGCCGCCGCAGATTCGTCGTCTGCGCGTGATTGCAAAT
CGTCGTTGGCAGCGCAAAGCACATCATCTGCGCGCAGCAATTGGTGCCGCAATTGCC
GAATATCGTCGTCGTGATGAAGACTTCAATGATGTGGTGAGCGCCCTGATTCGTCAT
ACCGATCCGGAAACCGGCGCACGCCTGAGTGATGAACATATTATTGATGAAGCCATT
CTGATGCTGGCCGGTGGTGTGGGCAGTATGGCCAGTCTGACCGGTTGGCTGTGGCAT
GAAGTGATGCGCCGTCCGGATATTGCCGAACGCATTCGCGCCGAACTGGATGAAGTG
GTTGGTGCAGGCCCGGTTCGTGCCGAACATATTGCAGAACTGACCTATCTGAAACAG
GTGGTGAGCGAAACCTTACGCTTCTGGGGCCCGTGGATTAGCGCAGGCAATGCCAGT
GGCCCGGTGACCGTTGGTGGTCTGACCATTCCGGATGGTACCGCCATTATGTTCAGC
CCGTATCTGGTTCAGCATGATCCGCGTCACTTCCCGAATCCGGAAGCCTTCGATCCG
GATCGCTGGAGCCCGGATCGCACCGGTGAAATTGATAAACAGGCCAATCTGAGCTTC
GGCGTTGGTCGCCGCCGTTGTCTGGGTGATCACTTCGCCCTGCTGGAAATTACCCTG
GCCAGCGCCGCACTGCTGGCCAGATGGCGTCCGGAACCGGATCCGGCATATATTGTG
CGCGCAAGTAATCGTGACTTCGTTCTGAGTCCGAGTGCAATTCCGGTTACCCTGCAT
CGCCGCTAACTCGAG-3′

WP_084494404 Wild type translation 859 aa

MHIAMFMDYHPATLGGVQTAVAALRRGLERAGHRVTLFVAPLSGTPSPATEADDGVI
ELAPLAGAVVNGFPMVLPTARNAALIDAAFAARGPVDLVHTHTTYGVAISGMKAARR
HRIPLVHTAQSRDDAFIEHTAPAPYLSALALRILHGRFVNHPARMPRGAQSRAARHA
WHTIVGQGQAADRVIAPTRHFADLLLDHGLTRPIRVVSNGVDDALVDIRCAVERDAD
GPLRLIWCGRLSAEKRLLESIDAVRRVPDCTLDVYGEGELFDRAAAVVAEHGLADRI
TLHGRVSQEQCLDAMAAADALLFPSWGFDTQGMVLLEAIATGLPVLYCDPDLAESIP
AGGGLRADDASPAALADVIARLATDRAELLAMRAEMAGYRDSVRQSGRIDAIVEVYR
QARADTEPAAQPPVPRRLSDVPTAPGALPLIGHSLRALRDAPGFVTSLADLGPIVRI
RFGRKTGYVLTTPELVREVGLGDAELNRDDLREAIADVAGGSVNVLRGAEHKLRRRM
IAPALRQSRLAEYTVTAAGLADTWSAALPAGGRVDLMDEAHGLILDTVSSTLFTAEF
SETARRRIRDNVPWLLSQVILRTALPPQIRRLRVIANRRWQRKAHHLRAAIGAAIAE
YRRRDEDFNDVVSALIRHTDPETGARLSDEHIIDEAILMLAGGVGSMASLTGWLWHE
VMRRPDIAERIRAELDEVVGAGPVRAEHIAELTYLKQVVSETLRFWGPWISAGNASG
PVTVGGLTIPDGTAIMFSPYLVQHDPRHFPNPEAFDPDRWSPDRTGEIDKQANLSFG
VGRRRCLGDHFALLEITLASAALLARWRPEPDPAYIVRASNRDFVLSPSAIPVTLHR
R*LE

WP_030573948 Wild type 1374 bp DNA

5′–ATGAGTACCCAGACCGGCCCGGCCCTGGATACCCCTAGCCGTGGTCATGCATTC
GTTGCCGGCCCGAAAGGCCTGCCGCTGCTGGGTAATCTGCCGCAGTTCGGTAAAAAT
CCGCTGGCATTCTTCGAACTGCTGCGTGGTCGTGGTGATATGGTGCGCTGGCGCTTC
GGTCGCAATCGCTGCGTGTTCATTAGCGATCCGGATTGCATTGGCGAACTGCTGACC
GAAACCGAACGCACCTTCGATCAGCCGGCCCTGGGTGTTGCCTTCCGTGCAGTTCTG
GGTAATGGTCTGATTGTGGCACGCGGTCGCGATTGGCGTCGTAAACGTAGTCTGGTG
CAGCCGAGCGTTCGTCCGAAACAGGTTAAAAGTTATGCCGCAACAATGGCTAGCAGT
GCCGTTGAACTGGCCGATGCCTGGAGTGATGGTGAACGTGTTGATATTAAACGTGAA
ATGGCCGCCCTGACCCAGAAAATTGCCGTTCGTACCATCTTCGGCGTGGATACCCCG
GCCGATAGTGAAGCAATGGGTCGTGCAATGGATGTTGCCCAGCTGGAAATTGGTAAA
GAGTTCGCCGGCATTGGCGCCCTGCTGCCGGATTGGGTTCCGACCCCGGGTCGTGCA
CGTATTCGTAAAGCCGCAGCAGTGATTGATAGCGAAGTTGGTCGTGTGGTTGCACGT
CATCGTGATGGCGAAACCGAACGTCCGGATCTGCTGAGCCGCCTGCTGACCGCCGTG
GATGAAAGCGGCGAACGCCTGAGCGATGAAGAAATTCGTGATGAAACCGTGACCCTG
TATATTGGCGGTCATGAAACCACCAGCAGTACCCTGGTGTGGGCATGGTATCTGCTG
AGTCGTAATCCGCGTGTTCGTGCCGCCCTGGCCGAAGAACTGGATCGTGTGCTGGGT
GATCGTGAACCGGGCTTCGAAGATTATGCCCAGCTGACCTATGCACAGGCAGTTGTT
AAAGAAACCTTACGTCTGTATCCGACCATCTGGCTGATTACCGGCGTGGCCAAAGAA
GGCGCCCGTCTGGGTGGTCTGCCGATTGAAGAAGGTACCCGTGTGTGGAGCAGCCAG
TGGAGTACCCATCGTGATGCACGTTGGTTCCCGGAACCGGAAGAGTTCCGTCCGGAA
CGCTGGGATCCGGAAGATGGTGATGAAATTGCAGAATATGCCTGGTTCCCGTTCGGT
GGCGGCCCGAGAGTGTGTCTGGGCACCCGCTTCGCAATGGTTGAAAGTGTGCTGATT
CTGGCAGTTCTGGCACGCCGCTTCGAACTGGATGTGGAACCGGGCGTTATTGAACCG
GTTCCGAGCCTGACCCTGCAGCCGGATCGCGATGTGCTGGCCACCGTGCGTGCACGT
TAACTCGAG–3′

WP_030573948 Wild type translation 458 aa

MSTQTGPALDTPSRGHAFVAGPKGLPLLGNLPQFGKNPLAFFELLRGRGDMVRWRFG
RNRCVFISDPDCIGELLTETERTFDQPALGVAFRAVLGNGLIVARGRDWRRKRSLVQ
PSVRPKQVKSYAATMASSAVELADAWSDGERVDIKREMAALTQKIAVRTIFGVDTPA
DSEAMGRAMDVAQLEIGKEFAGIGALLPDWVPTPGRARIRKAAAVIDSEVGRVVARH
RDGETERPDLLSRLLTAVDESGERLSDEEIRDETVTLYIGGHETTSSTLVWAWYLLS
RNPRVRAALAEELDRVLGDREPGFEDYAQLTYAQAVVKETLRLYPTIWLITGVAKEG
ARLGGLPIEEGTRVWSSQWSTHRDARWFPEPEEFRPERWDPEDGDEIAEYAWFPFGG
GPRVCLGTRFAMVESVLILAVLARRFELDVEPGVIEPVPSLTLQPDRDVLATVRAR*
LE

WP_012378285 Wild type 1374 bp DNA

5′–ATGAGCACCCAGACCGGTCCGGCCCTGGGCACACCTCCTCGTGGTCATGCATTC
GTTCCGGGCCCGCGTGGCCTGCCTCTGGTTGGTAATCTGCCGCAGTTCGGCAAAAAT
CCGCTGGCATTCTTCGAACTGCTGCGCGGTCATGGTGATATGGTGCGCTGGCGCTTC
GGTCGTAAACGCTGCGTGTTCCTGGCAGATCCGGATCTGGTTGGCGAACTGCTGACC
GAAACCGAACGCACCTTCGATCAGCCGCGTCTGGGTATTGCATTCCGTACCGTTCTG
GGCAATGGCATGCTGGTGGCCCGTGGCCGCGATTGGCGTCGTAAACGTAGTCTGGTT
CAGCCGAGCGTTCGCCCGAAACAGGTGACCAGCTATGCCACCACAATGGCTGGCTGC
GCCGTTGAACTGGCCGATCGCCTGGCAGATGGCCAGCGTATTGATGTGAAACGTGAA
ATGAGCGCACTGACCCAGAAAATTGCCGTGCGCACCATCTTCGGTGTTGATACCCCG

GCCGATAGTGAAGCAATGGGTCGTGCAATGGATGTGGCACAGATGGAAATTGGTAAA
GAGTTCGCAGGCCTGGGCGCCCTGCTGCCTGATTGGGTGCCGACCCCGGGTCGCACC
CGTATTCGCAAAGCCGCAGGCGTGATTGATGCCGAAGTGCGCCGTGTTGTGGCCCGT
CATCGCGATGGTGATGAAGAACGTCCGGATCTGCTGAGCCGCCTGCTGACCGCAGTG
GATGAAAGCGGTACCCGTCTGAGCGATGAAGAAATTCGTGATGAAGCCGTTACCCTG
TATATTGGCGGTCATGAAACCACCAGCACCACCCTGGTGTGGGCATGGTATCTGCTG
GCACGCAATCCGCGTGTTCGCGAAGCCCTGGCAGAAGAACTGGATCGTGTTCTGGGT
GATCGTGATCCGGGCTTCGGCGATTATGCCCAGCTGACCTATGCCCAGGCCGTGGTG
AAAGAAACCTTACGTCTGTTCCCGGCCGTGTGGCTGATTACCGGCATTGCAAAAGAA
GGCGCCACCATTGGTGGCCTGCCGGTTGCAGAAGGTACCCGCGTGTGGAGCAGTCAG
TGGGCAACCCATCGCGATGCCCGTTGGTTCCCGGAACCGGAAGAGTTCCGTCCGGAA
CGCTGGGATGCAGAAAGTGGCGATGCAATTCCGGAATATGCATGGTTCCCGTTCGGT
GGTGGCCCGCGTGTGTGTATTGGTACCCGCTTCGCAATGGTTGAAAGTGTGCTGCTG
CTGGCAGTTCTGGCACGCCGCTTCACCCTGGATGTTGATCCGGGTGAAATTACCCCG
CTGACCGGTCTGACCCTGCAGCCGGATCGCGATGTGCTGGCAACCGTGCGCGCCCGT
TAACTCGAG—3′

