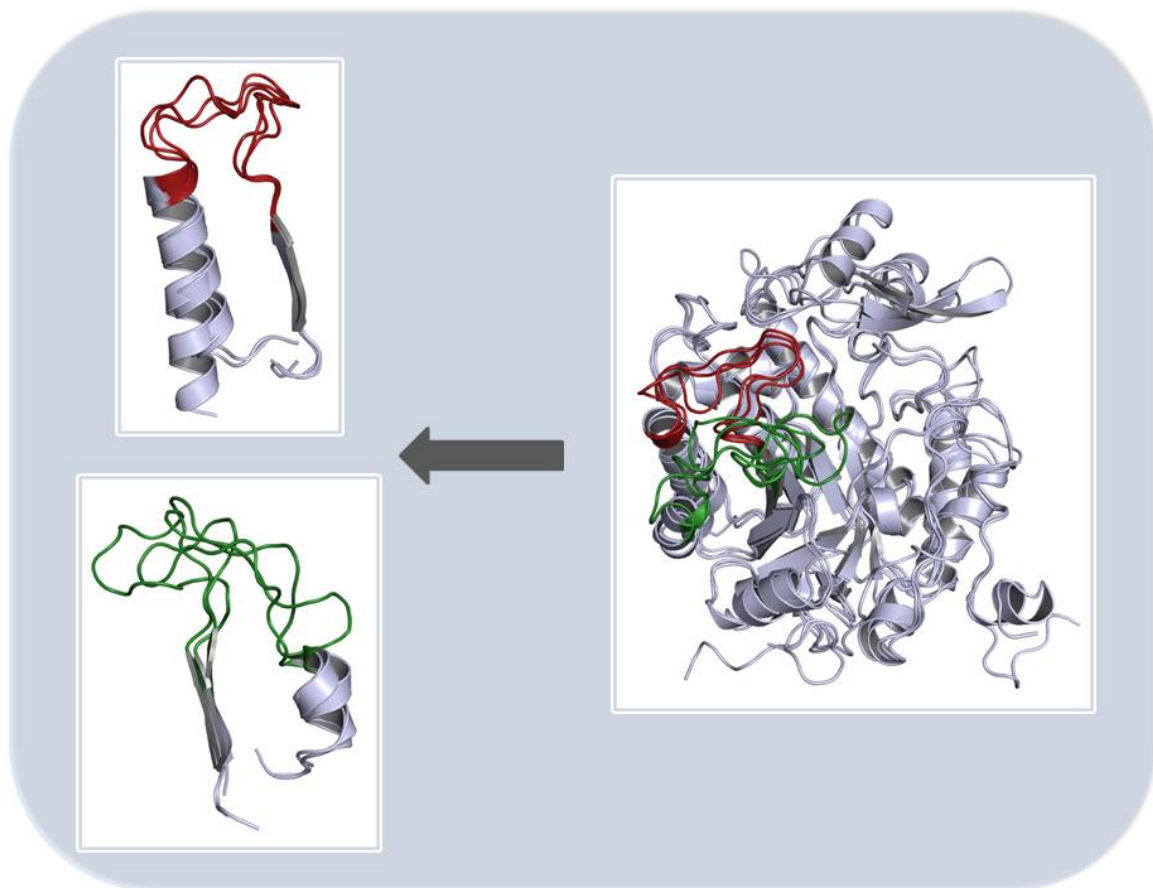


Studies on variable surface loop regions of the ene reductase NCR from *Zymomonas mobilis*

Studien über variable Oberflächenloops der En-Reduktase NCR aus *Zymomonas mobilis*



Sabrina Reich

Institut für Technische Biochemie
Universität Stuttgart

**Studies on variable surface loop regions of the ene reductase NCR
from *Zymomonas mobilis***

Studien über variable Oberflächenloops der En-Reduktase NCR aus
Zymomonas mobilis

Von der Fakultät 4 Energie-, Verfahrens- und Biotechnik der Universität Stuttgart zur
Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigte Abhandlung

Vorgelegt von

Sabrina Reich

aus Calw

Hauptberichter: Prof. Dr. Bernhard Hauer

Mitberichter: Prof. Dr. Karl-Heinrich Engesser

Tag der mündlichen Doktorprüfung:

04.04.2014

Institut für Technische Biochemie der Universität Stuttgart

2014

Die vorliegende Arbeit entstand auf Anregung und unter Anleitung von Prof. Dr. Bernhard Hauer in der Zeit von November 2010 bis Oktober 2013 am Institut für Technische Biochemie der Universität Stuttgart.

Cover:

Strukturelle Überlagerung der drei En-Reduktasen NCR (pdb Datei 4a3u), OYE1 (pdb Datei 1oya) und MR (pdb Datei 1gwj) mit Vergrößerung der beiden ausgewählten Loopbereiche A (rot) und B (grün). Die Bilder wurden mit PyMol generiert.

Im Rahmen dieser Dissertation wurden folgende Publikationen vorab veröffentlicht:

- S. Reich, H. W. Hoeffken, B. Rosche, B. M. Nestl and B. Hauer. „Crystal Structure Determination and Mutagenesis Analysis of the ene Reductase NCR.” *ChemBioChem* 2012, 13 (16), 2400.
- S. Reich, N. Kress, B. M. Nestl and B. Hauer. „Variations in the stability of NCR ene Reductase by rational enzyme loop modulation.” *J. Struc. Biol.* 2014, 185 (2), 228.
- S. Reich, M. Widmann, M. Genz, B. M. Nestl and B. Hauer. „Surface Loops in Proteins are Enzyme Modifying Elements.“, *manuscript in preparation*

Erklärung über die Eigenständigkeit der Dissertation

Ich, Sabrina Reich, versichere, dass ich die vorliegende Arbeit mit dem Titel „Studien über variable Oberflächenloops der En-Reduktase NCR aus *Zymomonas mobilis*“ selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Passagen und Gedanken, die aus fremden Quellen entnommen worden, sind als solche kenntlich gemacht. Des Weiteren bestätige ich in ausdrücklicher Form, dass die hier vorliegende Dissertation nicht in gleicher oder ähnlicher Art und Weise bei einer anderen Institution zur Erlangung eines akademischen Grades eingereicht wurde.

Declaration of Authorship

I, Sabrina Reich, hereby declare that the present thesis entitled „Studies on variable surface loop regions of the ene reductase NCR from *Zymomonas mobilis*“ is the result of my own work, that all sources used or quoted have been indicated, and that I have not used any illegitimated means. I further declare that I have not submitted this thesis for a degree in some form or another.

Name: Sabrina Reich

Ort und Datum/ Place and Date: Althengstett, den 16.12.2013

Meiner Mutter Elke Reich,
meinem Bruder Michael Reich
und in Angedenken an meinen Vater Günter Reich

Acknowledgements

Ich möchte mich an dieser Stelle bei vielen Personen bedanken, die maßgeblich dazu beigetragen haben, dass diese Dissertation in der vorliegenden Form entstanden ist.

Mein erster und größter Dank gilt meinem Doktorvater Prof. Dr. Bernhard Hauer, der es mir ermöglichte mich in den vergangenen drei Jahren mit einem unglaublich spannenden und herausfordernden Projekt im Rahmen meiner Dissertation zu beschäftigen. Des Weiteren möchte ich mich für die ausgezeichnete Betreuung, sowie die vielen hilfreichen Diskussionen und Gespräche bedanken.

Daneben gilt mein Dank im Besonderen Dr. Bettina Nestl für ihre exzellente Betreuung, das intensive Korrekturlesen aller über die Jahre erstellten, schriftlichen Abhandlungen, sowie die vielen Stunden konstruktiver und inspirierender Gespräche. Sie war mir eine große fachliche, wie auch persönliche Stütze während der letzten drei Jahre.

Danken möchte ich auch meinen Kooperationspartnern Maika Genz von der Universität Leipzig und Dr. Michael Widmann aus der Bioinformatikgruppe des Instituts für Technische Biochemie der Universität Stuttgart, die mir in kristallographischen, sowie bioinformatischen Aufgaben wertvolle Gesprächspartner und Unterstützer waren. Des Weiteren danken möchte ich Dr. Daniel Scheps, Nico Kress, Timon Eckes und Lisa Morlock, die in verschiedenen Phasen der Dissertation mitgeholfen haben diese zu einem Erfolg zu machen. Zusätzlich möchte ich mich bei allen Mitgliedern des ITB für eine wunderbare und inspirierende Arbeitsatmosphäre, sowie für viele schöne persönliche Momente bedanken.

Für die finanzielle Förderung dieser Dissertation möchte ich mich bei der Landesgraduiertenförderung Baden-Württemberg bedanken, die mich über diese drei Jahre mit einem Stipendium finanziert hat.

Mein letzter Dank gilt meinen Hauptstützen während der drei Jahre meiner Dissertation, meiner Mutter, meinem Bruder und meinem Freund. Sie waren zu jeder Zeit für mich da und haben mich in all meinen schweren Zeiten, sei es im Rahmen

meiner Dissertation oder in privaten Situationen begleitet und mich jeder Zeit mit voller Kraft unterstützt. Ich möchte mich bei ihnen für ihre grenzenlose Geduld und die vielen Stunden bedanken, in denen sie mein Übungspublikum für all meine Präsentationen waren; oder meine wissenschaftlichen Publikationen, sowie die vorliegende Arbeit korrigiert, haben.

Sabrina Reich

Table of contents

ACKNOWLEDGEMENTS	I
TABLE OF CONTENTS	III
ABBREVIATIONS	VI
ABSTRACT	IX
ZUSAMMENFASSUNG	XII
1 INTRODUCTION	1
1.1 Enzymes	1
1.1.1 Structure and function of enzymes	1
1.1.2 Exploration of protein evolution	3
1.1.3 Enzymes as biocatalysts	4
1.2 Flavoproteins	6
1.3 The family of the Old Yellow Enzymes	8
1.3.1 The subclasses of Old Yellow Enzymes	8
1.3.2 Structural features and reaction mechanism	12
1.3.3 Substrate und reaction spectra of Old Yellow Enzymes	16
1.4 Enzyme engineering of biocatalysts	19
1.4.1 Rational enzyme design	20
1.4.2 Directed, laboratory evolution of enzymes	21
1.4.3 Fusion methods of rational design and directed evolution	24
1.5 Aim of the work	27
2 EXPERIMENTAL SECTION	28
2.1 Genes, proteins, vectors, primers and strains	28
2.2 Methods	28
2.2.1 General molecular biological methods	28
2.2.2 Golden Gate Shuffling	28
2.2.3 Rational loop analysis	33
2.2.4 Expression and purification of wild type enzymes and enzyme variants	36

2.2.5	Biotransformation with wild type enzyme and created variants	38
2.2.6	GC analytic	40
3	RESULTS	41
3.1	Semi-rational loop design – Golden Gate Shuffling	41
3.1.1	Enzyme selection and fragment definition of the ene reductases for the Golden Gate Shuffling	41
3.1.2	Screening of the Golden Gate Shuffling variants	44
3.1.3	Biotransformation with Golden Gate Shuffling variants	46
3.2	Rational loop design	48
3.2.1	Determination of loop regions and model enzymes	48
3.2.2	Systematic and bioinformatic analysis of the OYE family and the selected β/α loop regions	54
3.2.3	Design of rational loop variants	64
3.2.4	Biotransformation with the designed loop variants	72
4	DISCUSSION	89
4.1	Golden Gate Shuffling-valuable tool for semi-rational loop exchange	89
4.2	Surface loops as small elements in the evolution of enzymes	94
4.3	Influence of surface loop regions on enzyme properties	101
5	CONCLUSION AND OUTLOOK	106
6	REFERENCES	109
7	SUPPLEMENTARY MATERIAL	121
7.1	Proteins, vectors, vector constructs, primers and strains	121
7.1.1	Proteins	121
7.1.2	Vectors	122
7.1.3	Vector constructs	123
7.1.4	Primers	124
7.1.5	Strains	127
7.2	Golden Gate Shuffling	128
7.2.1	Multiple sequence alignment	128
7.2.2	Supplementary tables	129
7.2.3	Supplementary figures	130

7.3	Rational loop modulation	131
7.3.1	Crystallization, data collection and structural determination	131
7.3.2	Supplementary tables	133
7.3.3	Supplementary figures	137
	CURRICULUM VITAE	144

Abbreviations

x g	gravitational acceleration
°C	degrees Celsius
µl	microliter
<i>A. thaliana</i>	<i>Arabidopsis thaliana</i>
aa	amino acid
Amp	ampicillin
AMP	adenosine monophosphate
ASM	alanine scanning mutagenesis
ATP	adenosine triphosphate
<i>B. subtilis</i>	<i>Bacillus subtilis</i>
bp	base pair
DNA	deoxyribonucleic acid
<i>E. cloacae</i>	<i>Enterobacter cloacae</i>
<i>E. coli</i>	<i>Escherichia coli</i>
EC	Enzyme Commission
EWG	electron withdrawing group
F	fragment
FAD	flavin adenine dinucleotide
FMN	flavin mononucleotide
g	gram
<i>G. kaustophilus</i>	<i>Geobacillus kaustophilus</i>
GKOE	Old Yellow Enzyme from <i>Geobacillus kaustophilus</i>
GC	gas chromatography
GC/FID	gas chromatography coupled to flame ionization detector
GC/MS	gas chromatography coupled to mass spectrometry
GGs	Golden Gate Shuffling
h	hour

HF	homologous family
IPTG	isopropyl- β -d-thiogalactopyranoside
Kan	kanamycin
kDa	kilo Dalton
l	liter
LB	lysogeny (Luria-Bertani) broth
mg	milligram
min	minute
ml	milliliter
mM	millimolar
mPa	millipascal
MR	morphinone reductase from <i>Pseudomonas putida</i> M10
MTBE	<i>tert</i> -butyl-methylether
NAD(P)H	nicotin adenine di (phosphate) nucleotide
NCR	2-cyclohexen-1-one reductase from <i>Zymomonas mobilis</i>
nm	nanometer
n.d.	not determined
OD ₆₀₀	optical density measured at 600 nm
OPR1	Oxophytodienoate reductase 1 from <i>Arabidopsis thaliana</i>
OYE	Old Yellow Enzyme
OYE1	Old yellow enzyme 1 from <i>Saccharomyces pastorianus</i>
<i>P. putida</i> M10	<i>Pseudomonas putida</i> M10
PCR	polymerase chain reaction
PETN	pentaerythritol tetra nitrate
PETNR	pentaerythritol tetra nitrate reductase from <i>Enterobacter cloacae</i>
PMSF	phenylmethanesulfonyl fluoride
RbCl	rubidium chloride
RE	restriction enzyme
rpm	revolutions per minute

RT	room temperature
<i>S. lycopersicum</i>	<i>Solanum lycopersicum</i>
<i>S. pastorianus</i>	<i>Saccharomyces pastorianus</i>
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
<i>T. pseudethanolicus</i>	<i>Thermoanaerobacter pseudethanolicus</i>
TB	terrific broth
THF	Tetrahydrofuran
TIM	triosephosphate isomerase
TRIS	tris-(hydroxymethyl)-aminoethan
U	units
WT	wild type
X-Gal	5-bromo-4-chloro-3-indolyl- β -galactopyranoside
YqjM	ene reductase from <i>Bacillus subtilis</i>
<i>Z. mobilis</i>	<i>Zymomonas mobilis</i>

Abstract

Enzymes represent powerful biocatalysts, which are next to their natural, indispensable role as fundamental building block of all biochemical processes within a cell also applied in industrial applications. The properties of an enzyme, such as binding and reaction specificity, selectivity and stability, are determined by the individual amino acid sequence and its local spatial arrangement. The present work now deals with the flavin-dependent ene reductases from the Old Yellow Enzyme family which catalyze a *trans*-hydrogenation of activated C=C double bonds with absolute stereospecificity due to the architecture of their active site. All family members possess as a structural backbone a barrel made of eight parallel β -sheets, which are connected by loops with eight surrounding α -helices (TIM barrel structure). Despite great structural and sequential similarities, the individual family members demonstrate significant differences in substrate specificity, thermostability and enantioselectivity. This fact leads to the question where the source of these different preferences of the individual ene reductases comes from. Comparing the amino acid composition as well as the spatial arrangement of the various family members, it is noticeable that the differences between these enzymes are located especially in so-called flexible β/α loop regions on the surface of the catalytic interface. The central question of this work is the influence of these β/α surface loop regions on enzyme's properties and the evolutionary relationship and development of such loop regions within this enzyme family.

In a first approach addressing the role of loops within an enzyme, the simultaneous shuffling of four β/α loop regions between five selected ene reductases was successfully performed by applying the semi-rational Golden Gate Shuffling method. For the loop shuffling monomeric 2-cyclohexen-1-one reductase NCR from *Zymomonas mobilis* was chosen as the structural scaffold. Based on a developed photometric activity assay, a total of five loop shuffling variants were selected for the characterization of the substrate spectrum. This characterization has shown that the random simultaneous shuffling of up to four β/α surface loop regions resulted in enzyme variants demonstrating compared to wild type NCR an increased reduction activity towards standard substrates. Therefore, alterations in β/α loop regions are

not causing inactivation of the enzyme, but instead may have positive effects on the catalytic activity of the enzyme.

In a second approach it was investigated, whether β/α surface loop regions represent repetitive appearing elements with conserved commonalities. A further key component of this approach was the flexibility of these loops determined by the structure and the related amino acid composition. Therefore, by using phylogenetic analysis based on the overall sequence identity, the family of the Old Yellow Enzymes was divided into five homologous subfamilies. One of these subfamilies was investigated in more detail with regard to the occurrence, as well as the evolutionary relationship of β/α surface loops. For this purpose, two β/α surface loops, Loop A and Loop B, in three different ene reductases, namely 2-cyclohexen-1-one reductase NCR from *Zymomonas mobilis*, Old Yellow Enzyme 1 OYE1 from *Saccharomyces pastorianus* and Morphinone Reductase MR from *Pseudomonas putida* M10, were defined by a sequence-based secondary structure prediction. The generation of Hidden Markov loop profiles led to the conclusion that β/α surface loops are composed of conserved, as well as variable amino acid positions and that individual family members can be assigned to certain loop profiles. Thus, each loop possesses a specific amino acid profile which can be found again in other ene reductases of the same homologous subfamily. Additionally to that Hidden Markov loop profiles pointed out that the presence of a common loop region at two different ene reductases is not depending on the overall sequence identity. Two widely related family members could exhibit the same loop profile. β/α Loop regions can therefore be assigned to specific profiles based on their amino acid length and composition. The occurrence of conserved regions within a loop profile is indicating towards an evolutionary context.

In a third approach the question, whether the loop length and/or the amino acid composition of a loop area possess a significant influence on the properties of an enzyme was addressed by the development of a rational loop design method. Therefore, a total of seven loop variants of the two structurally defined loop regions A and B were designed: On the one hand by the rational loop length variation of the intrinsic NCR loops and on the other hand by specific loop grafting between NCR and OYE1 and MR, respectively. The crystal structure determination of the loop grafting variant of Loop A from NCR against the corresponding loop of OYE1 revealed that

the exchange of loop regions led to changes that affect the complete enzyme structure. Furthermore, it was shown by the seven rational variants that both, the amino acid composition as well as the loop length of β/α surface loops of TIM barrel proteins possess a significant effect on the properties of an enzyme. Thus, in contrast to Loop B changes in the Loop A region in length or amino acid composition led to a decreased stability of the enzyme. Additionally, it could be shown that it was possible to transfer properties of one enzyme to another by the grafting of β/α surface loop regions, for example the *cis/trans* substrate specificity.

In summary, the present work could demonstrate that TIM barrel based enzymes are able to tolerate large structural as well as sequential alterations within their β/α surface loops without losing their catalytic activity. Based on the results obtained within this thesis it could be concluded that β/α surface loop regions of Old Yellow Enzymes are representing "Enzyme Modifying Elements" that possess a significant influence on the properties of the entire enzyme. Furthermore, it was also shown for the first time that surface loops can be assigned to specific loop profiles consisting of conserved and variable regions. Based on the acquired understanding of the influence of loop regions on enzyme properties, it should in the near future be possible to create a tailor-made reductase with desired properties by rational loop design.

