



TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

Dissertation

Discovery, Redesign, and Stabilization of Cyclohexanone monooxygenase (CHMO)

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines
Doktors der technischen Wissenschaften unter der Leitung von

Prof. Dr. Marko D. Mihovilovic

Asst. Prof. Dr. Florian Rudroff

Institut für Angewandte Synthesechemie, E163

Und unter Mitbetreuung im Rahmen des OeaD Sonderstipendien Technischen Universität
Wien

eingereicht an der Technischen Universität Wien

Fakultät für Technische Chemie

von

Project. Asst. MSc. Hamid Reza Mansouri Khosravi

Schlachthausgasse 23-29/101, 1030 Wien

Wien, August, 2020

*All your dreams can come true if you have the courage to pursue
them*

Walt Disney

*To my love,
Who is simply amazing.*

*To my parents,
Who supported me in each moment.*



Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Front Matter

Table of Contents

Front Matter	i
Table of Contents	i
Acknowledgments	iv
Abstract	v
Kurzfassung	vii
A Introduction	10
A I History of biocatalysis	10
A II Biocatalyst classification	12
A III Monooxygenases (EC 1.13.x.x and EC 1.14.xx)	13
A III.1 Baeyer-Villiger Monooxygenases	14
A III.1.1 Available BVMO crystal structure	21
A III.1.1 Biotechnological application of BVMOs	24
A III.1.2 Cyclohexanone monooxygenase (CHMO)	25
A III.1.2.1 Mode of action	27
A IV Stability of enzymes	28
A IV.1 Thermodynamic stability	29
A IV.2 Kinetic stability	30
A V Stabilization approaches	31
A V.1 Immobilization	32
A V.2 Additives and lyophilization	35
A V.3 Chemical modification	36
A V.4 Metagenomics (Genome mining)	37
A V.5 Protein engineering	38
A V.5.1 Random mutagenesis (Directed evolution)	39
A V.5.2 Rational design	39
A V.5.3 Semi-rational design	39
A V.5.3.1 Consensus approach	40
B Aim of this study	45
C Results and Discussion	47
C I Enzyme discovery by genome mining	47
C I.1 Genome mining	47
C I.1.1 Text-based Searches Using Enzyme Name	48
C I.1.2 Sequence-driven Approaches	49

Front Matter

C I.1.2.1	Signature/Key Motif-based Strategy	49
C I.2	Blasting and phylogenetic tree	51
C I.3	Multiple structure alignment	55
C I.4	Plasmid construct	57
C I.5	Enzyme expression and purification	57
C I.6	Characterization	58
C I.6.1	Optimum pH for activity	58
C I.6.2	The optimum temperature for activity	59
C I.6.3	Thermodynamic stability (T_m)	60
C I.6.4	Kinetic stability (Half-life)	61
C I.6.5	Solvent tolerance	63
C I.6.6	Kinetic measurement	64
C I.6.7	Substrate profile	66
C II	Redesign	69
C II.1	Substrate profile	73
C III	Protein engineering	77
C III.1	General workflow	78
C III.1.1	First-generation mutants	84
C III.1.1.1	The dissociation constant (Kd)	88
C III.2	Site saturation mutagenesis of G14	89
C III.2.1	MD simulation	94
C III.2.1.1	Analysis of the RMSF of the loop and FAD	95
C III.2.2	Second-generation variants	99
C III.2.3	Third-generation variants	103
C III.2.3.1	Enantioselectivity of third-generation variants	110
C III.2.3.2	Kinetic measurement	111
C III.3	Discussion of the protein engineering approach	115
D	Conclusion and perspective	118
D I	Summary	118
D II	Perspective	121
E	Experimental part	122
E I	Materials and methods-standard microbiological techniques	122
E I.1	General stock solutions	122
E I.2	Media Preparation	122
E I.3	Strain cultivation	123
E I.4	Bradford assay	124
E I.5	Preparation of CaCl_2 competent cells	124
E I.6	Preparation of RbCl competent cells	125
E I.6.1	Preparation of reagents	125
E I.7	Plasmid map and cloning plan	126
E I.8	Transformation of <i>E. coli</i> competent cell	127
E I.8.1	Heat shock transformation	127
E I.8.2	Electroporation transformation	127
E I.8.3	Preparation of SOC medium for cell recovery	128
E I.9	Plasmid purification and sequencing	128
E I.9.1	Nuclease free water preparation	128

Front Matter

E I.9.2	Purification	128
E I.10	Site saturation mutagenesis	130
E I.10.1	Polymerase chain reaction	130
E I.10.2	KLD reaction	131
E I.11	Enzyme expression	131
E I.12	SDS-PAGE	132
E I.12.1	Reagents and buffer preparation	132
E I.12.2	Gel staining	134
E I.13	Enzyme purification	134
E I.13.1	Buffer preparation for enzyme purification	134
E I.13.2	Column recharging	135
E I.14	Kinetic measurement	136
E I.14.1	Activity measurement	136
E I.14.2	K_m measurement	136
E I.14.3	K_d measurement	137
E I.15	Stability measurement	137
E I.16	Melting temperature evaluation	138
E I.17	Biotransformation	139
E I.17.1	List of substrates used for biotransformation	139
E I.17.2	General procedure	139
E I.17.3	Chemical reference reaction	140
E I.18	Bioinformatic tools and methods	140
E I.19	Circular dichroism (CD) spectroscopy	141
F	Appendix	143
F I.1	The consensus sequence for 31 sequences. ^[191c]	143
F I.2	List of sequences	153
F I.3	Melting chromatograms for mutants	163
F I.4	GC methods	168
F I.5	List of primers	170
F I.6	List of sequences used for the phylogenetic tree analysis	171
F I.7	Multiple sequence alignment of BVMO _{Flava} with known BVMOs	174
F I.8	Kd measurements of G14A and CHMO _{Acineto}	177
F II	Publications resulting from this thesis	178
F III	Curriculum vitae	179
F IV	List of abbreviations	183
F V	References	185

Acknowledgments

First of all, I would like to express my gratitude to my supervisor, Marko who gave me the opportunity to conduct my thesis in his lab and also believing in me and encouraging me throughout all my PhD. I also would like to thank Florian for his supervision and all he have taught me.

I also would like to thank my thesis examiners, Roland Ludwig and Uwe Bornscheure, for their commitment and support. To Roland my sincere thanks for supporting and collaborating in my research and also short stays and experiments, which I have performed in his lab. Many thanks to Su in Ludwig lab who helped me to perform my experiments there and supported me during my stay.

I would like also express my gratitude to all people who the conduction of my thesis could not have been possible without them. My former colleagues Thomas, Sofia, Wiesi, Anna, Erna, Patricia. And my special thanks go to Dr. Goncalves who have made my PhD research more interesting and challenging ☺. I also would like to thank all bio people, Julia, Tom, Clemens, Richard, Lydia, Fredy, who have made the biolab a really nice and more fun to work.

I also would like to thank my students Alex, Matina, and Nazanin for their help and support for my thesis and their dedication to the project.

I am also happy, which I had the chance to work with all the people in the MDM/BSC group. The colleagues in the girls lab, Blanca, Rafaela, Eleni, Kathi. I thank David, Viktor, Nik, Christoph, Hubert, Laszlo, Drasi and Dominik for being such a nice colleagues and enjoyed a lot working beside you in the last 5 years. I also want to thank Heci for his warm hugs, energy and super funny and nice small talks. I thank Christian and Michael for being always supportive.

Special thank to Resi and Dani for being always supportive and nice and all the enjoyable dinner and talks we had.

I would love to thank the most special person in my life, who have been beside my through all these time, who encouraged me, loved me and was the reason I could continue my path. I thank Shadi, I would like you know that without you it was almost impossible to go through all the frustrating times and challenges.

Last but not least I would like to thank my Family that supported me no matter what. Specially my father and mother for giving me hope and motivation till now, my brother for the time I was super desperate and did help me to do not quit, my sisters, who have been always giving me hope.

Abstract

Among the high number of flavin-dependent monooxygenases, Baeyer–Villiger monooxygenases (BVMOs) have been studied most for their application as a biocatalyst. A prominent conversion of these biocatalysts is the stereoselective oxidation of aliphatic and cyclic ketones to corresponding esters/lactones that are essential building blocks for the synthesis of bioactive compounds. The prototype BVMO, cyclohexanone monooxygenase from *Acinetobacter calcoaceticus* NCIMB9871 (CHMO_{Acineto}), is a promising biocatalyst for industrial application owing to its broad substrate spectrum and excellent regio-, chemo-, and enantioselectivity. However, the low stability of CHMO_{Acineto} and many other Baeyer–Villiger monooxygenases is an obstacle for their exploitation in the industry. This project is aiming to overcome this significant obstacle in transit to improve stability by using protein engineering and genome mining.

CHMO_{Acineto}, as the most studied BVMO, was chosen for this study. Mutations to stabilize helical structure motifs and also stabilizing enzyme-FAD binding were predicted. Combinatorial and rational protein design approaches were applied for the development of three generations of mutants. These variants were screened for their thermostability and activity. Selected thermally and kinetically stable variants were also screened with several substrates to indicate if they retained their selectivity after the modification. Kinetic stability of CHMO_{Acineto} at 30°C improved from 34 min to 275 min (3rd_{mut-4}); this was achieved by no loss in activity, which turns this variant to a valuable biocatalyst with a reasonable half-life and a high total turnover number (TTN) for industrial applications. The results obtained after the screenings showed that our best kinetically stable variant does not show any significant changes in its substrate scope. The reason for the improvement of stability was addressed by rational protein engineering and molecular dynamic simulations, which revealed a hot spot in the amino acid position 14 of CHMO_{Acineto}.

In a side project, metagenome mining was used to explore the vast data collection of BVMOs to find a thermostable biocatalyst as an alternative for CHMO_{Acineto} as well. This approach resulted in a novel BVMO (BVMO_{Flava}) that showed moderate kinetic stability and high thermodynamic stability. The high deviation between thermodynamic and kinetic stability is a significant problem in the field of BVMO research, since often only T_m values are reported without any perspective to the actual operational stability. This study helps to emphasize how important it is to determine both stabilities for future comparison and putative industrial applications of BVMOs.

Major findings of the thesis are that an increased affinity of FAD to the enzyme leads to a significant increase in kinetic stability; furthermore the power of protein engineering demonstrates how to improve enzyme characteristics and shows the high potential of genome databases to investigate and identify novel biocatalysts as well.

Kurzfassung

Unter der hohen Anzahl an Flavin-abhängigen Monooxygenasen wurden Baeyer-Villiger-Monooxygenasen (BVMOs) am häufigsten für ihre Anwendung als Biokatalysator untersucht. Eine herausragende Anwendung dieser Biokatalysatoren ist die stereoselektive Oxidation von aliphatischen und cyclischen Ketonen zu entsprechenden Estern / Lactonen, welche wesentliche Bausteine für die Synthese bioaktiver Verbindungen sind. Die Prototyp BVMO, Cyclohexanonmonooxygenase aus *Acinetobacter calcoaceticus* NCIMB9871 (CHMO_{Acineto}), ist aufgrund ihres breiten Substratspektrums und seiner hervorragenden Regio-, Chemo- und Enantioselektivität ein vielversprechender Biokatalysator für die industrielle Anwendung. Die geringe Stabilität von CHMO_{Acineto} und vielen anderen Baeyer-Villiger-Monooxygenasen ist jedoch ein Hindernis für ihre Nutzung in der Industrie. Dieses Projekt zielt darauf ab, dieses bedeutende Hindernis auf dem Transitweg zu überwinden, um die Stabilität durch Protein-Engineering und Genom-Mining zu verbessern.

CHMO_{Acineto} wurde als am meisten untersuchte BVMO für diese Studie ausgewählt. Mutationen zur Stabilisierung von Helixstrukturmotiven und zur Stabilisierung der Enzym-FAD-Bindung wurden vorhergesagt. Kombinatorische und rationale Proteindesign-Ansätze wurden für die Entwicklung von drei Generationen von Mutanten angewendet. Diese Varianten wurden auf ihre Thermostabilität und Aktivität untersucht. Ausgewählte thermisch und kinetisch stabile Varianten wurden ebenfalls mit mehreren Substraten untersucht, um anzuzeigen, ob sie nach der Modifikation ihre Selektivität beibehielten. Die kinetische Stabilität von CHMO_{Acineto} bei 30 ° C verbesserte sich von 34 min auf 275 min (3. Mut-4). Dies wurde ohne Aktivitätsverlust erreicht, was diese Variante zu einem wertvollen Biokatalysator mit einer angemessenen Halbwertszeit und einer hohen Gesamtumsatzzahl (TTN) für industrielle Anwendungen macht. Die nach den Screenings erhaltenen Ergebnisse zeigten, dass unsere beste kinetisch stabile Variante keine signifikanten Änderungen in ihrem Substratprofil zeigt. Der Grund für die Verbesserung der Stabilität wurde durch rationales Protein-Engineering und molekulardynamische Simulationen erforscht, wobei ein Hot Spot in der Aminosäureposition 14 von CHMO_{Acineto} enthüllt wurde.

In einem Nebenprojekt wurde der Metagenomansatz verwendet, um die umfangreiche Datenerfassung von BVMOs zu untersuchen und einen thermostabilen Biokatalysator als Alternative für CHMO_{Acineto} zu finden. Dieser Ansatz führte zu einer neuartigen BVMO (BVMO_{Flava}), die eine moderate kinetische Stabilität und eine hohe thermodynamische Stabilität aufwies. Die hohe

Abweichung zwischen thermodynamischer und kinetischer Stabilität ist ein bedeutendes Problem auf dem Gebiet der BVMO-Forschung, da häufig nur T_m -Werte ohne Bezug zur tatsächlichen Betriebsstabilität angegeben werden. Diese Studie unterstreicht, wie wichtig es ist, sowohl Stabilitäten für zukünftige Vergleiche als auch mutmaßliche industrielle Anwendungen von BVMOs zu bestimmen.

Eines der wichtigsten Ergebnisse der Arbeit ist, dass eine erhöhte Affinität von FAD zum Enzym zu einer signifikanten Erhöhung der kinetischen Stabilität führt. Darüber hinaus zeigt die Leistungsfähigkeit des Protein-Engineerings, wie die Enzymeigenschaften verbessert werden können, und zeigt das hohe Potenzial von Genomdatenbanken, auch neuartige Biokatalysatoren untersuchen und identifizieren zu können.

A Introduction

A I History of biocatalysis

Biocatalysis refers to the use of a biological system (enzymes) to catalyze a chemical reaction and to perform this task, one can either use isolated enzymes or enzymes that are inside the living cell (whole-cell catalysis).^[1] The origin of biocatalysis dates backs to more than 7000 years ago, when Sumerians were brewing different types of beer by using fermentation. There are also evidences that in China alcoholic drinks were produced by using fermentation, already 9000 years ago^{[2] [3]} Both cases were unintentionally using biocatalysts by fermentation in whole-cells. There are few differences between fermentation and using enzyme as biocatalyst especially in regard to application to non-native substrates. Fermentation is mostly referred to use of a microorganism to extract energy from carbohydrates in absence of oxygen. In biocatalysis, reactions can be performed in presence of oxygen and the non-native substrate can be used for production of product by biocatalyst.

The question is what is the benefit of using biocatalysts while we have already established and optimized chemical reaction by organic synthesis. The answer is that biocatalysts (enzymes) are offering many valuable features that make them superior to the old fashioned catalysts.

- Biocatalysts are mostly highly selective. They exhibit high chemo-, enantio-, and regioselectivity, which is highly essential due to the high demand of enantiopure compounds in industrial applications.
- Biocatalysts mostly perform the reactions under mild reaction conditions, which is safer and cost-effective than traditional methods.
- Multistep reactions can be performed under highly similar conditions or in a single strain with the ability to regenerate the cofactors, which decreases the cost of production.
- Biocatalysts are environmentally friendly in comparison to their chemical equivalents.
- Due to highly developed genome mining and bioinformatics tools, a vast number of new biocatalysts have been discovered, which have the potential to carry out novel reactions that were not possible by chemical means so far.

Due to the mentioned reasons above, biocatalysts are attracting more attention from scientists and industries. Also, this area of study got an exceedingly active field at the crossroads of organic synthesis, fermentation technology, molecular biology, and microbiology.

Baeyer-Villiger oxidation can be a great example to show the advantages of biocatalysis in contrast to organic catalysis. Baeyer-Villiger oxidation is a widely known method in organic chemistry, but still, several unsolved difficulties diminish attractiveness and also the applicability of this reaction. This reaction is proceeding conventionally by using peroxides. The problem is the shock-sensitivity and explosiveness of most peroxides. [4] Commonly used peracids for Baeyer-Villiger oxidations are made from their corresponding acids and using concentrated hydrogen peroxide. These solutions in high concentrations are susceptible to violent decomposition [5], so they have been mainly avoided for industrial applications. [6] Oxidation with peracids and peroxides also leads to stoichiometric quantities of waste. The other difficulties are the low selectivity and reactivity of these catalysts. [7] To overcome these difficulties, using biocatalyst can be a promising alternative. Baeyer-Villiger monooxygenases (BVMOs) are a group of enzymes, which can perform Baeyer-Villiger oxidations. BVMOs use free and green oxidant O₂, and generate water as a byproduct, which makes them green and effective biocatalysts to perform Baeyer-Villiger reactions. So they attracted considerable attention to be used instead of peroxides and peracids. In addition, the reaction is achieved within the active site of the enzyme representing a strong chirally inducing environment, hence, the biotransformation enables access to optically pure products of high value.

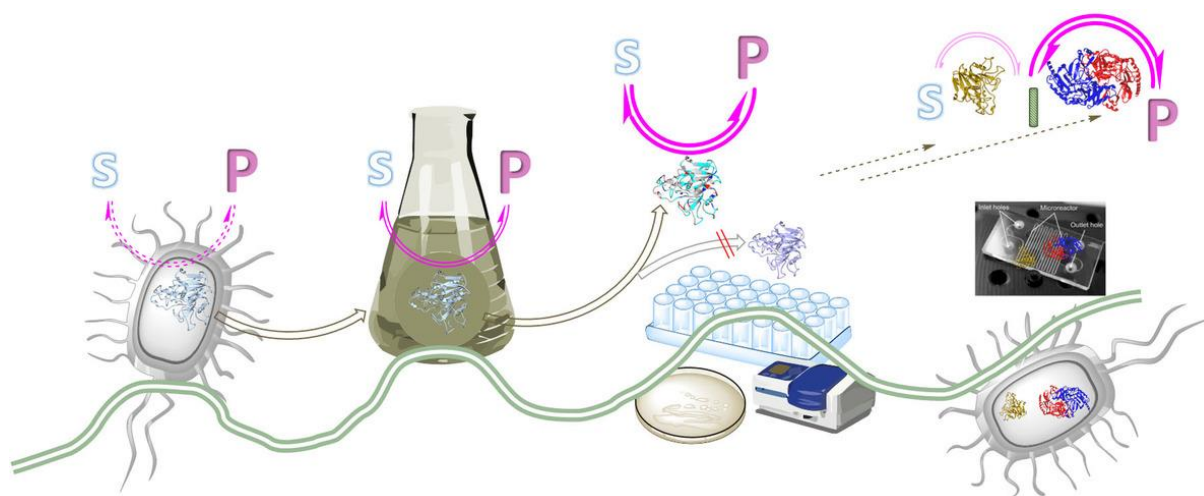
Scientists proposes that biocatalysts developed during different waves, which was reviewed nicely by Bornscheur [8] and Poppe [9]. The first wave of biocatalysis (Figure A.1) started more than a century ago when scientists realized that components of living cells could accomplish chemical transformations such as production of hydrogen cyanide and (R)-mandelonitrile from benzaldehyde using plant-derived biocatalysts, hydroxylation of steroids happening within microbial whole cells or production of l-ephedrine precursor by the aid of yeast cells. [10]

The second wave started (Figure A.1) in the last quarter of the 20th century. During this period, enzymes, media, and substrate screening were developed and initial protein engineering technologies, including site-directed mutagenesis and chemical modifications, allowed scientists to extend the substrate range of enzymes to unnatural compounds. These advances developed several classes of enzymes to be used as a synthetic toolbox and entered biocatalysis into the field of pharmaceutical compounds and fine chemical production. [9-10]

In the mid and late 1990s, the third wave of biocatalysis started (Figure A.1). During the third wave, researchers were able to use advanced molecular biology methods such as error-prone PCR and DNA shuffling combined with a high-throughput assay to perform quick and extensive alteration of biocatalysts. [9-10]

Even though the achievements are not sufficient to predict future developments, it is apparent that we are approaching the fourth wave of biocatalysis. ^[11] The fourth wave combines advanced molecular genetics, bioinformatics tools, and metagenomics, to discover novel enzymes or even generate novel enzyme properties that were not present in nature ^[11], by using advances in immobilization engineering, multienzyme-catalyzed processes, and microreactor technology. ^[12]

Figure A.1. Four waves of biocatalysis.^[9]



A II Biocatalyst classification

Due to the high inconsistency and complexity in enzyme naming, the International Union of Biochemistry did set up the enzyme commission to solve the mentioned issue. Based on the enzyme commission system, enzymes addressed in a four-part enzyme commission (EC) number. Enzymes can be classified based on the type of chemical reaction they carry out. Based on this method, enzymes categorized into seven different categories, which shown in Table A.1. These seven categories are the most abundant forms of enzymes. The first part of the EC number state the reaction, which is catalyzed by the enzyme. Each group of enzymes can further divided into several subgroups according to the characteristics of bonds or the functional groups in the substrates. ^[13]

Table A.1. Different categories of enzymes.

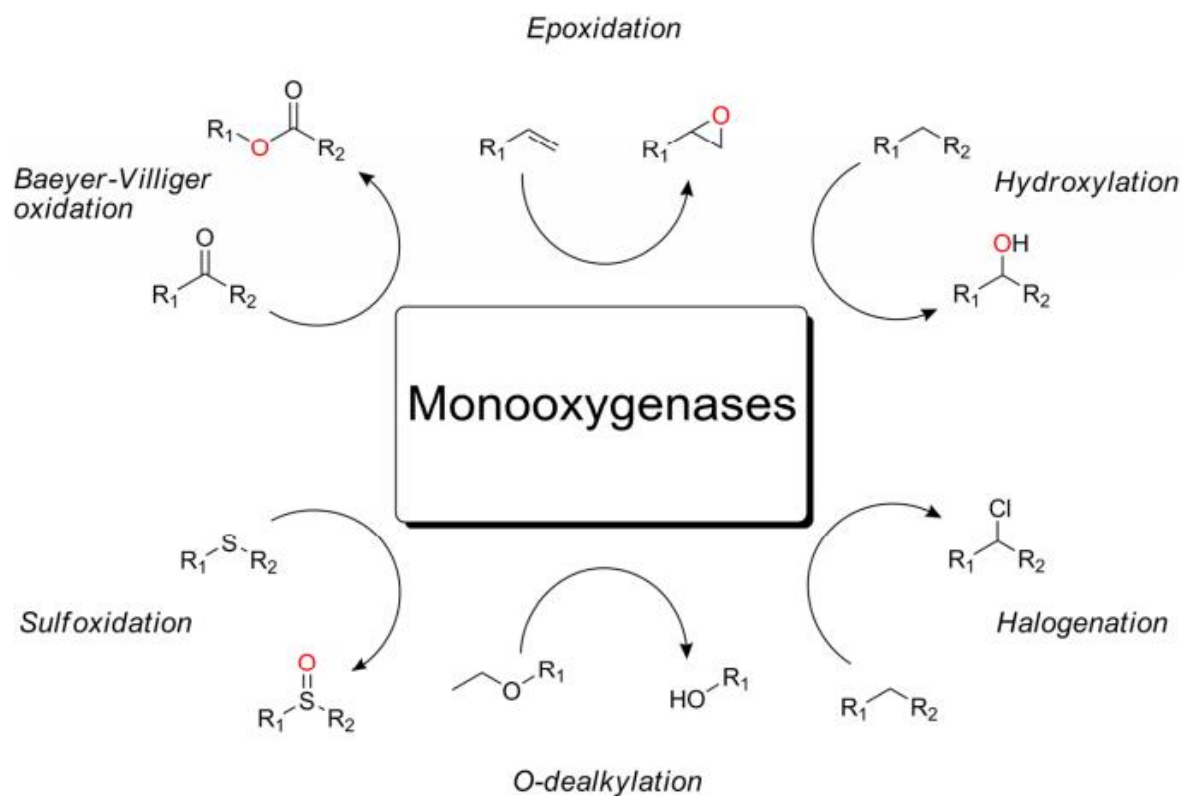
Class of enzyme	Enzyme commission number	Description
Oxidoreductases	EC 1	Catalyzing the redox reaction and subcategorized to reductase and oxidase
Transferases	EC 2	Catalyzing the exchange or transfer of some groups among substrates
Hydrolases	EC 3	catalyze the hydrolysis of starting material
Lyases	EC 4	Removal of a group from the substrate and leave a double bond reaction or catalyze its reverse reaction
Isomerases	EC 5	Enable the conversion of optical isomers, geometric isomers or isoisomers
Ligases	EC 6	Accelerate the conversion of two substrate molecules into a substrate with the release of energy
Translocases	EC 7	Catalyze the separation of ions or molecules within the membranes or movement of them across membranes

A III Monooxygenases (EC 1.13.x.x and EC 1.14.xx)

Monooxygenases belong to the category of oxidoreductases (EC 1), which are responsible for catalyzing redox reactions. These enzymes are capable of inserting a single oxygen atom from O₂ into the substrate. Monooxygenases can be divided into 22 different sub-classes, based on the reaction they are catalyzing and can be further subdivided based on the co-factor they are utilizing. ^[14] The majority of these biocatalysts contain either a flavin or a heme as a cofactor ^[15] but also, pterin-, copper-, non-heme, and iron-dependent monooxygenases have been identified. ^[16] Furthermore,

some monooxygenases can catalyze the insertion of an oxygen atom without using a cofactor. [17] Monooxygenases are capable of catalyzing a wide range of reactions, e.g., epoxidations, hydroxylations, (de)halogenations, Baeyer-Villiger oxidations, heteroatom dealkylations, and sulfoxidation (Figure A.2). Besides oxygen, monooxygenases require two electrons for the reduction of their cofactor. Generally, these two electrons are provided by reduced nicotinamide coenzymes, like NADPH and NADH. Nevertheless, as these coenzymes are expensive and stoichiometric amounts of them are required, it can be a limiting factor for biotechnological applications of monooxygenases.

Figure A.2. Reactions can be performed by monooxygenases. [18]



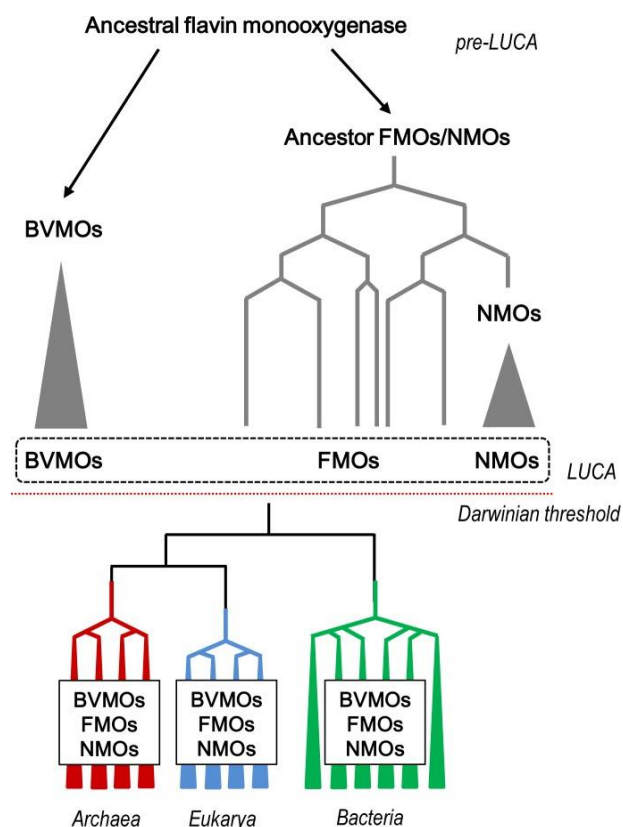
A III.1 Baeyer-Villiger Monooxygenases

Baeyer-Villiger Monooxygenases (BVMOs) belong to the family of oxidoreductases belonging to the "Class B" of flavin monooxygenases. Single domain flavin-dependent monooxygenases are divided into two subclasses (A and B), which perform the flavin reduction and monooxygenation with one

polypeptide chain. The main difference of subclass B in comparison to other subclass is that members of subclass B are strictly NADPH dependent and contain two different domains for FAD and NADPH. They have been identified, isolated, and characterized in the late 1960s.^{[19] [20]}

It is proposed that ancestral flavin monooxygenases genes existed at a pre-LUCA (Last Universal Common Ancestor of all life) time. Flavin monooxygenases genes gave rise to the monophyletic group of BVMOs and to a group of FMOs / NMOs-like genes that later originated the monophyletic group of NMOs and a paraphyletic group of modern FMOs. The multifunctional flavin containing monooxygenases or FMOs are actually eukaryotic flavoenzymes that are able to oxidize a broad range of heteroatom containing compounds and amines and NMO or N-hydroxylating monooxygenases are a class of microbial enzyme that perform the N-hydroxylation of long chain primary amines.^[21] After crossing the Darwinian threshold, a different proportion of flavin monooxygenase genes inherited from the three life domains: Archaea, Bacteria, and Eukarya (Figure A.3).^[22]

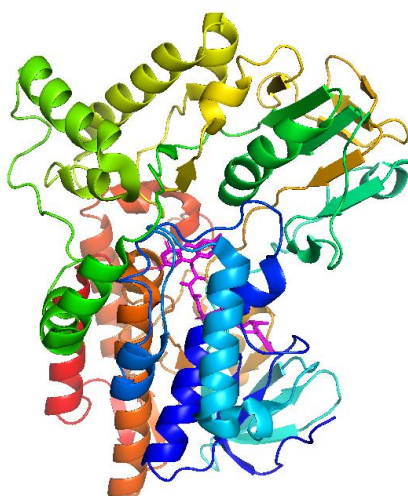
Figure A.3. Scheme of the proposed model for the evolutionary history of “Class B” flavin monooxygenases.^[22]



The first crystal structure of a BVMO was (Figure A.4) determined for Phenylacetone monooxygenase (PAMO) in 2004 ^[23], which helped further studies to understand the mechanism of these biocatalysts that is explained in section A III.1.2.1. Comprehensive knowledge about the structure and the key residues in the enzyme can facilitate the protein engineering studies that can facilitate the alteration in the enzyme cofactor specificity, enzyme activity and stability.

The mechanistic studies in PAMO (A III.1.2.1) that facilitated and made possible by the first crystal structure, ^[24] highlighted the catalytic importance of Arginine residue in position 337 as well as active site, FAD, and NADP binding domains that all became possible. The important information acquired from the structure of PAMO has helped Pazmino and coworkers to identify the role of crucial active site residue (R337) that is critical for the catalytic activity of PAMO. This Arginine is highly conserved in all other type I BVMOs. This key amino acid is required for oxygenation-reduction of the flavin cofactor and the reaction of the C4a-peroxyflavin intermediate with organic substrates. ^[24a] To gain a better a protein engineering study was performed in this position. The strictly conserved R337 in PAMO was replaced with lysine that resulted in drastically reduced rate of flavin reduction. But this mutation did not affect NADPH binding, which means that R337 is responsible for proper alignment of reduced nicotinamide moiety with respect to the isoalloxazine part of the flavin cofactor. This position may also may modulate the redox properties of the flavin cofactor ^[24a].

Figure A.4. Crystal structure of PAMO bound to FAD.



All known BVMOs are utilizing a flavin cofactor (FAD or FMN) that is crucial for their activity, and they also need an electron donor, which can be either NADH or NADPH. There are two types of BVMOs which are classified based on the cofactor they utilize to perform the catalysis. [25] Type I Baeyer-Villiger monooxygenases, which are the most abundant BVMOs. [21] Type I BVMOs are FAD dependent and requires NADPH as the source of electrons. Type I BVMOs are composed of only one polypeptide chain. Type I BVMOs are responsible for the oxidation of ketones to the corresponding ester or lactone, and they carry out this reaction by insertion of an oxygen atom into the C-C bond. The oxidation capacity is not limited to Baeyer-Villiger oxidation, BVMOs are also capable of sulfoxidation, and epoxidations. [25-26],[21, 27] Besides the promiscuity in reactivity, BVMOs are mostly very regio-, chemo- and enantioselective while accepting a broad range of substrates. [28] These monomeric biocatalysts (Table A.2) contains two Rossmann-fold motifs (GxGxxG/A) (Figure A.5). Rossmann-fold motifs indicate that BVMOs bind FAD and NADPH using separate dinucleotide binding domains. In addition, these biocatalysts contain BVMO-specific sequence motifs ([AG]-G-x-W-x-x-x-x-[FY]-[GM]-x-x-x-D) and (FxGxxxHxxxWD/P), which can be used to identify Type I BVMOs in (meta)genome databases (Figure A.5). [29] In some BVMOs, like (MoxY and CPDMO) minor deviations from the consensus for the nucleotide-binding sequence were reported. [30] The exact functional role of these residues is not clear. What is just known is that the long consensus sequence ([AG]-G-x-W-x-x-x-x-[FY]-[GM]-x-x-x-D) (Figure A.5) entails the conserved active-site aspartate and the short fingerprint (FxGxxxHxxxWD/P) is related to the linker connecting the NADP and FAD-binding domains. [23, 24b]

Type II BVMOs are FMN dependent, require NADH as the source of electrons and are composed of two different subunits. The crystal structures of type II BVMOs did not provide a clear picture of how the structure of these biocatalysts looks like, but sequence data reveals a sequence relationship with Flavin dependent luciferases [15b], so it is expected to contain a TIM-barrel fold in the oxygenase subunit. By now, most of the studies were performed with type I BVMOs [21] since they do represent uncomplicated monooxygenase system, being typically soluble and having several developed expression systems, while type II BVMOs are generally more complicated. They show a low solubility and so far not so many recombinant expression systems have been developed for this type of BVMOs. There is a list of characterized type I and II BVMOs depicted in Table A.2. A phylogenetic tree shows the relation of several known BVMOs and is shown in Figure A.7

Figure A.5. Multiple sequence alignment of some of the known BVMOs.

```

BVMOflava  TTRTP-----DVD A I V I GAGFGG I Y M L H K L R N E L G L - S V T A F E K G G G V GG
TmCHMO     TTQTP-----D L D A I V I GAGFGG I Y M L H K L R N D L G L - S V R V F E K G G G V GG
CHMOAcinet  QKM-----D F D A I V I GGGFGG L Y A V K K L R D E L E L - K V Q A F D K A T D V AG
PAMO       SRRQP-----P E E V D V L V V GAGFSG L Y A L Y R L R - E L G R - S V H V I E T A G D V GG
CPMO       NSVDD-----T L D V L L I GAGFTG L Y Q L H H L R - K L G F - K V H L V D A G A D V GG
CDMO       TPREP-----K L D H V T F A F I GGGFSG L V T A A R L R - E S G V E S V R I I D K A G D F GG
HAPMO      TAEEDLRAPRWKDHVASGRDFKVV I I GAGESG M I A A L R F K - Q A G V - P F V I Y E K G N D V GG
          . . . * . * * : : : . . . : . . . *
BVMOflava  TWYFNRYPGAKSD T E G F V Y R Y S F D K D L L R E W N W T T R Y L E Q A D V L A Y L E H V V E R F D L G R D I
TmCHMO     TWYWNKYPGAKSD T E G F V Y R Y S F D K E L L R E Y D W T T R Y L D Q P D V L A Y L E H V V E R Y D L A R D I
CHMOAcinet  TWYWNRYPGALTD T E T H L Y C Y S W D K E L L Q S L E I K K K Y V Q G P D V R K Y L Q Q V A E K H D L K K S Y
PAMO       VWYWNRYPGARCD I E S I E Y C Y S F S E E V L Q E W N W T E R Y A S Q P E I L R Y I N F V A D K F D L R S G I
CPMO       IWHWNCYPGARVD T H C Q I Y Q Y S M - P E L W G E F N W K E L F P N W A Q M R E Y F Y F V D K K L E L S K D I
CDMO       VWYWNRYPGAMCD T A A M V Y M P L L E E T - - - G Y M P T E K Y A H G P E I L E H C Q R I G K H Y D L Y D D A
HAPMO      TWRENTYPCRVD I N S F W Y S F S F A R G I - - - - - W D D C F A P A P Q V F A Y M Q A V A R E H G L Y E H I

BVMOflava  R L N T E V T G A V F D E E S D L W T V - - - T T A T G E T T T A R Y L V N A L G L L A R S N I P D I P G R D G FAGR
TmCHMO     Q L N T E V T D A I F D E E T E L W R V - - - T T A G G E T L T A R F L V T A L G L L S R S N I P D I P G R D S FAGR
CHMOAcinet  Q F N T A V Q S A H Y N E A D A L W E V - - - T T E Y G D K Y T A R F L I T A L G L L S A P N L P N I K G I N Q FKGE
PAMO       T F H T T V T A A A F D E A T N T W T V - - - D T N H G D R I R A R Y L I M A S G Q L S V P Q L P N F P G L K D FAGN
CPMO       S F N T R V Q S A V F D E Q R R E W T V - - - R S L G H Q P I R A K F V I A N L G F G A S P S T P K V E G I E K FKGE
CDMO       L F H T E V T D L V W Q E H D Q R W R I - - - S T N R G D H F T A Q F V G M G T G P L H V A Q L P G I P G I E S FRGK
HAPMO      R F N T E V S D A H W D E S T Q R W Q L L Y R D S E G Q T Q V D S N V V V F A V G Q L N R P M I P A I P G I E T FKGP
          : : * * : : * * * : : : : : * . * . * . * *

BVMOflava  LVHTNAWP D D - - - - - L D - I T G K R V G V I GTGSTG T Q F I I A A A K T A S H L T V F Q R S P Q
TmCHMO     LVHTNAWP E D - - - - - L D - I T G K R V G V I GTGSTG T Q F I V A A A K M A E Q L T V F Q R T P Q
CHMOAcinet  LHHTSRWP D D - - - - - V S - F E G K R V G V I GTGSTG V Q V I T A V A P L A K H L T V F Q R S A Q
PAMO       LYHTGNWP H E - - - - - P V D - F S G Q R V G V I GTGSSG I Q V S P Q I A K Q A A E L F V F Q R T P H
CPMO       WYHTALWP Q E - - - - - G V D - M A G K R V A I I GTGSSG V Q V A Q E A A L N A K Q V T V F Q R T P N
CDMO       SFHTSRWD Y D Y T G G D A L G A P M D K L A D K R V A V I GTGATA V Q C V P E L A K Y C R E L Y V V Q R T P S
HAPMO      MFHSAQWD H D - - - - - V D - W S G K R V G V I GTGASA T Q F I P Q L A Q T A A E L K V F A R T T N
          * : * : : . . : : * : * * * * : : * . . : * . * : .

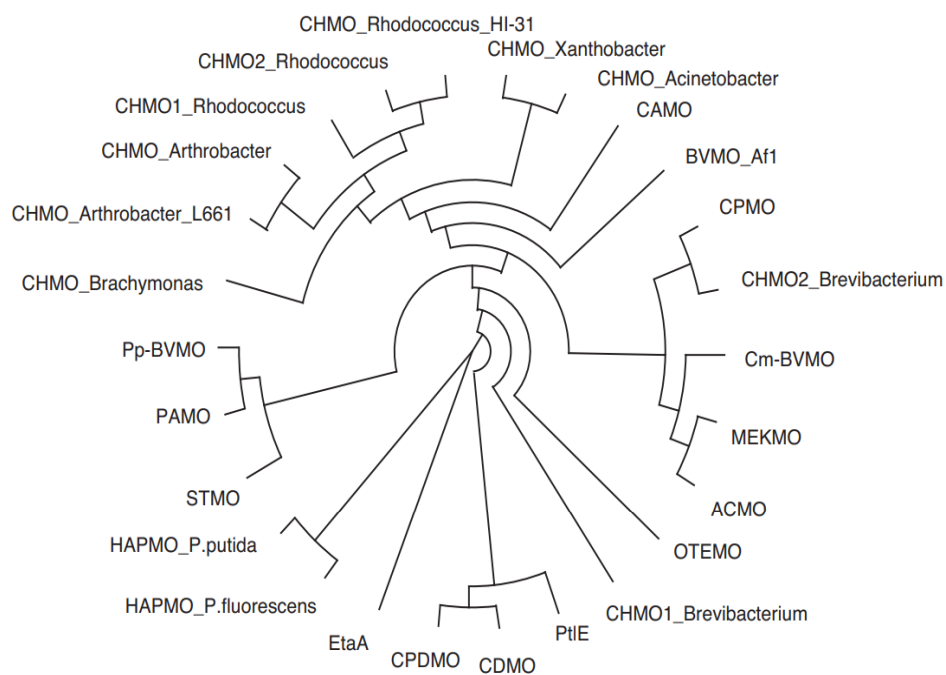
```

Table A.2. Identified and characterized BVMOs.

Biocatalyst	Origin	Reference
CHMO _{Acineto}	<i>Acinetobacter</i> NCIMB 9871	[31]
CPMO _{Coma}	<i>Comamonas</i> NCIMB 9872	[32]
STMO	<i>Rhodococcus rhodochrous</i>	[33]
CHMO _{Brevi 1&2}	<i>Brevibacterium</i> HCU	[34]
HAPMO	<i>Pseudomonas fluorescens</i> ACB	[35]
CDMO	<i>Rhodococcus ruber</i> SC1	[36]
BVMO _{P.putida}	<i>Pseudomonas putida</i> KT2440	[37]
CHMO _{Xantho}	<i>Xanthobacter sp.</i> strain ZL5	[38]
CHMO _{Arthro}	<i>Arthrobacter</i> BP2	[39]
CHMO _{Brachy}	<i>Brachymonas petroleovorans</i>	[40]
CHMO _{Rhodo 1&2}	<i>Rhodococcus</i> Phi1 and Phi2	[39b]
BVMO EtaA	<i>Mycobacterium tuberculosis</i>	[41]
PAMO	<i>Thermobifida fusca</i>	[42]
ACMO	<i>Gordonia sp.</i> strain TY-5	[43]
AKMO	<i>Pseudomonas fluorescens</i> DSM 50106	[44]
MEKMO	<i>Pseudomonas veronii</i> MEK700	[45]
HAPMO	<i>Pseudomonas putida</i> JD1	[46]
BVMO _{ptIE}	<i>Streptomyces avermitilis</i> MA4680	[47]
CHMOR.sp HI-31	<i>Rhodococcus sp.</i> strain HI-31	[48]
BVMO _{penE}	<i>Streptomyces exfoliatus</i>	[49]
OTEMO	<i>Pseudomonas putida</i> ATCC 17453	[50]
SAPMO	<i>Comamonas testosteroni</i> KF-1	[51]
CAMO	<i>Cylindrocarpon radicolica</i> ATCC 11011	[52]

2,5-DKCMO, 3,6-DKCMO	<i>Pseudomonas putida ATCC17453</i>	[53]
SAFMO	<i>Staphylococcus aureus</i>	[54]
BVMO _{AFL}	<i>Aspergillus Flavus</i>	[55]
TmCHMO	<i>Thermocrisum municipale</i>	[56]
PockeMO	<i>Thermothelomyces thermophila</i>	[57]
BVMO _{I.biflexa}	<i>Leptospira biflexa</i>	[58]
MtmOIV	<i>Streptomyces argillaceus</i> ATCC 12956	[59]
BVMO _{Sle_13190 & Sle_62070}	<i>Streptomyces leeuwenhoekii</i>	[60]
BoBVMO	<i>Bradyrhizobium oligotrophicum</i>	[61]
AmBVMO	<i>Aeromicrobium marinum</i>	[61]
BVMO _{Rp}	<i>Rhodococcus pyridinivorans</i>	[62]
BVMO _{Flava}	<i>Amycolaptosis thermoflava</i>	[63]

Figure A.6. BVMOs phylogenetic tree.^[64]



A III.1.1 Available BVMO crystal structure

The determination of BVMO crystal structures is quite beneficial, especially for the investigation of their mode of action. Fifteen years ago, the first crystal structure of a BVMO was solved for PAMO, which made this enzyme to be the first and only structural prototype. [23] The structure shed light on a two-residue insertion displayed by PAMO, which was found to be placed in the active site and called “the bulge” that can be seen in other BVMOs which crystalized later and mentioned in Table A.3. Critical residues could also be identified by resolving the crystal structure of more BVMOs (Table A.3). One of the well studied key residues is a conserved arginine in active site. This crucial active-site arginine acts as an anchoring component for the binding of the ketone substrate. The positively charged guanidinium group of Arginine can enhance the propensity of the substrate to undergo a nucleophilic attack by the flavin-peroxide intermediate. Besides, the arginine side chain, in cooperation with the ribose group of FAD, forms the niche that hosts the negatively charged Criegee intermediate, which is generated upon the reaction of the flavin-peroxide with the substrate. [24b] Furthermore, the presence of an aspartate and an aromatic amino acid can be seen in all BVMOs that is shown in Table A.3.

The available crystal structures of BVMOs are summarized in Table A.3. Cyclohexanone monooxygenase (CHMO) from *Acinetobacter* (CHMO_{Acineto}) has come to be the most-studied and number one prototype BVMO for decades, despite the failure to achieve its structure. Finally, in 2019, Zhang and colleagues could crystalize a mutant of CHMO_{Acineto}. [65] Nevertheless, it remains to be seen whether this structure can serve as a precise approximation to wild type since this mutant contains ten substitutions in the active site. The crystal structure revealed that CHMO_{Acineto} critical residues for catalytic activity are placed in position D57, R327 and aromatic active site residue is located in position 490 (Table A.3). The bulge residues are in position P431 and F432. Generally, the structures of BVMOs are surprisingly similar, despite the low sequence identities of less than 40%. Except for PAMO, many BVMOs are often rather unstable; though, no prominent structural features could acknowledge as the source of this instability. A study, which compared CHMO_{Acineto} and PAMO's tolerance towards cosolvents, a feature commonly shown to be linked to thermostability [66], suggested PAMO's increased number of ionic bridges may cause the higher solvent tolerance, as it could prevent damage to the structure. [67] The same reasoning was given for the higher stability of a recently crystalized thermostable CHMO from *Thermocrispum municipal* (TmCHMO). [56]

BVMOs have a multi-domain architecture consisting of an NADP-binding, a FAD-binding, and a helical domain. The helical domain differentiates BVMOs from other class B flavoprotein monooxygenase

families and leads to a partial shielding of the active site and the formation of a tunnel towards the active site. The other feature in some BVMO subgroups is the N-terminal extensions with varying lengths. This structure established in PockeMO, where it forms a long helix and several loops that wrap around the enzyme. ^[68] PockeMO shows higher thermostability than most BVMOs, but it is not clear whether the extension plays a role in the stability or not. Such a function proposed for 4-hydroxyacetophenone monooxygenase (HAPMO), where removal of the extension was not tolerated when exceeding a few amino acids. ^[69] The deletion of nine amino acids already reduced stability and also reduced the enzyme's tendency to dimerization.

Besides nicotinamide, FAD was found in all BVMOs crystal structures following its tight binding to the enzymes. ^[15b] There is specific structural mobility of cofactors (NADPH/NADP and FAD/FMN), and loops in BVMOs structure that have detected and the debate on its role in catalysis properties have recently reviewed. ^[70]

Table A.3. Available BVMO structures. [71]

Abbreviation	Name	Strain	PDB ID	Critical residues ^a				Ref
				D	R	Ω	Bulge	
CHMO _{Acineto}	Cyclohexanone monooxygenase	<i>Acinetobacter calcoaceticus</i> NCIMB9871	6A37	57	327	W 490	P–F 431- 432	[65]
Af838MO	<i>Aspergillus flavus</i> monooxygenase 838	<i>Aspergillus flavus</i> NRRL3357	5J7X	63	337	W 502	PTAF 441- 444	[72]
RhCHMO	Cyclohexanone monooxygenase	<i>Rhodococcus sp.</i> HI-31	3GWD, 3GWF, 3UCL , 4RG3c , 4RG4	59 W	329	W 492	P–F 433- 434	[48, 73]
RpCHMO	Cyclohexanone monooxygenase	<i>Rhodococcus sp.</i> Phi1	6ERA , 6ER9	60	330	W 493	P–F 434- 435	[74]
TmCHMO	Cyclohexanone monooxygenase	<i>Thermocrispum municipale</i> DSM 44069	5M10 , 5M0Z, 6GQI	59	329	W 492	P–F 433- 434	[56]
OTEMO	2-oxo-Δ ³ -4,5,5-trimethylcyclopentenyl-acetylcoenzyme A monooxygenase	<i>Pseudomonas putida</i> ATCC 17453	3UOV, 3UOX, 3UOY, 3UOZ, 3UP4, 3UP5	59	337	W 501	GSTF 440- 443	[50]
PAMO	Phenylacetone monooxygenase	<i>Thermobifida fusca</i> YX	1W4X, 2YLR, 2YLS, 2YLT , 2YLWb , 2YLBx, 2YLZ , 2YM1 , 2YM2 , 4C74, 4C77 , 4D03b , 4D04 , 4OVI	66	337	W 501	PSAL 440- 443	[23, 24b, 75]
PIBVMO	Parvibaculum lavamentivoran monooxygenase	<i>Parvibaculum lavamentivorans</i>	6JDK	67	340	W 504	PSGF 443- 446	[76]
PockeMO	polycyclic ketone monooxygenase	<i>Thermothelomyces thermophila</i> ATCC 42464	5MQ6	133	426	Y 600	S–Q 536- 537	[68]
STMO	Steroid monooxygenase	<i>Rhodococcus rhodochrous</i>	4AOS, 4AOX, 4AP1 ,4AP3	71	342	W 506	PSVL 445- 448	[77]

^aD: active-site aspartate, R: active-site arginine, Ω: active-site aromatic residue, bulge: active site insertion loop.

A III.1.1 Biotechnological application of BVMOs

Due to the low stability and dependency on expensive coenzymes, only a few examples of industrial-scale applications of these valuable biocatalysts are known and usually did not proceed beyond the pilot scale (Table A.4). In one of the few industrial processes, a heavily mutated CHMO_{Acineto} variant was used for the enantioselective sulfoxidation of pyrimetazole to produce esomeprazole. [78] The wild-type CHMO_{Acineto} has hardly any activity on pyrimetazole. But after 19 rounds of partially random and partially rational protein engineering evolution, the activity highly improved (~140000 fold improvement in productivity over wild type) with excellent enantioselectivity (>99% ee), and the low percentage of overoxidation to sulfone. Also, the cofactor efficiency was improved, which means less NADP⁺ was needed. The drawback of this example is that most companies and research groups do not have enough financial aid or proper equipment to perform such an extensive protein engineering evolution and to do such a complicated process optimization. While researchers are finding promising results, the academic route of biocatalyst development is quite slow, which partially describes the lack of industrial applications of BVMOs. [71]

Table A.4. Preparative scale reaction with BVMOs. [71]

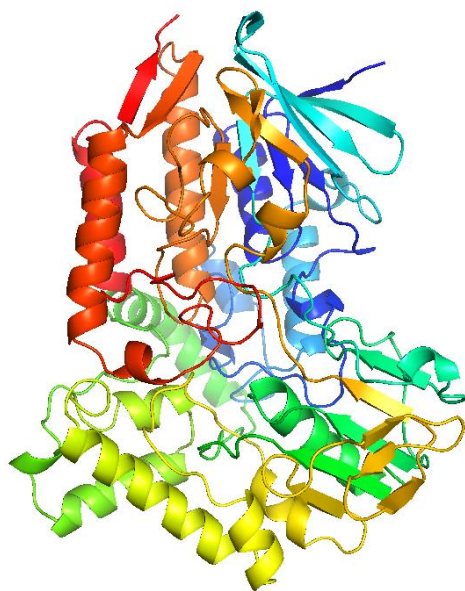
Enzyme	Product	Product concentration	Biocatalyst yield
CHMO _{Acineto} (multiple mutant)	Esomeprazole	50 (~151 mM) 87% yield (28.7 g)	50
CHMO _{Acineto}	bicyclo[3.2.0]hept-2-en-6-one lactone	4.5 (~41 mM) 55% yield (0.49 kg)	3 ^a
<i>Pseudomonas putida</i> (E6-BVMO C302L)	BVMO (Z)-11-(heptanoyl-oxy)undec9-enoic acid	41 (132 mM) 68 % yield (75 g)	1.6 ^a
TmCHMO	3,3,5-trimethyl-caprolactone	24.4 (~156 mM) 76% yield (1.9 kg)	0.6 ^a
CHMO _{Acineto} C376L/M400I/T415C/A463C	6-hydroxy-hexanoic acid	20 (~151 mM) 81% yield (8.1 g)	0.7 ^a
CHMO_Phi1	Lactone of (2R, 5R, 6R)-6-methylidihydrocarvone	0.82 (4.5 mM) 90% yield (49 mg)	6.7
CPDMO	Precursor of Nylon-9	(70 mM) 70% yield (33g)	2.3 ^a

^a gproduct / gcdw = gram product per gram cell dry weight.

A III.1.2 Cyclohexanone monooxygenase (CHMO)

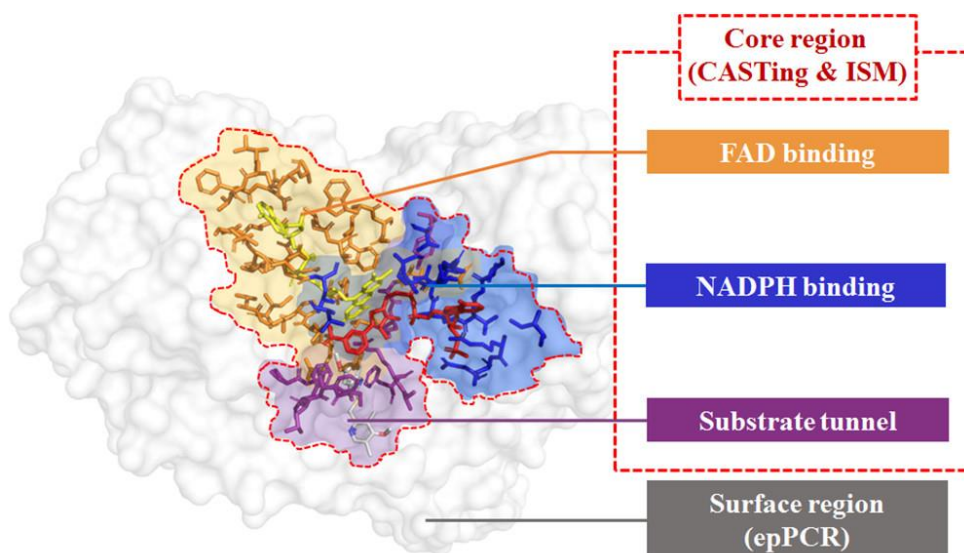
The most well studied BVMO is a bacterial flavoenzyme called cyclohexanone monooxygenase (Figure A.7) from *Acinetobacter* sp. NCIMB 9871 (CHMO_{Acineto}, EC 1.14.13.22), which was first identified and purified by Trudgill and co-workers. [32a]. They were studying microorganisms that can grow on non-naturally occurring aliphatics. They recognized oxygen- and NADPH-dependent biocatalysts from *Acinetobacter calcoaceticus* (NCIMB 9871), which was involved in the microbial metabolism of cyclohexane. They suggested that this biocatalyst catalyzes the conversion of cyclohexanone to ϵ -caprolactone. Their finding was confirmed by isolating the protein and proven that this protein contains a FAD cofactor as a prosthetic group. [32a] This BVMO, which is called cyclohexanone monooxygenase (CHMO_{Acineto}) quickly attracted attention because of its broad substrate scope and also because its product (caprolactone) was already well known as an essential precursor to nylon 6. [31, 79]. Moreover, this enzyme is able to oxidize structurally diverse compounds with high regio-, chemo- and enantioselectivity, which leads to production of optically pure products. This ability will turn this biocatalyst economically favorable, as it can reduce the cost of product purification. Apart from the classical transformations of racemic substrates to form chiral products upon kinetic resolutions and of the desymmetrization of prochiral starting materials, BVMOs are also capable to conduct regiodivergent biooxidations governed by stereoelectronic effects. Beside, this is an enzyme which is capable of performing dynamic kinetic resolution, which facilitate the total use of different enantiomer of a substrate for production of pure enantiomer of the product. This again reduce the cost for substrate and product purification. The homology model was made by SWISS-MODEL and we have used the crystal structure of first ever crystalized CHMO_{Acineto} that is a mutant enzymes and TmCHMO as the model for homology preparation (Figure A.7).

Figure A.7. Homology model of cyclohexanone monooxygenase. The homology model was made using SWISS-MODEL and visualized by PyMOL.



In the first-ever crystal structure of CHMO_{Acinetobacter}, Zhang and coworkers did divide the structure into two main parts (Figure A.8). The core section which contains the FAD binding site, the NADPH binding site, and a substrate tunnel. A substrate-binding site can be identified close to the flavin ring in CHMO_{Acinetobacter}.^[48] The second part is the surface region, which is not involved in the catalytic action of the enzyme.

Figure A.8. First-ever crystal structure of CHMO (mutated).

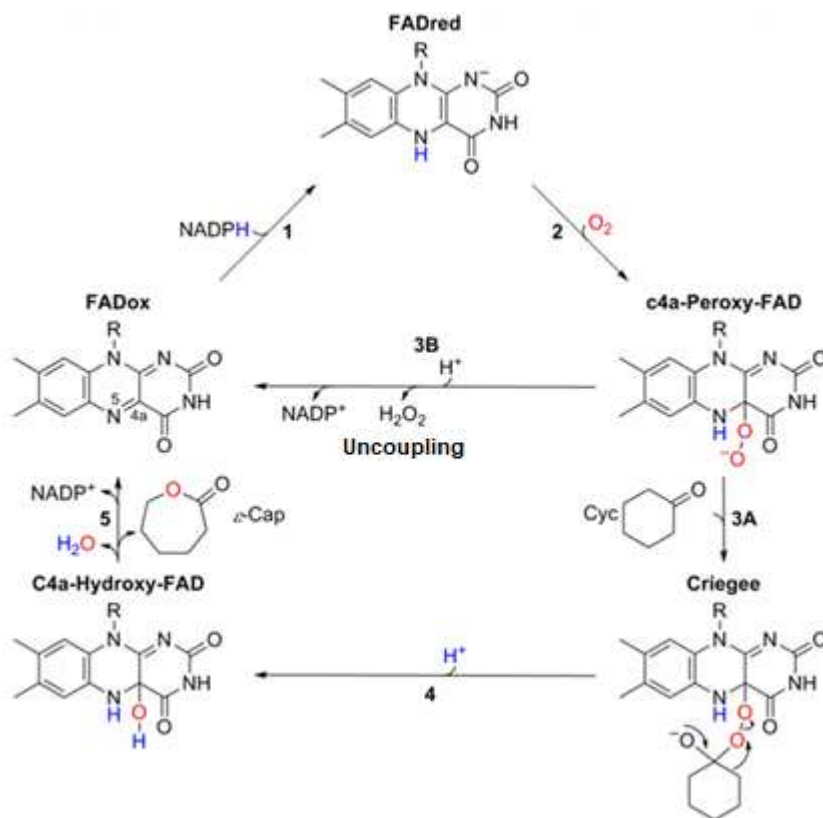


There are several reasons, which makes this enzyme an interesting biocatalyst for further research. CHMO_{Acineto} is an enzyme with an amazingly diverse range of accepted substrates (more than 100 non-natural substrates). It can oxidize structurally diverse compounds with high chemo, regio-, and enantioselectivity [28b, 63, 80]. Like many other BVMOs, CHMO_{Acineto}'s oxidizing capacity is not limited to the Baeyer-Villiger reaction. It also can perform sulfoxidation, amine oxidations, and epoxidations. [56, 81] It produces highly valued lactones and other crucial chiral building blocks while producing only water as a clean byproduct. Due to valuable features of CHMO_{Acineto}, many different industrial applications have suggested for this biocatalyst [82], but the main barriers to exploiting CHMO_{Acineto} as a biocatalyst for industrial applications is its lack of stability. [83]

A III.1.2.1 Mode of action

Pre-steady-state kinetic studies could elucidate the CHMO mechanism (Figure A.9). [84] While the enzyme is in the resting mode, the non-covalently bound FAD is in its oxidized form (FAD_{ox}). It is worth to mention here that the affinity of biocatalyst to the FAD and NADPH will affect several characteristic of enzyme and also reaction. This mainly affect the stability of enzyme and also catalytic efficiency. To perform the reaction, this complex needs to be reduced by NADPH to the enzyme-NADP⁺ complex. Then, the enzyme-NADP⁺ complex will be oxidized by molecular oxygen and form 4 α -flavin peroxide. In the presence of a ketone (cyclohexanone), a nucleophilic attack on the carbonyl group takes place by the reactive 4 α -flavin peroxide that will lead to the formation of Criegee intermediate. Later on it will rearrange to the corresponding lactone and a Flavin hydroxide. By the elimination of water and the release of the product and NADP⁺, the cycle will be completed. What is important to mention when we talk about the BVMOs mode of action is the uncoupling. It has been detected for BVMOs, which in the lack of suitable substrate, the peroxyflavin intermediate can decompose and forming hydrogen peroxide. [24a] Thus, a BVMO can act as an NADPH oxidase and show fake activity especially when you measure the activity by NADPH consumption approach. This procedure was called uncoupling that is unfavorable for a few reasons. First, a high uncoupling rate may show high signals in the screening while conversion is performed at a lower rate and this makes the comparison unreliable. In some cases, due to uncoupling, the activity can be tracked while the enzyme is inactive in the Baeyer-Villiger reaction. This can result in the isolation of false-positive clones in the NADPH-consumption-based screening method. Additionally, enzymes with high uncoupling rates, work less efficiently because of the higher coenzyme (NADPH) requirements and often presenting reduced catalytic rates.

Figure A.9. Catalytic cycle of cyclohexanone monooxygenase with cyclohexanone (cyc) as substrate.^[56]



A IV Stability of enzymes

Protein stability has come to have some different meaning, but most frequently refers to the tendency of an enzyme to reversibly unfold (thermodynamic stability) (Table A.5). In other explanations, protein stability refers to the length of time that the enzyme will remain active before going to deactivation (kinetic stability) (Table A.5). The stability of the biocatalyst is a crucial factor, which determines whether the industrial application of the biocatalyst will be commercially successful or not^[85]. Despite many desirable features of enzymes, lack of stability in most of the BVMO enzymes prevented them from exploitation in industrial applications. Although it is more desirable to perform the reaction under mild conditions, sometimes extreme process conditions cannot be avoided. For example, the use of elevated temperatures and cosolvents in some reactions can increase the efficiency of the industrial process^[85b, 86]. Employing the enzyme to work under harsh reaction conditions (high temperature, high

content of organic solvent, and harsh pH condition), will result in enzyme deactivation, which is not favorable. This issue is addressed more comprehensively in the recent reviews by Bommarius and Paye^[87] and Polizzi *et al.*^[88]

Table A.5. Different types of stability and definitions.