WP_012378285 Wild type translation 458 aa

MSTQTGPALGTPPRGHAFVPGPRGLPLVGNLPQFGKNPLAFFELLRGHGDMVRWRFG
RKRCVFLADPDLVGELLTETERTFDQPRLGIAFRTVLGNGMLVARGRDWRRKRSLVQ
PSVRPKQVTSYATTMAGCAVELADRLADGQRIDVKREMSALTQKIAVRTIFGVDTPA
DSEAMGRAMDVAQMEIGKEFAGLGALLPDWVPTPGRTRIRKAAGVIDAEVRRVVARH
RDGDEERPDLLSRLLTAVDESGTRLSDEEIRDEAVTLYIGGHETTSTTLVWAWYLLA
RNPRVREALAEELDRVLGDRDPGFGDYAQLTYAQAVVKETLRLFPAVWLITGIAKEG
ATIGGLPVAEGTRVWSSQWATHRDARWFPEPEEFRPERWDAESGDAIPEYAWFPFGG
GPRVCIGTRFAMVESVLLLAVLARRFTLDVDPGEITPLTGLTLQPDRDVLATVRAR*
LE

WP_012377925 Wild type 1356 bp DNA

5′—ATGCGTACCGATCCGCCGGGTCCGCCGGTTAGTGCACTGCCTGGCCTGCTGCGT
AAACTGGCCGTTGATCGCCTGGAAATGATGAAAGATGCCGCAGCCCTGGGCGATGCC
GTGCGTGTTAGTATGGGCCCGAAAAAACTGTATATCTTCAATCGTCCGGATTATGCC
AAACATGTGCTGGCAGATAATAGTGATAATTATCATAAGGGTATCGGTCTGGTTCAG
AGTCGCCGTGTTCTGGGCGATGGTCTGCTGACCAGCGATGGTGAAGTGTGGCGTGCC
CAGCGTCAGACCGTGCAGCCGGCATTCAAACCGGGCCGCATTAATCGTCAGGCCAAT
GCAGTTGCCGAAGAAGGTGCAAAACTGGTTGCACTGCTGCGCGCCCATGAAGGCGGT
GGTCCGGTGGATGTTCTGCATGAAGTTACCGGCCTGACCCTGGGTGTTCTGGGTCGC
ACCCTGCTGGATAGTGATCTGAGTAGTCAGGATACCCTGGCACCGAGCTTCGAAGAA
GTTCAGGATCAGGCCATGCTGGAAATGGTGAGCCAGGGTATGGTTCCGGGTTGGCTG
CCGCTGCCGCCGCAAGCACGCTTCCGTCGTGCCCGTCGTGAACTGTATCGTGTGGCC
GATCTGCTGGTGGCCGATCGTAGTGCACGCATGGCCGATGGTGAACCGGGCGATGAT
GCACTGGCACGTATTATTGAAGCCGCAGGTCGTGGCAATGGCCCGCCGAGACGCGTT
CGTGGTAAACTGCGTGAAGAACTGGTTACCCTGCTGCTGGCAGGCCATGAAACCACC
GCAAGTACCCTGGGTTGGACCCTGCATCTGCTGGAACGTCATCCGGAAGTTCGCGCC
GCAGTTCGTGAAGAAGCCCGCAGCGTTCTGGGTGAACGCCTGCCGGATCTGGATGAT
CTGCATCGTCTGACCTGGACCACCAAAGTGGTGCAGGAAGCCATGCGTCTGTATCCG
CCGGTGTGGGTTCTGCCGCGCGTGGCACAGCGCGAAGATGAAGTGGGCGGTTATACC

GTTAGCGCACGCGCCGATGTGCTGATCTGTCCGTATATTATGCATCGTAATCCGCGT
CTGTGGGAAGATCCGGAACGCTTCGATCCGGAACGCTTCGACCCGCAGGCAGTGGCA
AGCCGTCCGCGTTATGCCTATATTCCGTTCGGTGCCGGCCCGCGCTTCTGCGTTGGC
TCAAATCTGGGTATGATGGAAGCCGTGTTCGTTACCGCACTGATTACCCGCGATCTG
GATCTGCGCACCGTGCCGGGCCATCGCGCTGTTGCTGAACCGATGCTGAGCCTGCGC
ATGCGTGGCGGCCTGCCTATGACCGTGAGCGTGGCAGGTTAACTCGAG—3′

WP_012377925 Wild type translation 452 aa

MRTDPPGPPVSALPGLLRKLAVDRLEMMKDAAALGDAVRVSMGPKKLYIFNRPDYAK
HVLADNSDNYHKGIGLVQSRRVLGDGLLTSDGEVWRAQRQTVQPAFKPGRINRQANA
VAEEGAKLVALLRAHEGGGPVDVLHEVTGLTLGVLGRTLLDSDLSSQDTLAPSFEEV
QDQAMLEMVSQGMVPGWLPLPPQARFRRARRELYRVADLLVADRSARMADGEPGDDA
LARIIEAAGRGNGPPRRVRGKLREELVTLLLAGHETTASTLGWTLHLLERHPEVRAA
VREEARSVLGERLPDLDDLHRLTWTTKVVQEAMRLYPPVWVLPRVAQREDEVGGYTV
SARADVLICPYIMHRNPRLWEDPERFDPERFDPQAVASRPRYAYIPFGAGPRFCVGS
NLGMMEAVFVTALITRDLDLRTVPGHRAVAEPMLSLRMRGGLPMTVSVAG*LE

FG081 Wild type 1626 bp DNA

5′—ATGGCCGTGATGAGCAAACTGCTGAACTTCCCGAGTCTGACCCTGGCCGCATGT
ATTGAAGGCTTCTTCGCCATTAAACTGTTCCCGAATTATTATAGTACCCAGAGCCAT
CTGGCAGCCGTGGTTACCATTCTGCTGATTAATTATGCATTCGGCGTTGTGTTCTGG
GCCGTGCTGTATCCGCGTCTGTTCAGCCCGCTGCGTCGCATTCCGGGTCCGAAAGCC
TATCTGAGTGCAGCACATCATAGTCTGGTGGTGAAAGGTCGTCCGAGTGGTGATCTG
TTCCTGGATCTGGCCAAAGAATATCCGGGCAAAGATGTTATTATGCTGAATAGCTTC
CGTAATCAGCTGTGCATTATGAATCCGCAGCTGCTGGCAGATCTGCTGGTGCATAAT
TGCTATGACTTCGCCAAACCGAAACGTATTAGTGGCTTCCTGCGTCATGTGCTGGGT
GATGGTCTGATTATTGTGGAAGGTGAACAGCATAAATTCCTGCGCAAAAATAGCACC
CCGGCCTTCCACTTCCGCCATATTAAAGAACTGTATCCGATGATGTGGACCAAAAGC
GAAACCTTAACCAAAGCAATTGCCCAGGATATTACCGCCAGCCGCAGCCCGGTGGTG
GAACTGAATGGTTGGGCCAGTAAAGTTACCCTGGATATTATTGGCATTGCAGGTCTG
GGTCGTAAATTCGATGCCGTGGAAAAAAAAATTGATCCGCTGGCCGATATCTATGAA
CAGCTGCTGGAACCGGATCGCGAAAAACTGATCTTCGCCATGCTGAGTCTGGCCATT
GGTCTGCCGATTATTCGCATGATTCCGTGGAAAATGAATGATCTGTTCAATTATCTG
ACCGGCAGCCTGAATGATCTGTGTTATCCGATGATTAAAGAAAAAAAGGCAGCCATT
ATCGAAAAAGGCGATGATCACTTCGATGTGCTGAGTCTGCTGATTAAAAGTAATAAC
TTCAGTGATGAGGCACTGAAAGATCAGCTGCTGACCTTCCTGGCAGCCGGTCATGAA
ACCACCGCCAGCGCACTGACCTGGGCATGCTATCTGCTGACCCAGTATCCGGATATT
CAGAGCAAACTGCGTGATGAAGTGCGTGATAGCCTGCCGGCAGATGTGGATTGTAAT
ACCCCGGATCTGGCCAGCATTCTGGAACAGATGCCGTATCTGAATGGCGTGATGCAT
GAAACCTTACGCCTGTATCCGACCGTTCCGCTGACCATGCGTAGCGCCCTGCGTGAT
ACCCGCATTGGCGATCAGTATATTCCGGAAGGCACCGATGTTATTGTTAGTATCTGG
TATATTAACCGTGCACCGGAAATCTGGGGCCCGGATGCCGCCGAGTTCCGTCCTGAA
CGCTGGATGACCGAAGATGGCAAACCGAATCAGAATGGCGGCGCCAGTAGCAATTAT
AACTTCCTGACCTTCTTACATGGTCCGCGTAGCTGCATTGGTCAGGGCTTCGCCAAA
GCAGAAATGCGCTGCCTGCTGGCAAATATGGTTAAAAGCTTCGAATGGACCCTGGCA
ATGGATAATAAACTGGTGCTGCCGCGCGGTGTGATTACCATTAAACCGGAAAATGGT
ATGTATCTGAATATGAAAGCAATCTAACTCGAG—3′