Zusammenfassung

Enzyme sind leistungsstarke Biokatalysatoren, die aufgrund ihrer vielen wünschenswerten Eigenschaften neben ihrer natürlichen, unverzichtbaren Rolle als Grundbaustein aller biochemischen Prozesse innerhalb einer Zelle auch Einzug in industrielle Anwendungen gefunden haben. Die Eigenschaften eines Enzymes wie Bindungs- und Reaktionsspezifität, Selektivität und Stabilität werden durch die Abfolge der einzelnen Aminosäuren, sowie deren lokale räumliche Anordnung bestimmt. Die vorliegende Arbeit beschäftigt sich nun mit Flavin-abhängigen En-Reduktasen der *Old Yellow Enzyme* Familie, die aufgrund der Architektur ihres aktiven Zentrums eine *trans*-Hydrierung von aktivierten C=C Doppelbindungen mit absoluter Stereospezifität katalysieren. Als strukturelles Grundgerüst besitzen alle Familienmitglieder ein Fass gebildet aus acht parallelen β -Faltblätter, die mittels Loops mit acht umgebenden α -Helices verbunden sind (*TIM barrel* Struktur). Trotz großer struktureller, sowie sequenzieller Gemeinsamkeiten, weisen die einzelnen Familienmitglieder deutliche Unterschiede bezüglich Substratspezifität, Thermostabilität und Enantioselektivität auf. Daraus ergibt sich die Frage, worin der Ursprung dieser unterschiedlichen Präferenzen der einzelnen En-Reduktasen begründet liegt. Vergleicht man nun dazu die Aminosäurezusammensetzung, sowie die räumliche Struktur der verschiedenen Familienmitglieder, fällt auf, dass sich diese Enzyme vor allem in sogenannten flexiblen β/α -Loopregionen an der Oberfläche des katalytischen *Interface* unterscheiden. Die zentrale Frage die sich nun stellt, und die in dieser Arbeit mittels drei verschiedenen Ansätze beantwortet werden soll ist, welchen Beitrag diese oberflächlichen β/α -Loopregionen für die Eigenschaften eines Enzymes leisten, sowie ob sich eine evolutive Entwicklung der Loopregionen erkennen lässt.

In einem ersten Ansatz zur Beantwortung der Rolle nach Loops innerhalb eines Enzymes ist es mittels der Methode des semi-rationalen *Golden Gate Shufflings* gelungen vier β/α -Loopregionen zwischen fünf ausgewählten En-Reduktasen simultan zu durchmischen. Als strukturelles Grundgerüst für den Loopaustausch wurde die monomere 2-Cyclohexen-1-on Reduktase NCR aus *Zymomonas mobilis* gewählt. Basierend auf einem im Rahmen dieser Arbeit entwickelten, photometrischen Aktivitätsassay wurden fünf Loopaustausch Varianten ausgewählt,

die hinsichtlich ihres Substratspektrums charakterisiert wurden. Es konnte gezeigt werden, dass durch das zufällige, simultane Austauschen von bis zu vier β/α -Loopregionen Enzymvarianten erhalten wurden, die im Vergleich zum NCR Wildtyp eine erhöhte Reduktaseaktivität gegenüber Standardsubstraten zeigten. Änderungen in β/α -Loopregionen führen also nicht automatisch zur Inaktivierung des Enzyms, sondern können positive Einflüsse auf die katalytische Aktivität des Enzyms haben.

In einem zweiten Ansatz wurde der Frage nachgegangen, ob es sich bei β/α -Loopregionen um wiederholt auftretende Elemente mit konservierten Gemeinsamkeiten handelt, oder ob deren Form und Zusammensetzung völlig flexibel ist. Dazu wurde die Familie der *Old Yellow Enzyme* mittels phylogenetischer Analysen basierend auf der Gesamtsequenzidentität in fünf homologe Unterfamilien unterteilt. Eine dieser Unterfamilien wurde hinsichtlich des Auftretens und der evolutiven Zusammenhänge von β/α -Oberflächenloops genauer untersucht. Dazu wurden in insgesamt drei En-Reduktasen, 2-Cyclohexen-1-one Reduktase NCR aus *Zymomonas mobilis*, *Old Yellow Enzyme 1* OYE1 aus *Saccharomyces pastorianus* und der Morphinon Reduktase MR aus *Pseudomonas putida* M10 je zwei β/α -Loopregionen durch Sekundärstrukturvorhersage bestimmt, Loop A und Loop B. Es konnte mittels der Erstellung von Hidden Markov Loop Profilen gezeigt werden, dass sich β/α -Loops aus konservierten, sowie variablen Aminosäuren zusammensetzen und dass sich einzelne Familienmitglieder bestimmten Loop-Profilen zuordnen lassen. Jeder Loop besitzt also ein bestimmtes Aminosäureprofil, das sich auch in anderen En-Reduktasen derselben homologen Unterfamilie wieder finden lässt. Anhand der Loop-Profile wurde auch deutlich, dass das Vorkommen eines gemeinsamen Loopbereiches bei zwei verschiedenen En-Reduktasen nicht abhängig ist von der Gesamtsequenzidentität der beiden. Zwei weit verwandte Familienmitglieder können das gleiche Loop-Profil besitzen. β/α -Loopregionen können anhand ihrer Aminosäurelänge und Zusammensetzung also bestimmten Profilen zugeordnet werden. Das Auftreten konservierter Bereiche innerhalb eines Loop-Profils deutet auf einen evolutiven Zusammenhang hin.

In einem dritten Ansatz wurde mittels der Entwicklung einer rationalen Loopdesign Methodik untersucht, ob die Looplänge sowie die Aminosäurezusammensetzung eines Loopbereiches einen entscheidenden Einfluss auf die Eigenschaften eines

Enzyms ausübt. Dazu wurden insgesamt sieben Loopvarianten der beiden strukturell definierten Loopbereiche A und B von NCR aus *Z. mobilis*, basierend auf rationaler Looplängenveränderung der intrinsischen NCR Loops, sowie des gezielten Loopaustausches zwischen NCR und OYE1, beziehungsweise NCR und MR, erstellt. Die Kristallstrukturaufklärung der Loopaustauschvariante Loop A aus NCR gegen OYE1 brachte die Erkenntnis, dass der Austausch von Loopregionen zu Veränderungen führt, die die ganze Enzymstruktur betreffen. Des Weiteren konnte anhand der sieben rationalen Loopvarianten gezeigt werden, dass sowohl die Aminosäurezusammensetzung, als auch die Looplänge der β/α -Loops von *TIM barrel* Enzymen einen deutlichen Effekt auf die Eigenschaften eines Enzyms haben. So haben im Gegensatz zu Loop B Veränderungen im Loop A zu einer verminderten Stabilität des Enzymes geführt. Es konnte auch gezeigt werden, dass man anhand des Loopaustausches Eigenschaften von einem Enzym auf ein anderes übertragen kann wie zum Beispiel die *cis/trans* Substratspezifität.

Zusammenfassend konnte anhand der hier vorliegenden Arbeit gezeigt werden, dass *TIM barrel* basierte Enzyme große strukturelle, wie auch sequentielle Änderungen in ihren β/α -Oberflächenloops tolerieren können ohne ihre katalytische Aktivität zu verlieren. Man kann dank der in dieser Arbeit gewonnenen Ergebnisse folgern, dass es sich bei β/α -Oberflächenloops von *Old Yellow Enzymes* um „Enzym modifizierende Elemente“ handelt, die einen deutlichen Einfluss auf die Eigenschaften des gesamten Enzyms haben. Außerdem wurde ebenfalls zum ersten Mal gezeigt, dass sich Oberflächenloops bestimmten Loop-Profilen zuordnen lassen und aus konservierten, sowie variablen Bereichen bestehen. Dank des hier erlangten Verständnisses des Einflusses von Loopregionen auf Enzymeigenschaften sollte es in naher Zukunft möglich sein anhand der designten Loopbereiche eine Reduktase mit gewünschten Eigenschaften rational zu erstellen.

1 Introduction

1.1 Enzymes

1.1.1 Structure and function of enzymes

Proteins, especially enzymes, are amazing tools of nature. A myriad of chemical reactions is catalyzed by a variety of enzymes, which are present in all living cells building the indispensable basis of life. Enzymes are typically globular macromolecules representing the efficient, powerful and highly specific catalysts of biological systems which facilitate almost all biochemical reactions within a cell by the reduction of the reaction activation energy. The most outstanding properties of enzymes are their catalytic strength and high specificity which are directly linked to the enzyme structure since the loss of structure results in the destruction of the catalytic activity. The spatial structure of all proteins, including enzymes, can be divided into four levels (figure 1.1):

(a) The primary structure is the undermost level of proteins which is represented by the linear chain of amino acids linked by uncharged, planar peptide bonds. The amino acid sequence of a protein is encoded by the DNA sequence of the corresponding gene.

(b) The secondary structure is characterized by frequently, local structural elements which are formed by hydrogen bridge bonds between the C=O and NH-group of amino acids in the polypeptide backbone located in close proximity to each other. Three secondary structure elements are distinguished: α -helix, β -sheet and variable loop elements. The appearance of each secondary element depends on the side chains of the present amino acids. There are amino acids, which demonstrate a preference for occurring in α -helices like glutamic acid, methionine, and leucine, in β -sheets like valine, isoleucine and phenylalanine or in loop regions like proline, glycine and aspartic acid. The three secondary structure elements normally have different functions within a protein, while α -helices and β -sheets are often shape forming and stability conferring, variable loop elements are important for

the function of the protein, including catalysis, substrate specificity and protein-protein interactions.¹

(c) The tertiary structure represents the general arrangement of amino acids located far apart from each other in the linear sequence building a polypeptide chain. The driving force of the tertiary structure formation is the distribution of polar and unpolar amino acids over the protein in dependence from the surrounding environment. In an aqueous media, for example, all unpolar amino acids are located inside the protein, while the hydrophilic, polar amino acids are located at the surface interacting with the environment. The tertiary structure represents already a complete, catalytically active protein which is consisting of just one amino acid chain, called a monomeric enzyme.

(d) The quaternary structure is the spatial arrangement of two or more monomeric enzymes possessing a defined tertiary structure. This is demonstrated by complex proteins consisting of several subunits. The simplest quaternary structure is a dimer consisting of two identical subunits. The different subunits are held together by hydrogen bridge bonds, van der Waals forces as well as Coulomb forces.²

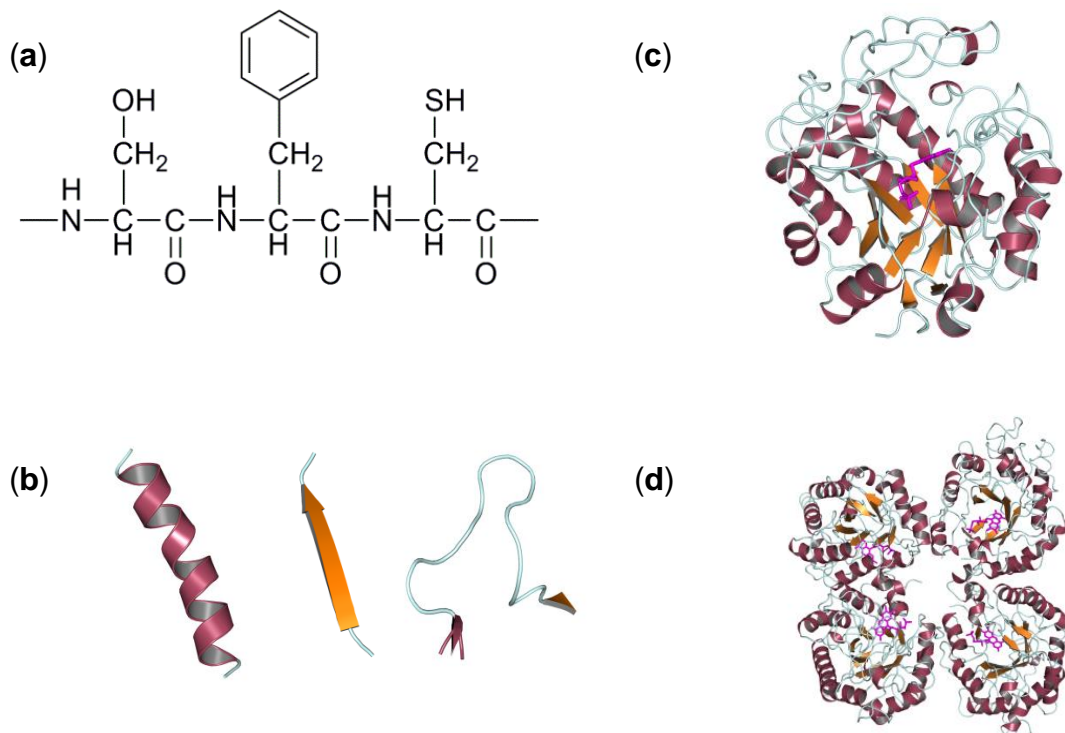


Figure 1.1: The four levels of protein organization. (a) Primary structure of proteins consisting of the amino acids linked by peptide bonds (b) Secondary structure elements represented by α -helices, β -sheets and variable loop regions (c) Tertiary structure of the 2-cyclohexen-1-one reductase NCR from *Zymomonas mobilis* (pdb file 4a3u) which is built by the sequence of the secondary structure elements (d) Quaternary structure of the ene reductase TOYE from *Thermoanaerobacter pseudethanolicus* E39 (pdb file 3krz) which is set up by four monomers.

Under physiological conditions it is mandatory that a defined primary structure results in a defined tertiary structure. The amino acid sequence determines the final structure and thereby also the catalytic activity of the enzyme.

Next to structural features, enzymes can be distinguished according to their catalyzed reaction type leading to a classification of six different enzyme classes numbered from EC 1 to EC 6 (table 1.1).

Table 1.1: The six enzyme classes classified after the reaction type^{2,3}

Enzyme class	Reaction type	Example
EC 1 Oxidoreductases	oxidation/reduction reactions	ene reductases alcohol dehydrogenases
EC 2 Transferases	transfer of functional groups	methyl transferases, transaminases
EC 3 Hydrolases	hydrolysis reactions	lipases, proteases
EC 4 Lyases	addition/elimination of groups for the formation of double bonds	aldolases, dehydratases
EC 5 Isomerases	isomerisation reactions	racemases topoisomerases
EC 6 Ligases	ligation of two substrates by ATP hydrolysis	DNA-ligases, carboxylases

1.1.2 Exploration of protein evolution

Just as every macromolecule present in a living cell, also proteins undergo a constant process of variation and selection which is called protein evolution. The evolution of proteins is an ongoing process giving rise to ever new protein variants, which could lead to the development of new functionalities or even new proteins. Due to the fact that the three dimensional structure is that property of a protein which correlates most closely with its function, it is best suited to investigate the evolutionary relationships among different proteins. If two proteins demonstrate the same structural scaffold it can be assumed that they are descended from a common ancestor and thus are related with each other. Proteins which are originating from a common ancestor are called homologous.⁴ Due to the fact that next to structural commonalities, homologous proteins also possess similarities on the level of amino

acid sequence it is possible to generate phylogenetic trees of aligned proteins presenting the relationship of the proteins among each other.

Proteins and their associated functions have evolved as a consequence of inherited alterations to genes. The genetic events in the evolution of the protein world are duplications, gene fusions as well as point mutations, insertions and deletions. The basic foundation for the evolution of proteins is given by their modular composition of different structural elements, like on the lowest level α -helices, β -sheets and loop regions, followed by the level of protein domains and finally the protein fold level. If one now looks at the lowest level of modular composition, the secondary structure elements, than flexible loop regions play an important role on the properties and function of a protein. For example loops are involved in the catalysis of enzymes or in the substrate specificity. Sequence changes, such as insertions, deletions or module exchanges are frequently localized in such loop regions. Therefore, loop elements can provide the basis for functional diversity and the route to evolutionary divergence of new functions.¹ The next level of protein organization is built by protein domains. A protein domain represents a folding unit of approximately 175 amino acid residues, which can obtain is three dimensional structure independently from the rest of the protein sequence.⁵ The last level of modular composition is formed by protein folds, which stand for the overall spatial arrangement of the entire protein. On the basis of a few successful formed protein folds, evolution was able to bootstrap its way forward by using several processes like insertions and deletions in loop regions, domain duplication, domain shuffling or the recombination of existing proteins to generate new functional proteins.

1.1.3 Enzymes as biocatalysts

In the last couple of years enzymes, apart from their natural function within living cells, become more and more important as biocatalysts in the asymmetric syntheses of chiral chemicals and drugs for pharmaceutical and industrial applications.^{6,7} The demand of enzymes catalyzing novel chemical reactions is constantly increasing due to the fact that biocatalysts possess several great advantages. They are highly efficient, combinable catalysts, accelerating reaction up to a value of 10^{12} by requiring only a low amount of enzyme in the reaction mixture under environmental

friendly conditions regarding temperature, pressure and solvents.³ Additionally, enzymes demonstrate three different kinds of desired selectivities, in particular chemo-, regio- and enantioselectivity. The catalytic activity as well as the selectivity of an enzyme is determined by the architecture and composition of the active site in which amino acids are positioned to enable substrate binding and to stabilize reaction intermediates. However, the usage of enzymes as biocatalysts in industrial applications also has to deal with limitations. Notwithstanding of their commonly broad substrate spectra, enzymes often demonstrate low specific activities towards non-physiological substrates and are restricted to narrow reaction conditions with regard to temperature, organic solvent tolerance and pH value. Furthermore, the catalytic activity of many enzymes is depending on the presence of additional small molecules, so-called cofactors. These cofactors can be divided into two groups: inorganic metals like Zn^{2+} (carboanhydrase) or Mg^{2+} (hexokinase) as well as small, organic molecules, like coenzyme A (acetyl-CoA-carboxylase) or flavin mononucleotide (2-cyclohexen-1-one reductase^{8,9}) which are derived from vitamins. For example, vitamin B₂ (riboflavin) represents the precursor for flavin mononucleotide (FMN), as well as flavin adenine dinucleotide (FAD), which are both included in many oxidoreductases as cofactors. The organic cofactors can be tightly bound to the enzyme and are therefore termed prosthetic groups. The presence of a cofactor within the active site is mandatory for enzyme activity due to the fact that they are involved in the reaction cycle and its removal result in an inactive biocatalyst. During a catalytic reaction cycle a cofactor undergoes a chemical change followed by a subsequent regeneration resulting in the same cofactor status as at the reaction beginning. The cofactor's chemical change is often implemented by additional coenzymes or cosubstrates as, for example, the reduction of FMN to $FMNH_2$, which is enabled by NAD(P)H as cosubstrate.¹⁰ The involved cosubstrates itself also often require a further regeneration, which could be achieved by an enzyme-coupled assay using glucose-6-phosphat dehydrogenase for the regeneration of $NAD(P)^+$ to NAD(P)H.¹¹ Enzymes which are possessing the same organic molecule as cofactor, also demonstrate normally a related mechanism and can therefore be united in one protein family, for example the flavoprotein family.

1.2 Flavoproteins

Enzymes, belonging to the flavoprotein family are yellow colored oxidoreductases which all possess a flavin cofactor as prosthetic group. The general structure of all flavin cofactors is represented by riboflavin (vitamin B₂) consisting of the redox active tricyclic isoalloxazine chromophore being able to perform one- or two-electron transitions from and to a substrate and a ribityl moiety.¹² The redox potential for the two-electron reduction of a free flavin is about -200 mV.¹³ The most common flavin cofactors present in enzymes are flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN), which are both synthesized *in vivo* from the vitamin riboflavin as starting material.¹⁴ In a first step the riboflavin becomes phosphorylated through the action of a flavin kinase by the consumption of one molecule ATP resulting in FMN. Then subsequently, in a second step FAD is formed from FMN by the addition of an adenosine monophosphate (AMP) catalyzed by the FAD pyrophosphorylase utilizing again ATP as cosubstrate. Flavoproteins are ubiquitous, widely distributed enzymes catalyzing a large variety of different reaction types (figure 1.2).