Measure	Type	Symbol	Definition
Free energy of unfolding	Thermodynamic	ΔG_u	Change in Gibbs free energy going from the folded to the unfolded state
Melting temperature	Thermodynamic	T_m	The temperature at which half of the protein is in the unfolded state
Unfolding equilibrium constant	Thermodynamic	K_u	The concentration of unfolded species divided by the concentration of folded species
Half-concentration	Thermodynamic	$C_{1/2}$	The concentration of denaturant needed to unfold half of the protein (the chemical equivalent of T_m)
Observed deactivation rate constant	kinetic	$K_d, \text{ obs}$	The overall rate constant for going from native to deactivation species
Half-life	kinetic	$t_{1/2}$	The time required for the residual activity to reduce to half
The temperature of half- inactivation	kinetic	T_{50}	The temperature of incubation to reduce residual activity by half during a defined period
Optimum temperature	kinetic	T_{op}	Temperature leading to the highest activity
Total turnover number	kinetic	TTN	Moles of product produced over the lifetime of the catalyst

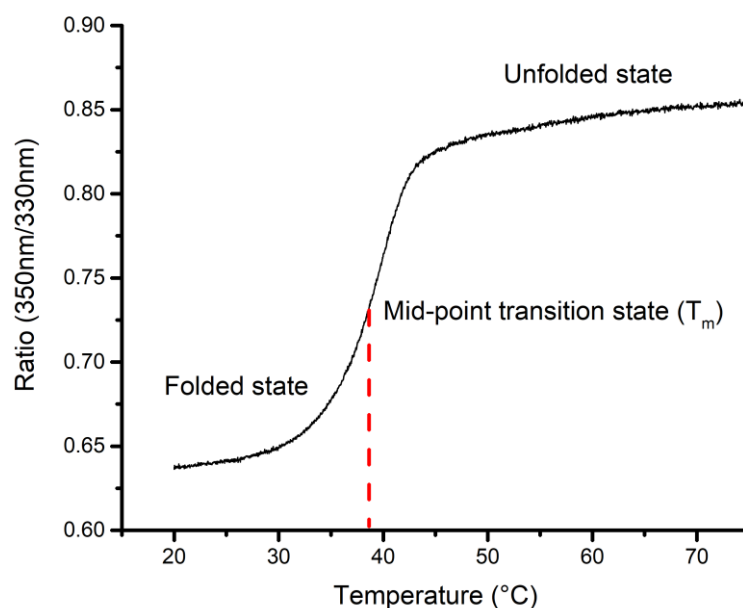
This difference in addressing the stability is leading to confusion and makes the comparison in the literature data impossible. So here is a summary of these two terms provided for better understanding throughout this thesis.

A IV.1 Thermodynamic stability

Thermodynamic stability defined by the enzymes free energy of stabilization (ΔG_{stab} reflecting the difference between the free energies of the folded and the unfolded conformation of the protein) Thermodynamic stability is often measured by melting temperature (Figure A.10) (T_m , the temperature at which 50% of the protein is unfolded).^[89] This type of stability gives more information about the structural stability of the enzyme and less information about the stability under reaction conditions.

Thus, it cannot be used as an absolute determining factor to evaluate the enzyme stability under the reaction condition.

Figure A.10. Thermodynamic stability. This type of stability is represented by T_m or melting temperature. This helps to understand the temperature the enzyme starts unfolding.

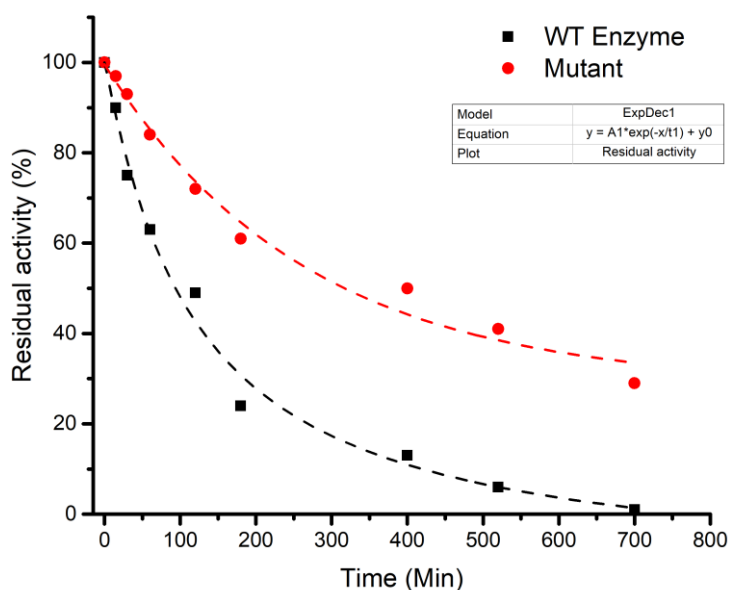


A IV.2 Kinetic stability

Kinetic stability refers to the period on which a protein remains in the functional form before undergoing irreversible denaturation. Kinetic stability generally expressed as the enzyme's half-life ($t_{1/2}$) at a defined temperature, which is the monitoring of enzymes catalytic deactivation by measuring the residual activity (Figure A.11). Kinetic stability is more valuable for the determination of the enzyme stability for the industrial application because it gives valuable information about the enzyme catalytic efficiency under the reaction condition. [89-90]

It is assumed and has been shown in some cases [91] that these two stabilities correlate since increasing the thermodynamic stability (enzyme resistance to unfolding) also increases kinetic stability (resistance to inactivation). However, this correlation is not absolute, particularly not when it comes to denaturation processes that do not or to a minor extent depend on folding stability (e.g., temperature-induced deamidation of Asn and Gln, oxidation of surface residues). [85b]

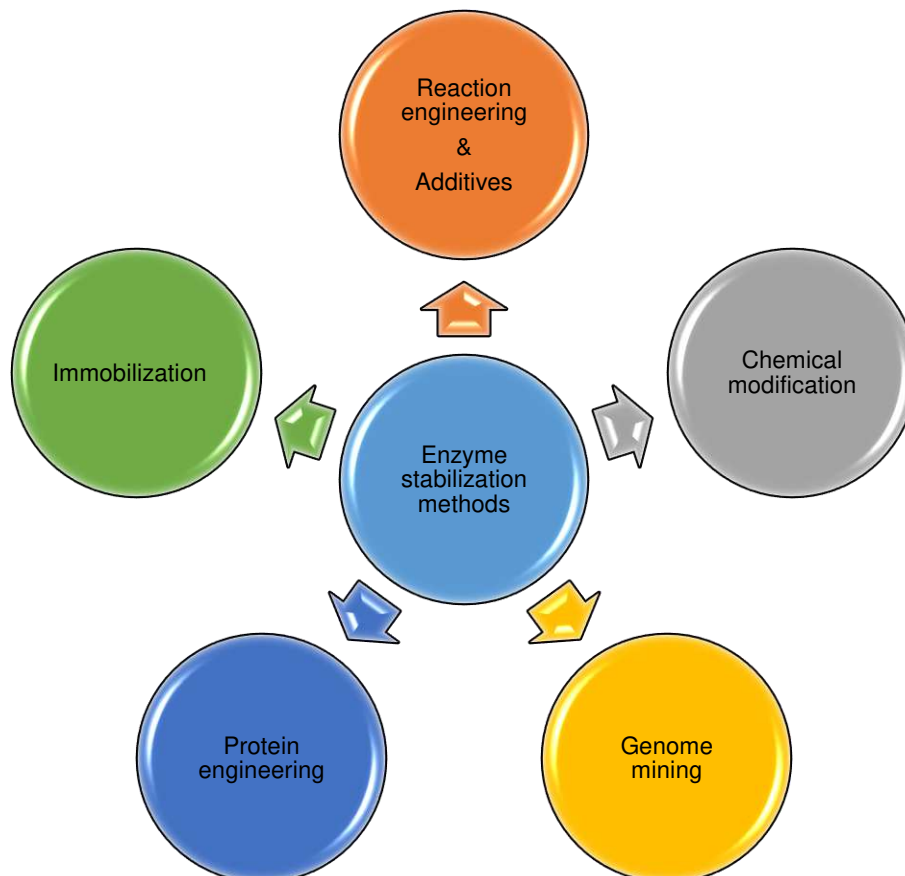
Figure A.11. Kinetic stability. This type of stability represented mostly by Half- life ($t_{1/2}$) which will be calculated by measuring the activity of the enzyme during the time while it is incubated at a certain temperature. The residual activity will be drawn versus time and the half-life will be measured by using exponential decay.



A V Stabilization approaches

To use the high potential of biocatalysts that suffer from low stability, several different approaches were established, which can help to overcome this obstacle. Stabilization of protein, in general, is referring to any process involved in conserving the structure and integrity of a protein and maintaining it from aggregation or degradation. Different approaches that are traditionally in use to improve stability shown in Figure A.12. These approaches will be discussed more in detail in the next few pages.

Figure A.12. Enzyme stabilization approaches.



A V.1 Immobilization

The term “Immobilized enzymes” is defined as “Enzymes that are physically attached to specific solid supports and thus confined, and which can be used repeatedly and continuously while maintaining their catalytic activities”. [92]

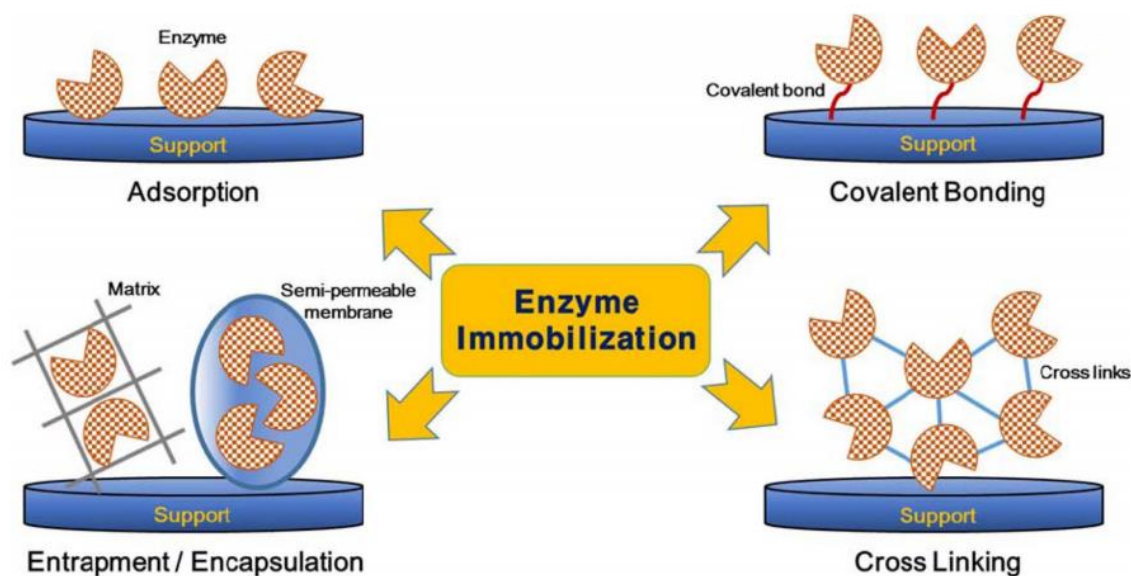
During the past decades, scientists have been actively performing biochemical and biophysical studies to enhance the catalytic efficiency and stability of enzymes through immobilization. [93] Enzyme immobilization significantly improved the functional efficiency of the enzymes and performance of industrial processes, thereby it can be an approach to overcome the low stability problem of enzymes and helps to implement enzymes for industrial applications (Figure A.13). [94] The use of immobilized enzyme has several advantages such as: First, a single batch of enzymes can be used several times which decreases the cost of enzyme production. Second, immobilization of enzymes will help the

enzyme to get more stable than free enzymes. Third, the enzymes can separate from the final product, which facilitates downstream processing and helps to avoid the contamination of product. Last but not least, the use of immobilized enzyme will facilitate the development of multi-enzymatic reaction systems. ^[95]

Alongside the advantages, the immobilization approach is facing a few limitations that need to be considered. For instance, the high activity of native enzymes such as amylases or proteases is intensely reduced when these enzymes are immobilized, which is due to diffusion restrictions. In big-scale production, the cost for the carrier that is needed for immobilization needs to be considered that can make can make the immobilized enzyme less economical. ^[96]

Adsorption approach can protect enzymes against proteolysis, aggregation, and hydrophobic interactions. ^[97] Enzyme adsorption can be performed by salt linkages or hydrophobic bonding between carrier material and enzyme. To reach this purpose support can be dipped in an enzyme solution and let it physically adsorbed. It is also possible to adsorb the enzyme by drying it on the surface of an electrode. The other approach for immobilization is covalent bonding. Enzymes have several different side chain amino acid residues with different reactivity, which can be used for covalent binding of enzymes to support materials. ^[98] For example, silanized silica gel carriers with removed unreacted aldehyde groups can bound covalently to enzymes that is resulting in highly stable and active biocatalysts. ^[99] Considerable improvement of enzyme stability and heat endurance can be seen in enzymes covalently bound to mesoporous silica and chitosan. Covalent binding of enzyme to electrospun nanofibers will lead to an increase in porosity and surface area that results in tremendously increased enzyme residual activity. Entrapment is also used as a method for immobilization that involves the detention of enzymes in fibers or gel by non-covalent or covalent interactions ^[100]. Entrapment in carriers like alginate–gelatin– calcium can prevent enzyme leakage and improve mechanical stability. Nanoparticles such as electrospun nanofibers are used as a support material for entrapment that has revolutionized the field of enzyme immobilization with their wide-ranging applications in the field of biosensors, chemistry, biofuels, and biomedicine. ^[101] Cross-linking is another irreversible approach to immobilize enzymes. This can be achieved by cross-linking of enzymes by covalent binding The method is also called carrier-free immobilization. ^[102] The cross linking is performed with the assistance of a multifunctional reagent that acts as linkers to connect enzymes. The immobilized enzyme will be present in the reaction mixture and does not require support. This is a quite simple method, which provides minimum leakage. ^[103]

Figure A.13. Immobilization methods^[103a]



A lack of stability typically hinders the industrial application of Baeyer-Villiger monooxygenases. Immobilization of these enzyme class can help to overcome this problem. What is challenging in the field of BVMOs immobilization and all other redox enzymes is the accessibility of cofactors that can be limited due to immobilization. The access of immobilized enzyme cofactor binding pocket can sometimes be limited in respect to what orientation the attached enzyme has on the support and where the binding pocket is faced. This issue can be addressed also to the substrate binding pocket as well. When we are thinking of BVMOs immobilization, it is also important to consider the oxygen supply that can be diminished due to immobilization.

Delgova and co-workers could successfully immobilize the thermostable CHMO (TmCHMO) and co-immobilize it to glucose dehydrogenase and could synthesize ϵ -caprolactone derivative from 3,3,5-trimethylcyclohexanone. They could reuse immobilized TmCHMO up to 15 times and achieve 11.2 mg product/mg biocatalyst and 83 % conversion. By this approach, they also could regenerate the cofactor, which decreases the final cost for the product formation. ^[104]

A V.2 Additives and lyophilization

The other options for stabilization and long term storage of enzymes are lyophilization. To increase the stability for long-term storage, lyophilizing the isolated enzymes with appropriate additives or shock freezing them with added glycerol proved to be efficient. ^[105] Lyophilization has been used for decades for long term storage of the biocatalysts. Lyophilization is an interesting approach for long term stability of enzymes as it simplifies shipping and storage. Besides prolongation of storage time, lyophilization allows the screening of new substrates and simplifies the use of enzymes in non-aqueous media as well. ^[106]

Deactivation of an enzyme can occur due to several reasons such as freezing, agitation, heating, and chemical degradation in an aqueous formulation. Lyophilization is one of the approaches that can help to protect enzymes against some of these effects. ^[107] Lyophilization will remove the water, which protects enzyme against water-mediated processes such as hydrolysis. This also can result in the formation of an amorphous phase. It has also been seen that below the glass transition temperature of this amorphous state, unfolding and aggregation are significantly reduced ^[108]. It is also important to mention that additives can also help to achieve better lyophilization quality of proteins. Additives will prevent aggregation and rehydration during the lyophilization process. Additives also can compensate for the loss of essential waters during lyophilization. ^[108-109]

This approach is generally a mild method to the enzyme, but for an unstable enzyme like CHMO_{Acineto}, it has been seen that after lyophilization with no additives, it loses almost half of its activity. It was reported that 6.5-45 mg/ml of sucrose could increase the stability of CHMO_{Acineto} during the lyophilization. This resulted to preserve enzyme activity when stored lyophilized at 50 °C while the enzyme lyophilized without sucrose was inactive at 50 °C. This finding then was transferred to other monooxygenases like CPDMO and P₄₅₀ BM₃ and with similar effects. Complementary to sucrose, additives like serine can also increase the stability of CHMO_{Acineto} upon lyophilization. ^[108]

The additives also can boost the stability upon the storage and also in the reaction condition. Goncalves and coworkers could successfully increase the stability of the flavin-dependent monooxygenases under reaction condition by using an engineered formulation of additives (the natural cofactors NADPH and FAD, and superoxide dismutase and catalase as catalytic antioxidants). By this approach, they could improve the stability by a 10³- to 10⁴-fold, which is by far the most improved stability in the family of monooxygenases to date. ^[110] In another attempt, they could boost the kinetic stability of an enoatereductase (XenB) by more than 20 fold in MOPS buffer compared to that in Tris-HCl buffer, and a pronounced positive effect on thermodynamic stability observed. ^[111]

A V.3 Chemical modification

Chemical modification was the first method developed for improving enzyme stability. This approach was demonstrated in the early 60s. Chemical modification of a biocatalyst can affect function and stability related properties of the protein. ^[112] This approach can be used as a replacement or complementary method to recombinant DNA technology for improving the protein properties. Chemical modification has some benefits for protein stabilization, in particular in context to commercial exploitation due to the growing complexity and difficulty in obtaining regulatory approval for genetically modified (GMO) products. ^[112] This approach is feasible as it is relatively inexpensive, can be performed in most laboratory settings, and offers rapid improvements. ^[113]

Amino and carboxyl groups as the reactive groups are mostly modified ^[114], although aliphatic groups can also be chemically modified. ^[113c] Chemical modification by targeting protein surface amino acids and avoiding modification of the active-site may lead to improvements in protein stability, activity, and efficiency. ^[113c, 115]

Chemical modification with various polyethylene glycol derivatives and other polymers was used by DeSantis and co-workers to increase the stability of biocatalysts toward organic solvents, high temperature, and extreme pH. ^[116] Villalonga *et al.* ^[117] reported the thermal stabilization of phenylalanine dehydrogenase by glycosidation with functionalized β -cyclodextrin derivatives. Modifications resulted in retaining >60% of the activity and also resulted in a 10°C increase in the optimum temperature. ^[117] Chemically modified cysteine protease papain showed an increase of 20°C for optimum reaction temperature and 3-fold higher thermostability. ^[118] Also, the pH optimum increased from pH 7 to pH 9. All this happened with the modification of lysine residues with various dicarboxylic anhydrides. Modification with citraconic anhydride exhibited the maximum impact on temperature stability, while the catalytic proficiency remained largely unchanged. All tested anhydrides produced a shift in pH optimum. ^[118] Barbas and colleagues expanded a rapid thiol-selective protein modification approach in a recombinant HSA and a fusion maltose-binding-HA peptide protein (MBP-C-HA), producing fluorophore and PEGylated conjugates, which illustrated superior stability in comparison to maleimide-conjugated proteins in human plasma. ^[119] Indeed, this novel approach also enabled the chemical modification of engineered cysteine and selenocysteine residues of antibody conjugates that successfully improved the stability of the modified protein. ^[120] Disulfide modification is another chemical modification, which allows the site-selective conjugation of polymers like polyethylene glycol (PEG) to proteins from which optimal pharmacokinetics were accomplished, therefore ensuring protein maintenance of the tertiary structure, purity, and stability. ^[121] Siar and

coworkers, aminated the biocatalyst using carbodiimide and ethylenediamine to convert all exposed carboxylic groups into amino groups. Besides, the activity was significantly higher than the reference or the free enzyme in 8 M urea. The enzyme exhibited a much higher specific activity than the non-aminated enzyme. Finally, the enzyme reused for five cycles of casein hydrolysis without any decrease in enzyme activity. ^[122] The functionalization of such a complex enzyme with retention of the quaternary structure illustrates the promise of chemical modification to improve enzyme properties.

A V.4 Metagenomics (Genome mining)

Instead of modification of the existing enzymes to improve the properties (activity or stability), we can find enzymes that already have the characteristic we are aiming for. This alternative strategy is based on an *in silico* approach by sequence similarity analysis, which is called metagenome mining or metagenomics. Metagenomics is a culture-independent technique, which helps to isolate and identify novel enzymes coupled with their catalytic understanding. The exploitation of genome databases is an interesting approach since genome sequence databases are a rich and largely untapped resource for biocatalysts with attractive biocatalytic features and novel chemistries. This approach increases the chances of obtaining the ideal biocatalysts by mining fully sequenced and annotated genomes of laboratory cultivable strains. ^[123] The use of metagenome mining helps us to search through the big pool of genes in the database and find probably more stable and more promiscuous enzymes while avoiding tedious protein engineering and other stabilization methods. The use of metagenomics to identify novel enzymes and biocatalysts for industrial applications is steadily growing but still in its infancy. ^[124] One of the main limitations of this method is the number of annotated available species. The other limitation is the poor transcription and translation by the host enzyme, which in most cases, is *E. coli*. ^[123b] To be more specific the enzymes that have been identified in eukaryotes or the one, which belongs to thermophile bacteria are not easy to be expressed in *E. coli* or other common hosts. This is because the gene sequence is optimized for the microorganism that the enzyme originally belonged to and not to the host. Also, there is always the chance that the host will not be able to express the enzyme of interest. Furthermore, it is quite tedious to acquire the proper condition for the expression of your enzyme of interest in the host microorganism.

Several BVMOs have been discovered using this method. ^[42, 47, 49, 63, 125] However, many sequence-related genes that represent flavoprotein monooxygenase may not catalyze Baeyer-Villiger reactions. Identification of Type I BVMO-specific motif (FxGxxxHxxxWD/P) helped to have a more reliable identification method for BVMO genes. This motif was identified by comparing sequences of characterized BVMOs. ^[29] This allows a feasible survey of all metagenome databases concerning the

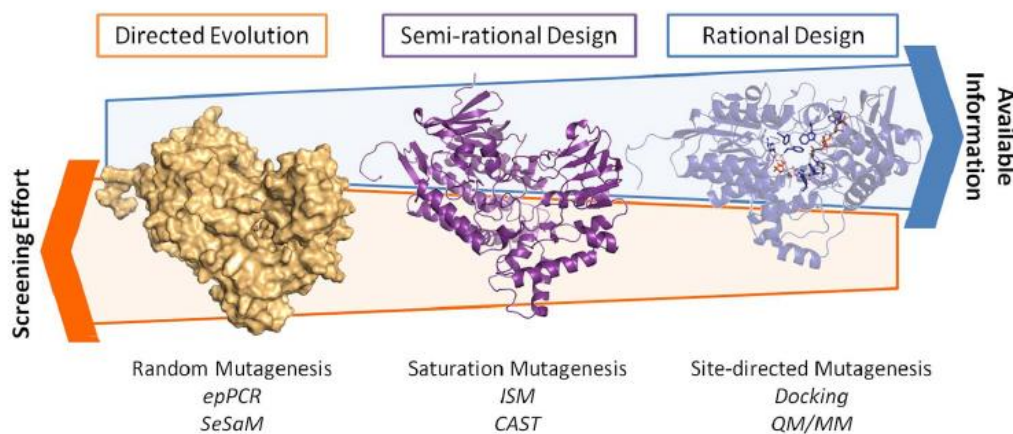
occurrence of Type I BVMOs. Estimations suggest that there is more than 400 novel Type I BVMOs genes present in the sequence database (including the unfinished genomes), which can be a great source to find BVMOs with high stability and even better substrate profile. [18]

This approach was used by many researchers and became a reliable method to identify novel BVMOs with the desired characteristic, especially the more stable enzymes. One of the most stable BVMOs to date, phenylacetone monooxygenase (PAMO) from thermophilic actinomycete *Thermobifida fusca*, was found by this method. [126] Recently, genome mining also guided Mattevi and co-workers to find two other thermostable cyclohexanone monooxygenases (TmCHMO & PockeMO), which were isolated from thermophilic bacteria. [56-57] In another study, Gran Scheuch and coworkers could find two thermostable BVMOs by using genome mining. [60] Other examples are two novel BVMOs, which have also identified and characterized by Mansouri *et al.* and Ceccoli *et al.* with high thermodynamic stability and broad substrate profile, respectively. [58, 63]

A V.5 Protein engineering

Modification of the protein sequence/structure or the synthesis of new proteins to achieve desired functions is called protein engineering. [127] This approach is the most powerful method of regulating macromolecular architecture. [128] There are three main approaches for performing protein engineering studies, which extensively depend on the available knowledge of the enzyme's structure and mechanism (Figure A.14). [129]

Figure A.14. Different protein engineering approaches sorted by screening effort vs. available knowledge.[129]



A V.5.1 Random mutagenesis (Directed evolution)

In the cases when there is not much information about the enzyme of interest, random mutagenesis (directed evolution) ^[130] by methods like error-prone PCR (epPCR) ^[131] is the only and best choice to carry on protein engineering. Nevertheless, it should take into account that this approach is quite tedious and necessitates a massive subsequent screening effort to evaluate the variants and identify the best mutant. ^[129] Spectrophotometric and fluorimetric assays are the most frequently used high-throughput screening methods, which can perform to evaluate the high number of variants. Mentioned high-throughput assays will allow efficient and fast evaluation of a high number of variants. Other methods for identification of promising variants in such extensive libraries are Fluorescence-activated cell sorting (FACS), phage display, and selection methods, such as growth assays. ^[132] Robotic platforms can simplify the screening for such big libraries of mutants using these methods. ^[133] So, in the case of a random mutagenesis approach, besides having an efficient strategy to introduce mutations, developing a high-throughput assay is necessary. The screening strategy also may impose a bias and will affect the ultimate end-point of the evolution campaign (“you get what you screen for”). Moreover, this strategy will always lead to a relative optimum, as based on possible sequence space hitting the absolute optimum is very unlikely.

A V.5.2 Rational design

Comprehensive information about the structure and mechanism and also crystal structures facilitate rational engineering (site-directed mutagenesis) of the protein, which limits the number of mutants that have to be evaluated and leads to a faster evaluation and determination of the best variants with less effort. This approach also helps to focus on the critical regions of the protein and helps to get more influence in the protein properties in most cases.

A V.5.3 Semi-rational design

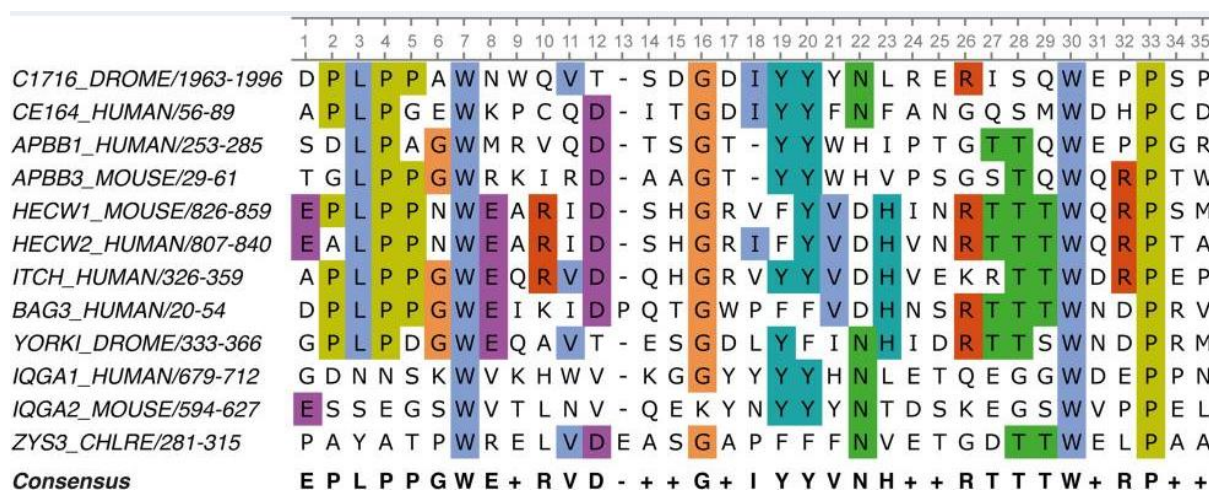
The last and third approach is the semi-rational design, which is a combination of random mutagenesis and rational design. This can be used for enzymes, which we have information from their mechanism and structure, but not enough to be able to introduce specific and effective mutations rationally. Semi-rational or knowledge-based hybrid method is leading to create small-sized libraries of very high quality, which have gained significant momentum. ^[134] In this approach, information from protein function, protein structure, sequence homology, and predictive computational algorithms will be combined to preselect residues for more focused mutagenesis with limited amino acid diversity. This focus will result in dramatically reduced library sizes with a significant increase in functional content.

The combinatorial active-site saturation test (CAST) [135] in combination with iterative saturation mutagenesis (ISM) [136], the protein sequence-activity relationships (ProSAR) strategy [137] or a 3DM database [138] and consensus approach [139] are the promising semi-rational approach for the generation of desired BVMO variants.

A V.5.3.1 Consensus approach

The consensus approach is a popular and successful strategy that classified as the semi-rational protein engineering for improving protein stability, which is based on the use of evolutionary information encapsulated in consensus protein sequences. The approach builds on the assumption that in a specific position, the consensus amino acids will contribute more to the stability of the protein in comparison to the non-conserved amino acids (Figure A.15). [140]

Figure A.15. An example of multiple sequence alignment of different species to reach a consensus sequence. [140]



The main strategy in the consensus approach, as was mentioned before, is the use of evolutionary information encapsulated in homologous protein conserved sequences. Multiple sequence alignments (MSAs) and phylogenetic analyses are standard tools for exploring consensus sequences [139] and ancestral relationships amongst protein homologs. [141] Such sequences and alignments can be acquired from natural sequence databases. [142]

The consensus design approach has been widely successful in improving the stabilities of functional and non-functional proteins, for example increasing thermodynamic stability by 10–32°C [134a, 140, 143] However, only half of the conserved amino acids are associated with improved stability, about ~10%

identified as stability neutral and 40% showed destabilizing effect, leading to challenges and trade-offs during implementation. [139, 143a, 143b, 144] This reflects the still poor understanding of molecular mechanisms of protein stabilization, which often prohibit purely knowledge-based approaches.

All of these three methods proved to be beneficial and successful for stabilizing proteins. However, the degree of success, limitations, and amount of effort vary significantly between the methods, which is directly connected to which method is chosen for your protein of interest. [145] It is also quite critical how properly the method is chosen for engineering. For instance, the specific influence of mutations, mainly when they are not located in critical positions, is not easy to predict, thus choosing the rational design for such a positions faces with a low degree of success and random mutagenesis might lead to more exciting and unpredictable variants and high success rate.

It is pretty hard to say which approach is the best and while all the implemented techniques found stabilizing mutations, it is simply a tradeoff between the amount of screening required, prerequisite knowledge (structure), and whether special modeling methods are needed. [145]

Here is a summary of studies, which conducted different protein engineering approaches to improve the stability of BVMOs (Table A.6), which some are explained in more detail. It is important to mention that except for altering the stability, several studies have been performed to modify other properties of BVMOs like selectivity and change of cofactor specificity as well. These studies could give valuable hints to the mechanism of the BVMOs and the key residues were identified.

Attempting to increase the enzyme's oxidative stability, Opperman and Reetz (2010) introduced point mutations at sulfur-containing residues. To improve the enzyme's tolerance to oxidative stress, they mutated all the methionine and cysteine residues due to their tendency towards oxidation by H₂O₂ following uncoupling reactions. Because oxidative damage promotes at higher temperatures, they aimed to improve the thermostability as well to reduce the oxidative damage. The mutants showed either higher thermostability or oxidative stability, a proof that it is not uncomplicated to increase oxidative stability and thermostability at the same time. They reported that 50% of activity remained after incubation at various temperatures. [80]

Two studies conducted by Van Beek *et al.* (2014) and Schmidt *et al.* (2015) both aimed to increase the thermostability by inducing disulfide bonds. The best mutants displayed an increase in the melting temperature (T_m) of 6°C and 5°C, respectively. [105, 146]

Van Beek *et al.* (2014) used a computational protocol for the prediction of disulfide bonds based on a CHMO_{Acineto} homology model. The mutants showed a higher melting point than the wild-type enzyme.

The disulfide bonds located in the same region of the protein, suggesting that this area is essential for thermostability. Unfortunately, the catalytic efficiency was decreased by 50%.⁶

Schmidt *et al.* (2015) used software for the prediction of disulfide bonds in CHMO_{Acineto}. Activity measurements of the disulfide mutants showed that variant DS3 (T415C-A463C) nearly kept its activity compared to the wild-type enzyme. It showed increased long-term stability and thermostability. The melting temperature increased by 5°C as well.⁶

By combining the most beneficial of these reported mutations, a recent study reported that none of the thermostabilized mutants could outperform the wild-type enzyme in terms of ϵ -caprolactone production.^[147]

A study from Füst *et al.* (2019) used the method FRESCO, which utilizes two independent algorithms to calculate the difference in folding free energy ($\Delta\Delta G^{\text{Fold}}$) between a wild-type structure and a single point mutant. The FRESCO protocol foresees the exclusion of residues in a 5 Å radius from the active site. The core algorithms restricted to the proteinogenic component of the enzyme; the cofactor-dependency represents an additional limitation. Therefore, approximately 19% of the enzyme's residues had to be excluded from the calculations.

Following an *in silico* screening, the selected mutants are expressed and tested for improved stability, and successful hits combined to a final stabilized mutant. This study conducted using RhCHMO, and it could successfully increase the T_m by up to 15°C. This subsequent screening confirmed the reliability of the FRESCO predictions, as nearly half of the tested mutations exhibited a stabilizing effect. In a previous study, a combinatorial library of wild type and mutant residues was created by golden gate shuffling using synthetic genes.^[148] At first, a fully mutated plasmid created, and then a mixed wild-type and mutated PCR fragments with primers remote from the mutated site were generated. After sequencing, successful randomization found. Previous FRESCO projects have demonstrated a T_m increase accompanied by long-term stability, which retained longer. Measurements from this study showed a strong correlation between T_m and stability over time, as the enzyme's half-life of activity at 30 °C was improved 33-fold from 22 min in the wild-type to 12.3 h.^[86]

Table A.6. BVMOs stability values and comparison with few variants.^[149]

Enzyme	Thermodynamic stability (T _m)	Kinetic stability (t _{1/2})	T ₅₀
TmCHMO	48 ^a		
Cm-BVMO	56 ^a		
PockeMO	47 ^a		
hFMO5	47 ^a		
PAMO	60.5-61 ^a	24 h ^e	
STMO	39 ^a		
RhCHMO	36 ^a	22min ^c	
CHMO _{Acineto}	37-39 ^a / 31.6 ^b	4min ^c / 1.82.h ^d /57h ^f /17h ^d /11h ^c /17h ^e	40.5/ 32.5
PAMO P3			
PASTMO	49 ^a		
PACHMO	55 ^a		
PAMEMO1	51 ^a		
PAMO B-PV	57.5 ^a		
PAMO CBA-GPTQ	58.5 ^a		
RhCHMO M8b	48.8 ^a	12.3h ^c	
CHMO _{Acineto} Mut15			47.3
CHMO _{Acineto} Mut16			43.4
CHMO _{Acineto} A255C-A293C	40.5 ^a		
CHMO _{Acineto} A325C-L483C	40.5 ^a		
CHMO _{Acineto} L323C-A325C	44 ^a	45min ^c	

Introduction

CHMO _{Acineto} T415C-A463C	DS3	36.5 ^a /36.4 ^b	5.2h ^d	38
CHMO _{Acineto} T415C		37.5 ^b	14.6h ^d	35

^a Determined by Thermo-FAD method.
^b Determined by CD-spectroscopy.
^c Measured after incubation at 30 °C.
^d Measured after incubation at 25 °C.
^e Measured after incubation at 52 °C.
^f Measured after incubation at 20 °C

There are some challenges in the protein engineering of BVMOs that hamper the generation of active enzyme mutants with the desired property. First of all, BVMOs show complex domain movements during the catalytic cycle, which makes it difficult to predict which residues are precisely interacting with the substrate and co-factor. Second of all, the mutation in the active site of BVMOs can lead to an increase in the uncoupling reaction rate (the generation of H₂O₂ instead of product formation) as an undesired side reaction. One other major problem is the fact that many stabilization efforts lead to lower activity of stabilized variants, which is troublesome. This results at the end in a variant with high stability but low catalytic efficiency that is not favorable for industrial purposes. This is due to the fact that many protein engineering efforts to improve the stability are focusing on rigidifying the enzyme, which can result to lower flexibility and leads to lower activity. So protein engineering is a challenging approach to improve the stability of BVMOs but can provide valuable data about enzyme critical residues and also information about the role of these residues in enzyme function, which can help to understand this family of enzymes more better.

B Aim of this study

BVMOs are valuable biocatalyst for the production of highly valued lactone and other chiral building blocks, but this potential can not be used for large scale production in industry. This is mainly because of low stability of these enzymes. This study aimed to tackle this problem by protein engineering approach and also we attempted to discover a thermostable BVMO. Last but not least, we tried to get a better understanding of these enzymes selectivity and have modified the selectivity of TmCHMO for the production of industrially valuable compounds (Figure B.1).

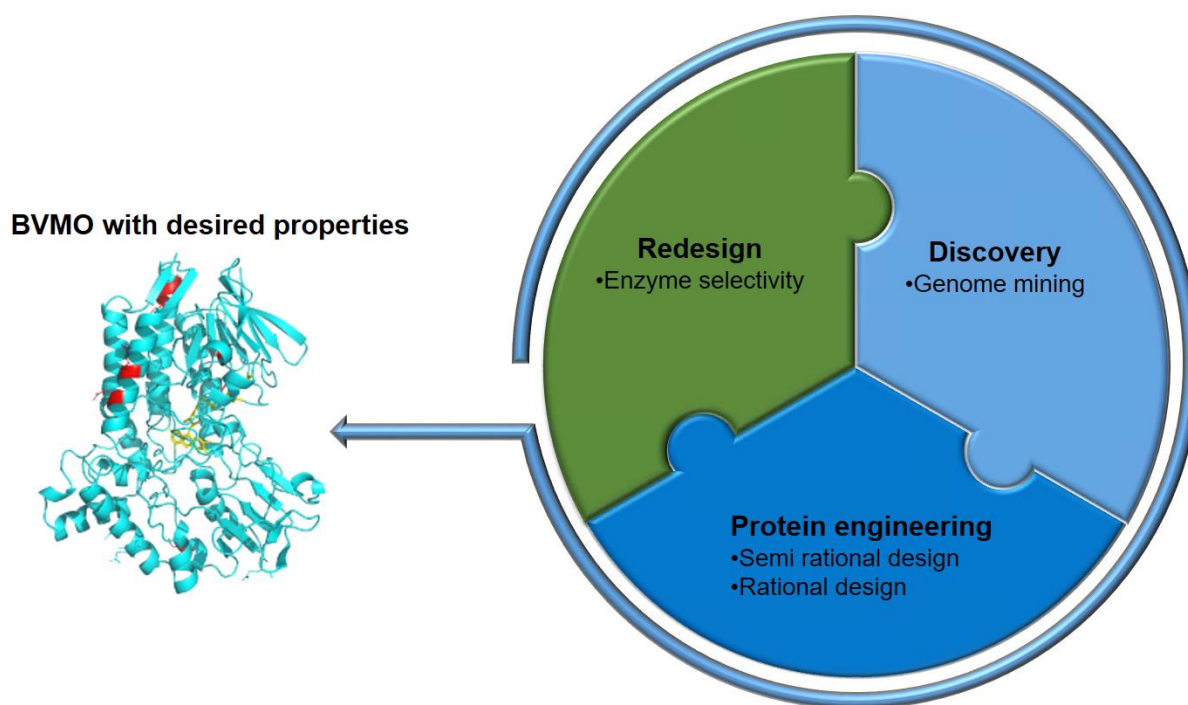


Figure B.1. The attempts considered in this thesis to modify and improve the enzyme properties.

Chapter C I.1 focuses on enzyme discovery based on genome mining. This was aimed to overcome the enzyme stability not by modification of enzyme itself but by identifying a novel BVMO employing the TmCHMO sequence as a search template. With such an improved enzyme at hand, concomitantly displaying a broad substrate profile, further manipulations to affect the catalytic performance were expected to process more facile. To improve the probability of finding a stable enzyme we have focused on the enzyme originated from thermophile bacterium. The enzyme

sequence with the highest similarity to TmCHMO was chosen. The chosen sequence originated from a thermophile bacterium called *Amycolatopsis thermoflava*. To evaluate the characteristic of this novel enzyme, it was cloned in pET22b+ vector and expressed in *E.coli* BL21DE3. The purified enzyme was characterized and its properties were determined. The enzyme kinetics and substrate scope were also determined.

Chapter C I.2 was not following the main objective of the project and was aimed to get a better understanding of TmCHMO selectivity. This was performed by applying protein engineering (Iterative saturation mutagenesis) and variants were studied for their enantioselectivity by using gas chromatography. This study provided information about several possible hits regarding selectivity of BVMO enzymes and can be used as the basis for further research in the field of BVMOs selectivity.

Chapter C I.3 is the main chapter and is focusing on enzyme stability improvement by protein engineering. The semirational and rational design were used as the prediction tool for choosing the mutation sites. The consensus approach was used to predict the majority of mutations, especially in the first and second generation libraries. 1st_{mut-1} (G14A) was the only rationally predicted mutation. This mutation was predicted based on our previous finding in the enzyme-FAD binding affinity properties and was aimed to improve this characteristic. All in all three generations of mutants were created with 17 variants. Activity, total turnover number (catalytic efficiency), thermodynamic, and kinetic stability were measured for all variants. In the last generation library of mutants, the best of first and second generation were combined with literature known hits. The best variant was determined and was studied more comprehensively. The effect of different pH and cosolvents was determined on this variant to evaluate the effect of mutations on cosolvent tolerance and pH tolerance. Molecular dynamic simulation and 3D structure analysis were used to understand the mechanism that resulted in the stabilization of this variant. Finally, we determined the effect of mutations on enzyme selectivity to get a better understanding of mutations effect on enzyme characteristics.

C Results and Discussion

C I Enzyme discovery by genome mining

C I.1 Genome mining

Nature is the expert protein engineer, and it began its bioengineering trial billions of years ago, which can obviously be detected in the sequences of proteins. ^[150] During the evolution, nature was proactively engineering the existing enzymes and create better variants, which can be considered as the optimized biocatalysts. Historically, strategies to find a novel enzyme were based on in vivo selection on collections of strains or individual and for a decade, on metagenomes. Briefly, the genetic material extracted from the mixed microbial community, then size-selected inserts cloned into expression vectors. ^[151] The next step was to screen for enzymatic activities, which was generally performed in situ and based on an indicator medium. To identify the gene of interest, positive clones will be sequenced. This approach is very effective and useful but faces several limitations. For example, it is restricted to enzymes, which have a generic assay for their activity (e.g., lipases, amylases, and oxidases). ^[152] Besides, it shows low sensitivity since the screening was performed without overexpression of the protein. Moreover, the isolation of enzymes is usually time-consuming, and it is the case that most of the time, the protein loses much of its catalytic activity. Because of this limitation in using traditional screening techniques, it is estimated that we missed up to 99% of existing microbial resources. ^[153]

In contrast to traditional screening techniques, genome sequence information enables direct cloning of the desired genes using the PCR and, therefore, an efficient expression in a host strain, even if PCR errors and/or expression drawbacks may be encountered. Next-generation sequencing (NGS) did change the game after emerging by simplifying the sequencing of the non-culturable microbes and mixed microbial population. NGS helped the scientists to extract an incredible amount of protein and gene sequences, which deposited in the databases, and it is still increasing. Till the end of 2016, it was estimated that more than 67,000,000 protein sequences from 509,000 species have been deposited in the TrEMBL database, <http://www.uniprot.org/uniprot/> TrEMBL source. The valuable information in the sequences makes the DNA and protein databases gold mines for discovering novel

enzymes. Moreover, the sequencing of metagenomes from various environments can be a huge reservoir of novel enzymes, which can offer specific catalytic activity required for a defined process. [154] It is necessary to rationalize the exploration of the vast amount of available genomic resources to decrease the experimental effort, which can be possible by computational methods that reveal the sequence/function relationships of proteins. The ability of a biocatalyst to transform a range of different unnatural substrates expands the conversion possibilities. This characteristic, inherent to many enzymes and outlined as a potential for enzyme discovery for dozens of years, increases the capability of enzymes and the biotransformation performed by living cells. [155] The search for a novel enzyme is then mostly based on the substrate promiscuity since unnatural substrates are often the main target in organic synthesis. From a genome sequence, a vast amount of information is available. The function can be reflected in the name of an enzyme to conserved patterns/ signatures and also the (predicted) structure; all these features can be used to search for new biocatalysts. [156] Often, to analyze the huge amount of information, the data will be processed by using computational approaches and comprehensive, organized databases. One of the public databases is BRENDA (BRaunschweig ENzyme DAtabase; www.brenda-enzymes.org), which contains extensive details on a full collection of known enzyme substrates, which is providing indications about the biocatalytic potential of the biocatalyst. [157] All this information provides a great potential to identify novel enzymes for biotechnological applications like discovery of novel enzymes with new characteristic, enhanced or inverted (chemo-, regio-, stereo-) selectivity, improved stability (temperature, solvent, etc.), altered pH- or temperature profile, expanded substrate profile, and improved catalytic efficiency.

Here is a short overview of different genome mining approaches are presented that can be useful for enzyme discovery.

C I.1.1 Text-based Searches Using Enzyme Name

DNA sequences that obtained from the sequencing of a single gene, entire genome, and microorganism consortia are available in databases like UniProtKB and NCBI. It is possible to predict the function of proteins, which can be performed mainly by multiple sequence comparisons with already known enzymes.

It has long been one of the easiest ways to find new enzymes just by in silico screening of databases for a specific enzyme using the name of the enzymatic function as the query. Though, this method suffers from two main downsides: (1) this approach is limited by the lack of novelty in the features of the newly identified enzymes. [158] The huge amount of data that is produced by the NGS generates a growing number of sequences with no reliable annotation. (2) the experimentally established functions

only concern a small fraction of the enzymes, since enzyme function mainly concluded from a few numbers of characterized enzymes. About 40% of the sequences, which are stored in the UniProtKB can be considered an “uncharacterized protein”, and at least 20% of allocated functions are estimated to be wrong. ^[158] So, this approach does not generate the most out of the data potential. However, it has successfully been used in many projects regarding various enzyme families, including cytochromes P450, nitrilases, halogenases, and glycosylhydrolases. ^[159]

C I.1.2 Sequence-driven Approaches

This approach can lead to the discovery of novel enzymes by using at least one described protein or nucleotide sequence. Sequence driven methods indeed belong to the most common ones to discover new biocatalysts.

This method is generally based on the analysis of the primary sequence of the enzyme as a whole or in specific portions. By using a characterized enzyme and its related gene that will be used as a starting point, new enzymes, which are performing the same or similar reactions, can be identified. The sequence encoding the known enzyme can be used as a template to aim for unsequenced proteins by designing primers for their amplification or to aim for already sequenced genes by pairwise protein sequence alignment tools. These methods explore the steadily expanding protein-sequence database and open the path for the efficient discovery of novel enzymes.

C I.1.2.1 Signature/Key Motif-based Strategy

Rather than searching for new enzymes using the primary sequence of a protein as a whole, one can focus on specific portions of this sequence. Proteins mainly classified into families based on the presence of crucial conserved sequence or domain features. These allocated motifs can be exploited using various computational approaches that typically use multiple sequence alignment of proteins that are sharing a set of characteristics as a starting point. Specific sequence motifs can allow the identification of the targeted family from the vast number of other sequences among a superfamily. Therefore, protein motifs can be very useful for the discovery of novel enzymes. Few studies employed this approach and proved that this could be a convenient approach for enzyme discovery. Fraaije *et al.* could identify a novel BVMO (PAMO, a thermostable monooxygenase from *Thermobifida fusca*) using the protein sequence motif [FXGXXXHXXXW(DP)] which described earlier by the same group. ^[29, 42] What authors pointed out also in this study was the difficulty of predicting the enzyme substrate profile based on its sequence; indeed, PAMO, despite its high sequence identity (53%) with steroid monooxygenase (STMO), has no activity towards progesterone. ^[126] Recently, Wetzl and colleagues used the Hidden Markov Models (HMM) of the C-terminal domain of six known IREDs to find novel

IREDs of bacterial origin. After collecting the enzyme sequences that are matching this HMM hypothesized to be responsible for the catalytic activity, enzymes representative of the IRED sequence space were selected and tested. Interestingly, relationships between substrate structure, stereochemistry, and clustering have observed.^[160] Two very recently identified IRED-specific motifs, GLGxMGx5[ATS]x4Gx4[VIL]WNR[TS]x2[KR] the cofactor binding motif and Gx[DE]x[GDA]x[APS]x3^[161]x[ASL]x[LMVIAG] active site motif, should help to discover many more IREDs.^[162] The first discovery of Fe-type nitrile hydratase also achieved using a conserved motif as a probe located in alpha-subunit. This motif (KNVIVCSLCSTAWPILGLPPTWYKSFEYRARVVREPR) is also containing the iron-binding motif (CSLCSC), which identified after sequence alignment of all characterized Fe-type NHases. The nitrile hydratase from *Pseudomonas putida* F1 showed efficient catalytic properties on small aliphatic nitriles but also on some aromatic nitriles.^[163] In a comprehensive survey, which was performed by Maimanacos *et al.*, they could identify sequence patterns in AMDases that allowed them to discover 58 new AMDases from genomes and metagenomes.^[164] Another interesting example is the carboxylic acid reductase (CAR) from *Mycobacterium marinum* that discovered by Akhtar *et al.* This enzyme was selected because it contains three consensus motifs characteristic of a previously characterized CAR enzyme: (i) ATP domain, (ii) phosphopantetheine attachments site (LGGXSXXA) and (iii) Rossmann fold for NADPH binding. This enzyme was found to convert a wide range of aliphatic fatty acids (C6–C18) into corresponding aldehydes. This feature is making it a useful catalyst for the synthesis of fatty acid-derived chemical commodities.^[165] It can be expected that in the next few years, this method will substantially proliferate by relying on an increased number of enzyme motifs associated with the enzymatic activity (co-factor binding sites, binding residues, and catalytic sites). This is thanks to the growing number of experimentally validated sequences, which should lead to the discovery of new motifs/signatures inside enzyme families.

Inspired by the work of Romero *et al.*, a study to find a novel BVMO with high stability, high activity, and broad substrate acceptance started. This study based on the sequence-similarity genome mining approach. The big pool of genome sequences blasted by using the TmCHMO sequence. Among BVMOs sequences found in the NCBI genome database, a new putative BVMO sequence that belongs to the thermophilic organism *Amycolatopsis thermoflava* selected. This organism isolated from a heat-treated soil^[166], which increases the chance to have a thermostable BVMO.

C I.2 Blasting and phylogenetic tree

First, the TmCHMO sequence used as the search query for blasting in the NCBI databank. The reason why this enzyme was used as the starting point originated from this BVMO being among one of the most stable BVMOs to date, and finding a BVMO with high sequence similarity with this enzyme resembles a high chance of a BVMO with similar characteristic. We focused on the thermophilic bacteria since it also increases the chance of expressing the enzymes with higher stability. The most similar sequence to TmCHMO among thermophilic bacteria was a new putative BVMO from *Amycolaptosis thermoflava* (BVMO_{Flava}), which was selected for further experiments (Figure C.1). The sequence similarity between BVMO_{Flava} and TmCHMO was 83 %. The multiple sequence alignment with several different type I BVMOs (**Figure C.2**) confirms that this BVMO belongs to the type I BVMO family as it shows the conserved region belongs to this family.

Figure C.1. Blasting NCBI database using TmCHMO.

Sequences producing significant alignments							Download	Manage Columns	Show	100	?	
select all 100 sequences selected							GenPept	Graphics	Distance tree of results	Multiple alignment		
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession					
<input checked="" type="checkbox"/>	NAD(P)/FAD-dependent oxidoreductase [Thermococcus municipale]	1116	1116	100%	0.0	100.00%	WP_028849141.1					
<input checked="" type="checkbox"/>	NAD(P)/FAD-dependent oxidoreductase [Amycolaptosis thermoflava]	858	858	99%	0.0	82.75%	WP_027929089.1					
<input checked="" type="checkbox"/>	NAD(P)/FAD-dependent oxidoreductase [Amycolaptosis methanolica]	853	853	99%	0.0	82.37%	WP_017982805.1					
<input checked="" type="checkbox"/>	NAD(P)/FAD-dependent oxidoreductase [Pseudonocardia autotrophica]	908	908	98%	0.0	81.90%	WP_051736657.1					
<input checked="" type="checkbox"/>	NAD(P)/FAD-dependent oxidoreductase [Pseudonocardia thermophila]	907	907	99%	0.0	80.89%	WP_073457557.1					
<input checked="" type="checkbox"/>	cyclohexanone monooxygenase [Pseudonocardia sp. 73-21]	912	912	98%	0.0	80.56%	OJY43802.1					
<input checked="" type="checkbox"/>	NAD(P)/FAD-dependent oxidoreductase [Prauserella endophytica]	904	904	100%	0.0	80.22%	WP_137095890.1					
<input checked="" type="checkbox"/>	MULTISPECIES: NAD(P)/FAD-dependent oxidoreductase [Prauserella]	903	903	100%	0.0	80.22%	WP_112272984.1					
<input checked="" type="checkbox"/>	NAD(P)/FAD-dependent oxidoreductase [Pseudonocardia sp. N23]	901	901	99%	0.0	79.55%	WP_098956603.1					

This novel BVMO contained the Type I BVMO family fingerprints (G/AGxWxxxxF/YPG/MxxxD and FxGxxxHxxxWP/D). Multiple sequence alignments showed the presence of both Rossmann-fold motifs (GxGxxG/A) (**Figure C.2**). These motifs are responsible for NADP⁺/NADPH and FAD binding and play an essential role in the catalytic cycle of the BVMO family. The full alignment depicted in appendix E I.6.

Figure C.2. Multiple sequence alignment analysis of BVMO_{Flava} with known BVMOs. 4-

Hydroxyacetophenone monooxygenase (HAPMO)^[35] [*Pseudomonas fluorescens*], cyclododecanone monooxygenase (CDMO)^[36] [*Rhodococcus ruber*], cyclopentanone 1,2-monooxygenase (CPMO)^[167] [*Acidovorax sp. SCN 65-28*], phenylacetone monooxygenase (PAMO)^[126] [*Thermobifida fusca*], cyclohexanone 1,2-monooxygenase (CHMO)^[31] [*Acinetobacter sp. NCIMB9871*], NAD(P)/FAD-dependent oxidoreductase (TmCHMO)^[168] [*Thermocrispum municipale*], NAD(P)/FAD-dependent oxidoreductase (BVMO_{Flava}) [*Amycolatopsis thermoflava*]. (G/AGxWxxxxF/YPG/MxxxD and FxGxxxHxxxWP/D) and (GxGxxG/A) are highlighted.

```

BVMOflava TTRTP-----DVDAIVIGAGFGGIYMLHKLRLNELGL-SVTAFEKGGGVGG
TMCHMO    TTQTP-----DLDAIVIGAGFGGIYMLHKLRLNDLGL-SVRVFEKGGGVGG
CHMOAcineto --QKM-----DFDAIVIGGGFGLLYAVKKLRDELEL-KVQAFDKATDVAG
PAMO      SRRQP-----PEEVDVLVGAGFSGLYALYRLR-ELGR-SVHVIETAGDVGG
CPMO      NSVDD-----TLDVLLIGAGFTGLYQLHHLR-KLGF-KVHLVDAGADVGG
CDMO      TPREP-----KLDHVTFAFIGGGFSGLVTAAARL-ESGVEVRIIDKAGDFGG
HAPMO     TAEEDLRAPRWKDHVASGRDFKVVIGAGESGMIAALRFK-QAGV-PFVIYEKGNDVGG
          . . . * . * : : : . . . : . . . *

BVMOflava TWYFNRYPGAKSDTEGFVYRYSFDKDLLREWNWTRYLEQADVLAYLEHVVERFDLGRDI
TMCHMO     TWYWNKYPGAKSDTEGFVYRYSFDKELLREYDWTTRYLDQPDVLAYLEHVVERYDLARDI
CHMOAcineto TWYWNRYPGALTDTETHLYCYSWDKELLQSLI KKKYVQGPDKRKYLQOVAEKHDLKXSI
PAMO       VWYWNRYPGARCDIESIEYCYSFSEEVLEQWNTERYASQPEILRYINFVADKFDLRSIGI
CPMO       IWHWNCYPGARVDTHCQIYQYSM-PELWGEFNWKELFPNWAQMREYFYFVDKKLELSKDI
CDMO       VWYWNRYPGAMCDTAAVMYMPLEET---GYMPTEKYAHGPEILEHCQRIGKHYDLYDDA
HAPMO      TWRENTYPGCRVDINSFWYSFSFARGI-----WDDCFAPAPQVFAYMQAVAREHGLYEHI
          * * * * . * * : : : : : . *

BVMOflava RLNTEVTGAVFDEESDLWTV---TTATGETTTARYLVNALGLLARSNIPIPIGRDGFAGR
TMCHMO     QLNTEVTDAIFDEETELWRV---TTAGGETLTARFLVTALGLLSRSNIPIPIGRDSFAGR
CHMOAcineto QFNTAVQSAHYNEADALWEV---TTEYGDKYRTARFLITAGLLSAPNLPNIKGINQFKGE
PAMO       TFHTTVTAAAFDEATNTWTV---DTNHGDRI RARYLIMASGQLSVPQLPNFPGLKDFAGN
CPMO       SFNTRVQSAVFDEQRREWTV---RSLGHQPIRAKFVIANLGFASPSPTPKVEGIEKFKGE
CDMO       LFHTEVTDLVWQEHDQRWRI---STNRGDHFTAQFVGMGTGPLHVAQLPGIPGIESFRGK
HAPMO      RFNTEVSDAHWDESTQRWQLLYRDSGQTQVDSNVVVFVAVGQLNRPMPAIPGIETFKGP
          : : * * : : * * : : : : : * . * . * . * *

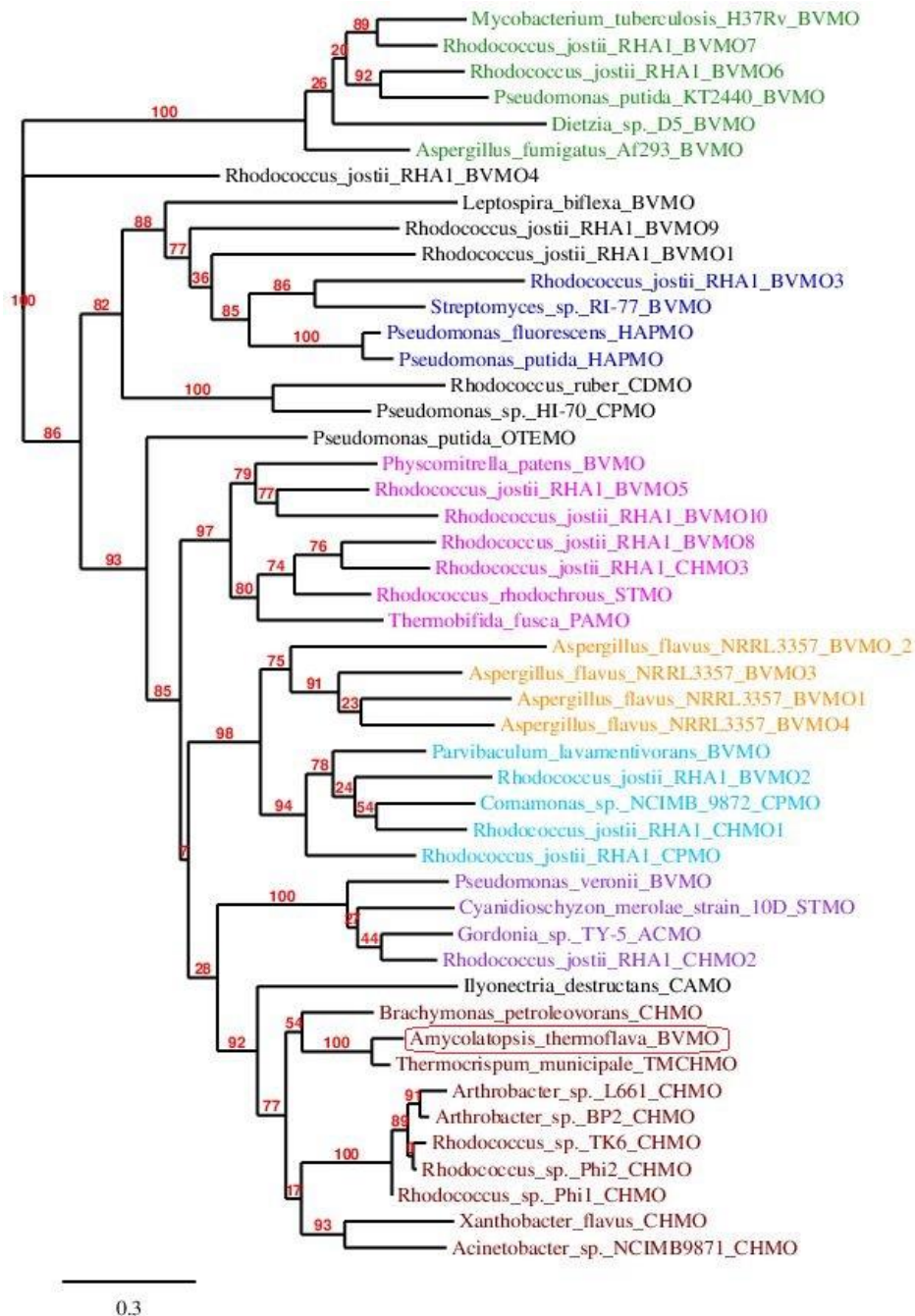
BVMOflava LVHTNAWPDD-----LD-ITGKRVGVIGTGSTGTQFIIAAAKTASHLTVFQSPQ
TMCHMO     LVHTNAWPED-----LD-ITGKRVGVIGTGSTGTQFIVAAAKMAEQLTVFQRTPO
CHMOAcineto LHHTSRWPDD-----VS-FEGKRVGVIGTGSTGVQVITAVAPLAKHLTVFQRSAQ
PAMO       LYHTGNWPHE-----PVD-FSGQRVGVIGTGSSGIQVSPQIAKQAAELFVQRTPH
CPMO       WYHTALWPQE-----GVD-MAGKRVAIGTGSSGVQVAQEAALNAKQVTVFQRTPN
CDMO       SFHTSRWDYDYTGGDALGAPMDKLDKRVAVIGTGATAVQCVPELAKYCRELYVVQRTPS
HAPMO      MFHSAQWDHD-----VD-WSGKRVGVIGTGASATQFIPQLAQTAELKVFARTTN
          * : * : : : . : * * : * * : : . * * . : * . * :

```

To make a better understanding of BVMO_{Flava} evolution and position of BVMO_{Flava} between different groups of BVMOs, we performed a phylogenetic tree analysis. The phylogenetic tree was created by PhyML (Figure C.3) and visualized by TreeDYN (Figure C.3). The mid-point rooted maximum likelihood phylogram shows the diversity of different BVMOs from different groups, I to VII.^[55, 169] As shown in the phylogram (Figure C.3), the TmCHMO sequence is the closest neighbor of the

BVMO_{Flava} sequence with a robust bootstrap statistical support of 100 %. Moreover, the phylogram indicates that BVMO_{Flava} placed in the clade of the CHMO-family, and especially it is a close neighbor of CHMO_{Acineto}; this is of particular relevance when aiming for a biocatalyst with a prospectively broad substrate profile.