FG081 Wild type translation 542 aa

MAVMSKLLNFPSLTLAACIEGFFAIKLFPNYYSTQSHLAAVVTILLINYAFGVVFWA
VLYPRLFSPLRRIPGPKAYLSAAHHSLVVKGRPSGDLFLDLAKEYPGKDVIMLNSFR
NQLCIMNPQLLADLLVHNCYDFAKPKRISGFLRHVLGDGLIIVEGEQHKFLRKNSTP
AFHFRHIKELYPMMWTKSETLTKAIAQDITASRSPVVELNGWASKVTLDIIGIAGLG
RKFDAVEKKIDPLADIYEQLLEPDREKLIFAMLSLAIGLPIIRMIPWKMNDLFNYLT
GSLNDLCYPMIKEKKAAIIEKGDDHFDVLSLLIKSNNFSDEALKDQLLTFLAAGHET
TASALTWACYLLTQYPDIQSKLRDEVRDSLPADVDCNTPDLASILEQMPYLNGVMHE
TLRLYPTVPLTMRSALRDTRIGDQYIPEGTDVIVSIWYINRAPEIWGPDAAEFRPER
WMTEDGKPNQNGGASSNYNFLTFLHGPRSCIGQGFAKAEMRCLLANMVKSFEWTLAM
DNKLVLPRGVITIKPENGMYLNMKAI

FG022 Wild type 1632 bp DNA

5′—ATGGCACCGATTGCCGTGTTCCTGAGCTATTATCGCCCGTGGGCAGCCATTGCC
ATTGTGGTTGCCGTGTATTATCAGATTATTGTTGATCAGAACTACAGCGTGAAAAAT
GCCCCGTTCTATCTGGGTAAATATCTGCTGGGCTTCCTGAGTATCTGGGCCGTGTAT
ACCGTGTTCCTGTATCCGGCCCTGCTGAGCCCGCTGCGTCATCTGCCGCAGCCGAAA
CCGGGCAGCTTCATTAATGGTCATTGGGCAGAAAGTGTGCGTGAACCGGCCGGCCTG
CCGTTCCGTCATTGGATGCGTACCATTGAAAATAATGGCCTGATTCGTTATAAGCAT
CTGTTCAATCAGGAACGTATTCTGGTTACCAGCCCGGAAGGCCTGAAAGAAGTTCTG
GGTCAGAATAGCTATGATTATGTGAAACCGCATCTGCTGCGCGCAATGGTTGGCAAA
ATTCTGGGTTATGGTCTGCTGCTGAGTGAAGGCAATGTTCATAAAATGCAGCGTAAA
AATCTGATGCCGGCATTCAGCTTCCGCCATATTAAAGAACTGTATCCGGTGTTCTGG
AGTAAAGCCCAGGAACTGGTGCATGGCATTGAAAAAGAAATGAGCGAAGCACCGGGT
AGCCAGATTGATATTGCAGATTGGGCCAGTCGCGCCAGTCTGGATATTATTGGCAGT
GCAGGTATGGGCCATGAGTTCAAAAGTCTGAGCGATCCGAGTATTGAAGATACCATG
AAAATGTATGGTAGTATGGTTAAACAGAGCGGCGGCGCAAAACTGCTGACCGTTCTG
CAGCTGGTGCTGCCGAGTATGATTACCGATTATCTGCCGTTCCAGCGCAATATGGGC
GTTCTGGCCGCCAGTAAAGCAGCACGCGATACCAGCCAGCGCCTGATTAATGCAAAA
AAAGTGCAGATGGCCGCCAAAGAAAAACTGAGCCCGGATATTATTAGCACCGCCCTG
GAAAGCGGCCACTTCACCGATGAAGGTCTGGTTGATACCATGATGACCTTCCTGGCC
GCAGGTCATGAAACCAGCGCCGCAGCACTGACCTGGACCATCTTCCTGCTGGCAAAA
AATCATGGCATTCAGGATCGTCTGCGTGAAGAAATTCGTCAGAATGTTGATGGTCTG
GCCGATGATGTTGATAGTAAAAAACTGGATGGTCTGAGTTATCTGCATGCCGTGTGC
CAGGAAAGTCTGCGCCTGTATGCACCGATTCCGTTCACCGTTCGCGATAGCCTGCGC
GATACCACCATTCTGGGTACCTTCGTTCCGAAAGGCACCATGATTATTCTGTGCCCG
TGGGCCATTAATCGCGCCCATGAAAGCTGGGGCGCCGATGCAGATGACTTCTATCCG
GAACGCTGGATGGTGCCGGGTCAGGCCAATAGCGGCGGTGCCAAAAATAATTATGCC
AATCTGACCTTCCTGCATGGTCCGCGTAGTTGTATTGGTCAGAAATTCAGTCTGGCC
GAACTGATGGCACTGACCTGCGCCCTGGTGGGTCGCTATCGCTTCGATATTGATAAA
GATTATGAGGTTAAGGACCTGACCGATGGTATTGTTGCAAAACCGCGTGAAGGCCTG
AAGGTTAGCGTTGAAGAAATTCAGGGCTGGTAACTCGAG—3′

FG022 Wild type translation 544 aa

MAPIAVFLSYYRPWAAIAIVVAVYYQIIVDQNYSVKNAPFYLGKYLLGFLSIWAVYT
VFLYPALLSPLRHLPQPKPGSFINGHWAESVREPAGLPFRHWMRTIENNGLIRYKHL
FNQERILVTSPEGLKEVLGQNSYDYVKPHLLRAMVGKILGYGLLLSEGNVHKMQRKN
LMPAFSFRHIKELYPVFWSKAQELVHGIEKEMSEAPGSQIDIADWASRASLDIIGSA
GMGHEFKSLSDPSIEDTMKMYGSMVKQSGGAKLLTVLQLVLPSMITDYLPFQRNMGV
LAASKAARDTSQRLINAKKVQMAAKEKLSPDIISTALESGHFTDEGLVDTMMTFLAA
GHETSAAALTWTIFLLAKNHGIQDRLREEIRQNVDGLADDVDSKKLDGLSYLHAVCQ
ESLRLYAPIPFTVRDSLRDTTILGTFVPKGTMIILCPWAINRAHESWGADADDFYPE
RWMVPGQANSGGAKNNYANLTFLHGPRSCIGQKFSLAELMALTCALVGRYRFDIDKD
YEVKDLTDGIVAKPREGLKVSVEEIQGW

FG048 Wild type 1176 bp DNA

5′—ATGCAGCGCCGCAGCCTGACCCCGGCATTCGCATTCCGTCATATTAAAAATCTG
TATCCGGTGTTCTGGCGTAAAGCACGCGAAGTGACCCGCACCATGATGGCAGAGTTC
GGTCAGCAGGAAGAAACACAAGTTGAAATTAGTGGCTGGGCCAGCCGTGCAACCCTG
GATATTATTGGCCTGGCAGGCATGGGTCGTGACTTCGGCGCCATTCAGAATCCGAAT
AATACCCTGGCACAGACCTATAGCAAAATCTTCAAACCGAGCCGTCAGGCCCAGATT
CTGGCCTTCGTTGGTATGATTATTCCGATGGAGTTCATTACCAAACTGCCGTTCCGC
CGCAATGAAGATATTGCCAAAGCCGCAAGTGATATTCGCGCAATCTGTCGCGATCTG
ATTCAGGAAAAAAAAGCAAAAATGGCCAATAAAGAGCAGGCCGATGTTGATATTCTG
AGCGTTGCCCTGGAAAGTGGTGGCTTCACCGATGAAAATCTGGTTGATCAGCTGATG
ACCTTCCTGGCAGCCGGCCATGAAACCACCGCAAGTGCAATGACCTGGGCCATCTAT
ATGATGGCACGTCATAGCGATATTCAGACCCGTCTGCGTGAAGAAATTCGTGAACAT
CTGCCGAGTGTGGATAGTGATGTGGATATTACCAGCCTGGATATTGATCGTATGCCG
TATCTGAATGCCGTGTGCAGCGAAGTGCTGCGCTATTATGCCCCGGTGCCGCTGACC
ATGCGTGATGCCGCATATGATACCACCATTCTGGGCCAGACCATTCGCAAAGGTACC
CGCATTGTGATTGTTCCGTGGGCAACCCACTTCGATCATGATCTGTGGGGTCCGGAT
GCCGATCAGTTCAATCCGGATCGCTGGCTGAGTGCAGGCGGCGAAAATAAAATTGGC
GCCGATCGCAAAGCCGCAAGCGGCGGTGCCAATAGCAATTATGCCTTCCTGACCTTC
CTGCATGGTCCGCGTAGTTGTATTGGTAGCAGCTTCGCCAAAGCAGAGTTCGCATGC
CTGCTGGCAGCATGGATTGGTCGCTTCGAGTTCAGTCTGGCAAATCCGGAAGAAATG
GATGAAAAAAATGTTGAAATCCGCGGCGGTGTGACCGCACGTCCGGCTAAAGGTATG
CATGTGAAAGTGAAAGTGATTGGTGGTTATTAACTCGAG—3′

FG048 Wild type translation 392 aa

MQRRSLTPAFAFRHIKNLYPVFWRKAREVTRTMMAEFGQQEETQVEISGWASRATLD
IIGLAGMGRDFGAIQNPNNTLAQTYSKIFKPSRQAQILAFVGMIIPMEFITKLPFRR
NEDIAKAASDIRAICRDLIQEKKAKMANKEQADVDILSVALESGGFTDENLVDQLMT
FLAAGHETTASAMTWAIYMMARHSDIQTRLREEIREHLPSVDSDVDITSLDIDRMPY
LNAVCSEVLRYYAPVPLTMRDAAYDTTILGQTIRKGTRIVIVPWATHFDHDLWGPDA
DQFNPDRWLSAGGENKIGADRKAASGGANSNYAFLTFLHGPRSCIGSSFAKAEFACL
LAAWIGRFEFSLANPEEMDEKNVEIRGGVTARPAKGMHVKVKVIGGY*LE