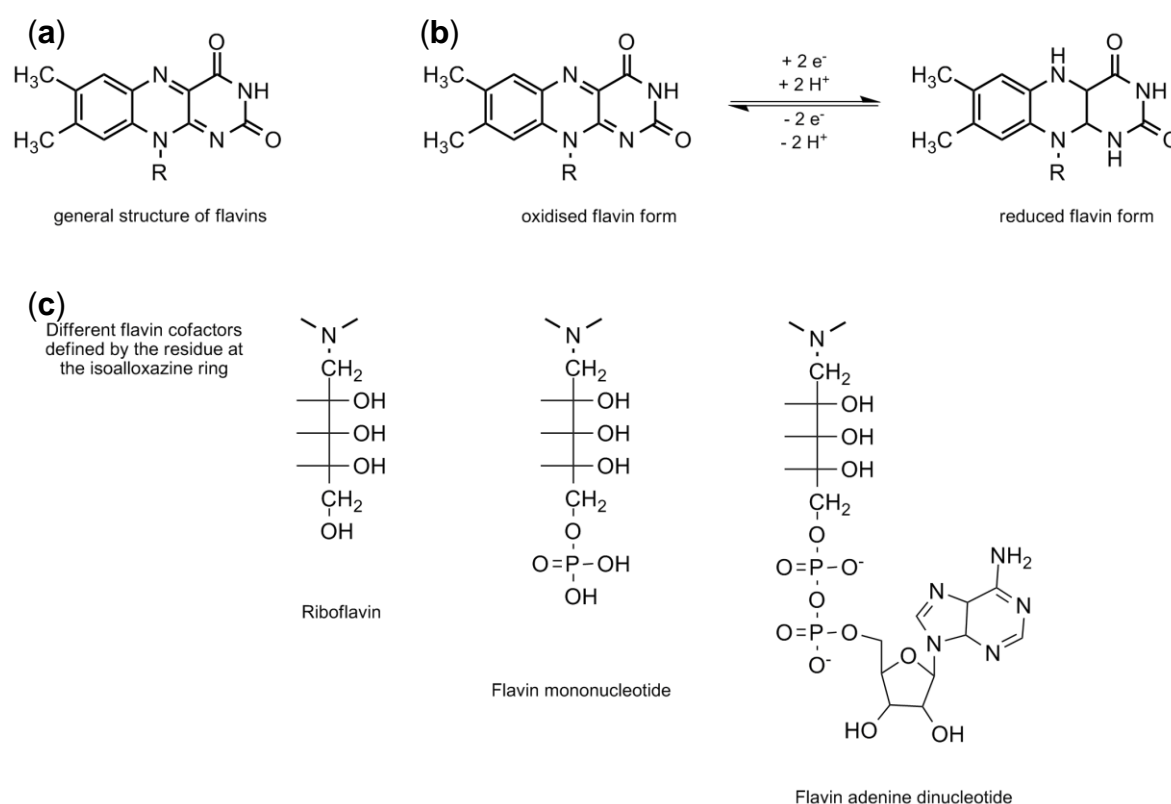


Figure 1.2: Structure of the flavin cofactors. (a) General structure of the isoalloxazine ring (b) Two-electrons uptake and delivery of the isoalloxazine ring (c) Residues at the isoalloxazine ring defining the different flavin cofactors riboflavin, flavin mononucleotide and flavin adenine dinucleotide.

They play a central role in the aerobic metabolism for energy production, are involved in light emission, detoxification and biodegradation processes, as well as in the oxygen activation for oxidation and hydroxylation reactions.^{12,15} It is assumed that, on average 1-3 % of all genes in eukaryotic, as well as bacterial genomes are coding for flavin-binding proteins.¹⁴ Therefore, on basis of their large reaction variety and applicability in biocatalytic reactions, as well as their wide natural distribution it is not surprising that these enzymes are representing one of the best studied enzyme families in terms of reactivity, protein structure and substrate scope.¹²⁻¹⁶ Furthermore, flavoproteins are divided into three subgroups according to their main reaction type after Massey¹²:

(1) *Simple flavoproteins*: All members of this subgroup are proteins containing the flavin cofactor and interact somehow with oxygen. They are further subdivided into oxidases, electron transferases and flavoprotein monooxygenases on the basis of their reactivity towards molecular oxygen. Oxidases, like glucose oxidase, react very fast with molecular oxygen to yield hydrogen peroxide and the oxidized protein without detectable intermediate. Electron transferases, like flavodoxin, demonstrate a single-electron transfer and react slowly with molecular oxygen by forming O_2^- and the flavin semiquinone. Flavoprotein monooxygenases¹⁷, like phenylacetone monooxygenase¹⁸, transfer one electron from the reduced flavin to molecular oxygen leading to the formation of a complex consisting of a superoxide and the flavin radical.

(2) *Flavoproteins with auxiliary redox centers*: All members of this subgroup are more complex enzymes consisting of several prosthetic groups or cofactors. They are also further subdivided, particularly into flavoprotein-disulfide oxidoreductases, heme-containing flavoproteins and metal-containing flavoproteins. One of the most prominent, as well as industrial applied representative of this subgroup of flavoproteins is the cytochrome P450 monooxygenase family. These enzymes possess next to a flavin cofactor (either FAD or FMN or a combination of FAD/FMN) also an essential heme and, in some cases an additional iron-sulfur cluster as cofactor catalyzing the cleavage of molecular oxygen by the incorporation of one atom of molecular oxygen into a substrate while reducing the second one to water.¹⁹

(3) *Flavoenzymes of unknown function*: This subgroup contains all flavoproteins demonstrating a difficult classification with regard on physiological function as well as physiochemical characteristics. Of all things, interestingly, the first of all described

flavoproteins, the Old Yellow Enzyme 1 from brewer's bottom yeast belong to this subgroup. Much of this enzyme and its protein family, known as the Old Yellow Enzyme family, is known and will be presented in detail in the following pages.

1.3 The family of the Old Yellow Enzymes

The family of the Old Yellow Enzymes is named after the first, in 1932 in *Saccharomyces pastorianus* by Warburg and Christian²⁰ discovered family member, a flavin-dependent enzyme which was called a yellow enzyme ("das gelbe Ferment"). When later, in 1938 by Haas and coworkers a second yellow enzyme was identified, the Warburg enzyme was called "Old Yellow Enzyme", and this name sticks until nowadays to name the entire enzyme family. The Old Yellow Enzyme family is still growing in size and widely distributed in nature being present in prokaryotes, yeasts, fungi, plants as well as in parasitic eukaryotes.²¹ Old Yellow Enzymes are ubiquitous in nature catalyzing the reduction of activated carbon-carbon double bonds via a hydride transfer followed by a proton addition resulting in the formation of up to two new chiral centers. Members of this family are also called ene reductases and possess a rather broad substrate scope including ketones, aldehydes, nitroalkenes, nitroesters, nitroaromatics, lactones, carboxylic acids, esters, terpenoids, nitriles and steroids.^{10,22-31} Till today for most of the family members no physiological role could be discovered. However, in nature some of them occur in several well defined pathways, for example in the jasmonic acid biosynthesis^{32,33} of plants or the morphinone biosynthesis³⁴ in *Pseudomonas putida* M10. Additionally there are family members known being involved in the oxidative stress response, like YqjM from *Bacillus subtilis*, which was firstly described in 2003³⁵ and represent the first member of a new subfamily of Old Yellow Enzyme, the *thermophilic-like* subfamily.

1.3.1 The subclasses of Old Yellow Enzymes

In 2005 the crystal structure of an OYE homologous from the mesophilic soil bacterium *Bacillus subtilis*, called YqjM, was solved indicating that next to the classical Old Yellow Enzymes a second subgroup of ene reductases exists, demonstrating several considerable differences particularly on the structural level as

well as the amino acid chain length.³⁶ The first subfamily is called the “classical” subfamily with members like the 2-cyclohexen-1-one reductase NCR from *Zymomonas mobilis*⁸, OYE1 from *Saccharomyces pastorianus*^{37,38} and OPR1 from *Arabidopsis thaliana*³⁹ possessing an amino acid sequence of 385 to 400 amino acids.

	OYE1	MSFVKDFKPQALGDTNLFKPIKIGNNELLHRAVIPPLTRMRALHPGNI PN RDWAVEYYTQ	60
	OYE2	MPFVKDFKPQALGDTNLFKPIKIGNNELLHRAVIPPLTRMRAQHPGNI PN RDWAVEYYAQ	60
classical subclass	OPR1	-----MENGEAKQSVPLLPYKMGFRNLSHRVVLAPLTRQR--SYGNVP-QPHAAIYYSQ	52
	OPR2	---MEMVNAEAKQSVPLLPYKMGFRNLSHRVVLAPLTRQK--SYGSVP-QPHAILIYYSQ	54
	MR	-----MPDTSFSNPGFLTPLQLGSLSLPNRVIMAPLTRSR--TPDSVP-GRLQQIYYGQ	51
	PETNR	-----MSAEKLFPLKVGAVTAPNRVFMAPLTRLRSIEPGDIP-TPLMGGEYYRQ	48
	NCR	-----MPSLFDPIRFGAFTAKNRIWMAPLTRGR-ATRDHVP-TEIMAEYYAQ	45
thermophilic-like subclass	YqjM	-----MARKLFTPTITIKDMTLKNRIVMSPMCYSSEHEKDGKLPFFHMAHYISR	48
	GkOYE	-----MNTMLFSPTIRGLTLKNRIVMSPMCYSKDTKDGAVRTHKLIHYPAR	48
	Crs	-----MALLFTPLELGGRLRLKNRLAMSPMCQYSAT-LEGEVTDWHLHLYPTR	46
	TOYE	-----MSILHMLPKIKDITIKNRIMSPMCYSAS-TDGMNDWHIVHYATR	46
classical subclass	OYE1	WAAFP----DNLARDGLRYDSASDNVFMDAEQEAK-----AKKANPQHSITKDEIKQY	170
	OYE2	WAAFP----DTLARDGLRYDSASDNVYMNAEQEEK-----AKKANPQHSITKDEIKQY	170
	OPR1	RVSN-----SGFQPNKAPISCSKPLMPQIRSN-----IDEALFTPPRRLGIEEIPGI	161
	OPR2	RVSN-----RGFQPRRQAPISCTGKPIPMQMRANG-----IDEARFTPPRRLSIEEIPGI	163
	MR	RVSH-----ELVQPDGQQPVAPSALKAEGAECFVEFEDGTAGLHPTSTPRALETGDIPIGI	164
	PETNR	RISH-----SSIQPGQAPVSASALNANTRTSLRD-ENGNAIRVDTTTPRALELDEIPGI	160
	NCR	RMVP-----SNVS--GMQPVAPSASQAPGLGHTYD-----GKKPYDVARALRLDEIPRL	150
thermophilic-like subclass	YqjM	RKA----ELEG-----DIFAPSAIAFDEQ-----SAPVEMSAEKVKET	142
	GkOYE	RKS----QVPG-----EIIAPSAVPFDD-----SPTPKEMTKADIEET	142
	Crs	RKAGTARWEGGKPLGWRVVGPSPIPFDEG-----YVPEPLDEAGMERI	150
	TOYE	RKCN--ISYED-----VVGPSPIKAGDR-----YKLPRELSVEEIKSI	141
classical subclass	OYE1	IKEYVQAANKNSIAAGADGVEIHSANGYLLNQFLDPHSNTRTDEYGGSIENRARFTLEVVD	230
	OYE2	VKEYVQAANKNSIAAGADGVEIHSANGYLLNQFLDPHSNNRTDEYGGSIENRARFTLEVVD	230
	OPR1	VNDFRLAARNAMEAGFDGVEIHSANGYLIDQFMKDTVNDRTDEYGGSLQNRCKFPLEIVD	221
	OPR2	VNDFRLAARNAMEAGFDGVEIHSANGYLIDQFMKDKVNDRTDEYGGSLQNRCKFALEIVD	223
	MR	VEDYRQAARAKRAGFDMVEVHAANAACLPNQFLATGTNRRTDQYGGSIENRARFPLEVVD	224
	PETNR	VNDFRQAVANAREAGFDLVELHSANGYLLHQFLSPSSNQRTDQYGGSVENRARLLEVVD	220
	NCR	LDDYEKAARHALKAGFDGVEIHSANGYLIDEFIRDSTNHRHDEYGGAVENRIRLLKDVTE	210
thermophilic-like subclass	YqjM	VQEFKQAARAKEAGFDVIEIHSANGYLIEHFLSPLSNHRTDEYGGSPENRYRFLREIID	202
	GkOYE	VQAFQNGARRAKEAGFDVIEIHSANGYLIEHFLSPLSNRQDEYGGSPENRYRFLGEVID	202
	Crs	LQAFVEGARRALRAGFQVIELHMAHANGYLLSSFLSPLSNQRTDAYGGSLNRMRFPLQVAQ	210
	TOYE	VKAFGEAAKRANLAGYDVVEIHSANGYLIEHFLSPLSNKRKDEYGNSEIENRARFLIEVID	201
classical subclass	OYE1	FISNPDLVDRLEKGLPLNKYDRDTFYQ-MSAHGYIDYPTYEEALKLGDWKK	400
	OYE2	FISNPDLVDRLEKGLPLNKYDRDTFYK-MSAEGYIDYPTYEEALKLGDWKN	400
	OPR1	FLANPDLKRFQVDAPLNKYDRPTFYTSDPVVGTYDYPFLESTA-----	372
	OPR2	FLANPDLKRFQVDAPLNKYNRSTFYTSDPVVGTYDYPSLESTA-----	374
	MR	FIANPDLPERFRLGAALNEPDPSTFYG-GAEVGYTDYPFPLDNGHDLRG---	377
	PETNR	YIANPDLVARLQKKAELNQPPESEFYG-GGAEGYTDYPSL-----	365
	NCR	FIGNPDLPRRFFFEKAPLTKDVIEWT- QTPKGYTDYPLLGD-----	358
thermophilic-like subclass	YqjM	LLRDPFFARTAAQLNTEIPAPVQYER-----GW-----	338
	GkOYE	LLRNPYWPYAAARELGAKISAPVQYER-----GWR-----	340
	Crs	LLRDPYFPLRAAKALGVAPEVPPQYQR-----GF-----	349
	TOYE	LLRNPYWLHTYTSK---EDWPKQYER-----AFKK-----	337

Figure 1.3: Excerpt of a multiple sequence alignment of eleven ene reductases divided into two subfamilies. Highlighted in red is the catalytic active amino acid, tyrosine 177 (numbering after NCR). Highlighted in green, respectively cyan are the amino acids which are conserved within the active site and involved in substrate binding: T25, H172 and N175 (NCR numbering). In the *thermophilic-like* subclass T25 is exchanged against C26 and N175 against H169 (YqjM numbering). Marked in blue is an area in which the *thermophilic-like* subclass is clearly shorter in length than the *classical* one and marked in pink is the overall amino acid length of all aligned ene reductases. The multiple sequence alignment was performed with ClustalW.

The second subfamily is called the “*thermophilic-like*” subfamily represented for example by YqjM from *Bacillus subtilis*^{35,36}, GkOYE from *Geobacillus kaustophilus*⁴⁰ and Crs from *Thermus scotoductus* SA-01^{41,42} exhibiting considerably shorter amino acid chain lengths of 337 to 340 residues. The differences in the chain length are mainly attributed to shorter loop regions, which are common features of thermophilic enzymes. Differences, as well as commonalities of the two subfamilies, which could be seen on the amino acid sequence level, are highlighted in figure 1.3. Both subfamilies have in common that they possess a conserved, catalytic active tyrosine in the active site (Y177 at NCR and Y169 at YqjM) and two amino acids which are important for the substrate binding in close proximity to the tyrosine (H172 and N175 at NCR, as well as H164 and H167 at YqjM). The *thermophilic-like* ene reductases need for their activity two histidine residues for the substrate binding, while in the *classical* Old Yellow Enzymes the second histidine could also be exchanged against an asparagine. If one looks now at the conserved N-terminal amino acid threonine (T24 at NCR) of the *classical* subfamily, which is known to regulate the redox potential of the flavin mononucleotide to drive the hydride transfer from NAD(P)H to FMN⁴³, it strikes out that it is exchanged against a conserved cysteine (C26 at YqjM) within the *thermophilic-like* subfamily pointing towards altered redox properties for these enzymes. In addition, a conserved tyrosine (Y343 at NCR) located at the C-terminus of the *classical* ene reductases is not present in the *thermophilic-like* subclass members. Instead the *thermophilic-like* Old Yellow Enzymes possess a conserved tyrosine (Y28) at the N-terminus which has the same role like the C-terminal one in the *classical* subfamily, namely the formation of a hydrogen bridge bond with the activating group of the substrate. Apart from redox modulation of the flavin and substrate binding the two subfamilies also possess large differences in the architecture of the active site, which is caused by an alternative quaternary structure organization. In contrast to the *classical* subfamily, which consists of monomers (NCR figure 1.4a) and dimers of functional monomers (OYE1, MR), the *thermophilic-like* subclass members exhibiting higher oligomerization states, so is for example YqjM a tetramer of catalytically depending dimers. The *thermophilic-like* dimers demonstrate a shared active site architecture in which a C-terminal located, conserved arginine finger (R336 at YqjM) extends into the active site of the adjacent monomer (figure 1.4c). The arginine finger is directly involved in substrate recognition and therefore important for the activity of the enzyme. The two dimers of YqjM are

arranged in such a way that their active sites open up in opposite directions to the solvent, however connected with each other by the C-terminal arginine finger. By finally looking at the accessibility to the active site, it is apparent that due to the architecture of the active site, as well as smaller loop regions, the *thermophilic-like* subfamily members exhibit a considerably wider access tunnel into the active site than the *classical* family members (figure 1.4b & c).

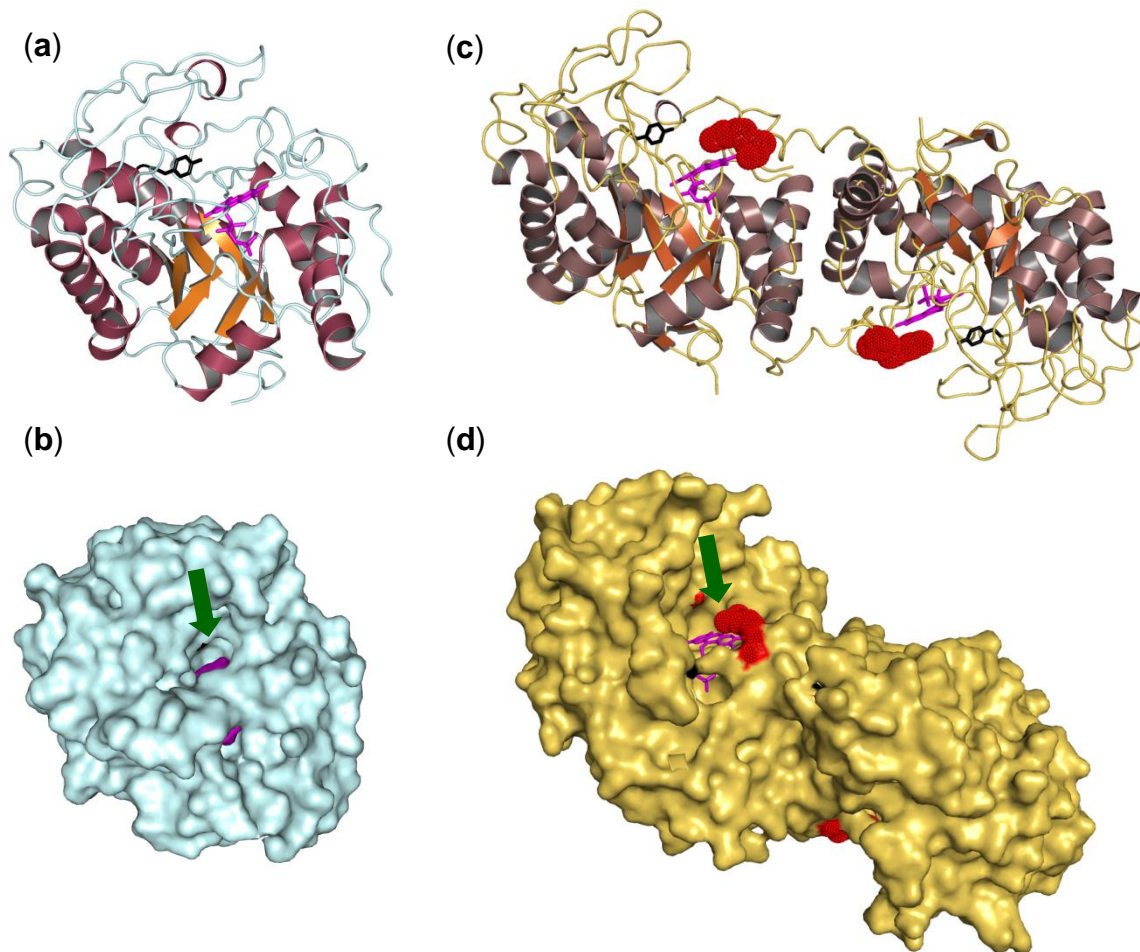


Figure 1.4: Structural representation of the functional unit of the *classical* and the *thermophilic-like* subfamily. (a) Cartoon representation of the monomeric, *classical* family member NCR (pdb file 4a3u) (b) Surface representation of NCR colored in pale cyan with the narrow active site accessing channel marked by a green arrow (c) Cartoon representation of a functional depending dimer of YqjM (pdb file 1z41). The conserved residue R336, representing the in the substrate binding involved arginine finger is colored in red. (d) Surface representation of YqjM colored in yellow orange with the wide access to the active site marked by a green arrow. The catalytic active tyrosine is colored in black and the FMN in magenta.