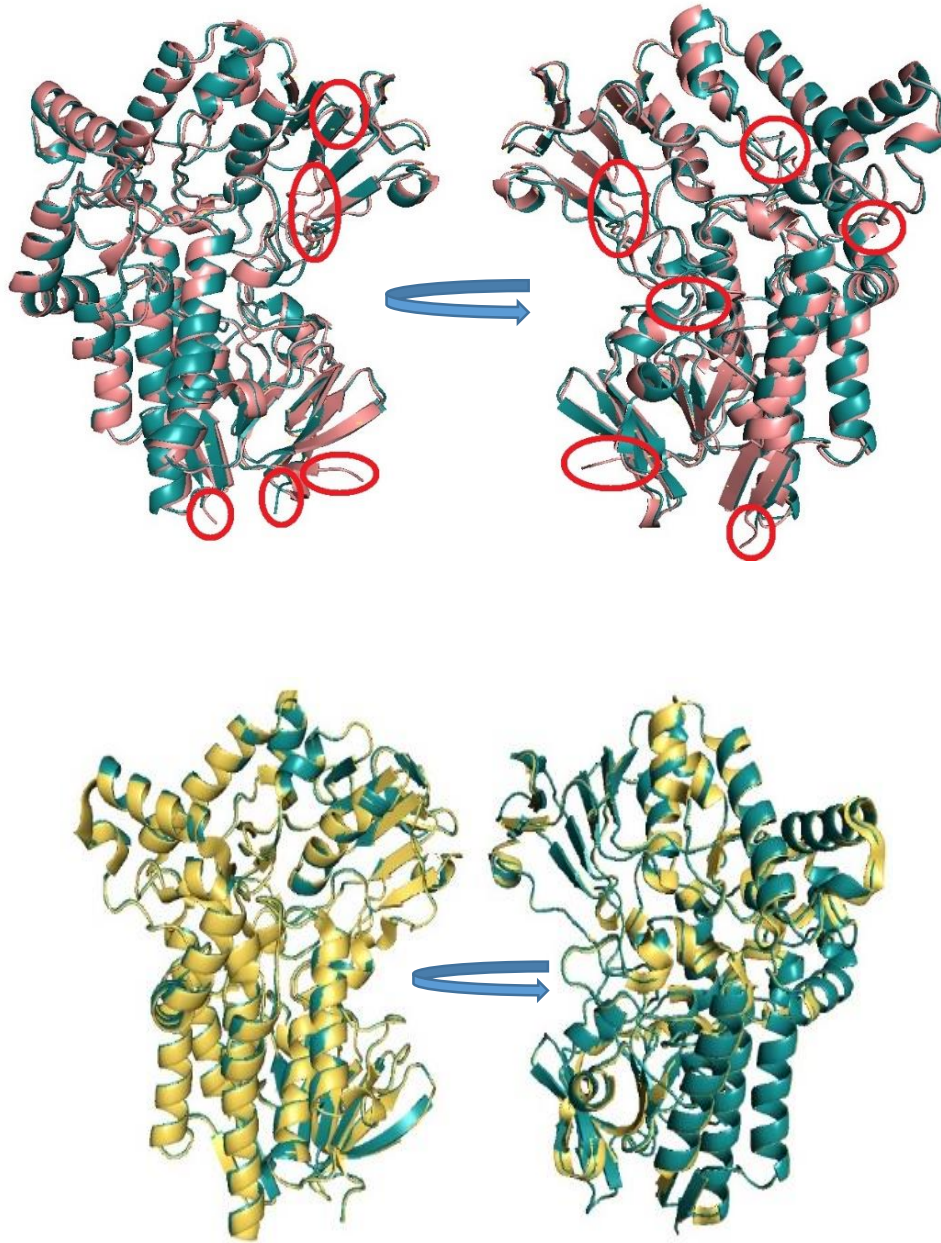
Figure C.3. Phylogenetic tree. The sequences of recombinantly expressed BVMOs have used to make the phylogeny tree, which was constructed by PhyML and visualized by Inkscape. Different BVMOs are color-coded based on their group they belong to. group 1 (light blue) group 2 (pink) group 3 (maroon) group 4 (blue) group 5 (green) group 6 (orange) and group 7 (violet). BVMO_{Flava} located in group 3 (red). The accession code of the sequences can be found in the appendix E I.5.



C I.3 Multiple structure alignment

The structures for the studied enzymes (BVMO_{Flava}, CHMO_{Acineto}, and TmCHMO) are visualized by PyMOL (Figure C.4). As expected, based on the high sequence similarity, TmCHMO (gold) and BVMO_{Flava} (green) showed high similarity in structure as well. This increases the possibility of BVMO_{Flava} for being a thermostable enzyme. The structural comparison of CHMO_{Acineto} (pink) and BVMO_{Flava} (green) also performed. The structure alignment shows several differences, which are pointed out by the red circles. There are several nonoverlapping regions, which can especially be seen on the surface and outer region of the protein and mostly in the loops. This could be a hint that the flexibility of BVMO_{Flava} is hampered, and hence structural stability could be increased. The loops are a critical part of proteins as they are responsible for the flexibility of the enzyme. It is worth to mention that the more flexible the enzyme is, the more unstable it can be. Generally, more tightly structured proteins are more stable.

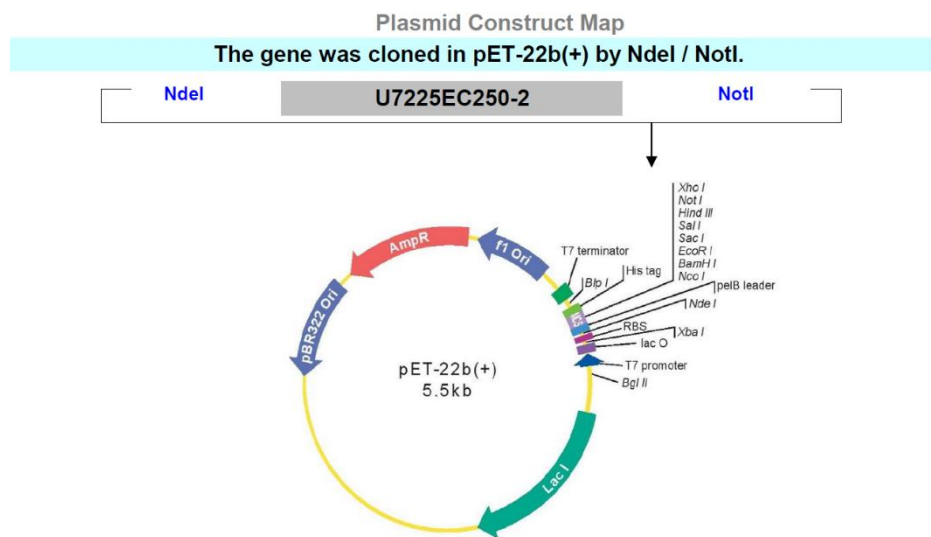
Figure C.4. Multiple structure alignment of BVMOFlava (green), TmCHMO (gold), and CHMO_{Acineto} (pink). The non-matching regions are highlighted (red oval shape) for CHMO_{Acineto} and BVMOFlava.



C I.4 Plasmid construct

With this putative novel type I BVMO sequence in hand, we ordered the synthetic gene already cloned into a pET22b(+) expression vector with a His-Tag on the C-terminus (Figure C.5). pET22b(+) is a vector that encodes a signal sequence for inducible expression of proteins in the periplasm, which facilitates the expression of the soluble enzyme. NdeI and NotI were used as the cloning site for the insertion of the insert to the plasmid backbone (Figure C.5).

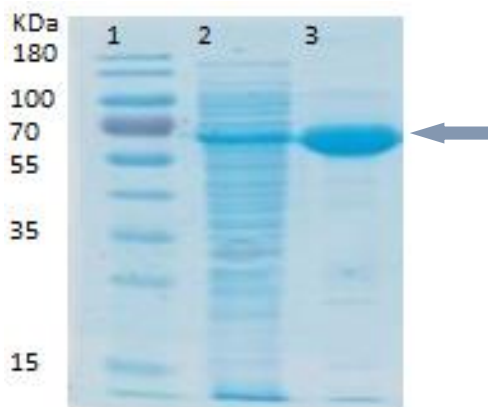
Figure C.5. Plasmid construct map pET-22b(+).



C I.5 Enzyme expression and purification

E. coli BL21 DE(3) was used as the expression host and was transformed by using a heat shock protocol (D I.9). Isopropyl- β -D-thiogalactopyranoside (IPTG, 50 μ M) was used as the inducing agent for protein expression. Expression was conducted by adding IPTG at 20 $^{\circ}$ C for 20-22 h. The expression was confirmed by using SDS-PAGE analysis (Figure C.6). The molecular weight of this BVMO was predicted to be 59.5 kDa by using https://www.bioinformatics.org/sms/prot_mw.html. The concentrated purified and cell-free extract fractions of the enzyme were running into the SDS-PAGE, and a band of the purified enzyme found at around 60 kDa that belongs to the new BVMO_{Flava}. Purification performed by standard His-Trap affinity column explained in experimental part D I.14.

Figure C.6. SDS-PAGE result for the expression of BVMO_{Flava}: lane (1) belongs to standard protein marker, lane (2): cell-free extract solution, and lane (3) displayed concentrated and purified enzyme.



C I.6 Characterization

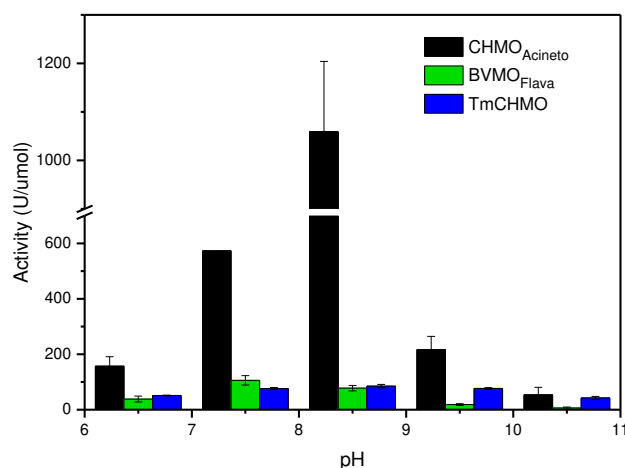
After we could successfully performed protein expression and purification, the characterization of this novel BVMO was carried out. Characterization of an enzyme means investigating different physical and chemical properties (characteristics) of an enzyme. It includes the use of a series of different laboratory procedures like the determination of the effect of changes in pH on enzyme activity, the effect of temperature on enzyme activity and also how susceptible an enzyme is to different solvents. Since the natural substrate for this enzyme was unknown, first, we had to find a suitable substrate. Based on the sequence similarity to CHMO_{Acineto} and TmCHMO and also phylogeny analysis, the assumption was that this novel BVMO might have a comparable substrate profile to these two BVMOs, so cyclohexanone was used as a model compound. Indeed, GC conversion showed BVMO_{Flava} fully converted cyclohexanone to the corresponding ϵ -caprolactone.

C I.6.1 Optimum pH for activity

As mentioned before, enzymes are affected by changes in pH. The most suitable pH value, which is the point where the enzyme is showing the highest activity, is known as the optimum pH. Extremely low or high pH generally results in total loss of activity for most enzymes. Determination of pH optimum for enzymes is vital as an enzyme will work most efficiently at or near its optimum pH. To evaluate the catalytic efficiency of this enzyme at different pH, the pH optimum was determined, and in parallel, the same experiment was performed for CHMO_{Acineto} and TmCHMO as the reference BVMOs. The experiment to

evaluate the optimum pH for the activity measured at pH values ranging from 6.5 to 10.5 with an interval of 1 (Figure C.7). The highest activity for BVMO_{Flava} was at pH 7.5, while CHMO_{Acineto} has its optimum at 8.5, and TmCHMO is equally active from 7.5-9.5 (Figure C.7). It is worth to mention that at higher pH values (10.5), TmCHMO outperforms CHMO_{Acineto} and BVMO_{Flava} by maintaining 50% of its initial activity while BVMO_{Flava} was almost inactive and CHMO_{Acineto} showed less than 10% initial activity. This is explained by the fact that TmCHMO is generally a stable variant.

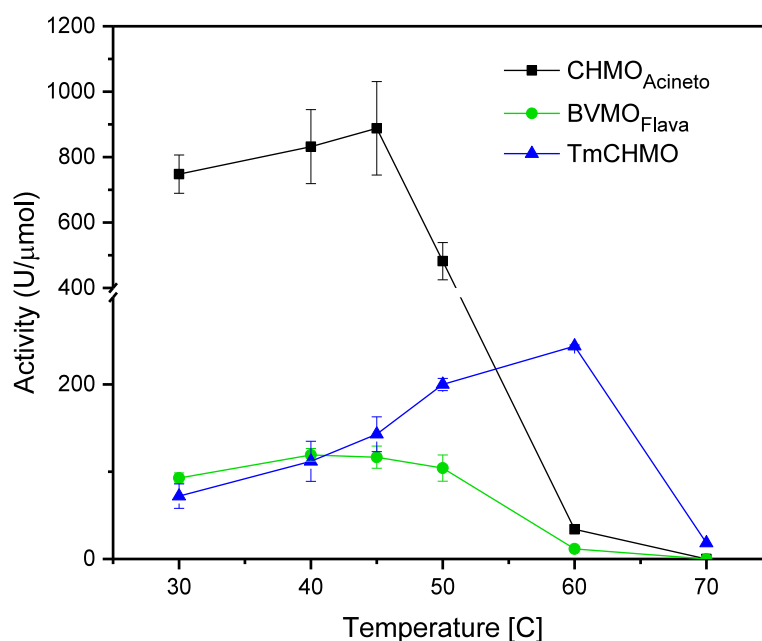
Figure C.7. Effect of pH on the activity at 30 °C in 50 mm Tris·HCl+10 µm FAD, 0.5 mm cyclohexanone, and 100 µm NADPH.



C I.6.2 The optimum temperature for activity

There is a specific temperature at which an enzyme's catalytic activity is at its highest, which is called optimum temperature. The optimum temperature was determined to investigate at which temperature the highest catalytic activity can be achieved and also investigate the applicability of the enzyme in different temperatures (Figure C.8). So we determined the temperature optimum for all three enzymes, which was 45 °C for both CHMO_{Acineto} and BVMO_{Flava}, while TmCHMO showed the highest activity at 60 °C (Figure C.8). This finding was actually in contrast to our expectations since we expected that BVMO_{Flava} is going to show higher optimum temperature as the sequence of BVMO_{Flava} originated from a thermophilic organism.

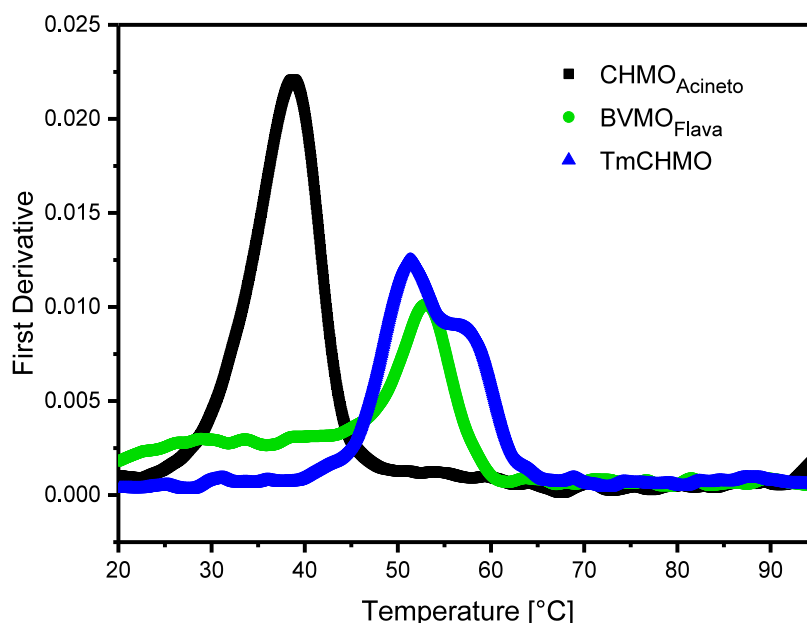
Figure C.8. Activity measurements at various temperatures from 30 to 70 °C (same conditions as those used for the pH study).



C I.6.3 Thermodynamic stability (T_m)

Thermodynamic stability was the next property of these three enzymes, which was investigated. Melting temperature (T_m) was determined as the representative of thermodynamic stability. By comparison of the thermodynamic stability, a different picture was observed (Figure C.9). BVMO_{Flava} showed the highest melting temperature (53.1 ± 0.2 °C) while CHMO_{Acineto} and TmCHMO showed a T_m of 38.5 ± 0.1 °C and 52.1 ± 0.6 °C, respectively (Figure C.9). This finding is well in agreement with the fact that the origin of the sequence based on the thermostable TmCHMO. It is also worth to mention that TmCHMO showed a second transition midpoint, which probably indicates a second unfolding and deactivation process with two active native states. It can be observe that when the temperature exceeds the second transition midpoint, TmCHMO goes into the unfolded and deactivated state.

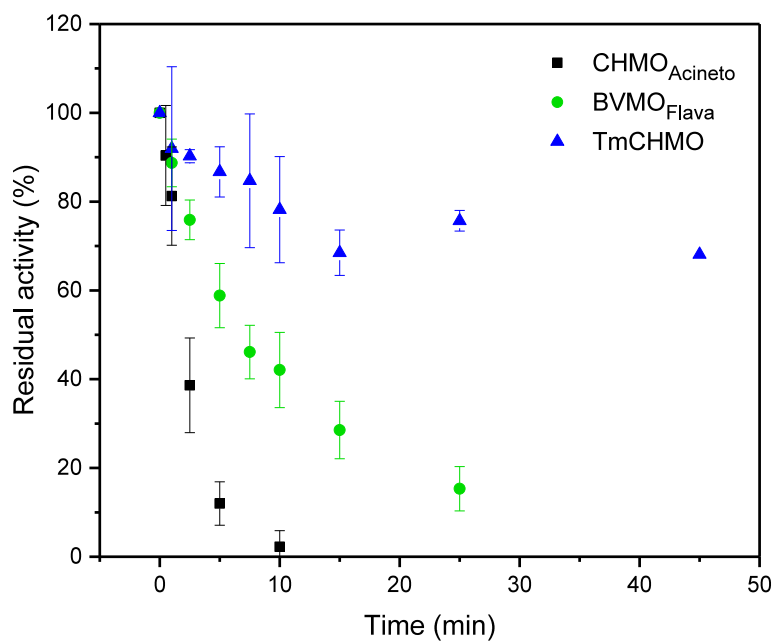
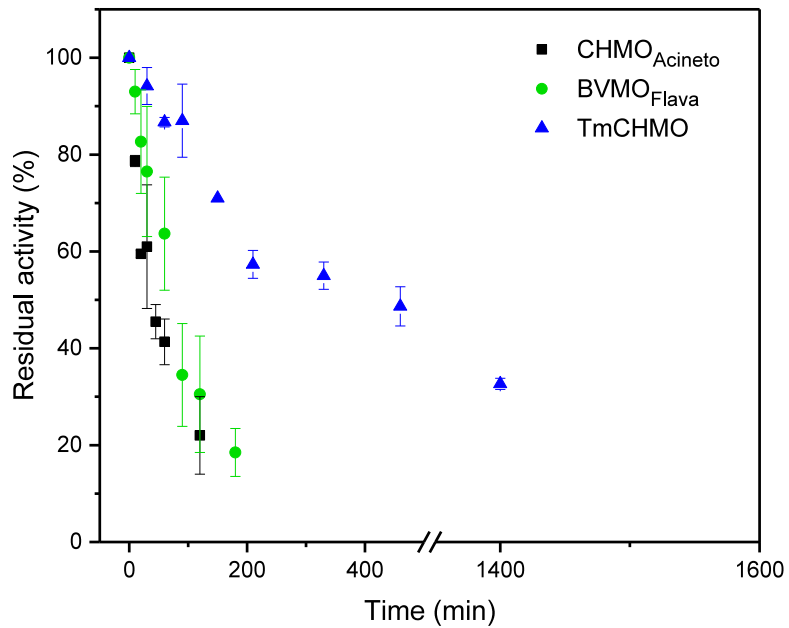
Figure C.9. Melting temperature determination performed employing nano differential scanning fluorimetry (nanoDSF): 50 mM Tris·HCl, 10 μ M FAD, 2 mg mL⁻¹ enzyme.

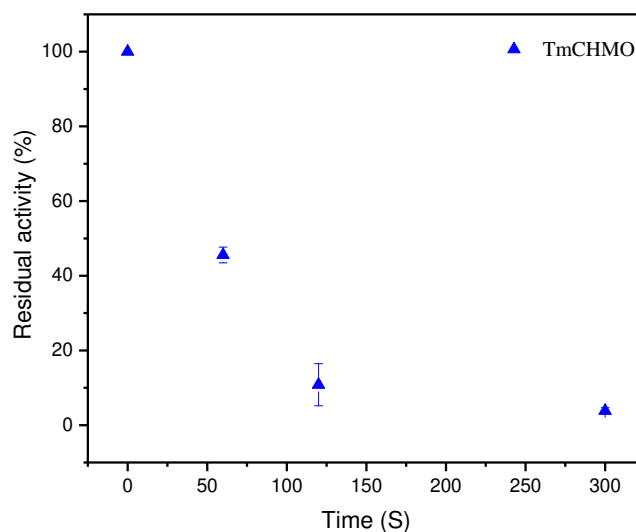


C I.6.4 Kinetic stability (Half-life)

In parallel to thermodynamic stability, it was also necessary to determine the kinetic stability as this value gives more information about the enzyme catalytic stability during the reaction condition. So we investigated the kinetic stability (half-life = $t_{1/2}$) of all three BVMOs at 30 °C, 40 °C and 60 °C (for detailed reaction conditions see D I.16). At 30 °C, BVMO_{Flava} is almost 2-fold (73 ± 10 min) more thermostable than CHMO_{Acineto} with a half-life of 46 ± 6 min, but this enzyme is 7.5-fold less stable than TmCHMO (549 ± 51 min) (Figure C.10) which was unexpected as in the case of thermodynamic stability BVMO_{Flava} and TmCHMO showed almost similar melting temperature. A similar picture was observed when the half-life measured at 40 °C. CHMO_{Acineto} showed a $t_{1/2}$ for 2.02 ± 0.45 min, while BVMO_{Flava} is three folds more stable with a $t_{1/2}$ of 6.00 ± 0.80 min. At 40 °C, TmCHMO still shows the highest stability by showing more than 60 % of its initial activity after 1 hour. The $t_{1/2}$ of TmCHMO was also investigated at 60 °C and showed a half-life of 0.8 ± 0.2 min (Figure C.10). These results indicate and also confirmed our previous finding that thermodynamic stability does not necessarily correlate to kinetic stability within BVMO biocatalysts.^[110]

Figure C.10. A) Half-life measurements: incubation at 30 °C, 10 μm enzyme, 50 mm Tris·HCl, 10 μm FAD, pH 7.5. B) Determination of half-life at 40 °C. C) Determination of half-life for TmCHMO at 60 °C.

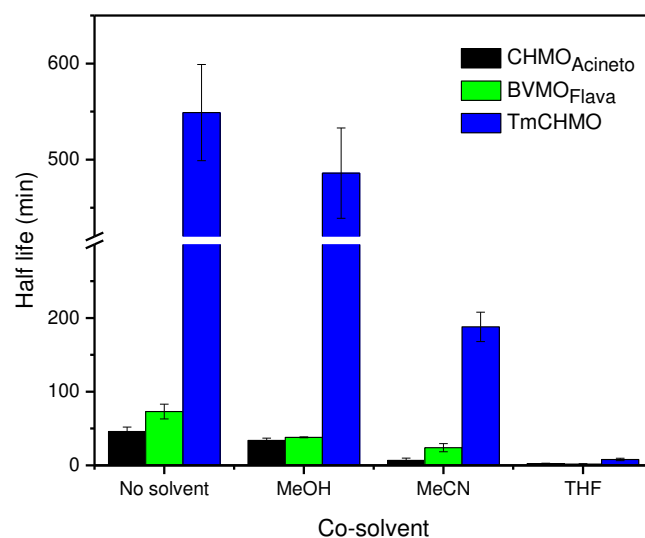




C I.6.5 Solvent tolerance

To investigate the efficiency of this novel BVMO for the industrial application, it was also essential to determine its stability in the presence of different solvent as they vastly used in biotransformation, especially for better substrate solubilization. So we investigated the stability of all three enzymes in the presence of three different organic solvents with the final concentration of 5 % v/v (Figure C.11). All three enzymes showed decent stability in the presence of MeOH; however, the presence of 5 % v/v ACN affected CHMO_{Acineto} the most and resulted in almost complete loss of activity. The most destructive organic solvent for all three enzymes was THF that led immediately to the complete deactivation of all three biocatalysts.

Figure C.11 Determination of half-life in the presence of 5 % cosolvent (same conditions as those used for half-life measurements).



C I.6.6 Kinetic measurement

For practical purposes, the determination of enzyme turnover number and also the K_m value is necessary. Turnover number or k_{cat} is the number of times each enzyme site converts the substrate to product per unit time. K_m is the concentration of the substrate, which permits the enzyme to achieve half V_{max} . An enzyme with a high K_m has a low affinity for its substrate and requires a higher concentration of the substrate to achieve V_{max} . The rate of reaction when the enzyme saturated with the substrate is the maximum rate of reaction, which is called V_{max} . In order to measure the catalytic constants (K_m and k_{cat}), reactions started by mixing the enzyme solution ($0.05 \mu\text{M}$) with pre-warmed solutions of Tris·HCl 50 mM (30°C) containing NADPH ($100 \mu\text{M}$) and variable concentrations of cyclohexanone as it can observe in Figure C.12. [110] Kinetic values have been studied for all three enzymes. The K_m value for cyclohexanone and TmCHMO it was below $<1 \mu\text{M}^{[56]}$ (Table C.1), BVMO_{Flava} (Table C.1, Figure C.12) was $0.53 \pm 0.1 \mu\text{M}$, and for CHMO_{Acineto} was one order of magnitude higher ($6.74 \pm 2 \mu\text{M}$) (Figure C.12, Table C.1). Next, we determined k_{cat} , which resulted in the highest value for CHMO_{Acineto} $15 \pm 1.3 \text{ s}^{-1}$ in comparison to TmCHMO 2.0 s^{-1} and BVMO_{Flava} $1.5 \pm 0.1 \text{ s}^{-1}$ (Table A.1). The ratio k_{cat}/K_m regularly referred to as the "specificity constant" that is a beneficial index for comparing the relative rates of an enzyme acting on alternative, competing substrates. Though an

alternative explanation, "catalytic efficiency", is often used, and on occasions misused, to relate the reactivity of two enzymes acting on the same substrate. [170]

Figure C.12. K_m and k_{cat} measurement A) CHMO_{Acineto} B) BVMO_{Flava}.

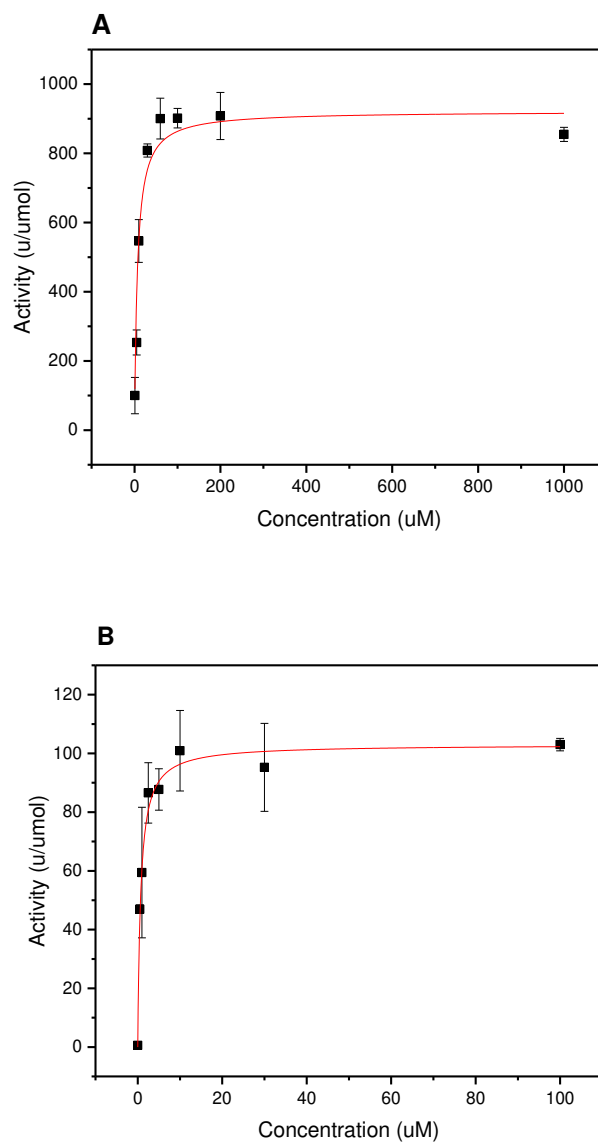


Table C.1. K_m and k_{cat} value measurement.

	K_m	StDEv	k_{cat}	StDEv	k_{cat}/K_m
Cyclohexanone	[μM]		[s^{-1}]		[$\text{mM}^{-1} \text{s}^{-1}$]
CHMO _{Acineto}	6.74	± 2	15.0	± 1.3	2220
BVMO _{Flava}	0.53	± 0.1	1.5	± 0.1	2932
TmCHMO ^[56]	<1	-	2	-	>2000

C I.6.7 Substrate profile

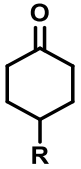
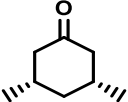
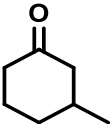

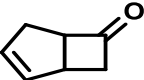
The next characteristic of the enzyme, which is highly essential for further application of the enzyme in the industry, is the substrate acceptance of the enzyme. Enzymes with a high substrate profile will be more attractive for industrial applications as they can produce different products and convert unnatural substrates. To determine the substrate acceptance of the novel BVMO_{Flava}, we elucidated its substrate profile for potential industrial applications and compared it to literature data for TmCHMO and CHMO_{Acineto}. We used whole-cell biotransformations under non-growing conditions and analyzed the conversion and enantiomeric excess by chiral gas chromatography. Positive control experiments were performed with cyclohexanone as the known, accepted substrate. We did perform Baeyer-Villiger oxidation of differently substituted cyclohexanones and cyclobutanones.

BVMO_{Flava} showed almost the same conversion rate and enantioselectivity toward substrates **1a** and **1b** in comparison to TmCHMO and CHMO_{Acineto}. In comparison to CHMO_{Acineto}, BVMO_{Flava} showed quite better acceptance of a bulky substrate like 4-tertbutylcyclohexanone (**1c**) with the conversion rate of more than 80% and high enantioselectivity toward S product while this substrate not well accepted by CHMO_{Acineto} (Table C.2). A similar trend was detected for substrates **2** and **3**. All three enzymes showed the same conversion rate and enantioselectivity (Table C.2) with a minor exception for BVMO_{Flava}, which showed a slightly lower conversion rate of substrate **2**.

The kinetic resolution of 4-phenylcyclohexanone (**4a**) and 2-benzylcyclohexanone (**4b**) was performed and resulted in the formation of the R enantiomer in up to 99 % ee after almost 50 % conversion. Next, the conversion rate and enantioselectivity of all three enzymes for four different cyclobutanones tested (5-7), where again, almost the same conversions rate and enantioselectivities observed for all

three biocatalysts with minor exceptions. BVMO_{Flava} showed poor acceptance toward compounds **5a** and **5b**. BVMO_{Flava} converted substrate **5a** to almost racemic lactone, but it converted **5b** to the desired lactone with 77 % ee optical purity. For the compound **6**, full conversion and perfect optical purities for both the normal and the abnormal lactone achieved. The normal lactone is an intermediate in the synthesis of Corey lactone, which is a building block for prostaglandin synthesis^[171], and the abnormal product is a starting material of brown algae pheromone synthesis.^[28b] The cyclobutanone seven was well accepted, and the same enantioselectivity seen in all three enzymes.

Table C.2. Substrate profile for the novel BVMO.

Substrate	R	BVMO _{Flava}		Reference reaction CHMO _{Acineto}		Reference reaction TmCHMO	
		Conv (%) ^A	ee (%) ^B	Conv (%)	ee (%)	Conv (%)	ee (%) ^[56]
	1a R=Me	+++ ^C	99 (<i>S</i>)	+++	98 (<i>S</i>) ^[172]	+++	99 (<i>S</i>)
	1b R=OH	+++	8 (<i>R</i>)	+++	10 (<i>R</i>) ^[173]	+++	18 (<i>R</i>)
	1c R= <i>t</i> Bu	+++	96 (<i>S</i>)	+	>98 (<i>S</i>) ^[173]	+++	93 (<i>S</i>)
	1d R=Ph	+	89 (-)	+	60 (-) ^[174]	+++	88 (-)
	2	++	99 (4 <i>S</i> ,6 <i>R</i>)	+++	99 (4 <i>S</i> ,6 <i>R</i>) ^[175]	+++	99 (4 <i>S</i> ,6 <i>R</i>)
	3	+++	P:D/41:59 ^D >99(-),96(-)	+++	P:D/49:51 99(-),99(-) ^[176]	+++	P:D/49:51 99(-),99(-)
	4a R=Ph	+	94 (<i>R</i>)	+	98 (<i>R</i>) ^[176]	+	97 (<i>R</i>)
4b R=Bn	+	99 (<i>R</i>)	+	96 (<i>R</i>) ^[176]	+	98 (<i>R</i>)	
	5a R=Ph	+	17 (<i>R</i>)	+++	62 (<i>R</i>) ^[174]	+++	49 (<i>R</i>)
	5b R=Cl-Ph	+	77 (<i>S</i>)	+++	81 (<i>S</i>) ^[174]	+++	95 (<i>S</i>)
	6	+++	N:ABN/50:50 ^E >99(-),>99(-)	+++	N:ABN/51:49 95(-),>99(-) ^[176]	+++	N:ABN/50:50 >99(-),>99(-)
	7	+++	N:ABN/59:41 72(-), >99(-)	+++	N:ABN/65:35 60(-),>95(-) ^[176]	+++	N:ABN/55:45 79(-),>98(-)

^ARelative conversion (Conv) of the substrate to product ^BEnantiomeric excess (ee) of product ^C10-50%: +, 50-80%: ++, 80-100%: +++ ^DP:D ratio of proximal to distal lactone ^EN:ABN ratio of normal to abnormal lactone

C II Redesign

Asymmetric catalysis plays an important part in organic chemistry. ^[177] Many therapeutic compounds, fragrances, and plant-protecting agents and many natural products are chiral, most applying their biological effect only in one enantiomeric form. ^[177a] Enzymes are products of evolution and may be expected to function with high enantioselectivity toward specific enantiomers and also only with natural substrates. However, it is already well known that this is not the case because enzymes can convert a large number of unnatural substrates. ^[178] Nevertheless, the problem of substrate specificity and also enantioselectivity persists. In such cases, several methods to enhance enzyme-substrate and enantioselectivity have been described, including site-specific mutagenesis and directed evolution (Figure C.13). ^[179]

The development of chiral biocatalysts for the enantioselective synthesis of optically pure active organic compounds is of considerable interest, especially in the area of pharmaceuticals, fragrance, and plant protecting agents, as was mentioned earlier. ^[180]

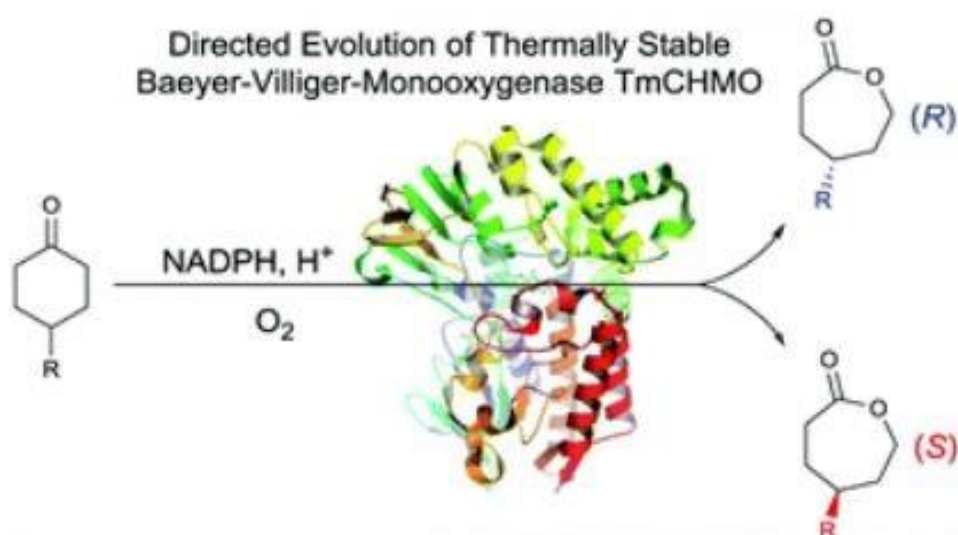
BVMOs have shown the capability of the conversion of numerous structurally different ketones with high stereoselectivity. ^[24b, 27, 28b, 82, 181] In those cases in which moderate or poor stereoselectivity has been observed, protein engineering approach like directed evolution can be applied to inverse or improve the enantioselectivity. ^[182] For example, this has been applied to CHMO_{Acineto}. ^[80, 177a, 180, 182-183] But, the major obstacle for wide industrial applications seems to persist, which is the lack of stability of most BVMOs under operating conditions. ^[110] One approach is to apply protein engineering to improve the thermostability of CHMO_{Acineto}, which have met limited success. The next alternative is to use protein engineering to improve the selectivity and promiscuity of already known thermostable BVMOs like recently discovered thermostable CHMO from *Thermocrispum municipale* DSM 44069 (TmCHMO). ^[56] This BVMO has high thermostability and accepts ketones such as cyclohexanone, phenylacetone, acetophenone, and 2-octanone, but stereoselectivity has not been studied to date. ^[56] Besides, the TmCHMO crystal structure has been solved, which can provide beneficial information for guiding protein engineering. Therefore, TmCHMO was considered to be an attractive starting point for exploring and possibly engineering the regio- and stereoselectivity as well as substrate acceptance.

To achieve this purpose, we have used molecular dynamic simulation to predict the mutations that can result in alteration of selectivity. The predicted mutations were implemented by using directed evolution based on iterative saturation mutagenesis (ISM). To see if we can see a general trend toward different substrates and if we found a hot spot for stereoselectivity, the substrate profile for mutants

was studied. This resulted in the inverse selectivity of some variants toward different enantiomer in comparison to wild type TmCHMO but it was not a general trend for all the substrates that have been studied. This all happened without affecting the thermostability of the biocatalyst.

Baeyer-Villiger oxidation of 4-methylcyclohexanone was used as the model reaction for assessing the enantioselectivity of TmCHMO (Figure C.13). Whole-cell reactions using wild type TmCHMO illustrates excellent enantioselectivity and conversion in favor of (*S*)-**2a** (99% ee; conversion 100%/24 h).

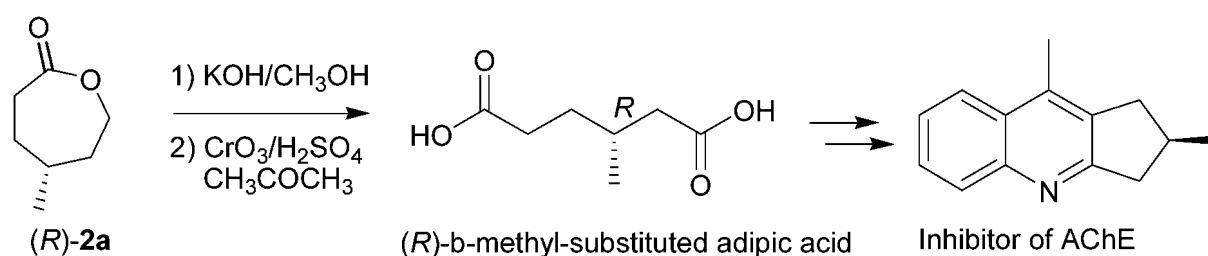
Figure C.13. Directed evolution of TmCHMO.^[184]



We tried to invert stereoselectivity by protein engineering toward lactone (*R*) as this is a chiral compound of particular interest because it is a precursor of (*R*)- β -methyl-substituted adipic acid that is an important intermediate in the synthesis of an effective inhibitor of acetylcholinesterase (AChE)^[185] (Figure C.14). This was tried before, by inverting the selectivity of cyclopentanone monooxygenase from *Comamonas* sp toward (*R*) product, which was successful.^[181a] The mutant was shown to be (*R*)-selective (96% ee) in this reaction^[181a], but then again this BVMO lacks sufficient thermostability under operating condition. In another study, a variant of PAMO was created that was also shown to be (*R*)-selective (98% ee), nevertheless, the activity was poor.^[186] So, it is quite important and interesting to change the selectivity of TmCHMO toward R-product as a valuable biocatalyst.

All the experiments for this section (redesign) were designed and performed in our collaboration partner lab (Reetz group, Max-Planck-Institut für Kohlenforschung, Mülheim an der Ruhr, Germany), and only assigning the substrate profile was performed as a side project during this Ph.D. thesis.

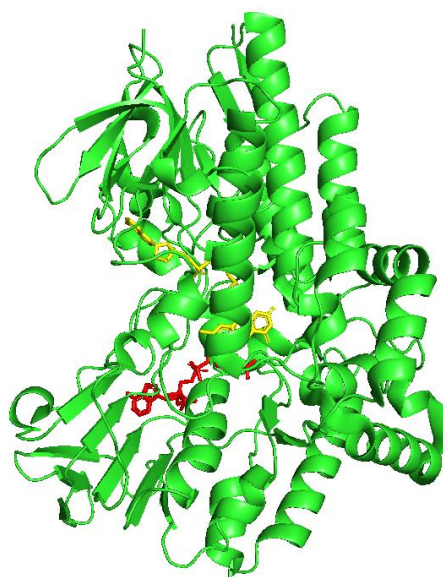
Figure C.14. Synthesis of (*R*)- β -methyl-substituted adipic acid, the precursor of an AChE-inhibitor.



The investigations were started with directed evolution based on saturation mutagenesis at sites, which are surrounding the binding pocket (CAST-sites; Combinatorial Active-site Saturation Test)¹⁴ and iterative saturation mutagenesis (ISM).^[187] Utilizing the X-ray crystal structure of TmCHMO (pdb accession code 5M10),^[56] 4-Methylcyclohexanone docked in the active site of the enzyme (Figure C.15).

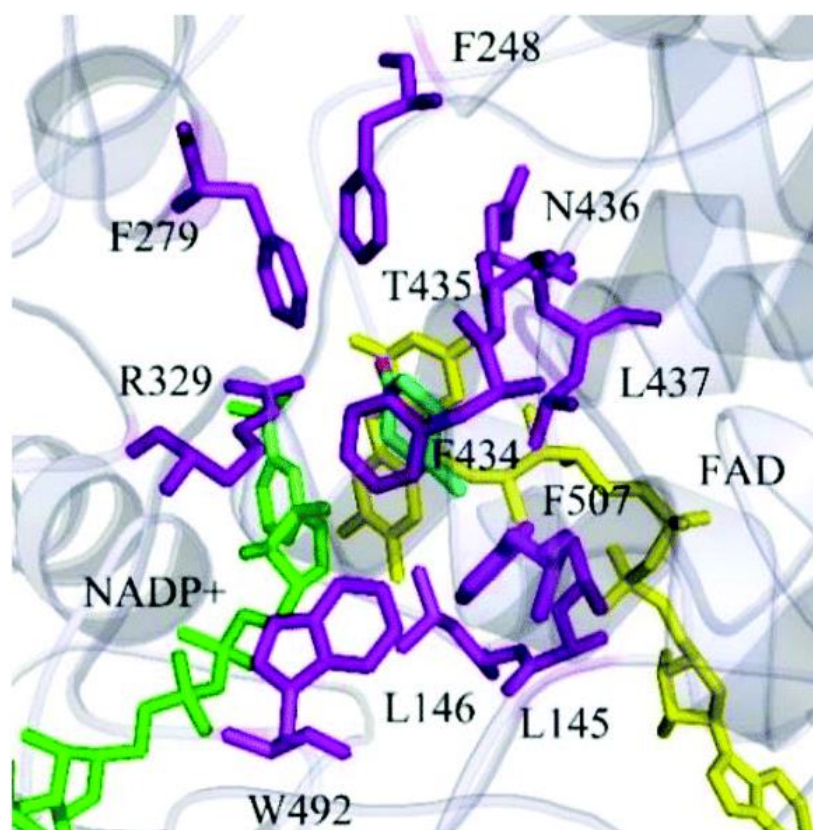
Figure C.15. TmCHMO crystal structure model docked with 4-Methylcyclohexanone (Cyan).

This is the model used for the prediction of mutagenesis involved in the selectivity.



Subsequently, 11 amino acids, which were close enough to substrate binding site (within 5 Å from the docked substrate) chosen for saturation mutagenesis (L145, L146, F248, F279, R329, F434, T435, N436, L437, W492, and F507) (Figure C.16). The mentioned positions subjected individually to site saturation mutagenesis (NNK-based randomization) in which all 20 amino acids substituted, followed by the screening of about 96 transformants for 95% library coverage (one microtiter plate) in each case. Such exploratory experiments can provide valuable information for subsequent mutagenesis steps.

Figure C.16. TmCHMO structure model showing docked 4-methylcyclohexanone (1a) (in cyan) based on the crystal structure of wild-type TmCHMO (5M10),¹¹ which served as a guide for choosing amino acids for saturation mutagenesis (in purple).



Several mini-libraries generated by saturation mutagenesis at predicted positions (146, 434, 435, 437, and 507). Many variants were revealed decreased (*S*)-selectivity (L146E, F434I, T435F, T435Y, T435W, L437G, L437T, L437A, and F507W), thus pointing the way towards the reversal of enantioselectivity. The other libraries that created at the other six positions revealed no positive variants.

At this stage, two mutagenesis approaches were investigated in parallel. In the first strategy, saturation mutagenesis at a 5-residue randomization site defined by the above hot spots was performed. This design was based on the information provided by the positive mutants. 146E/L, 434I/F, 435F/Y/W/T, 437T/A/G/L and 507W/F were chosen for creating a 5-residue saturation mutagenesis library. 95% of the created library was screened by investigating 384 transformants. Only three (*R*)-selective variants were showing moderately reversed enantioselectivity (50%–66% ee).

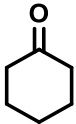
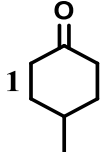
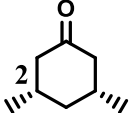
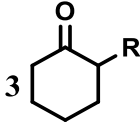
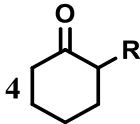
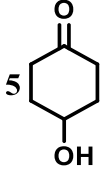
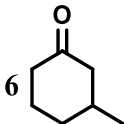
In the second approach, the best variant that identified in the exploratory NNK-based mini-libraries, (L437A), was chosen for performing iterative saturation mutagenesis (ISM) at the remaining four hot spot residues (L146, F434, T435 and F507). Subsequently, several ISM pathways were explored using the respective best variants as templates and NDT codon degeneracy encoding only 12 amino acids (Leu, Phe, Tyr, Ile, Val, Asn, His, Arg, Asp, Ser, Cys, and Gly). Two best (*R*)-selective variants, were identified, LGY3-4-E5 (F434I/T435L/L437A/F507V) (91% ee) and LGY3-4-D11 (L146F/F434I/T435L/L437A/F507C) (94% ee). The next step was to determine if this is a general trend and can be seen concerning different substrates, and also verify if these positions are hot spots for selectivity of TmCHMO or not. This was carried out by using the whole-cell biotransformation method in a non-growing condition as described in (D I.18) using different substrates.

C II.1 Substrate profile

Nine compounds from the family of cyclohexanones were chosen for the selectivity determination experiments. In parallel to all reactions, a control reaction with cyclohexanone has been performed to make sure that enzymes are active and expressions have been performed successfully. The results summarized in Table C.3.

Regarding the compound 4-methylcyclohexanone due to the mutations, the enantioselectivity of the enzyme is changed. Wt is producing S product while mutants are producing R enantiomer of the product, but for the substrate dimethylcyclohexanone, this does not happen. This was not the case for the substrate **2** and **3**. It worth mentioning that variants showed a considerable improvement in enantioselectivity toward compound **5**. Results indicate that mutations improved the tendency of the variants to produce more of the distal product by conversion of compound **6**.



Table C.3. Substrate scope for wt TmCHMO and variants.

Substrate	WT		D11		E5		Reference Biocatalyst
	Con	ee [%]	Con	Ee	Con	ee	
	+++	-	+++	-	+++	-	
1 	+++	98 (S)	+++	94 (R)	+++	91 (R)	CHMO _{Acineto} ^[188]
2 	+++	99 (4S, 6R)	++	99 (4S, 6R)	+++	99 (4S, 6R)	CHMO _{Acineto} ^[175]
3  R=Ph	50%	97 (R)	6%	75 (R)	38%	96 (R)	CHMO _{Acineto} ^[176]
4  R=Bn	39%	98(R)	3%	66(R)	5%	58(R)	CHMO _{Acineto} ^[176]
5 	+++	18(R)	+++	99(R)	+++	99(R)	CHMO _{Xantho} ^[189]
6 	+++	Ratio P: D 45:55, ee% 99(-) ,>99(-)	+++	Ratio P:D 35:65, ee% 28(-),44(-)	+++	Ratio P:D 30:70, ee% 86(-),55(-)	CHMO _{Acineto} ^[176]

4-tert butylcyclohexanone was consumed totally by wt TmCHMO while it almost consumed nor with purified whole-cell D11 and E5 neither, but 4-phenylcyclohexanone was well accepted and consumed by wt TmCHMO and poorly consumed by D11 and E5 (Table C.4). Wt TmCHMO and variant D11



were producing in favor of (-) enantiomer with ee 88 and 16 % respectively while E5 produced another (+) enantiomer of the product by ee of 30% (Table C.4).

Table C.4. Substrate scope for wt TmCHMO and variants.

Substrate	WT		D11		E5		Biocatalyst
	Con	ee	Con	ee	Con	ee	
 R=tBu	+++	93 (S)	n.c	n.d.	n.c	n.d.	CHMO _{Xantho} [190]
 R=Ph	++	88 (-)	+	16(-)	+	30(+)	CHMO _{Acineto} [174]

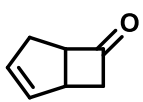
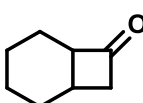
As was mentioned, four different cyclobutanones were selected for the experiments. The results are compiled in Table C.5. Wt enzyme and both mutants showed the same enantioselectivity toward 3-phenylcyclobutan-1-one. The information which can address here is that ee almost increased two times in mutants, which is a good improvement in enzyme characteristic. However, different behavior has monitored regarding substrate 3-(4-Chlorophenyl)-cyclobutanone, and mutants are producing S product while wt enzyme is producing R product, again here it can be seen that rate of conversion in mutant D11 is lower than others.

Table C.5. Substrate scope for wt TmCHMO and variants.

Substrate	WT		D11		E5		Biocatalyst
	Con	ee	Con	ee	Con	ee	
 R=Cl-Ph	+++	95 (S)	++	94 (R)	+++	95 (R)	CHMO _{Acineto} [174]
 R=Ph	+++	49 (R)	+++	99 (R)	+++	98 (R)	CHMO _{Acineto} [174]

All three enzymes were converting the substrate (\pm)-cis-Bicyclo[3.2.0]hept-2-en-6-one with the same rate and almost the same ratio of N: ABN % and ee of the product (Table C.6). Regarding substrate Bicyclo[4.2.0]octan-7-one both mutants showed 48:52 ratio of N: ABN with 97 and 98 % ee respectively in favor of (-) product which is slightly different from wt enzyme which produced the product by ratio of N:ABN (55:45) and ee of 79 and 98 in favor of (-) product. Here, it can be seen that mutations gave the mutants the ability to produce slightly more ABN than N in comparison to wt, which is producing more N product.

Table C.6. Substrate scope for wt TmCHMO and variants.

Substrate	WT		D11		E5		Biocatalyst
	Con	ee	Con	ee	Con	ee	
	+++	N: ABN 50:50 ee% >99(-), >99(-)	+++	50:45 83(-), 99(-)	+++	50:50 98(-), 99(-)	CHMO _{Acineto} [174]
	+++	N: ABN 55:45 ee% 79(-), 98(-)	+++	48:52 97(-), 97(-)	+++	48:52 97(-), 97(-)	CHMO _{Acineto} [174]

This study illustrated that TmCHMO, and the two variants LGY3-4-D11 and LGY3-4-E5, are exceptional biocatalysts in the asymmetric conversion of a variety of different ketones. The two variants were made by directed evolution utilizing iterative saturation mutagenesis (ISM). In some cases, the great enantioselectivity in desymmetrization reactions is comparable to those reported for CHMO_{Acineto}, but TmCHMO has the advantage of being moderately stable in comparison to CHMO_{Acineto}. Furthermore, the reversal of enantioselectivity, especially in case of 4-methylcyclohexanone, allows access to chiral compounds of particular synthetic value.

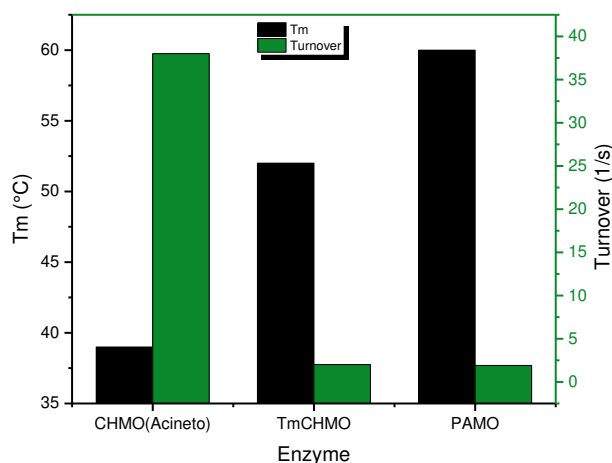
C III Protein engineering

The main aim of this study was to improve the stability of CHMO_{Acineto} using protein engineering. To achieve this goal, we have used two different protein engineering approaches (Semi-rational design and rational design). These two approaches have the benefit of reducing the size of mutant libraries, which leads to the reduction of screening efforts and material.

It is crucial to mention here that alteration of the enzyme stability should be monitored in parallel to enzyme activity as well. An enzyme with high stability and no activity cannot be the right candidate for the industrial applications (Figure C.17). The screening for the stability should be performed in parallel with the catalytic efficiency of the enzyme. In the best case, an improved variant should display the same or even a higher activity and have an increased kinetic stability.

The evaluation of these two parameters are also important because in most of the studies the aim is to rigidify the enzyme structure by introducing the disulfide bond, salt bridges or higher number of hydrogen bond which all can result in a substantially rigidified enzyme, which loses the activity because of the lack of elasticity for movements, especially in or close to the active site. So checking activity is crucial to avoid an improvement in stability in the cost of diminishing the activity.

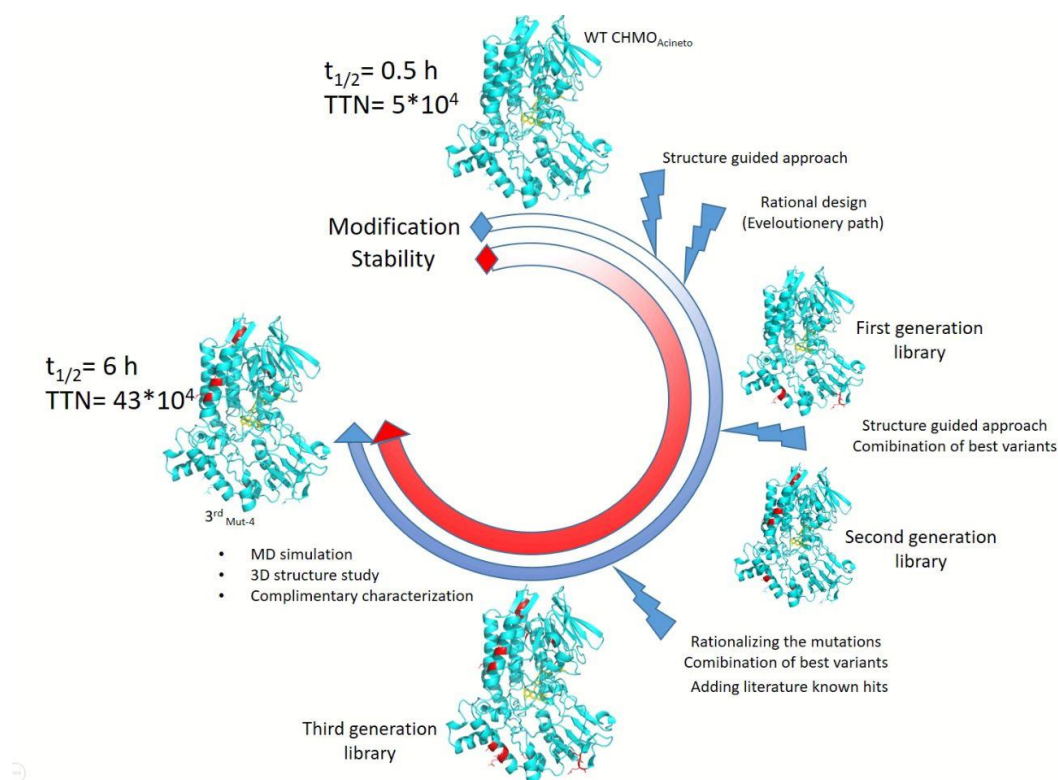
In Figure C.17, the two most stable BVMOs, with respect to T_m values (TmCHMO and PAMO) are compared to the most active one (CHMO_{Acineto}). Although the thermodynamic stability of TmCHMO and PAMO is much higher the activity will restrict any further industrial application.

Figure C.17. The comparison of stability and activity of different BVMOs.

C III.1 General workflow

To improve the characteristics of CHMO_{Acineto}, especially its thermostability, three generations of mutants created (Figure C.18). The first generation was mainly predicted using semirational design (Consensus approach) and rational design (1st_{mut-1}). After the characterization of the 1st generation library, the best variants were detected and used for the next generation. The second generation is a combination of best variants from the first library and a few more mutations, which were predicted by the consensus approach. Characterization was carried out afterward, and the best variants were determined (Figure C.18). The third generation of variants created using the combinatorial method. The best mutants were combined with a few literature-known mutation hits (Figure C.18).

Figure C.18. Overview of the individual steps followed in the CHMO stabilization workflow. Three generations of libraries created with 17 individual variants. Mutations are labeled in red.



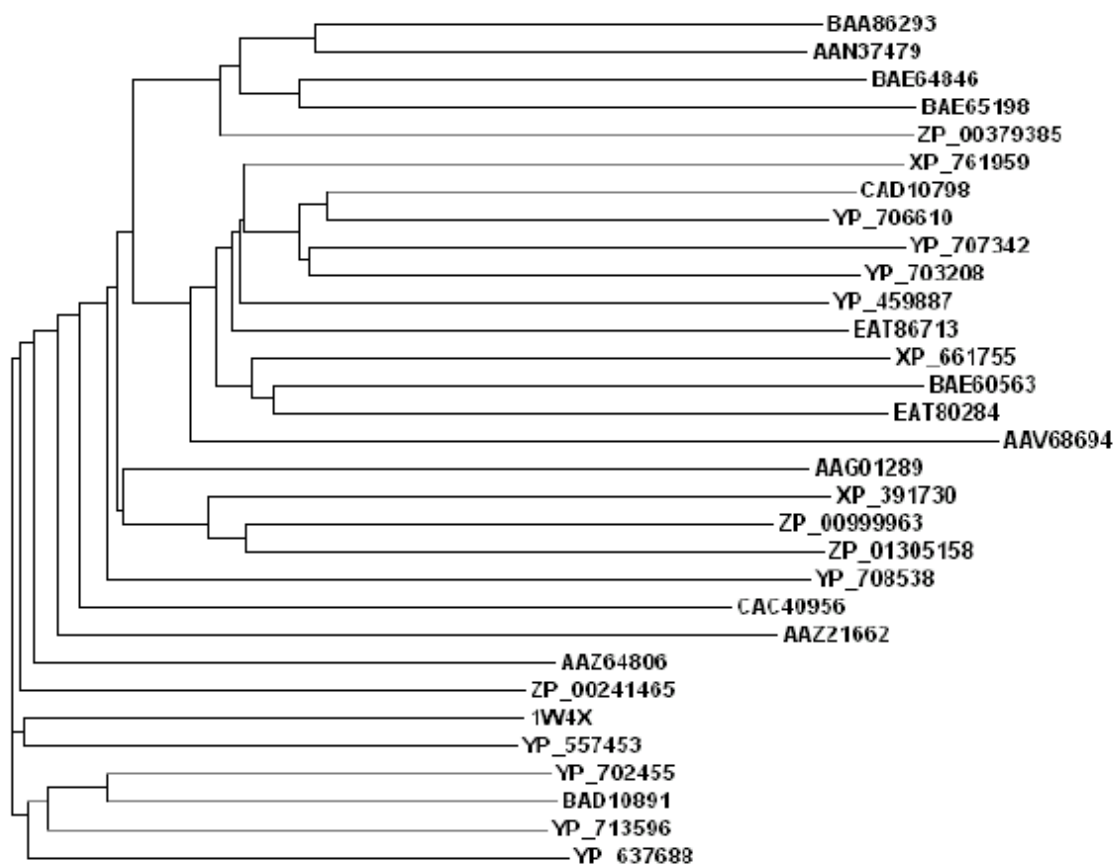
Structure-guided consensus approach was applied to the sequence of the CHMO_{Acineto} by (J. F. Chaparro-Riggers, A. S. Bommarius, Georgia Institute of Technology, Atlanta) (Figure F.1). The structure-guided consensus concept, which was based on sequence alignment information, is classically applied to homologous protein families. In the mid-nineties, the consensus approach was established by Steipe *et al.* [139] to create thermostable immunoglobulins. Later the research group of Bommarius *et al.* advanced this approach by the implementation of structural information to gain higher stability, especially thermostability of enzymes. [144a]. This approach is based on the hypothesis that conserved amino acids at a given position of enzyme sequence contribute more to enzyme stability than the non-conserved amino acids. So this approach focusses on finding conserved amino acids and modification of these amino acids. This approach increasingly used because it can limit the size of a library based on extensive protein knowledge and make the screening faster and easier.

The consensus approach was already conducted within the PhD thesis of our formal colleague Dr. Saima Feroz to improve the stability of CHMO_{Acineto}. But it turned out that during her efforts back then,

the available screening method only insufficiently described stabilization effects. Later on, upon another work by former colleague Dr. Leticia Goncalves,^[110] the appropriate assay conditions were established to provide reliable and reproducible data. Her study clearly showed that there are several variables that can affect the stability of CHMO_{Acineto} in the reaction condition that need to be considered to achieve reliable and reproducible results. These variables were firstly the enzyme concentration. Secondly, the cofactors (NADPH and FAD) concentration are quite critical and needs to be optimized for the experiments. One other thing that was reconsidered by Dr. Goncalves was the enzyme purification optimization. She noticed that due to low stability of CHMO_{Acineto}, the purification needs to be optimized timewise and chemicalwise. She shortened the purification time to achieve the active enzyme and also added FAD as a protective compound throughout the purification that helped enzyme to survive the purification steps. Based on this, the previous consensus approach was revisited within this thesis.

This blast-search based on the BVMO family sequence pattern F-x-G-x-x-x-H-x-x-x-W resulted in more than 150 sequences. As an additional filter for further experiments, sequences that showed only less than 8% of gaps selected. It was also the case that the sequences chosen to be very different from each other. Therefore the branches in a phylogenetic tree will be deeply divided. By applying the above selection criteria, the number of enzyme sequence candidates reduced to 31, and they all aligned with PRALINE.^[191] These sequences were used for drawing the phylogenetic relationship between these BVMOs as well (Figure C.19).^[191c] The phylogenetic tree (Figure C.19) shows that the branches are very deeply divided as desired. These sequences were aligned using ClustalW (<http://www.ebi.ac.uk/clustalw/>), and the multiple sequence alignment subjected to predict the consensus sequence using the (<http://coot.embl.de/Alignment/consensus.html>) (Figure F.1).

Figure C.19. Phylogram obtained after aligning the protein sequences used for the consensus approach. ^[191c] The list of sequences can be found in the appendix.



The position selected for the rational design was the glycine in position 14. This is a hot and critical position in the enzyme structure as it is located precisely in the Rossmann fold, and it is assumed to be involved in FAD binding (Figure C.20). From our previous studies, we already knew, that the binding of FAD to the enzyme has a significant influence both in the activity and stability of the enzyme. ^[110]. Thus, improving enzyme FAD binding affinity will help us to reach our aim that is improving the stability of CHMO_{Acineto}.

The Rossmann fold is a super secondary structure that can be seen in proteins binding to nucleotides. Rossmann fold is composed of beta strands and alpha helical sections. The alpha helices surround both faces of the sheet to produce a three-layered sandwich and the beta strands are hydrogen bonded to each other forming an extended beta sheet. The main function of this motif in enzymes is to bind nucleotide cofactors. It also often contributes to substrate binding. It is worth to mention that in

the Rossmann fold, first, third, and sixth positions are highly conserved, and they are highly susceptible to changes.

Based on this knowledge, we believe that by improving the affinity of the enzyme to FAD, we will improve the properties of the enzyme. Interestingly, this position was suggested by the consensus approach as a position that can be involved in the stability of CHMO_{Acineto}. Also, multiple sequence alignment with different BVMOs (Figure C.21 **Error! Reference source not found.**) indicates that BVMOs with higher stability are carrying alanine amino acid in this position, which encouraged us to replace the glycine in the position 14 with all amino acids and especially with Alanine. So this position was our interest in performing site saturation mutagenesis to evaluate the effect of different amino acids.