FG094 Wild type 1680 bp DNA

5′–ATGAGTGTTATCGGCCTGTGGCTGGTTACCGTGACCGCAACCCTGAGCCTGTTC
GTGTGGCAGCTGATCTTCCTGCTGAGTATTCCGAAAAGTATTGTTGTGTGTCTGATT
GCCGAAAGTCTGTTCTTCGTGGCATGGTTCTTCTATTGGACCGTGATCTATCCGCGT
TATCTGACCCCGTTCCGTCATCTGCCGACCCCGGCCAGCCGCTCAATTCTGACCGGT
AATCAGAATGGTCTGTTCACCGAAAATAGCTGGGATGTGGCCCGTCGCGTGAGCCAG
ACCGTGCCTAATAGCGGTCTGATTCGTTATTATGTGGCCCTGAGTAATGAACGCATT
CTGGTTACCAATACCCGCGCCCTGAGTGATGTTCTGACCAATCATAGTCATGACTTC
GGCAAAAGTAATCTGGCAAAATTCGCACTGAAACGTCTGACCGGCAATGGTCTGGGC
TTCCTGGAAGGTAATGAACATAAAGTGCATCGCAAAAATCTGATGCCGGCATTCACC
CGCAAACATGTGAAAGAACTGACCCCGATCTTCTGGGATAAAGCAATGGAAATGGTT
AAAGGCATGGAAGCCGAAGTTCGTTGCGGTAAAGATACCAGCACCCAGGGCACCGGT
ATTGTTGAAATTCATGATTGGGCAACCCGCGCAACCCTGGATATTATTGGTACCGCA
GGCTTCGGTTATGACTTCGGCACCCTGCATAATCCGAGCAATGAAATTGGTCAGCAG
TATAAAAAAAATGTTCCTGGAACCGAGTACCGCATTCAATTGGCTGGAACTGCTGGGT
AATTATATTGACTTCCGCTTCCTGATGACCCTGCCGGTGAAAAAAAATCGCGATCTG
ACCGCCGGCAGTAACTTCATGCGCGAAATTGCAAAAAAAGTGATTCGCGAACGCCGT
CATGAACTGTTCCAGCGTATGACCAGTCAGGCCGGTAATATGAAAAATACCAAAAAA
GATATCATCACCACCGCCCTGGCAAGTGATTGCTTCACCGATGATCAGCTGGTGGAT
CATGTTATGGCATTCCTGGTGGCAGGTCATGAAAGTACCGCCACCGCCTTCGAATGG
GCCATGTATGAACTGGGTCATCGCCCGGAAATGCAGAAACGTGTTCGTGATGAAGTT
CGCACCTATCTGCCGAGTCCGAGCGCCGGCGGCGTGAAAAATATTACCTTCGAAAGT
GTTCCGTATCTGCAGGCCATCTGTAATGAAGTGCTGCGTCTGTATCCGTTCCTGCCG
TTCGCAACCCGCGTGGCCGAAAAAGATACCTGGGTGGCCGATCAGTTCGTGCCGAAA
GGCACCATTGTGGCCTATGCAGCACATATTAGCAATCGCGATAGCGAACTGTGGAGT
GGTCCGGCCCTGGATGCATTCGATCCGGAACGTTGGATGGAACCGGGCAAAGAAAGC
AGCGGTGGTGCCAATAGCAATTATGCAATGCTGACCTTCAGTGCCGGTCCGAAAAGT
TGTATTGGTGAAGCATGGACCCGTGCCGAACTGCCGTGCCTGGTGGGCGCTATGGTG
GGCAGCTTCGAAATTGAACTGGTTGAAGGCAAACAGGCAGATGGCACCGTGTATCCG
ACCGTGGACTTCAAAATGGGTAAAGTTCTGAAAAGTCGTGATGGCGTGTTCGTGCGT
CTGCGCCGTCTGGAAGATTGGTAACTCGAG–3′

FG094 Wild type translation 560 aa

MSVIGLWLVTVTATLSLFVWQLIFLLSIPKSIVVCLIAESLFFVAWFFYWTVIYPRY
LTPFRHLPTPASRSILTGNQNGLFTENSWDVARRVSQTVPNSGLIRYYVALSNERIL
VTNTRALSDVLTNHSHDFGKSNLAKFALKRLTGNGLGFLEGNEHKVHRKNLMPAFTR
KHVKELTPIFWDKAMEMVKGMEAEVRCGKDTSTQGTGIVEIHDWATRATLDIIGTAG
FGYDFGTLHNPSNEIGQQYKKMFLEPSTAFNWLELLGNYIDFRFLMTLPVKKNRDLT
AGSNFMREIAKKVIRERRHELFQRMTSQAGNMKNTKKDIITTALASDCFTDDQLVDH
VMAFLVAGHESTATAFEWAMYELGHRPEMQKRVRDEVRTYLPSPSAGGVKNITFESV
PYLQAICNEVLRLYPFLPFATRVAEKDTWVADQFVPKGTIVAYAAHISNRDSELWSG
PALDAFDPERWMEPGKESSGGANSNYAMLTFSAGPKSCIGEAWTRAELPCLVGAMVG
SFEIELVEGKQADGTVYPTVDFKMGKVLKSRDGVFVRLRRLEDW*LE

FG101 Wild type 1593bp DNA

5′–ATGCTGGTGCCGTTCCTGAGCGCAGTTGCCACCCTGATGCCGAGCCTGCTGATG
GGCATCTTCATTCTGTGCCTGCTGGGCTTCAGCGCAATTACCTTCAATATCTTCGTG
TATCCGTTCTATCTGAGCCCGCTGCGCCATCTGCCGGGTCCTACCGATAATGCCTTC
TTCTTCGGTCAGGCAGCCAAATTCCTGCAGGTTCCGTGGTTCCCGGAACTGTTCTGC
CAGTGGAGCCGTGATCATCCGGGTGCACCGTTCATTCGTTATCTGAACTTCGCCAAT
AGCGAAACCTTATTCGTGAATACCATTGGCGCATATAAACAGGTTCTGCAGACCAAA
AGCGCCTTCTTCGTTAAACCGGCCTTCGCCCGCCAGTTCGCACATGAAATTATTGGT
GATGGTCTGCCGTTCGTGGAAGGTAATCTGCATAAACTGCGTCGTGCAGCCATTAGT
CAGCCGTTCAGTGCCCCGCGTCTGCGCGAGTTCTATCCGACCGTTCAGGGCAAAGCA
GAACAGCTGGTGCGCGTTCTGAGCCAGCGCCATGATAAAAATGGCAATGTGGAAATT
GAAACCAATGTGTGGAAAACCGTTCTGGATGTTATTGGTCTGGAAACCTTCGGCCTG
GATCTGAATCATCTGGAAAGTAATGAAAGCCCGCTGTTCGAAACCTTCACCACCATG
ATGCAGCCGAGTGCACTGGGTCATATTGTGAATTATCTGAATAGCCTGGTTCCGATT
CGTCAGTTCATTCCGATTGCAGAATGTGTGGAGTTCAGCAAAAGTTGCACAATGGTT
CGCGAGTTCATTCTGTGTCTGATTAATATTCGTCGTAATCTGGGTGAAAAAGGCTTC
ATTAATAATCAGGATGCACTGCAGTGCCTGCTGGAACATAATGATGCCGATTGGAAT
GATGAAAGCCTGGTGGAATATGTGCTGAATCTGCTGGTGCTGGGTCATGATACCACC
GCCTGCAGCATTACCTGGGCAATTCATGAACTGAGTCGTCGCCCGGATTGTCAGCAG
CGCCTGCGTGATGAAATTAAAATTCTGGATGATACCTGTCCGATGATTAGCTTCAAT
GATATTGATAAACTGCCGTATCTGCATAACTTCGTTCGTGAAGTTCTGCGTCTGTAT
TGCGCAGTGGCAATGGCCCCGCGTCAGGCAACCGAAGATGTTGAAATTGAAGGTGTT
ATTATCCCGAAAGGTACCGTTGTTCAGCTGAGTCCGGCAGTGATGAATACCCATCCG
AGTGTGTGGGGTAGCGATGCACAGGTGTTCAATCCGGATCGTTGGAATGGTCTGAAA
GGTGATGCAGCAAGCGCCTATGCATTCGAAACCTTCCATAATGGTCCGCGTATGTGT
ATTGGTAAACAGCTGAGCGTTATGGAAATGAAAGTGATGCTGGTTGAAATTGTGCAT
AAATTCGAAATTCACAAGCCGTTCGGTGATGAGTTCAAAGAAGTGGCAGTGGCAGGC
CCGACCTTCACCCTGCGTCCGAAAGAAAAACTGGTTGTGCGCCTGGTGGAACTGCGC
CACGGTATTCTGGATACCCTGCGTGATAGCGATTGTGTGTGGGGTGTGTAACTCGAG
–3′

FG101 Wild type translation 531 aa

MLVPFLSAVATLMPSLLMGIFILCLLGFSAITFNIFVYPFYLSPLRHLPGPTDNAFF
FGQAAKFLQVPWFPELFCQWSRDHPGAPFIRYLNFANSETLFVNTIGAYKQVLQTKS
AFFVKPAFARQFAHEIIGDGLPFVEGNLHKLRRAAISQPFSAPRLREFYPTVQGKAE
QLVRVLSQRHDKNGNVEIETNVWKTVLDVIGLETFGLDLNHLESNESPLFETFTTMM
QPSALGHIVNYLNSLVPIRQFIPIAECVEFSKSCTMVREFILCLINIRRNLGEKGFI
NNQDALQCLLEHNDADWNDESLVEYVLNLLVLGHDTTACSITWAIHELSRRPDCQQR
LRDEIKILDDTCPMISFNDIDKLPYLHNFVREVLRLYCAVAMAPRQATEDVEIEGVI
IPKGTVVQLSPAVMNTHPSVWGSDAQVFNPDRWNGLKGDAASAYAFETFHNGPRMCI
GKQLSVMEMKVMLVEIVHKFEIHKPFGDEFKEVAVAGPTFTLRPKEKLVVRLVELRH
GILDTLRDSDCVWGV*LE