1.3.2 Structural features and reaction mechanism

The overall protein structure is fundamental important for the activity as well as the stability of an enzyme. Furthermore, a stable, but variable and evolvable protein structure is representing a key feature in the capability of proteins to tolerate changes on their constitution. An evolvable protein structure is made up of a tightly packed scaffold leading to stability and robustness and a versatile active site enabling catalytic plasticity. Additionally, diverse and variable loop regions standing for structural flexibility of the evolvable fold.¹ Therefore, an enzyme is divided into a stable face and a variable catalytic face. In 1994 the first crystal structure of an Old Yellow Enzyme, the OYE1 from *Saccharomyces pastorianus*, was solved showing that the ene reductases possess an $(\alpha/\beta)_8$ barrel, a so-called TIM barrel, as structural scaffold with a size of 40-45 kDa.³⁸ This barrel motif is named after the first described enzyme with such a protein structure, the triose phosphate isomerase.⁴⁴ All TIM barrel based proteins exhibit the structural subdivision with a catalytic face located at the C-terminal end of the β -strand barrel in the versatile β/α loop region and a stable face built by the barrel itself and the α/β loops at the N-terminal end⁴⁵ (figure 1.5).

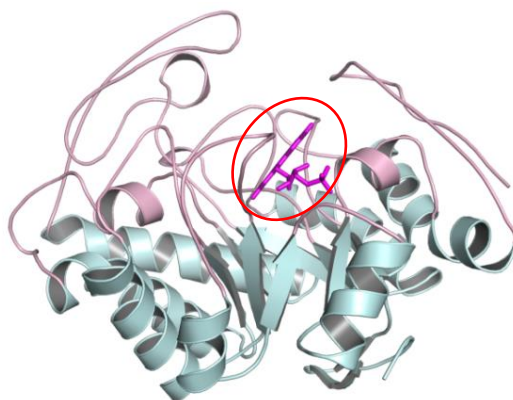


Figure 1.5: Crystal structure of the two face TIM barrel ene reductase NCR from *Zymomonas mobilis* (pdb 4a3u). The stability face is built up by the $(\beta/\alpha)_8$ barrel structure and the α/β loops and is colored in pale cyan. The catalytic face is consisting of the variable β/α loop segments and is colored pink. The prosthetic FMN is shown as sticks and colored in magenta and highlighted in red.

Therefore, the TIM barrel structure motif demonstrates great functional and structural diversity as well as a high catalytic versatility and active sites with high plasticity. In addition and due to the fact that the substrate binding and the catalytic important amino acids are located in separate regions of the protein, TIM barrel based enzymes are considered to represent the ideal scaffolds for the reshaping of the

binding site or for the grafting of new catalytic function by loop swapping.^{46,47} So for instance it was possible to introduce a new isomerase activity in the TIM barrel scaffold of a synthase by *in vitro* mutagenesis and recombination.⁴⁸ Additionally, it could be shown that TIM barrel proteins tolerate large alterations in their catalytic interface obtained via loop swapping experiments.⁴⁷ Generally speaking, loop regions are representing a part of an enzyme in which alterations can be tolerated more easily leading to new enzyme properties. Therefore, such flexible areas in a protein are optimal targets for directed evolution and rational design. It could be shown that rational loop design and loop swapping have an influence on substrate specificity⁴⁹, enantioselectivity^{50,51} or solvent stability⁵² of several enzymes.

The TIM barrel is by far the most common tertiary fold observed in high resolution protein crystal structures and belongs to the class of α/β proteins. It is estimated that approximately 10 % of all known enzymes have at least one Tim barrel domain in their structure.⁵³ Next to the eight parallel β -sheets, forming the barrel and the eight surrounding, stabilizing α -helices, Old Yellow Enzymes demonstrate several additional secondary elements (figure 1.6). A highly conserved additional structural element is a β -hairpin structure at the N-terminus of the protein representing a covering of the bottom of the stability face of the barrel. Furthermore, all ene reductases possess additional secondary structure elements in the large β/α loop region connecting β -sheet β_3 with α -helix α_3 .

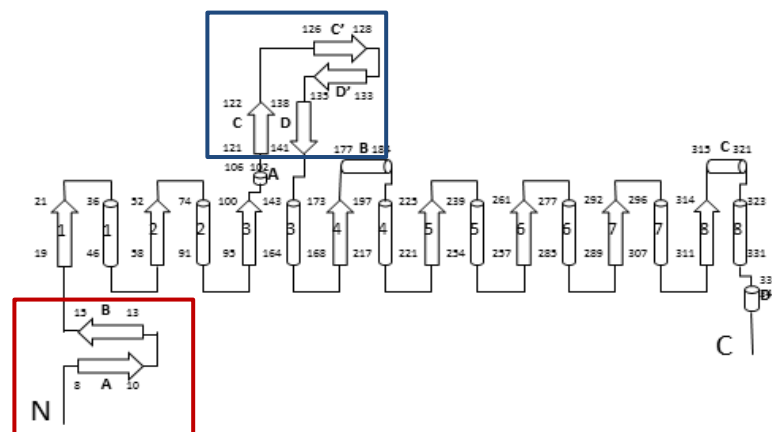


Figure 1.6: Topology diagram for NCR from *Z. mobilis*. The cylinders are representing α -helices and the arrows beta strands. All α -helices and β -sheets are numbered and the amino acid position of the beginning and end of each secondary element is labeled.⁹ Highlighted in red is the additional beta hairpin structure. Highlighted in blue is a large β/α loop region possessing additional β -sheets.

Looking now at the architecture of the catalytic interface, or more precisely, the design of the active site, it strikes out that the active site is highly conserved through the family of the Old Yellow Enzymes. However, each member demonstrates a slightly altered substrate binding region resulting in an altered enzyme specificity and cofactor preference.^{21,38,54} In total there are three catalytically important features present in the active site:

(1) The prosthetic flavin mononucleotide, which is non-covalently bound at the carboxyterminal end of the beta barrel forming the bottom of the active site. The isoalloxazine ring of the FMN is buried with its *re*-face in the beta barrel structure by extensive hydrogen bonding interactions while its *si*-face is exposed to the solvent in the active site. Amino acid residues which are in contact with the FMN by side chain interactions are often highly conserved across both subfamilies, like T25, Q98, R224 and R314 (NCR numbering, figure 1.7a). The redox potential of the flavin cofactor is influenced by the, in the *classical* subfamily highly conserved T25 which forms a hydrogen bond with the FMN. Hence, it is assumed that T25 stabilizes the negative charge of the reduced flavin leading to an increased redox potential.⁴³

(2) The next significant feature is the presence of a catalytic important, proton transferring amino acid which is in all ene reductases, except MR possessing a cysteine, a tyrosine (NCR Y177). During the reduction reaction of carbon-carbon double bonds, it is necessary to transfer a proton on the C_α of the double bond. It could be revealed by site-directed mutagenesis studies of OYE1 that the involved amino acid is the tyrosine 196 of OYE1.⁵⁵ However, it was also shown that a mutation of the tyrosine 177 of NCR against alanine (Y177A) still possesses reduction activity, thus in a reduced yield⁹ leading to the assumption that the required proton is finally derived from the solvent.

(3) The third catalytically important feature in the active site is the existence of two, highly conserved amino acids, which are involved in the binding and proper orientation of the substrate in the active site by interaction with its electron withdrawing group. For the *thermophilic-like* subfamily both substrate binding amino acids are histidines (H164 and H167 of YqjM), while for the *classical* subfamily members just the first amino acid is a histidine (H172 of NCR). The second amino acid could be either a histidine or an asparagine (N175 of NCR). The importance of these two amino acids in the binding of the activating group of the substrate was shown by site-directed mutagenesis studies of OYE1.⁵⁶

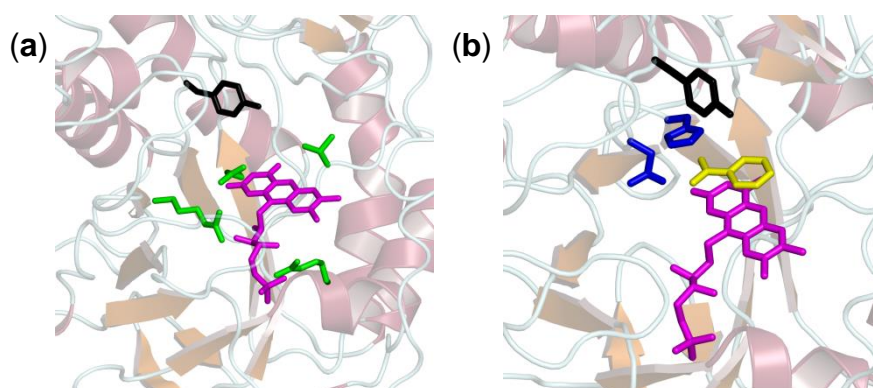


Figure 1.7: Representation of the active site of NCR with the catalytic important tyrosine 177 colored in black and the FMN colored in magenta. **(a)** Localization of the FMN cofactor by the four hydrogen bonding amino acids T25, Q98, R224 and R316 colored in green. **(b)** Substrate binding site, indicated by the bound inhibitor hydroxybenzaldehyde colored in yellow and the two substrate binding amino acids H172 and N175 colored in blue.

The architecture of the active site as well as the presence of the catalytically important amino acids is the basis of the catalytic mechanism demonstrated by Old Yellow Enzymes. The overall reaction of these enzymes is the NAD(P)H dependent reduction of activated carbon-carbon double bonds via a Michael-type addition of $[H_2]$ following a ping pong bi bi mechanism (figure 1.8).⁵⁷

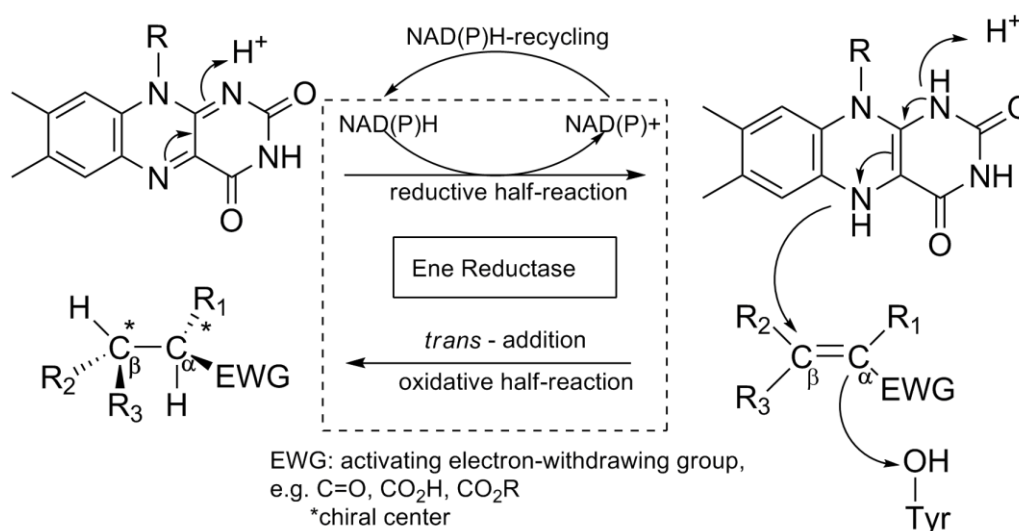


Figure 1.8: Ping pong bi bi reaction mechanism of the Old Yellow Enzyme family consisting of an oxidative and a reductive half-reaction.

The catalytic mechanism is divided into two parts with regard to the flavin cofactor. In the first part, the reductive half-reaction, the prosthetic flavin cofactor is reduced by the physiological cosubstrate NAD(P)H which is bound in an optimal position for the

hydride transfer with the nicotinamide C4 atom, the hydride donor, close to the N5 acceptor atom of the FMN.³⁸ Afterwards, the oxidized NAD(P)⁺ is released from the active site leaving the enzyme in a reduced state for the subsequent second reaction part, the oxidative half-reaction. The physiological substrate for this half-reaction is unknown due to the large variety of compounds which could be reduced by Old Yellow Enzymes. Next to the known [H₂] and electron acceptors of a reduced FMN like molecular oxygen, methylene blue, ferricyanide and quinones⁵⁸, a large number of α,β -unsaturated alkene substrates have been found to act as oxidants for ene reductases.^{8,10,30,59–63} In the oxidative half-reaction a substrate possessing an activated carbon-carbon double bond is bound in the active site through hydrogen bonding interactions of the activating group with the conserved amino acid pair H172 and N175 (NCR numbering). The hydrogen bond interactions result in a polarization of the carbon-carbon double bond and thereby activating the C_β for a nucleophilic attack of the FMNH₂.⁶⁴ The presence of an activating or electron-withdrawing group as substituent at the C_α is absolutely necessary for the reduction reaction. Described activating groups are ketones, aldehydes, nitro groups, carboxylic acids, esters, anhydrides, lactones or imide functionalities¹⁰. After the substrate binding a hydride is transferred onto the C_β of the olefinic bond from the reduced FMNH₂ followed by a proton addition onto the C_α facilitated through the conserved tyrosine 177 from the opposite site. Actually, the proton is ultimately derived from the solvent.⁵⁷ Hence, based on the architecture of the active site, with the FMN localized below and the catalytic active tyrosine above of the bound substrate, the reduction of the carbon-carbon double bond always proceeds in a *trans*-specific fashion. Overall, an additional benefit of the *trans*-specific reduction of activated carbon-carbon double bonds by Old Yellow Enzymes is that the reaction often proceeds with a very high stereo- and enantioselectivity.¹⁰

1.3.3 Substrate und reaction spectra of Old Yellow Enzymes

To this day, a common, overall physiological substrate or role for Old Yellow Enzymes within the cell is still unknown, however, there are some family members known with defined physiological function. For example the plant reductase OPR3 from *Arabidopsis thaliana* which catalyzes an intermediate step in the production of

jasmonic acid, a phytohormone involved in the plant response towards wounding and stress.⁶⁵ Or the bacterial YqjM from *Bacillus subtilis* known to be involved in the oxidative stress response which was proven by an increased *in vivo* expression level of the ene reductase after the addition of the toxic xenobiotic trinitrotoluene or the oxidative stress inducing hydrogen peroxide in the media.³⁵ In addition to their natural function within the cell, some of the reduced products obtained by Old Yellow Enzymes are of interest for industrial production. For example, the production of the semisynthetic opiate drugs hydromorphone and hydrocodone acting as a strong analgesic or mild antitussive by using the morphinone reductase from *Pseudomonas putida* M10.⁶⁶ Furthermore, next to the reduction of natural, also the reduction of non-physiological substrates by ene reductases results in products possessing several industrial interesting applications. For instance, at least eleven Old Yellow Enzymes are able to catalyze the enantiomeric pure production of the diketone (*R*)-levodine, representing a key intermediate in the carotenoid synthesis from the substrate ketoisophorone.^{30,67} Or the ability of the yeast enzyme OYE1 to reduce β -nitro acrylates which represents a key step in the production of optically active chiral β^2 -amino acids.²⁹ If one looks now at the character of non-physiological substrates and reactions which could be performed by Old Yellow Enzymes, it strikes out that these enzymes catalyze next to the so far reported classical activated alkene reduction, a variety of promiscuous reactions.⁶⁸ Figure 1.9 illustrates an excerpt of substrate types and promiscuous reactions which can be catalyze by the ene reductase family. For example, it was reported that the yeast enzymes OYE1, OYE2 and OYE3 are able to accept ionones as substrates, with OYE3 being the best catalyst demonstrating even a higher activity towards the alkyne than towards the corresponding alkene substrate.⁶⁹ In addition, the three yeast enzymes are also able to reduce the activated 2-cyclohexen-1-one without the cofactor NAD(P)H to the corresponding alkane 2-cyclohexanone by using a second substrate molecule of 2-cyclohexen-1-one as electron donor in a so-called disproportionation reaction.^{58,70} In a disproportionation reaction one of two equal substrate molecules gets reduced to the corresponding alkane while the second one becomes oxidized to the according phenol-like product without any required cofactor. Furthermore, there are several ene reductases known, with PETNR from *Enterobacter cloacae* PB2 being the best investigated one, which can catalyze the reduction of xenobiotic, man-made nitro-

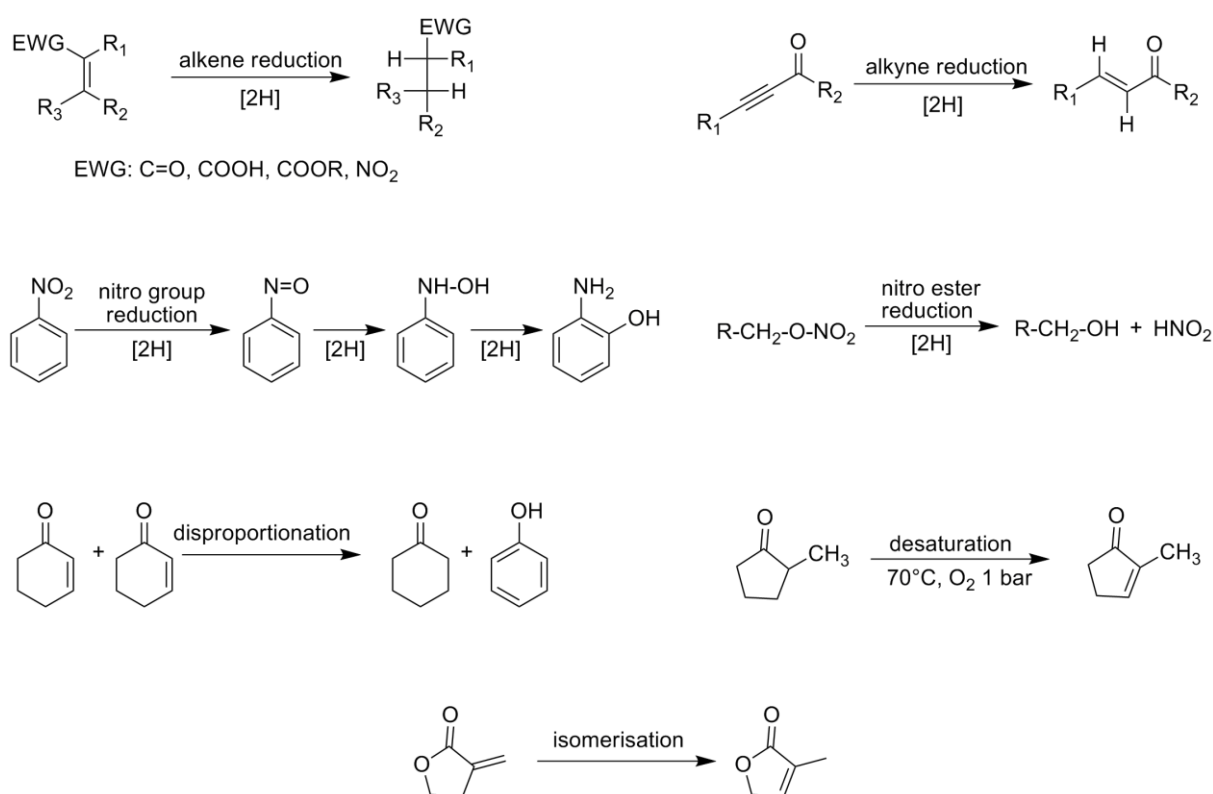


Figure 1.9: Selection of reactions catalyzed by ene reductases. For a general overview of promiscuous reactions catalyzed by ene reductases see Faber *et al.* 2013⁶⁸.

containing explosives like nitro aromatic and nitro ester compounds as glycerol trinitrate (GTN) and pentaerythritol tetranitrate (PETN).^{31,63,71,72} The OYE homologous PETNR, as well as the other described nitro compounds reducing ene reductases, catalyze the stereochemical liberation of one or two equivalents of nitrite depending on the substrate. Moreover, a thermophilic ene reductase, GkOYE from *Geobacillus kaustophilus*, was described being able to convert 2-methyl-cyclopentanone to 2-methyl-cyclopentenone at high reaction temperature and under oxygen atmosphere, representing the reverse reaction, a desaturation of alkanes.⁴⁰ Till today this is the only family member known being capable of catalyzing the unfavored back reaction. Recently, a promiscuous reaction for Old Yellow Enzymes, the isomerization of γ -butyrolactones was reported.⁷³ This reaction can be catalyzed by the yeast enzymes OYE1 and OYE2, as well as the estrogen-binding protein EBP1 from *Candida albicans*.^{74,75} In addition to the already from nature conferred properties it is often necessary to alter an enzyme for application as biocatalyst in industrial processes with regard to substrate acceptance and specificity, selectivity, as well as thermo- and solvent stability. As a consequence of this necessity an entire series of

molecular biological techniques was developed to engineer the properties of enzymes.