Figure C.20. Rossmann fold sequence in the CHMO_{Acineto} sequence.



These all tempted us to focus on this position and to have a better understanding of enzyme FAD binding properties and better understanding of this position, we have performed site saturation mutagenesis in position 14.

Figure C.21. Multiple sequence alignment of different BVMOs.

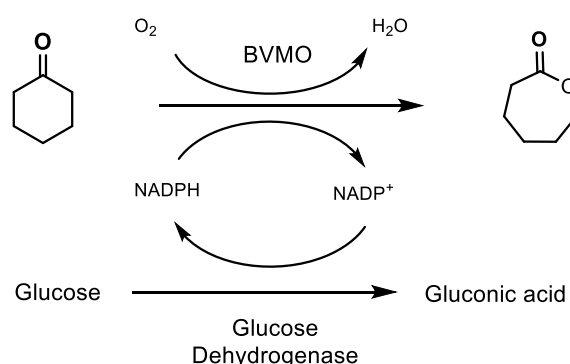
HAPAMO	TAEEDLRAPRWHKDHVASGRDFKVVIIIGAGESGMIAALRFK-QAGVPFVIYEKGNVGGT	179
CPDMO	-----DPVVRETDVFIIGGGFGGLLAAVRLQQAGVSDYVMVERAGDYGGT	111
CDMO	-----EPKLDHVTFAFIIIGGGFSGLVTAARLRESGVESVRIIDKAGDFGGV	95
CPMO	-----NSVDDTLDVLLIGAGFTGLYQLHHLR-KLGFKVHLVDAGADVGGI	56
PAMO	-----RQPPEEVDVLVVGAGFSGLYALYRLR-ELGRSVHVIETAGDVGGV	54
CAMO	-----IPEVLNVDALVVGAGVAGIYSTYRLS-RAGLNVQCIDTAGDVGGT	47
CHMO	-----MSQKMFDAIVIGGGFGGLYAVKKLRDELELKVQAFDKATDVAGT	45
TmCHMO	-----TTQTPDLDAIVIGAGFGGIYMLHKLRLNDLGLSVRVFEKGGVGGT	47
	.:*.* *: : : . . . *	

All the mutations introduced using the QuikChange® II XL Site-Directed Mutagenesis kit from Stratagene following the protocol provided with the kit. A complete list of the oligonucleotides utilized as mutagenic primers is available in the Ph.D. thesis from former colleague Dr. Saima Feroz. ^[191c] All the primers were designed using the online primer designing tool that recommended in the kit. Three transformant colonies from each variant were isolated using the Wizard Plus SV Miniprep DNA Purification kit (Promega). Plasmid samples were sequenced, and the introduction of mutations confirmed by sequence analysis. These plasmids were transformed into *E. coli* BL21 (DE3) cells as the final host using the heat shock described in the experimental part.

Overexpression of the mutant carried out as described in (D I.12) and the purity determined by crude SDS-page gel analysis.

What is so crucial in enzyme kinetic measurement is to have a reliable and reproducible method to evaluate the enzyme kinetics, and to do that, the assay method should be chosen carefully. In the whole-cell measurement, since the cellular machinery is still intact, the host will provide a natural recycling system for all coenzymes. However, working with the isolated enzyme require cofactor recycling, which can be achieved by adding either the cofactors directly to the reaction mixture in over-stoichiometric quantities or using a suitable regeneration system like dehydrogenase, which can reduce NADP⁺ and provide the NADPH at the expense of an additional substrate (Figure C.22).

Figure C.22. Baeyer-Villiger oxidation of cyclohexanone and NADP⁺/NADPH cofactor.



Since it is quite complicated and also somehow impossible to investigate the thermostability of our variants in *E. coli* whole cells as the cell will disintegrate at elevated temperatures, protein purification and the direct use of cofactors was mandatory for the assay. So, the purified enzyme with a suitable

screening condition ^[110] used to evaluate the activity and thermostability of the new CHMO_{Acineto} variants according to established protocols, as mentioned in the experimental section.

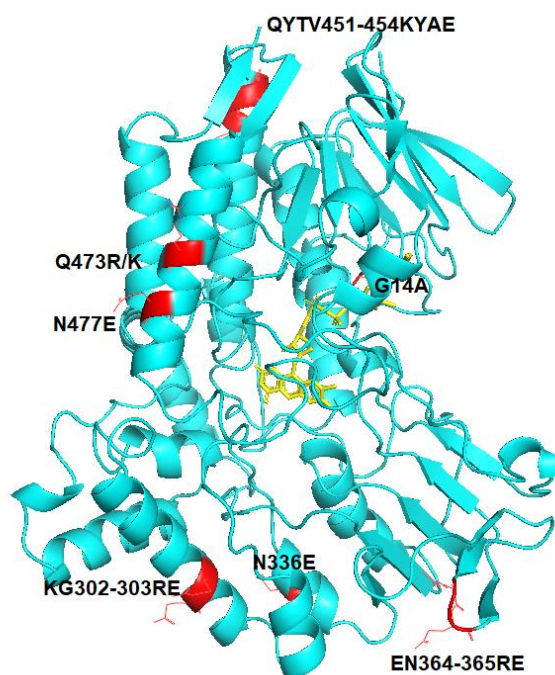
C III.1.1 First-generation mutants

The mutations were designed far enough from the active site, more than 6Å. This is because there was no intention to interfere directly with the activity or the affinity for the substrates in the first place. A list of variants that was made as the first-generation library is shown in Table C.7, the rationally designed G14A variant, is also included in the first generation variants. A model of CHMO_{Acineto} with the mutations labeled, which made based upon the homology approximation with TmCHMO, is shown in Figure C.23.

Table C.7. List of first-generation variants.

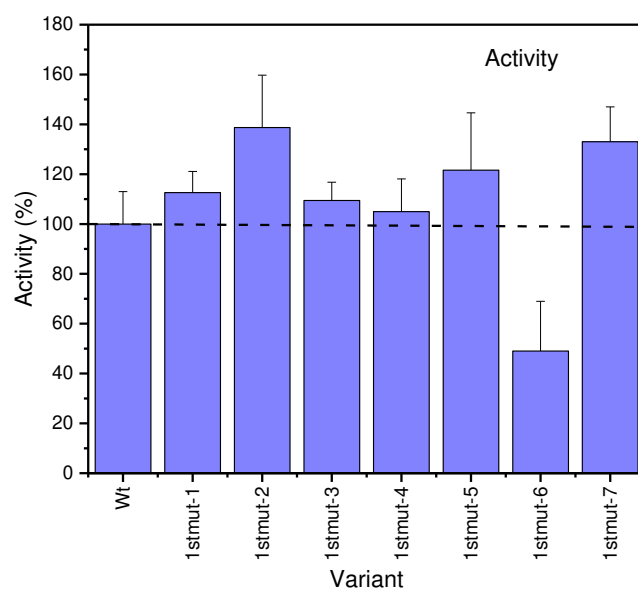
Variant	Abbreviation
G14A	1 st _{mut-1}
N336E	1 st _{mut-2}
KG302-303RE	1 st _{mut-3}
EN364-365RE	1 st _{mut-4}
Q473R/N477E	1 st _{mut-5}
Q473K/N477E	1 st _{mut-6}
QYTV451-454KYAE	1 st _{mut-7}

Figure C.23. CHMO_{Acineto} structure with the first-generation mutations.



All the screening was performed in triplicate unless otherwise stated, and always the same experiment was performed for wild type CHMO_{Acineto} as the reference. The activity was measured as described in the experimental part. The first interesting result obtained by screening the activity at 30 °C for the first generation variants (Figure C.24). All the variants either showed the same or even higher activity in comparison to the wild type CHMO_{Acineto} except 1st_{mut-6}, which showed about 50% less activity, which means this position can be critical for the activity and it is better to avoid for the future experiments. The most exciting variant was 1st_{mut-2}, which contains N336E mutation by showing about 40 % higher activity. This makes this position interesting to be used in further experiments as it can increase enzyme activity (Figure C.24).

Figure C.24. Activity measurement for first-generation variants. Enzyme activity was measured by monitoring the decrease of NADPH absorbance at 340 nm. The activity assay mixture contained 0.05 μM CHMO, 100 μM NADPH, 0.5 mM cyclohexanone in 50mM Tris HCl pH 8.5 at 30°C.



Stability data reported in the literature for CHMO_{Acineto} is not consistent and can vary by order of magnitude, reported half-life varies from 4 min [105] to 1 day [83] (incubated at 30 °C, pH 8.5). This considerable discrepancy cannot be seen in the case of thermodynamic stability as this method is quite simple and easy to reproduce in comparison to kinetic stability. Thermodynamic stability was determined by nano differential scanning fluorimetry (nanoDSF) using 2 mg mL⁻¹ enzyme in 50 mM Tris HCl, 10 μM FAD, pH 8.5. The enzyme concentration was different from the kinetic stability as the Nanotemper machine has a limitation for enzyme concentration (Figure C.25). In parallel to all measurements, the evaluation was performed for CHMO_{Acineto} (reference) as well. Fortunately, all the variants showed either higher or same T_m value as the wild type CHMO_{Acineto}. The best mutant was rationally designed 1st_{mut-1} (G14A) by increasing almost two °C in T_m value that again makes it the best variant among the first generation of mutants and also most promising one to be used for the next generation of variants. The kinetic stability measurement shows quite different results in comparison to thermodynamic stability as it was expected (Figure C.26). 1st_{mut-1}, 1st_{mut-2}, and 1st_{mut-7} were the only variants that showed improvement in the half-life of the enzyme, and again the best variant was 1st_{mut-1} with about 40% higher half-life (Figure C.26).

Figure C.25. Thermodynamic stability of first-generation variants. Thermodynamic stability was measured by nano differential scanning fluorimetry (nanoDSF) using 2 mg mL^{-1} enzyme in 50 mM Tris HCl , $10 \text{ }\mu\text{M FAD}$, $\text{pH } 8.5$. The standard deviation is less than 3% for all thermodynamic stability measurement.

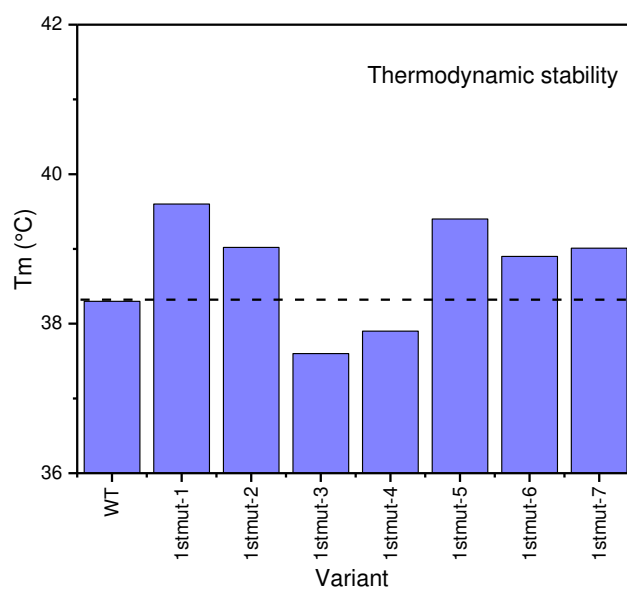
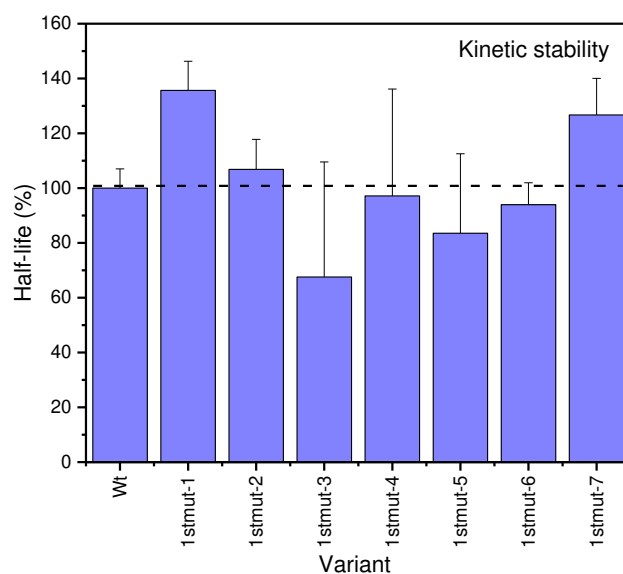


Figure C.26. Kinetic stability for first-generation variants. Kinetic stability was measured by incubation of $1 \text{ }\mu\text{M}$ isolated enzyme at 30°C in 50 mM Tris HCl containing $10 \text{ }\mu\text{M FAD}$, $\text{pH } 8.5$.



C III.1.1.1 The dissociation constant (Kd)

Kd measurement helps to understand the binding affinity of the enzyme to its cofactors, which is significant to the appreciation of the intermolecular interactions driving biological processes, structure-function relationships, and structural biology. Kd, or in other words, binding affinity, is determining the strength of the binding interaction between enzyme to its ligand/binding partner. The smaller the Kd value, the higher the binding affinity of the ligand for its target. The larger the Kd value, the weaker the target molecule and ligand are attracted to and bind to one another. Kd is influenced by non-covalent intermolecular interactions such as hydrogen bonding, hydrophobic, electrostatic interactions, and Van der Waals forces between the enzyme and its ligand. Additionally, binding affinity may be affected by the existence of other molecules.

Based on the previous study, we know that the higher the enzyme-FAD binding affinity is, the more stable the enzyme is. ^[110]. As mentioned earlier, position 14 located in the Rossmann fold, and modification in this position can affect the cofactor binding property of the enzyme. Dissociation constant (Kd) evaluation is an approach to check if the binding affinity is affected by mutation in this position or not. As can be seen in results, variant G14A (1st_{mut-1}) is increasing the stability and this mutation located quite close to the FAD binding site. So there is a chance that this mutation is altering the Kd value. So the determination of the Kd value will help to understand if there is an alteration in Kd value and also verify or falsify our hypothesis.

The result showed that this variant has a higher binding affinity (>8-fold) than the wildtype CHMO_{Acineto} (Table C.8), which means it is more tightly bound to FAD, and it can be one of the reasons why this variant is more stable than wild type. The graphs for Kd measurements can be seen in appendix E I.7.

Table C.8. Kd measurement. The Kd was determined by fitting the data of catalytic activity of the holoenzyme versus concentration of FAD with a logistic function (Origin 8.5 for Windows)

Variant	Kd (μM)	StDEv
Wt	1.6	0.06
G14A	0.19	0.07

C III.2 Site saturation mutagenesis of G14

As was discussed before, position G14 seems to be a critical position that can directly affect FAD binding affinity and indirectly improving thermostability of the enzyme. So we considered to study and evaluate in more detail how important this position is. Site saturation mutagenesis was used as the method to introduce the mutations in this position. By this approach, we could substitute all the amino acids in this position and study their effect. The procedure for site saturation mutagenesis is explained in the experimental part (D I.11). The results for the introduction and confirmation of substitution summarized in

Table C.9. Except for the Met, Val, and Ser, we could introduce all other amino acids successfully, which made possible by using the primers, which listed in appendix E I.4. To facilitate the expression of these variants, we transformed the plasmid containing the gene of interest to *E. coli* BL21 (DE3) as the expression host.

Table C.9. Site saturation in position G14.

NO.	Mutagenesis	Template	PCR	Transformation (<i>E. coli</i> Top10)	Sequencing	Transformation to <i>E. coli</i> BL21 (DE3)
1	G14F	CHMO _{Acineto}	✓	✓	✓	✓
2	G14L	CHMO _{Acineto}	✓	✓	✓	✓
3	G14I	CHMO _{Acineto}	✓	✓	✓	✓
4	G14M	CHMO _{Acineto}	✓	✗	✗	-
5	G14V	CHMO _{Acineto}	✓	✓	✗	-
6	G14S	CHMO _{Acineto}	✓	✓	✗	-
7	G14P	CHMO _{Acineto}	✓	✓	✓	✓
8	Control	Neb control plasmid	✓	✓	✓	✓
9	G14T	CHMO _{Acineto}	✓	✓	✓	✓
10	G14Y	CHMO _{Acineto}	✓	✓	✓	✓
11	G14H	CHMO _{Acineto}	✓	✓	✓	✓
12	G14Q	CHMO _{Acineto}	✓	✓	✓	✓
13	G14N	CHMO _{Acineto}	✓	✓	✓	✓
14	G14K	CHMO _{Acineto}	✓	✓	✓	✓

Results and Discussion

15	G14D	CHMO _{Acineto}	✓	✓	✓	✓
16	G14E	CHMO _{Acineto}	✓	✓	✓	✓
17	G14C	CHMO _{Acineto}	✓	✓	✓	✓
18	G14W	CHMO _{Acineto}	✓	✓	✓	✓
19	G14R	CHMO _{Acineto}	✓	✓	✓	✓

All the variants were characterized, and their activity and also thermodynamic stability were determined (Table C.10). Among all 15 variants, only 8 of them showed conclusive melting temperature (Appendix. E I.2), and the rest did not show a detectable value. It also worth to mention that out of these eight variants with T_m value, only 2 (G14A, G14R) showed melting temperature comparable with wildtype CHMO_{Acineto}, and the rest showed a drastic decrease in the T_m value, which means this position is so critical that any modification can hugely alter the protein characteristic. The activity was also determined, and the only variants showed a specific activity by measuring the NADPH consumption are G14A and G14R. The activity of variants was also checked by GC using whole-cell biotransformation to avoid the chance of losing enzyme activity through the purification steps. All the variants with T_m value showed GC conversion, but again only G14A and G14R showed comparable results to the wildtype CHMO_{Acineto}. This means that the other variants also folded adequately, but they are so fragile, which they cannot survive the purification.

Table C.10. Characterization of variants from site saturation mutagenesis. Enzyme activity was measured by monitoring the decrease of NADPH absorbance at 340 nm. The activity assay mixture contained 0.05 μM CHMO, 100 μM NADPH, 0.5 mM cyclohexanone in 50mM Tris HCl pH 8.5 at 30°C. Thermodynamic stability was measured by nano differential scanning fluorimetry (nanoDSF) using 2 mg mL⁻¹ enzyme in 50 mM Tris HCl, 10 μM FAD, pH 8.5.

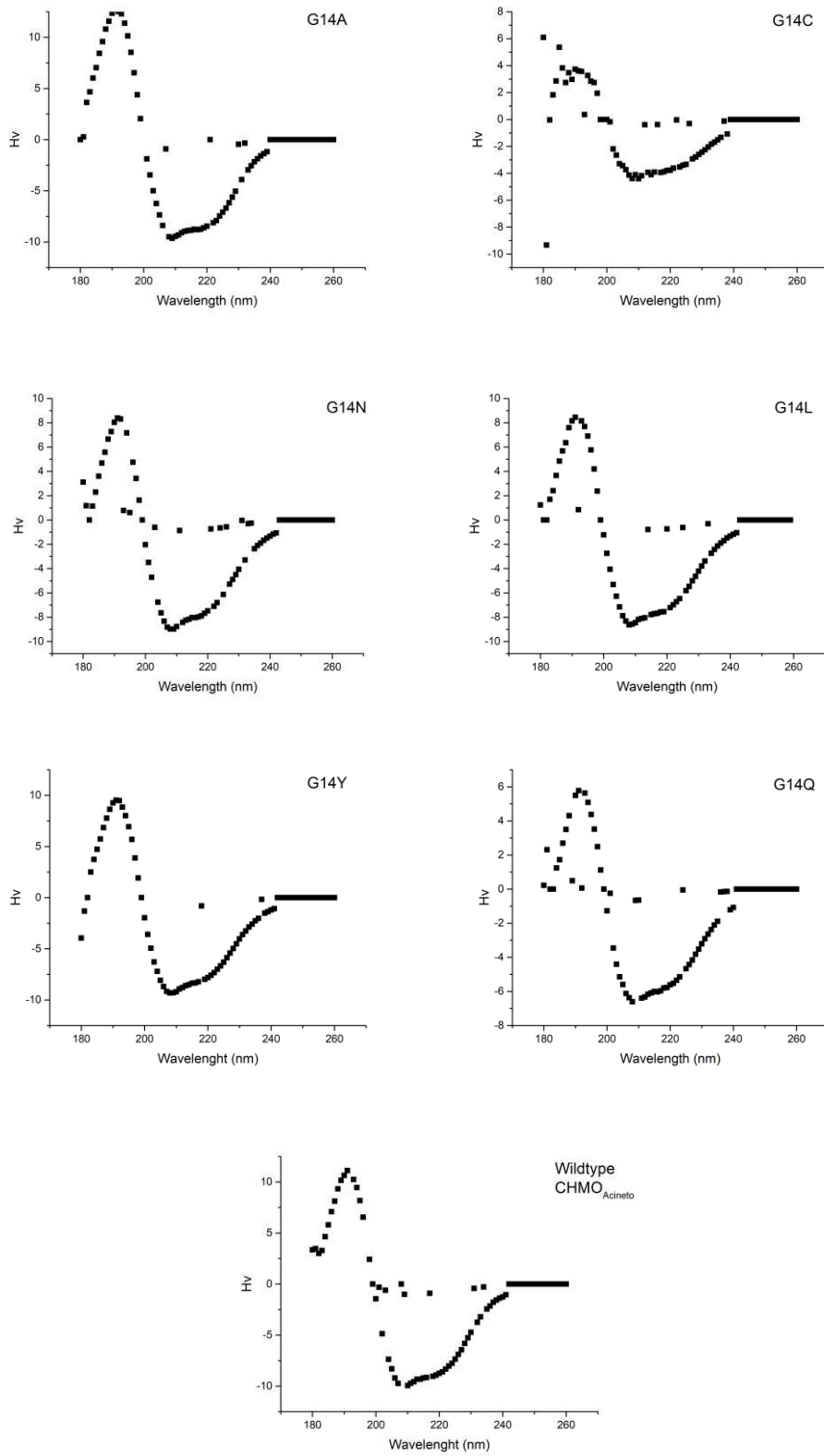
Variant	T _m (°C)	Activity (U/mg)	GC (relative conversion %)
WT	38.9	16.4±1.1	>99
G14A	39.8	18.5±1.6	92
G14L	n.a*	n.a	n.a
G14F	n.a	n.a	n.a
G14Y	28.3	n.a	5.5
G14H	n.a	n.a	n.a
G14N	30.6	n.a	24
G14Q	n.a	n.a	n.a
G14I	25	n.a	<3
G14T	27.6	n.a	5.6
G14E	n.a	n.a	n.a
G14R	39.2	11.5±0.3	94
G14W	28.9	n.a	13.7
G14C	n.a	n.a	n.a
G14D	23	n.a	30
G14P	n.a	n.a	n.a
G14K	n.a	n.a	n.a

n.a= not applicable

CD spectroscopy was used to understand why several variants did not show any catalytic activity (Figure C.27). CD spectroscopy helps to understand the quality of folding and can tell us if the variants not appropriately folded that they are not active or there is another reason which is involved. This method is explained more comprehensively in the experimental section. So seven samples were

chosen for this experiment, 4 variants (CHMO_{Acineto}, G14N, G14A, and G14Y) with measurable T_m value and activity, three variants (G14C, G14Q, and G14L), which showed neither activity nor a T_m value. CHMO_{Acineto} wildtype was used as a reference. As can be seen in Figure C.27 except for G14C, which shows improper folding, the rest are all appropriately folded. That is the reason why the G14C variant is not active. The CD spectroscopy is explained more comprehensively in section E 1.19 This means that the reason behind that most variants are inactive is not only because of improper folding and other mechanisms are involved in the deactivation of these variants.

Figure C.27. CD spectra of mutants and wt enzyme.



The best variants G14A and G14R, which showed a decent structure were studied in more details by performing their kinetic parameters, such as K_m , k_{cat} , and K_d (Table C.11). Both of the variants showed comparable K_m and k_{cat} values.

Table C.11. Kinetic measurement for G14A and G14R. Catalytic rates (K_m , k_{cat}) observed upon incubation of the isolated enzyme with varying amounts of the substrate (cyclohexanone) fitted to the Michaelis-Menten equation.

	K_m	StDEv	k_{cat}	StDEv	k_{cat}/K_m	Ratio
Cyclohexanone	[μM]		[s^{-1}]		[$\text{mM}^{-1} \text{s}^{-1}$]	
CHMO _{Acineto}	6.74	2	15.0	1.3	2220	1
G14A	3.5	0.3	24.2	1.4	7058	3.17
G14R	3.2	1.1	19.7	3.5	6037	2.79

C III.2.1 MD simulation

To have a better understanding of the mechanism involved in the improvement of binding affinity to FAD, five molecular dynamics simulations of 20 ns have performed on the homology model of wildtype CHMO_{Acineto} (Reference), G14A, G14R and G14T (Negative control) variants. These have been performed by our collaboration partner in Ludwig Lab (BOKU University, Vienna, Austria) MD simulations will be beneficial as the mutations did not make sense to the blank eye we especially in the case of G14R. Substitution of glycine with alanine is one mutation that proved to be stabilizing the enzyme and also improving enzyme FAD binding affinity. alanine and glycine are both small, non-polar amino acids and their aliphatic side chains do not allow any specific chemical interactions with other molecules. As these two amino acids are very small in size, they can be found almost anywhere in a protein. glycine, is often found in tight turns because its small hydrogen side chain does not cause steric interference. Glycine is a unique amino acid as it contains a hydrogen as its side chain, which means that there is much more conformational flexibility in glycine. This is also very important for our study because this high flexibility can destabilize the FAD binding in the rosmann fold.

As mentioned the substitution of glycine with arginine (G14R) resulted in an unexpected outcome. Arginine is a bulky amino acid with a polar side chain that makes it quite different from glycine. But as mentioned in Table C.10, Table C.11, this variant is illustrating very similar characteristics compared to wildtype enzyme. This was unexpected as it was assumed that such a bulky amino acid would interfere with proper folding of the Rossmann fold and result in an inactive variant. So carrying out MD simulation would give important information for further research in this position.

C III.2.1.1 Analysis of the RMSF of the loop and FAD

The simulations containing the mutation (G14A) displayed both similar fluctuations to the wildtype loop and also similar average positions (Figure C.28). These results suggested that the mutant variant is not affecting the flexibility of the loop. However, the adenosine part of the FAD cofactor fluctuated significantly more in the wildtype protein than in the mutant one (Figure C.28, **Figure C.29**), suggesting that the observed increase in the stability of the mutant seems to derive from a change in the stability of the cofactor-enzyme complex. The simulations also suggested that the observed loss of the cofactor may be initiated by a detachment of the adenosine part of the FAD.

Figure C.28. 3D representation of the average position of the mutated loop and the FAD cofactor for the G14A mutant (A) and the wildtype (B), with their respective RMSF represented in the color gradient ranging from blue (small fluctuations) to red (higher fluctuations).

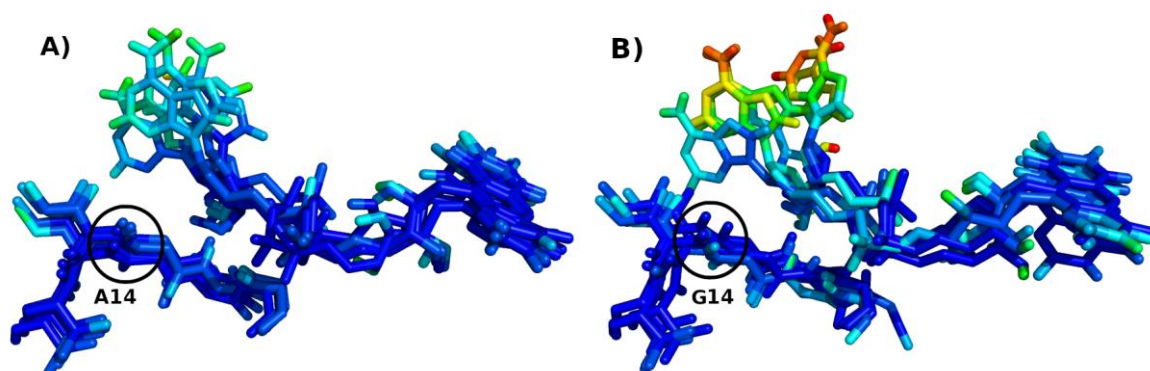
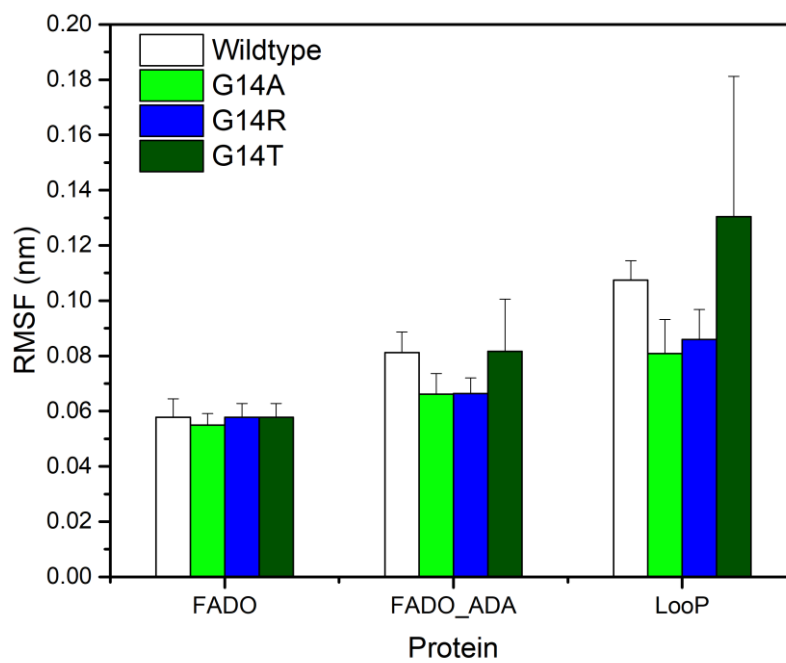
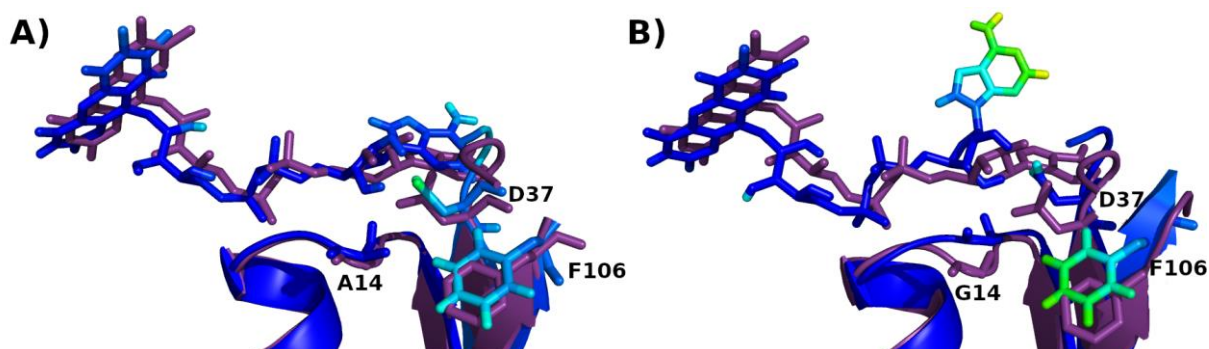


Figure C.29. Bar plot of the average fluctuations of the mutated loop and the FAD cofactor for the wildtype (orange) and G14A mutant (blue).



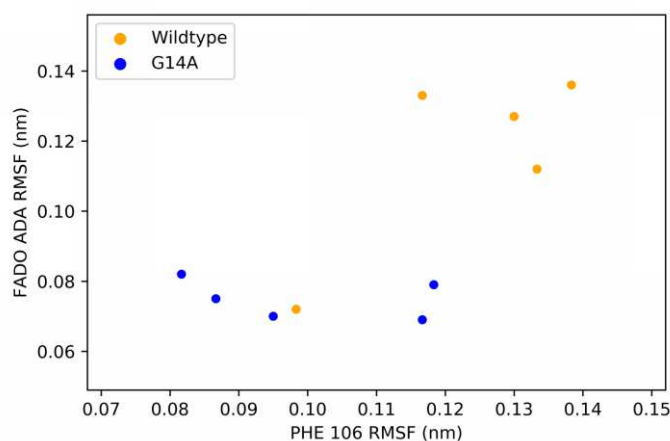
A closer examination of the trajectories indicated that the change in position of the adenosine part of the FAD seems to be initiated by the movement of the residue ASP 37. When this aspartic acid moves, the FAD backbone bends to be able to maintain two hydrogen bonds with the ASP 37. For both the wildtype and the G14A mutant, there was a higher shift of the FAD adenosine and also shifting in the position of the ASP 37 (Figure C.30).

Figure C.30. 3D representation of the average position of one of the simulations compared to the reference structure (purple) for the G14A mutant (A) and the wildtype (B). The color scale indicates the magnitude of the fluctuations, ranging from blue (low fluctuation) to red (higher fluctuation).



There is one residue, PHE 106, which is placed in a loop that is directly interacting with multiple hydrogen bonds to the loop where ASP 37 is located. The simulations in which this phenylalanine seems to move more are also the ones where the FAD drifted more from its original position Figure C.30(Figure C.31).

Figure C.31. Scatter plot of the PHE 106 fluctuations against the fluctuations of the adenosine part of the FAD cofactor, for the wildtype (orange) and the G14A mutant (blue). The obtained average fluctuations seem to cluster quite well, except for one simulation of the wildtype protein where both the PHE 106 and the FAD fluctuated less than in the rest.



The presence of an additional methyl group in the mutated variant seems to generate a steric repulsion that makes the movement of this PHE 106 more difficult, ending up in fewer fluctuations than in the wildtype protein (Figure C.32). This difference becomes more significant if the wildtype simulation that differs more from the rest (outlier) is removed (Figure C.33). It should be noted that probably, with longer simulation times, the wildtype simulation that is not fluctuating yet will start to fluctuate, since fluctuations are difficult to converge. The outlier simulation can easily be spotted in (Figure C.31).

Figure C.32. Bar plot of the average fluctuations of the Aspartic acid 37 and the Phenylalanine 106 for the wildtype (orange) and G14A mutant (blue).

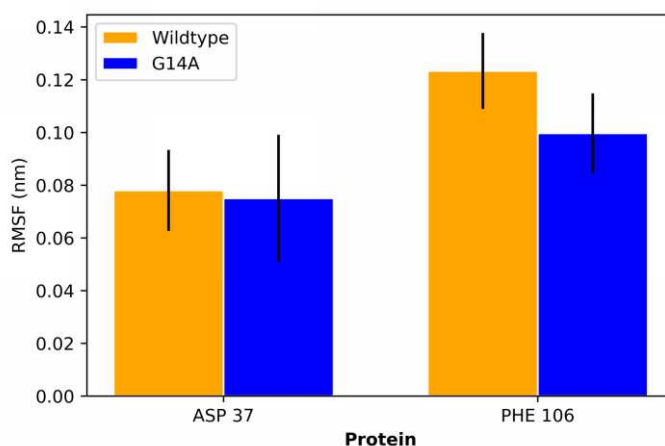
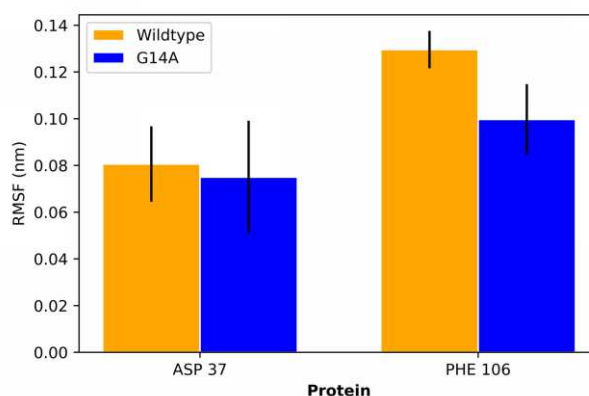


Figure C.33. Bar plot of the average fluctuations of the Aspartic acid 37 and the Phenylalanine 106 for the wildtype (orange) and G14A mutant (blue), without the outlier wildtype simulation found in figure 4.



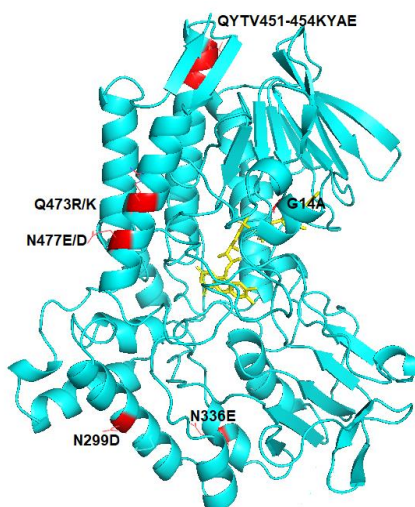
C III.2.2 Second-generation variants

The second generation was the result of a combinatorial effort. The second-generation variants were made by combining the best variants from the first generation and also adding a few more mutations, which seemed to be beneficial for the enzyme stability based on the consensus approach. So five more mutants were created and studied for their characteristic (Table C.12). Again for these variants, mutations designed to be far from the active site to avoid interference with selectivity and activity. The position of mutations was shown in a homology model of CHMO_{Acineto} using the crystal structure of TmCHMO^[56] and mutant CHMO_{Acineto}^[65] as the template (Figure C.34).

Table C.12. List of second-generation variants.

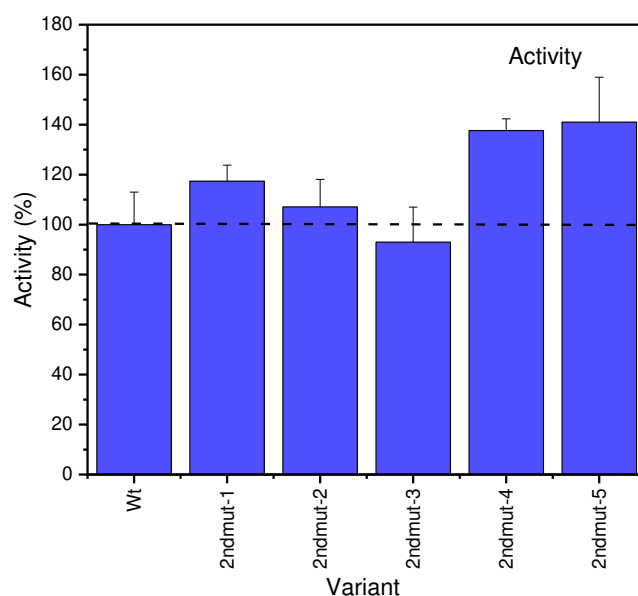
Variant	Abbreviation
G14A/Q473R/N477E	2 nd _{mut-1}
G14A/N299D/Q473K/N477D	2 nd _{mut-2}
G14A/N299D/Q473R/N477D	2 nd _{mut-3}
G14A/QYTV451-454KYAE/Q473R/N477E	2 nd _{mut-4}
G14A/N336E/QYTV451-454KYAE/Q473R/N477E	2 nd _{mut-5}

Figure C.34. CHMO_{Acineto} structure with the second-generation mutations.



The activity was the first feature, which was measured for the new variants (Figure C.35). 2nd_{mut-1} (G14A/N299D/Q473K/N477D) was the only variant that showed a decrease in the activity while the rest of the mutants showed higher activity in comparison to the wildtype CHMO_{Acineto}. 2nd_{mut-5} which is a combination of the most active and most stable mutations from first-generation and also the following mutations QYTV451-454KYAE/Q473R/N477E, showed the highest activity with 40% higher activity, which is also noticeable in the 1st_{mut-2} that means this can come either from only 1st_{mut-2} or the synergy of different mutations can be involved in this improvement as the rest of mutations in this variant also showed improvement in activity but this was not the aim to understand it, and we did not study that further.

Figure C.35. Activity measurement for second-generation variants. Enzyme activity was measured by monitoring the decrease of NADPH absorbance at 340 nm. The activity assay mixture contained 0.05 μM CHMO, 100 μM NADPH, 0.5 mM cyclohexanone in 50mM Tris HCl pH 8.5 at 30°C.



Thermodynamic stability was measured as described in the experimental section. All the variants either showed higher or same melting temperatures, which supports the data from the consensus approach in increasing the thermodynamic stability (Figure C.36). The best variant was $2^{\text{nd}}_{\text{mut-5}}$, which also showed the highest activity among all variants. This finding needed to be confirmed by measuring the kinetic stability as the fundamental value to check the real stability of the biocatalyst under reaction conditions. Fortunately, the $2^{\text{nd}}_{\text{mut-5}}$ also showed the highest kinetic stability with an almost two-fold longer half-life than wild type CHMO_{Acineto}. The second best variant, which also showed more or less the highest value in all measurements after $2^{\text{nd}}_{\text{mut-5}}$, was the $2^{\text{nd}}_{\text{mut-4}}$. The rest of the variants also showed longer half-life except $2^{\text{nd}}_{\text{mut-1}}$ and $2^{\text{nd}}_{\text{mut-2}}$ that illustrated lower kinetic stability value (Figure C.37). That is considered being an excellent achievement because, in other studies, the stability increased at the cost of diminished activity, which results in a poor biocatalyst for industrial applications. In our study, we could successfully create variants that showed higher stability while they either have kept their activity or even shows more activity that cannot observe in the other studies.

Figure C.36. Thermodynamic stability for the second-generation of variants. Thermodynamic stability was measured by nano differential scanning fluorimetry (nanoDSF) using 2 mg mL⁻¹ enzyme in 50 mM Tris HCl, 10 μM FAD, pH 8.5. The standard deviation is less than 3% for all thermodynamic stability measurement.

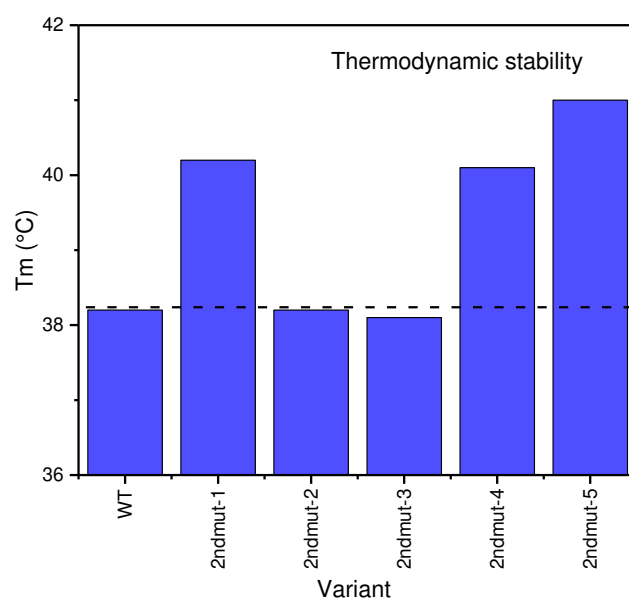
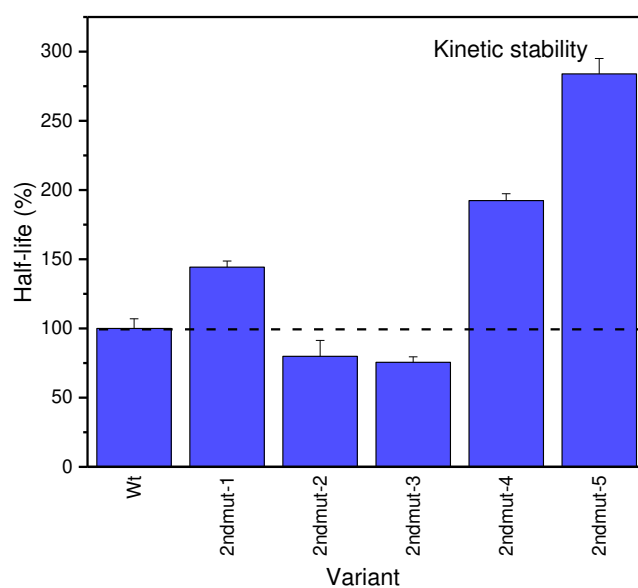


Figure C.37. Kinetic stability for second-generation variants. Kinetic stability was measured by incubation of 1 μM isolated enzyme at 30°C in 50 mM Tris HCl containing 10 μM FAD, pH 8.5.



C III.2.3 Third-generation variants

After evaluating the second-generation variants, the best mutations were identified. The best variants considered to be 2nd_{mut-4} and 2nd_{mut-5}, and they have been used as the basis for the creation of the third-generation library. Thus we have used the combinatorial method and introduced the two of best literature known mutation, which confirmed to be beneficial for increasing the stability, to our best variants in the second-generation library. This combination resulted in the generation of the third library, which consists of 5 more mutants (Table C.13). The position of these mutations shown in the homology model of CHMO_{Acineto}, which made using TmCHMO [56] and mutant CHMO_{Acineto} [65] as a template (Figure C.38).

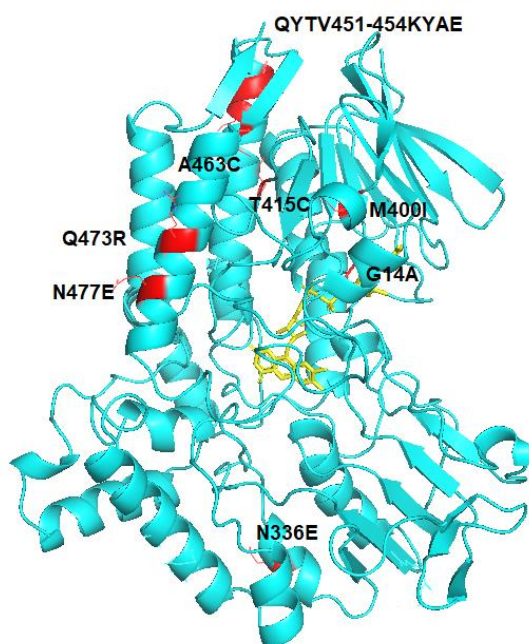
The first literature known mutation (T415C/A463C) that we decided to add to our best variant was introduced by Schmidt *et al.* [146] This mutant consists of two cysteine substituents and was designed to enable a disulfide bond in the enzyme structure. Interestingly the disulfide bond could not be established due to a misconception in the distance between the two cysteines. The flexibility and movement of the BVMO structure makes accurate predictions difficult. Nevertheless, the mutant showed an increased kinetic stability by about three fold. A possible explanation for the increased half-life of this mutant, is the potential of cysteines to function as antioxidant groups. By scavenging reactive oxygen species the enzyme will be protected and avoid structural damage, which causes a loss of catalytic function. [146]

The second mutation (M400I) that was considered to be beneficial for our purpose, was designed by Opperman and colleagues. [80] This mutation was introduced based on the fact that oxidation of Met residue at remote positions can lead to altered structural conformations, which can lead to the lower stability of enzymes due to faster unfolding. [192] Furthermore, oxidation of Met residues can decrease enzyme activity when Met residue located at the active sites of biocatalyst [80]. Thus substitution of Met residues, which can be easily affected by oxidation, with amino acids that are less sensitive to oxidation may lead to higher oxidative and thermal stability of enzymes. [80]

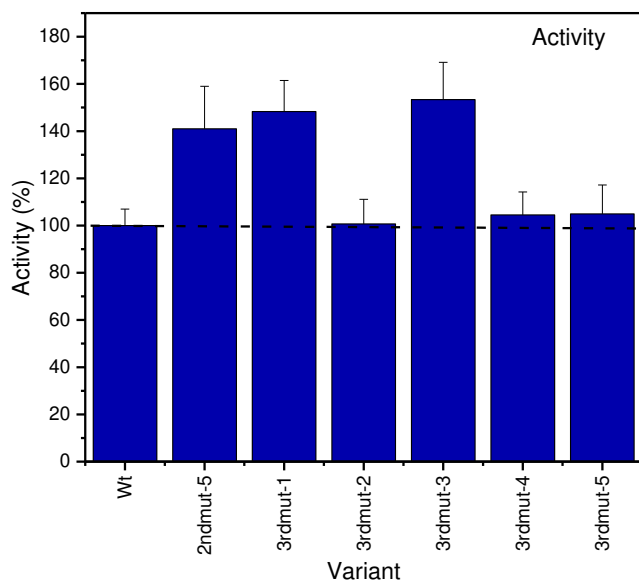
Table C.13. List of third-generation variants.

Variant	Abbreviation
QYTV451-454KYAE+Q473R+N477E+G14A+T415C+A463C	3 rd _{mut-1}
QYTV451-454KYAE+Q473R+N477E+G14A+M400I	3 rd _{mut-2}
QYTV451-454KYAE+Q473R+N477E+G14A+N336E+T415C+A463C	3 rd _{mut-3}
QYTV451-454KYAE+Q473R+N477E+G14A+N336E+M400I	3 rd _{mut-4}
QYTV451-454KYAE+Q473R+N477E+G14A+ N336E+ T415C+A463C+ M400I	3 rd _{mut-5}

Figure C.38. CHMO_{Acineto} structure with third-generation mutations.



All the variants in the third-generation library showed either the same activity as the wild type enzyme or improved activity (Figure C.39). 3rd_{mut-1} and 3rd_{mut-2} showed more than 50% higher activity while the rest of the variants showed almost the same activity as wild type CHMO_{Acineto}. This improvement was so far the highest in all of our variants and cannot observe in the literature for the BVMO enzymes.

Figure C.39. Activity measurement of the third-generation of variants.

Both thermodynamic and kinetic stability measured for all the new variants. (Figure C.40, Figure C.41) All the mutants illustrated higher T_m value in comparison to the previous mutants, which means the combination of mutation did increase the thermodynamic stability. The highest melting temperature observed in 3^{rd}_{mut-5} (QYTV451-454KYAE+Q473R+N477E+G14A+ N336E+ T415C+A463C+ M400I), which is the combination of the best variant from 2nd generation mutants with all literature known mutations that we used in this study. The melting temperature in this mutant increased by 5 °C from 38 to 43 that is among the highest improvement achieved for the CHMO_{Acineto} enzyme by protein engineering. However, as discussed before, the more valuable property of the enzyme for us is kinetic stability, which was measured by determining the half-life for all variants (Figure C.41). The measurement showed improvement in all the variants. 3^{rd}_{mut-1} showed the lowest increase in half-life by more than 2-fold improvement, and 3^{rd}_{mut-4} showed the highest increase in the half-life by more than 8-fold. This achieved by the combination of 2^{nd}_{mut-5} from second generation mutants and M400I from literature. This combination had a tremendous synergetic effect as 2^{nd}_{mut-5} showed only a 2-fold higher half-life that reached to 8-fold after combination with M400I. One other thing which makes this finding more interesting is the fact that this improvement did not result in loss of activity. To do further studies 3^{rd}_{mut-4} as the most thermostable variant was chosen.

Figure C.40. Thermodynamic stability for the third-generation of variants. Thermodynamic stability was measured by nano differential scanning fluorimetry (nanoDSF) using 2 mg mL⁻¹ enzyme in 50 mM Tris HCl, 10 μM FAD, pH 8.5. The standard deviation is less than 3% for all thermodynamic stability measurement

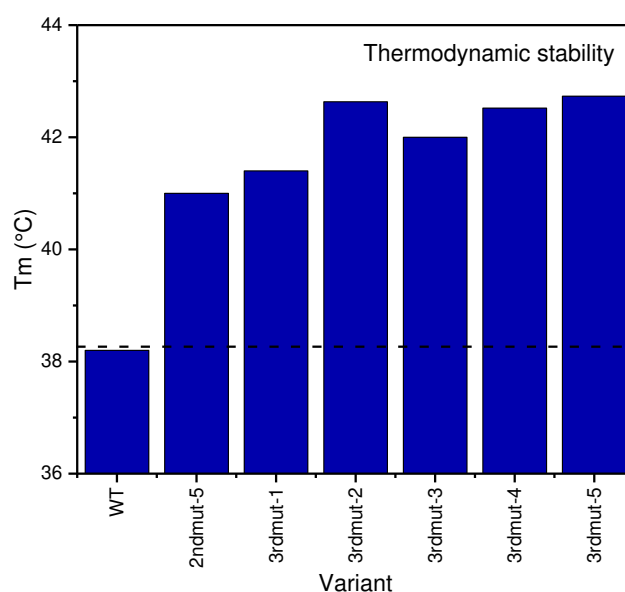
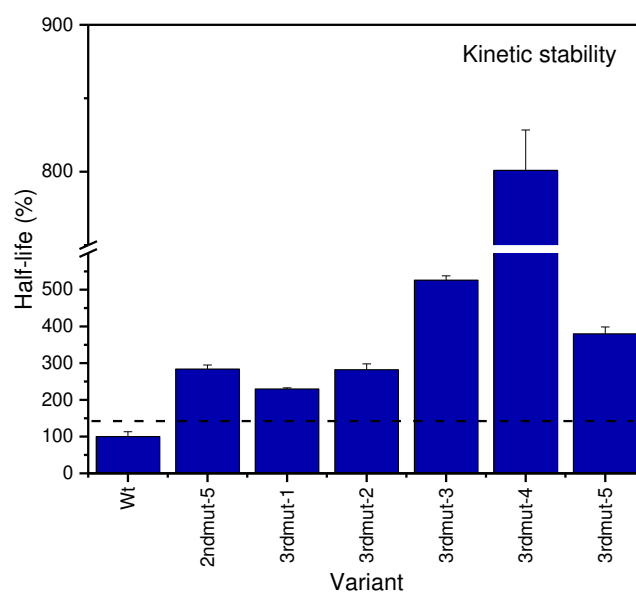


Figure C.41. Kinetic stability for the third-generation of variants. Kinetic stability was measured by incubation of 1 μM isolated enzyme at 30°C in 50 mM Tris HCl containing 10 μM FAD, pH 8.5.



It is crucial to understand more about the variants and also their stability under process conditions. Therefore, the thermodynamic stability (Figure C.42) and kinetic stability (Figure C.43, Figure C.41) of wildtype CHMO_{Acineto} and its mutants 3rd_{mut-4} as the most stable variant in our library was evaluated at different pH. As can be seen in Figure C.42, both wildtype CHMO_{Acineto} and 3rd_{mut-4} showed the same trend of stability in different pH except for the pH 5.5, which caused more decrease in the stability of 3rd_{mut-4} than the wildtype. Figure C.42 shows that the wildtype and variant are more stable at pH 6.5-8.5. The pH 7.5 seems to be the optimum pH for the enzyme to perform the reaction for the longer time, and pH 10.5 is least favorable for the stability of both variants. The same behavior can be observed in Figure C.43 regarding kinetic stability. It can clearly see that the high pH (9.5-10.5) are so destructive for both wildtype and 3rd_{mut-4}. These variants are not appropriate choice to do reactions which are occurring in this pH range. What is important to mention is the high stability of 3rd_{mut-4} in pH 5.5 ($t_{1/2}$ 155 min), while wildtype CHMO_{Acineto} will lose half of its activity after 16 min. This means that to carry on the reaction in such a low pH, 3rd_{mut-4} would be a better option than the wildtype enzyme.

Figure C.42. pH effect on the thermodynamic stability of 3rd_{mut-4}. Thermodynamic stability was measured by nano differential scanning fluorimetry (nanoDSF) using 2 mg mL⁻¹ enzyme, 10 μ M FAD.

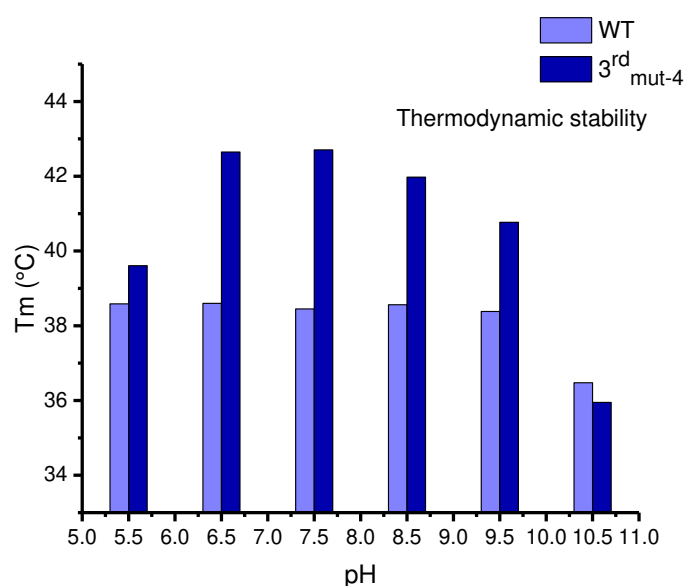
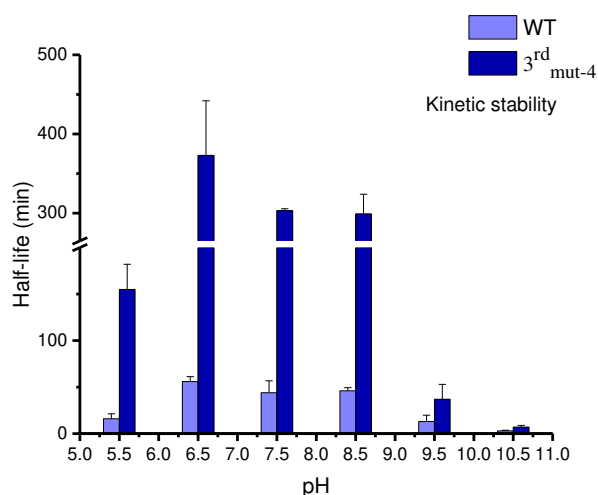


Figure C.43: pH effect on kinetic stability of 3rd_{mut-4}. Kinetic stability was measured by incubation of 1 μ M isolated enzyme at 30°C with 10 μ M FAD.



The effect of different co-solvents on the thermodynamic and kinetic stability of wild type CHMO_{Acineto} and 3rd_{mut-4} was also investigated (Figure C.44, Figure C.45). The addition of 5% dioxane, methanol (MeOH), ethanol (EtOH), acetonitrile (MeCN) and THF was compared to the curve obtained without co-solvent in 50 mM Tris·HCl pH 8.5, 30 °C. It was observed that both the wild type enzyme and 3rd_{mut-4} indicate similar behavior towards the co-solvent in the reaction condition. MeCN, dioxane, and THF have displayed the most destructive effect on the stability, and the most suitable co-solvents to use for both variants are MeOH and EtOH. It is also worth to mention that again, there is a difference in the result obtained from thermodynamic stability and kinetic stability as dioxane showed less destructive effect than MeCN in the T_m value of wild type CHMO_{Acineto} and 3rd_{mut-4} while it is the other way around when determining half-life measurement. This was not totally surprising, as there is no direct correlation between the thermodynamic stability and kinetic stability, as seen before.

Figure C.44. Solvent effect on the thermodynamic stability of 3rd_{mut-4}. Thermodynamic stability was measured by nano differential scanning fluorimetry (nanoDSF) using 2 mg mL⁻¹ enzyme in 50 mM Tris HCl, 10 μM FAD, pH 8.5, 5%cosolvent.

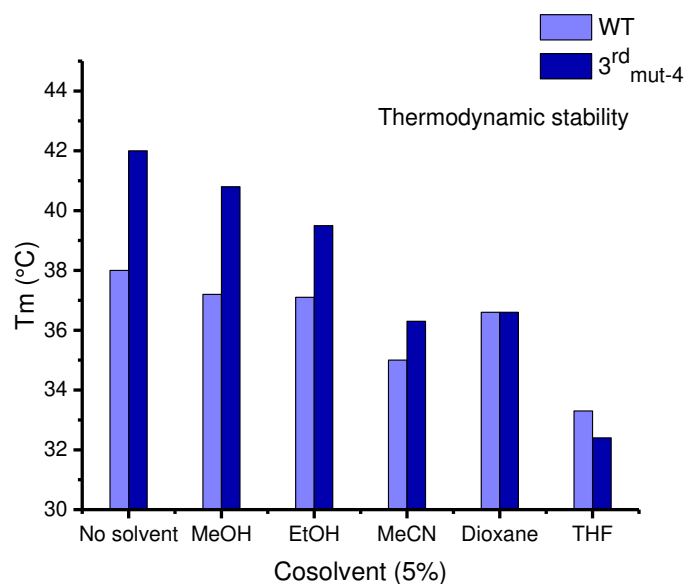
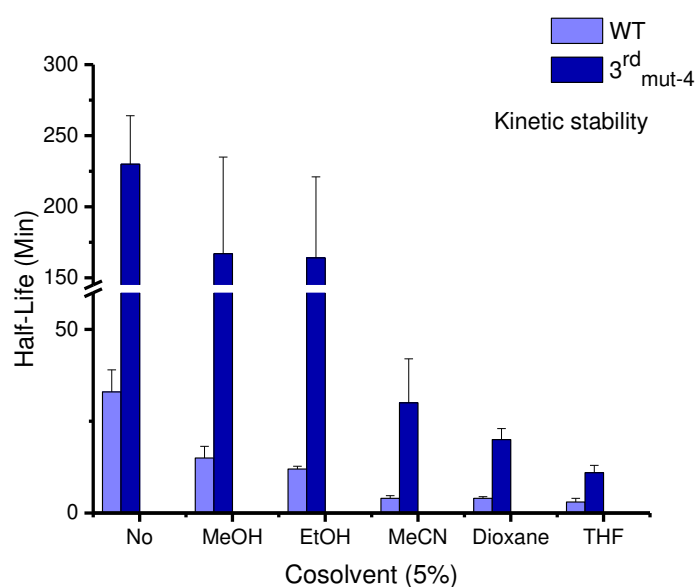


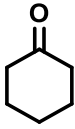
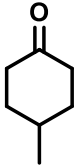
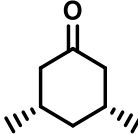
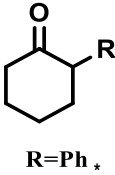

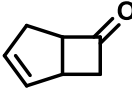
Figure C.45. Solvent effect on kinetic stability of 3rd_{mut-4}. Kinetic stability was measured by incubation of 1 μM isolated enzyme at 30°C in 50 mM Tris HCl containing 10 μM FAD, pH 8.5, 5% cosolvent.



C III.2.3.1 Enantioselectivity of third-generation variants

A final critical aspects in the stabilization efforts needed to be addressed, and this was the effect of the mutations on the enantioselectivity of the enzymatic oxidation step. So the selectivity of our third generation variants toward several cyclohexanones and cyclobutanone was determined as described in the experimental part. The results are summarized in Table C.14. The results indicate that the mutations did not change the enantioselectivity of our variants. This was in part expected as the mutations are far from the catalytic site of the enzyme. The only difference, which is worth mentioning is the decrease in the enantioselectivity of 3rd_{mut-3} toward 3-(4-Chlorophenyl) cyclobutanone. This variant shows less enantioselectivity toward S-product by 49% than the wild type enzyme, which shows 76(S) ee. However, this substrate class displays high susceptibility on stereochemical outcome, so actually is was partly expected to identify any effects most probably for such structures.

Table C.14. Substrate profile for selected variants. Biotransformation carried out by using the resting cell approach under nongrowing condition.

										 R=Ph *			 R=Cl-Ph					
	Con	Con ^A	ee ^B	Con	ee	Con	ee	Con	ee	Con	ee	Con	ee	Con	ee			
WT	+++	+++	99(S)	+++	99 (4S,6R)	+	95 (R)	+	76 (S)	+++	N:ABN ^C 51:49 95(-):99(-)							
2 nd _{mut-5}	+++	+++	99(S)	+++	99 (4S,6R)	+	95 (R)	+	72 (S)	+++	N:ABN 51:49 95(-):99(-)							
3 rd _{mut-1}	+++	+++	99(S)	+++	99 (4S,6R)	+	95 (R)	+	76 (S)	+++	N:ABN 51:49 95(-):99(-)							
3 rd _{mut-2}	+++	+++	99(S)	++	99 (4S,6R)	+	95 (R)	+	80 (S)	+++	N:ABN 51:49 95(-):99(-)							
3 rd _{mut-3}	+++	+++	99 (S)	++	99 (4S,6R)	+	95 (R)	+	49 (S)	+++	N:ABN 51:49 95(-):99(-)							
3 rd _{mut-4}	+++	+++	99(S)	++	99 (4S,6R)	+	95 (R)	+	74 (S)	+++	N:ABN 51:49 95(-):99(-)							
3 rd _{mut-5}	+++	+++	99 (S)	++	99 (4S,6R)	+	95 (R)	+	75 (S)	+++	N:ABN 51:49 95(-):99(-)							

^ARelative conversion (Conv) of substrate to product + < 50% ++ 50-80% +++ > 80%,^BEnantiomeric excess (ee) of product^CN:ABN ratio of normal to abnormal lactone

C III.2.3.2 Kinetic measurement

The other property of the enzyme which can be altered after the mutation is the enzyme affinity towards the substrate and cofactor. These features are essential, especially for the application of the

enzyme for industry. For this purpose, the Michaelis constant (K_m) of the 3rd generation variants toward the substrate (cyclohexanone) and also the catalytic constant (k_{cat}) were measured. The results summarized in **Table C.15**.