FG110 Wild type 1728 bp DNA

5′–ATGACCCATGAAAGTAAACGCCGCCCGGAAAGTGATGATCCGTGCATGCGCCAT
TGGGCCCGCCCGGTTAGTATTAGTAATGGCGGTGTTAGTCAGGAACAGGAAACACAT
GTGTATGAAAAATACTTCATCCCGTTCTTCCTGCAGGCATTCGTTGTGCAGTATATG
GGCGATGTTCATATGAAAGCAATGGAAGCCAGCTATGCCACCATTCTGTGGGGTGTT
GCACTGGCAGTGGTTGGCTTCCTGAGCCATCGTATTATTGCAATTGCCCTGCGCCCG
CATAGTAGTCGCTTCAATGGCCTGCCGCGCCCGAAAGGCCAGCAGCCGTTCGCTGGC
CATGCACTGCGTATTCTGCGCGGCGGTGGCCCGAATGATGTGTATCTGCAGTGGACC
CGTCAGTGGCCGGATGCCCCGTTCATTCGTTGTCTGAGCTGGCTGAATGAAGAAATT
CTGCTGGTTAATAGTCTGGAAGCATGCCGTGAAGTGCTGCAGACCAATGCATACTTC
TTCGCAAAACCGGGCTTCTTCCATACCCTGGTGGGCGAGTTCCTGGGCCTGGGCCTG
CTGTTCAGCGTGGGTGAACAGCATAAACGCCTGCGTCGTATTATTGCCGGTCCGCTG
AGCCGCCCGAGTATTCGTAAACTGTTCCCGACCTTCGTGACCTATAGTCAGAAACTG
AATCGTGAAATTGGTGAAGCCCTGGAACGTAGTAAAAGTGGTATTGTGGAAATTGAA
GATCTGATTATTCGTGTTACCCTGGATATTATTGGTGTGAGTCTGCTGGGCCGTGAA
CTGCGCGACTTCCGCAGCGAAAGCAGTCCGCTGAGCTTCGAACAGTGCTATAATGCC
ATTCTGGCACAGCCGCTGGCCGGCCAGATTATTAGCTTCATTAATCCGTTCATTCCG
CTGCGTTGGCTGCCGGTGAGCGCCAATCTGAACTTCATTCGCGCCAAAAGTGCACTG
AAAACCATGATGGAAGGCCTGATTGAACAGCGCACCGCCGAAGTTGGTGCAGCCAAA
CTGATGAATGAAGATGATAAACTGAGTGATGATCTGCTGACCCGCATGATTGAAGCC
AGCGCAGAAGAAAGTCAGAAATTAAGTAAAGAAGAGCTGATTGATATCACCATGCAG
GTGATTGCCGCAGGCCATGAAACCACCGCAAGTGCCCTGGTGTGGACCGCCTATAGC
CTGGCCAAAGATCCGGCAAGCCAGCAGAGCCTGCGTGCAGAAATTCATAGCCTGGGT
ACCGAAATGAGCGCAAAAGGTATTGATGAACTGCCGTTCCTGGATAATGTGATTCGT
GAAGCAATGCGCGTTCATAGCCCGACCCTGATTATTCCGTGGGAAGCACAGAAAGAT
ATGACCATTGCAGGCACCCATATTCCGAAAGGTACCACCGTGCAGATTGTGCCGGCA
ATGATTCAGCTGAATCCGGAAATCTGGGGTAGTGATGCCGATGTGTTCCGTCCGGGT
CGTTGGGAAGATATGGGTGGTAATGCCAGCAGCCCGTATGCAATGGAAACCTTCAGT
AATGGTCCGCGTATGTGTCCGGGCAAAGCACTGGCACTGCTGAATATGAAAGTGCTG
CTGGTGGGCCTGATTCGCGACTTCGAAATGGAAATTGTGGATGATGGTAAAGAAGTT
GAACTGCGTAATCCGAGTCTGACCCTGAAAGCAAAAAGTCTGATTCAGTTCAAAATG
CGTAAAGTTAGTTAACTCGAG–3′

FG110 Wild type translation 576 aa

MTHESKRRPESDDPCMRHWARPVSISNGGVSQEQETHVYEKYFIPFFLQAFVVQYMG
DVHMKAMEASYATILWGVALAVVGFLSHRIIAIALRPHSSRFNGLPRPKGQQPFAGH
ALRILRGGGPNDVYLQWTRQWPDAPFIRCLSWLNEEILLVNSLEACREVLQTNAYFF
AKPGFFHTLVGEFLGLGLLFSVGEQHKRLRRIIAGPLSRPSIRKLFPTFVTYSQKLN
REIGEALERSKSGIVEIEDLIIRVTLDIIGVSLLGRELRDFRSESSPLSFEQCYNAI
LAQPLAGQIISFINPFIPLRWLPVSANLNFIRAKSALKTMMEGLIEQRTAEVGAAKL
MNEDDKLSDDLLTRMIEASAEESQKLSKEELIDITMQVIAAGHETTASALVWTAYSL
AKDPASQQSLRAEIHSLGTEMSAKGIDELPFLDNVIREAMRVHSPTLIIPWEAQKDM
TIAGTHIPKGTTVQIVPAMIQLNPEIWGSDADVFRPGRWEDMGGNASSPYAMETFSN
GPRMCPGKALALLNMKVLLVGLIRDFEMEIVDDGKEVELRNPSLTLKAKSLIQFKMR
KVS*LE

8.2 CO-difference spectra

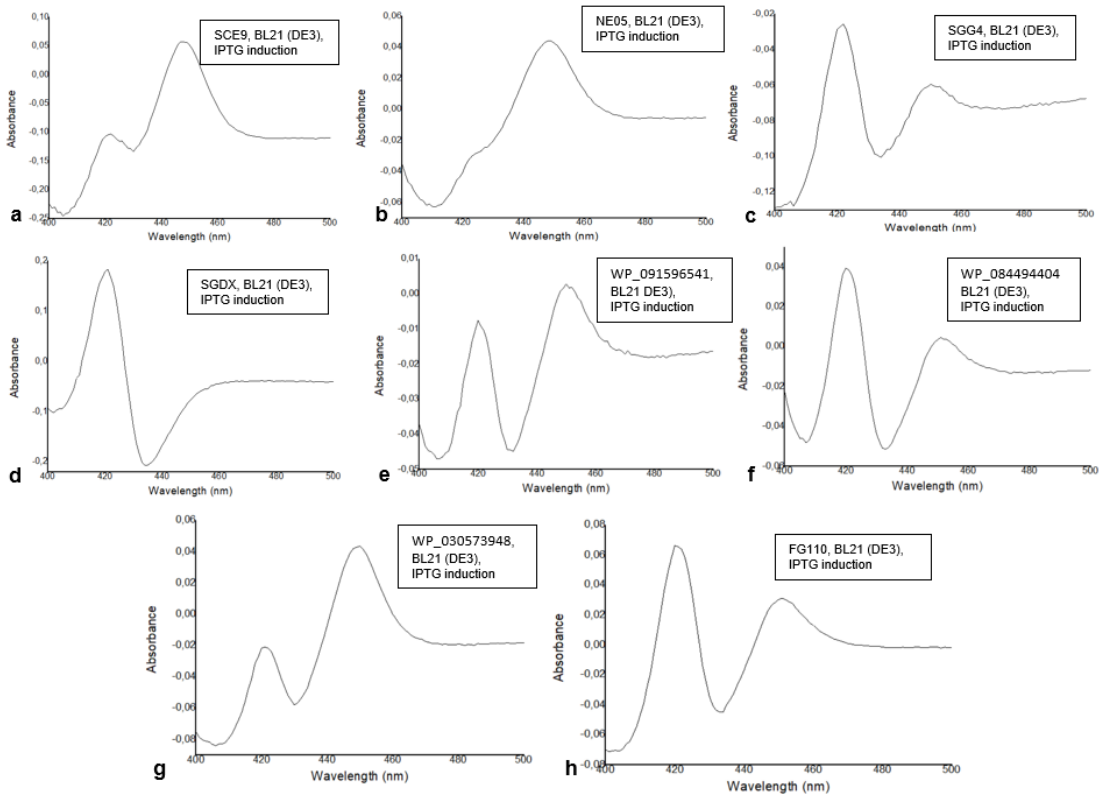


Figure 8.1: Spectra of the CO assay for the cytochrome CYPs SCE9, NE05, SGG4, SGDX, WP_091596541, WP_084494404, WP_030573948 and FG110 using IPTG, in *E. coli* BL21(DE3). On the x-axis the wavelength is displayed from 400 nm to 500 nm and on the y-axis the absorbance is displayed ranging from -0.25 to 0.10 (a), -0.06 to 0.06 (b), -0.12 to -0.02 (c), -0.2 to 0.2 (d), -0.05 to 0.01 (e),-0.06 to 0.04 (f), -0.06 to 0.06 (g) and -0.08 to 0.08 (h).

As displayed in figure 8.1, from the CO-difference spectra, is determined that the P450s SCE9 (a), SGG4 (c), WP_091596541 (e), WP_084494404 (f), WP_030573948 (g) and FG110 (h), expressed in *E. coli* BL21(DE3), using IPTG to boost the induction, are partially properly folded and partially misfolded, as indicated from the signal at 450 nm and at 420 nm. On the other hand, the CYP NE05 is folded correctly, as there is a Soret band at 450 nm confirming the successful result (b) and lastly, the P450 SGDX is misfolded, as resulted by the signal at 420 nm (d). The concentrations of the folded and denatured P450s have been calculated and displayed in tables 3.3 and 3.4.

8.3 HPLC chromatograms

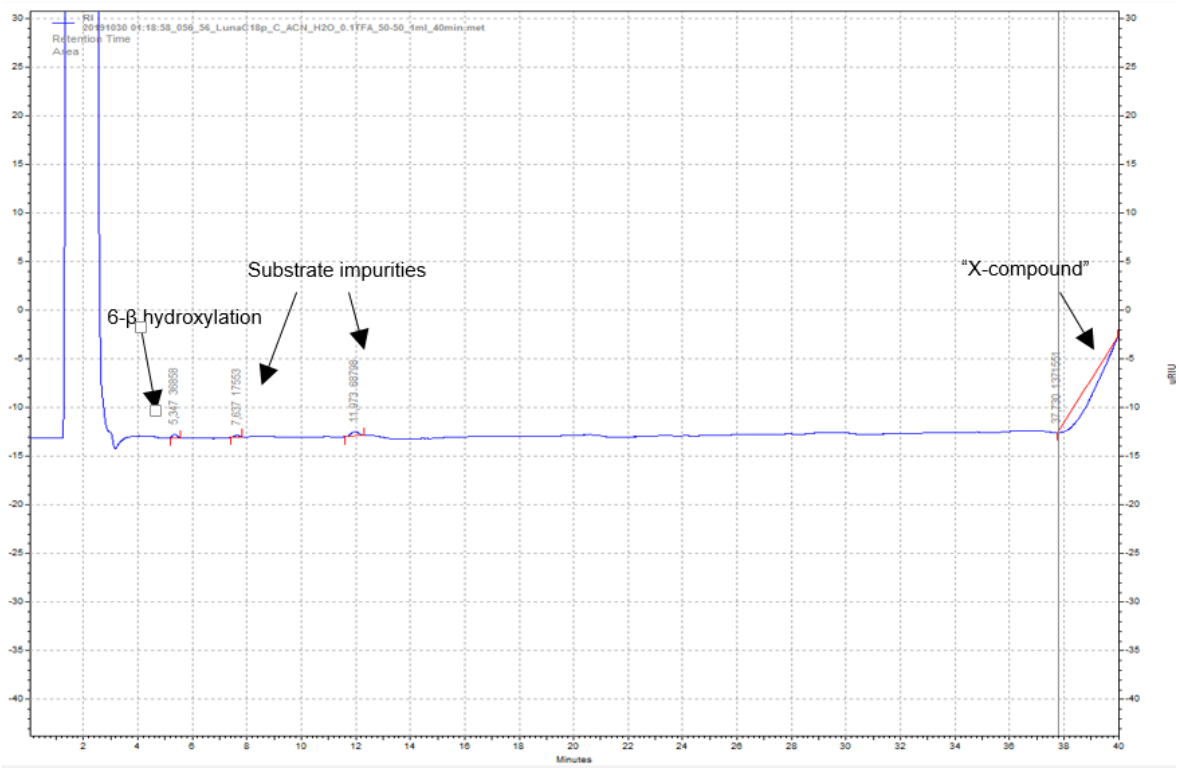


Figure 8.2: Chromatogram of OleP-camA/B in *E. coli* C41(DE3) after HPLC analysis. On the x-axis the time is displayed from 0 min to 40 min and on the y-axis the absorbance (mAU) is displayed ranging from -40 to 30.