1.4 Enzyme engineering of biocatalysts

Enzymes are remarkable, energy-efficient and environmentally friendly molecular machineries performing their natural reaction often with extreme high selectivity and without any byproduct formation which represent industrial desired properties. However, one should keep in mind that enzymes naturally are optimized to guarantee the survival and reproduction of their originating organism and not, to catalyze industrial applied reactions. This feature of enzymes is just a side product, so to say a fortunate coincidence. Therefore, it is not surprising that naturally occurring enzymes applied as catalysts often lacking necessary features for commercial applications. The ideal industrial biocatalyst should combine high specific activity with high specificity towards the applied substrate and, if desired high enantioselectivity, as well as a high and long stability under industrial process conditions like high reaction temperatures and the presence of organic solvents.⁷⁶ Furthermore, an ideal biocatalyst should not be inhibited by its own product resulting in merely low turnover numbers. The identification and further molecular biological engineering of industrial applicable biocatalysts can be divided into three waves.⁷⁷ For a long time, during the first wave until the 1980s, the identification of new biocatalysts was limited to natural biodiversity by work-intensive screening of microbial cultures for the desired activity, and stability limitation could only be overcome by immobilization of the enzyme. An example for an applied biocatalyst of the first wave of biocatalyst engineering is an immobilized glucose isomerase for the industrial scale production of the sweeter-tasting fructose. Later on, in the 1980s and 1990s, while the second wave initial protein engineering technologies were developed and applied in order to extend the substrate range or stability of the enzyme. With the availability of mainly structure-based rational enzyme design methods two central questions need to be answered, in particular “what would be the ideal protein template to start for concerning a specific question” and “what tools from the toolbox of protein engineering technologies should be applied”.⁷⁶ The basis of the rational design methods generated during the second wave was the resolution and understanding of enzyme structures and reaction mechanisms, which is no more than the connection of protein sequence and function.⁷⁸ Down to the present day the number of protein sequences

and structures stored in databases are still considerably increasing with, for example, over 77.000 stored protein crystal structures in the RCSB Protein Data Bank (<http://www.pdb.org>).⁷⁷ The understanding of the connection between enzyme structure and function is absolutely necessary for the identification of the active site with the catalytic important amino acids which then could be addressed by developed mutagenesis methods like the QuikChange site-directed mutagenesis.⁷⁹ With rational site-directed mutagenesis based on structural, as well as sequential information it was possible to turn a desaturase into a hydroxylase by just four amino acid substitutions.⁸⁰⁻⁸² The third, and by today last wave of biocatalyst engineering was started in the late 1990s by Frances H. Arnold⁸³⁻⁸⁶ and Willem P. C. Stemmer⁸⁷⁻⁹⁰ with the development of “evolutionary” protein design methods inspired by the Darwinian evolution through random mutagenesis, natural DNA recombination and an external selection pressure. All techniques, which mimic the natural evolution in the laboratory, are summed up under the term “directed evolution”. Directed evolution techniques were applied to improve several enzyme properties as catalytic activity⁸⁷, enantioselectivity⁹¹ and thermostability.⁹² Several techniques of the second and third wave of biocatalyst engineering are highlighted below.

1.4.1 Rational enzyme design

Rational enzyme design was the first tool for the fine-tuning of enzyme properties which was developed during the second wave of biocatalyst engineering between the 1980s and 1990s. It is a quite information-intensive method due to the fact that it requires information about the protein structure, as well as the knowledge about the relationship between protein sequence, structure and function. Usually, single amino acid substitutions, likely involved in particular enzyme properties, are selected by sequence comparisons of related enzymes and inserted via site-directed mutagenesis in order to alter the properties by just a few “logical” variations.⁹³ Therefore, on great advantage of this method is minimal number of trials and errors of enzyme variant generation in a reasonable period of time.⁹⁴ The design of new biocatalysts with rational design is generally limited to the substrate specificity, the catalytic mechanism and the reaction range of the compared related parental genes. One of the main goals of protein engineering is, however, the development of new enzyme features which are not present in the natural sequences. Therefore,

nowadays rational enzyme design can apply a combination of sequence-based protein engineering with structural, as well as mechanistic information and additional computational *in silico* tools like molecular modeling or Rosetta. With molecular modeling it is now possible to predict how to increase the selectivity, activity, as well as stability of an enzyme *in silico*, even without a solved crystal structure, just based on a structural homology model. In addition, with programs like Rosetta it is possible to design within an existing template structure a new, artificial active site around the transition state of the desired reaction in order to create a novel biocatalyst. So for example, the development of a non-natural occurring kemp eliminase was achieved by transition-state stabilization applying RosettaMatch.⁹⁵ Nonetheless of all successful engineered biocatalysts obtained by rational design this method exhibits several considerable limitations. Firstly it is restricted to a quiet small pool of well-studied enzymes concerning structure and catalytic mechanism as unfortunately these informations are unavailable for the large majority of enzymes. And secondly, rational design is limited to alterations, which appear striking by looking and comparing the amino acid sequence, as well as the protein structure of related enzymes. However, many important properties are not present in the small number of conserved amino acids, but consist of contributions from many residues distributed over large parts of the protein, for example stability properties. Therefore, in the third wave of biocatalytic protein engineering a new approach was developed called directed evolution.

1.4.2 Directed, laboratory evolution of enzymes

Directed evolution is a collective term combining several *in vitro* molecular biological methods mimicking natural Darwinian evolution in the laboratory with a necessary subsequent high-throughput screening system. The basis of directed evolution is the performance of several cycles of random gene mutagenesis creating libraries of mutated genes, followed by enzyme expression in a suitable host organism and a high-throughput screening or selection system for the desired property until the required degree of improvement is achieved. Iterative rounds of random mutagenesis analyzed with the same screening system result in the accumulation of beneficial mutations distributed over the entire enzyme.^{83–86} However, one should keep in mind that with this engineering approach a large number of enzyme variants are created

which need to be screened with a suitable screening system for the desired property. So, for example, if one amino acid is substituted randomly in a 300 residue enzyme 5700 variants are possible, but the number of variants increases to 16 million when just two amino acids are simultaneously substituted.⁹⁶ Therefore, the bottleneck of this protein engineering technique is the development of a simple and fast high-throughput screening method for the desired property. The first enzyme property which was addressed by directed evolution was the thermo-, as well as the solvent stability. The most simple and commonly applied method for random gene mutagenesis is the error-prone polymerase chain reaction (epPCR). Most of the 19 possible amino acid substitutions at a specific position within a protein are deleterious and decrease the enzyme activity, just a few amino acid exchanges result in the desired property change. Therefore, low mutation rates of only a few amino acid substitutions per gene are favored. The mutation rate of the applied polymerase in the epPCR can be controlled by the inserted $MgCl_2$ and nucleotide concentration. The epPCR technique is most successful in the improvement of enzyme stability and activity due to the fact that these properties are often increased by additive mutations which can be found in constitutive mutagenesis rounds. It was possible to increase considerably the robustness of the protease subtilisin E towards the enzyme damaging solvent dimethylformamide by several rounds of epPCR and a suitable screening system.⁹⁷ Next to the quite simple methods of epPCR, which mimics the occurrence of natural random point mutations, a second, completely different and more challenging method, the DNA shuffling, was developed imitating the natural gene recombination.^{87–90} For DNA shuffling one single parental gene or, also a number of homologous parental genes, can be used as template for the fragmentation with *DNase I*. After the *DNase I* digestion DNA fragments possessing the desired length (for example 100 – 200 bp) are isolated and subsequently reassembled to the initial gene length. The reassembly reaction is a self-priming PCR reaction without any additional primers. The fragments are reassembled during the normal, repetitive PCR cycles of denaturation, annealing and elongation in a combinatorial manner in areas of sequence homology. This process possesses an additional mutation rate of seven point mutations per kilobase pair.⁸⁹ Finally, the obtained reassembled gene products are amplified in a PCR reaction with appropriate designed terminal primers (figure 1.10). For the first time, the DNA shuffling of homologous genes, also called family DNA shuffling, was applied at four

related, bacterial cephalosporinase genes in order to obtain moxalactamase activity.⁸⁷ When DNA shuffling was applied to just one of the selected four cephalosporinase genes, one single cycle of shuffling yielded in an eightfold improvement of the moxalactamase activity.

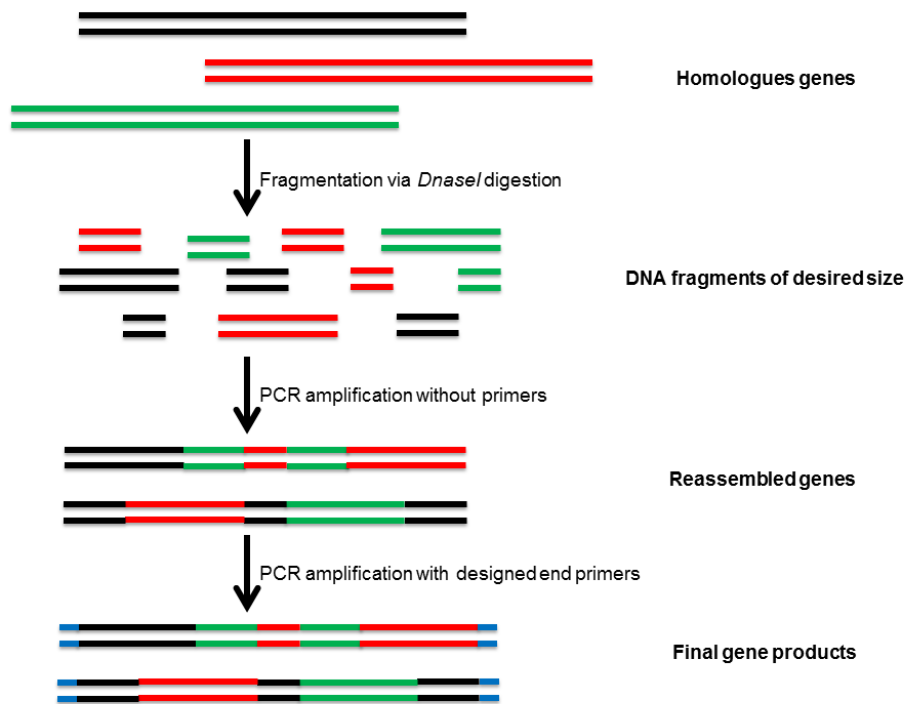


Figure 1.10: Schematic representation of the DNA shuffling of homologous genes developed by Stemmer *et al.*⁸⁷ consisting of four steps: (1) selection of the homologous genes, (2) *DNase I* digestion, (3) primerless reassembly of the gene fragments, (4) amplification of the achieved reassembly products via PCR reaction

If, however, all four genes were included in the shuffling approach, the best obtained variant demonstrated a 270- to 540-fold improvement of the moxalactamase activity compared to the four wild type enzymes. The best mutant possessed eight fragments from three of the four parental homologous genes and 33 additional point mutations. The power of the family DNA shuffling can be explained with considerable larger differences in the overall amino acid sequence of the obtained variant compared to the DNA shuffling of just one gene, which results in only a few amino acids substitutions.

1.4.3 Fusion methods of rational design and directed evolution

However, nowadays the distinction between rational design and directed evolution techniques becomes more and more blurred. Most of the new developed methods are consisting of a combination of rational design and directed evolution, which can be called semi-rational methods, due to the fact that more and more sequences as well as structural data are accessible. A good example for this trend represent the two methods CASTing^{96,98} and B-Fit^{92,99} developed by Manfred T. Reetz, which combine saturation mutagenesis with structural knowledge. CASTing, which stands for Combinatorial Active-site Saturation Test, is a simple, systematic and knowledge-driven saturation mutagenesis of all relevant amino acid positions within the binding pocket of an enzyme. For selection of such relevant positions, structural as well as mechanistic information is necessary, representing the rational part of this method. In a model study 2005 Reetz and coworkers were able to expand the substrate scope of the lipase from *Pseudomonas aeruginosa* (PAL) by CASTing at five defined randomization site towards bulky ester which are not accepted by the wild type enzyme.⁹⁸ Thus, it should be mentioned that a randomization site could consist of more than one amino acid in order to allow cooperative effects.⁹⁶ Next to the substrate scope it was also shown that the enantioselectivity of an enzyme was successfully altered by CASTing.^{100–102} However, if thermostability of an enzyme is the desired property for improvement the CASTing method is useless, due to the fact that stability is not conferred by amino acids located in the binding pocket. Therefore, a second method, called B-Fit method, was developed which utilizes as randomization site the amino acid residues demonstrating the highest B-factors. The B-factor is a value which indicates high thermal motion and therefore stands for the flexibility of amino acids within the protein. A high structural flexibility of an enzyme goes in hand with a high thermo-, as well as solvent sensitivity. Thus, the saturation mutagenesis at positions with high B-factors should result in more rigidity and thus higher stabilization of the enzyme. With the B-Fit method it was possible to increase the T_{50}^{60} (temperature at which after 60 min of heat treatment 50 % enzyme activity was still present) value of lipase A from *Bacillus subtilis* by 45°C applying iterative saturation mutagenesis at the ten amino acids demonstrating the highest B-factor values.⁹² Beside the combination of saturation mutagenesis and structural data leading to new semi-rational protein engineering methods it is also possible to

combine the DNA shuffling technique with rational design methods. In 2009 Carola Engler and coworkers published the so-called Golden Gate Shuffling, which is a simple, robust and efficient semi-rational high-throughput DNA shuffling technique allowing the directional ligation of up to nine different gene fragments in one pot.¹⁰³ This cloning method is based on a kind of site-specific recombination by type IIS restriction enzyme¹⁰⁴ digestion and subsequent ligation. DNA restriction enzymes, also called endonucleases, can be divided into four types according to the following characteristics: relation of recognition to cleavage site, requirement of ATP hydrolysis for nucleolytic activity, methyltransferase activity as well as methylation state of the DNA sequence.¹⁰⁵ Due to their importance for recombinant DNA shuffling methods the best investigated and characterized type, with over 3500 enzymes, are the type II endonucleases. All type II restriction enzymes have in common that they recognize a short, usually palindromic nucleotide sequence of four to eight bp and cleave in the presence of Mg^{2+} cofactor the DNA within, or in close proximity to the recognition site. None of them requires ATP hydrolysis for their activity or possesses a methylation activity.¹⁰⁶ Type II restriction enzymes can be divided into eleven subtypes¹⁰⁵, with type IIS being the most important one for the Golden Gate Shuffling. Type IIS endonucleases possess an asymmetric recognition site with the DNA cleavage occurring at a defined distance to that recognition site enabling the seamless ligation of two digested compatible DNA fragments. The Golden Gate Shuffling consists of three main steps:

(1) *Selection of homologous genes and type IIS restriction enzyme:* As a first step the genes acting as shuffling partners within the Golden Gate Shuffling need to be selected based on desired properties. Furthermore the type IIS endonuclease which will be used has to be specified with regard to the obtained recombination/cleavage site. For example *SapI* and *LglI* possess a seven base pair long recognition sequence and a three nucleotide cleavage site, while *BsaI* and *Esp3I* have a six base pair long recognition sequence and a four nucleotide cleavage site.¹⁰⁷ The most common used type IIS restriction enzyme for the Golden Gate Shuffling is *BsaI* with its recognition sequence being GAGACC (figure 1.11) and a four nucleotide overhang allowing 256 different overhangs.¹⁰⁴

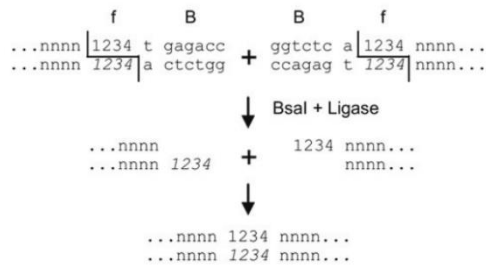


Figure 1.11: Schematic representation of the *Bsal* recognition site GAGACC, the created four nucleotide overhang after digestion, and the ligation. Graphic adapted from Engler *et al.*¹⁰³

(2) *Selection of gene fragments and the recombination sites:* Based on a multiple sequence alignment, as well as a structural superposition of the chosen shuffling partners for the Golden Gate Shuffling the desired amount of fragments is determined and the genes accordingly divided. A gene fragment consists of a core sequence, which is variable among the homologous and is flanked by a four nucleotide recombination site with the adjacent *Bsal* recognition site at each end. A recombination site is located where the *Bsal* digestion and subsequent ligation will take place. It needs to fulfill the following criteria to ensure directional ligation of the fragments. A recombination site needs to be conserved between all chosen parental sequences at the same fragment and differ from all other selected recombination sites within the same gene to avoid illegitimate recombination. In addition it is beneficial to avoid a match for three of the four consecutive nucleotides within two different sites to ensure proper ligation.

(3) *Selection of cloning and expression vectors:* In a third step, it is necessary to choose the appropriate vectors for the Golden Gate Shuffling. Two different vectors are required: (1) a cloning vector for the insertion of each single fragment possessing an antibiotic resistance and a lac Z alpha gene including the multiple cloning site for the blue/white screening (2) an expression vector with a different antibiotic resistance, a promoter region for the expression of the religated gene, as well as two *Bsal* restriction sites with recombination sites compatible with the beginning of the first and the end of the last fragment set. After the cloning of each desired fragment in a cloning vector and the insertion of the two *Bsal* restriction sites in the selected expression vector, all vectors are mixed together in one pot and the digestion ligation reaction can take place.