Enzyme kinetics is the study of the chemical reaction, which is catalyzed by the enzyme. Studying the enzyme kinetics will help to investigate the effects of variables in the conditions of the reaction and their effect on the enzyme reaction rate and also understand the enzyme performance better. Three main enzyme kinetic values are Michaelis constant (K_m), catalytic constant (k_{cat}) and k_{cat}/K_m . The K_m is defined as the substrate concentration at which the enzyme activity is half of its maximal value. It generally provides an estimation of the affinity of the enzyme for a substrate.^[193] The higher K_m is, the more substrate is needed for the enzyme to reach the maximum reaction rate. k_{cat} is a measure of how many substrates one enzyme can convert into a product per second. The k_{cat}/K_m ratio has been widely used as a measure of the performance of an enzyme.^[193] Different studies indicate that an enzyme displaying a higher k_{cat}/K_m value will, at specific substrate concentrations, perform a reaction at higher rates than an enzyme having a lower k_{cat}/K_m .^[193], but it is essential to mention that when comparing an enzyme acting on two different substrates, a comparison of k_{cat}/K_m parameters does not provide an unequivocal diagnosis of efficiency. It is essential to know that the use of k_{cat}/K_m relationships as efficiency assistance can also be misleading in acquiring conclusions about the performance of industrial enzymes that generally operate under extreme and non-natural conditions, including a high substrate and/or product concentration.^[193-194] It is also worth to mention that the actual reaction rate of an enzyme does not only depends on the value of k_{cat}/K_m but also on substrate concentration [S], product concentration [P] (hence, prospective inhibition effect) and its dissociation constants (K_d).^[193]

Results illustrate that all variants are showing the same tendency toward cyclohexanone as the natural substrate as they all show similar K_m values (**Table C.15**). The same trend was seen in regard to catalytic efficiency of the variants (k_{cat}). We have evaluated the affinity of wildtype CHMO_{Acineto} and 3rd_{mut-4} toward NADPH as the cofactor was investigated as well. We again observed almost the same result for both variants, which means that the mutations were not affecting the catalytic efficiency or affinity toward the substrate (**Table C.15**).

Table C.15. Kinetic measurement. Catalytic rates (K_m , k_{cat}) observed upon incubation of the isolated enzyme with varying amounts of the substrate (cyclohexanone) fitted to the Michaelis-Menten equation.

	K_m	StDEv	k_{cat}	StDEv	k_{cat}/K_m	Ratio
Cyclohexanone	μM		$[\text{s}^{-1}]$		$[\text{mM}^{-1} \text{s}^{-1}]$	
WT	6.74	± 2	15.0	± 1.3	2220	1
2 nd _{mut-5}	3.97	± 0.76	13.8	± 1.5	3499	1.5
3 rd _{mut-1}	8.53	± 0.15	18.7	± 1.7	2191	0.9
3 rd _{mut-2}	3.37	± 0.07	11.8	± 1.8	3527	1.5
3 rd _{mut-3}	4.60	± 0.4	17.9	± 5.4	3902	1.7
3 rd _{mut-4}	5	± 1.5	16.3	± 1.5	3187	1.4
3 rd _{mut-5}	4.5	± 1.4	13.7	± 2.3	2941	1.3
NADPH						
WT	1.67	± 0.17	13.07	± 1.1	7827	1
3 rd _{mut-4}	2.15	± 0.6	15.01	± 2	7039	0.9

We tried to rationalize why 3rd_{mut-4} indicating high stability and to understand what could be the reason, we had a close look in the the K_d . K_d evaluation helps to clarify and understand the mechanisms, which are involved in stabilization of the 3rd_{mut-4}, we did analyze the enzyme affinity toward FAD (Table C.16) that can alter the stability and also the structure of this variant, especially the surface electrostatic potential. The K_d measurement indicates that 3rd_{mut-4} has a higher affinity toward FAD (11-fold) in comparison to the wildtype enzyme. This can come from mutation G14A, which is one of the alterations in 3rd_{mut-4} and can be one of the reasons this variant shows significantly higher stability than wildtype.

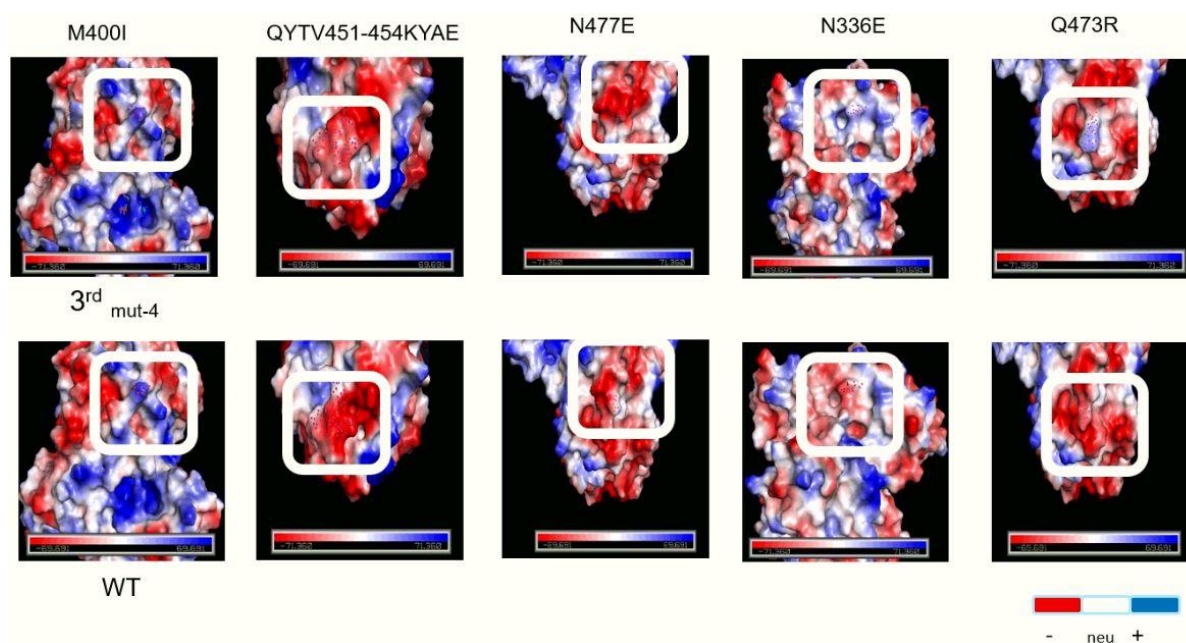
Table C.16. Kd measurement

Variant	Kd (μM)	StDEv
Wt	1.6	0.06
3 rd _{mut-4}	0.14	0.02

Next, the surface electrostatic potential of 3rd_{mut-4} was studied to see if there is a change in this property. This is because the electrostatic potential is affecting the enzyme-enzyme interaction, enzyme-environment interaction and enzyme folding and all these can affect the enzyme stability. Analyzing the structure of the 3rd_{mut-4} variant as the most stable mutant and comparing the electrostatic potential at the surface with the wildtype CHMO_{Acineto}, we did find that some of the mutations involved in 3rd_{mut-4} are changing this value in different extent (Figure C.46). As mentioned above, changes in electrostatic potential can affect the stability of enzyme. So one possible reason for altered thermostability can be modifications in electrostatic potential in the surface of the mutant.

M400I did show no effect on the electrostatic potential. Mutation QYTV451-454KYAE did small change of potential from negative to the positive. The changes were more prominent and more evident in the case of N477E, this mutation caused the electrostatic potential of a significant portion of the area in its surroundings to change from neutral to negative. As is shown in Figure C.46, mutations N336E and Q473R had the highest effect in the alteration of surface electrostatic potential. N336E caused a total change of its surrounding surface potential from negative to positive in a significant area of the surface. The same also happened when Q in position 473 changed to R. In the wildtype CHMO_{Acineto} the potential in this position is negative, while the introduction of Q473R mutation altered this potential completely to positive. These changes can results in altered protei-protein interaction or interaction of 3rd_{mut-4} with the reaction solution that are both affecting the enzyme stability.

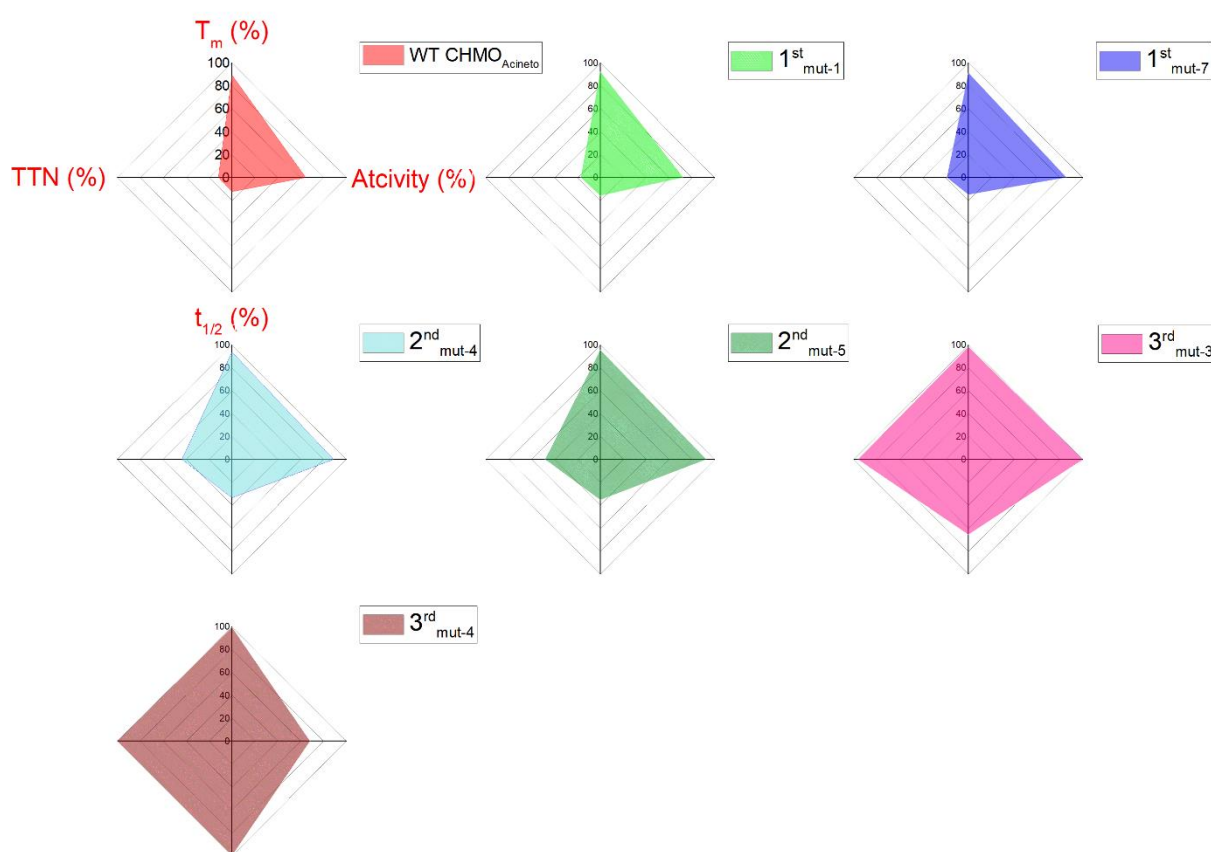
Figure C.46. Comparison of the surface electrostatic potential of original amino acids and mutated amino acids in most stable variant ($3^{\text{rd}}_{\text{mut-4}}$). The calculation and visualization was performed using PyMOL.



C III.3 Discussion of the protein engineering approach

We started with unstable WT CHMO_{Acinetobacter} and aimed to improve stability. The other properties like activity and also catalytic efficiency were determined (Figure C.47). The properties were slightly increased in the first-generation library by 30-40% for total turnover number, kinetic stability, and activity. The thermodynamic stability was increased by 1.4 °C (Figure C.47). These improvements were boosted up to 400% for TTN, 180% higher kinetic stability, 2 °C more T_m value, and 40% higher activity within the next generation (Figure C.47). This was achieved by the synergic effect of combining the most stable variant of the first generation library (G14A), most active one (N336E), and most catalytically efficient variant (QYTV451-454KYAE). This shows the combinatorial approach helped to improve the properties and helped to use the synergy of mutations. The characteristics were boosted again by using the combination of the best of 2nd generation by literature known hits. The properties were improved by 800% for TTN and kinetic stability. The thermodynamic stability was improved from 38.2 °C to 42.7 °C that is the second highest improvement of thermodynamic stability in the family of BVMOs. These improvements achieved by no loss of activity which is the substantial novelty of this study as in other attempts the activity was diminished by the improvement of stability.

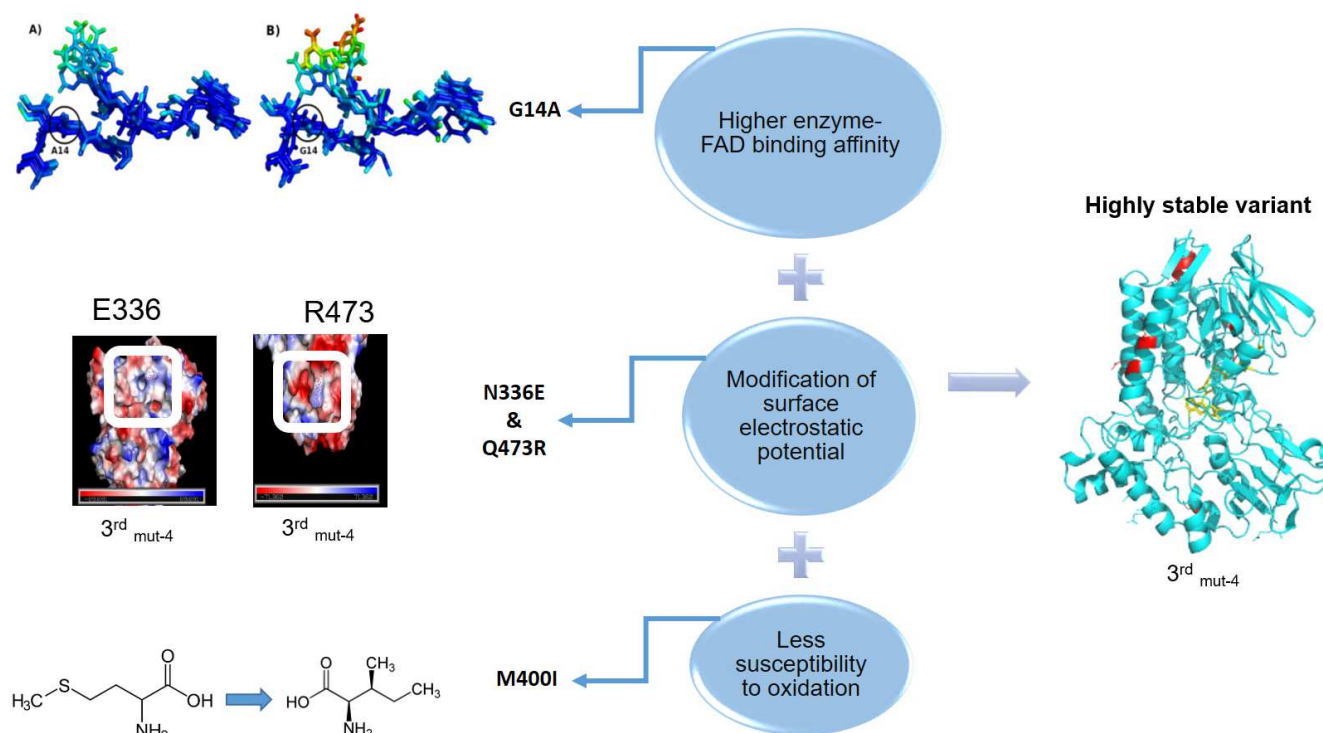
Figure C.47. Characterization of best variants. Kinetic stability of 1 μ M isolated CHMO at 30 °C in 50 mM Tris HCl buffer, pH 8.5. Thermodynamic stability measured by nano differential scanning fluorimetry (nanoDSF): 50 mM Tris HCl, 10 μ M FAD, 2 mg mL⁻¹ enzyme. Enzyme activity was measured by monitoring the decrease of NADPH absorbance at 340 nm. Standard assays contained the enzyme (0.05 μ M), cyclohexanone (0.5 mM), and NADPH (100 μ M) in 50 mM Tris HCl pH 8.5. TTN values were obtained from the exponential fit of catalytic enzyme activity under turnover conditions



Conclusion

Understanding the mechanism behind the stabilization can be as important as improving the properties as well. By understanding the mechanism, we can generalize the finding in other BVMO family members. Several mechanisms were found to be involved in the stabilization of our best variant ($3^{\text{rd}}_{\text{mut-4}}$). The biggest contributor to higher stability is probably the replacement of glycine with alanine in position 14 (Figure C.48). This mutation increased the enzyme-FAD binding affinity that resulted in higher stability. This was confirmed by measuring the K_d value and the mechanism explained nicely by MD simulation. The other reason is the presence of M400I mutation that decreased the oxidation susceptibility of enzyme in the remote position 400 (Figure C.48). Last but not least, the changes in the surface electrostatic potential resulted in higher stability. The changes in potential resulted in alteration of enzyme-enzyme and enzyme-reaction solution interaction (Figure C.48). These change can help enzyme first to probably avoid the distractive effect of radicals in enzyme reaction solution and probably helped enzymes to interact higher and we know the more enzymes are packed together the more stable they can be.

Figure C.48. Mechanisms involved in stabilization of most stable variant $3^{\text{rd}}_{\text{mut-4}}$.



D Conclusion and perspective

D I Summary

This project aimed at improving the features of BVMOs, primarily kinetic stability, a drawback that is preventing the industrial application of such impressive biocatalysts, so far. Different approaches like protein engineering, enzyme discovery, and also enzyme redesign have used to reach this aim.

The first project (Enzyme discovery): The rise of Next Generation Sequencing has generated an incredible pool of gene and protein sequences, which deposited in databases like the National Center for Biotechnology Information (NCBI), and this pool is continuously increasing. These valuable sequences turn databases to gold mines for discovering novel enzymes, which can perform new reactions or enzymes with better characteristics. Due to having access to such a comprehensive and valuable resource, we did try to find new BVMOs with higher stability and also better substrate profile. So, the newly discovered TmCHMO, which is a stable BVMO and also BVMOs fingerprint sequences, was used as a template to search through the metagenome mine and find novel BVMOs. By this approach, we could identify a novel BVMO from *Amycolaptosis thermoflava*, which is a semithermophile bacterium, which showed high sequence similarity with TmCHMO, so it was assumed to be a stable BVMO. This novel BVMO was expressed and purified for characterization. Based on the sequence and phylogenetic tree analysis, this novel BVMO (BVMO_{Flava}) belongs to type BVMOs group 3, and it was placed in the same clade as that of TmCHMO and CHMO_{Acineto}. As expected, based on the sequence similarity, a similar substrate profile for all three enzymes was determined. Even though BVMO_{Flava} originates from a thermophilic bacterium, the kinetic stability (half-life) at slightly elevated temperatures dropped from 72 (at 30 °C) to 6 min at 40 °C.

In contrast, the thermodynamic stability (T_m value) was comparable to that of TmCHMO and significantly higher than that of CHMO_{Acineto}. This apparent deviation between thermodynamic and kinetic stability is a big problem in the field, since often only thermodynamic stabilities are published, without any context to the actual operational stability of the new catalyst. This finding is in agreement with this critical discussion that there is no direct correlation between thermodynamic and kinetic stability, which needs to consider for future comparison and putative industrial applications of BVMOs.

The second project (Redesign): The capacity to differentiate between optical isomers is vital for living systems. The selective production of only one product stereoisomer from achiral substrates is a sophisticated task for enzymes. Since stereoselectivity is very crucial in biology, the mechanistic and

Conclusion

structural basis underlying these phenomena has become the focus of intensive studies. With the improvements in protein engineering and, especially, rational computational design, researchers start to provide various approaches and tools to manipulate the stereoselectivity of enzymes. This study shows that the TmCHMO, and the two mutants LGY3-4-E5 and LGY3-4-D11, are outstanding biocatalysts in the asymmetric conversion of a variety of structurally diverse ketones. The two variants generated by our collaboration partner using directed evolution utilizing iterative saturation mutagenesis (ISM). As demonstrated by the determination of thermodynamic stability (T_m), the stability of the two TmCHMO variants retained relative to WT TmCHMO.

The high enantioselectivity in desymmetrization reactions is similar to those reported for CHMO_{Acineto}, but this prototype BVMO is suffering from the lack of stability. Furthermore, the reversal of enantioselectivity, as reported herein, allows access to some chiral compounds of particular synthetic importance. This study shows that TmCHMO is an attractive and potential enzyme for future mutagenesis.

These studies did provide a better understanding of BVMOs on the molecular level and facilitated the chiral Baeyer Villiger bio-oxidation to become an easily applicable tool in stereoselective synthesis. Besides, novel indications on the critical structural areas of high impact for biocatalyst efficiency are expected for BVMOs, in general, providing a better understanding of this fascinating enzyme family.

Third approach (main attempt): Certain combinatorial (semi-rational) and rational protein design approaches applied for the development of several generations of mutants. These were used to improve both the activity and stability of our enzyme of interest (CHMO from *Acinetobacter*). Parallel screening methods both on whole-cells as well as purified proteins were utilized to assess variants performance. Rational design and semi-rational design helps to limit the size of variants library that needs to be screened, which at the end leads to decrease the time, effort, and material needs for screening. Position 14 was indicated as a hot spot and spotted to be altered using the rational design. Rationally designed 1st_{mut-1} (G14A) showed higher activity, kinetic and thermodynamic stability and, as a successful mutation, did apply for the further round of mutagenesis. Molecular dynamics simulations explained these experimentally proven improvements by 1st_{mut-1}. The simulation indicated these alterations are coming from the effect of enzyme-cofactor binding modifications, which happened after mutation in this position. Also, several mutants, which predicted by structure-guided approach (consensus approach), have been created and studied. These mutations aimed to stabilize helical structure motifs and predicted by sequence comparison with the structure of phenylacetone monooxygenase (PAMO) from *Thermobifida fusca*. These modifications were designed to be placed far enough from the active site, more than 6Å, so that they do not affect the activity of the enzyme or

Conclusion

with its affinity for the substrates. After the first round of site-directed mutagenesis based on the structure-guided consensus concept, an improvement of thermodynamic stability, kinetic stability, and activity of CHMO_{Acineto} obtained. Almost all the variants showed better characteristics except for a few of them. After successive screenings at the given conditions, finally, the three best performing, thermostable mutants were selected. We created a second generation of mutants by a combination of best-performing candidates of the single-site modification study. The best variants from the first experiments were chosen and did go through another round of mutagenesis (second-generation library). These new mutants also studied for their characteristics. Again the most stable and active mutants did chose for further improvement. In the last stage (Third-generation), we did combine the best mutations from first and second generations with literature known mutations, which experimentally proved to be beneficial to increase the stability of CHMO_{Acineto}. All the third generation variants showed significant improvement of thermodynamic and kinetic stability and either showed higher or similar activity in comparison to wild type CHMO_{Acineto}. This finding is quite essential as we could always keep the activity at the same level or even higher in comparison to wild type enzyme and increasing the stability as well, which is unprecedented in previous literature to the best of our knowledge, so far. This leads to an increase in the TTN and makes the variants great biocatalysts for large scale industrial applications. Variant 3rd_{mut-4} showed the most significant improvement concerning stability and also TTN, which made it interesting for further studies. The effect of different pH and also co-solvent in the stability of this variant investigated. Also, all kinetic values were measured for this variant to study its importance and potential for industrial application. This variant showed almost the same behavior to different pH and also co-solvent when compared with wild type CHMO_{Acineto}. Kinetic measurements also revealed that this variant shows an almost same tendency toward substrate and NADPH as a cofactor in comparison to wild type. This was not the case when the affinity of 3rd_{mut-4} was checked toward the FAD. 3rd_{mut-4} showed an 11-fold higher affinity toward FAD, which can be the reason why this variant is significantly more stable than the wild-type enzyme.

D II Perspective

In the field of enzyme discovery and especially for BVMOs so far, studies suggest that, on average, roughly one out of two microbial genomes contains a BVMO gene, which means that there is a vast potential for novel BVMOs to be discovered. Now more than 400 novel Type I BVMOs genes are present in the genome sequence database. This large quantity of BVMO sequences has a great potential to contain BVMOs with a larger or altered substrate profile and also significantly increased stability, which makes these novel biocatalysts potential candidates to be used for industrial applications.

Regarding the TmCHMO redesign, it will be exciting to investigate and see how TmCHMO and the two generated mutants will perform in Baeyer-Villiger reactions of other ketones, and also in enantioselective sulfoxidation. Further iterative saturation mutagenesis (ISM) studies starting from WT, or the two variants may also be a rewarding task in future studies. Last but not least, a theoretical study, may hopefully be unveiling the molecular basis of stereoselectivity, would not only be of substantial interest, but it could also guide future directed evolution of this promising enzyme.

The results acquired from the protein engineering approach (Consensus approach) indicate how useful this method can be for the improvement of enzyme characteristics, especially stability. This study also helped to determine some putative hot spots in CHMO_{Acineto}, which can function as templates for further protein engineering studies in related BVMO class enzymes. The most interesting hot spot can be position 14, which is involved in the enzyme FAD binding properties. As it is already mentioned this property of the enzyme is quite important and has a direct relation with the stability. By improving the enzyme FAD binding, we can improve the stability of CHMO_{Acineto} and probably other BVMOs, that makes this hot spot discovery more important. Also, MD simulations suggested that position ASP37 and PHE106 are involved in the FAD binding and can be a critical position that can be investigated and altered by rational protein engineering.

E Experimental part

E I Materials and methods-standard microbiological techniques

Unless otherwise notified, all chemicals, reagents, and enzymes were purchased from commercial suppliers and used without any further purification. All the glass equipment and plastic consumables were either sterilized before use by autoclaving (121 °C, 20 min, >15 psi) or were sterile upon purchase.

E I.1 General stock solutions

All the aqueous stock solutions prepared using distilled water and either sterilized by autoclave or using the 0.2 µm cellulose syringe filter whenever applicable. All the stock solutions were kept at -20 °C and added before use. The stock and working concentration shown in Table E.1.

Table E.1. Stock solutions.

Reagent	Stock concentration (dH ₂ O)	Working concentration
IPTG	100 mM	0.1 mM
Amp	100 mg mL ⁻¹	100 µg mL ⁻¹
Kan	50 mg mL ⁻¹	50 µg mL ⁻¹

E I.2 Media Preparation

Quantity of reagents refers to the preparation of 400 ml medium unless noted otherwise. All the media were stored at RT in the dark. Once the medium bottle was opened, it was stored at 4 °C and visually checked for the possible contamination. Terrific broth (TB) and Lysogeny broth (LB) were both employed for bacteria cultivation, enzyme expression, and biotransformation (Table E.2).

Table E.2. Medium preparation.

LB	Quantity (g)	TB	Quantity (g)
Bacteriological peptone	4	Bacteriological peptone	4.8
Yeast extract	2	Yeast extract	9.6
NaCl	4	K ₂ HPO ₄	5
		KH ₂ PO ₄	0.91
Fill up to 400 mL dH ₂ O			
Autoclaved at 121 °C for 20 minutes			

The antibiotic was added after autoclave and when the solution reaches to RT.

E I.3 Strain cultivation

To grow *E. coli* strains 4 mL LB medium were inoculated with 10 µL of cryo stock of respective strain, supplemented with the antibiotic(s) if applicable and incubated at 37°C with shaking (200 rpm; InforsHT Multitron 2 Standard) for 12–24 h. From this culture, cells were streaked on LB agar plates (LB-Miller medium containing 1.5 % (w/v) bacteriological agar in standard Petri dishes [94x16 mm] supplemented with an antibiotic if applicable). Plates were incubated upside down at 37 °C for 12–24 h (Heraeus Function line incubator). Agar plates were stored in the dark at 4 °C. *E. coli* strains were propagated every 4–6 weeks onto freshly prepared agar plates.

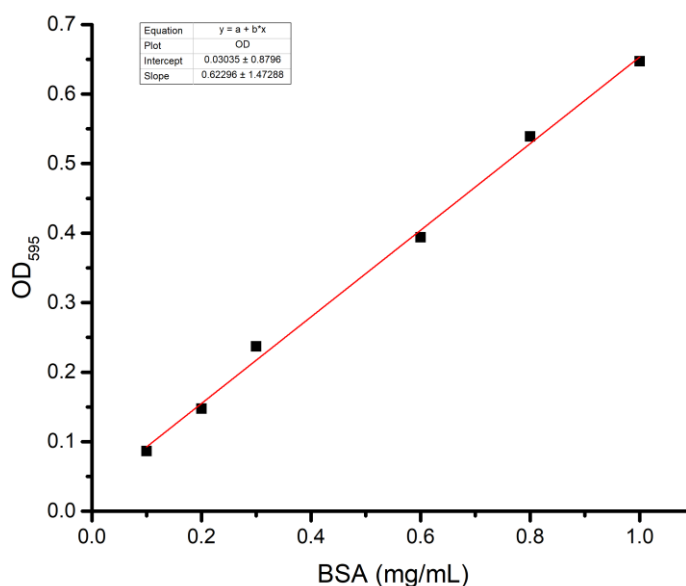
Preparation of -80°C (cryo) stocks

The desired strain of bacteria was inoculated in 5 mL of LB-medium supplemented with the appropriate antibiotic and incubated overnight at 37 °C with shaking 200 rpm. The cells were dispensed in the Eppendorf tube, and 50% glycerol was added to the final concentration of 25 %. The resulting permanent culture was mixed gently and snap-frozen by liquid nitrogen. The vial was immediately stored at –80 °C for further use.

E I.4 Bradford assay

In order to determine the concentration protein standard Bradford assay used. The protein solution was diluted 1:30 in (ddH₂O) and 5 μ L of diluted protein solution transferred to the 96 well plates and mixed with 200 μ L of 1:5 diluted Bradford reagent. (Protein Assay Dye Reagent Concentrate; 500-0006, Bio-Rad) and mixed for 30 s (1 350 rpm; Heidolph Titramax 1000). The plate was incubated at room temperature for 15 min for the color to develop. The absorbance was determined with the plate reader (AnthosZenyth3100) at 590 nm. The protein concentration determined by using the bovine serum albumin (BSA) calibration curve. The calibration curve shown in Figure E.1.

Figure E.1. An example of the Bradford calibration curve. The calibration curve made by using the BSA (0-1 mg.mL⁻¹) in dH₂O.



E I.5 Preparation of CaCl₂ competent cells

A single colony of the desired *E. coli* strain was used to inoculate 5 mL LB medium and incubated overnight in an orbital shaker (200 rpm; InforsHT Multitron 2 Standard) at 37 °C . 1% (v/v) of the overnight culture was inoculated in a fresh LB culture and incubated until it reaches an OD₅₉₀ = 0.3–0.4. The cell culture was centrifuged (6000 g, 4°C, 10 min; Sigma Laboratory Centrifuge 3k15), and

the supernatant was removed, cells were resuspended in 2 mL ice-cold 0.1 M CaCl₂ and placed on ice for 15 min. Cells were centrifuged and resuspended in 0.5 mL ice-cold 0.1 M CaCl₂ and dispensed in 1.5 mL tubes in 50 µL aliquots, snap-frozen in liquid nitrogen and stored at -80 °C for further use.

E I.6 Preparation of RbCl competent cells

A single colony of the desired *E. coli* strain was incubated overnight in 5 mL LB medium at 37 °C with shaking at 200 rpm. After inoculating a fresh LB culture (100 mL) with 1 % (v/v) of the overnight culture, it was grown to an OD₅₉₀ = 0.3-0.4. Cells were harvested by centrifugation (6000 g, 4 °C, 10 min) and subsequently resuspended in 1/5 volume of the main culture in RF1 buffer (20 mL; 100 mM RbCl, 50 mM MnCl₂, 30 mM KOAc, 10 mM CaCl₂, 15% (w/v) glycerol). After incubating the cells for 15 min, they were centrifuged and resuspended in RF2 buffer (10 mM RbCl, 10 mM MOPS, 75 mM CaCl₂, 15% (w/v) glycerol) using 1/5 volume of the RF1 suspension (4 mL). Cells dispensed in 1.5 mL tubes in 50 µL aliquots, snap-frozen in liquid nitrogen, and stored at -80 °C.

E I.6.1 Preparation of reagents

The preparation of the RF1 and RF2 buffers is shown in Table E.3 and Table E.4. The reagents were combined for 100 mL of RF1 and RF2 buffer, respectively. Acetic acid and NaOH have used for adjusting the pH RF1 and RF2, respectively. After pH adjustment, buffers sterilized both and stored at 4°C.

Table E.3. RF1 buffer pH 5.8.

Chemical	Quantity (gr)
RbCl	1.21
MnCl ₂	0.6292
KOAc	0.294
CaCl ₂	0.1109
Glycerol	15

Fill up to 100 mL dH₂O, Autoclaved at 121 °C for 20 minutes, pH was adjusted to 5.8.

Table E.4. RF2 buffer pH 6.8.

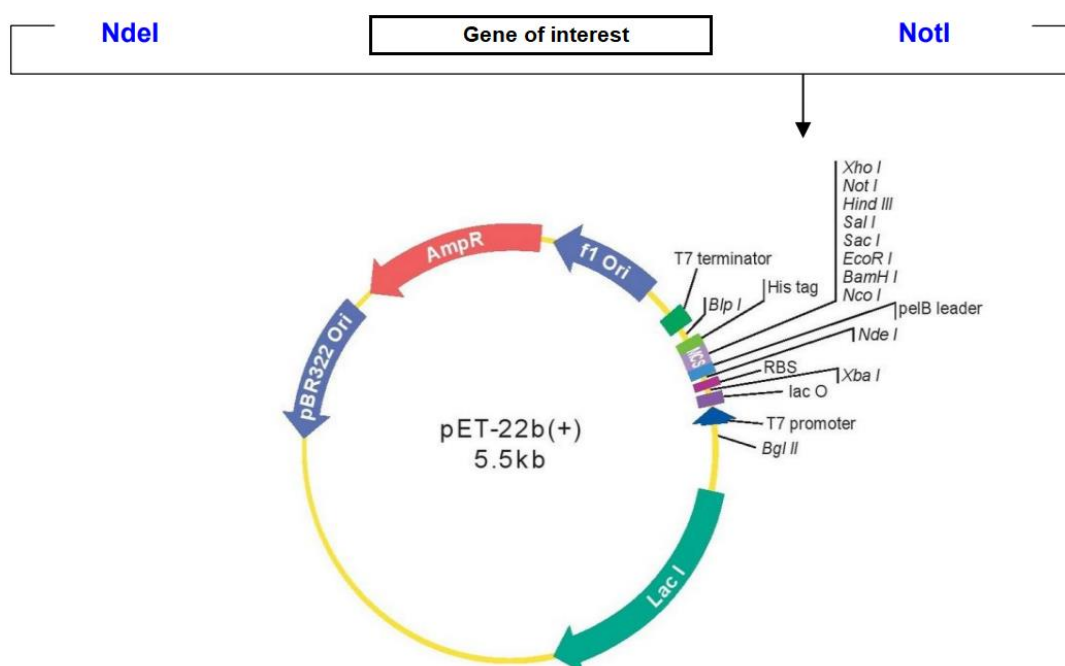
Chemical	Quantity (gr)
RbCl	0.12
MOPS	0.21
CaCl ₂	0.83235
Glycerol	7.5

Fill up to 100 mL dH₂O, Autoclaved at 121 °C for 20 minutes, pH was adjusted to 6.8.

E I.7 Plasmid map and cloning plan

Unless otherwise notified, all enzymes in this study were cloned in pET22 b (+) by using the NdeI and NotI restriction sites, and all contained a His-Tag at the C-terminus. The first and second series of mutants have been prepared by Dr. Saima Feroz and the rest of mutants and new putative BVMOs synthesized by Genscript company with the same plan (Figure E.2).

Figure E.2. Plasmid plan.



E I.8 Transformation of *E. coli* competent cell

E I.8.1 Heat shock transformation

The NEB transformation protocol for *E. coli* BL21 (DE3) has been used for the plasmid transformation. First, 50 μL of the *E. coli* BL21 (DE3) competent cell were thawed on ice until all ice crystals disappeared, then 1-5 μL of plasmid DNA (1 pg-100 ng) were added to the cells and flicked for 5-7 times. The mixture was incubated on ice for 30 min. Heat shock was performed at 42 °C in an Eppendorf thermomixer for 20 s. The cells were placed back on ice for 5 min. In order to recover the cells, 950 μL of SOC medium were added and incubated for 1 h at 37 °C, 200 rpm. After the recovery, the cells were diluted 10times in LB medium and 50-100 μL were stroke-plated on an agar plate with the appropriate antibiotic. The selection plates were incubated overnight upside down at 37 °C.

E I.8.2 Electroporation transformation

Electroporation cuvette (Cuvette Plus, 1 mm gap, 90 μL , sterile) and solutions were placed on ice for 30min before use. The cuvette was filled with 40 μL electrocompetent cells and 3 μL of the plasmid. Electroporation was carried out at 2.5 kV and 4.0 msec. Immediately after the shock, 950 mL of prewarmed (37 °C) SOC media were added to the cuvette and mixed. The solution was moved to a 1.5 mL Eppendorf tube and incubated at 37 °C, 200rpm for 45 min. After incubation, 50 μL of the cells were plated in a selective LB-agar plate and incubated overnight at 37 °C. Experiments were usually accompanied by a negative control (no plasmid solution) and positive control (parent plasmid).

E I.8.3 Preparation of SOC medium for cell recovery

SOC medium (Table E.5) used for the recovery of the cells during the plasmid transformation.

Table E.5. SOC medium.

Chemical	Quantity (gr)
Yeast extract (0.5% W/V)	2.5
Trypton (2 % W/V)	10
NaCl (10 mM)	0.29
KCl (2.5 mM)	0.09
MgSO ₄ (10 mM)	0.60
MgCl ₂ (10 mM)	0.48

pH adjusted to 7.5 by 1 mM NaOH, filled up 490 mL, and autoclaved. It cooled down to room temperature, and then 10 mL filter-sterilized 20% Glucose was added to the solution. These quantities are for the preparation of 500 mL SOC medium.

E I.9 Plasmid purification and sequencing

E I.9.1 Nuclease free water preparation

Unless otherwise notified, nuclease-free water (NFW) was used for the plasmid purification and all of the other molecular biology purposes. NFW was prepared by adding 0.05-0.1 % of diethylpyrocarbonate (DEPC) to deionized water and mixed at 37 °C for a few hours, and then it autoclaved and stored at - 20 °C.

E I.9.2 Purification

In order to confirm the sequence of the desired enzyme and also the mutations plasmid containing the gene of interest needed to be purified. For this purpose, we have used the GeneJet plasmid miniprep kit from ThermoScientific and purified the plasmid with the following procedure. One single

colony was inoculated to the 5 mL LB-medium containing the appropriate antibiotic if applicable and incubated overnight at 37 C in shaking incubator with 200 rpm. After incubation, the cells have been collected by centrifuge (Sigma 1-14) at the 16000 g for 2 min. The supernatant was removed, and the pellet was resuspended in 250 μ L resuspension solution provided by the kit. The bacteria were resuspended entirely by pipetting up and down. 250 μ L lysis solution (provided by the kit) was added and mixed by inverting the tube 5-7 times. 350 μ L Neutralization Solution was added and mixed by inverting 7-5 times. To separate pellet cell debris and chromosomal DNA, the mixture centrifuged for 8 min at 16000 g. The supernatant transferred to the GeneJET columns by pipetting without disturbing the white precipitate. It continued by 1 min centrifugation at 16000 g and discarding the flow-through afterward. The column placed back into the same collection tube. 500 μ L Wash Solution added to the column, and it centrifuged for 30-60 seconds. The flow-through discarded, and the column placed back into the same collection tube, and the same step repeated. The flow was discarded and additionally centrifuged for 1 min to remove residual EtOH. The column transferred into a new Eppendorf tube, and 35 μ L NFW added to the filter. After incubation at room temperature for 2 min, it centrifuged for 2 min at 16000 g, and the collected DNA stored at -20 °C. The utilization of NFW instead of elution buffer was due to a better quality of purified plasmid.

For the sequencing, we have used Microsynth AG. The sample's quality and quantity was checked with the nanodrop machine. For the best result, the plasmid concentration was adjusted to 80 ng/ μ L and sent for sequencing. Geneious analyzed the sequence's data.

E I.10 Site saturation mutagenesis

E I.10.1 Polymerase chain reaction

The following components in Table E.6 were assembled in a thin wall PCR tube and mixed gently. The PCR tube was transferred to the thermocycler (Biometra T_{ADVANCED} analytikjena), and the cycling was performed based on the condition in Table E.7. The list of primers used for SSM listed in the appendix. The primers were designed using the NEBaseChanger (<https://nebasechanger.neb.com>).

Table E.6. PCR reaction components and reagents.

Component	Volume 25 μ l reaction (μ l)
Q5 Hot Start High-Fidelity 2X Master Mix	12.5
10 μ M Fw Primer	1.25
10 μ M Rev Primer	1.25
Template DNA (1-25 ng/ μ l)	1
Nuclease Free Water	9

Table E.7. Thermocycling Conditions for routine PCR.

Step	Temperature ($^{\circ}$ C)	Time
Initial Denaturation	98	30 sec
25 cycles	98	10 sec
	x*	30 sec
	72	10-30 sec/kb
Final Extension	72	2 min
Hold	4–10	

* Annealing Temperature see suggested T_m of Software NEBaseChanger.

E I.10.2 KLD reaction

The KLD reaction was assembled as shown in Table E.8, following the protocol in the kit.

Table E.8. KLD reaction protocol and components.

Component	Volume (μ l)
PCR Product	1
2x KLD Reaction Buffer	5
10X KLD Enzyme Mix	1
Nuclease-free water	3
Incubate for 5 minutes at room temperature	

E I.11 Enzyme expression

E. coli strain BL21(DE3) was used as an expression host for all enzymes in this study. LB medium (5 mL) was supplemented with the appropriate antibiotic (Table E.1) and was inoculated with *E. coli* BL21(DE3) pET22b(+)_BVMO_{Flava}/CHMO_{Acineto}/TmCHMO/CHMO_{Variants} and incubated in an orbital shaker at 37 °C, 200 rpm over-night. Pre-cultivated bacteria (2 % v/v) transferred to a 1 L flask containing 250 mL LB with the same concentration of antibiotic as before. They were incubated at 37 °C, 200 rpm for 2- 3 hours to reach an optical density between 0.6-0.8 at 590 nm. Then Isopropyl β - d- thiogalactopyranoside (IPTG) was added to the final concentration of 50 μ M, and the flask was transferred to 20 °C and incubated for 18-22 h.

E I.12 SDS-PAGE

E I.12.1 Reagents and buffer preparation

Table E.9. SDS-PAGE buffers preparation.

SDS-PAGE sample buffer		10X SDS-PAGE running buffer		Stacking gel buffer (0.5 M Tris-HCl, pH 6.8)		Resolving gel buffer (1.5 M Tris-HCl, pH 8.8)	
Chemicals	Quantity(m L)	Chemical s	Quantity(g r)	Chemical s	Quantity(g r)	Chemical s	Quantity(g r)
dH ₂ O	7.1	Tris	30.3	Tris	6	Tris	18.15
stacking gel buffer	2.5	glycine	144				
glycerol	5.0	SDS	10				
10% SDS	4.0						
0.5 % bromophen ol blue	0.4						

The amount of tris for resolving and stacking gel accounts for 100 mL. For the 10X running buffer, the amount of chemicals accounts for 1000 mL: The pH adjusted by 2 M HCl: In order to make the sample buffer ready to use, it is necessary to add 50 μ L of β -mercaptoethanol to 950 μ L sample buffer.

Table E.10. SDS-PAGE reagent preparation.

30% Acrylamide		10% SDS		0.5% Bromophenol blue		10% APS	
Chemicals	Quantity (gr)	Chemicals	Quantity (gr)	Chemicals	Quantity (gr)	Chemicals	Quantity (gr)
acrylamide	29.2	SDS	1	bromophenol blue	0.05	APS	1
N',N'-bis-methyleneacrylamide	0.80						

The amount for 30 %acrylamide accounts for 100 mL, and the rest of the solution accounts for 10 mL.

Table E.11. SDS-PAGE gel preparation.

Stacking gel		Resolving gel	
Chemical	Quantity (mL)	Chemical	Quantity (mL)
stacking gel buffer	1	resolving gel buffer	2
30% (w/v) acrylamide	0.52	30% (w/v) acrylamide	3.2
dH ₂ O	2.42	dH ₂ O	2.7
10% SDS	0.04	10% SDS	0.08
10% APS	0.02	10% APS	0.04
TEMED	0.004	TEMED	0.004

This volume accounts for the preparation of 2 gels. TEMED and 10% APS added at the end, and after adding them, the gel poured immediately.

E I.12.2 Gel staining

The gel was removed from the cassette after electrophoresis, and the stacking gel was removed carefully. The resolving gel was placed in the staining box, and it was covered with dH₂O and microwaved at 750 W for 1 min. After the microwave, the gel was shaken for 2 min at RT (PSU-10i; Grant-bio). The dH₂O was removed and replaced with fresh dH₂O. Then, the mixture was microwaved at 500 W for 1 min, and shaken at for 2 min. This time water with dyeing solution (SimplyBlue™ SafeStain; LC6065, Novex®) and microwave at 350 W for 1 min. The gel was shaken at RT for 10 min. The dyeing solution removed, and the gel was placed in dH₂O and shaken for 10 min. The water removed, and the gel scanned for documentation.

E I.13 Enzyme purification

E I.13.1 Buffer preparation for enzyme purification

All the buffers prepared by using the ddH₂O, and the pH was adjusted afterward (Table E.12).

Table E.12. Enzyme purification buffers.

Binding buffer (50 mM Tris·HCl, 0.5 M NaCl, 25 mM Imidazole, pH 8)		Elution buffer (50 mM Tris·HCl, 0.5 M NaCl, 250 mM Imidazole, pH 8)		Washing buffer (50 mM Tris·HCl, pH 8.5)	
Chemical	Quantity (gr)	Chemical	Quantity (gr)	Chemical	Quantity (gr)
NaCl	14.61	NaCl	14.61	Tris	3.02
Tris	3.02	Tris	3.02		

The amount of chemicals accounts for 500 mL buffer. The buffers made using dH₂O. Imidazole was added before use from the 1 M stock: FAD was added to the corresponding buffer freshly before use from a 100 mM FAD stock solution. The Ph adjusted with 2 M HCl and 3 M NaOH.

All further steps were carried out at 4 °C to protect the enzyme against inactivation. The overnight culture containing expressed recombinant cells was centrifuged at 8000 g, 4 °C for 10 min, and cells

were collected. Cell pellets were resuspended in 50 mM Tris·HCl, pH 7.5 containing 100 μ M FAD, and 100 μ M of PMSF (phenylmethylsulfonyl fluoride). The crude cell extract was sonicated by a Bandelin KE76 sonotrode connected to a Bandelin Sonoplus HD 3200 in 9 cycles (5s pulse, 55s break, amplitude 50 %). Cell debris and aggregates were removed by centrifugation (15000 g, 25 min, 4 °C, JA-17 Beckmann rotor). The supernatant was filtered using a 0.25 μ m filter, equilibrated with 50 mM Tris·HCl, pH 7.5, 0.5 M NaCl, 100 μ M FAD, and applied on 1 mL of Ni- sepharose column (1 mL, GE Healthcare Bioscience). The unwanted non-attached proteins were washed by using a 5 column volume of 50 mM Tris·HCl, 0.5 M NaCl, 40 mM imidazole, 100 μ M FAD at pH 7.5. The elution step was performed by applying 5 column-volumes of 50 mM Tris·HCl, 0.5 M NaCl at pH 7.5 containing 400 mM imidazole, and 100 μ M FAD. The eluted enzymes were washed by 50 mM Tris·HCl, 100 μ M FAD, at pH 8, and concentrated with an ultracentrifugal tube with a cut-off of 10 kDa ^[110].

E 1.13.2 Column recharging

The column was recharged after each use, and they can be used until they reach their expiry date or visible damage is detectable. The recharging was done immediately after purification. Reagents and buffer for recharging the column are compiled in Table E.13. First, the column was washed with a 5 column volume of stripping buffer and this was followed by the same volume of binding buffer and finished by adding the 5-10 column volume of dH₂O. In this stage, the column must be colorless as everything washed out by stripping and binding buffer. In this step, the column was charged with 0.6 mL of 0.1 M NiSO₄ and incubated for 2 min. After incubation, the non-attached NiSO₄ washed with 5 column volume of dH₂O and binding buffer, respectively. The column washed with 20% EtOH and stored at RT.

Table E.13. Buffer and reagents for recharging the column.

Stripping buffer (50 mM Tris·HCl, 0.5 NaCl, 50 mM, 50 mM EDTA, pH 7.5)		Binding buffer (50 mM Tris·HCl, 0.5 NaCl, 50 mM, 100 μM FAD, pH 8)		Nickle tag (0.1 M NiSO ₄)	
Chemical	Quantity (gr)	Chemical	Quantity (gr)	Chemical	Quantity (gr)
Tris	3.02	Tris	3.02	NiSO ₄	0.77
NaCl	14.61	NaCl	14.61		
EDTA	7.3				

The quantity for the stripping and the binding buffer accounts for 500 mL. The amount for the nickel solution corresponds to 50 mL. Imidazole added before use from the 1 M stock. FAD was added to the corresponding buffer freshly before use from a 100 mM FAD stock solution. The pH adjusted with 2 M HCl and 3 M NaOH.

E I.14 Kinetic measurement

E I.14.1 Activity measurement

Enzyme activity was measured by monitoring the decrease of NADPH absorbance at 340 nm. Standard assays contained the enzyme (0.05 μM), cyclohexanone (0.5 mM), and NADPH (100 μM) in 50 mM Tris·HCl, adjusted to the desired pH. All the measurements have been carried out at 30 °C [110]. The reaction was started immediately after enzyme addition by mixing 4 μL NADPH (25 mM stock solution) to the cuvette (final volume 1 mL). Oxidation of NADPH followed at 30 °C in a Lambda 35 spectrophotometer (Perkin–Elmer, Waltham, MA, USA) for 120 seconds.

E I.14.2 K_m measurement

To evaluate the catalytic activity of the enzymes and also affinity toward the natural substrate, we did measure the K_m value as described below.

The activity of the enzyme was measured in different concentrations of substrate (0-1000 μM) to determine the K_m value. The activity measurement was done based on section D I.14 with just different

concentrations of cyclohexanone as the natural substrate. The value determined by fitting the value of activity versus concentration of substrate with a Michaelis-Menten function.

E I.14.3 K_d measurement

All these procedures were performed at 4 °C to protect the enzyme from deactivation. The column chromatography method was used to generate the CHMO FAD-free apoenzyme. Clear cell-free extract of the enzyme made and equilibrated 1:1 with binding buffer (50 mM Tris·HCl, 0.5 M NaCl, 25 mM imidazole, pH 8). The mixture loaded at a flow rate of 1 mL min⁻¹ into a Ni²⁺-Sepharose HP column (1 mL, GE Healthcare). After loading the extract, the FAD was removed by washing the column with a 5-8 column volume of deflavination buffer (250 mM phosphate buffer, 3 M KBr, pH 8.0). The KBr weakens the bonding between the enzyme and FAD and helps for making the apoenzyme. The column bound apoenzyme was eluted at a flow rate of 5 mL min⁻¹ with 5 column volume of elution buffer (50 mM Tris·HCl, 0.5 M NaCl, 250 mM imidazole, pH 8.0). The apoenzyme was desalted with PD MiniTrap desalting column (PD MiniTrap G-25, GE Healthcare) using 50 mM Tris·HCl, pH 8.5. Protein concentration was determined by the Bradford assay as described in section D I.5 To measure the dissociation constant, the catalytic activity of the enzyme determined after deflavination. For that, 1 μM apoenzyme was incubated with different amounts of FAD (0–100 μM, 5 min incubation at RT), and activity was measured. The activity measured according to section D I.15 and the values plotted versus the concentration of FAD. The measurements performed in triplicate. The K_d determined by fitting the data of activity versus concentration of FAD with a logistic function (Origin 8.5 for Windows).

E I.15 Stability measurement

Unless otherwise notified, stability measurement was performed by incubating 10 μM enzyme at 30 °C in 50 mM Tris·HCl, 10 μM FAD, pH 7.5. Samples were taken at different time points and added to a cuvette containing 100 μM NADPH and 0.5 mM substrate to test for catalytic activity. The experimental data were fitted to an exponential decay equation using Origin Pro software (Origin 9.1 for Windows). The thermostability of enzymes has also been measured in the presence of different co-solvent and also different pH values. For the effect of co-solvent on stability, the incubation buffer supplemented with a total amount of 5 % and mixed properly. The enzyme added and incubated at 30 °C. In order to evaluate the effect of pH on stability, the enzyme incubated in buffer with different pH and stability

measured as mentioned in section D I.16 for the pH 5.5 and 6.5, the buffer of choice was 50 mM phosphate buffer. For pH 7.5 and 8.5 50 mM, Tris·HCl used, and for pH 9.5 and 10.5 50 mM, CHES buffer used.

E I.16 Melting temperature evaluation

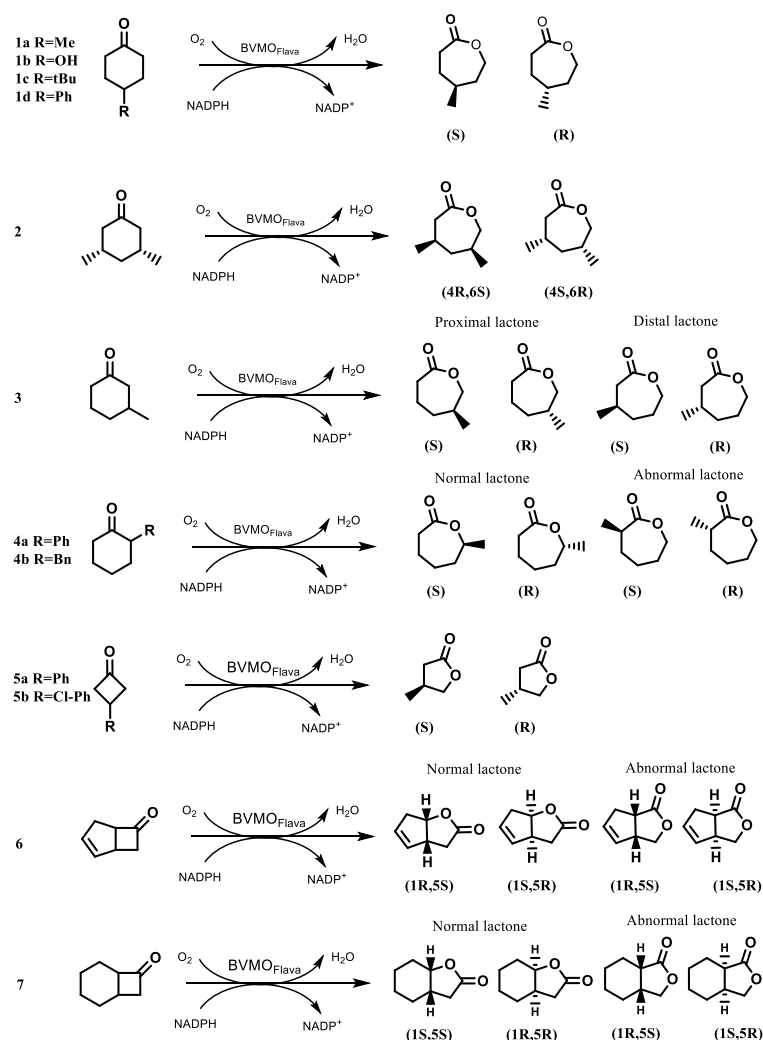
A Prometheus NT.48 device was used to determine the melting temperature (T_m) of all three enzymes. The samples were prepared in Tris·HCl 50 mM pH 8.5, 10 μ M FAD with a final enzyme concentration of 2 mg mL⁻¹. The glass capillaries filled with 10 μ L of the enzyme solution and the samples ran from 20 to 95 °C.

E I.17 Biotransformation

E I.17.1 List of substrates used for biotransformation

All the following substrates have been purchased from commercial suppliers with high purity and used without purification. The stock solution of the following chemicals was prepared in EtOH and stored at -20 °C.

Figure E.3. List of chemicals used for substrate profile study.



E I.17.2 General procedure

Unless otherwise stated, recombinant protein expression was performed in LB and induced with IPTG with the final concentration of 50 μ M at 20 °C. Cells were centrifuged (8000 g, 4°C, 10 min) and

resuspended and washed in 50 mM PBS buffer pH 7.4. After washing, the cells were centrifuged (8000 g, 4 °C, 10 min) and resuspended again with the same buffer to reach OD 30. 1 mL (OD 590=30) of recombinant expressed cells suspended in PBS buffer (pH 7.4, 50 mM), and 10 mM substrate final concentration, (methanol as co-solvent (5 % of total volume)) The components of the reaction (1.02 mL in total) were added into 25 mL flask, and the reaction was performed at 30 °C by shaking (220 rpm) for 24 h.^[57] The product was extracted with ethyl acetate containing 0.1 mM methyl benzoate as the internal standard for the GC analysis. The product analysis was performed with GC (Thermo Scientific Trace or Focus GC, Thermo Fisher Scientific, Waltham, MA, USA) using the chiral/achiral column. Product validation was performed by literature known reference biotransformations. The information of columns and methods for the GC experiments listed in the appendix.

E I.17.3 Chemical reference reaction

Chemical reference reactions were performed for the substrates 4-tert-butylcyclohexanone (**1c**), 4-phenylcyclohexanone (**1d**), 2-benzylcyclohexanone (**4b**), and bicycle[4.2.0]octan-7-one (**7**). For the mentioned substrates, we could not determine their enantiomers by enzymatic reference reactions. For this purpose, 3-chloroperbenzoic acid [8.9 μmol, 20 μl of a 10%w/v stock solution of reagent grade 3-chloroperbenzoic acid (77% w/w) in dichloromethane, 2.2 equiv] and substrate (4.0 μmol, 8 μL of a 0.5 M stock solution in dioxane) were mixed and moved to a GC vials. This resulted in a final concentration of 0.360 M peracid and 0.160 M ketone. The colorless solution was shaken for 18 h at RT. The solution was diluted with dichloromethane (100 μl) and transferred into a 1.5 mL Eppendorf tube. A solution of water (500 μl) and triethylamine (ca. 45 μmol, 1000 μl of 0.6% v/v solution in dichloromethane, ca. 9 equiv.) added. The biphasic mixture was shaken at RT for 30 min and centrifuged for 45 s at 1500 g. After removing the aqueous layer, the organic phase dried over Na₂SO₄. Methylbenzoate as standard added and the solution analyzed via GC.

E I.18 Bioinformatic tools and methods

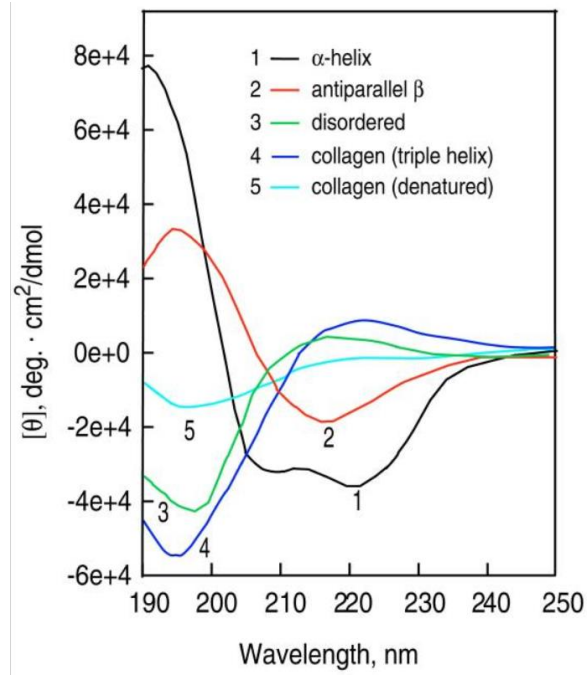
Different bioinformatics tools have used to investigate more deeply inside the enzymes sequences and also structure. Multiple sequence alignment was performed by MUSCLE (Multiple Sequence Comparison by Log-Expectation).^[195] The phylogenetic tree was generated by using phylogeny.fr.^[196] The homology model for all the enzymes was created by SWISS-MODEL^[197]. The protein 3D structure

was visualized and investigated by Swiss PDB viewer. ^[198] Multiple structure alignment has been carried out by PyMOL. Multiple structure alignment did help to see the differences in the structure level better and helps to plan better for further experiments.

E I.19 Circular dichroism (CD) spectroscopy

Circular dichroism is very sensitive and useful approach to study secondary structure of polypeptides and proteins . CD spectroscopy is also an excellent method for rapid determination of folding properties of proteins . The most widely used applications of this method is to determine whether a purified enzyme is folded properly, or if a mutation affects its conformation or stability. ^[199] Furthermore, CD measurements can be used to study protein interactions. This approach is a form of light absorption spectroscopy, which measures the difference in absorbance of left and right circularly polarized light by a substance. CD spectra between 260 and approximately 180 nm can be analyzed for the different secondary structural types: turn, parallel and antiparallel beta sheet, alpha helix and other. When the chromophores of the amides of the polypeptide backbone of proteins are aligned in arrays, their optical transitions are shifted or split into multiple transitions due to “exciton” interactions The consequence is that different structural elements have characteristic CD spectra (Figure E.4). For example, α -helical proteins have a positive band at 193 nm and negative bands at 222 nm and 208 nm ^[200]. Proteins with well-defined antiparallel β -pleated sheets have positive bands at 195 nm and negative bands at 218 nm ^[201], while disordered proteins have very low ellipticity above 210 nm and negative bands near 195 nm ^[202]. By comparison of your reference well folded protein or wild type protein cd spectra with the cd measurement of mutated variants, you can easily determine if mutations have effected the folding or not.