As displayed in figure 8.2 at the HPLC chromatogram of OleP-camA/B after biocatalysis in BL21(DE3), the substrate is shown at 37.73 min and the product of 6-β hydroxylation at 5.34 min.

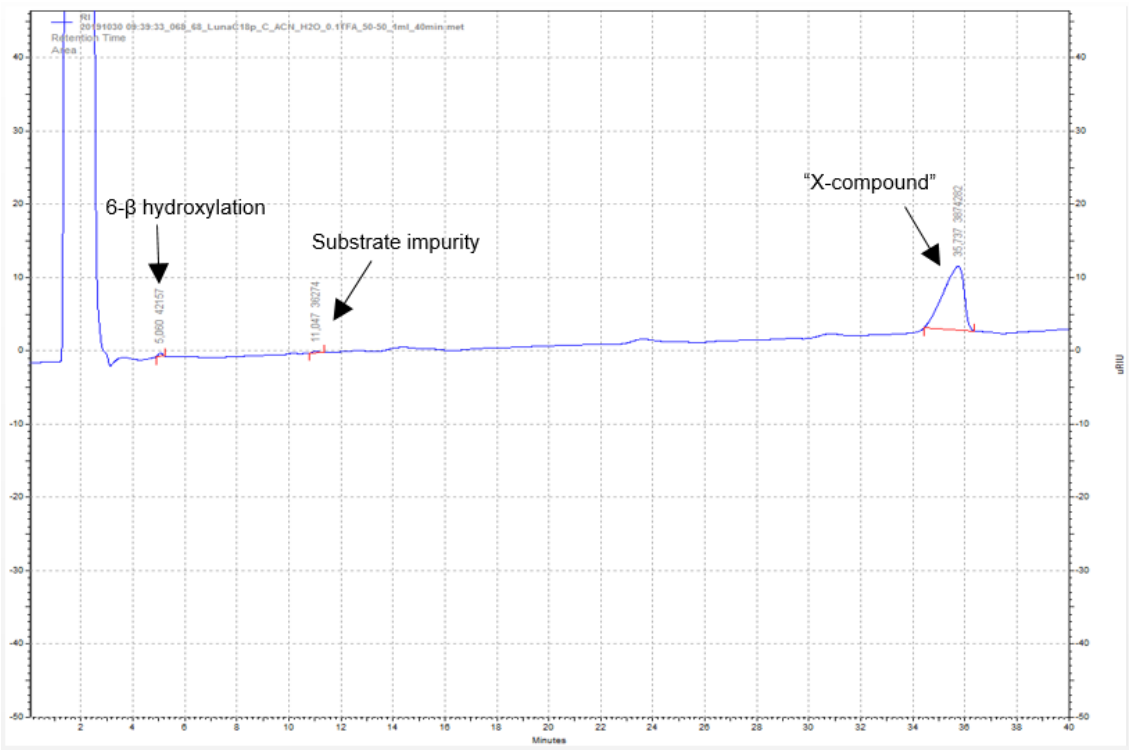


Figure 8.3: Chromatogram of OleP-camA/B in *E. coli* C43(DE3) after HPLC analysis. On the x-axis the time is displayed from 0 min to 40 min and on the y-axis the absorbance (mAU) is displayed ranging from -10 to 25.

As displayed in figure 8.3 at the HPLC chromatogram of OleP-camA/B in C43(DE3), the substrate is shown at 35.73 min and the product of 6-β hydroxylation at 5.06 min.

8.4 TLC plates

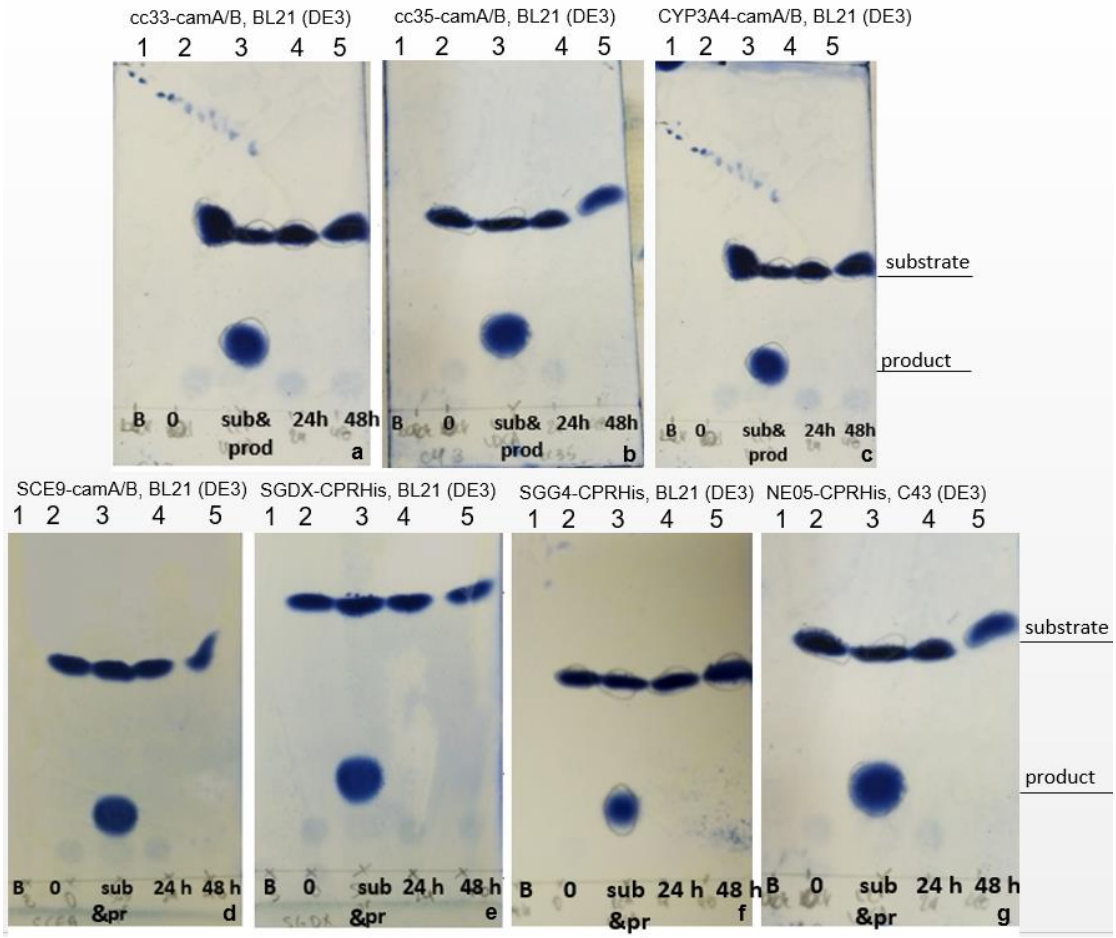


Figure 8.4: TLC plates for the biotransformation of the steroid of interest using (a) cc33-camA/B in BL21(DE3), (b) cc35-camA/B in BL21(DE3), (c) CYP3A4-camA/B in BL21(DE3), (d) SCE9-camA/B in BL21(DE3), (e) SGDX-CPR_{His} in BL21(DE3), (f) SGG4-CPR_{His} in BL21(DE3), and (g) NE05-CPR_{His} in C43(DE3). The samples where loaded from left to right with the order: background (B), time zero (0), compound/product (which were used as standards), sample after 24 hours of reaction (24 h) and sample after 48 h of reaction (48 h).

As indicated in figure 8.4, there is no product formation in any of the plates, as there is no signal indicating it at the corresponded height of the standard (lanes 3) after 24 h and 48 h of reaction, except of the substrate's (lanes 4, 5 (a), lanes 4, 5 (b), lane 4, 5 (c), lanes 4, 5 (d), lanes 4, 5 (e), lanes 4, 5 (f), lanes 4, 5 (g), figure 9.4.1). Additionally, in all plates there is no signal in the background sample, as there was no substrate added yet (lanes 1) and only substrates signal in lanes 2, representing 5 min of incubation.