1.5 Aim of the work

In the present thesis, the TIM barrel based Old Yellow Enzyme NCR (2-cyclohexen-1-one reductase) from the gram negative bacterium *Zymomonas mobilis* was chosen as studying object for the investigation of the influence, the presence of cooperative effects and the role of β/α surface loop regions on enzyme properties like reduction activity, substrate selectivity, substrate specificity, enantioselectivity, thermo- and solvent stability as well as cofactor interaction. Additionally, a systematical phylogenetic analysis of the Old Yellow Enzyme family was performed in order to determine the relationship of several β/α surface loop regions within the reductase family. For these purposes, the subsequent tasks were established:

- Generation of semi-rational designed loop variants by Golden Gate Shuffling of several Old Yellow Enzyme family members selected based on biochemical and structural literature data followed by the development of a suitable screening assay.
- Generation of rational designed loop variants of NCR by two different approaches: (1) loop length variation of intrinsic NCR loops, (2) loop grafting between NCR and other Old Yellow Enzyme family members, which were also selected based on biochemical and structural literature data.
- Phylogenetic analysis of the Old Yellow Enzyme family based on overall sequence identity, as well as the generation and distribution of several loop motifs within the entire enzyme family.
- Biochemical investigation of the influence of the designed loop variants on reduction activity, substrate selectivity, substrate specificity, enantioselectivity, thermo- and solvent stability as well as cofactor interaction.

2 Experimental section

2.1 Genes, proteins, vectors, primers and strains

Protein sequences, vectors, vector constructs, designed primers and used strains are detailed in supplementary material section 7.1.

2.2 Methods

2.2.1 General molecular biological methods

Materials and methods for general molecular biological methods like overlapping extension PCR, restriction enzyme digest, SDS gel electrophoresis, ligation, preparation of RbCl competent cells, heat shock transformation and protein concentration determination after Bradford are described elsewhere.^{9,108,109}

2.2.2 Golden Gate Shuffling

2.2.2.1 Selection of vectors, enzymes and shuffling fragments

Vector selection: The technique of the Golden Gate Shuffling¹⁰³ depends on the ability of type IIS restriction enzymes like *BsaI* (New England Biolabs, Ipswich, MA, USA) cutting outside of the recognition site¹⁰⁴ enabling a ligation of the gene without any intrinsic restriction site. Required for this type of shuffling is a suitable cloning, as well as an expression vector. As a cloning vector the pUC19 plasmid (Carl Roth, Karlsruhe, Germany) was selected (figure 7.1). As expression vector for the Golden Gate Shuffling a modified pET28a(+) vector (Novagen, Wisconsin, USA) was used (pET28a(+)_NCR_*BsaI*). It already included NCR wild type gene with two additionally *BsaI* restriction sites (figure 7.2). The two *BsaI* restriction sites were introduced by overlapping extension PCR in the NCR gene (table 7.2).

Enzyme selection: Based on a comparison of published biochemical as well as structural data on the ene reductases, four members of the Old Yellow Enzyme family were selected as shuffling partner for NCR from *Zymomonas mobilis*. Next to NCR, two of the selected shuffling partners, OYE1 from *Saccharomyces pastorianus* and PETNR from *Enterobacter cloacae*, belong to the subclass of the *classical* OYE, whereas the other two, YqjM from *Bacillus subtilis* and GkOYE from *Geobacillus kaustophilus*, are part of the *thermophilic-like* subclass.

Fragment selection: Bases on a structural and multiple sequence alignment (figure 7.3) the five enzymes were divided into seven fragments (figure 7.4). A main criterion for this shuffling technique is that two consecutive fragments are connected by four shared nucleotides (table 7.5). Additionally each set of the complimentary fragments of the five selected ene reductases needs to end with the same four nucleotides. For the preservation of the order of the fragments it is necessary that each of the four nucleotide linkers is only once present within one gene. In the Golden Gate Shuffling the N- and C-terminal fragments, as well as the fragment possessing the catalytic important tyrosine (F4 amino acid 164-226), are part of the NCR scaffold. The other four fragments (F2, F3, F5 and F6) will be shuffled among the five enzymes. With a total of 21 fragments the Golden Gate Shuffling leads to 1024 different enzyme variants.

2.2.2.2 Cloning of the selected fragments

Amplification of the fragments: The selected 21 fragments of the five different ene reductases were amplified with the appropriate primers (table 7.2) by the use of the KOD Hot Start Polymerase (Merck Millipore, Darmstadt Germany) according to user's manual. Afterwards the PCR products were purified with the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, Irvine, CA, USA) and the DNA concentration was determined with Nanodrop ND 1000 Spectrophotometer.

Blunt end digestion: The cloning vector pUC19 was digested with *SmaI* FastDigest restriction enzyme (Fermentas, Schwerte, Germany) at 37°C for 1 h and afterwards inactivated at 65°C for 5 min followed by a dephosphorylating step with FastDigest Alkine Phosphatase (Fermentas, Schwerte, Germany) at 37°C for 10 min. The digested vector was purified with the Zymoclean™ Gel DNA Recovery Kit (Zymo

Research, Irvine, CA, USA) and the DNA concentration was determined with a Nanodrop ND 1000 spectrophotometer.

Sticky end digestion: Double digestion of the appropriate PCR fragments, as well as the cloning vector pUC19, was performed with the two restriction enzymes *HindIII* and *XbaI* (Fermentas, Schwerte, Germany) at 37°C for 3 h. Purification of the digested fragments and the cloning vector was carried out with the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, Irvine, CA, USA). Determination of the DNA concentration was done with a Nanodrop ND 1000 spectrophotometer.

Ligation: The ratio of digested vector to PCR fragment product for the ligation with T4 ligase (Promega, Madison, WI, USA) was 1:7 according to user's protocol. The ligation was performed at RT for 4 h, afterwards transformed in *E. coli* XL-1 blue cells, plated on LB-Amp plates containing 40 µg/ml X-Gal and incubated overnight at 37°C.

Blue/white screening: After the overnight incubation of the LB-Amp-X-Gal plates the white colonies were picked and used to inoculate a 5 ml LB culture, which was cultivated at 37°C and 180 rpm overnight. The plasmid was isolated according to the manufacturer's protocol with the Zyppy Plasmid MiniPrep Kit (Zymo Research, Irvine, CA, USA). The isolated plasmids with the desired inserted fragments were verified by sequencing (GATC Biotech, Köln, Germany) and transformed in *E. coli* DH5α cells for amplification.

2.2.2.3 One pot digestion-ligation reaction of the Golden Gate Shuffling

For the digestion-ligation reaction 100 ng of each fragment set and 100 ng of the constructed expression vector pET28a(+)_NCR_Bsal were used as templates. The fragment sets for F2, F3, F5 and F6 consist of each of the corresponding fragments from NCR, OYE1, PETNR, YqjM and GkOYE in equal amounts (25 ng). Fragment set F4 contained just 100 ng of the fragment from NCR. For the reaction 20 U *Bsal* (New England Biolabs, Ipswich, MA, USA) and 9 U T4 DNA ligase (Promega, Madison, WI, USA) were used with the T4 ligation buffer in a total volume of 20 µl. Table 2.1 shows the temperature program for the digestion- ligation reaction.

Table 2.1: Temperature program for the digestion-ligation reaction of the Golden Gate Shuffling

Step	Reaction temperature	Reaction time	Repeats
1	digestion	37°C	} 70 x
2	ligation	16°C	
3	final digestion	50°C	1 x
4	heat inactivation	80°C	1 x

After the reaction the digestion-ligation mixture was transformed by heat shock transformation in RbCl competent *E. coli* BL21 (*DE3*) cells, plated on LB-Kan plates and incubated overnight at 37°C. For the verification of the Golden Gate Shuffling several on plate grown colonies were picked to inoculate a LB culture consisting of 5 ml media with kanamycin and incubated overnight at 37°C and 180 rpm. After plasmid isolation of the cultivated LB-culture with the Zyppy Plasmid MiniPrep Kit (Zymo Research, Irvine, CA, USA) according to user's manual the performance of the Golden Gate Shuffling was verified by DNA sequencing (GATC-Biotech, Köln, Germany) of in total 28 created variants.¹¹⁰

2.2.2.4 Developed NADH assay for the activity screening of the Golden Gate Shuffling variants

In total 1100 colonies were screened with the subsequent developed NADH dependent spectro-photometric assay. Therefore, a three step screening assay was developed:(i) cultivation of the Golden Gate Shuffling (GGS) variants (ii) cell disruption and (iii) 96-well NADH assay.¹¹⁰

Cultivation of the Golden Gate Shuffling variants: The cultivation of all GGS variants was done in 15 ml falcon tubes (Greiner Bio-one, Frickenhausen, Germany). A 4 ml LB-Kan preculture was inoculated from plate and incubated at 37°C and 180 rpm overnight. Additionally, two negative and one positive control for the following NADH assay were cultivated. As the two negative controls *E. coli* BL21 (*DE3*) was used without any plasmid (1) plus the NCR variant Y177A (2), in which the catalytic active tyrosine is exchanged against an alanine. As positive control *E. coli* BL21 (*DE3*) was used containing the wild type NCR. With the obtained overnight preculture the main

culture consisting of 4 ml TB-Kan media was inoculated (with 50 μ l preculture) and incubated at 37°C and 180 rpm for 3 h. Afterwards the protein expression was induced with 1 mM IPTG and cultivated at 30°C and 180 rpm overnight.

Cell disruption: After the overnight cultivation the main culture was harvested by centrifugation (for 15 min with 4.000 x g at 4°C), the supernatant was discarded, the obtained pellet was resuspended in 800 μ l BugBuster protein extraction reagent (Merck Millipore, Darmstadt, Germany) and transferred in 1.5 ml Eppendorf tubes (Eppendorf, Hamburg, Germany). The suspension was incubated at room temperature for 20 min at 300 rpm and afterwards centrifuged for 20 min with 16.000 x g at 4°C. The supernatant was transferred in another Eppendorf tube and used for the subsequent NADH assay.

96-well NADH assay: For the 96-well plate activity assay the α,β -unsaturated ketoisophorone was chosen as model substrate. The assay was performed in a total volume of 100 μ l, containing 94 μ l protein lysate, 1 mM ketoisophorone and 2 mM NADH. The reaction was performed at 30°C, shaken in-between each measurement, and the NADH consumption was measured at 340 nm for 25 min (a measurement each 30 sec). Catalytic active variants were sequenced (GATC-Biotech, Köln, Germany) and applied for biotransformation reactions.

2.2.2.5 Expression and purification of the Golden Gate Shuffling variants for biotransformation

In total five GGS variants (GGS 10, GGS 174, GGS 222, GGS 225 and GGS 229) were chosen for biotransformation reactions with standard α,β -unsaturated substrates. Each of them was expressed in *E. coli* BL21(DE3) in 2 L TB media containing 50 μ g/ml kanamycin and the enzyme expression was induced at an OD₆₀₀ of 0.5 – 0.6 with 0.2 mM IPTG at 30°C and 180 rpm for 20 h. After harvesting by centrifugation (30 min, 9.000 x g, 4°C), the cells were resuspended in 5 ml 50 mM Tris-HCl pH 7.5, disrupted via sonification (6 x 1 min with 30 sec break) and centrifuged (30 min, 21.000 x g, 4°C). Then the lysate was filtered through a 0.45 μ m filter and purified via Ni-based immobilized metal affinity chromatography by using 1 ml His GraviTrap TALON columns (GE Healthcare, Freiburg, Germany). As loading buffer 50 mM Tris-HCl was used and for a washing step the same buffer containing

10 mM imidazole was applied. The bound GGS variant was eluted with a 50 mM Tris-HCl pH 7.5 buffer containing 50 mM imidazole. The purification was followed by a filtration and desalting step carried out with Vivaspin ultrafiltration spin column (Vivaspin 10 kDa, Sartorius, Göttingen, Germany) and concentrated in Tris-HCl reaction buffer (50 mM, pH 7.5). The purity was controlled by SDS-PAGE and the protein concentration was determined with the Bradford Coomassie Protein Assay Kit (Thermo Scientific, Schwerte, Germany) according to user's protocol. The purified enzyme was stored at -20°C until further use.

2.2.2.6 Biotransformation with Golden Gate Shuffling variants

In total six different standard α,β -unsaturated substrates (ketoisophorone, 2,4-heptandienal, neral, geranial, cinnamyl alcohol, and 3-methyl-cyclohexenone) were selected to investigate the substrate scope of the GGS variants. The biotransformation reactions were performed in triplicates with the following conditions: 50 μ g purified protein (purity > 70 %), 2 mM substrate, 2 mM NADH, filled up to 1 ml 50 mM Tris-HCl pH 7.5 at 30°C with 180 rpm for 24 h. The reactions were extracted with 2 x 500 μ l MTBE and analyzed with GC-FID with the standard programs for cyclic and aliphatic substrates (table 7.6).

2.2.3 Rational loop analysis

2.2.3.1 Generation of loop profiles

For the two selected loop regions Loop A and Loop B of the three reference sequences NCR from *Zymomonas mobilis*, OYE1 from *Saccharomyces pastorianus* and MR from *Pseudomonas putida* M10, Hidden Markov (HMM) profiles were created using HMMER.¹¹¹ In the three reference sequences the two loop regions were manually identified based on structural information. The first amino acid of the loop is the last amino acid of the leading β -sheet and the final loop position is the last amino acid of the unstructured region before the following α -helix. For the generation of the HMM profiles a three step algorithm was developed (figure. 2.1). Based on one

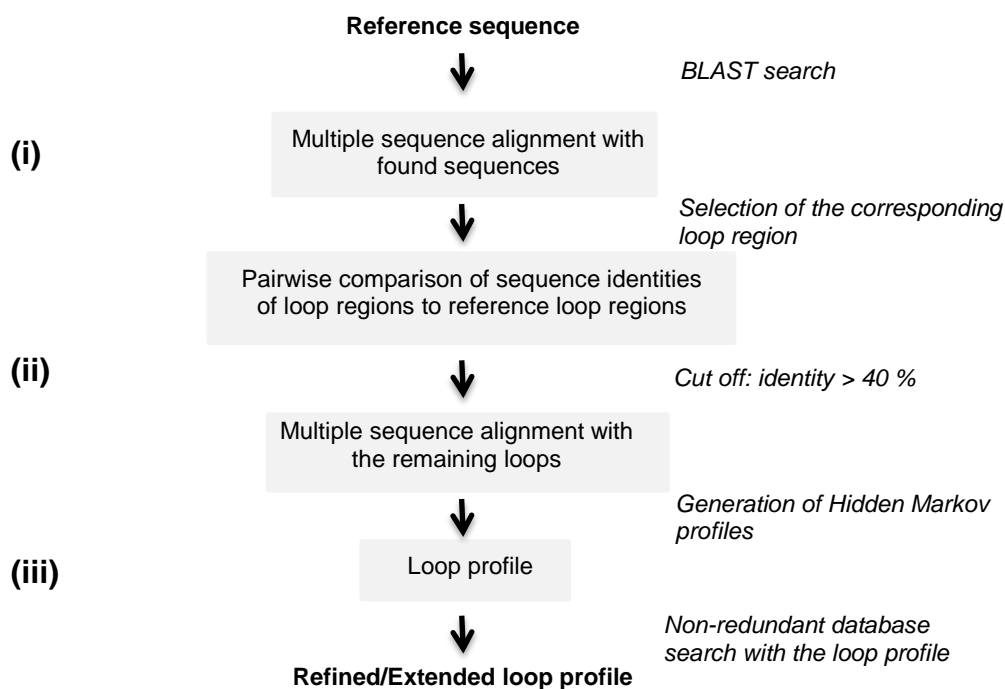


Figure 2.1: Three step algorithm for the construction of loop HMM profiles based on a reference sequence.

of the three reference sequences NCR, OYE1 and MR, a local BLAST¹¹² search using standard parameters was performed against a collection of 4483 ene reductases collected from Genebank.¹¹³ With the first 500 hits a multiple sequence alignment was generated using ClustalW.¹¹⁴ Based on the before defined loop region (A or B) in the reference sequence the corresponding loop regions in the aligned sequences were extracted. Each obtained loop region of each sequence was analyzed and all loops possessing more than 30 % gaps were discarded. With the remaining loop regions a pairwise sequence alignment using Needleman-Wunsch with standard parameters was performed against the reference sequence¹¹⁵. For the subsequent steps just the sequences with an identity > 40 % in the loop region were picked out to generate of a first HMM profile. For the increase of the number of sequences used for the building of the HMM profiles a profile search against the 4483 ene reductase collection was performed. All received sequence hits were used to generate the final HMM loop profiles.

2.2.3.2 Generation of a phylogenetic tree

Phylogenetic tree of Old Yellow Enzyme superfamily: The superfamily of the Old Yellow Enzymes consists of 4483 different proteins which can be found in the non-redundant sequence database at NCBI. For the phylogenetic analysis of the entire superfamily a multiple sequence alignment using ClustalW¹¹⁴ was created for all 4483 sequences. Based on that alignment a phylogenetic tree for the family of ene reductases was created using the PHYLIP package.¹¹⁶

Phylogenetic tree based on NCR, OYE1 and MR: Based on the three reference sequences NCR, OYE1 and MR used for the creation of the HMM profiles, a phylogenetic tree for the homologous family H1 of the ene reductase superfamily was created. A BLAST search against the non-redundant sequence database at NCBI was carried out for each reference sequence of NCR, MR and OYE1 with an E-value cut off of 10^{-5} , resulting in a total of 2558 unique sequences representing the sequence space for the phylogenetic analysis. A multiple sequence alignment using ClustalW¹¹⁴ was created for the 2558 sequences and a phylogenetic tree for this subset of the ene reductase family was created using also the PHYLIP package.¹¹⁶

2.2.3.3 Alanine scanning mutagenesis of the Loop A region

Alanine scanning mutagenesis (ASM) was performed as site-directed mutagenesis according the QuikChange standard protocol. Therefore the plasmid pET28a(+) including the N-terminal His₆-tagged NCR wild type enzyme was mutated. In total 16 amino acids (L225–E240) were exchanged against alanine using the oligonucleotides listed in supplementary material tab. 7.4. The PCR mixture was digested with *DpnI* at 37°C for 3 h and afterwards transformed in competent *E. coli* DH5α cells. After plasmid isolation with the Zyppy Plasmid MiniPrep Kit (Zymo Research, Irvine, CA, USA) according to user's manual the desired mutation was verified by DNA sequencing (GATC-Biotech, Köln, Germany). Finally the plasmid was transformed in *E. coli* BL21 (*DE3*) and the protein was expressed in TB media with standard conditions.⁹ The variants were purified using a Co-based Talon Superflow 10 ml column (GE Healthcare, Freiburg, Germany) and an imidazole concentration of 50 mM in the 50 mM Tris-HCl pH 7.5 buffer. The fractions were

desalted via ultrafiltration with vivaspin columns (6 ml, 10 kDa, PES membrane, Sartorius, Göttingen, Germany); the purity was controlled by SDS-PAGE and the protein concentration was determined with the Coomassie (Bradford) Protein Assay Kit (Thermo Scientific, Schwerte, Germany) according to user's protocol. The purified enzymes were stored at -20°C until use. Biotransformation reactions with the ASM variants were performed with the following conditions: 100 µg/ml purified enzyme (protein purity > 90 %), 10 mM ketoisophorone as substrate, 10 mM NADH ad 1 ml 50 mM Tris-HCl pH 7.5 at 30°C and 180 rpm for 2.5 h. The reactions were performed at least in triplicates and were terminated by the extraction with MTBE. The resulting samples were analyzed by GC-FID with the standard program for cyclic substrates (table 7.6).

2.2.3.4 Determination of the crystal structure of variant Loop A_OYE1

The crystallization was performed by cooperation partners with the procedure described in supplementary material section 7.3.2.

2.2.4 Expression and purification of wild type enzymes and enzyme variants

2.2.4.1 Standard expression and purification wild types and loop variants

NCR and MR wild type enzymes, as well as all designed loop variants were cloned in pET28a(+), expressed in *E. coli* BL21 (DE3) in 400 ml TB media containing 50 µg/ml kanamycin and the enzyme expression was induced at an OD₆₀₀ of 0.5 – 0.6 with 0.2 mM IPTG at 30°C for 20 h. After harvesting by centrifugation (30 min , 9.000 x g, 4°C), the cells were disrupted via a French press (EmulsiFlex-C5, Avestin, Mannheim, Germany) and centrifuged (30 min, 21.000 x g, 4°C). Then the obtained lysate was filtered through a 0.45 µm filter and purified with a continuous imidazole gradient (up to 300 mM) via Ni-based immobilized metal affinity chromatography by using 5 ml HisTrap HP column (GE Healthcare, Freiburg, Germany) with an ÄKTA system. The purification was followed by a filtration and desalting step carried out with Vivaspin ultrafiltration spin column ns (Vivaspin 10 kDa, Sartorius, Göttingen,

Germany) and concentrated in Tris-HCl reaction buffer (50 mM, pH 7.5). The purified enzyme was stored at -20°C until further use.