Figure E.4. Circular dichroism (CD) spectra of polypeptides and proteins with representative secondary structures. [199]



F Appendix

F I.1 The consensus sequence for 31 sequences. ^[191c]

Residues highlighted in red are known or thought to participate in the catalysis.

```

AA      RRQPPEEVDVLVVGAGFSGLYALYRLR  ELGRSVHVIETAGDVGGVWYWNRYPGARCDIESIEY
SS      CCCCCCEEEEEEECCCHHHHHHHHHHH  HTTCCEEEECCCCCCCCCHHHHCCCTTCECCCCCTTC
RSA     967128802000000100000001103  537240000151330022122010220201130001
consensus/90%.....hDslllGuGhuGhh.lh.L+.t.Ghps.hh-tuts.GGhWahNpYPGsh.D....Y
consensus/80%.....phDslllGuGFuGly.lapLR.phGhpsphh-tusshGGsWaWNpYPGAsDsps..Y
consensus/70%.....phDslllGAGFuGlytla+LR.chGhpsphh-susslGGsWaWNRYPGARsDs-s.hY
consensus/60%.st.pchDslllGAGFuGlytLa+LR.chGhpV+lhEuusslGGTWTWYWNRYPGARsDoEuhhY
consensus/50%tspspclDALVIGAGFuGlytL++LR.clGhcV+lLEAusDlGGTWTWYWNRYPGARsDoEshhY
AA prop          L      G      K      K      -      S

MSQKMDFAIVIGGGFGGLYAVKKLRDELELKVQAFDKATDVAGTWTWYWNRYPGALTD TETHLY 63
CCCCCEEEEEEECCCHHHHHHHHHHHCCCCCEEEEECCCCCCCCCHHHHCCCCCECCCCCCCC
791676140000101010000001027508050000162420033022130250110120101
BAA86293      KMDFAIVIGGGFGGLYAVKKLRDELELKVQAFDKATDVAGTWTWYWNRYPGALTD TETHLY
ZP_00379385   YLKC DVIIIGARVSGIYAAHKLSDNLGLDVGIEKGS DAGGTWYWNRYPGVQADTDSHVY
YP_345732     DTSLD AIVIGAGFAGIYALHKLRLNELGLAVRCFDKADG VGGTWHWNRYPGAKSDSEGFVY
YP_552312     QTDYDA VVVGAGFGGLYMLHKLRLNELGMNVRVFDKAGD VGGTWTWYWNRYPGALSDTESFVY
CAD10801      GADYDA VVVGAGFGGLYAVHKLRLNEQGMNVKAYDNAAD IGGTWFWNRYPGAVSDTESFVY
AAN37479      FQTVDA VVIGAGFGGIYAVHKLHNEQGLTVVGFDKADG PGGTWTWYWNRYPGALSDTESHVY
XP_755274     EHYLDA LVVGTGFSGIYALQSL-LKLN LKVKAIDAASDVG GTWYWSRYPGAMSDSWSHLY
AAR99068      AIHFDA IVVAGFGGMYMLHKLRLDQLGLKVVFDTAGG IGGTWTWYWNRYPGALSDTHSHVY
EAT88085      EQTIDA LVVAGFGGVYQLKKLR-DQGLKTKLIDAASDVG GTWYCTASFM-----Y
BAE64846      LPSYDA LVVAGFGGIYQLYSL-LKGLTVKLVRAEGP GGTTWYWNRYPGATSDTPSHLY
ZP_01384653   -----MHTASQALAIERQASAQPSLAERVS RPKTRAILGLMRLN-FLSVIRDTESFVY
BAE65198      EHHLDA LVVAGFGGIYTLYSLV-KEGLNVKAIDTAGD VGGTWTWYWNRYPGALSDTWSHLY
consensus/90% ..hDslllGuuhuGhY.h.pLh.p.thpshhh-tuts.uGTWahsta.us.tDo.saly
consensus/80% ..phDALVlGAGFuGlytLppLh.c.shpVhsh-pAsshGGT WaWNRYPGshsDo.oaly
consensus/70% ..phDALVlGAGFuGlytLppLt.c.GlpVpshDpAsslGGT WYWNRYPGAhoDT.Saly
consensus/60% thphDALVlGAGFGGLYtL+KLppc.GlpV+shDpAsslGGT WYWNRYPGAhsDTc.Saly
consensus/50% ppchDALVlGAGFGGIYALHKLRLs-LGLcVKuaDKAuDVGGT WYWNRYPGALSDTESHVY
AA prop          V  G      I  LH      G  K      G      S  S  L
RosDes          +  -      -  +-      -  -      +
Distance        +  5      +  ++      +  +      -      -  -
Interaction                    D30
Substitution     +  A      +      +  +      S
Position         14      L23      K
Final            A

```

Appendix

```

AA      CYSFSEEVLQEWNWTERYASQPEILRYINFVADKFDLRSGITFHHTTVTAAAFDEATNTWT
SS      CCCCCHHHHHCCCCCECCCEHHHHHHHHHHHHHHHTT333HEECCCCCEEEEEETTTTTEE
RSA     001215500441605220010300160022007327022000160304102144841101
consensus/90% ta.h..pl.papatphastt.phhtYht.lscphslppth.htstl.tstasc ttt.W.
consensus/80% taoh..-lhppasWppasst.ElhtYhpalsc+hDLpcsh.Ftopl.sApas-tst.Wp
consensus/70% paSh..ElhppwsWscpasspsElhpYhpaVs-+hDL++chpFpTpVpuApaDEtsshWp
consensus/60% paSaspElhp-WsWoc+YsspsElhpYhpaVs-+hDL++DhpFsTcVpoApaDEssspWp
consensus/50% sYsFscElhp-WsWoERySsQsElhcYlcaVu-+aDLR+DhpFsT+VsSAcadEsospWs
AA prop  W      E- W      Q E      -      -
          CYSWDKELLQSLKIKKKYVQGGPDVRKYLQQVAEKHDLKKSYPQFNTAVQSAHYNEADALWE 123
          CCCCCHHHHHCCCCCECCCEHHHHHHHHHHHHHHCHHHHEECCCCCEEEEEEECCCCCEE
          001107501662713310030300450045005526026100250106102145850003
BAA86293 CYSWDKELLQSLKIKKKYVQGGPDVRKYLQQVAEKHDLKKSYPQFNTAVQSAHYNEADALWE
ZP_00379385 RYSGDPTVSPSWDRARYQKGSQIRDYLDQFMKRHDLKIFHFETTATTADFKEEQARWE
YP_345732    RYsFDKEMLQQWSWTNRYLQAEVLEYLNAVVDRLDRDIQLETAVTsARWDDSLARWE
YP_552312    CYSFDKALLQEQWQDTRYVVTQPQILSYLNHVADRLDLRRDIEFNTGVTGARFDEKRNLWE
CAD10801    RFSFDRELLQGRWKNRYVTQPEILAYLNEVADHLDLRRSYEFNTKVSAAQFDDATGLWK
AAN37479    RFSFDKGLLQDGTWKHTYITQPEILEYLEDVVDRLRRHFRFGTEVKSATYLEDEGLWE
XP_755274    RYsFD----YEYPLRYRYVVSQPEMLAYLRHVVEKYDLRGMQFNTDMTSAVWDEGTSTWR
AAR99068    QYSFDEAMLQEWTKNKYLTQPEILAYLEYVADRLDLRPDIQLNTVVTSMHFNEVHNIWE
EAT88085    RYsWDLEDLRTYPWSNSYVVSQKEVQGYLRHVSErHDLIKDMVFEAALESADWDEASSSWI
BAE64846    RYsWDKEDLQSYsWSHNYLERKEVLAYLEHVVERHDLRRHMqFHTeVVSIAWNDDSCTWT
ZP_01384653 CYSWDKELLQEMHITTRYVVTQPQILSYLEHVADRHNLRPDIQLNTGITAAHFNEATNLWE
BAE65198    RFLFDQEFLOQTYPWKRWYLTQPEIMQYLRDVERYHLRKHMQFNTKMQRAEWNDETkiWE
consensus/90% paSaD.t...ph.httpYlpt.plhtYLptVs-+hcLc.phphpTthpsA.ap-tpshWp
consensus/80% paSaD.t.Lpph.hpppYlppsplhtYLppVs-+hDL+tpphpTthpuApas-tpshWc
consensus/70% pYsADpthLQphphpppYlpQs-lhtYLpcVs-RaDLR+chpFpTtlpuApas-spshWc
consensus/60% pYsADcphLQpasWpp+YlQPEILsYLpcVs-RaDLR+chpFsTslsoAcasEspshWE
consensus/50% pYsFDKELLQsasWcsRYVTQPEILuYLcaVuDRHDLR+chQFNTsVTsAcacEuoulWE
AA prop  R  W      W      Q  E  L      E      T
RosDes   -  -
Distance -  +      +  +  +  -  +  -      +
Interaction 3xK problematic QG instable: 1)QtoE with K78 salt bridge, 2)QGto TQ
Substitution -  -      +  +  +  ++  +      -
Position   74  E      83E
Final     EWSW      TQ
    
```

Appendix

```

AA          VDTNHGDRIRARYLIMASGQLSVPQLPNFPGLKDFAGNLYHTGNWPHEPVDFSGQRVGVIG
SS          EEETTCCEEEEEEEEECCCCCCCCCCCCCTT333CCCCCCCCCECEEEEEEC
RSA         0205591504020000034131426885252095074431301603788160581300002
consensus/90% 1..ttut hpspahl.shGhhu...hPph.GhppFtGt.hHou.WP.t.hshtGc+lullG
consensus/80% VpsppGphhps+ahl.shGhhuts.hPph.GhppFtGphhHTutWP pslshtG+RVuVIG...
consensus/70% VpocpGphhpA+allhAsGhhSts.hPslsGhcpFpGphaHTutWp-sl-hsG+RVuVIG..
consensus/60% VpT-sGcshpA+alIsAsGhLSpsphPslsGl-sFpGphaHTupWp-uVDhsGKRVuVIG..
consensus/50% VpT-cGcshsARaLIsAsGhLSspsphPcIFGl-sFcGchaHTupWp-cGVdhsGKRVGVIG..
AA prop          P - G -

VTTEYGDKYTARFLITALGLLSAPNLPNIKGINQFKGELHHTSRWP DDVSFEGKRVGVIG 183
EEEECCCCEEEEEEEECCCCCCCCCCCCCHHCCCEEEEECCCCC CCCCCCCCCCEEEEEEC
0307673601030000023220427685373194074331202702 97270743100012
BAA86293      VTTEYGDKYTARFLITALGLLSAPNLPNIKGINQFKGELHHTSRWP DDVSFEGKRVGVIG
ZP_00379385   VRTDRGQTFsARYLIGAAGVLSKpVLPNIPGLASFTGQTIHTARWP EGVSLAGARVAVFG
YP_345732     VRTDSSKVYRSKYLITALGVLSEpNTPEIPGIEQFSGQVVHTSRWP EGLDVAGRKRVGVIG
YP_552312     IQTDTDKVTARYLITALGLLSATNVPNIKGIETFOGAQYHTGAWP EGVDfKGRVGVIG
CAD10801      VTTDKGQAVTAKYLITGLGLLSATNLPKFKGMDTFKGRILHTGAWP EGVELAGKRVGIIG
AAN37479      VTTGGGAVYRAKYVINAVGLLSAINFPNLPIDTfEGETIHTAAWP QGKSLAGRRVGVIG
XP_755274     VSCKTGdVfHVRyLLTALGLLTKANYPDLPLGLQTFRGEIRHTSAWD TDLDLKGKRVGLVG
AAR99068      VRTDRGGYyTARFIVTALGLLSAINWPNIPGRESFOGEMyHTAAWP KDVELRGKRVGVIG
EAT88085     VETSAGITFRTRYLVtALGILSKRNVPDIAGLSSFGGEKfHTAAWP ENVALEGKRVGVIG
BAE64846     VESSQGS-FMSRYLITSLGIITePNWPNIPGRDQFQGSlyHTARWP DQYDLKGKRVGLIG
ZP_01384653  VKTDTGEAYTAKFLVtALGLLSATNVPKIKGLDfTQGEWlHTGNWP DDVQYAGKRVGVIG
BAE65198     VQCETGDVfHVRyLFTALGLLVKNYPDIpGMDTFKGMNHTSAWN PDVELENKRVGVIG
consensus/90% Vpsptuthhps+allsulGlLot.NhPpl.GhppFpGphhHTutWs pshphtG+RVGllG
consensus/80% VpoptGthaps+aLlTALGlLot.NhPpl.GhppFpGphhHTutWP pslphtG+RVGlIG
consensus/70% VpTcPgsapu+YLLtALGllStsNhPsI.GhpcFpGphhHTutWP cslplTgKRVGVIG
consensus/60% VpT-pGpsapuRYLlTALGllStsNhPsIsGl-oFpGchhHTusWP -sVsLtGKRVGVIG
consensus/50% VcTdsGcsaoARyLlTALGllSAsNlPNIpGl-pFpGEhhHTuAWP -sV-LcGKRVGVIG
AA prop          D - P Q S -
RosDes
Distance        + + + - - -
Substitution     - - P - - -
Position         - 136 153
Final            Y P
    
```


Appendix

```

AA      TGSSGIQVSPQIAKQA  AELFVFQRTPHFAVPARNAPLDPEFLADLKKRYAEFREES  R
SS      CCHHHHHHHHHHHHHE  CEEEEEECCCCCEEECCCECCHHHHHHHHTHHHHHHHH  H
RSA     0230001000400640  3300000641300000423815674045018405600330  2
consensus/90%  sGuoGlQhh..hu..s  tphhlf.Rospshs..pt.ht. t..t...t.h.thht...
consensus/80%  TGuoGlQhh.thup.s  tplhVFQRoFphslPhtpp.hssp..t.hptpasphphpth h
consensus/70%  TGuoGlQll.thActA  pcLhVFQRTPphulPhtpp.lsscph.h+tpYsphhpt t
consensus/60%  TGusGlQlI.plAcpA  ccLsVFQRTPsaulPhtpcslss-ctpph+spYs-hhcps p
consensus/50%  TGSSGIQlIspIA+pa  c+LTVFQRTPNaolPhtN+slss--tcchKsYs-lhccs p
AA prop      T          K          SPN          E

TGSTGVQVITAVAPLA--KHLTVFQRSAQYSVPIGNDPLSEEDVKKIKDNYDKIWDGV-W  240
CCHHHHHHHHHHHHHHC--CEEEEEEECCCCCEEECCCECCHHHHHHHHHCHHHHHHHHH-H
0320000001200640--6300000551400011426624873187228607630340-3
BAA86293
ZP_00379385  TGSTGCQLIANAATEV--EEMVVFQRTAQYVVPAGQRRTDSESHHYEQSFAELWDAR-R
YP_345732    TGSTGTQFICTAAETA--QQLTVFQRTAQYSIPSGNGPIDQEYLDRCRSNYDAIWDQV-R
YP_552312    TGSTGLQVITALAPQA--SHLTVFQRSPQYSVPVGNPVPSPDYVKSVKQNYEQIWKDV-R
CAD10801     TGSTGVQVITATAPIA--KHLTVFQRSAQFVVPIGNTPQDAETIARQKATYDDIWKQV-K
AAN37479     TGSTGQQVITALAPEV--EHLTVFVRTPQYSVVPVGRKRVTTQQIDEIKADYDNIWAQV-K
XP_755274    VGSSGVQLVPAVADTV--QSLHVFIRRPQFCVPSGDRALTAEEERAIFRDFPQIWSEA-R
AAR99068     TGSTGVQLITAIAPEV--KHLTVFQRTPQYSVPTGNRPVSAQEIAEVKRNFSKVMWQV-R
EAT88085     NGSSGVQVVTAIAKDAKHKHLISFQRNPQYSVPSGLRVPSSERQEINEKYPEIWKAEFE
BAE64846     NGSSGVQVITAIADQV--ESLVCFRHPQYIVPAGKRAVSQEERNTINKAYDEIWKQV-K
ZP_01384653  TGSTGTQVITAIAPKV--EHLTVFQRSPQYSVVPVGNPVTPEYVAEVKKNYDEIWDQV-M
BAE65198     VGSSGVQVVTAIADKV--KSLHVFVRRPQYTVPSGNRDVTPKERALVNKNYPALIADA-R
consensus/90%  sGSoGhQllsShA..s..ppLhF.RpsQasVP.Gptshs.p.htthptsastlWtts.h
consensus/80%  sGSoGsQllsAhA.ps..ppLhVF.RpsQYsVPsGptsls.p.htphppsasplWpps.c
consensus/70%  sGSoGsQlITALAsps..cpLsVFQRssQYsVPsGppslospphtplppsYspIWpps.+
consensus/60%  TGSTGLQVITALAsps..ccLsVFQRcFQYoVPsGscPloscphpcpcpsYspIWppV.+
consensus/50%  TGSTGVQVITALAPcV..cHLTVFQRcFQYSVPuGNRFVosEElcllKcNYDcIWcpV.R
AA prop      KA          P          D -          E R
RosDes      -
Distance    ++          -?          ++          +
Interaction          near problematic region
Substitution          P          E R
Position     K          209          --          238 240
Final       P          E R
    
```

Appendix

```

AA      NTPGGTHRYQGPKSALEVSDEELVETLERYWQEGGE DI LAAYRDILRDRDANERVAEF
SS      CCCCCCCCCCCCCCTTTCCHHHHHHHHHHHHHHCC3 33 33CCCCCCCCCHHHHHHHHHH
RSA     414200326505310450587304510351055000 30 11006001638700420030
consensus/90% .s..uh.h....thhphs.tpRtthh-that.suh..hh.tsa.Dhhhs.tuNt.h.pF
consensus/80% po.suh.h....ptshphs.c-RpthaEphWt.GGh.thhhtsatDhhhs.pANc.hhpF
consensus/70% po.suh.h....spssh-lotEERpphaEphWptGGh.phhhtsapDlhhs.cANcphhcF
consensus/60% pohuGhsa..sspssh-lotEERpchaEphWppGGh.phhhsFpDlhssccANcpst-F
consensus/50% cosuGasastsscsuhDVo-EERcchaEctWpcGGa.caahusFsDlIsDc-ANcpAs-F
AA prop      G      K      -      -      -      Pcheck      -Rcheck

NSALAFGLNESTVPAMSVSAEERKAVFEKAWQTGGGFRFMFETFGDIATNMEANIEAQNF 300
CCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHCCCCCHHHHCCCCCCCCCHHHHHHHHHH
423101705529340561568613530270174000111001006102436500320050
NSALAFGLNESTVPAMSVSAEERKAVFEKAWQTGGGFRFMFETFGDIATNMEANIEAQNF
LTRLACGFEEPQTGAQDVDEKTRREEVFEKAWEEGGGFNFMFGTFSDLAFNRASNNAADF
NSIVGCGFEESTVSATSSEAEERTRVFEEESWQRGNAFHFMFGTFNDIIFDPAANLAAADF
SSVVAFGFKESVVPAMSVSEERRAVYQKAWDTGGGFRFMFETFCDIATSEEANETAAAF
SSAVAFGFEESTIPAETASPEERDRVFEAAWQRGGGFYFMFGTFSDIATSEANETAAAF
RSGVAFGFEESTVPAMSVTEERRQVYEKAWYGGGFRFMFETFSDIATDEEANETAAAF
TSRVAIGFPELSREAMALCPADREAAFERLWEAGSGLQLFLFGGFTDLLTDPVANEAEACKF
ESAVAFGFEESTVPAMSVSEAEERQVFEAWNQNGFYMFGTFCDIATDPQANEAAATF
ESAFVAFGFKEASRKTDFDVDEKERNRIYEEAWNKGGGFRFMFETFSDIATDEANNAAGDF
QEIGGMGVVEEAKTSAMSVVDEERERIYQAAWDDGGAFRFLGTFNDLILNEASNRAACDF
GSVVAFGFKESTVSAMSVSEERQAVFQKAWDNGGGFRFMFETFCDIATDERANKAAQDF
TSVFAMGTPEPKRTFMSLSPEDREELLEQQWNIGNGFQFMFGGFSDIATNEVANEVCRF
tohhuhGh.Esph.s.sls.t-RptlapttWp.GsuF.FhFtsFsDlhhs..uN.tAtpF
pShhAhGhpEsphsAhsIs.t-RptlappuWptGsGFpFMFtTFsDlhhs..ANptAssF
pShlAhGfcEuohsAhoVstcERptlapcAWppGuGFpFMFtTFsDIsTs.tANpsAssF
pSslAhGfcEuossAMSVo-cERcpVaEcAWppGGGFpFMFtTFsDIATsptANcsAscF
sSsVAFGFEEESTVsAMSVSEERcVFE+AW-pGGGFRFMFGTFoDIATDEsAN-AAADF
AA prop      -      -N      E      E      E      DE      EE      D
RosDes      +      -
Distance    +      ++      +      +      +      ++      ++      +
Substitution -      -      E      E      -      DE      E      D
Position    260      E      290      E      299
Final      E      DR      D
    
```


Appendix

```

AA      IRNKIRNTVRDPEVAERLVKGYPFGTKLLILEIDYYEMFNDRNVHLVDTLSPAPIETITPRG
SS      HHHHHHHHCCCHHHHHHHCCCCCCCCCCCCCEEECCCHHHHTTCTTEEEETTTCCCEEECCCE
RSA     00410371085662034000593001054000046003002491031010650416200550
consensus/90%  httpKhtthlpdsththLhP..h.hhs+R..h-ptaa..hspsNVp1lsh.ttsI.pls.pg
consensus/80%  httpKhcthlpDPthschLhP..ashus+R.sh-psYa-haNpsNVp1lshptsPIpplo.pg
consensus/70%  hcpKlRthlpDFphAchLsP..ashGsKRssh-psYa-haNpsNVp1VdlptsPIpclT.pg
consensus/60%  h+cKlRpplcDPchA-hLsPpsasaGsKRssl-ssYaEsaN+sNVcIVDl+psPIpcITspG
consensus/50%  lRcKIRs+V+DPclA-pLsPpsHsaGsKRssLEssYYEsFNDRNVcLVdl+CsPIpcITssG
AA prop  E/Dcheck
IKGKIAEIVKDPAlAQAQLMP  QDLYAKRPLCDsGYNTFNDRNVRLedVKANPIVEITENG  360
HHHHHHHHCCCHHHHHHHCC  CCCCCCCCCCEEECCCHHHCCCCCEEECCCEEECCCE
02310572075642074050  8301154010054005002491061110731314400530
BAA86293  IKGKIAEIVKDPAlAQAQLMP  QDLYAKRPLCDsGYNTFNDRNVRLedVKANPIVEITENG
ZP_00379385  IKQKINIVIRDSDTAKDLTP  TEPYARRPVAVDNYEYTFNDRNVSLVNTRRSPIRSVTPHG
YP_345732  VRDKIAAIVDDPDTARKLMP  SGYYATRPiANKGYEYTFNRPNVSLVSIKDNPITRLTENA
YP_552312  IRGKIAEIVKDPETARKLTP  TDLYAKRPLCDsGYAYIYNRDNVALVDVKATPITEITPRG
CAD10801  IKRKLKQIVKDPETARKLTP  SDLYAKRPLCGDDYGVYNRDNVTLADVKADPIAEFTLTG
AAN37479  IRNKIVETIKDPETARKLTP  TGLFARRPLCDDGYFQVFNRPNVEAVAIKENPIREVTAKG
XP_755274  IRRKIKAEIVKDPQKRKVLTP  DEPYARRPLCANGFYEQFNDRNVFAVDIRAHPIVKIEAEG
AAR99068  IRNKIAEIVKDPETARKLTP  TDVYARRPLCDsGYRTYNRNSVSLVDVKATPISAMTPRG
EAT88085  IKAKIREKVKDPEKARKLMP  SEFYARRPLCDTGFYEQFNRENVDIVNLKETPIERITEKG
BAE64846  LKKKIDQIVQDPEKRRKLT  SELYARRPVCADGYEQFNRENVDVVDIAESPMLEFTRDG
ZP_01384653  IRSKIAEIVKDPETARKLMP  QDLYAKRPLCDsGYATYNRPNVDLVDVKANPIVEITPKG
BAE65198  LRKKIASIVKDPQKRDVLT  KELYGRRPLCDAGFYEIFNKENVFAVDIKKSPITEVTPSG
consensus/90%  l+tKItthVcDPpptchLhP  pt.YA+RPlsssaYt.aNR.NV.hssl+tsPI.phT.pg
consensus/80%  l+tKIttIV+DPppt+cLhP  p-hYA+RPlCcssGaYt.aNR.NVthVsl+tsPIhchT.pg
consensus/70%  I+pKIsPivKDF-pA+KLhP  s-hYA+RPLCcssGYyphaNRsNVp1VslKtsPIhclT.pg
consensus/60%  I+pKIsPivKDF-pARKLTP  o-lyA+RPLCcssGYyppFNR-NVslVDlKtsPIsclT.cG
consensus/50%  IAsKIAEIVKDPETARKLTP  o-LYARRPLCDsGYEsFNDR-NVslVDlKAoPIsElTs+G
AA prop  - E R T E V
RosDes
Distance  + + + - + +
Substitution  -+ + + - + + +
Position  302D 313R 336 346 358
Final  RD E R E V PR
If more stable than I295E=> slat bride possible
    
```

Appendix

```

AA      VRTSER EYELDSLVLATGEDALTGALFKIDIRGV GNVALKEKW AAGPRTYLGLSTAG
SS      EEECCC EEECCCCCCCCCCCCCTTHHHHCCEEEEC3 33CEHHHHT TTCCECCCECCECTT
RSA     040775 2040400000253330001025040303 964003720 53002000000002
consensus/90% 1.htpt..h.hDhllhATGaDuhsGsh.thtl.sp.ts..lt-.W ttu.pohhGht..t
consensus/80% lhssst..hphDhllhATGFDAhGshhthsIpGp.sG.tlt-hW ttG.pohhGhtsts
consensus/70% ltTs-u..aElDhllhATGFDAhTGultplcIpGp.sGhsLp-tW tpGspTaLGlssu
consensus/60% IpTu-u..aElDsIlaATGFDAITGuLppIDl+G+.sGhsLp-cW ssGscTaLGlussG
consensus/50% IpTuDGptaELDlIVFATGFDAITGuLssIDIRG+.sGhsLcDcW ucGP+TYLGlucG
AA prop  - - - - - D L - - - L K S M G

VKLENGDFVELDMLICATGEDAVDGNYVRMDIQGK-NGLAMKDYW-KEGPSSYMGVTVNN 418
EEEECCCECCCCCEEECCCCCCCCCHHHHHCEEECH-HHCEHHHHC-CCCCCECCCCCECC
04068351540300010253221001025040304-654104730-6300201000001
VKLENGDFVELDMLICATGFDAVDGNYVRMDIQGK-NGLAMKDYW-KEGPSSYMGVTVNN
ITLDDGRFFEVDVIVFATGFDAVDGAYRDFHVQGR-EGTNLLEHW-GDAPSSYLGIATPK
VVTADGTEHEIDLLVLATGFDA---GYKKMLTGR-DGTPISELW-NETTAAYLGIATHQ
IKTSDGVEHELDVLIIFATGFDAVDGNYTKIDLRGR-NGKTIKDKW-KAGPSTYLGVASAD
IRLASGAHELDVVIFATGFDAVDGNYTRMDMRGR-NGVSLRDMW-KEGPLGLGIMEAE
VVTEDGVLHELDVIVFATGFDAVDGNYRMEISGR-DGVNINDHW-DGQPTS YLGVSTAK
IRTADGVEHELDMLILATGYDAVDGNYRRIDLGR-GGQTINEHW-NDTPTSYVGVSTAN
ILTSDGKEHELDVLIIFATGFDSVDGSTRLAIKGK-EG-TLKDYWSAEGPTS YLGVSVPS
IKLADGTVHKLDVVICATGFNAFDGAYRRIDVVGR-EGKTLNEYW-KDGPTTNMAVATAG
VKTTDGVHEHELDMLIFATGFDAVDGNYTKIDIRGR-NGLTIQEOW-KSGPSSYMGVANAN
ICTADGTTHELDVLIIFATGFDAVDGTAMVDIRGR-DGKNLYDMWKPSGSPSTYVGM SVHG
lhhtsGh.aElDhllhATGa-AhDGsYhchclpG+.pG.sl.-hW.ttsPssYhGl.s.t
lhhsDhGh.HELdhlhATGFDAhDGsYt+hclpGE.pG.slp-hW.tpsPsoYhGlsst
lphsDGs.HELdllhATGFDAVDGsYp+h-lpGR.sGhsLp-hW.tpuPooYlGlussp
IpTuDGs.HELdvlIIFATGFDAVDGsYp+hDlpGR.sGhsLp-hW.pcGPooYlGluss
I+TADGsEHELDVLIIFATGFDAVDGNYsRlDI+GR.-GpTLpDaW.c-GPoSYLGVuoAs
AA prop  - --D H V - - - - Q - E A- - - - N
RosDes  - - - - - - - - - - - - - +
Distance + +++ + 9 4.7 + + + ++ +
Interaction conserved region
Substitution I T D H
Position 364 368 395 T 412
Final AD EH RE L
(K262R + D367E=> possible salt bridge)

```

Appendix

```

AA      FPNLFFIAGPGSP SALS2N2MLVSI2EQHVEWVTDHAIYMFK2TTL      TRSEAVLEKE
SS      CTTEEECCCTCC 333CCHHHHHHHHHHHHHHHHHHHHHHCCC      CCEEECHHHH
RSA     0000000000000 01210000000000200040023029582      3101036620
consensus/90% FPNhhhhhGP.ss...hsNh..shp.pspal.phlt.hpttth      t.hEsp.tu.
consensus/80% FPNhhhhhGPTuP...hsNhs.shE.ps-2alhphlthhcppsh      thhEsp.pup
consensus/70% FPNhFhlhGPTuPo.shsNh2sshE.ps-Wlschlphhcppsh      ptlEsp.cAc
consensus/60% FPNhFhlhGPTuPo.shsNhPsulE.ps-WIs-hlpahccpsh      splEust-AE
consensus/50% FPNhFhlhGPpoPS.sauNhPsulEhQs-WIs-sIcahccs2sl      ssIEAos-AE
AA prop      F      G

YPNMFVLGPNGP      FTN2PPSIESQVEWISDTIQYTVENNV-----ESIEATKEAE      468
CCCEEECCCCCCC      CCCHHHHHHHHHHHHHHHHHHHHHHCCC-----CCEEECHHHH
0000000000000      200200000100200020033047491-----4102036600
BAA86293      YPNMFVLGPNGP      FTNLPPSIESQVEWISDTIQYTVENNV-----ESIEATKEAE
ZP_00379385      FPNFFTVLGPLGV      FSNLPPGIEAQVDFIAEMIAVADSTEA-----ATVEVHPDAV
YP_345732      FPNMFVYGPNSV      FTNLPPGIETQVEWITELIRQAETQGS-----ATVEADETAV
YP_552312      YPNLFMILGPNGP      FTNLPPSIETQVEWISDLIKHMNDTGR-----QLVEATHEGE
CAD10801      FPNLFMILGPNGP      FTNLPPSIETQVEWIADMVKTMEARKL-----KTSEPTAQR
AAN37479      FPNWFVLGPNGP      FTNLPPSIETQVEWISDTVAYAEENGI-----RAIEPTPEAE
XP_755274      FPNLMMINGPQTP      FANIPPVSEENVNFIVDLIKRAEEISQR-----MNRPCLVEATEEAQ
AAR99068      FPNMFMILGPNGP      FTNLPPSIEAQVEWITDLVAHMRQHGL-----ATAEPTRDAE
EAT88085      FPNLFMLGGPNGP      FSNIPPALEVHARFISDLIESAEAKRKAE----SARAPVIEATREAE
BAE64846      FPNLFMIFGPQTP      LTNGPPAIEAHVEFITGAISR2AEK2HRKEQSTLPTPAKIVIESTE2EEGE
ZP_01384653      FPNMFVLGPNGP      FTNLPPSIESQVEWIAALIKDVNAKDL-----KTVEATTAAE
BAE65198      FPNLLLVNGPHMA      FANIP2TSGETNTEFIMDLVRR2AEK2ISKQ-----TGRQCEIEALEEAE
consensus/90%      aPNhhhl.GPpss      FsNlPPuhEpscaIsthlt.hpt.t.....hEsp.pu.
consensus/80%      FPNhFmlhGPpss      FoNlPPulEspV-aIs-hlt.hptpth.....thlEsstpAp
consensus/70%      FPNhFmlhGPPuP      FoNlPPuIEspVEaIs-hlpps2cppsh.....tslEsTp-Ac
consensus/60%      FPNhFmlhGPNGP      FTNLPPuIEsQVEaIsDhIpps-ppsh.....tslEuTc-AE
consensus/50%      FPNLFMVLGPNGP      FTNLPPSIEoQVEWIoDLI2+cAEcpup.....soleATcEAE
AA prop      -      L K AE
RosDes      +      - +
Distance     +      + + ++
Substitution  F      F
Position     419      451
Final        F      KYAE
    
```

Appendix

```

AA      DEWVEHVNEIADETLVPM  TASWYTGANVPGKPRVFMPLYVGGFHRYRQICDEVAAK  GY
SS      HHHHHHHHHHTTCC333  C333CCCCCCTTCCCCCCCCCHHHHHHHHHHHHT  TC
RSA     440053015107610125  1740118775897520000022102601420550276  40
consensus/90% ttW.t.h.ph.t.olh..  stuWh.usNh.tc.....a.us..Yht.htp.h.t  th
consensus/80% ptW.tthtphttoLhsp  spSWhhGsNls+.h.hh.a.uGhs.Yhthtpshtt  sa
consensus/70% ptWstpstChsstTLhsp  spSWahGuNlPGKsp.hh.Y.GGhs.Ytphtcsttp  sY
consensus/60% cpWspclschustTLasp  usSWahGuNlPGKe+phhsYhGGlstYppthp-sssp  GY
consensus/50% cpWsc+ls-lustTLaPc  AcSWYhGANlPGKsRphlsYlGGlupY+cpec-lssp  GY
AA prop      E   E           A   -   PN           E   G

EQWTQTCANIAEMTLFPK--AQSWIFGANIPGKKNTVYFYLGGLKEYRSALANCKNH-AY  525
HHHHHHHHHHHHHCCHHH--CHHHECCCCCCCCCECCCCCHHHHHHHHHHHHC-CC
560263024207721145--2841236774795820100024105602530450475-40
EQWTQTCANIAEMTLFPK--AQSWIFGANIPGKKNTVYFYLGGLKEYRSALANCKNH-AY
NRWAVESTEMAQGSVFAE--VKSWIFGSNVHTDSPRALFYFGGLGYSRQRLRSEIDN-RL
RTWAQLCDDIANASLFPK--AHSWIFGANIPGKTSRALFYFAGLGNRYRRLADEADA-DY
DGWTATCNEIAGYTLFPK--ADSWIFGANIPGKARTVMFYLAGLGAYTQKLNVTST-GY
DQWVELCRTIANMTLFPK--AESWIFGANIPGKKNTVMFYLAGLGNRYKVLGLSLES-GY
AEWTETCTQIANMTVFTK--VDSWIFGANVPGKPSVLFYLGGLGNRYGVLLDDVTAN-GY
RKWGEHCDEIANSTLLKQ--VPQWLFGNVPGKKVSTLPHYFGGVGRFRALIADIKAT-GF
DAWGRTCAEIAEQTLFGQ--VESWIFGANS PGKKHTLMFYLAGLGNRYKQLADVANA-QY
KGWTDLCNQVSAGSLFRK--TESWMFAANIPGKKRYIQFWLGGLSGFGKQLQDIAAQ-GW
AAWGGLCNAISDAALFRT--AASYFNGVNVDGKPRL----PGLVRM-----
AGWTKTCQDIANMTLFPK--ADSWIFGANIPGKNTVYFWLNRDSSGNGSCLMGSMPEPRV
RAWTARSRSSIAGTMLAPCFPSDLVLAALNLLQSCPQ-----
consensus/90% ttWsthstphut.olh.p..stsahhusNl.sct.....sh.t.....
consensus/80% ttWsthCpplutholF.p..spSWlFGsNlsGKp.hh.Fahutlthtt.h.....t.th
consensus/70% ptWsthCspIashoLFsp..spSWlFGuNlPGKp.phhFahuGlutapt.ltt.ttt.th
consensus/60% ctWseppCspIashTLFsc..scSWIFGANlPGKpssshFYhuGLusapt.Ltshtsp.ta
consensus/50% cuWTcpCs-IAsuTLFsK..A-SWIFGANIPGKKsolhFYLGGLGsYRptLs-lsss.uY
AA prop      E/R  E           E           E           E
RosDes
Distance      +   +           +           +
Substitution   E   E           E           E
Position      E   477           500           519
Final         E           PR           E
                If better than also Q473K/R=> salt bride possible
    
```


Appendix

AA	EGFVLT
SS	TTEEEC
RSA	410426
consensus/90%	tsa
consensus/80%	puF.h
consensus/70%	pGFpfs
consensus/60%	pGFpls
consensus/50%	cGFsIs
	EGFDIQ
	CCEEEC
	610628
BAA86293	EGFDIQLQRSDIKQPANA-----
ZP_00379385	PSFQLSTRSRLTQQREGAHHVHQ
YP_345732	PNFTFRSDDHTSLEREHAS-IPQ
YP_552312	PGFEIR-----
CAD10801	PTIIFDRAVECVA-----
AAN37479	RGFELKSEAAVAA-----
XP_755274	TGFKEPLSRKSEPRPLL-----
AAR99068	QGFAFQPL-----
EAT88085	KGYSITPAAS-----
BAE64846	-----
ZP_01384653	TG-----
BAE65198	-----
consensus/90%
consensus/80%	.s.....
consensus/70%	.sh.....
consensus/60%	.uathp.....
consensus/50%	sGFphp.ttp.....
AA prop	
RosDes	
Distance	
Substitution	
Position	

F I.2 List of sequences

CHMO-WT

ATGTCACAAAAAATGGATTTTGATGCTATCGTGATTGGTGGTGGTTTTGGCGGACTTTATGCAGTCAAAAAATTAAGAGACGAGCTC
 GAACTTAAGGTTTCAGGCTTTTGATAAAGCCACGGATGTCGCAGGTACTTGGTACTGGAACCGTTACCCAGGTGCATTGACGGATAC
 AGAAACCCACCTCTACTGCTATTCTTGGGATAAAGAATTACTACAATCGCTAGAAATCAAGAAAAATATGTGCAAGGCCCTGATGT
 ACGCAAGTATTTACAGCAAGTGGCTGAAAAGCATGATTTAAGAAGAGCTATCAATCAATACCGCGGTCAATCGGCTCATTACAA
 CGAAGCAGATGCCTTGTGGGAAGTCAACCTGAATATGGTGATAAGTACACGGCGCTTTCCTCATCTACTGCTTTAGGCTTATTGTC
 TGGCCTAECTTGCCTAACATCAAAGGCATTAATCAGTTTAAAGGTGAGCTGCATCATAACCGCGCTGGCCAGATGACGTAAGTTT
 TGAAGGTAAACGTGTGGCGTGATTGGTACGGGTTCCACCGGTGTTACAGGTTATTACGGCTGTGGCACCTCTGGCTAAACACCTCA
 CTGTCTTCCAGCGTTCTGCACAATACAGCGTTCCAATTGGCAATGATCCACTGTCTGAAGAAGATGTTAAAAAGATCAAAGACAATT
 ATGACAAAATTTGGGATGGTGTATGGAATTCAGCCCTTGCCTTTGGCCTGAATGAAAGCACAGTGCCAGCAATGAGCGTATCAGCT
 GAAGAACGCAAGGCAGTTTTTAAAAGGCATGGCAAACAGGTGGCGGTTCCGTTTCATGTTTAAAACCTTCGGTGATATTGCCAC
 CAATATGGAAGCCAATATCGAAGCGCAAATTTTCATTAAGGGTAAAATTGCTGAAATCGTCAAAGATCCAGCCATTGCACAGAAGC
 TTATGCCACAGGATTTGTATGCAAAACGTCCGTTGTGTGACAGTGGTACTACAACACCTTTAACCGTGACAATGTCCGTTTAGAAG
 ATGTGAAAGCCAATCCGATTGTTGAAATTACCGAAAACGGTGTGAAACTCGAAAATGGCGATTTTCGTTGAATTAGACATGCTGATA
 TGTGCCACAGGTTTTGATGCCGTCGATGGCAACTATGTGCGCATGGACATTCAGGTA AAAACGGCTTGGCCATGAAAGACTACTG
 GAAAGAAGGTCCGTCGAGCTATATGGGTGTCACCGTAAATAACTATCCAAACATGTTTCATGGTGCTTGGACCGAATGGCCCGTTTA
 CCAACCTGCCGCCATCAATTGAATCACAGGTGGAATGGATCAGTGATACCATTCAATACACGGTTGAAAACAATGTTGAATCCATTG
 AAGCGACAAAAGAAGCGGAAGAACAATGGACTCAAACCTGCGCCAATATTGCGGAAATGACCTTATCCCTAAAGCGCAATCCTGG
 ATTTTTGGTGCGAATATCCCGGGCAAGAAAACACGGTTACTTCTATCTCGGTGGTTTAAAAGAATATCGCAGTGCCTAGCCAAC
 TGCAAAAACCATGCCTATGAAGGTTTTGATTTCAATTACAACGTTTCAGATATCAAGCAACCTGCCAATGCCTAA

1 MSQKMFDAI VIGGGFGLY AVKKLRDELE LKVQAFDKAT DVAGTWYWNR YPGALTDDET
 61 HLYCYSWDKE LLQSLEIKKK YVQGPVDRKY LQQVAEKHDL KKSYPQNTAV QSAHYNEADA
 121 LWEVTTEYGD KYTARFLITA LGLLSAPNLP NIKGINQFKG ELHHTSRWPD DVSFEGKRVG
 181 VIGTGSTGVQ VITAVAPLAK HLTVFQRSAQ YSVPIGNDPL SEEDVKKIKD NYDKIWDGVW
 241 NSALAFGLNE STVPAMSVSA EERKAVFEKA WQTGGGFRFM FETFGDIATN MEANIEAQNF
 301 IKGKIAEIVK DPAIAQKLM QDLYAKRPLC DSGYYNTFNR DNVRLDVDKA NPIVEITENG
 361 VKLENGDFVE LDMLICATGF DAVDGNVVRM DIQKNGLAM KDYWKEGPSS YMGVTVNNYP
 421 NMFVVLGPNP PFTNLPPSIE SQVEWISDTI QYTVENNVES IEATKEAEEQ WTQTCANIAE
 481 MTLFPKAQSW IFGANIPGKK NTVYFYLGLL KEYSALANC KNHAYEGFDI QLQRSDIKQP
 541 ANA

CHMO-QYTV451-454KYAE+Q473R+N477E+G14A (2nd_{mut-4})

ATGAGCCAGAAGATGGACTTTGATGCGATTGTGATTGGT GCGGGTTTTGGTGGTCTGTATGCGGTGAAGAAGCTGCGTGACGAAC
 TGGAGCTGAAAGTGCAGGCGTTTGACAAGGCGACCGATGTTGCGGGCACCTGGTACTGGAACCGTTATCCGGGTGCGCTGACCGA
 CACCGAAACCCACCTGTAAGTGTATAGCTGGGATAAAGAGCTGCTGCAGAGCCTGAAATCAAGAAAAAGTATGTGCAAGGTCCG
 GACGTTCTGAAATACCTGCAGCAAGTTGCGGAGAAGCACGATCTGAAAAAGAGCTATCAGTTAACACCGCGTTCAAAGCGGCAC
 TACAACGAAGCGGACGCGCTGTGGGAAGTGACCACCGAATACGGCGATAAGTATACCGCGGTTTCTGATTACCGCGCTGGGTCT
 GCTGAGCGCGCCGAACCTGCCGAACATCAAAGGCATTAACAGTTAAGGGCGAGCTGCACCACACCAGCCGTTGGCCGGACGAT
 GTGAGCTTCGAAGGCAAACGTGTGGGTGTTATCGGCACCGGTAGCACCGCGTGCAAGTTATTACCGCGGTGGCGCCGCTGGCGA
 AGCACCTGACCGTTTTTCAGCGTAGCGCAATATAGCGTGCCGATCGGTAACGACCCGCTGAGCGAGGAAGATGTTAAAAAGATC
 AAGGACAACACGATAAGATTTGGGATGGTGTGTGGAACAGCGCGCTGGCGTTTGGTCTGAACGAGAGCACCGTGCCGGCGATGA
 GCGTTAGCGCGGAGGAACGTAAAGCGTTTTTCGAGAAGGCGTGCCAGACCGGTGGCGGTTTCCGTTTTATGTTCAACACCTTTGGC
 GACATCGCGACCAACATGGAGGCGAACATCGAAGCGCAAACCTTCATTAAGGCAAGATCGCGGAAATTGTGAAAGATCCGGCGA
 TTGCGCAGAAGCTGATGCCGAAGACCTGTACGCGAAACGTCCGCTGTGCGATAGCGTTACTATGAAACCTTCAACCGTGACAAC
 GTGCGTCTGGAAGATGTTAAAGCGAACCCGATCGTGGAGATTACCGAAAACGGCGTTAAGCTGGAGAACGGTGACTTTGTTGAAC
 TGGATATGCTGATCTGCGCGACCGTTTTGATGCGGTGGATGGTAACTACGTTCTGATGGACATTACGGGCAAGAACGGTCTGGCG
 ATGAAAGATTACTGAAAGAGGGTCCGAGCAGCTATATGGGTGTGACCGTTAACAACACTACCGAACATGTTTATGGTCTGGGCC
 GAACGGTCCGTTACCAACCTGCCGCCGAGCATCGAGAGCCAAGTTGAATGGATCAGCGACACCATT AAATATGCGGAGGAAAAC
 AACGTGGAGAGCATCGAAGCGACCAAAGAGGGCGGAGGAACAGTGGACC CGTACCTGCGCG GAGATTGCGGAAATGACCCTGTT
 CCGAAAAGCGCAAAGCTGGATCTTCGGCGGAACATTCCGGGTAAAAGAACACCGTTTACTTCTATCTGGGCGGTCTGAAAGAATA
 CCGTAGCGCGCTGGCGAACTGCAAGAACCACGCGTATGAGGGTTTTGACATCCAACCTGCAACGTAGCGACATTAACAACCGCGG
 AACGCG

1 MSQKMFDAI VIGAGFGGLY AVKKLRDELE LKVQAFDKAT DVAGTWYWNR YPGALTDDET
 61 HLYCYSWDKE LLQSLEIKKK YVQGPVDRKY LQQVAEKHDL KKSYPQFNTAV QSAHYNEADA
 121 LWEVTTEYGD KYTARFLITA LGLLSAPNLP NIKGINQFKG ELHHTSRWPD DVSFEGKRVG
 181 VIGTGSTGVQ VITAVAPLAK HLTVFQRSAQ YSVPIGNDPL SEEDVKKIKD NYDKIWDGVW
 241 NSALAFGLNE STVPAMSVSA EERKAVFEKA WQTGGGFRFM FETFGDIATN MEANIEAQN
 301 IKGKIAEIVK DPAIAQKLMQ QDLYAKRPLC DSGYNTFNR DNVRELDVKA NPIVEITENG
 361 VKLENGDFVE LDMLICATGF DAVDGNVVRM DIQKKNGLAM KDYWKEGPSS YMGVTVNNYP
 421 NMFVVLGPNP PFTNLPPSIE SQVEWISDTI KYAENNIVES IEATKEAEEQ WRTCAEIAE
 481 MTLFPAQSW IFGANIPGKK NTVYFYLGLL KEYRSALANC KNHAYEGFDI QLQRSDIKQP
 541 ANA

CHMO-QYTV451-454KYAE+Q473R+N477E+G14A+N336E (2nd_{mut-5})

ATGAGCCAGAAGATGGACTTTGATGCGATTGTGATTGGT GCGGGTTTTGGTGGTCTGTATGCGGTGAAGAAGCTGCGTGACGAAC
 TGGAGCTGAAAGTGCAGGCGTTTGACAAGGCGACCGATGTTGCGGGCACCTGGTACTGGAACCGTTATCCGGGTGCGCTGACCGA
 CACCGAAACCCACCTGTACTGCTATAGCTGGGATAAAGAGCTGCTGCAGAGCCTGGAAATCAAGAAAAAGTATGTGCAAGGTCCG
 GACGTTTCGTAATACTGCAGCAAGTTGCGGAGAAGCACGATCTGAAAAAGAGCTATCAGTTTAACACCGCGGTTCAAAGCGGCAC
 TACAACGAAGCGGACGCGCTGTGGGAAGTGACCACCGAATACGGCGATAAGTATACCGCGCTTTCCTGATTACCGCGCTGGGTCT
 GCTGAGCGCGCCGAACCTGCCGAACATCAAAGGCATTAACAGTTTAAAGGGCGAGCTGCACCACACCGCCGTTGGCCGGACGAT
 GTGAGCTTCGAAGGCAAACGTGTGGGTGTTATCGGCACCGGTAGCACCGCGTGCAAGTTATTACCGCGGTGGCGCCGCTGGCGA
 AGCACCTGACCGTTTTTCAGCGTAGCGCGCAATATAGCGTGCCGATCGGTAACGACCCGCTGAGCGAGGAAGATGTTAAAAAGATC
 AAGGACAACCTACGATAAGATTTGGGATGGTGTGTGGAACAGCGCGCTGGCGTTTGGTCTGAACGAGAGCACCGTGCCGCGATGA
 GCGTTAGCGCGGAGGAACGTAAAGCGGTTTTCGAGAAGGCGTGGCAGACCGGTGGCGTTTTCCGTTTTATGTTCAACCTTTGG
 CGACATCGCGACCAACATGGAGGCGAACATCGAAGCGCAAACCTTATTAAAGGCAAGATCGCGGAAATTGTGAAAGATCCGGCG
 ATTGCGCAGAAGCTGATGCCGCAAGACCTGTACGCGAAACGTCCGCTGTGCGATAGCGGTTACTATGAAACCTTCAACCGTGACAA
 CGTGCGTCTGGAAGATGTTAAAGCGAACCCGATCGTGGAGATTACCGAAAACGGCGTTAAGCTGGAGAACCGTGACTTTGTTGAA
 CTGGATATGCTGATCTGCGGACCGTTTTGATGCGGTGGATGGTAACTACGTTTCGTATGGACATTAGGGCAAGAACGGTCTGGC
 GATGAAAGATTACTGGAAAGAGGGTCCGAGCAGCTATATGGGTGTGACCGTTAACTACCCGAACATGTTTATGGTGCTGGGCC
 CGAACCGTCCGTTACCAACCTGCCGCCGAGCATCGAGAGCCAAGTTGAATGGATCAGCGACACCATTAAATATGCGGAGGAAAA
 CAACGTGGAGAGCATCGAAGCGACCAAAGAGGCGGAGGAACAGTGGACCGTACCTGCGCGGAGATTGCGGAAATGACCCTGTT
 TCCGAAAGCGCAAAGCTGGATCTTCGGCGCGAACATTCCGGGTAAAAAGAACACCGTTTACTTCTATCTGGGCGGTCTGAAAGAAT
 ACCGTAGCGCGCTGGCGAACTGCAAGAACCACGCGTATGAGGGTTTTGACATCCAACCTGCAACGTAGCGACATTAAACAACCGGC
 GAACGCG

1 MSQKMDFDAI VIGAGFGGLY AVKCLRDELE LKVQAFDKAT DVAGTWYWNR YPGALTDDET
 61 HLYCYSWDKE LLQSLEIKKK YVQGPVDRKY LQQVAEKHDL KKSYPFNTAV QSAHYNEADA
 121 LWEVTTEYGD KYTARFLITA LGLLSAPNLP NIKGINQFKG ELHHTSRWPD DVSFEGKRVG
 181 VIGTGSTGVQ VITAVAPLAK HLTVFQRSAQ YSVPIGNDPL SEEDVKKIKD NYDKIWDGVW
 241 NSALAFGLNE STVPAMSVSA EERKAVFEKA WQTGGGFREFM FETFEDIATN MEANIEAQN
 301 IKGKIAEIVK DPAIAQKLMQ QDLYAKRPLC DSGYYETFN RDNVRLEDVKA NPIVEITENG
 361 VKLENGDFVE LDMLICATGF DAVDGNVVRM DIQKNGLAM KDYWKEGPSS YMGVTVNNYP
 421 NMFVVLGPNQ PFTNLPPSIE SQVEWISDTI KYAEENNVES IEATKEAEEQ WRTCAEIAE
 481 MTLFPKAQSW IFGANIPGKK NTVYFYLGLL KEYSALANC KNHAYEGFDI QLQRSDIKQP
 541 ANA

CHMO-QYTV451-454KYAE+Q473R+N477E+G14A+T415C+A463C (3rdmut-1)

ATGAGCCAGAAGATGGACTTTGATGCGATTGTGATTGGT GCGGGTTTTGGTGGTCTGTATGCGGTGAAGAAGCTGCGTGACGAAC
 TGGAGCTGAAAGTGCAGGCGTTTGACAAGGCGACCGATGTTGCGGGCACCTGGTACTGGAACCGTTATCCGGGTGCGCTGACCGA
 CACCGAAACCCACCTGTACTGCTATAGCTGGGATAAAGAGCTGCTGCAGAGCCTGGAATCAAGAAAAAGTATGTGCAAGGTCCG
 GACGTTCTGAAATACCTGCAGCAAGTTGCGGAGAAGCACGATCTGAAAAAGAGCTATCAGTTTAAACCCGCGGTTCAAAGCGGCAC
 TACAACGAAGCGGACGCGCTGTGGGAAGTGACCACCGAATACGGCGATAAGTATACCGCGGTTTCTGATTACCGCGCTGGGTCT
 GCTGAGCGCGCCGAACCTGCCGAACATCAAAGGCATTAACAGTTTAAAGGGCGAGCTGCACCACACAGCCGTTGGCCGGACGAT
 GTGAGCTTGAAGGCAAACGTGTGGGTGTTATCGGCACCGGTAGCACCGCGTGCAAGTTATTACCGCGGTGGCGCCGCTGGCGA
 AGCACCTGACCGTTTTTCAGCGTAGCGCGCAATATAGCGTGCCGATCGGTAACGACCCGCTGAGCGAGGAAGATGTTAAAAAGATC
 AAGGACAACACTACGATAAGATTTGGGATGGTGTGTGGAACAGCGCGCTGGCGTTTGGTCTGAACGAGAGCACCGTGCCGGCGATGA
 GCGTTAGCGCGGAGGAACGTAAAGCGGTTTTCGAGAAGGCGTGGCAGACCGGTGGCGGTTTTCCGTTTTATGTTAACACCTTTGGC
 GACATCGCGACCAACATGGAGGCGAACATCGAAGCGCAAACCTTCATTAAAGGCAAGATCGCGGAAATTGTGAAAGATCCGGCGA
 TTGCGCAGAAGCTGATGCCGCAAGACCTGTACGCGAACGTCCGCTGTGCGATAGCGGTTACTATGAAACCTTCAACCGTGACAAC
 GTGCGTCTGGAAGATGTTAAAGCGAACCCGATCGTGGAGATTACCGAAAACGGCGTTAAGCTGGAGAACGGTGACTTTGTTGAAC
 TGGATATGCTGATCTGCGGACCGGTTTTGATGCGGTGGATGGTAACTACGTTTCGATGGACATTCAGGGCAAGAACGGTCTGGCG
 ATGAAAGATTACTGAAAGAGGGTCCGAGCAGCTATATGGGTGTGTCGTTAACAACACTACCGAACATGTTTATGGTGTGGGCC
 GAACGGTCCGTTACCAACCTGCCGCCGAGCATCGAGAGCCAAGTTGAATGGATCAGCGACACCATTAAATATGCGGAGGAAAAC
 AACGTGGAGAGCATCGAATGACCAAAAGAGGCGGAGGAACAGTGGACCCTACCTGCGCGGAGATTGCGGAAATGACCCTGTTTC
 CGAAAGCGCAAAGCTGGATCTTCGGCGCGAACATTCCGGGTAAAAAGAACACCGTTTACTTCTATCTGGGCGGTCTGAAAGAATAC
 CGTAGCGCGCTGGCGAACTGCAAGAACCACGCGTATGAGGGTTTTGACATCCAACCTGCAACGTAGCGACATTAACAACCGGCGA
 ACGCG

1 MSQKMDFDAI VIGAGFGGLY AVKCLRDELE LKVQAFDKAT DVAGTWYWNR YPGALTDDET
 61 HLYCYSWDKE LLQSLEIKKK YVQGPVDRKY LQQVAEKHDL KKSYPFNTAV QSAHYNEADA
 121 LWEVTTEYGD KYTARFLITA LGLLSAPNLP NIKGINQFKG ELHHTSRWPD DVSFEGKRVG
 181 VIGTGSTGVQ VITAVAPLAK HLTVFQRSAQ YSVPIGNDPL SEEDVKKIKD NYDKIWDGVW
 241 NSALAFGLNE STVPAMSVSA EERKAVFEKA WQTGGGFRFM FETFGDIATN MEANIEAQN
 301 IKGKIAEIVK DPAIAQKLMQ QDLYAKRPLC DSGYNTFNR DNVRLDVDKA NPIVEITENG
 361 VKLENGDFVE LDMLICATGF DAVDGNVVRM DIQKKNGLAM KDYWKEGPSS YMGVAVNNYP
 421 NMFVVLGPNQ PFTNLPPSIE SQVEWISDTI KYAENNIVES IETKEAEEQ WTRTCAEIAE
 481 MTLFPKAQSW IFGANIPGKK NTVYFYLGLL KEYSALANC KNHAYEGFDI QLQRSDIKQP
 541 ANA

CHMO-QYTV451-454KYAE+Q473R+N477E+G14A+M400I (3rd mut-2)

ATGAGCCAGAAGATGGACTTTGATGCGATTGTGATTGGT GCGGGTTTTGGTGGTCTGTATGCGGTGAAGAAGCTGCGTGACGAAC
 TGGAGCTGAAAGTGCAGGCGTTTGACAAGGCGACCGATGTTGCGGGCACCTGGTACTGGAACCGTTATCCGGGTGCGCTGACCGA
 CACCGAAACCCACCTGTACTGCTATAGCTGGGATAAAGAGCTGCTGCAGAGCCTGGAATCAAGAAAAAGTATGTGCAAGGTCCG
 GACGTTCTGAAATACCTGCAGCAAGTTGCGGAGAAGCACGATCTGAAAAAGAGCTATCAGTTTAAACCCGCGGTTCAAAGCGGCAC
 TACAACGAAGCGGACGCGCTGTGGGAAGTGACCACCGAATACGGCGATAAGTATACCGCGGTTTCTGATTACCGCGCTGGGTCT
 GCTGAGCGCGCCGAACCTGCCGAACATCAAAGGCATTAACAGTTTAAAGGGCGAGCTGCACCACACAGCCGTTGGCCGGACGAT
 GTGAGCTTGAAGGCAAACGTGTGGGTGTTATCGGCACCGGTAGCACCGCGTGCAAGTTATTACCGCGGTGGCGCCGCTGGCGA
 AGCACCTGACCGTTTTTCAGCGTAGCGCGCAATATAGCGTGCCGATCGGTAACGACCCGCTGAGCGAGGAAGATGTTAAAAAGATC
 AAGGACAACCTACGATAAGATTTGGGATGGTGTGTGGAACAGCGCGCTGGCGTTTGGTCTGAACGAGAGCACCGTGCCGGCGATGA
 GCGTTAGCGCGGAGGAACGTAAAGCGGTTTTCGAGAAGGCGTGGCAGACCGGTGGCGTTTTCCGTTTTATGTTCAACACCTTTGGC
 GACATCGCGACCAACATGGAGGCGAACATCGAAGCGCAAACCTTCATTAAAGGCAAGATCGCGGAAATTGTGAAAGATCCGGCGA
 TTGCGCAGAAGCTGATGCCGCAAGACCTGTACGCGAACGTCCGCTGTGCGATAGCGGTTACTATGAAACCTTCAACCGTGACAAC
 GTGCGTCTGGAAGATGTTAAAGCGAACCCGATCGTGGAGATTACCGAAAACGGCGTTAAGCTGGAGAACGGTGACTTTGTTGAAC
 TGGATATGCTGATCTGCGGACCGGTTTTGATGCGGTGGATGGTAACTACGTTTCGATGGACATTCAGGGCAAGAACGGTCTGGCG
 ATTAAAGATTACTGAAAGAGGGTCCGAGCAGCTATATGGGTGTGACCGTTAACTACCCGAACATGTTTATGGTGTGGGCC
 GAACGGTCCGTTACCAACCTGCCGCCGAGCATCGAGAGCCAAGTTGAATGGATCAGCGACACCATTAAATATGCGGAGGAAAAC
 AACGTGGAGAGCATCGAAGCGACCAAAGAGGCGGAGGAACAGTGGACCCTGACCTGCGCGGAGATTGCGGAAATGACCCTGTTT
 CCGAAAGCGCAAAGCTGGATCTTCGGCGCGAACATCCGGGTAAAAAGAACCCTTACTTCTATCTGGGCGGTCTGAAAGAATA
 CCGTAGCGCGCTGGCGAACTGCAAGAACCACGCGTATGAGGGTTTTGACATCCAACCTGCAACGTAGCGACATTAACAACCGGCG
 AACGCG

1 MSQKMDFDAI VIGAGFGGLY AVKCLRDELE LKVQAFDKAT DVAGTWYWNR YPGALTDDET
 61 HLYCYSWDKE LLQSLEIKKK YVQGPVDRKY LQOVAEKHDL KKSYPFNTAV QSAHYNEADA
 121 LWEVTTEYGD KYTARFLITA LGLLSAPNLP NIKGINQFKG ELHHTSRWPD DVSFEGKRVG
 181 VIGTGSTGVQ VITAVAPLAK HLTVFQRSAQ YSVPIGNDPL SEEDVKKIKD NYDKIWDGVW
 241 NSALAFGLNE STVPAMSVSA EERKAVFEKA WQTGGGFRFM FETFGDIATN MEANIEAQN
 301 IKGKIAEIVK DPAIAQKLMQ QDLYAKRPLC DSGYNTFNR DNVRLDVDKA NPIVEITENG
 361 VKLENGDFVE LDMLICATGF DAVDGNVVRM DIQKNGLA KDYWKEGPSS YMGVTVNNYP
 421 NMFVVLGPNQ PFTNLPPSIE SQVEWISDTI KYAENNVEE IEATKEAEEQ WTRTCAEIAE
 481 MTLFPKAQSW IFGANIPGKK NTVYFYLGLL KEYSALANC KNHAYEGFDI QLQRSDIKQP
 541 ANA

CHMO-QYTV451-454KYAE+Q473R+N477E+G14A+N336E+T415C+A463C (3rd_{mut-3})

ATGAGCCAGAAGATGGACTTTGATGCGATTGTGATTGGTGGTTTTGGTGGTCTGTATGCGGTGAAGAAGCTGCGTGACGAAC
 TGGAGCTGAAAGTGCAGGCGTTTGACAAGGCGACCGATGTTGCGGGCACCTGGTACTGGAACCGTTATCCGGGTGCGCTGACCGA
 CACCGAAACCCACCTGTACTGCTATAGCTGGGATAAAGAGCTGTCTGAGAGCCTGAAATCAAGAAAAAGTATGTGCAAGGTCCG
 GACGTTTCGTAATACTGCAGCAAGTTGCGGAGAAGCACGATCTGAAAAAGAGCTATCAGTTTAACACCGCGGTTCAAAGCGGCAC
 TACAACGAAGCGGACGCGCTGTGGGAAGTGACCACCGAATACGGCGATAAGTATACCGCGGTTTCTGATTACCGCGCTGGGTCT
 GCTGAGCGCGCCGAACCTGCCGAACATCAAAGGCATTAACAGTTTAAAGGGCGAGCTGCACCACACCAGCCGTTGGCCGGACGAT
 GTGAGCTTGAAGGCAAACGTGTGGGTGTTATCGGCACCGGTAGCACCGCGTGCAAGTTATTACCGCGGTGGCGCCGCTGGCGA
 AGCACCTGACCGTTTTTCAGCGTAGCGCGCAATATAGCGTGCCGATCGGTAACGACCCGCTGAGCGAGGAAGATGTTAAAAAGATC
 AAGGACAACACTAGATAAGATTTGGGATGGTGTGTGGAACAGCGCGCTGGCGTTTGGTCTGAACGAGAGCACCGTGCCGGCGATGA
 GCGTTAGCGCGGAGGAACGTAAAGCGGTTTTCGAGAAGGCGTGGCAGACCGGTGGCGGTTTTCCGTTTTATGTTCAACCTTTGG
 CGACATCGCGACCAACATGGAGGCGAACATCGAAGCGCAAACCTTCATTAAGGCAAGATCGCGGAAATTGTGAAAGATCCGGCG
 ATTGCGCAGAAGCTGATGCCGCAAGACCTGTACGCGAAACGTCCGCTGTGCGATAGCGGTTACTATGAAACCTTCAACCGTGACAA
 CGTGCGTCTGGAAGATGTTAAAGCGAACCCGATCGTGGAGATTACCGAAAACGGCGTTAAGCTGGAGAACCGTGACTTTGTTGAA
 CTGGATATGCTGATCTGCGCGACCGTTTTGATGCGGTGGATGGTAACTACGTTCTGATGGACATTACGGGCAAGAACGGTCTGGC
 GATGAAAGATTACTGGAAGAGGGTCCGAGCAGCTATATGGGTGTGTGGTTAACTACCCGAACATGTTTATGGTGTGGGCC
 CGAACCGTCCGTTACCAACCTGCCGCCGAGCATCGAGAGCCAAGTTGAATGGATCAGCGACACCATTAAATATGCGGGAGAAAA
 CAACGTGGAGAGCATCGAATGCCACCAAAGAGGCGGAGGAACAGTGGACCGTACCTGCGCGGAGATTGCGGAAATGACCCTGTTT
 CCGAAAGCGCAAAGCTGGATCTTCGGCGCGAACATCCGGGTAAAAAGAACCCTTACTTCTATCTGGGCGGTCTGAAAGAATA
 CCGTAGCGCGCTGGCGAACTGCAAGAACCACGCGTATGAGGGTTTTGACATCCAACCTGCAACGTAGCGACATTAACAACCGGCG
 AACGCG