References

1. Bornscheuer, U.T., et al., *Engineering the third wave of biocatalysis.* Nature, 2012. **485**(7397): p. 185-94.
2. *Morlock L., Investigation on the general activity and uncoupling of the catalytic cycle of different Cytochrome P450 monooxygenases, Ph.D thesis*, in *Institute of Biochemistry*. 2018, University of Greifswald: Greifswald. p. 161.
3. Turner, N.J., *Directed evolution drives the next generation of biocatalysts.* Nat Chem Biol, 2009. **5**(8): p. 567-73.
4. Poppe, L. and B.G. Vertessy, *The Fourth Wave of Biocatalysis Emerges-The 13 th International Symposium on Biocatalysis and Biotransformations.* Chembiochem, 2018. **19**(4): p. 284-287.
5. Wachtmeister, J. and D. Rother, *Recent advances in whole cell biocatalysis techniques bridging from investigative to industrial scale.* Curr Opin Biotechnol, 2016. **42**: p. 169-177.
6. Bornscheuer, U.T., *The fourth wave of biocatalysis is approaching.* Philos Trans A Math Phys Eng Sci, 2018. **376**(2110).
7. Schmid, A., et al., *Industrial biocatalysis today and tomorrow.* Nature, 2001. **409**(6817): p. 258-68.
8. Woodley, J.M., *New opportunities for biocatalysis: making pharmaceutical processes greener.* Trends Biotechnol, 2008. **26**(6): p. 321-7.
9. McDonald, A.G. and K.F. Tipton, *Fifty-five years of enzyme classification: advances and difficulties.* FEBS J, 2014. **281**(2): p. 583-92.
10. Guengerich, F.P., et al., *Measurement of cytochrome P450 and NADPH-cytochrome P450 reductase.* Nat Protoc, 2009. **4**(9): p. 1245-51.
11. Omura, T. and R. Sato, *The Carbon Monoxide-Binding Pigment of Liver Microsomes. Ii. Solubilization, Purification, and Properties.* J Biol Chem, 1964. **239**: p. 2379-85.
12. Bernhardt, R. and V.B. Urlacher, *Cytochromes P450 as promising catalysts for biotechnological application: chances and limitations.* Appl Microbiol Biotechnol, 2014. **98**(14): p. 6185-203.
13. *https://cyped.biocatnet.de/sequence-browser (as seen in 29.09.2020)*. 2020; Available from: https://cyped.biocatnet.de/sequence-browser.
14. Schewe, H., M.A. Mirata, and J. Schrader, *Bioprocess engineering for microbial synthesis and conversion of isoprenoids.* Adv Biochem Eng Biotechnol, 2015. **148**: p. 251-86.
15. Grobe, S., et al., *Highly selective bile acid hydroxylation by the multifunctional bacterial P450 monooxygenase CYP107D1 (OleP).* Biotechnol Lett, 2020. **42**(5): p. 819-824.
16. Axelrod, J., *The enzymatic demethylation of ephedrine.* J Pharmacol Exp Ther, 1955. **114**(4): p. 430-8.
17. Brodie, B.B., J.R. Gillette, and B.N. La Du, *Enzymatic metabolism of drugs and other foreign compounds.* Annu Rev Biochem, 1958. **27**(3): p. 427-54.
18. Ryan, K.J. and L.L. Engel, *Hydroxylation of steroids at carbon 21.* J Biol Chem, 1957. **225**(1): p. 103-14.

19. *Konia E., Investigation and expression optimization of cytochromes P450 with biotechnological interest, B.Sc thesis*, in *Department of Chemistry*. 2019, University of Crete: Heraklion, Crete. p. 63.
20. Poulos, T.L., et al., *The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450.* J Biol Chem, 1985. **260**(30): p. 16122-30.
21. Poulos, T.L., B.C. Finzel, and A.J. Howard, *High-resolution crystal structure of cytochrome P450cam.* J Mol Biol, 1987. **195**(3): p. 687-700.
22. Estabrook, R.W., et al., *A new spectral intermediate associated with cytochrome P-450 function in liver microsomes.* Biochem Biophys Res Commun, 1971. **42**(1): p. 132-9.
23. Yasui, H., S. Hayashi, and H. Sakurai, *Possible involvement of singlet oxygen species as multiple oxidants in p450 catalytic reactions.* Drug Metab Pharmacokinet, 2005. **20**(1): p. 1-13.
24. Okey, A.B., et al., *Induction of drug-metabolizing enzymes: mechanisms and consequences.* Clin Biochem, 1986. **19**(2): p. 132-41.
25. Coon, M.J., *Cytochrome P450: nature's most versatile biological catalyst.* Annu Rev Pharmacol Toxicol, 2005. **45**: p. 1-25.
26. Zhang, X. and S. Li, *Expansion of chemical space for natural products by uncommon P450 reactions.* Nat Prod Rep, 2017. **34**(9): p. 1061-1089.
27. Podust, L.M. and D.H. Sherman, *Diversity of P450 enzymes in the biosynthesis of natural products.* Nat Prod Rep, 2012. **29**(10): p. 1251-66.
28. Rudolf, J.D., et al., *Cytochromes P450 for natural product biosynthesis in Streptomyces: sequence, structure, and function.* Nat Prod Rep, 2017. **34**(9): p. 1141-1172.
29. Dietrich, M., et al., *Cloning, expression and characterisation of CYP102A7, a self-sufficient P450 monooxygenase from Bacillus licheniformis.* Appl Microbiol Biotechnol, 2008. **79**(6): p. 931-40.
30. Parisi, G., et al., *Substrate-induced conformational change in cytochrome P450 OleP.* FASEB J, 2019. **33**(2): p. 1787-1800.
31. Li, Z., et al., *Engineering cytochrome P450 enzyme systems for biomedical and biotechnological applications.* J Biol Chem, 2020. **295**(3): p. 833-849.
32. Zhang, W., et al., *New reactions and products resulting from alternative interactions between the P450 enzyme and redox partners.* J Am Chem Soc, 2014. **136**(9): p. 3640-6.
33. *Büchsenschütz H. C., Investigation of two putative P450 monooxygenases from marine habitats, M.Sc thesis.* 2017: p. 80.
34. Agematu, H., et al., *Hydroxylation of testosterone by bacterial cytochromes P450 using the Escherichia coli expression system.* Biosci Biotechnol Biochem, 2006. **70**(1): p. 307-11.
35. McIntosh, J.A., et al., *Structural Adaptability Facilitates Histidine Heme Ligation in a Cytochrome P450.* J Am Chem Soc, 2015. **137**(43): p. 13861-5.
36. Girvan, H.M., et al., *Structural and spectroscopic characterization of P450 BM3 mutants with unprecedented P450 heme iron ligand sets. New heme ligation states influence conformational equilibria in P450 BM3.* J Biol Chem, 2007. **282**(1): p. 564-72.

37. Furukawa, Y., K. Ishimori, and I. Morishima, *Electron transfer reactions in Zn-substituted cytochrome P450cam.* Biochemistry, 2000. **39**(36): p. 10996-1004.
38. Hrycay, E.G., *Monooxygenase, peroxidase and peroxygenase properties and mechanisms of cytochrome p450*. 2016, [Place of publication not identified]: SPRINGER INTERNATIONAL PU.
39. Dawson, J.H., et al., *Letter: Oxidized cytochrome P-450. Magnetic circular dichroism evidence for thiolate ligation in the substrate-bound form. Implications for the catalytic mechanism.* J Am Chem Soc, 1976. **98**(12): p. 3707-8.
40. Auclair, K., P. Moenne-Loccoz, and P.R. Ortiz de Montellano, *Roles of the proximal heme thiolate ligand in cytochrome p450(cam).* J Am Chem Soc, 2001. **123**(21): p. 4877-85.
41. Tang, S.C., et al., *Axial ligation modes in iron(III) Porphyrins. Models for the oxidized reaction states of cytochrome P-450 enzymes and the molecular structure of iron(III) protoporphyrin IX dimethyl ester p-nitrobenzenethiolate.* J Am Chem Soc, 1976. **98**(9): p. 2414-34.
42. Munro, A.W., et al., *What makes a P450 tick?* Trends Biochem Sci, 2013. **38**(3): p. 140-50.
43. *https://www.rcsb.org/3d-view/2CPP/1 (as seen in 29.09.2020).* 2020.
44. Whitehouse, C.J., S.G. Bell, and L.L. Wong, *P450(BM3) (CYP102A1): connecting the dots.* Chem Soc Rev, 2012. **41**(3): p. 1218-60.
45. Bernhardt, R., *Cytochromes P450 as versatile biocatalysts.* J Biotechnol, 2006. **124**(1): p. 128-45.
46. Guengerich, F.P. and W.W. Johnson, *Kinetics of ferric cytochrome P450 reduction by NADPH-cytochrome P450 reductase: rapid reduction in the absence of substrate and variations among cytochrome P450 systems.* Biochemistry, 1997. **36**(48): p. 14741-50.
47. Spolitak, T., J.H. Dawson, and D.P. Ballou, *Reaction of ferric cytochrome P450cam with peracids: kinetic characterization of intermediates on the reaction pathway.* J Biol Chem, 2005. **280**(21): p. 20300-9.
48. Egawa, T., H. Shimada, and Y. Ishimura, *Evidence for compound I formation in the reaction of cytochrome P450cam with m-chloroperbenzoic acid.* Biochem Biophys Res Commun, 1994. **201**(3): p. 1464-9.
49. Omura, T., *Forty years of cytochrome P450.* Biochem Biophys Res Commun, 1999. **266**(3): p. 690-8.
50. Peterson, J.A. and S.S. Boddupalli, *P450BM-3: reduction by NADPH and sodium dithionite.* Arch Biochem Biophys, 1992. **294**(2): p. 654-61.
51. Munro, A.W., H.M. Girvan, and K.J. McLean, *Variations on a (t)heme--novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily.* Nat Prod Rep, 2007. **24**(3): p. 585-609.
52. Estabrook, R.W., *A passion for P450s (rememberances of the early history of research on cytochrome P450).* Drug Metab Dispos, 2003. **31**(12): p. 1461-73.
53. Arnold, F.H., *Innovation by Evolution: Bringing New Chemistry to Life (Nobel Lecture).* Angew Chem Int Ed Engl, 2019. **58**(41): p. 14420-14426.