2.2.4.2 Expression and purification of OYE1 wild type enzyme

OYE1 wild type was cloned in a pDHE vector and introduced into *E. coli* BL21 (DE3). The resulting strain was grown at 37°C and 180 rpm in 400 ml TB medium containing 100 µg/ml ampicillin. After reaching an OD₆₀₀ of 1.3 – 1.6, the culture was supplemented with 0.2 % l-rhamnose. After 12 – 14 h of incubation at 30°C and 160 rpm, the cells were harvested by centrifugation (15 min, 8.000 x g, 4°C) and resuspended in 50 mM potassium phosphate buffer pH 7.4 buffer containing 0.1 mM PMSF. Cell pellets were disrupted in 2-3 cycles on a French press (EmulsiFlex-C5, Avestin, Mannheim, Germany) at 4°C. The resulting crude extract was centrifuged (30 min, 37.000 x g, 4°C) and the supernatant with the soluble proteins was recovered. The protein purification was performed in three steps: (i) an ammonium sulfate precipitation, (ii) a FPLC chromatography and (iii) a HIC chromatography.

Ammonium sulfate precipitation: Protein precipitation of the lysate was performed with 24 % ammonium sulfate (end concentration). Subsequently the lysate was centrifuged for 15 min with 7.000 x g at 4°C and protein pellet was discarded. Then the ammonium sulfate concentration was increased to an end concentration of 45 % followed by another centrifugation step (15 min, 7.000 x g, 4°C) and the supernatant was discarded. Afterwards the proteins were resuspended in 50 mM potassium phosphate buffer pH 7.8.

FPLC chromatography: The second part of protein purification was carried out by FPLC using a column with Q-sepharose FF (GE Healthcare, Freiburg, Germany) packed to a volume of 275 ml and a maximum flow of 20 ml/min. The column was washed (10 ml/min working flow) with a step gradient protocol with 50 mM potassium phosphate buffer pH 7.8 containing 0 – 1.4 M ammonium sulfate. The elution of the OYE1 protein occurred at 400 mM ammonium sulfate. In addition to the characteristic total protein detection at 280 nm, OYE1 wild type was identified by its absorbance at 455 nm.

HIC chromatography: The last purification step was a hydrophobic interaction chromatography performed by using a phenyl sepharose HP (GE Healthcare, Freiburg, Germany) as material. The column was packed with a maximum flow of 18 ml/min to a volume of 240 ml. 50 mM Potassium phosphate buffer pH 7.2 containing 1.4 – 0 M ammonium sulfate was used. The elution of the OYE1 protein occurred at 850 mM ammonium sulfate. The purification was followed by a filtration and desalting step carried out with Vivaspin ultrafiltration spin column (Vivaspin 10 kDa, Sartorius, Göttingen Germany) and concentrated in Tris-HCl reaction buffer (50 mM, pH 7.5). The purified enzyme was stored at -20°C until further use.

2.2.4.3 Two-step purification of variant Loop A_OYE1

The loop shuffling variant Loop A_OYE1 was expressed as described elsewhere¹⁰⁸ and purified in two steps for crystallization. The first step was a Ni-based immobilized metal affinity chromatography using a 5 ml HisTrap HP column (GE Healthcare, Freiburg, Germany) with an ÄKTA system and a continuous imidazole gradient (up to 300 mM in 50 mM Tris-HCl pH 7.5). The fractions were concentrated via ultrafiltration with vivaspin columns (6 ml, 10 kDa, PES membrane, Sartorius, Göttingen, Germany) to a final volume of 1.4 ml. The second step was a size exclusion chromatography with a XK16/70 column (Pharmacia Biotech, Uppsala, Sweden), packed with sephacryl S-200 high resolution material (bed size 35 cm, maximal pressure 0.25 mPa, GE Healthcare, Freiburg, Germany) and a column flow of 0.5 ml/min of 25 mM Tris-HCl pH 7.5 using an ÄKTA explorer system. A total sample volume of 1.4 ml pre-purified protein solution was loaded on the column; the fractions (fraction size 1 ml) were merged and again concentrated via ultrafiltration with vivaspin columns to a final concentration of 35 mg/ml for the crystallization.

2.2.5 Biotransformation with wild type enzyme and created variants

2.2.5.1 Reduction of α,β -unsaturated standard substrates

All biotransformation reactions were carried out with purified enzyme (protein purity > 95 %) and in triplicates. For the reduction of standard substrates the following

conditions were used: 50 µg/ml purified enzyme, 2 mM substrate (ketoisophorone, 2-methyl-2-pentenal, cinnamaldehyde, 2,4-heptandienal, neral, geranial), 2.5 mM NADH filled up to 1 ml with 50 mM Tris-HCl pH 7.5 at 30°C and 180 rpm for 2.5 h. The reactions were terminated by the extraction with MTBE. The resulting samples were analyzed by GC-FID with the standard programs for cyclic or aliphatic substrates (table 7.6).

2.2.5.2 Thermostability

The biotransformation reactions at elevated reaction temperature¹⁰⁸, as well as the determination of the denaturation temperature with the ThermoFAD method¹¹⁷ were described in the corresponding reference and performed in triplicates.

2.2.5.3 Solvent stability

The biotransformation reactions with four different solvents (acetone, ethyl acetate, isopropyl alcohol and THF) at different solvent concentrations were described elsewhere.¹⁰⁸

2.2.5.4 Bi-enzymatic allyl alcohol reduction

All biotransformation reactions for the bi-enzymatic allyl alcohol reduction were carried out with purified enzyme (protein purity > 95 %) and in triplicates. For the allyl alcohol reduction the following conditions were used: 100 µg/ml purified enzyme, 2 mM substrate (cinnamyl alcohol, perillyl alcohol, geraniol), 0.2 U ADH equine (Sigma Aldrich, Taufkirchen, Germany), 5 mM NADH, filled up to 1 ml with 50 mM Tris-HCl pH 7.5 at 30°C and 180 rpm for 24 h. The reactions were terminated by the extraction with MTBE. The resulting samples were analyzed by GC-FID with the standard programs for cyclic or aliphatic substrates (table 7.6).

2.2.5.5 NADH/NAD⁺ cofactor handling

For the analysis of the cofactor handling of wild type enzymes and loop variants all reactions were performed in triplicates and the following conditions were used: 100 µg/ml purified enzyme (protein purity > 95 %), 2 mM cinnamaldehyde, 5 mM NADH/NAD⁺ cofactor (ratio NADH/NAD⁺: 1/0; 4/1; 1/1), filled up to 1 ml with 50 mM Tris-HCl pH 7.5 at 30°C with 180 rpm for 24 h. The resulting samples were analyzed by GC-FID with the standard programs for cyclic or aliphatic substrates (table 7.6).

2.2.6 GC analytic

2.2.6.1 GC-FID

All GC-FID measurements were performed on a GC-2010 (Shimadzu, Kyoto, Japan) equipped with an AOC-20i auto injector and a DB-5 (Agilent Technology, Waldbronn, Germany), respectively a chiral Hydrodex β-TBDAC (Machery-Nagel, München, Germany), capillary column with hydrogen as carrier gas (1.38 ml/min⁻¹). Details on the achiral and chiral developed GC programs, as well as the used columns, are given in table 7.9 in supplementary material.

2.2.6.2 GC-MS

All GC-MS measurements were performed on a GC-MS QP 2010 (Shimadzu, Kyoto, Japan) equipped with an AOC-5000 auto injector and a DB-5 MS (Agilent Technology, Waldbronn, Germany) capillary column with helium as carrier gas (0.67 ml/min⁻¹). Details on the achiral developed GC programs, as well as the used column, are given in table 7.9 in supplementary material.

3 Results

3.1 Semi-rational loop design – Golden Gate Shuffling

3.1.1 Enzyme selection and fragment definition of the ene reductases for the Golden Gate Shuffling

For the generation of new reduction biocatalysts variable surface loop regions building the catalytic interface of several selected Old Yellow Enzyme (OYE) family members were shuffled by using the technique of Golden Gate Shuffling. Therefore, four homologous shuffling partners for the 2-cyclohexen-1-one reductase NCR from *Zymomonas mobilis* within the two subfamilies of the Old Yellow Enzymes were chosen³⁰: (1) OYE1 from *Saccharomyces pastorianus* and PETNR from *Enterobacter cloacae* which are members of the *classical* subfamily; (2) YqjM from *Bacillus subtilis*, as well as GkOYE from *Geobacillus kaustophilus* being members of the *thermophilic-like* subfamily. The selection of the four shuffling partners is based on comparisons of described members within the OYE family regarding differences in amino acid composition, structure, stability as well as activity. The yeast enzyme OYE1 is selected due to the fact that it represents one of the best investigated family members in terms of structure,^{38,43,55,56} as well as substrate spectrum^{8,28,29,60,72,118,119} and possesses the property to perform promiscuous catalytic reactions like a disproportionation⁷⁰ and an isomerization⁷³ in addition to the natural reductase activity. The bacterial PETNR is selected as shuffling partner for NCR because, next to its reduction activity towards α,β -unsaturated substrates,^{59,63} it is also able to degrade high explosive molecules like nitrate esters by liberating nitrite.^{31,71} The first described gram positive bacterial enzyme YqjM is attractive due to its enhanced thermostability and the fact, that it is the first representative of the new *thermophilic-like* subclass^{35,36} acting as a tetramer of catalytically dependent dimers. The extremophile GkOYE is so far the only ene reductase described which is able to perform, beside the expected enone reduction also the reverse reaction, the desaturation of carbon-carbon double bonds next to a carbonyl group resulting in the corresponding α,β -unsaturated ketones at elevated temperature. The four selected

shuffling partners possess compared to NCR an amino acid sequence identity that ranges from 25.3 % to 40.8 % (table 3.1).

Table 3.1: Amino acid sequence identity of NCR in comparison with the four selected ene reductases based on pairwise sequence alignments using Needleman-Wunsch with standard parameters.¹²⁰

Pairwise alignment of NCR with the four selected ene reductases				
	<i>NCR/OYE1</i>	<i>NCR/PETNR</i>	<i>NCR/YqjM</i>	<i>NCR/GkOYE</i>
Sequence identity	29.4 %	40.8 %	25.6 %	25.3 %

The closest relative to NCR is the bacterial PETNR, while the two *thermophilic-like* subfamily members, as well as the yeast reductase are more distantly related. However, by performing a multiple sequence alignment of the five chosen Old Yellow Enzymes it could be shown that the two subclasses possess beside very different regions regarding amino acid chain length and composition, also areas with a high conservation of amino acids (figure 3.1).

<i>classical</i>	NCR	G-----KKPYDVARALRLDEIPRLDDYEKAARHALK AG FDGVQ IHA ANG 176
<i>subfamily</i>	PETNR	ENGNAIRVDTTTTPRALELDEIPGIVNDFRQAVANARE AG FDLVEL HS AHG 186
	OYE1	K----AKKANNPOHSLTKDEIKQYIKEYVQAAKNSIA AG ADGV EIHS ANG 196
<i>thermophilic</i>	GkOYE	D-----SSPTPKEMTKADIEETVQAFQNGARRAKE AG FDV IEIHA AHG 168
<i>subfamily</i>	YqjM	E-----QSATPVEMSAEKVKETVQEFKQAAARAKE AG FDV IEIHA AHG 168
<i>classical</i>	NCR	YL ID E FIR D ST N HR H DE Y GG A VE N RIRLLKDV T ERVIATIGK E RTAV R LS 226
<i>subfamily</i>	PETNR	YL L H Q F LS P SS N Q R TD Q Y GG S VE N RARLVLEVVD A VCNE S AD R IGIR V S 236
	OYE1	YL LN Q FLD P HS N TR T DE Y GG S IE N RAR F TLEVVDALVE A IGHE K VGL R LS 246
<i>thermophilic</i>	GkOYE	YL INE F LS P LS N RR Q DE Y GG S PE N RY R FLGEV I DAVRE V WDG-PLF V RI S 217
<i>subfamily</i>	YqjM	YL I H E F LS P LS N HR T DE Y GG S PE N RY R FLRE I IDE V KQ V WDG-PLF V RV S 217
<i>classical</i>	NCR	PNGEIQGTVD S HP---EQV F IPAA K ML S DL D -----IA F LG M REG A VD G T 268
<i>subfamily</i>	PETNR	PI G TF Q N-V D N G PN-EEADAL Y LIEELAK R G-----I A Y L H M SE T DL A G- 278
	OYE1	P Y GV F NS M SG G A E T G IV A Q Y AY V AGELE K RA K AG K RL A F V HL V E P RV T NP 296
<i>thermophilic</i>	GkOYE	AS D Y H PD G L T A K D-----Y V PY A K R M K EQ G -----V D L V D V SS G A I V P A 256
<i>subfamily</i>	YqjM	AS D Y T DK G L D I A D-----H I GF A K W M K EQ G -----V D L I D C SS G AL V H A 256

Figure 3.1: Excerpt of the multiple sequence alignment of the five selected OYE family members NCR from *Z. mobilis*, PETNR from *E. cloacae*, OYE1 from *S. pastorianus*, GkOYE from *G. kaustophilus* and YqjM from *B. subtilis*. Highlighted in bold are conserved amino acids among all enzymes. Highlighted in red is the catalytic active tyrosine. Marked in blue is a diverse region with respect to the number of amino acids and composition. The multiple sequence alignment was performed with ClustalW. For the complete alignment see figure 7.3.

For the fragment definition two main points need to be taken in account:

(1) At the beginning and at the end of each fragment one conserved amino acid is required to obtain four common nucleotides within the five shuffling partners for the *Bsa*I restriction site.

(2) One fragment can only comprise one β/α surface loop region. Based on a combination of structural superposition of NCR with the *classical*, as well as the *thermophilic-like* subclass members (figure 3.2) and a multiple sequence alignment, each of the five ene reductase sequences was divided into seven fragments (F1-F7) (figures 4.1, 7.3).

To ensure the presence of a common, conserved four nucleotide overhang at the beginning and end of each of the fragments, three point mutations have to be inserted: (1) the first point mutation was inserted in NCR, D254G, at the transition between fragment 5 (F5) and fragment 6 (F6); (2) the second mutation was inserted in OYE1, A277G, also at the transition between F5 and F6; (3) and the third mutation was again inserted in OYE1, A328T, at the transition between F6 and F7 (supplementary material table 7.5). The basic scaffold for all new created variants by the Golden Gate Shuffling was unchanged and is delivered from NCR. Therefore, in total three out of the seven defined fragments for the shuffling were taken from NCR, which were the first and last fragment, being the N-, respectively the C-terminus of the enzyme, as well as the fragment containing the catalytically essential amino acids (fragment 4).

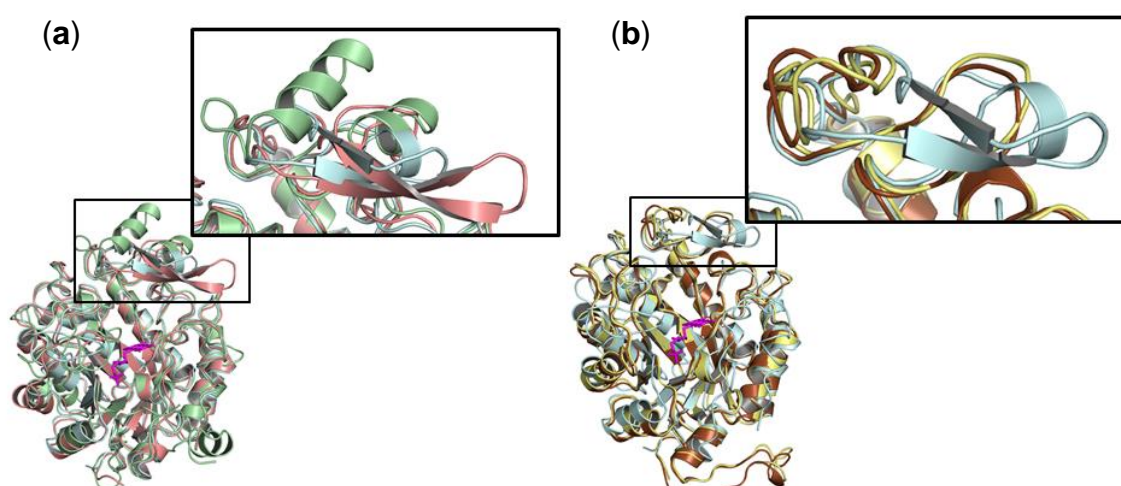


Figure 3.2: Superposition of NCR with the four chosen shuffling partners presenting structural differences in one selected loop region. (a) Structural alignment of NCR (pale cyan) with the two *classical* subfamily members OYE1 (pale green) and PETNR (pink). (b) Structural alignment of NCR (pale cyan) with the two *thermophilic-like* subclass members YqjM (yellow) and GkOYE (brown).

The remaining four fragments from all five ene reductases (F2, F3, F5 and F6) along with F4 from NCR, in total 21 fragments, were cloned in the chosen cloning vector

pUC19 (figure 7.1). As expression vector for the Golden Gate Shuffling variants pET28a(+) with the already included NCR wild type gene was used containing two additionally inserted *Bsal* restriction sites: One after F1 from NCR and the other in front of F7 from NCR (figure 7.2). The 21 different fragments allow the creation of in total 1024 different enzyme variants by the Golden Gate Shuffling that needed to be screened.

3.1.2 Screening of the Golden Gate Shuffling variants

With the developed photometric assay, based on the consumption of NADH during the reduction reaction, in total 1100 Golden Gate Shuffling (GGS) variants were screened with regard to reduction activity. Furthermore, 29 clones were sent for sequencing to verify the success of the shuffling and to determine the composition of the most active ones (for fragment composition see supplementary material table 7.6). For the classification of the reduction activity of the created GGS variants several control reactions were performed: (1) a negative control consisted of NADH and the substrate ketoisophorone, (2) a second negative control using NADH, ketoisophorone and the lysate of the expression strain *E. coli* BL21 (DE3),

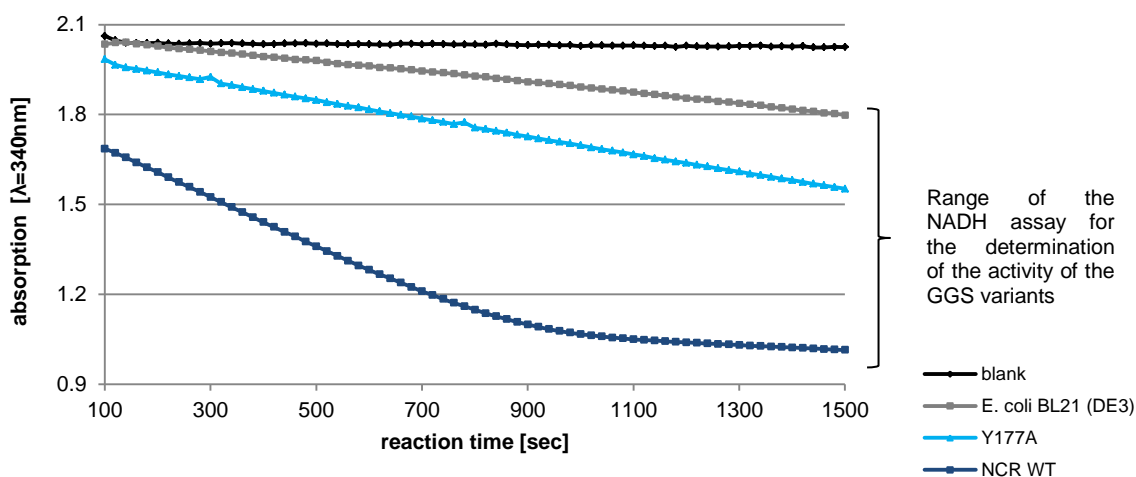


Figure 3.3: Diagram of the NADH consumption of the four control reactions starting at 100 sec reaction time. Colored in black and gray are the negative controls; colored in light and dark blue are the positive controls. Highlighted is the absorption range for the activity determination of the GGS variants.