1 MSQKMFDAI VIGAGFGGLY AVKCLRDELE LKVQAFDKAT DVAGTWYWNR YPGALTDDET
 61 HLYCYSWDKE LLQSLEIKKK YVQGPVDRKY LQQVAEKHDL KKSYPQNTAV QSAHYNEADA
 121 LWEVTTEYGD KYTARFLITA LGLLSAPNLP NIKGINQFKG ELHHTSRWPD DVSEFEGKRVG
 181 VIGTGSTGVQ VITAVAPLAK HLTVFQRSAQ YSVPIGNDPL SEEDVKKIKD NYDKIWDGVW
 241 NSALAFGLNE STVPAMSVSA EERKAVFEKA WQTGGGFRFM FETFEDIATN MEANIEAQNF
 301 IKGKIAEIVK DPAIAQKLMF QDLYAKRPLC DSGYYNTFNR DNVRLDVDKA NPIVEITENG
 361 VKLENGDFVE LDMLICATGF DAVDGNVVRM DIQKNGLAM KDYWKEGPSS YMGVQVNNYP
 421 NMFVVLGPNP PFTNLPPSIE SQVEWISDTI KYAENNVES IECTKEAEEQ WRTCAEIAE
 481 MTLFPKAQSW IFGANIPGKK NTVYFYLGLL KEYSALANC KNHAYEGFDI QLQRSDIKQP
 541 ANA

CHMO-QYTV451-454KYAE+Q473R+N477E+G14A+N336E+M400I (3rd mut-4)

ATGAGCCAGAAGATGGACTTTGATGCGATTGTGATTGGT GCGGGTTTTGGTGGTCTGTATGCGGTGAAGAAGCTGCGTGACGAAC
 TGGAGCTGAAAGTGCAGGCGTTTGACAAGGCGACCGATGTTGCGGGCACCTGGTACTGGAACCGTTATCCGGGTGCGCTGACCGA
 CACCGAAACCCACTGTACTGCTATAGCTGGGATAAAGAGCTGTCTGAGAGCCTGGAATCAAGAAAAAGTATGTGCAAGGTCCG
 GACGTTTCGTAATACTGCAGCAAGTTGCGGAGAAGCACGATCTGAAAAAGAGCTATCAGTTTAACACCGCGGTTCAAAGCGGCAC
 TACAACGAAGCGGACGCGCTGTGGGAAGTGACCACCGAATACGGCGATAAGTATACCGCGGTTTCCTGATTACCGCGCTGGGTCT
 GCTGAGCGCGCCGAACCTGCCGAACATCAAAGGCATTAACAGTTTAAAGGGCGAGCTGCACCACACCAGCCGTTGGCCGGACGAT
 GTGAGCTTCAAGGCAAACGTGTGGGTGTTATCGGCACCGGTAGCACCGCGTGCAAGTTATTACCGCGGTGGCGCCGCTGGCGA
 AGCACCTGACCGTTTTTCAGCGTAGCGCGCAATATAGCGTGCCGATCGGTAACGACCCGCTGAGCGAGGAAGATGTTAAAAAGATC
 AAGGACAACCTACGATAAGATTTGGGATGGTGTGTGGAACAGCGCGCTGGCGTTTGGTCTGAACGAGAGCACCGTGCCGGCGATGA
 GCGTTAGCGCGGAGGAACGTAAAGCGGTTTTCGAGAAGGCGTGGCAGACCGGTGGCGGTTTTCCGTTTTATGTTCAACCTTTGG
 CGACATCGCGACCAACATGGAGGCGAACATCGAAGCGCAAACCTTATTAAAGGCAAGATCGCGGAAATTGTGAAAGATCCGGCG
 ATTGCGCAGAAGCTGATGCCGCAAGACCTGTACGCGAAACGTCCGCTGTGCGATAGCGGTTACTATGAAACCTTCAACCGTGACAA
 CGTGCGTCTGGAAGATGTTAAAGCGAACCCGATCGTGGAGATTACCGAAAACGGCGTTAAGCTGGAGAACCGTGACTTTGTTGAA
 CTGGATATGCTGATCTGCGCGACCGTTTTGATGCGGTGGATGGTAACTACGTTCTGATGGACATTACGGGCAAGAACGGTCTGGC
 GATTAAAGATTACTGGAAGAGGGTCCGAGCAGCTATATGGGTGTGACCGTTAACTACCCGAACATGTTTATGGTGTGGGGC
 CGAACCGTCCGTTACCAACCTGCCGCCGAGCATCGAGAGCCAAGTTGAATGGATCAGCGACACCATTAAATATGCGGAGGAAAA
 CAACGTGGAGAGCATCGAAGCGACCAAAGAGGCGGAGGAACAGTGGACCCTGACCTGCGCGGAGATTGCGGAAATGACCCGTGT
 TCCGAAAGCGCAAAGCTGGATCTTCGGCGCAACATTCCGGGTAAGAAAGAACACCGTTTACTTCTATCTGGGCGGTCTGAAAGAAT
 ACCGTAGCGCGCTGGCGAACTGCAAGAACCACGCGTATGAGGGTTTTGACATCCAACCTGCAACGTAGCGACATTAAACAACCGGC
 GAACGCG

1 MSQKMFDAI VIGAGFGGLY AVKCLRDELE LKVQAFDKAT DVAGTWYWNR YPGALTDDET
 61 HLYCYSWDKE LLQSLEIKKK YVQGPVDRKY LQQVAEKHDL KKSYPNTAV QSAHYNEADA
 121 LWEVTTEYGD KYTARFLITA LGLLSAPNLP NIKGINQFKG ELHHTSRWPD DVSFEGKRVG
 181 VIGTGSTGVQ VITAVAPLAK HLTVFQRSAQ YSVPIGNDPL SEEDVKKIKD NYDKIWDGVW
 241 NSALAFGLNE STVPAMSVSA EERKAVFEKA WQTGGGFRFM FETFEDIATN MEANIEAQN
 301 IKGKIAEIVK DPAIAQKLMQ QDLYAKRPLC DSGYBTFNR DNVRLDVKI NPIVEITENG
 361 VKLENGDFVE LDMLICATGF DAVDGNVVRM DIQKNGLA KDWKKEGPSS YMGVTVNNYP
 421 NMFVVLGPNP PFTNLPPSIE SQVEWISDTI KYAENNVES IEATKEAEEQ WRTCAEIAE
 481 MTLFPKAQSW IFGANIPGKK NTVYFYLGLL KEYSALANC KNHAYEGFDI QLQRSDIKQP
 541 ANA

CHMO-QYTV451-454KYAE+Q473R+N477E+G14A+ N336E+ T415C+A463C+ M400I (3rdmut-5)

ATGAGCCAGAAGATGGACTTTGATGCGATTGTGATTGGT GCGGGTTTTGGTGGTCTGTATGCGGTGAAGAAGCTGCGTGACGAAC
 TGGAGCTGAAAGTGCAGGCGTTTGACAAGGCGACCGATGTTGCGGGCACCTGGTACTGGAACCGTTATCCGGGTGCGCTGACCGA
 CACCGAAACCCACCTGTACTGCTATAGCTGGGATAAAGAGCTGCTGCAGAGCCTGGAATCAAGAAAAAGTATGTGCAAGGTCCG
 GACGTTTCGTAATACTGCAGCAAGTTGCGGAGAAGCACGATCTGAAAAAGAGCTATCAGTTTAACACCGCGGTTCAAAGCGGCAC
 TACAACGAAGCGGACGCGCTGTGGGAAGTGACCACCGAATACGGCGATAAGTATACCGCGGTTTCTGATTACCGCGCTGGGTCT
 GCTGAGCGCGCCGAACCTGCCGAACATCAAAGGCATTAACAGTTTAAAGGGCGAGCTGCACCACACAGCCGTTGGCCGGACGAT
 GTGAGCTTCAAGGCAAACGTGTGGGTGTTATCGGCACCGGTAGCACCGCGTGCAAGTTATTACCGCGGTGGCGCCGCTGGCGA
 AGCACCTGACCGTTTTTCAGCGTAGCGCGCAATATAGCGTGCCGATCGGTAACGACCCGCTGAGCGAGGAAGATGTTAAAAAGATC
 AAGGACAACCTACGATAAGATTTGGGATGGTGTGTGGAACAGCGCGCTGGCGTTTGGTCTGAACGAGAGCACCGTGCCGGCGATGA
 GCGTTAGCGCGGAGGAACGTAAAGCGGTTTTCGAGAAGGCGTGGCAGACCGGTGGCGGTTTTCCGTTTTATGTTCAACCTTTGG
 CGACATCGGACCAACATGGAGGCGAACATCGAAGCGCAAACCTCATTAAAGGCAAGATCGCGGAAATTGTGAAAGATCCGGCG
 ATTGCGCAGAAGCTGATGCCGCAAGACCTGTACGCGAACGTCGCTGTGCGATAGCGGTTACTATGAAACCTTCAACCGTGACAA
 CGTGCGTCTGGAAGATGTTAAAGCGAACCCGATCGTGGAGATTACCGAAAACGGCGTTAAGCTGGAGAACCGTGACTTTGTTGAA
 CTGGATATGCTGATCTGCGGACCGTTTTGATGCGGTGGATGGTAACTACGTTCTGATGGACATTACGGGCAAGAACGGTCTGGC
 GATTAAAGATTACTGGAAGAGGGTCCGAGCAGCTATATGGGTGTGTGCGTTAACAACTACCCGAACATGTTTATGGTGTGGGCC
 CGAACCGTCCGTTACCAACCTGCCGCCGAGCATCGAGAGCCAAGTTGAATGGATCAGCGACACCATTAAATATGCGGGAGAAAA
 CAACGTGGAGAGCATCGAATGCCACCAAAGAGGCGGAGGAACAGTGGACCCTACCTGCGCGGAGATTGCGGAAATGACCCTGTTT
 CCGAAAAGCGCAAAGCTGGATCTTCGGCGCAACATCCGGGTAAAAAGAACCCTTTACTTCTATCTGGGCGGTCTGAAAGAATA
 CCGTAGCGCGCTGGCGAACTGCAAGAACCACGCGTATGAGGGTTTTGACATCCAACCTGCAACGTAGCGACATTAAACAACCGCGG
 AACGCG

1 MSQKMDFDAI VIGAGFGGLY AVKCLRDELE LKVQAFDKAT DVAGTWYWNR YPGALTDTET
 61 HLYCYSWDKE LLQSLEIKK YVQGPVVRKY LQOVAEKHDL KKS YQFNTAV QSAHYNEADA
 121 LWEVTTEYGD KYTARFLITA LGLLSAPNLP NIKGINQFKG ELHHTSRWPD DVSFEGKRVG
 181 VIGTGSTGVQ VITAVAPLAK HLTVFQRSAQ YSVPIGNDPL SEEDVKKIKD NYDKIWDGVW
 241 NSALAFGLNE STVPAMSVSA EERKAVFEKA WQTGGGFRFM FETFGDIATN MEANIEAQN
 301 IKGKIAEIVK DPAIAQKLMQ QDLYAKRPLC DSGYYETFN RDNVRLEDVKA NPIVEITENG
 361 VKLENGDFVE LDMLICATGF DAVDGNVVRM DIQKNGLA KDYWKEGPSS YMGVQVNNYP
 421 NMFVVLGPNP PFTNLPPSIE SQVEWISDTI KYAENNVES IECTKEAEEQ WRTCAEIAE
 481 MTLFPKAQSW IFGANIPGKK NTVYFYLGLL KEYRSALANC KNHAYEGFDI QLQRSDIKQP
 541 ANA

Appendix

TmCHMO

ATGAGCACCACCCAGACCCCGGACCTGGATGCGATCGTGATTGGCGCGGGCTTCGGCGGTATCTACATGCTGCACAACTGCGTAA
 CGACCTGGGTCTGAGCGTGCCTGTTTTGAGAAGGGTGGCGGTGTTGGCGGTACCTGGTACTGGAACAAGTATCCGGGCGCGAA
 AGCGATACCGAAGGTTTTGTGTACCGTTATAGCTTCGACAAAGAGCTGCTGCGTGAATACGATTGGACCACCCGTTATCTGGACCA
 GCCGGATGTTCTGGCGTACCTGGAGCACGTGTTGAACGTTATGACCTGGCGCGTGATATCCAACCTGAACACCGAGGTGACCGAC
 GCGATTTTTGATGAGGAAACCGAGCTGTGGCGTGTACCACCGCGGGCGGTGAAACCTGACCGCGCGTTTCTGGTGACCGCGCT
 GGGTCTGCTGAGCCGTAGCAACATCCCGGACATTCCGGGCCGTGATAGCTTCGCGGGTCTGCTGGTTCACACCAACGCGTGGCCGG
 AAGACCTGGATATCACCAGCAAGCGTGTGGGTGTTATTGGCACCGGTAGCACCGGTACCCAGTTTATCGTGCGGCGGCGAAAAAT
 GCGGAGCAACTGACCGTGTCCAGCGTACCCGCAATACTGCGTTCGAGCGGCAACGTCGATGGACCCGGATGAAGTTGCG
 CGTATCAAGCAGAACTTTGACAGCATTGGGATCAAGTGCCTAGCAGCACCGTTGCGTTCGGCTTTGAGGAAAGCACCGTGGAGGC
 GATGAGCGTTAGCGAGAGCGAACGTCAGCGTGTTCAGCAAGCGTGGGACAAAGGCAACGTTTCCGTTTTATGTTCCGGTACCT
 TTTGCGATATCGCGACCAACCCGAGGCGAACGCGGCGGCGGCGGCGTTCATTTCGTAGCAAGATCGCGGAGATTGTGAAAGACCC
 GGAAACCGCGCGTAAGCTGACCCGACCGATCTGTACGCGAAACGTCGCTGTGCAACGAGGGTACTATGAAACCTATAACCGTG
 ACAACGTGAGCCTGGTTAGCCTGAAGGAAACCCGATCGAGGAAATTGTGCCGAAGGCGTTCGTACCAGCGACGGTGTGGTTCA
 CGAACTGGATGTGCTGGTTTTTTCGACCGGCTTCGACGCGGTTGATGGTAACTACCGTGCATGAACCTGCGTGGCCGTGATGGTC
 GTCACATCAACGAGCACTGGACCGAAGGCCGACCGACTATCTGGGCGTGACCAAGGCGGGTTTTCCGAACATGTTTCATGATCCTG
 GGCCCGAACGGTCCGTTTACCAACCTGCCGCCGAGCATTGAGGCGCAGGTTGAATGGATCAGCGACCTGATTGATAAAGCGACCC
 GTGAGGGTCTGACCACCGTGAACCGACCGCGGATGCGGAGCGTGAATGGACCGAAACCTGCGCGGAAATTGCGAACATGACCCCT
 GTTCCGAAGGCGGATAGCTGGATCTTCGCGCGAACATTCCGGGTAACGTCACGCGGTTATGTTCTACCTGGGCGGTCTGGGTA
 ACTATCGTCGTAACCTGGCGGACGTGGCGGACGGC GGCTATCGTGGTTTCCAACCTGCGTGGTGAACGTGCGCAAGCGGTGGCG

1 MSTTQTPDL AIVIGAGFG IYMLHKL RND LGLSVRVFEK GGGVGGTWY NKYPGAKSDT
 61 EGFVYRYSFD KELLREYDWT TRYLDQPDVL AYLEHVVERY DLARDIQLNT EVTDAIFDEE
 121 TELWRVTTAG GETLTARFLV TALGLLSRSN IPDIPGRDSF AGRVHTNAW PEDLDITGKR
 181 VGVIGTGSTG TQFIVAAAKM AEQLTVFQRT PQYCVPSGNG PMDPDEVARI KQNFDSIWDQ
 241 VRSSTVAFGF EESTVEAMSV SESERQRFVQ QAWDKNGFR FFMGTFCDIA TNPEANAAAA
 301 AFIRSKIAEI VKDPETARKL TPTDLYAKRP LCNEGYYETY NRDNVSLVSL KETPIEIVP
 361 QGVRTSDGVV HELDVLVFAT GFDVAVDGNR AMNLRGRDGR HINEHWTEGP TSYLGVTKAG
 421 FPNMFILGP NGPFTNLPPS IEAQVEWISD LIDKATREGL TTVEPTADAE REWTETCAEI
 481 ANMTLFPKAD SWIFGANIPG KRHAVMFYLG GLGNYRRQLA DVADGGYRGF QLRGERAQAV
 541 A

Appendix

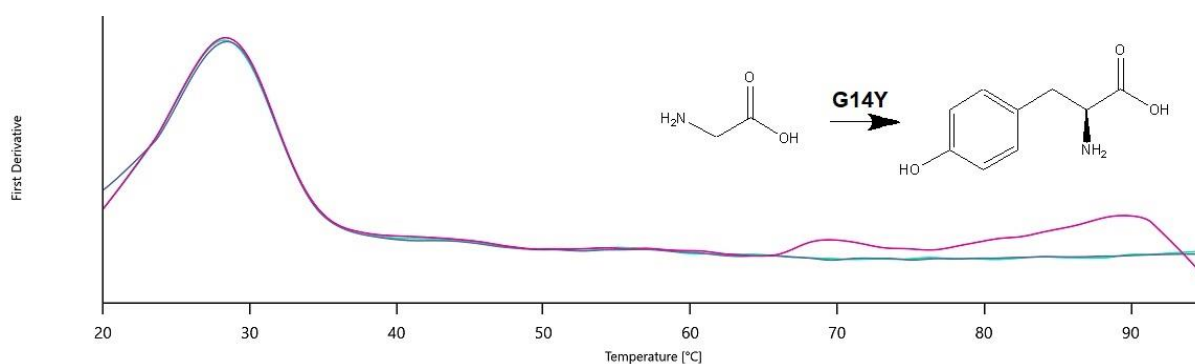
BVMO_{flava}

ATGAGCACCACCCGTACCCCGGACGTGGATGCGATCGTTATTGGCGCGGGTTTTGGTGGCATCTACATGCTGC
 ACAAGCTGCGTAACGAACTGGGTCTGAGCGTGACCGCGTTTGAGAAAGGTGGCGGTGTTGGCGGTACCTGGA
 CTTCAACCGTTATCCGGGTGCGAAGAGCGACACCGAAGGCTTTGTGTACCGTTATAGCTTCGACAAAGATCTG
 CTGCGTGAATGGAAGTGGACCACCCGTTACCTGGAGCAGGCGGATGTTCTGGCGTATCTGGAACACGTGGTTA
 GCGTTTTGACCTGGGTCTGATATCCGTCTGAACACCGAAGTGACCGGCGCGTTTTTCGACGAGGAAAGCGAT
 CTGTGGACCGTGACCACCGCGACCGGCGAGACCACCGCGCGTTATCTGGTTAACGCGCTGGGTCTGCTGG
 CGCGTAGCAACATCCCGGACATTCCGGGTCTGATGGTTTTGCGGGCCGTCTGGTGCACACCAACGCGTGGCC
 GGACGATCTGGACATCACCGGCAAGCGTGTGGGCGTTATTGGCACCGGTAGCACCGGTACCCAGTTTATCATT
 GCTGCGGCGAAAACCGCGAGCCACCTGACCGTGTCCAGCGTAGCCCGCAATATTGCGTTCGAGCGGTAACG
 GTCCGGTGGACCAGACCGAAGTTGATCGTACCAAGGAGAACTTTGACGCGATCTGGGATCAAGTGCGTAACA
 GCGTGGTTGCGTTCGTTTTGAGGAAAGCGGCGTGGAAAGCGATGAGCGTTAGCGAGGAAGAGCGTCGTAAG
 GTTTTCCAAGAGGCGTGGGACAAAGGCAACGGTTTTCCGTTTTATGTTTCGGTACCTTTTTCGATATCGCGACCAA
 CCCGGAAGCGAACGCTGCGGCGGCGGCGTTCATTCGTGCGAAGATCGCGGAAATTGTTGACGATCCGGAGAC
 CGCGCGTAAGCTGACCCCGACCGACCTGTACGCGAAACGTCCGCTGTGCAACGAAGGTTACTATGAGACCTAT
 AACCGTGATAACGTGGAAGTGGTTAGCATCAAAGAAAACCCGATCCGTGAGATTACCCCGCGGGTGTGCGT
 ACCGCGGATGGTATTGAGCACCCGCTGGATGTGCTGGTTTTTTCGACCGGTTTTTCGACGCGGTTGATGGCAACT
 ATCGTGCGATGGACCTGCGTGGTCTGCGGTCGTACATCAGCGAACACTGGACCGGCGGTTCCGACCAGCT
 ATCTGGGTGTGAGCACCGCGGCTTTCCGAACATGTTTCATGATCCTGGGTCCGAACGGTCCGTTACCAACCT
 GCCGCCGAGCATTGAGACCCAAGTTGATTGGATCGGTGAACTGATTCGTTCATGCGGAGCGTACCGGCGTGCG
 TACCGTTGAGCCGACCGCTGCGGCGGAAGAGGCGTGGACCGCGACCTGCGCGGAGATTGCGGACATGACCCCT
 GTTTCCGAAGGCGGATAGCTGGATCTTCGGTTCGGAACATTCCGGGCAAACGTAACGCGGTGATGTTTTACCTG
 GCGGGCCTGGGTGCGTATCGTGCGAACTGCGTGAGGTTGCGGACGCGGGTTACACCGGCTTCGAACTGACC
 CGTGAGAACGCGACCGCGGCGGTG

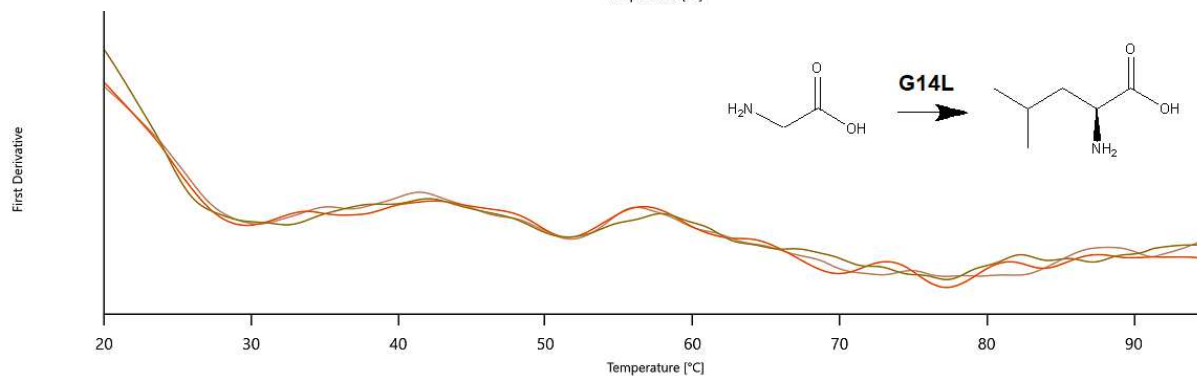
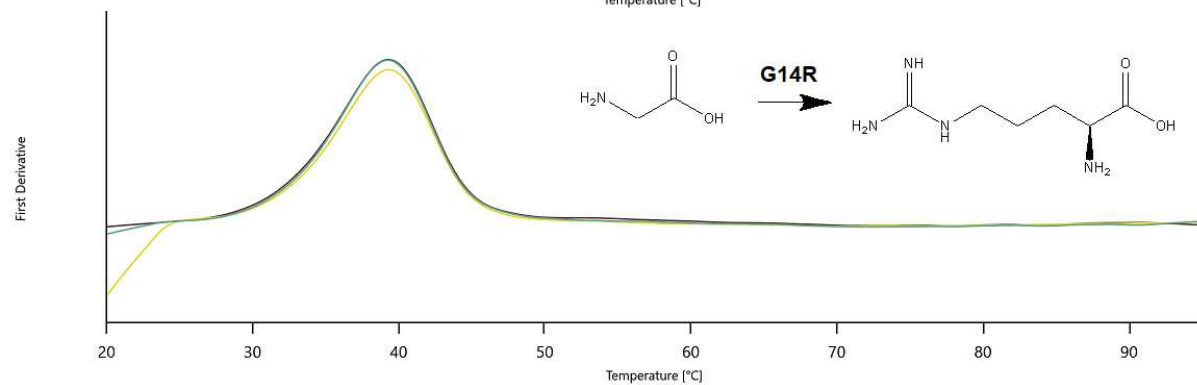
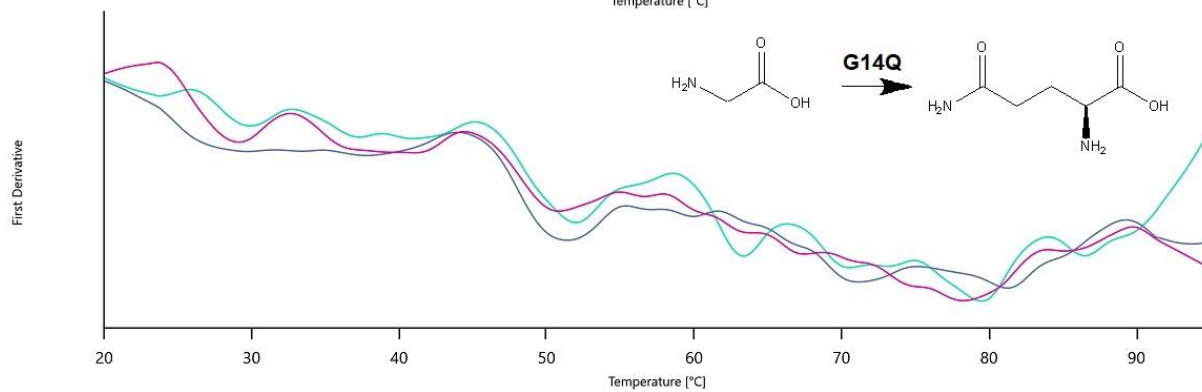
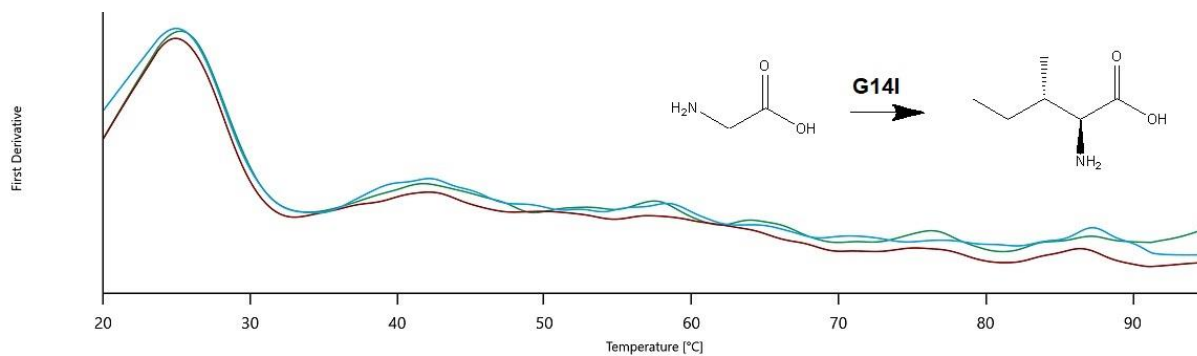
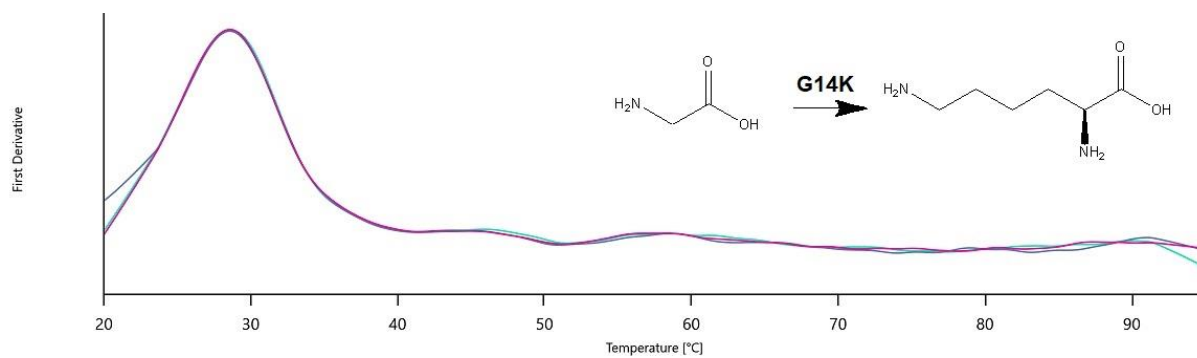
1 MSTTRTPDVD AIVIGAGFGG IYMLHKLRNE LGLSVTAFEK GGGVGGTWYF NRYPGAKSDT
 61 EGFVYRYSFD KDLLREWNWT TRYLEQADVL AYLEHVVERF DLGRDIRLNT EVTGAVFDEE
 121 SDLWTVTTAT GETTTARYLV NALGLLARSN IPDIPGRDGF AGRLVHTNAW PDDLITGKR
 181 VGVIGTGSTG TQFIIAAAKT ASHLTVFQRS PQYCVPSGNG PVDQTEVDRT KENFDAIWDQ
 241 VRNSVVAFGF EESGVEAMSV SEEERRKVFQ EAWDKGNGFR FFMGTFCDIA TNPEANAAAA
 301 AFIRAKIAEI VDDPETARKL TPTDLYAKRP LCNEGYEYTY NRDNVELVSI KENPIREITP
 361 AGVRTADGIE HPLDVLVFAT GFDAVDGNYR AMDLRGRGGR HISEHWTGGP TSYLGVSTAG
 421 FPNMFILGP NGPFTNLPPS IETQVDWIGE LIRHAERTGV RTVEPTAAAE EAWTATCAEI
 481 ADMTLFPAKAD SWIFGANIPG KRNAVVFYLA GLGAYRAKLR EVADAGYTGF ELTRENATAA
 541 V

F I.3 Melting chromatograms for mutants

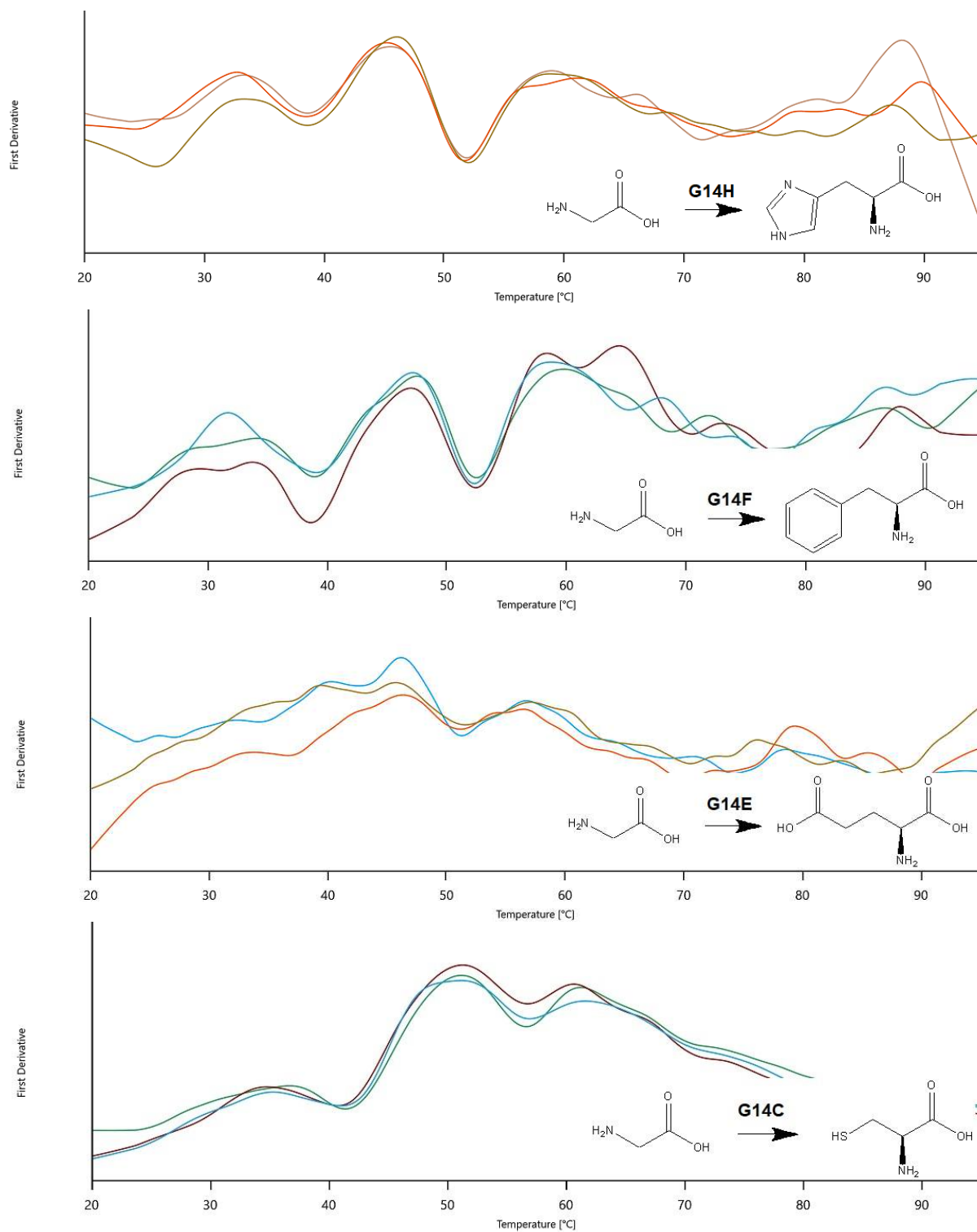
Variant	T _m (°C)
wt	38.2
1 st mut-1	39.6
1 st mut-2	38.7
1 st mut-3	37.8
1 st mut-4	38
1 st mut-5	39.4
1 st mut-6	38.9
1 st mut-7	39.1
2 nd mut-1	40.2
2 nd mut-2	38.2
2 nd mut-3	38.1
2 nd mut-4	40.1
2 nd mut-5	40.9
3 rd mut-1	41.4
3 rd mut-2	42.6
3 rd mut-3	42
3 rd mut-4	42.5
3 rd mut-5	42.7



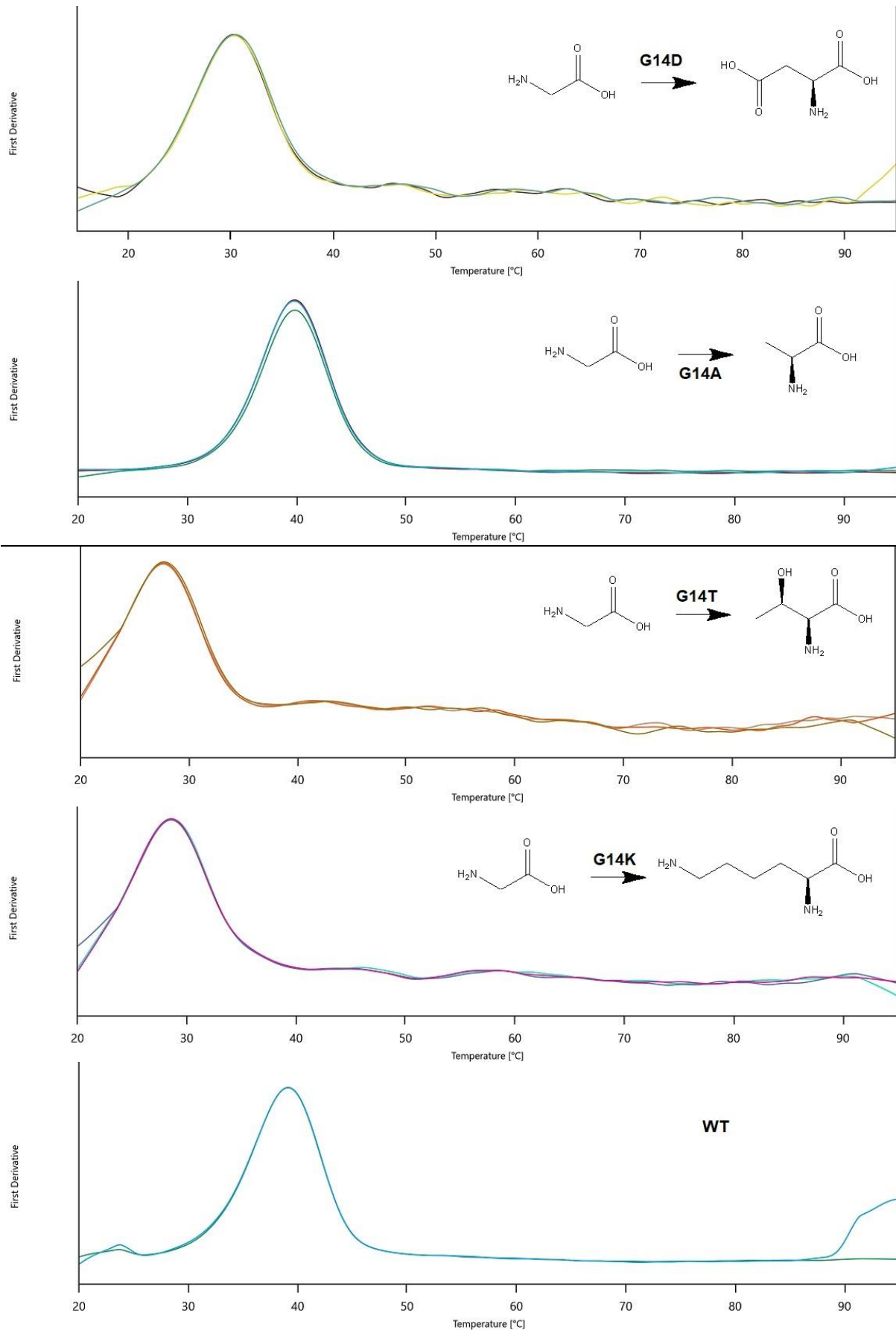
Appendix



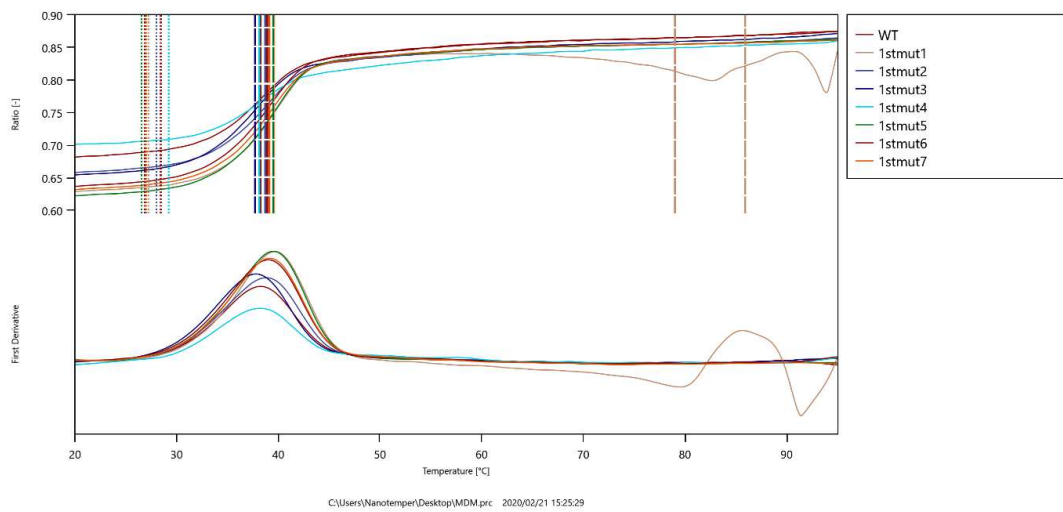
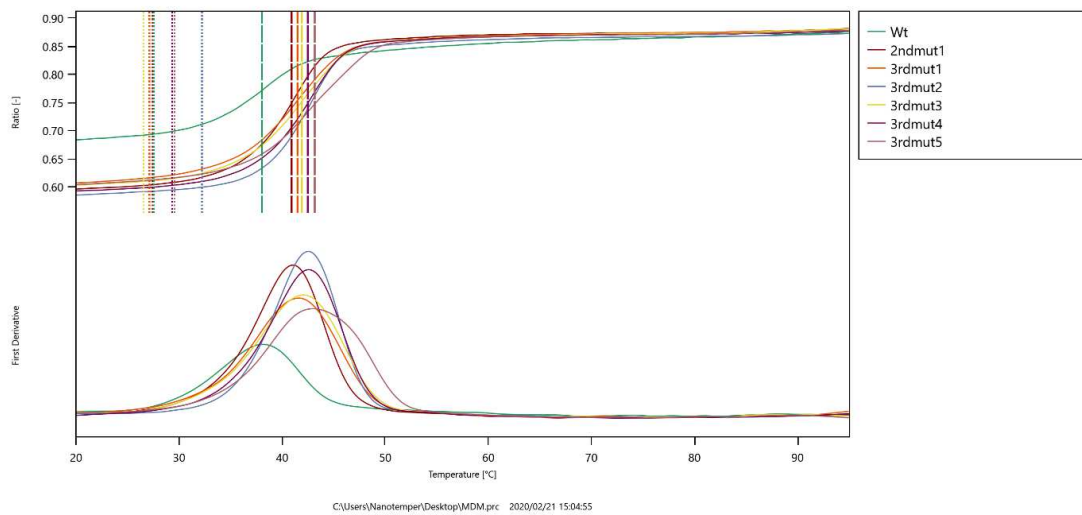
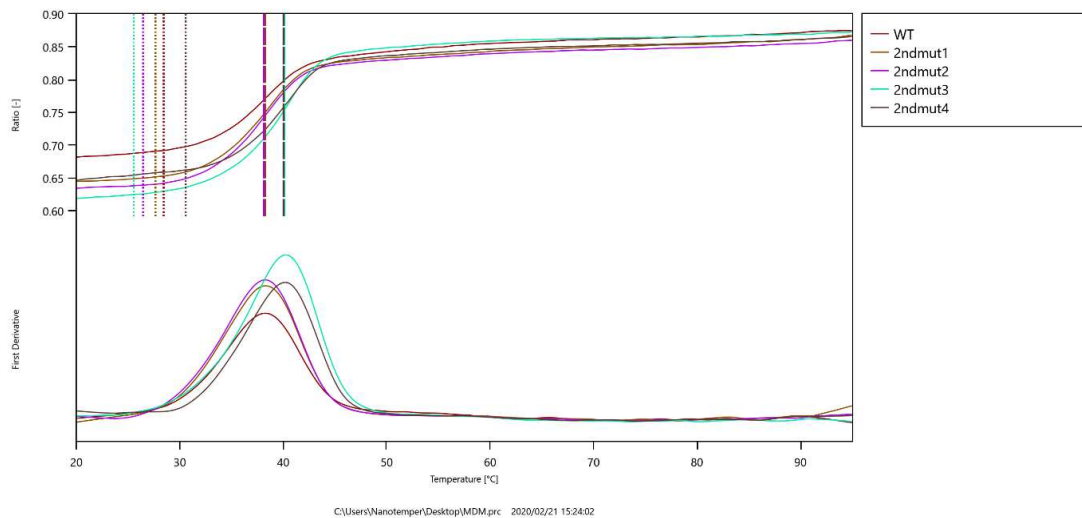
Appendix



Appendix



Appendix



F I.4 GC methods

Substrate	Condition	Retention time
1a	80 °C (2 min), 5 °C/min, 160 °C (1 min), 10 °C/min, 220 °C (8 min). Helium: 2 mL/min. Column: BGB173, 30 m x 0.25 mm ID, 0.25 µm.	1a: 6.08 min (S)-2a: 14.33 min (R)-2a : 14.75 min
1b	80 °C (2 min); 2 °C/min, 220 °C (8 min), Helium: 2 mL/min. Column: BGB175, 30 m x 0.25 mm ID, 0.25 µm.	1b: 34.8 (S)-2b: 50.997 (R)-2b: 51.282
1c	80 °C (2 min), 5 °C/min, 160 °C (1 min), 10 °C/min, 220 °C (8 min). Helium: 2 mL/min. Column: BGB173, 30 m x 0.25 mm ID, 0.25 µm.	1c: 13.143 min (R)-2c: 22.512 min (S)-2c: 22.583 mi
1d	80 °C (2 min), 5 °C/min, 160 °C (1 min), 10 °C/min, 220 °C (8 min). Helium: 2 mL/min. Column: BGB173, 30 m x 0.25 mm ID, 0.25 µm.	1d: 21.855 min (+)-2d: 27.8 min (-)-2d: 28.073 mi
2	80 °C (2 min), 5 °C/min, 160 °C (1 min), 10 °C/min, 220 °C (8 min). Helium: 2 mL/min. Column: BGB173, 30 m x 0.25 mm ID, 0.25 µm.	2: 5.342 min (4R;6S)-4: 18.53 min (4S;6R)-4: 18.62 min
3	80 °C (2 min), 5 °C/min, 160 °C (1 min), 10 °C/min, 220 °C (8 min). Helium: 2 mL/min. Column: BGB173, 30 m x 0.25 mm ID, 0.25 µm.	3: 7.305 or 7.492 min P(-)-12a: 19.508 min P(+)-12a: 19.073 min D(-)-12b: 19.41 min D(+)-12b: 19.635 mi
4a	110 °C (2 min); 10 °C/min, 118 °C; 2 °C/min, 122 °C; 25 °C/min, 200 °C (1 min); 50 °C/min, 220 °C (4 min), Helium: 2 mL/min. Column: BGB175, 30 m x 0.25 mm ID, 0.25 µm.	4a: 9.1min (S)-14a: 11.1 min

		(R)-14a: 11.5 min
		4b: 39.83 min
4b	80 °C (2 min); 2 °C/min, 220 °C (8 min), Helium: 2 mL/min. Column: BGB175, 30 m x 0.25 mm ID, 0.25 µm.	(S)-16a: 57.6 min (R)-16a: 57.88 min
		5a: 26.187 min
5a	80 °C (2 min); 2 °C/min, 220 °C (8 min), Helium: 2 mL/min. Column: BGB175, 30 m x 0.25 mm ID, 0.25 µm.	(S)-6a: 46.767 min (R)-6a: 46.997mi
		5b: 41.713 min
5b	80 °C (2 min); 2 °C/min, 220 °C (8 min), Helium: 2 mL/min. Column: BGB175, 30 m x 0.25 mm ID, 0.25 µm.	(S)-6b: 60.803 min (R)-6b: 61.1 min
		6: 4.647 min
		N(-)-8a: 14.233 min
6	80 °C (2 min), 5 °C/min, 160 °C (1 min), 10 °C/min, 220 °C (8 min). Helium: 2 mL/min. Column: BGB173, 30 m x 0.25 mm ID, 0.25 µm.	N(+)-8a: 14.31 min ABN(-)-8b: 13.133 min ABN(+)-8b:
		8: 7.357 min
		N(-)-10a: 16.37 min
		N(+)-10a: 16.22 min
7	80 °C (2 min), 5 °C/min, 160 °C (1 min), 10 °C/min, 220 °C (8 min). Helium: 2 mL/min. Column: BGB173, 30 m x 0.25 mm ID, 0.25 µm.	ABN(-)-10b: 15.28 min ABN(+)-10b: 17.19 min

F I.5 List of primers

Primers were designed by using the NEBBaseChanger online tool and synthesized by Merck KGaA:

List of primers.

Mutation	Forward	Reverse
G14F	CGTGATTGGTtttGGTTTTGGCG	ATAGCATCAAATCCATTTTTTG
G14L	CGTGATTGGTctgGGTTTTGGCG	ATAGCATCAAATCCATTTTTTG
G14I	CGTGATTGGTattGGTTTTGGCG	ATAGCATCAAATCCATTTTTTG
G14M	CGTGATTGGTatgGGTTTTGGCG	ATAGCATCAAATCCATTTTTTG
G14V	CGTGATTGGTgtgGGTTTTGGCG	ATAGCATCAAATCCATTTTTTG
G14S	CGTGATTGGTagcGGTTTTGGCG	ATAGCATCAAATCCATTTTTTG
G14P	CGTGATTGGTccgGGTTTTGGCG	ATAGCATCAAATCCATTTTTTG
G14T	CGTGATTGGTaccGGTTTTGGCG	ATAGCATCAAATCCATTTTTTG
G14Y	CGTGATTGGTtatGGTTTTGGCG	ATAGCATCAAATCCATTTTTTG
G14H	CGTGATTGGTcatGGTTTTGGCG	ATAGCATCAAATCCATTTTTTG
G14Q	CGTGATTGGTcagGGTTTTGGCG	ATAGCATCAAATCCATTTTTTG
G14N	CGTGATTGGTaacGGTTTTGGCG	ATAGCATCAAATCCATTTTTTG
G14K	CGTGATTGGTaaaGGTTTTGGCG	ATAGCATCAAATCCATTTTTTG
G14D	CGTGATTGGTgatGGTTTTGGCG	ATAGCATCAAATCCATTTTTTGTGACATG
G14E	CGTGATTGGTgaaGGTTTTGGCG	ATAGCATCAAATCCATTTTTTG
G14C	CGTGATTGGTtgcGGTTTTGGCG	ATAGCATCAAATCCATTTTTTG
G14W	CGTGATTGGTtggGGTTTTGGCG	ATAGCATCAAATCCATTTTTTGTG
G14R	CGTGATTGGTcgcGGTTTTGGCG	ATAGCATCAAATCCATTTTTTG

F I.6 List of sequences used for the phylogenetic tree analysis

BVMO sequences used to infer the phylogenetic tree.

Enzyme	Accession number	Amino acids identity
BVMO <i>Leptospira biflexa</i>	ABZ97795	-
CHMO <i>Acinetobacter</i> sp. NCIMB 9871	BAA86293	140/464 (30%)
BVMO <i>Acinetobacter radioresistens</i> S13	ADF32068	117/435 (27%)
CHMO <i>Arthrobacter</i> sp. L661	ABQ10653	136/472 (29%)
CHMO <i>Arthrobacter</i> sp. BP2	AAN37479	139/472 (29%)
CHMO <i>Brachymonas petroleovorans</i>	AAR99068	142/488 (29%)
CHMO1 <i>Brevibacterium</i> sp. HCU	AAG01289	133/464 (29%)
CHMO2 <i>Brevibacterium</i> sp. HCU	AAG01290	123/499 (25%)
CPMO <i>Comamonas</i> sp. NCIMB 9872	BAC22652	125/468 (27%)
BVMO3 <i>Dietzia</i> sp. D5	AHE80562	100/417 (24%)
BVMO4 <i>Dietzia</i> sp. D5	AGY78320	128/489 (26%)
ACMO <i>Gordonia</i> sp. TY-5	BAF43791	130/494 (26%)
BVMO1 <i>Mycobacterium tuberculosis</i> H37Rv	CAA97398	201/461 (44%)
BVMO2 <i>Mycobacterium tuberculosis</i> H37Rv	CAA17436	96/346 (28%)
BVMO3 <i>Mycobacterium tuberculosis</i> H37Rv	CAB06212	108/426 (25%)
BVMO4 <i>Mycobacterium tuberculosis</i> H37Rv	CAB02175	176/488 (36%)
BVMO5 <i>Mycobacterium tuberculosis</i> H37Rv	CAA16134	185/471 (39%)
BVMO6 <i>Mycobacterium tuberculosis</i> H37Rv	CAA16141	120/426 (28%)
BVMO <i>Oceanicola batsensis</i> HTCC2597	A3U3H1	129/486 (27%)
BVMO <i>Parvibaculum lavamentivorans</i> NCIMB13966	A7HU16	136/497 (27%)
CHMO <i>Polaromonas</i> sp. JS666	YP_552312	146/492 (30%)
CPDMO <i>Pseudomonas</i> sp. HI-70	BAE93346	130/450 (29%)
CHMO <i>Pseudomonas aeruginosa</i> PAO1	AAG04927	169/495 (34%)
BVMO <i>Pseudomonas fluorescens</i> DSM50106	AAC36351	172/481 (36%)
HAPMO <i>Pseudomonas fluorescens</i> ACB	AAK54073	163/504 (32%)
HAPMO <i>Pseudomonas putida</i> JD1	ACJ37423	157/477 (33%)

Appendix

BVMO <i>Pseudomonas putida</i> KT2440	AAN68413	113/420 (27%)
OTEMO <i>Pseudomonas putida</i>	AEZ35248	146/504 (29%)
BVMO <i>Pseudomonas veronii</i> MEK700	ABI15711	117/442 (26%)
CHMO <i>Rhodococcus</i> sp. HI-31	BAH56677	142/474 (30%)
CHMO <i>Rhodococcus</i> sp. TK6	AAR27824	143/474 (30%)
CHMO <i>Rhodococcus</i> sp. Phi2	AAN37491	141/474 (30%)
CHMO <i>Rhodococcus</i> sp. Phi1	AAN37494	139/462 (30%)
BVMO1 <i>Rhodococcus jostii</i> RHA1	ABG98452	128/495 (26%)
BVMO2 <i>Rhodococcus jostii</i> RHA1	ABG96095	168/464 (36%)
BVMO3 <i>Rhodococcus jostii</i> RHA1	ABG95050	135/506 (27%)
BVMO4 <i>Rhodococcus jostii</i> RHA1	ABG94866	120/468 (26%)
BVMO5 <i>Rhodococcus jostii</i> RHA1	ABG93916	162/481 (34%)
BVMO6 <i>Rhodococcus jostii</i> RHA1	ABG93685	164/484 (34%)
BVMO7 <i>Rhodococcus jostii</i> RHA1	ABG97785	183/488 (38%)
BVMO9 <i>Rhodococcus jostii</i> RHA1	ABH00079	139/491 (28%)
BVMO10 <i>Rhodococcus jostii</i> RHA1	ABH00083	148/504 (29%)
BVMO11 <i>Rhodococcus jostii</i> RHA1	ABG98471	128/477 (27%)
BVMO12 <i>Rhodococcus jostii</i> RHA1	ABG98876	167/457 (37%)
BVMO13 <i>Rhodococcus jostii</i> RHA1	ABG95573	102/374 (27%)
BVMO14 <i>Rhodococcus jostii</i> RHA1	ABG95240	121/464 (26%)
BVMO15 <i>Rhodococcus jostii</i> RHA1	ABG94297	130/488 (27%)
BVMO16 <i>Rhodococcus jostii</i> RHA1	ABG94724	114/430 (27%)
BVMO17 <i>Rhodococcus jostii</i> RHA1	ABG97009	191/491 (39%)
BVMO18 <i>Rhodococcus jostii</i> RHA1	ABG97176	150/499 (30%)
BVMO19 <i>Rhodococcus jostii</i> RHA1	ABG97302	166/483 (34%)
BVMO20 <i>Rhodococcus jostii</i> RHA1	ABG99184	123/492 (25%)
BVMO21 <i>Rhodococcus jostii</i> RHA1	ABH00380	130/488 (27%)
BVMO24 <i>Rhodococcus jostii</i> RHA1	ABG97104	145/489 (30%)
STMO <i>Rhodococcus rhodochrous</i> IFO 3338	BAA24454	146/497 (29%)
CDMO <i>Rhodococcus ruber</i> SC1	AAL14233	136/487 (28%)
BVMO PntE <i>Streptomyces arenae</i>	ADO85575	132/514 (26%)
BVMO PtlE <i>Streptomyces avermitilis</i> MA-4680	BAC70705	122/447 (27%)
BVMO1 <i>Streptomyces coelicolor</i> A3(2)	CAB55657	129/416 (31%)

Appendix

BVMO2 <i>Streptomyces coelicolor</i> A3(2)	CAB59668	178/485 (37%)
BVMO PenE <i>Streptomyces exfoliatus</i>	ADO85591	125/513 (24%)
BVMO (InfQ) <i>Streptomyces</i> sp. RI-77	BAU98044.1	161/501 (32%)
PAMO <i>Thermobifida fusca</i>	Q47PU3	144/527 (27%)
CHMO <i>Xanthobacter</i> sp. ZL5	CAD10801	129/442 (29%)
BVMO <i>Aspergillus clavatus</i> NRRL1	XP_001270542	149/527 (28%)
BVMO210 <i>Aspergillus flavus</i> NRRL3357	XP_002375343	123/466 (26%)
BVMO456 <i>Aspergillus flavus</i> NRRL3357	XP_002375466	115/490 (23%)
BVMO619 <i>Aspergillus flavus</i> NRRL3357	XP_002383043	133/492 (27%)
BVMO838 <i>Aspergillus flavus</i> NRRL3357	XP_002375657	130/512 (25%)
BVMO1 <i>Aspergillus fumigatus</i> Af293	XP_747160	136/483 (28%)
BVMO2 <i>Aspergillus fumigatus</i> Af293	XP_746949	98/397 (25%)
BVMO3 <i>Aspergillus fumigatus</i> Af293	XP_755274	135/489 (28%)
BVMO <i>Cyanidioschyzon merolae</i> 10D	BAM80902	148/543 (27%)
CAMO <i>Cylindrocarpon radicum</i> ATCC 11011	AET80001	138/483 (29%)
BVMO <i>Physcomitrella patens</i>	XP_001758613	154/532 (29%)

F I.7 Multiple sequence alignment of BVMO_{Flava} with known BVMOs

Figure F.1. Multiple sequence alignment. 4-hydroxyacetophenone monooxygenase (HAPMO)^[35] [Pseudomonas fluorescens] (Q93TJ5.1), cyclododecanone monooxygenase (CDMO)^{[36] [36] [36] [36] [36]} [36] 45 45 45 45 45 [36] [36] [35] [30] 38[2] [Rhodococcus ruber] (AAL14233.1); cyclopentanone 1,2-monooxygenase (CPMO)^[167] [Acidovorax sp. SCN 65-28] (ODS80058.1); phenylacetone monooxygenase (PAMO)^[42] [Thermobifida fusca] (WP_011291921.1); cyclohexanone 1,2-monooxygenase (CHMO)^[31] [Acinetobacter sp. NCIMB9871] ,(BAA86293.1); NAD(P)/FAD-dependent oxidoreductase (TmCHMO)^[56] [Thermocrispum municipale] (WP_028849141); NAD(P)/FAD-dependent oxidoreductase (BVMO_{Flava}) [Amycolatopsis thermoflava] (WP_027929099.1). Rossmann-fold motifs (GxGxxG/A) and consensus sequences of Type I BVMOs (G/AGxWxxxxF/YPG/MxxxD and FxGxxxHxxxWP/D) are highlighted. Multiple sequence alignment was performed by MAFFT V.7. ^[203]

```

BVMOflava MS-----
TMCHMO MS-----
CHMO MS-----
PAMO MA----GQTTVD-----
CPMO MSKVTPQQLSMN-----
CDMO MT-----TSID-----
HAPMO MSAFNTTLP SLDYDDDTLREHLQGADIPTLLLLTVAHLTGD LQILKPNWKPSIAMGVARSG
* :
BVMOflava -----
TMCHMO -----
CHMO -----
PAMO -----
CPMO -----
CDMO -----REALRRKYAEERDKRI-----RPDGNDQYIRLDHV DGW SHDPYMPI-----
HAPMO MDLETEAQVREFCLQRLIDFRDSGQPAPGRPTSDQLHILGTWLMG PVI EPYLP LIAEEAV

BVMOflava TTRTP-----DVDAIVI GAGFGGIYMLHKLRNELGL-SVTAF EKG GGVGG
TMCHMO TTQTP-----DLDAIVI GAGFGGIYMLHKLRNDLGL-SVRVFEKGGGVGG
CHMO --QKM-----DFDAIVI GGGFGLYAVKKLRDELEL-KVQAFDKATDVAG
PAMO SRRQP-----PEEVDVLVV GAGFSGLYALYRLR-ELGR-SVHVIETAGDVGG
CPMO NSVDD-----TLDVLLI GAGFTGLYQLHHLR-KLGF-KVHLVDAGADVGG
CDMO TPREP-----KLDHVTFAFI GGGFSGLVTAARLR-ESGVESVRIIDKAGDFGG
HAPMO TAEEDLRAPRWHKDHVASGRDFKVVI I GAGESGMIAALRFK-QAGV-PFV IYEKGN DVG

```

. .:*. * : : : . . : . . . *

Appendix

BVMOflava **TWYFNRYPGA**KSDETEGFVYRYSFDKDLLREWNWTTTRYLEQADVLAYLEHVVERFDLGRDI
 TMCHMO **TWYWNKYPGA**KSDETEGFVYRYSFDKELLREYDWTTRYLDQPDVLAYLEHVVERYDLARDI
 CHMO **TWYWNRYPGAL**TDTEHLYCYSWDKELLQSLEIKKKYVQGPDVRKYLQQVAEKHDLKKSYS
 PAMO **VWYWNRYPGAR**CDIESIEYCYSFSEEVLEQEWNTERYASQPEILRYINFVADKFDLRSGI
 CPMO **IWHWNCYPGAR**VDTHCQIYQYSM-PELWGEFNNWELFPNWAQMREYFYFVDKLELSKDI
 CDMO **VWYWNRYPGAM**CDTAAMVYMPLEET---GYMPTKEYAHGPEILEHCQRIGKHYDLYDDA
 HAPMO **TWRENTYPGCR**VDINSFWYSFSFARGI-----WDDCFAPAPQVFAYMQAVAREHGLYEH
 * * * * . * * : . : : : . *

BVMOflava RLNTEVTGAVFDEESDLWTV---TTATGETTTARYLVNALGLLARSNIPDIPGRDG**FAGR**
 TMCHMO QLNTEVTDIAIFDEETELWRV---TTAGGETLTARFLVTALGLLSRSNIPDIPGRDS**FAGR**
 CHMO QFNATVQSAHYNEADALWEV---TTEYGDKYTARFLITALGLLSAPNLPNIKGINQ**FKGE**
 PAMO TFHTTVTAAAFDEATNTWTV---DTNHGDRIRARYLIMASGQLSVPQLPNFPGLKD**FAGN**
 CPMO SFNTRVQSAVFEQRREWTV---RSLGHQPIRAKFVIANLGFASPSSTPKVEGIEK**FKGE**
 CDMO LFHTEVTDLVWQEHQWRWI---STNRGDHFTAQFVGMGTGPLHVAQLPGIPGIES**FRGK**
 HAPMO RFNTEVSDAHWDESTQRWQLLYRDSEGQTQVDSNVVVFVAVGQLNRPMIPAIPIGIET**FKGP**
 : : * : : * * : : : : : * . * . * . * *

BVMOflava **LVHTNAWP**DD-----LD-ITGKRVGVI**GTGSTG**TQFIIAAAKTASHLTVFQRSPO
 TMCHMO **LVHTNAWP**ED-----LD-ITGKRVGVI**GTGSTG**TQFIVAAAKMAEQLTVFQRTPO
 CHMO **LHHTSRWP**DD-----VS-FEGKRVGVI**GTGSTG**VQVITAVAPLAKHLTVFQRSAPQ
 PAMO **LYHTGNWP**HE-----PVD-FSGQRVGVI**GTGSSG**IQVSPQIAKQAAELFVQRTPH
 CPMO **WYHTALWP**QE-----GVD-MAGKRVAII**GTGSSG**VQVAQEAALNAKQVTVFQRTPN
 CDMO **SFHTSRWD**YDYTGGDALGAPMDKLADKRVAVI**GTGATA**VQCVPELAKYCRELYVVQRTPS
 HAPMO **MFHSAQWD**HD-----VD-WSGKRVGVI**GTGASA**TQFIPQLAQTAELKVFARTTN
 * : * : : : . : * * : * * : * . : * . * : .