54. Girvan, H.M. and A.W. Munro, *Applications of microbial cytochrome P450 enzymes in biotechnology and synthetic biology.* Curr Opin Chem Biol, 2016. **31**: p. 136-45.
55. Sakaki, T., *Practical application of cytochrome P450.* Biol Pharm Bull, 2012. **35**(6): p. 844-9.
56. Joseph, S., et al., *An amperometric biosensor with human CYP3A4 as a novel drug screening tool.* Biochem Pharmacol, 2003. **65**(11): p. 1817-26.
57. Hannemann, F., et al., *Cytochrome P450 systems--biological variations of electron transport chains.* Biochim Biophys Acta, 2007. **1770**(3): p. 330-44.
58. Sadeghi, S.J. and G. Gilardi, *Chimeric P450 enzymes: activity of artificial redox fusions driven by different reductases for biotechnological applications.* Biotechnol Appl Biochem, 2013. **60**(1): p. 102-10.
59. Puchkaev, A.V. and P.R. Ortiz de Montellano, *The Sulfolobus solfataricus electron donor partners of thermophilic CYP119: an unusual non-NAD(P)H-dependent cytochrome P450 system.* Arch Biochem Biophys, 2005. **434**(1): p. 169-77.
60. Rylott, E.L., et al., *An explosive-degrading cytochrome P450 activity and its targeted application for the phytoremediation of RDX.* Nat Biotechnol, 2006. **24**(2): p. 216-9.
61. Roberts, G.A., et al., *Identification of a new class of cytochrome P450 from a Rhodococcus sp.* J Bacteriol, 2002. **184**(14): p. 3898-908.
62. Sevrioukova, I.F. and T.L. Poulos, *Structural biology of redox partner interactions in P450cam monooxygenase: a fresh look at an old system.* Arch Biochem Biophys, 2011. **507**(1): p. 66-74.
63. Holden, M., et al., *Probing the interactions of putidaredoxin with redox partners in camphor P450 5-monooxygenase by mutagenesis of surface residues.* J Biol Chem, 1997. **272**(35): p. 21720-5.
64. Batabyal, D., L.S. Richards, and T.L. Poulos, *Effect of Redox Partner Binding on Cytochrome P450 Conformational Dynamics.* J Am Chem Soc, 2017. **139**(37): p. 13193-13199.
65. Katagiri, M., B.N. Ganguli, and I.C. Gunsalus, *A soluble cytochrome P-450 functional in methylene hydroxylation.* J Biol Chem, 1968. **243**(12): p. 3543-6.
66. Girhard, M., et al., *Cytochrome P450 reductase from Candida apicola: versatile redox partner for bacterial P450s.* Appl Microbiol Biotechnol, 2013. **97**(4): p. 1625-35.
67. Lottermoser, K., W.H. Schunck, and O. Asperger, *Cytochromes P450 of the sophorose lipid-producing yeast Candida apicola: heterogeneity and polymerase chain reaction-mediated cloning of two genes.* Yeast, 1996. **12**(6): p. 565-75.
68. Seifert, A., et al., *Rational design of a minimal and highly enriched CYP102A1 mutant library with improved regio-, stereo- and chemoselectivity.* Chembiochem, 2009. **10**(5): p. 853-61.
69. Narhi, L.O. and A.J. Fulco, *Characterization of a catalytically self-sufficient 119,000-dalton cytochrome P-450 monooxygenase induced by barbiturates in Bacillus megaterium.* J Biol Chem, 1986. **261**(16): p. 7160-9.

70. Seifert, A., et al., *An efficient route to selective bio-oxidation catalysts: an iterative approach comprising modeling, diversification, and screening, based on CYP102A1.* Chembiochem, 2011. **12**(9): p. 1346-51.
71. *https://www.rcsb.org/structure/2IJ2 (as seen in 29.09.2020).*
72. Miura, Y. and A.J. Fulco, *(Omega -2) hydroxylation of fatty acids by a soluble system from bacillus megaterium.* J Biol Chem, 1974. **249**(6): p. 1880-8.
73. Miura, Y. and A.J. Fulco, *Omega-1, Omega-2 and Omega-3 hydroxylation of long-chain fatty acids, amides and alcohols by a soluble enzyme system from Bacillus megaterium.* Biochim Biophys Acta, 1975. **388**(3): p. 305-17.
74. Capdevila, J.H., et al., *The highly stereoselective oxidation of polyunsaturated fatty acids by cytochrome P450BM-3.* J Biol Chem, 1996. **271**(37): p. 22663-71.
75. Li, H. and T.L. Poulos, *Fatty acid metabolism, conformational change, and electron transfer in cytochrome P-450(BM-3).* Biochim Biophys Acta, 1999. **1441**(2-3): p. 141-9.
76. Whitehouse, C.J., et al., *Structure, electronic properties and catalytic behaviour of an activity-enhancing CYP102A1 (P450(BM3)) variant.* Dalton Trans, 2011. **40**(40): p. 10383-96.
77. Clark, J.P., et al., *The role of Thr268 and Phe393 in cytochrome P450 BM3.* J Inorg Biochem, 2006. **100**(5-6): p. 1075-90.
78. Li, H. and T.L. Poulos, *The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid.* Nat Struct Biol, 1997. **4**(2): p. 140-6.
79. Kille, S., et al., *Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution.* Nat Chem, 2011. **3**(9): p. 738-43.
80. Graham-Lorence, S., et al., *An active site substitution, F87V, converts cytochrome P450 BM-3 into a regio- and stereoselective (14S,15R)-arachidonic acid epoxygenase.* J Biol Chem, 1997. **272**(2): p. 1127-35.
81. Narhi, L.O. and A.J. Fulco, *Identification and characterization of two functional domains in cytochrome P-450BM-3, a catalytically self-sufficient monooxygenase induced by barbiturates in Bacillus megaterium.* J Biol Chem, 1987. **262**(14): p. 6683-90.
82. Rodriguez, A.M., et al., *A cytochrome P450-like gene possibly involved in oleandomycin biosynthesis by Streptomyces antibioticus.* FEMS Microbiol Lett, 1995. **127**(1-2): p. 117-20.
83. Montemiglio, L.C., et al., *Functional analysis and crystallographic structure of clotrimazole bound OleP, a cytochrome P450 epoxidase from Streptomyces antibioticus involved in oleandomycin biosynthesis.* Biochim Biophys Acta, 2016. **1860**(3): p. 465-75.
84. *https://www.rcsb.org/3d-view/4XE3/1 (as seen in 29.09.2020).*
85. Niwa, T., et al., *Regioselective hydroxylation of steroid hormones by human cytochromes P450.* Drug Metab Rev, 2015. **47**(2): p. 89-110.
86. Monte, M.J., et al., *Bile acids: chemistry, physiology, and pathophysiology.* World J Gastroenterol, 2009. **15**(7): p. 804-16.

87. Staels, B. and V.A. Fonseca, *Bile acids and metabolic regulation: mechanisms and clinical responses to bile acid sequestration.* Diabetes Care, 2009. **32 Suppl 2**: p. S237-45.
88. Donova, M.V. and O.V. Egorova, *Microbial steroid transformations: current state and prospects.* Appl Microbiol Biotechnol, 2012. **94**(6): p. 1423-47.
89. Pavlovic, N., et al., *Bile Acids and Their Derivatives as Potential Modifiers of Drug Release and Pharmacokinetic Profiles.* Front Pharmacol, 2018. **9**: p. 1283.
90. Hofmann, A.F. and L.R. Hagey, *Bile acids: chemistry, pathochemistry, biology, pathobiology, and therapeutics.* Cell Mol Life Sci, 2008. **65**(16): p. 2461-83.
91. Lovering, F., J. Bikker, and C. Humblet, *Escape from flatland: increasing saturation as an approach to improving clinical success.* J Med Chem, 2009. **52**(21): p. 6752-6.
92. Brill, Z.G., et al., *Navigating the Chiral Pool in the Total Synthesis of Complex Terpene Natural Products.* Chem Rev, 2017. **117**(18): p. 11753-11795.
93. Crocker, M. and E. Santillan-Jimenez, *Chemical catalysts for biomass upgrading*. 2020.
94. Vottero, E., et al., *Role of residue 87 in substrate selectivity and regioselectivity of drug-metabolizing cytochrome P450 CYP102A1 M11.* J Biol Inorg Chem, 2011. **16**(6): p. 899-912.
95. van Beilen, J.B., et al., *Biocatalytic production of perillyl alcohol from limonene by using a novel Mycobacterium sp. cytochrome P450 alkane hydroxylase expressed in Pseudomonas putida.* Appl Environ Microbiol, 2005. **71**(4): p. 1737-44.
96. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.
97. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.
98. Evans, D.A., *History of the Harvard ChemDraw project.* Angew Chem Int Ed Engl, 2014. **53**(42): p. 11140-5.
99. Dumon-Seignovert, L., G. Cariot, and L. Vuillard, *The toxicity of recombinant proteins in Escherichia coli: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3).* Protein Expr Purif, 2004. **37**(1): p. 203-6.
100. Wagner, S., et al., *Consequences of membrane protein overexpression in Escherichia coli.* Mol Cell Proteomics, 2007. **6**(9): p. 1527-50.
101. Panek, H. and M.R. O'Brian, *A whole genome view of prokaryotic haem biosynthesis.* Microbiology, 2002. **148**(Pt 8): p. 2273-2282.
102. Ogun, A.S., N.V. Joy, and M. Valentine, *Biochemistry, Heme Synthesis*, in *StatPearls*. 2020: Treasure Island (FL).
103. Urlacher, V.B., S. Lutz-Wahl, and R.D. Schmid, *Microbial P450 enzymes in biotechnology.* Appl Microbiol Biotechnol, 2004. **64**(3): p. 317-25.
104. Correia, M.A., P.R. Sinclair, and F. De Matteis, *Cytochrome P450 regulation: the interplay between its heme and apoprotein moieties in synthesis, assembly, repair, and disposal.* Drug Metab Rev, 2011. **43**(1): p. 1-26.

105. Heinemann, I.U., M. Jahn, and D. Jahn, *The biochemistry of heme biosynthesis.* Arch Biochem Biophys, 2008. **474**(2): p. 238-51.
106. Brenna, E. *Synthetic methods for biologically active molecules : exploring the potential of bioreductions*. 2014; Available from: http://site.ebrary.com/id/10753405.
107. Gustafsson, J.A., L. Rondahl, and J. Bergman, *Iodosylbenzene derivatives as oxygen donors in cytochrome P-450 catalyzed steroid hydroxylations.* Biochemistry, 1979. **18**(5): p. 865-70.
108. Hrycay, E.G., et al., *Sodium periodate, sodium chloride, organic hydroperoxides, and H2O2 as hydroxylating agents in steroid hydroxylation reactions catalyzed by partially purified cytochrome P-450.* Biochem Biophys Res Commun, 1975. **66**(1): p. 209-16.
109. Wong, T.S., F.H. Arnold, and U. Schwaneberg, *Laboratory evolution of cytochrome p450 BM-3 monooxygenase for organic cosolvents.* Biotechnol Bioeng, 2004. **85**(3): p. 351-8.
110. Zangar, R.C., D.R. Davydov, and S. Verma, *Mechanisms that regulate production of reactive oxygen species by cytochrome P450.* Toxicol Appl Pharmacol, 2004. **199**(3): p. 316-31.