BVMOflava YCVPSGNGPVDQTEVDRTK-----ENFDAIWDQVRNSVVAFGFEESGVEAMS
 TMCHMO YCVPSGNGPMDPDEVARIK-----QNFDSIWDQVRSSVVAFGFEESTVEAMS
 CHMO YSVPIGNDPLSEEDVKKIK-----DNYDKIWDGVWNSALAFGLNESTVPAMS
 PAMO FAVPARNAPLDPEFLADLK-----KRYAEFREESRNTPGGTHRYQGPKSALE
 CPMO LALPMHQRLSADDNNRMR-----PMPAAFERRGKCFAGFDFDFVPKNATE
 CDMO AVDERGNHPIDEKWFAQIATPGWQKRWL----DSFTAIDWGVLTDPSELAI-----
 HAPMO WLLPTPD--LHEKISDSCKW-----LLAHVPHYSLWYRVAMAMPQSVGFLE-----
 : :

BVMOflava VSEEERRKVFQEAWDKNGFRFMFGTFCDIATN-----PEANAAAAAFIRAKIA
 TMCHMO VSESERQRVFQQAWDKNGFRFMFGTFCDIATN-----PEANAAAAAFIRSKIA
 CHMO VSAEERKAVFEKAWQTGGGFRFMFETFGDIATN-----MEANIEAQNFIKGKIA
 PAMO VSDEELVETLERYWQEGG--PDILAAAYRDILRD-----RDANERVAEFIRNKIR
 CPMO VSDTERNEILEELWNTGG-FRYWLANFQDYLFD-----DKANDYVVEFWRDKVR
 CDMO ----EHEDLVQDGWTALG--QRMRAAVGSVPIEQYSPENVQRALEEADDEQMERIRARVD
 HAPMO -----DVMVDVGYPP-TELAVSARNDRLRQDISAWME
 . : : :

BVMOflava -EIVDDPETARKLTPT----DLYAKRPLCNEG-YYETYNRDNVELVSIKENPIREITPAG
 TMCHMO -EIVKDPETARKLTPT----DLYAKRPLCNEG-YYETYNRDNVSLVKETPIEEIVPQG
 CHMO -EIVKDPPIAQLKMPQ----DLYAKRPLCDSG-YYNTFNDRDNVRLLEDVKANPIVEITENG
 PAMO -NTVRDPEVAERLVPK--GYPFGTKRLLILEID-YYEMFNDRDNVHLVDTLSAPIETITPRG
 CPMO -ARIKDPKVAEKLAPMKKPHYPYGTKRPSLEQW-YYEIFNQSNVKLVVNETPIQRISSETG
 CDMO -EIVTDPATAAQLKAW---FRQMCKRPFCHDD-YLPAFNRPNTHLVDTGGKGVVERITENG
 HAPMO PQFADRPDLREVLIPD---SPVGGKRIVRDNGTWISTLKRDNVSMI---RQPIEVITPKG
 * * . ** . : : * . : : * *

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar. The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



Appendix

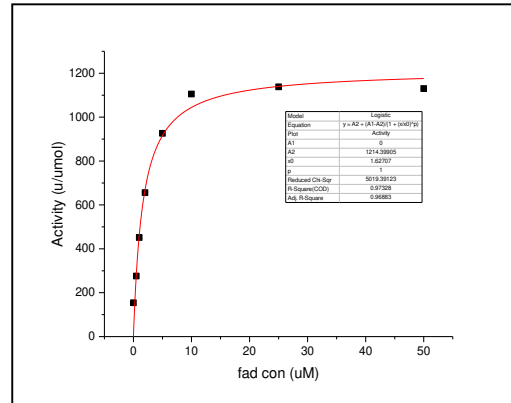
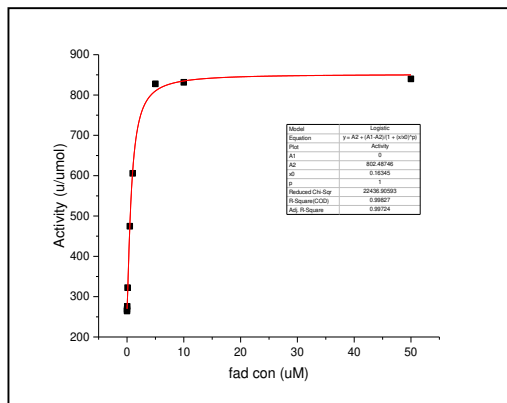
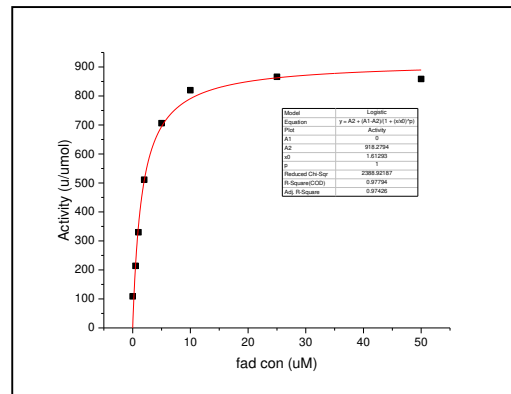
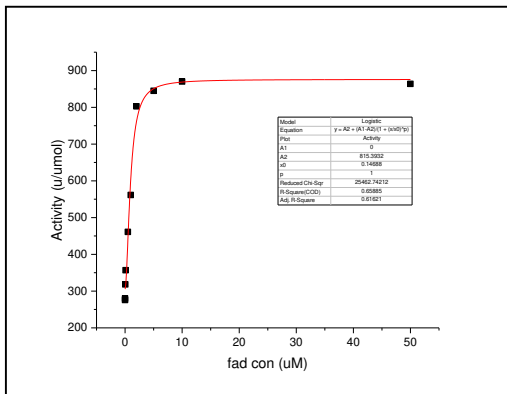
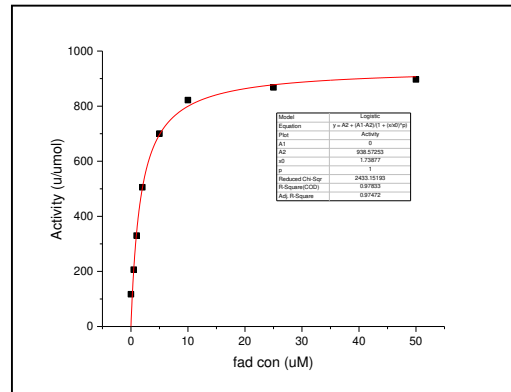
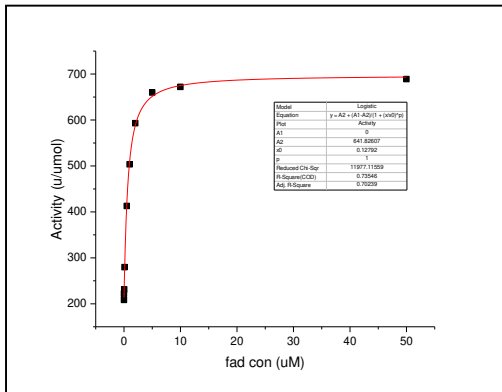
BVMOflava VRTADGIEHPLDVLVFATGFDAVDGNYR---AMDLRGRGRHISEHWT-GGPTS YLGVST
 TMCHMO VRTSDGVVHEL DVLVFATGFDAVDGNYR---AMNLRGRDGRHINEHWT-EGPTS YLGVTK
 CHMO VKLENGDFVELDMLICATGFDAVDGNYV---RMDIQKNGLAMKDYWK-EGPSSYMGVTV
 PAMO VRTSER-EYELDSLVLATGFDAL TGALF---KIDIRGVGNVALKEKWA-AGPRTYLG LST
 CPMO IVTQEG-ETEFDLIVFATGFDAVTGGLT---SIDFRNNEGQSFKDVWS-DGIRTQLGVAT
 CDMO VVVA-GVEYEVDCIVYASGF EFLGTGYTDRAGFDPTGRDGVKLSEHWA-QGTRTLHG MHT
 HAPMO ICCVDGTEHEFDLIVYGTGFHA-SKFLM---PINVTGRDGV ALHDVWKGDDARAYLGMTV
 : . * :: . : ** . : : . . : : * . : * :

BVMOflava AGFPNMFMI LGPNGPF---TNLPPSIETQVDWIGELIRHAERTGVRTVEPTAAAE EAWTA
 TMCHMO AGFPNMFMI LGPNGPF---TNLPPSIEAQVEWISDLIDKATREGLTTVEPTADAEREWTE
 CHMO NNYPNMFMI LGPNGPF---TNLPPSIESQVEWISDTIQYTVENNVESIEATKEAE EQWTQ
 PAMO AGFPNLF FFIAGPGSPS-ALSNMLVSI EQHVEWVTDHIA YMFKNGLTRSEAVLEKEDEWVE
 CPMO AGFPNLLFGYGPQSPA-GFCNGPSSAEYQGDLLIELMNHLRKN DITRIEAQPAAQEA WRK
 CDMO YGFPNLFV LQLMQGAA-LGSNIPHN FVEAARVVA AIVDHV LSTGTSSVETTKAEQAWVQ
 HAPMO PQFPNMF C MYGPNTGLVVYSTVIQFSEMTASYIVDAVRL LLEGGHQSM EVKTPVFESYNQ
 : ** : : . : : . * :

BVMOflava TC--AEIADMTLFPKAD-SWIFGANIPG--KRNAV MFYLAGLGAYRAKLRE-VADAGYT-
 TMCHMO TC--AEIANMTLFPKAD-SWIFGANIPG--KRHAVMFYLGGLGN YRRQLAD-VADGGYR-
 CHMO TC--ANIAEMTLFPKAQ-SWIFGANIPG--KKNTVYFYLGGLKEYRSALAN-CKNHAYE-
 PAMO HV--NEIADETLYPMTA-SWYTGANVPG--KPRVFMLYVGGFHRYRQICDE-VA AKGYE-
 CPMO LI--ADFDSSLFPRAK-SWYQGN NIPG--KKVESLNFPLGLPTYIAKFKE-SAEQGYA-
 CDMO LL--LDHGRPLGNPECTPGYNNNEGKPAELKDRLNVGYPAGSAAFFRMMDHWLAAGSFD-
 HAPMO RVDEGNALRAWGF SKVN-SWYKNS-----KGRVTQNF PFTAVEFWQ RTHS-VEPTDYQL
 : . . : . * : : :

BVMOflava GFELTRENAT----AAV
 TMCHMO GFQLRGERAQ----AVA
 CHMO GFDIQLQRSDIKQPANA
 PAMO GFVLT-----
 CPMO GFSL-----K
 CDMO GLTFR-----
 HAPMO G-----
 *

F I.8 Kd measurements of G14A and CHMO_{Acineto}



F II Publications resulting from this thesis

Journal articles

- i) Hamid Reza Mansouri Khosravi, Florian Rudroff, and Marko D Mihovilovic. Cloning and Characterization of a Type I Baeyer-Villiger Monooxygenase from *Amycolatopsis thermoflava*. *ChemBioChem*, 2019.
- ii) Guangyue Li, Maximilian J. L. J. Fürst, Hamid Reza Mansouri Khosravi, Anna K. Ressmann, Adriana Ilie, Florian Rudroff, Marko D. Mihovilovic, Marco W. Fraaije, and Manfred T. Reetz* Manipulating the stereoselectivity of the thermostable Baeyer–Villiger monooxygenase TmCHMO by directed evolution, *Organic & Biomolecular Chemistry*, 2017,15, 9824-9829.
- iii) Hamid Reza Mansouri Khosravi, Leticia C Goncalves, Saima Feroz, Su Ma, Andreas S Bommarius, Roland Ludwig, Florian Rudroff, and Marko D Mihovilovic Improving Operational Stability of BVMOs Employing the Consensus Approach, manuscript in preparation

Conference talks

- i) Hamid Reza Mansouri Khosravi; Leticia Christina P. Goncalves; Andreas S. Bommarius; Saima Feroz; Florian Rudroff and Marko D. Mihovilovic. Improving stability and activity of CHMO by protein engineering. Vienna young scientist symposium. 2019.

Poster Presentations

- Hamid Reza Mansouri Khosravi; Leticia Christina P. Goncalves; Andreas S. Bommarius; Saima Feroz; Florian Rudroff and Marko D. Mihovilovic. Discovery and characterization of a new Baeyer-Villiger monooxygenase using metagenome mining. 12th Conference on Protein Stabilization (ProtStab 2018), Vilnius, Lithuania. (Poster presentation)
- Hamid Reza Mansouri Khosravi; Leticia Christina P. Goncalves; Andreas S. Bommarius; Saima Feroz; Florian Rudroff and Marko D. Mihovilovic. Improving stability and activity of CHMO by protein engineering. 13th international symposium on biocatalysis and biotransformation (BioTrans 2017), Budapest, Hungary. (Poster presentation)

F III Curriculum vitae

Hamid Reza Mansouri Khosravi

Born 22.05.1989



EDUCATION&QUALIFICATION

- **2016-2020** **Doctorate of Philosophy** (Natural science), **TU Wien**, Project title (Discovery, Redesign, and stability improvement of Baeyer-Villiger Monooxygenases), under supervision of **Prof. Marko D. Mihovilovic. (Thesis submitted)**
- **2012-2013** **Master of Science** (Biotechnology), **Universiti Teknologi Malaysia** (UTM), Thesis Title (Partial purification and characterization of Uricase Isolated from a Thermophilic *Pseudomonas sp*) under supervision of **Prof. Dr. Shafinaz Shahir**
- **2007-2011** **Bachelor of Science** (Cellular and Molecular Biology-Microbiology), **Karaj Azad University**.

PUBLICATION

- **Hamid Reza Mansouri Khosravi**, Florian Rudroff, and Marko D Mihovilovic. Investigation of a novel Type I Baeyer-Villiger Monooxygenase from *Amycolatopsis thermoflava* revealed high thermodynamic but limited kinetic stability. *ChemBioChem*, 2019.
- Leticia C. P. Gonçalves, **Hamid Reza Mansouri Khosravi**, Shadi PourMehdi, Erick L. Bastos, Florian Rudroff, and Marko D. Mihovilovic. Boosting photobioredox-catalysis by morpholine electron donors under aerobic conditions, *Catalysis Science & Technology*, 2019, **9**, 2682-2688.
- Leticia C. P. Gonçalves, **Hamid Reza Mansouri Khosravi**, Erick L. Bastos, Florian Rudroff, and Marko D. Mihovilovic. Morpholine-based buffers activate aerobic

Appendix

- photobiocatalysis via spin correlated ion pair formation, *Catalysis Science & Technology*, 2019, **9**, 1365-1371.
- Guangyue Li, Maximilian J. L. J. Fürst, **Hamid Reza Mansouri Khosravi**, Anna K. Ressmann, Adriana Ilie, Florian Rudroff, Marko D. Mihovilovic, Marco W. Fraaije, and Manfred T. Reetz* Manipulating the stereoselectivity of the thermostable Baeyer–Villiger monooxygenase TmCHMO by directed evolution, *Organic & Biomolecular Chemistry*, 2017, **15**, 9824-9829.
 - Nor Sahlin Irwan Shah Lee, **Hamid Reza Mansouri Khosravi**, Norahim Ibrahim, and Shafinaz Shahir Isolation, partial purification and characterization of thermophilic uricase from *Pseudomonas otitidis* strain SN4, *Malaysian Journal of Microbiology*, 2015, **11.4**.
 - Azin fouladvand, Hamed Yari, Alireza Emami, **Hamid Reza Mansouri Khosravi**, and Shadi Pourmehdi Optimization of a Rapid DNA Extraction Protocol in Rice Focusing on Age of Plant and EDTA Concentration, *Journal of Medical and Bioengineering*, 2013 **2**, 3.
 - Hamid Sadegh, Hamed Yari, **Hamid Reza Mansouri Khosravi**, Nazanin Mohebbali, Mohammad Reza Ganjalikhany. Study of directed insilico mutations on polyphenol oxidase to improve enzyme-substrate binding efficiency- *Journal of Clinical Biochemistry*, 2011, **44**, 13.
 - 2 more papers under preparation on the topic of improving the thermostability of CHMO by protein engineering.

CONFERENCES & POSTER PRESENTATION

- Hamid Reza Mansouri Khosravi; Leticia Christina P. Goncalves; Andreas S. Bommarius; Saima Feroz; Florian Rudroff and Marko D. Mihovilovic. Improving stability and activity of CHMO by protein engineering. Vienna young scientist symposium. 2019. (**Lecture**)
- Hamid Reza Mansouri Khosravi; Leticia Christina P. Goncalves; Andreas S. Bommarius; Saima Feroz; Florian Rudroff and Marko D. Mihovilovic. Discovery and characterization of a new Baeyer-Villiger monooxygenase using metagenome mining. 12th Conference on Protein Stabilization (ProtStab 2018), Vilnius, Lithuania. (**Poster presentation**)
- Hamid Reza Mansouri Khosravi; Leticia Christina P. Goncalves; Andreas S. Bommarius; Saima Feroz; Florian Rudroff and Marko D. Mihovilovic. Improving stability and activity of CHMO by protein engineering. 13th international symposium on biocatalysis and biotransformation (BioTrans 2017), Budapest, Hungary. (**Poster presentation**)
- Hamid Reza Mansouri Khosravi, Sara Aryanfar, Nasim Saeedi. Get the anti-anxiety effect of aqueous-alcoholic extracts of *Lippa citridora* on NMRIs adult mice, 20th Iranian congress of Physiology and Pharmacology, Hamedan, Iran. Oct 2011 (**Poster presentation**)
- Hamid Reza Mansouri Khosravi, Sara Zavarmand, Nasim Saeedi. “Anti Acne and antibacterial effects of aqueous and alcoholic extract of Cumin plant,” 4th Iranian congress of Clinical Microbiology, Isfahan, Iran. November 2010 (**Presentation**)

Appendix

- Hamid Reza Mansouri Khosravi, Hamid Reza Sadegh, Alireza Emami, Hamed Yari. The anti-bacterial effect of water and alcoholic extracts of two plants and barberry shoe mountain- 3rd international conference of biology and future, May 2010 (Poster presentation)

WORK EXPERIENCES

- **Project assistant**, Institute of Applied synthetic chemistry, TU Wien, 01.2020-.
- **Chair**, Vienna Young Scientist Symposium (VSS2019), TU WIEN. 2019
- **Organization committee**, Vienna Young Scientist Symposium (VSS2018), TU WIEN. 2018
- **Research assistant**, NANOCAT (Nanotechnology and Catalysis Research Centre) the University of Malaya, project title: Design of Additive for Enhancement of Nitrile (NBR) Gloves: Antimicrobial, March-August 2015
- **Head**, Society of Microbiology, Karaj Islamic Azad University as Head of society January 2009-January 2011
- **Assistant** at Media Preparation laboratory, Karaj Islamic Azad University, Jan 2009-May 2010

WORKSHOPS & SUMMER SCHOOL

- PAcMEN summer school in protein engineering, 2018, University of Aachen, Germany
- Masterclass in applied biocatalysis, 2017, Groningen University, Netherland
- HPLC 2 Days workshop, 2014, Tehran, Iran
- Recombinant Protein Expression in E. Coli, 2014, Pasture Institute of Iran
- Spin Column Protocol for High-throughput Purification of His-tagged recombinant protein, 2013, Malaysia
- Medical Nanotechnology, 2008, Tehran, Iran

HONORS & AWARDS

- Ph.D. special scholarship, 2016-2019, European and international mobility and cooperation programs in education, science, and research (OeAD)
- Research funding scheme (international communication), 2018, Austrian research association
- Research funding scheme (international communication), 2017, Austrian research association

LANGUAGE

- Persian Mother tongue
- English Fluent
- German A2.2

SKILLS

- Microsoft office
- Bioinformatics software (Mega5, Jalview, Pymol, SPDVB viewer, etc.)
- International Computer Degree License (ICDL)
- **Laboratory Techniques**
 - Recombinant Enzyme Expression
 - Recombinant Enzyme Purification
 - Enzyme kinetics assay (Activity, stability and so on)
 - Cloning and Enzyme engineering (Directed evolution and rational design)
 - Biotransformation
 - SDS-PAGE Analysis and protein quantification methods
 - GC-MS, GC-FID, Spectrophotometry
 - Isolate and screen microorganisms and Bacterial growth curves, Media Preparation
 - RTPCR, PCR & Gel electrophoresis
 - ASTM (Antimicrobial susceptibility test method)

F IV List of abbreviations

A°	angstrom	DNA	deoxyribonucleic acid
ACN	acetontrile	dH ₂ O	deionized/distilled water
AChe	Acetylcholine esterase	dH ₂ O	deionized/distilled water
Ala	alanine	EC	enzyme commission number
Amp	ampicillin	EDTA	ethylene-diamineteraacetic acid
Arg	arginine	ee	enantiomeric excess
Asn	asparagine	EtOH	ethanol
Asp	aspartic acid	<i>E. coli</i>	<i>Escherichia coli</i>
BANA	benzoyl-d,l-arginine p-nitroanilide hydrochloride	FACS	fluorescens-activatedcellsorting
BVMO	Baeyer-Villiger monooxygenase	FAD	flavin adenine dinucleotide
BVMO _{Flava}	Baeyer-Villiger monooxygenase from <i>Amycolaptosis thermoflava</i>	FAD _{ox}	Oxidized flavin adenine dinucleotide
CAST	combinatorial active-site saturation test	FMN	flavin mononucleotide
CDMO	cyclododecanone monooxygenase	FMO	Flavin containing monooxygenases
CFE	Cell-free extract	GC	Gas chromatography
CHMO	Cyclohexanone monooxygenase	Gln	glycine
CHMO _{Acineto}	Cyclohexanone monooxygenase from <i>Acinetobacter</i> sp	GMO	Genetically modified microorganism
CPDMO	Cyclopentadecanone monooxygenase	H ₂ O ₂	hydrogen peroxide
CPMO	cyclopentanone monooxygenase	HAPMO	4-hydroxyacetophenone monooxygenase
HMM	hidden markov model	n.C	no conversion

Appendix

HSA	human serum albumin	NCBI	
IPTG	isopropyl β -D-1- thiogalactopyranoside	n.d	not detected/not determined
IREDs	Imine reductases	NMO	N-Hydroxylating monooxygenases
ISM	iterative saturation mutagenesis	PAGE	polyacrylamide gel electrophoresis
Kan	Kanamycin	PAMO	phenylacetone monooxygenase
LB medium	lysogeny broth medium	PCR	polymerase chain reaction
MBP-C-HA	maltose-binding-human albumin peptide-protein	PEG	polyethylene glycol
MCS	multiple cloning site	Phe	phenylalanine
Met	methionine	PockeMO	polycyclic ketone monooxygenase
MeOH	Methanol	ProSAR	protein sequence-activity relationships
MOPS	3-(N-morpholino)propane sulfonic acid	Ser	serine
mRNA	messenger RNA	STMO	steroid monooxygenase
MSA	multiple sequence alignment	THF	tetrahydrofuran
MS	Mass spectrometry	TmCHMO	thermostable cyclohexanone monooxygenase
NADP ⁺ (NADPH)	dinucleotide (reduced) nicotinamide adenine	Val	valine
n.a	not applicable/not available	xenB	xenobiotic reductase B
NAD ⁺ (NADH)	nicotinamide adenine		

F V References

- [1] a) A. J. J. Straathof, Adlercreutz. P., *Applied Biocatalysis*, CRC Press, **2000**; b) K. Faber, *Biotransformations in Organic Chemistry, 5th edition*, Springer, **2011**.
- [2] P. Grunwald, *Biocatalysis: Biochemical Fundamentals and Applications*, Imperial college press, **2009**.
- [3] a) T. Yoshida, *Applied Bioengineering: Innovations and Future Directions*, Wiley, **2017**; b) U. T. Bornscheuer, K. Buchholz, *Eng. Life Sci.* **2005**, *5*, 309-323.
- [4] H. Schubert, Kuznetsov, A., *Detection and Disposal of Improvised Explosives*, Springer, **2006**.
- [5] D. E. Clark, Elsevier Science Inc., **2001**, pp. 12-20.
- [6] G. Strukul, *Angew. Chem. Int. Ed.* **1998**, *37*, 1198-1209.
- [7] K. P. Bryliakov, *Chem. Rev.* **2017**, *117*, 11406-11459.
- [8] U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins, *Nature* **2012**, *485*, 185-194.
- [9] V. B. G. Poppe L., *ChemBioChem* **2018**, *19*, 284-287.
- [10] R. A. Sheldon, J. M. Woodley, *Chem. Rev.* **2018**, *118*, 801-838.
- [11] U. T. Bornscheuer, *Philos Trans A Math Phys Eng Sci* **2018**, *376*.
- [12] a) L. Tamborini, P. Fernandes, F. Paradisi, F. Molinari, *Trends Biotechnol* **2018**, *36*, 73-88; b) G. S. Paloma A. Santacoloma, Krist V. Gernaey, John M. Woodley, *Org. Process Res. Dev.* **2011**, *15*, 203-212.
- [13] P. K. Robinson, *Essays. Biochem.* **2015**, *59*, 1-41.
- [14] F. Xu, *Ind. Biochem.* **2005**, *1*, 38-50.
- [15] a) R. Bernhardt, *J. Biotechnol.* **2006**, *124*, 128-145; b) W. J. H. van Berkel, N. M. Kamerbeek, M. W. Fraaije, *J. Biotechnol.* **2006**, *124*, 670-689.
- [16] a) R. E. M. S. T. Prigge, B. A. Eipper, L. M. Amzel *Cell. Mol. Life. Sci.* **2000**, *57*, 1236-1259; b) J. D. L. Bradley J. Wallar, *Chem. Rev.* **1996**, *96*, 2625-2658.
- [17] S. Fetzner, *Appl. Microbiol. Biotechnol.* **2002**, *60*, 243-257.
- [18] D. E. Torres Pazmiño, University of Groningen (Netherland), **2008**.
- [19] H. E. Conrad, R. Dubus, M. J. Namtvedt, I. C. Gunsalus, *J. Biol. Chem.* **1965**, *240*, 495-503.
- [20] F. W. F. a. A. J. Markovetz, *Biochem. Biophys. Res. Commun.* **1969**, *37*.
- [21] M. L. Mascotti, W. J. Lapadula, M. Juri Ayub, *PLoS One* **2015**, *10*, e0132689.
- [22] E. Malito, Alfieri, A., Fraaije, M. W., Mattevi, A., *PNAS* **2004**, *101*, 13157-13162.
- [23] a) D. E. Torres Pazmiño, Baas, B., Janssen, D. B., Fraaije, M. W., *Biochemistry* **2008**, *47*, 4082-4093; b) R. Orru, H. M. Dudek, C. Martinoli, D. E. T. Pazmino, A. Royant, M. Weik, M. W. Fraaije, A. Mattevi, *J. Biol. Chem.* **2011**, *286*, 29284-29291.
- [24] A. Willetts, *Trends. Biotechnol.* **1997**, *15*, 55-62.
- [25] D. B. J. Nanne M. Kamerbeek, Willem J. H. van Berkel, Marco W. Fraaije, *Adv. Synth. Catal.* **2003**, *345*, 667-687.
- [26] B. M. Marko D. Mihovilovic, Peter Stanety *Eur. J. Org. Chem.* **2002**, 3711-3730.
- [27] V. Alphand, G. Carrea, R. Wohlgemuth, R. Furstoss, J. M. Woodley, *Trends. Biotechnol.* **2003**, *21*, 318-323.
- [28] a) K. Balke, B. Marcus, U. T. Bornscheuer, *ChemBioChem* **2017**, 1627-1638; b) H. Leisch, K. Morley, P. C. K. Lau, *Chem. Soc. Rev.* **2011**, 4165-4222.
- [29] N. M. K. Marco W. Fraaije, Willem J.H. van Berkel, Dick B. Janssen, *FEBS Lett.* **2002**, *518*, 43-47.
- [30] a) H. Iwaki, S. Wang, S. Grosse, H. Bergeron, A. Nagahashi, J. Lertvorachon, J. Yang, Y. Konishi, Y. Hasegawa, P. C. Lau, *Appl. Environ. Microbiol.* **2006**, *72*, 2707-2720; b) C. Tolmie, M. S. Smit, D. J. Opperman, *Toxins* **2018**, *10*.
- [31] Y. C. J. Chen, O. P. Peoples, C. T. Walsh, *J. Bacteriol.* **1988**, *170*, 781-789.
- [32] a) D. B. N. Nuala A. Donoghue, Peter W. Trudgill, *Eur. J. Biochem.* **1976**, *63*, 175-192; b) H. Iwaki, Y. Hasegawa, S. Wang, M. M. Kayser, P. C. Lau, *Appl. Environ. Microbiol.* **2002**, *68*, 5671-5684.
- [33] S. S. Shingo Morii, Yuhji Yamauchi, Masahiko Miyamoto, Masafumi Iwami, Eiji Itagaki, *J. Biochem.* **1999**, *126*, 624-631.
- [34] K. L. G. Patricia C. Brzostowicz, Stuart M. Thomas, Mary Sue Blasko, and Pierre E. Rouvière, *J. Bacteriol.* **2000**, *182*, 4241-4248.

Appendix

- [35] N. M. Kamerbeek, M. J. H. Moonen, J. G. M. van der Ven, W. J. H. van Berkel, M. W. Fraaije, D. B. Janssen, *Eur. J. Biochem.* **2001**, *268*, 2547-2557.
- [36] K. Kostichka, S. M. Thomas, K. J. Gibson, V. Nagarajan, Q. Cheng, *J. Bacteriol.* **2001**, *183*, 6478-6486.
- [37] a) J. Rehdorf, A. Kirschner, U. T. Bornscheuer, *Biotechnol. Lett.* **2007**, *29*, 1393-1398; b) C. W. K. E. Nelson, I. T. Paulsen, R. J. Dodson, H. Hilbert, V. A. P. Martins dos Santos, D. E. Fouts, S. R. Gill, M. Pop, M. Holmes, L. Brinkac, M. Beanan, R. T. DeBoy, S. Daugherty, J. Kolonay, R. Madupu, W. Nelson, O. White, J. Peterson, H. Khouri, I. Hance, P. Chris Lee, E. Holtzapple, D. Scanlan, K. Tran, A. Moazzez, T. Utterback, M. Rizzo, K. Lee, D. Kosack, D. Moestl, H. Wedler, J. Lauber, D. Stjepandic, J. Hoheisel, M. Straetz, S. Heim, C. Kiewitz, J. Eisen, K. N. Timmis, A. Düsterhöft, B. Tümmeler, C. M. Fraser, *Environ. Microbiol.* **2002**, *4*, 799-808.
- [38] F. M. Jan B. Van Beilen, Markus A. Seeger, Jasminka Kovac, Zhi Li, Theo H. M. Smits, Urban Fritsche, Bernard Witholt, *Environ. Microbiol.* **2003**, *5*, 174-182.
- [39] a) S. M. T. Qiong Cheng, Kristy Kostichka, James R. Valentine, Vasantha Nagarajan, *J. Bacteriol.* **2000**, *182*, 4744-4751; b) P. C. Brzostowicz, D. M. Walters, S. M. Thomas, V. Nagarajan, P. E. Rouviere, *Appl. Environ. Microbiol.* **2003**, *69*, 334-342.
- [40] M. Bramucci, P. Brzostowicz, K. Kostichka, V. Nagarajan, P. Rouviere, S. Thomas, *Chemical Abstracts, Vol. 138*, **2003**, p. 233997.
- [41] M. W. Fraaije, N. M. Kamerbeek, A. J. Heidekamp, R. Fortin, D. B. Janssen, *J. Biol. Chem.* **2004**, *279*, 3354-3360.
- [42] M. W. Fraaije, J. Wu, D. P. Heuts, E. W. van Hellemond, J. H. Spelberg, D. B. Janssen, *Appl. Microbiol. Biotechnol.* **2005**, *66*, 393-400.
- [43] T. Kotani, H. Yurimoto, N. Kato, Y. Sakai, *J. Bacteriol.* **2007**, *189*, 886-893.
- [44] A. Kirschner, J. Altenbuchner, U. T. Bornscheuer, *Appl. Microbiol. Biotechnol.* **2007**, *73*, 1065-1072.
- [45] a) C. Onaca, M. Kieninger, K. H. Engesser, J. Altenbuchner, *J. Bacteriol.* **2007**, *189*, 3759-3767; b) A. Volker, A. Kirschner, U. T. Bornscheuer, J. Altenbuchner, *Appl. Microbiol. Biotechnol.* **2008**, *77*, 1251-1260.
- [46] J. Rehdorf, C. L. Zimmer, U. T. Bornscheuer, *Appl. Environ. Microbiol.* **2009**, *75*, 3106-3114.
- [47] J. Jiang, C. N. Tetzlaff, S. Takamatsu, M. Iwatsuki, M. Komatsu, H. Ikeda, D. E. Cane, *Biochemistry* **2009**, *48*, 6431-6440.
- [48] B. J. Y. I. Ahmad Mirza, Shaozhao Wang, Stephan Grosse, Hélène Bergeron, Akihiro Imura, Hiroaki Iwaki, Yoshie Hasegawa, Peter C. K. Lau, Albert M. Berghuis, *J. Am. Chem. Soc.* **2009**, *131*, 8848-8854.
- [49] M. J. Seo, D. Zhu, S. Endo, H. Ikeda, D. E. Cane, *Biochemistry* **2011**, *50*, 1739-1754.
- [50] H. Leisch, R. Shi, S. Grosse, K. Morley, H. Bergeron, M. Cygler, H. Iwaki, Y. Hasegawa, P. C. Lau, *Appl. Environ. Microbiol.* **2012**, *78*, 2200-2212.
- [51] M. Weiss, K. Denger, T. Huhn, D. Schleheck, *Appl. Environ. Microbiol.* **2012**, *78*, 8254-8263.
- [52] R. W. Friedemann Leipold, Uwe T. Bornscheuer, *Appl. Microbiol. Biotechnol.* **2012**, *94*, 705-717.
- [53] S. G. Hiroaki Iwaki, Hélène Bergeron, Hannes Leisch, Krista Morley, Yoshie Hasegawa, Peter. K. Lau, *Appl. Environ. Microbiol.* **2013**, *79*, 3282-3293.
- [54] Q. X. William C. Hwang, Bainan Wu, Adam Godzik, *Proteins* **2014**, *86*, 269-269.
- [55] F. M. Ferroni, M. S. Smit, D. J. Opperman, *J. Mol. Catal. B-Enzym.* **2014**, *107*, 47-54.
- [56] E. Romero, J. Ø. Rub, A. Mattevi, M. W. Fraaije, *Angew. Chem. Int. Ed.* **2016**, *55*, 15852-15855.
- [57] G. de Gonzalo, M. Fürst, M. Fraaije, *Catalysts* **2017**, *7*, 288.
- [58] R. D. Ceccoli, D. A. Bianchi, M. J. Fink, M. D. Mihovilovic, D. V. Rial, *AMB Express* **2017**, *7*, 87.
- [59] F. Fiorentini, E. Romero, M. W. Fraaije, K. Faber, M. Hall, A. Mattevi, *ACS. Chem. Biol.* **2017**, *12*, 2379-2387.
- [60] A. Gran-Scheuch, M. Trajkovic, L. Parra, M. W. Fraaije, *Front. Microbiol.* **2018**, *9*, 1609.
- [61] F. L. Yan Zhang, Na Xu, Yin-Qi Wu, Yu-Cong Zheng, Qian Zhao, Guoqiang Lin, Hui-Lei Yu, Jian-He Xu, *Appl. Environ. Microbiol.* **2018**, *84*, 00638-00618.
- [62] Y. Y. Liu, C. X. Li, J. H. Xu, G. W. Zheng, *Appl. Environ. Microbiol.* **2019**, *85*.
- [63] H. R. Mansouri, M. D. Mihovilovic, F. Rudroff, *ChemBioChem* **2019**.
- [64] E. Beneventi, M. Niero, R. Motterle, M. Fraaije, E. Bergantino, *J. Mol. Catal.* **2013**, *98*, 145-154.
- [65] Y. Zhang, Y.-Q. Wu, N. Xu, Q. Zhao, H.-L. Yu, J.-H. Xu, *ACS. Sustain. Chem. Eng.* **2019**, *7*, 7218-7226.
- [66] N. Doukyu, H. Ogino, *Biochem. Eng.* **2010**, *48*, 270-282.
- [67] F. Secundo, S. Fiala, M. W. Fraaije, G. de Gonzalo, M. Meli, F. Zambianchi, G. Ottolina, *Biotechnol. Bioeng.* **2011**, *108*, 491-499.

Appendix

- [68] M. J. Fürst, S. Savino, H. M. Dudek, J. R. Gomez Castellanos, C. Gutierrez de Souza, S. Roviada, M. W. Fraaije, A. Mattevi, *J. Am. Chem. Soc.* **2017**, *139*, 627-630.
- [69] N. M. Kamerbeek, Van der Ploeg, R., Olsthoorn, A. J. J., Tahallah, N., Heck, A. J. R., Malito, E., Janssen, D. B., Fraaije, M. W., *Exploring the role of the N-terminus of 4-hydroxyacetophenone monooxygenase*, **2005**.
- [70] M. J. Fürst, F. Fiorentini, M. W. Fraaije, *Curr Opin Struct Biol* **2019**, *59*, 29-37.
- [71] M. J. L. J. Fürst, A. Gran-Scheuch, F. Aalbers, M. W. Fraaije, *ACS. Catal.* **2019**.
- [72] F. M. Ferroni, C. Tolmie, M. S. Smit, D. J. Opperman, *PLoS One* **2016**, *11*, e0160186.
- [73] a) B. J. Yachnin, T. Sprules, M. B. McEvoy, P. C. K. Lau, A. M. Berghuis, *J. Am. Chem. Soc.* **2012**, *134*, 7788-7795; b) B. J. Yachnin, M. B. McEvoy, R. J. D. MacCuish, K. L. Morley, P. C. K. Lau, A. M. Berghuis, *ACS Chem. Biol.* **2014**, *9*, 2843-2851.
- [74] H. L. Messiha, S. T. Ahmed, V. Karuppiah, R. Suardiaz, G. A. Ascue Avalos, N. Fey, S. Yeates, H. S. Toogood, A. J. Mulholland, N. S. Scrutton, *Biochemistry* **2018**, *57*, 1997-2008.
- [75] C. Martinoli, H. M. Dudek, R. Orru, D. E. Edmondson, M. W. Fraaije, A. Mattevi, *ACS. Catal.* **2013**, *3*, 3058-3062.
- [76] T. D. Nguyen, G.-E. Choi, D.-H. Gu, P.-W. Seo, J.-W. Kim, J.-B. Park, J.-S. Kim, *Biochem. Biophys. Res. Commun.* **2019**, *512*, 564-570.
- [77] S. Franceschini, H. L. van Beek, A. Pennetta, C. Martinoli, M. W. Fraaije, A. Mattevi, *J. Biol. Chem.* **2012**, *287*, 22626-22634.
- [78] Y. K. Bong, Clay, M. D., Collier, S. J., Mijts, B., Vogel, M., Zhang, X., et al, *Vol. 8895271*, US, **2014**.
- [79] O. Abril, C. C. Ryerson, C. Walsh, G. M. Whitesides, *Bioorg. Chem.* **1989**, *17*, 41-52.
- [80] D. J. Opperman, M. T. Reetz, *ChemBioChem* **2010**, *11*, 2589-2596.
- [81] a) G. de Gonzalo, M. D. Mihovilovic, M. W. Fraaije, *ChemBioChem* **2010**, *11*, 2208-2231; b) N. G. Stefano Colonna, Giacomo Carrea, Gianluca Ottolina, Piero Pastab, Francesca Zambianchib, *Tetrahedron Lett.* **2002**, *43*, 1797-1799.
- [82] M. Bucko, P. Gemeiner, A. Schenkmyerova, T. Krajcovic, F. Rudroff, M. D. Mihovilovic, *Appl. Microbiol. Biotechnol.* **2016**, *100*, 6585-6599.
- [83] F. Zambianchi, P. Pasta, G. Carrea, S. Colonna, N. Gaggero, J. M. Woodley, *Biotechnol. Bioeng.* **2002**, *78*, 489-496.
- [84] a) D. P. B. Carol Cummings Ryerson, Christopher Walsh, *Biochemistry* **1982**, *21*, 2644-2655; b) D. S. D. P. B. V. Massey, *Biochemistry* **2001**, *40*, 11156-11167.
- [85] a) M. A. Longo, D. Combes, *Analysis of the thermal deactivation kinetics of α -chymotrypsin modified by chemoenzymatic glycosylation*, Vol. 15, Elsevier Science Inc., **1998**; b) V. G. Eijsink, S. Gaseidnes, T. V. Borchert, B. van den Burg, *Biomol. Eng.* **2005**, *22*, 21-30.
- [86] M. Fürst, M. Boonstra, S. Bandstra, M. W. Fraaije, *Biotechnol. Bioeng.* **2019**, *116*, 2167-2177.
- [87] A. S. Bommarius, M. F. Paye, *Chem. Soc. Rev.* **2013**, *42*, 6534-6565.
- [88] K. M. Polizzi, A. S. Bommarius, J. M. Broering, J. F. Chaparro-Riggers, *Curr. Opin. Chem. Biol.* **2007**, *11*, 220-225.
- [89] V. Brissos, N. Goncalves, E. P. Melo, L. O. Martins, *PLoS One* **2014**, *9*, e87209.
- [90] N. Palackal, Y. Brennan, W. N. Callen, P. Dupree, G. Frey, F. Goubet, G. P. Hazlewood, S. Healey, Y. E. Kang, K. A. Kretz, E. Lee, X. Tan, G. L. Tomlinson, J. Verruto, V. W. Wong, E. J. Mathur, J. M. Short, D. E. Robertson, B. A. Steer, *Protein Sci.* **2004**, *13*, 494-503.
- [91] a) L. J. Wang, X. D. Kong, H. Y. Zhang, X. P. Wang, J. Zhang, *Biochem. Biophys. Res. Commun.* **2000**, *276*, 346-349; b) S. D'Amico, J. C. Marx, C. Gerday, G. Feller, *J Biol Chem* **2003**, *278*, 7891-7896; c) G. V. B. Van der burg, Oene R. Veltman, G. Venema, Vincent G. H. Eijsink, *PNAS* **1998**, *95*, 2056-2060.
- [92] a) A. Sassolas, L. J. Blum, B. D. Leca-Bouvier, *Biotechnol. Adv.* **2012**, *30*, 489-511; b) E. K. Katzir, *Trends. Biotechnol.* **1993**, *11*, 471-478.
- [93] Y. Zhang, J. Ge, Z. Liu, *ACS. Catal.* **2015**, *5*, 4503-4513.
- [94] A. K. S. Nisha S, Gobi N, *Chem. Sci. Rev. Lett.* **2012**, *1*, 148-155.
- [95] a) A. A. Homaei, R. Sariri, F. Vianello, R. Stevanato, *J. Chem. Biol.* **2013**, *6*, 185-205; b) I. Esmail, J. D. Vieira, A. C. Amaral, *Appl. Microbiol. Biotechnol.* **2015**, *99*, 2065-2082.
- [96] A. Basso, S. Serban, *Mol. Catal.* **2019**, *479*, 110607.
- [97] S. Cynthia, D. M. Shelley, *Recent Pat. Eng.* **2008**, *2*, 195-200.
- [98] S. F. D'Souza, *Curr. Sci.* **1999**, *77*, 69-79.
- [99] D. H. Lee, C. H. Park, J. M. Yeo, S. W. Kim, *J. Ind. Eng. Chem.* **2006**, *12*, 777-782.
- [100] S. Datta, L. R. Christena, Y. R. S. Rajaram, *Biotech.* **2013**, *3*, 1-9.

Appendix

- [101] Z.-G. Wang, L.-S. Wan, Z.-M. Liu, X.-J. Huang, Z.-K. Xu, *J. Mol. Catal.* **2009**, *56*, 189-195.
- [102] R. A. Sheldon, *Biochem. Soc. Trans.* **2007**, *35*, 1583-1587.
- [103] a) M. K. Hoang Hiep Nguyen, *Appl. Sci. Conver. Technol.* **2017**, *26*, 157-163; b) B. S. Chang, Mahoney, R., *Biotechnol. Appl. Biochem.* **1995**, *22*, 203-214.
- [104] M. A. F. Delgove, D. Valencia, J. Solé, K. V. Bernaerts, S. M. A. De Wildeman, M. Guillén, G. Álvaro, *Appl. Cata. A-Gen.* **2019**, *572*, 134-141.
- [105] H. L. van Beek, H. J. Wijma, L. Fromont, D. B. Janssen, M. W. Fraaije, *FEBS. Open. Bio.* **2014**, *4*, 168-174.
- [106] A. L. Serdakowski, J. S. Dordick, *Trends Biotechnol.* **2008**, *26*, 48-54.
- [107] J. F. Carpenter, M. J. Pikal, B. S. Chang, T. W. Randolph, *Pharm. Res.* **1997**, *14*, 969.
- [108] H. L. van Beek, N. Beyer, D. B. Janssen, M. W. Fraaije, *J. Biotechnol.* **2015**, *203*, 41-44.
- [109] S. D. Allison, B. Chang, T. W. Randolph, J. F. Carpenter, *Arch. Biochem. Biophys.* **1999**, *365*, 289-298.
- [110] L. C. P. Goncalves, D. Kracher, S. Milker, M. J. Fink, F. Rudroff, R. Ludwig, A. S. Bommarius, *Adv. Synth. Catal.* **2017**, *359*, 2121-2131.
- [111] L. C. P. Gonçalves, H. R. Mansouri, S. PourMehdi, M. Abdellah, B. S. Fadiga, E. L. Bastos, J. Sá, M. D. Mihovilovic, F. Rudroff, *Catal. Sci. Technol.* **2019**, *9*, 2682-2688.
- [112] R. Cavicchioli, P. M. G. Curmi, K. S. Siddiqui, T. Thomas, *Method. Microbiol.* **2006**, *35*, 395-436.
- [113] a) A. J. A. Saleem A. Bokhari, M. Hamid Rashid, M. Ibrahim Rajoka, Khawar S. Siddiqui, *Biotechnol. Prog.* **2002**, *18*, 276-281; b) S. A. A. Siddiqui K. S., Rashid M. H., Rajoka M. I., *Enzyme. Microb. Technol.* **2000**, *27*, 467-474; c) P. A. Siddiqui K. S., Cavicchioli R., *Cell. Mol. Biol.* **2004**, *50*, 657-667.
- [114] a) R. L. Lundblad, *Techniques in Protein Modification*, CRC Press Inc., **1995**; b) T. Imoto, Yamada, H., *Chemical modification. In Protein Structure: A Practical Approach*, Oxford University Press, **1989**.
- [115] a) V. G. Eijsink, A. Bjork, S. Gaseidnes, R. Sirevag, B. Synstad, B. van den Burg, G. Vriend, *J. Biotechnol.* **2004**, *113*, 105-120; b) V. A. Š. Vadim V. Mozhaev, Nikolay S. Melik-Nubarov, Nida Z. Galkaintaite, Gervydas J. Denis, Eugenius P. Butkus, Boris Yu. Zavlavsky, Nataliya M. Mestechkina, Karel Martinek, *Eur. J. Biochem.* **1988**, *171*, 147-154.
- [116] J. B. J. Grace DeSantis, *Curr. Opin. Biotechnol.* **1999**, *10*, 324-330.
- [117] R. Villalonga, S. Tachibana, R. Cao, H. L. Ramirez, Y. Asano, *Biochem. Eng.* **2006**, *30*, 26-32.
- [118] K. Sangeetha, T. E. Abraham, *J. Mol. Catal.* **2006**, *38*, 171-177.
- [119] N. Toda, S. Asano, C. F. Barbas, 3rd, *Angew. Chem. Int. Ed.* **2013**, *52*, 12592-12596.
- [120] J. T. Patterson, S. Asano, X. Li, C. Rader, C. F. Barbas, 3rd, *Bioconjug. Chem.* **2014**, *25*, 1402-1407.
- [121] S. Brocchini, A. Godwin, S. Balan, J. W. Choi, M. Zloh, S. Shaunak, *Adv. Drug. Deliv. Rev.* **2008**, *60*, 3-12.
- [122] E. H. Siar, R. Morellon-Sterling, M. N. Zidoune, R. Fernandez-Lafuente, *Int. J. Biol. Macromol.* **2019**, *133*, 412-419.
- [123] a) A. B. M. E. Guazzaroni, J. M. Vieites, Y. Al-ramahi, N. L. Cortés, A. Ghazi, P. N. Golyshin, M. Ferrer, *Handbook of Hydrocarbon and Lipid Microbiology. Springer, Berlin, Heidelberg* **2010**, 2911-2927; b) J. W. E. Jeffries, N. Dawson, C. Orengo, T. S. Moody, D. J. Quinn, H. C. Hailes, J. M. Ward, *ChemistrySelect* **2016**, *1*, 2217-2220.
- [124] B. Temperton, S. J. Giovannoni, *Curr. Opin. Microbiol.* **2012**, *15*, 605-612.
- [125] K. D. Park J., Kim S., Kim J., Bae K., Lee C., *J. Microbiol. Biotechnol.* **2007**, *17*, 1083-1089.
- [126] M. W. Fraaije, J. Wu, D. P. H. M. Heuts, E. W. van Hellemond, J. H. L. Spelberg, D. B. Janssen, *Appl. Microbiol. Biotechnol.* **2005**, *66*, 393-400.
- [127] B. U. T. Kazlauskas R. J, *Nat. Chem. Biol.* **2009**, *5*, 526-529.
- [128] J. K. Dhanjal, V. Malik, N. Radhakrishnan, M. Sigar, A. Kumari, D. Sundar, *Computational Protein Engineering Approaches for Effective Design of New Molecules*, **2019**.
- [129] U. Bornscheuer, R. J. Kazlauskas, *Curr Protoc Protein Sci.* **2011**, Chapter 26, Unit 26 27.
- [130] a) V. Stepankova, S. Bidmanova, T. Koudelakova, Z. Prokop, R. Chaloupkova, J. Damborsky, *ACS. Catal.* **2013**, *3*, 2823-2836; b) Q. Liu, G. Xun, Y. Feng, *Biotechnol. Adv.* **2019**, *37*, 530-537.
- [131] D. S. Wilson, A. D. Keefe, *Curr. Protoc. Mol. Biol.* **2000**, *51*, 8.3.1-8.3.9.
- [132] H. Xiao, Z. Bao, H. Zhao, *Ind. Eng. Chem. Res.* **2015**, *54*, 4011-4020.
- [133] M. Dorr, M. P. Fibinger, D. Last, S. Schmidt, J. Santos-Aberturas, D. Bottcher, A. Hummel, C. Vickers, M. Voss, U. T. Bornscheuer, *Biotechnol. Bioeng.* **2016**, *113*, 1421-1432.
- [134] a) S. Lutz, *Curr. Opin. Biotechnol.* **2010**, *21*, 734-743; b) H. J. Wijma, R. J. Floor, D. B. Janssen, *Curr Opin Struct Biol* **2013**, *23*, 588-594; c) W. M. Patrick, A. E. Firth, *Biomol. Eng.* **2005**, *22*, 105-112.

Appendix

- [135] M. B. Manfred T. Reetz, José Daniel Carballeira, Dongxing Zha, Andreas Vogel, *Angew. Chem. Int. Ed.* **2005**, *44*, 4192-4196.
- [136] P. S. Reetz M. T, Carballeira J. D, Gumulya Y, Bocola M., *J. Am. Chem. Soc* **2010**, *132*, 9144-9152.
- [137] R. J. Fox, S. C. Davis, E. C. Mundorff, L. M. Newman, V. Gavrilovic, S. K. Ma, L. M. Chung, C. Ching, S. Tam, S. Muley, J. Grate, J. Gruber, J. C. Whitman, R. A. Sheldon, G. W. Huisman, *Nat. Biotechnol.* **2007**, *25*, 338-344.
- [138] R. K. Kuipers, H. J. Joosten, W. J. van Berkel, N. G. Leferink, E. Rooijen, E. Ittmann, F. van Zimmeren, H. Jochens, U. Bornscheuer, G. Vriend, V. A. dos Santos, P. J. Schaap, *Proteins* **2010**, *78*, 2101-2113.
- [139] B. S. Boris Steipe, Andreas Plückthun, Stefan Steinbacher, *J. Mol. Biol.* **1994**, *240*, 188-192.
- [140] B. T. Porebski, A. M. Buckle, *Protein. Eng. Des. Sel.* **2016**, *29*, 245-251.
- [141] a) Z. E. Pauling L., Henriksen T., Lövsstad R. , *Acta. Chem. Scand.* **1963**, *17*, 9–16; b) S. K. Z Yang, M Nei, *Genetics* **1995**, *141* 1641-1650; c) E. N. Joseph W. Thornton, David Crews, *Science* **2003**, *301*, 1714-1717; d) J. W. Thornton, *Nat. Rev. Genet.* **2004**, *5*, 366-375.
- [142] a) C. UniProt, *Nucleic Acids Res.* **2008**, *36*, D190-195; b) N. R. Coordinators, *Nucleic Acids Res.* **2014**, *42*, D7-17.
- [143] a) W. M. Lehmann M., *Curr. Opin. Biotechnol.* **2001**, *12*, 371-375; b) C. L. Martin Lehmann, Anke Middendorf, Dominik Studer, Søren F. Lassen, Luis Pasamontes, Adolphus P.G.M. van Loon, Markus Wyss, *Protein Eng. Des. Sel.* **2002**, *15*, 403–411; c) T. J. Magliery, J. J. Lavinder, B. J. Sullivan, *Curr. Opin. Chem. Biol.* **2011**, *15*, 443-451; d) A. Paatero, K. Rosti, A. V. Shkumatov, C. Sele, C. Brunello, K. Kysenius, P. Singha, V. Jokinen, H. Huttunen, T. Kajander, *Biochemistry* **2016**, *55*, 914-926.
- [144] a) K. M. Polizzi, J. F. Chaparro-Riggers, E. Vazquez-Figueroa, A. S. Bommarius, *Biotechnol. J.* **2006**, *1*, 531-536; b) O. Khersonsky, G. Kiss, D. Rothlisberger, O. Dym, S. Albeck, K. N. Houk, D. Baker, D. S. Tawfik, *Proc. Natl. Acad. Sci.* **2012**, *109*, 10358-10363.
- [145] B. J. Jones, H. Y. Lim, J. Huang, R. J. Kazlauskas, *Biochemistry* **2017**, *56*, 6521-6532.
- [146] S. Schmidt, M. Genz, K. Balke, U. T. Bornscheuer, *J. Biotechnol.* **2015**, *214*, 199-211.
- [147] J. Engel, K. S. Mthethwa, D. J. Opperman, S. Kara, *Mol. Catal.* **2019**, *468*, 44-51.
- [148] C. Martin, A. Ovalle Maqueo, H. J. Wijma, M. W. Fraaije, *Biotechnol. Biofuels.* **2018**, *11*, 56.
- [149] K. Balke, A. Beier, U. T. Bornscheuer, *Biotechnol. Adv.* **2018**, *36*, 247-263.
- [150] L. Fernandez-Arrojo, M. E. Guazzaroni, N. Lopez-Cortes, A. Beloqui, M. Ferrer, *Curr. Opin. Biotechnol.* **2010**, *21*, 725-733.
- [151] K. N. Lam, J. Cheng, K. Engel, J. D. Neufeld, T. C. Charles, *Front. Microbiol.* **2015**, *6*, 1196.
- [152] G. Sandoval, *Lipases and Phospholipases : Methods and Protocols*, Humana Press Inc., **2012**.
- [153] L. K. Lorenz P., Niehaus F., Eck J., *Curr. Opin. Biotechnol.* **2002** *13*, 572-577.
- [154] M. Ferrer, F. Martinez-Abarca, P. N. Golyshin, *Curr. Opin. Biotechnol.* **2005**, *16*, 588-593.
- [155] a) I. Nobeli, A. D. Favia, J. M. Thornton, *Nat Biotechnol* **2009**, *27*, 157-167; b) U. T. Bornscheuer, R. J. Kazlauskas, *Angew. Chem. Int. Ed.* **2004**, *43*, 6032-6040; c) A. Babbie, N. Tokuriki, F. Hollfelder, *Curr. Opin. Chem. Biol.* **2010**, *14*, 200-207.
- [156] G. A. Behrens, A. Hummel, S. K. Padhi, S. Schätzle, U. T. Bornscheuer, *Adv. Synth. Catal.* **2011**, *353*, 2191-2215.
- [157] A. Chang, I. Schomburg, S. Placzek, L. Jeske, M. Ulbrich, M. Xiao, C. W. Sensen, D. Schomburg, *Nucleic Acids Res.* **2015**, *43*, D439-446.
- [158] A. M. Schnoes, S. D. Brown, I. Dodevski, P. C. Babbitt, *PLoS Comput. Biol.* **2009**, *5*, e1000605.
- [159] a) L. J. Basile, R. C. Willson, B. T. Sewell, M. J. Benedik, *Appl. Microbiol. Biotechnol.* **2008**, *80*, 427-435; b) H. Deng, L. Ma, N. Bandaranayaka, Z. Qin, G. Mann, K. Kyeremeh, Y. Yu, T. Shepherd, J. H. Naismith, D. O'Hagan, *ChemBioChem* **2014**, *15*, 364-368; c) H. Suzuki, F. Okazaki, A. Kondo, K. Yoshida, *Appl. Microbiol. Biotechnol.* **2013**, *97*, 2929-2938; d) T. Furuya, K. Kino, *ChemSusChem* **2009**, *2*, 645-649.
- [160] D. Wetzl, M. Berrera, N. Sandon, D. Fishlock, M. Ebeling, M. Muller, S. Hanlon, B. Wirz, H. Iding, *ChemBioChem* **2015**, *16*, 1749-1756.
- [161] M. Schrewe, M. K. Julsing, B. Buhler, A. Schmid, *Chem. Soc. Rev.* **2013**, *42*, 6346-6377.
- [162] S. Fademrecht, P. N. Scheller, B. M. Nestl, B. Hauer, J. Pleiss, *Proteins* **2016**, *84*, 600-610.
- [163] X. Pei, L. Yang, G. Xu, Q. Wang, J. Wu, *J. Mol. Catal.* **2014**, *99*, 26-33.
- [164] J. Maimanacos, J. Chow, S. K. Gasmeyer, S. Gullert, F. Busch, R. Kourist, W. R. Streit, *Front. Microbiol.* **2016**, *7*, 1332.
- [165] M. K. Akhtar, N. J. Turner, P. R. Jones, *Proc. Natl. Acad. Sci. U S A* **2013**, *110*, 87-92.

Appendix

- [166] J. S. Chun, S. B. Kim, Y. K. Oh, C. N. Seong, D. H. Lee, K. S. Bae, K. J. Lee, S. O. Kang, Y. C. Hah, M. Goodfellow, *Int. J. Syst. Bacteriol.* **1999**, *49*, 1369-1373.
- [167] R. S. Kantor, A. W. van Zyl, R. P. van Hille, B. C. Thomas, S. T. L. Harrison, J. F. Banfield, *Environ. Microbiol.* **2015**, *17*, 4929-4941.
- [168] E. Romero, J. R. G. Castellanos, A. Mattevi, M. W. Fraaije, *Angew. Chem. Int. Ed.* **2016**, *55*, 15852-15855.
- [169] C. Szolkowy, L. D. Eltis, N. C. Bruce, G. Grogan, *ChemBioChem* **2009**, *10*, 1208-1217.
- [170] R. Eisenthal, M. J. Danson, D. W. Hough, *Trends Biotechnol.* **2007**, *25*, 247-249.
- [171] J. Lebreton, V. Alphand, R. Furstoss, *Tetrahedron Lett.* **1996**, *37*, 1011-1014.
- [172] M. D. Mihovilovic, F. Rudroff, B. Grötzl, P. Kapitan, R. Snajdrova, J. Rydz, *Angew. Chem. Int. Ed.* **2005**, *44*, 3609-3613.
- [173] D. Black, Q.-z. Chen, *Tetrahedron-Asymmetry* **1993**, *4*, 1387-1390.
- [174] M. J. Fink, D. V. Rial, P. Kapitanova, A. Lengar, J. Rehdorf, Q. Cheng, F. Rudroff, M. D. Mihovilovic, *Adv. Synth. Catal.* **2012**, 3491-3500.
- [175] M. D. Mihovilovic, F. Rudroff, A. Winninger, T. Schneider, F. Schulz, M. T. Reetz, *Org. Lett.* **2006**, *8*, 1221-1224.
- [176] D. V. Rial, P. Cernuchova, J. B. V. Beilen, M. D. Mihovilovic, *J. Mol. Catal. B-Enzym.* **2008**, *50*, 61-68.
- [177] a) M. T. Reetz, *PNAS* **2004**, *101*, 5716-5722; b) R. Noyori, *Angew. Chem. Int. Ed.* **2002**, *41*, 2008-2022; c) W. S. Knowles, *Angew. Chem. Int. Ed.* **2002**, *41*, 1998-2007.
- [178] S. M. Miller, *Enzyme Catalysis in Organic Synthesis: A Comprehensive Handbook*, Wiley-VCH, **1995**.
- [179] a) A. M. Klibanov, *Nature* **2001** *409* 241-246; b) M. T. Reetz, *Curr. Opin. Chem. Biol.* **2002**, *6*, 145-150.
- [180] M. T. Reetz, *Methods. Enzymol.* **2004**, *388*, 238-256.
- [181] a) C. M. Clouthier, Kayser, M. M., Reetz, M. T., *J. Org. Chem.* **2006**, *71*, 8431-8437; b) D. M. Marko, *Curr. Org. Chem.* **2006**, *10*, 1265-1287.
- [182] Z.-G. Zhang, L. P. Parra, M. T. Reetz, *Chem. Eur. J.* **2012**, *18*, 10160-10172.
- [183] M. T. Reetz, S. Wu, *J. Am. Chem. Soc.* **2009**, *131*, 15424-15432.
- [184] G. Li, M. Furst, H. R. Mansouri, A. K. Ressmann, A. Ilie, F. Rudroff, M. D. Mihovilovic, M. W. Fraaije, M. T. Reetz, *Org. Biomol. Chem.* **2017**, *15*, 9824-9829.
- [185] D. dos Santos Pisoni, J. Sobieski da Costa, D. Gamba, C. L. Petzhold, A. C. de Amorim Borges, M. A. Ceschi, P. Lunardi, C. A. Saraiva Goncalves, *Eur. J. Med. Chem.* **2010**, *45*, 526-535.
- [186] S. Wu, J. P. Acevedo, M. T. Reetz, *PNAS* **2010**, *107*, 2775-2780.
- [187] a) M. T. Reetz, *Angew. Chem. Int. Ed.* **2011**, *50*, 138-174; b) Z. Sun, Y. Wikmark, J.-E. Bäckvall, M. T. Reetz, *Chem. Eur. J.* **2016**, *22*, 5046-5054.
- [188] M. D. Mihovilovic, F. Rudroff, B. Grötzl, P. Kapitan, R. Snajdrova, J. Rydz, *Angew. Chem. Int. Ed.* **2005**, *44*, 3609-3613.
- [189] F. Rudroff, M. J. Fink, R. Pydi, U. T. Bornscheuer, M. D. Mihovilovic, *Monatsh Chem.* **2017**, *148*, 157-165.
- [190] D. V. Rial, D. A. Bianchi, P. Kapitanova, A. Lengar, J. B. van Beilen, M. D. Mihovilovic, *Eur. J. Org. Chem.* **2008**, *2008*, 1203-1213.
- [191] a) J. Heringa, *Comput. Chem.* **1999**, *23*, 341-364; b) W. Pirovano, K. A. Feenstra, J. Heringa, *Bioinformatics.* **2008**, *24*, 492-497; c) S. Feroz, TU Wien (Austria), **2012**.
- [192] a) W. R. Perry L. J., *Protein Eng.* **1987**, *1*, 101-105; b) B. A. H. Kim Y. H., Spencer D. S., Stites W. E., *Protein Eng.* **2001**, *14*, 343-347.
- [193] N. Carrillo, E. Ceccarelli, O. Roveri, *Biotechnol. Genet. Eng. Rev.* **2010**, *27*, 367-382.
- [194] R. J. Fox, M. D. Clay, *Trends Biotechnol.* **2009**, *27*, 137-140.
- [195] R. C. Edgar, *Nucleic Acids Res.* **2004**, *32*, 1792-1797.
- [196] A. Dereeper, V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J. F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J. M. Claverie, O. Gascuel, *Nucleic Acids Res.* **2008**, *36*, W465-469.
- [197] K. Arnold, L. Bordoli, J. Kopp, T. Schwede, *Bioinformatics.* **2006**, *22*, 195-201.
- [198] N. Guex, M. C. Peitsch, *Electrophoresis.* **1997**, *18*, 2714-2723.
- [199] N. J. Greenfield, *Nat. Protoc.* **2006**, *1*, 2876-2890.
- [200] G. Holzwarth, P. Doty, *J. Am. Chem. Soc.* **1965**, *87*, 218-228.
- [201] N. J. Greenfield, G. D. Fasman, *Biochemistry* **1969**, *8*, 4108-4116.
- [202] S. Y. Venyaminov, I. A. Baikalov, Z. M. Shen, C. S. C. Wu, J. T. Yang, *Anal. Biochem.* **1993**, *214*, 17-24.

[203] K. Katoh, D. M. Standley, *Mol. Biol. Evol.* **2013**, *30*, 772-780.