(3) a positive control composed of NADH, ketoisophorone and the lysate of NCR wild type expressed in *E. coli* BL21 (DE3) and (4) a positive control with NADH, ketoisophorone and the lysate of variant Y177A expressed in *E. coli* BL21 (DE3). The

variant Y177A was chosen as additional positive control due to the fact that it demonstrated moderate activity of around 30 % towards the substrate ketoisophorone.⁹ These four control reactions determined the absorption range of the NADH consumption assay (figure 3.3). The limits of the photometric assay were restricted by NCR wild type enzyme showing a complete conversion of the substrate ketoisophorone and the negative control of the *E. coli* BL21 (DE3) lysate demonstrating a slight consumption of NADH due to the presence of NADH-consuming enzymes like alcohol dehydrogenases in the cell lysate. The variant Y177A represented a benchmark with around 30 % conversion of ketoisophorone. The blank reaction showed no NADH consumption at all. From the 1100 screened GGS variants, 18 demonstrated an at least slightly increased activity compared to the *E. coli* BL21 (DE3) lysate control reaction.

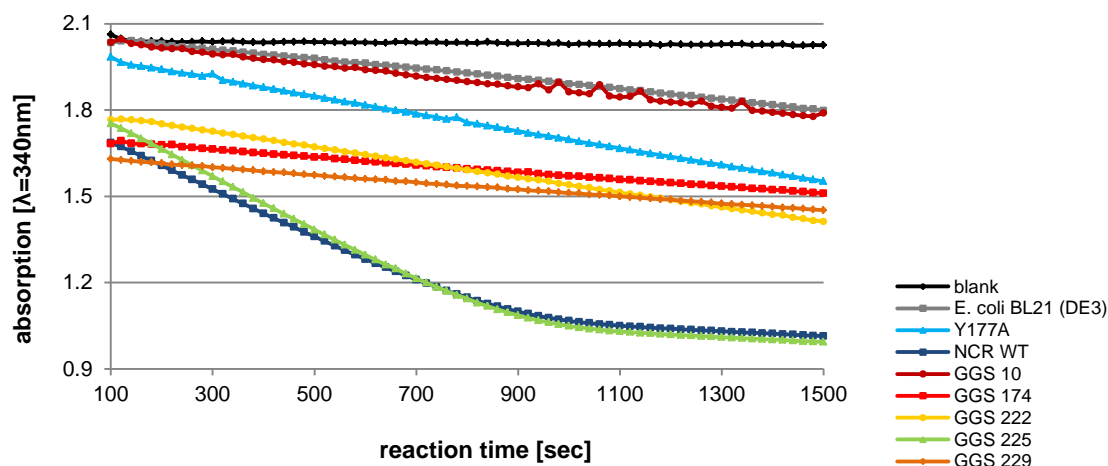


Figure 3.4: Diagram of the NADH consumption of the four control reactions with five selected GGS variants including the four variants demonstrating the greatest NADH consumption. Additionally one variant was selected showing activity similar to *E. coli* BL21 (DE3) lysate (GGS 10).

Figure 3.4 illustrates five GGS variants, including the four variants showing the greatest change in the absorption of NADH, with GGS 225 being the most active one. The three variants GGS 174, GGS 222 and GGS 229 are located in the activity range of variant Y177A, while GGS 10 shows NADH consumption near the negative lysate control. Based on the data obtained from the NADH assay, these five GGS variants (10, 174, 222, 225 and 229) were selected for biotransformation reactions in order to investigate their substrate scope and their loop composition.

The NADH-dependent photometric assay developed within this thesis represents a valuable and well applicable tool in the screening of created NADH-consuming

enzyme libraries in order to find new active variants by a fast screening method. It allows the parallel screening of 92 variants in comparison with four control reactions in a reproducible manner and a short reaction time.

3.1.3 Biotransformation with Golden Gate Shuffling variants

In this section the substrate scope of the generated Golden Gate Shuffling variants was investigated. Therefore, NCR wild type enzyme and the five active GGS variants (GGS 10, 174, 222, 225 and 229), selected by the developed NADH-dependent photometric assay, were expressed in *E. coli* BL21 (*DE3*) and purified by immobilized metal affinity chromatography (protein purity > 70 %). For the investigation of the substrate scope six substrates with different properties were selected to guarantee a wide range of substrates:

- (1) Three standard substrates for ene reductases (ketoisophorone, citral consisting of the two isomers neral and geranial)
- (2) A new activated α,β -unsaturated substrate (2,4-heptandienal)
- (3) A sterically difficult substrate for ene reductases (3-methyl-cyclohexenone)
- (4) A substrate being less activated and a non-natural substrate (cinnamyl alcohol)

The formation of reduced products is shown in table 3.2. First, it should be highlighted that all GGS variants tested were active in the reduction of α,β -unsaturated substrates in different quantities. For the two challenging substrates 3-methyl-cyclohexenone and cinnamyl alcohol no conversion was detectable at all. The GGS variant 229 was more active towards all tested α,β -unsaturated substrates than wild type NCR. It consists of fragment 2 from PETNR, fragment 3 also from PETNR and fragment 5 as well as fragment 6 from YqjM (table 3.3). Also variant GGS 225 exhibited compared to NCR a clearly increased activity in the reduction of the three substrates ketoisophorone (+25 %), 2,4-heptandienal (+38.5 %) and neral (+13 %). For geranial the reduction activity was similar to the wild type. In contrast to variant GGS 229, variant 225 still possessed two fragments of NCR (F3 and F5) while the fragments 2 and 6 were also from PETNR and YqjM, respectively. Interestingly, the presence of the two additional NCR fragments in GGS 225 compared to GGS 229 led to a decrease about 43 % in the reduction of the new

substrate 2,4-heptandienal, thus resulting in an activity in the same range as wild type NCR. Therefore, the presence of the two wild type fragments F3 and F5 demonstrate a negative effect on the reduction activity of the variant GGS 225. The variant GGS 174, possessing all four shuffled fragments from PETNR, also was active towards 2,4-heptandienal in the range of wild type NCR.

Table 3.2: Product formation table of the reduction of α,β -unsaturated substrates with five GGS variants

	Substrate/ Enzyme	Product formation in [%]			
		<i>Ketoisophorone</i>	<i>2,4-Heptandienal</i>	<i>Neral</i>	<i>Geranial</i>
controls	<i>E. coli</i> BL21 (<i>DE3</i>) lysate	0.9±0.01	0.01±0.05	4.6±0.01*	5.0±0.03*
	NCR wild type	54.0±1.2	27.5±1.4	37.7±1.6	43.9±1.1
Golden	GGS 10	39.3±3.8	60.9±1.6	30.3±0.8	30.6±6.3
Gate	GGS 174	44.1±3.8	34.0±7.6	29.6±3.7*	34.4±5.6*
Shuffling variants	GGS 222	n.d.	2.7±1.1	29.1±6.7*	16.8±6.7*
	GGS 225	67.7±4.2	38.1±1.65	48.1±0.04	43.8±0.2
	GGS 229	65.0±3.6	66.3±14.0	42.6±1.2	52.4±4.0

Reactions were performed in triplicates in a final volume of 1 ml and run at 30°C and 180 rpm for 24 h. * In the reduction of two isomers of citral, neral and geranial, also citronellol was produced as byproduct due to low protein purity. There are still alcohol dehydrogenases present in the protein mixture reducing the formed citronellal to the alcohol citronellol. For detailed information see supplementary material table 7.7, figure 7.4 and Morlock, L., 2013.¹¹⁰

GGS 10 instead, exhibited an over twofold increased conversion towards 2,4-heptandienal in the range of the variant GGS 229. Just GGS 222 had a clearly decreased activity for 2,4-heptandienal (-90 %). None of the tested GGS variants demonstrated a clear influence on the *cis/trans* substrate specificity for the two isomers of citral, neral and geranial. GGS 222 was the least active enzyme variant tested. In summary, almost all GGS variants tested, except GGS 222, were more active than wild type NCR in the reduction of the new substrate 2,4-heptandienal indicating that the fragment composition of the variants have an influence on the reduction activity of the enzyme.

Table 3.3: Fragment composition of the created GGS variants.

GGS variants	Fragment composition of the GGS variants			
	fragment 2	fragment 3	fragment 5	fragment 6
GGS 10	NCR	OYE1	NCR	NCR
GGS 174	PETNR	PETNR	PETNR	PETNR
GGS 222	n.d.	n.d.	n.d.	n.d.
GGS 225	PETNR	NCR	NCR	YqjM
GGS 229	PETNR	PETNR	YqjM	YqjM

The conformation of the fragment composition of the GGS variants was done by DNA sequencing (GATC-Biotech, Köln, Germany). n.d. not determined.

To sum up this chapter, the technique of the Golden Gate Shuffling enables the semi-rational design of new enzyme variants possessing a catalytic interface composed of different β/α surface loop regions as well as a defined library size based on the amount of defined fragments. It also allows the generation of biocatalysts with increased activity towards selected substrates. Therefore, the Golden Gate Shuffling represents a fast, efficient and applicable DNA shuffling method in order to generate new variants of homologous enzymes with improved properties based on differences in β/α surface loop regions.

3.2 Rational loop design

3.2.1 Determination of loop regions and model enzymes

All members of the family of the Old Yellow Enzymes consist of the same structural scaffold, the TIM barrel, possessing the active site within the catalytic interface built by the β/α loops at the C-terminal end of the β -sheet barrel. By comparing various ene reductase family members, it is apparent that the largest differences with respect to structure as well as amino acid composition and length of the sequence are located in the β/α loop regions of the enzymes. In the developed and combined approach consisting of a multiple sequence alignment and a structural alignment of Old Yellow Enzymes, three loop regions considerably differing in structure and amino acid composition were defined.

NCR	-----MPSLFDPIRFGAFTAKNRIWMAPLTRGRATR-DHVPTEIMAEYYAQR	46
MR	-----MPDTSFSNPGLFTPLQLGSLSLPNRVIMAPLTRSRTP--DSVPGRLQOIYYGQR	52
OYE1	MSFVKDFKPKQALGDTNLFKPIKIGNNEELLHRAVIPPLTRMRALHPGNI PNDRWAVEYYTQ	60
OYE2	MPFVKDFKPKQALGDTNLFKPIKIGNNEELLHRAVIPPLTRMRAQHPGNI PNDRWAVEYYAQ	60
OPR1	-----MENGEAKQSVPLLTYPYKMGFRNLSHRVVLAPLTRQRSY--GNVPQPHAAIYYSQR	53
OPR3	-----MASSAQDGNPNLFSYKMGKFNLSHRVVLAPMTRCRAL--NNIPQAAALGEYYEQR	53
YqjM	-----MARKLFTPTITIKDMTLKNRIVMSPCMYSSHEKDGKLTFFHMAHYISR	48
GkOYE	-----MNTMLFSPYTIRGLTLKNRIVMSPCMYSCDTKDGAVRTWHKIHYPAR	48
NCR	--ASAG-LIISEATGISQEGLGWPYAPGIWSDAQVEAWLPITQAVHDAGGLIFA	QLWHMG 103
MR	--ASAG-LIISEATNISPTARGVYVTPGIWTDQAEAGWKGVVEAVHAKGGRIAL	QLWHVG 109
OYE1	RAQRPGTMIITEGAFISPOAGGYDNAPGVWSEEQMVWTKIFNAIHEKKSFWVW	QLWVLG 120
OYE2	RAQRPGTMIITEGTFPSPQSGGYDNAPGIWSEEQIKWTKIFKAIHENKSAFWV	QLWVLG 120
OPR1	--TTPGGFLITEATGVSDTAQGYQDTPGIWTKHVEAWKPIVDVAVHAKGGIFFC	QIWHVG 111
OPR3	--ATAGGFLITEGTMISPTSAGFPHPVPGIFTKEQVREWKIVDVVHAKGAVIFC	QLWHVG 111
YqjM	AIGQVG-LIIVEASAVNPQGRITDQDLGIWSEDEHIEGFAKLTEQVKEQGSKIGI	QLAHAG 107
GkOYE	AVGQVG-LIIVEATGVTPQGRISERDLGIWSDDHIAGLRELVLVKEHGAAIGI	QLAHAG 107
Loop C		
NCR	RMVP-SNVS--GMQPVPASASQAPGLGHTYDGKK-----PYDVARALRLDE	IPRLDDY 154
MR	RVSH-ELVQPDGQOPVAPSALKAEGAECFVEFEDGTAGLHPTSTPRALETDG	IPGIVEDY 168
OYE1	WAAFDPNLRDGLRYSASDNVFMDAEQEAKAKKAN-----NPQHSITKDE	IKQYIKEY 174
OYE2	WAAFDPNLRDGLRYSASDNVYMNAEQEEKAKKAN-----NPQHSITKDE	IKQYVKEY 174
OPR1	RVSN-SGFQPNKAPISCSDKPLMPQIRSNID----EALFTPPRRLGIEE	IPGIVNDF 165
OPR3	RASH-EVYQPAAGAAPISSTKPIFNRRWIRLMPDGT---HGIYKPKPRAIGTYE	ISQVVEDY 167
YqjM	RKAE-----LEGDIFAPSAIAFDEQS-----ATPVMSAEK	VKETVQEF 146
GkOYE	RKSQ-----VPGEIIAPSAVFPDDSS-----PTPKEMTKAD	IETVQAF 146
NCR	EKAARHALKAGFDGVQIHAANGYLIDEFIRDSTNHRHDEYGGAVENRIRLLKDVTERVIA	214
MR	RQAAQRAKRAGFDMVEVHAANACLPNQFLATGTNRRTDQYGGSIENRARFPLEVVDVAVE	228
OYE1	VQAAKNSIAAGADGVEIHSANGYLLNQFLDPSNTRTDEYGGSIENRARFTELVVDALVE	234
OYE2	VQAAKNSIAAGADGVEIHSANGYLLNQFLDPSNTRTDEYGGSIENRARFTELVVDVVD	234
OPR1	RLAARNAMEAGFDGVEIHGANGYLDQFMKDTVNDRTDEYGGSLQNRCKFPLEIVDAVAK	225
OPR3	RRSALNAIEAGFDGIEIHGAHGYLIDQFLKDGINDRTDEYGGSLANRCKFITQVVQAVVS	227
YqjM	KQAAARAKEAGFDVIEIHAAGYLIHEFLSPLSNHRTDEYGGSPENRYRFLREIIDEVKQ	206
GkOYE	QNGARRAKEAGFDVIEIHAAGYLINEFLSPLSNRRQDEYGGSPENRYRFLGVEIDAVRE	206
Loop A		
NCR	TIGKERTAVRLSPN--GEIQGTVDSHPE	QVFI PAAKMLS DLD-----IAFLQ MREGAV 265
MR	VFGPERVGI R L T P F ---LELFG L T D D E P E	A M A F Y L A G E L D R R G -----L A Y L H F N E P D W 279
OYE1	AIGHEKVG L R L S P Y G V F N S M S G G A E T G I V A	Q Y A Y V A G E L E K R A K A G - K R L A F V H L V E P R V 293
OYE2	AIGPEKVG L R L S P Y G V F N S M S G G A E T G I V A	Q Y A Y V L G E L E R R A K A G - K R L A F V H L V E P R V 293
OPR1	EIGPDRVGI R L S P F ---A D Y M E S G D T N P G A	L G L Y M A E S L N K Y G -----I L Y C H V I E A R M 276
OPR3	AIGADRVG V R V S P A ---I D H L D A M S N P L S	L G L A V V E R L N K I Q L H S G S K L A Y L H V T Q P R Y 284
YqjM	VWDG-PLFVR V S A S -----D Y T D K G L D I A	D H I G F A K W M K E Q G -----V D L I I C S S G A L 253
GkOYE	VWDG-PLFVR I S A S -----D Y H P D G L T A	K D Y V P Y A K R M K E Q G -----V D L V I V S S G A I 253
Loop B		
NCR	DGTFGKTDQ-----PKLS	PEIRKVF K P P L V L N Q D Y T F E - T A Q A A L D S G V A D A I S F G R 316
MR	IG-GDITYP-----EGFR	EQMRQRFKGGLIYCGNYDAG-RAQARLDDNTADAVAFGR 329
OYE1	TNPFLTEGEG----EYEGGSN	DFVYSIWKGPIRAGNFALHPEVVREEVKDKRTLIGYGR 349
OYE2	TNPFLTEGEG----EYNGGSN	KFAYSIWKGPIRAGNFALHPEVVREEVKDPRTLIGYGR 349
OPR1	KTMGEVHA-----CPHTL	MPMRKAFKGTFTISAGGF TRE-DGNEAVSKGRDTLVAYGR 327
OPR3	VAYGQTEAGRLGSEEEEEARLM	RTLNRNAYQGTFTICSGGYTRE-LGIEAVAQGDADLVSYGR 343
YqjM	VHADINVPFG-----YQVSFA	EKIREQADMATGAVGMITDGSMAEEILQNGRADLIFIGR 308
GkOYE	VPARMNVYPG-----YQVPFA	ELIRREADIPTGAVGLITSGWQAEIILQNGRADLVFLGR 308
NCR	FFIGNPDLPRRF FEKAPLTKDVIETWYTQTPKG-YTDYPLLGD-----	358
MR	FFIANPDLPERFRLGAALNEPDPSTFYGGAEVG-YTDYPLDNHGRDLG----	377
OYE1	FFISNPDLVDRLEKGLPLNKYDRDTFYQMSAHG-YIDYPTYEEALKLGWDKK--	400
OYE2	FFISNPDLVDRLEKGLPLNKYDRDTFYKMSAEG-YIDYPTYEEALKLGWDKN--	400
OPR1	WFLANPDLPKRFQVDAPLNKYDRPTFYTSDPVVGYTDYPFLESTA-----	372
OPR3	LFISNPDLVMRIKLNAPLNKYNRKTFYTDQDPVVGYTDYPFLQNGSNGPLSRL	396
YqjM	ELLRDPFFARTAAKQLNTEIPAPVQYERGW-----	338
GkOYE	ELLRNPYWPYAAARELGAKISAPVQYERGWRF-----	340

Figure 3.5: Multiple sequence alignment of eight members of the Old Yellow Enzyme family performed with ClustalW. The three enzymes NCR, MR and OYE1 selected for the structural alignment are highlighted in the corresponding colors of the structural representation (figure. 3.6). Framed and marked in bold are the three defined β/α surface loop regions Loop A, Loop B and Loop C.

For the first part, the multiple sequence alignment, a total of eight family members with already published crystal structures were selected (figure 3.5). To ensure a wide palette of OYE enzymes the eight alignment partners consisted of two bacterial representatives (NCR from *Z. mobilis* and MR from *P. putida* M10), two yeast enzymes (OYE1 from *S. pastorianus* and OYE2 from *S. cerevisiae*), two members of the plant reductases (OPR1 from *A. thaliana* and OPR 3 from *S. lycopersicum*) as well as two Old Yellow Enzymes from the *thermophilic-like* subclass (YqjM from *B. subtilis* and GkOYE from *G. kaustophilus*). In the second part of the approach for the loop definition, the structural alignment, the selected ene reductases were superimposed to visualize the structural differences within β/α surface loop regions. To guarantee a clear visualization of the loop regions, only three out of the eight alignment partners were picked: NCR from *Z. mobilis*, MR from *P. putida* M10 and OYE1 from *S. pastorianus*. With regard to the superimposition, three β/α loop regions attracted attention by differing in secondary structure elements, loop amino acid compositions as well as loop length (figure 3.6). Each loop region is defined as the flexible coil region in-between a leading TIM barrel building β -sheet and the following α -helix, plus the amino acids anchoring the loop region in the previous, as well as subsequent secondary structure element.

The three defined loops were ordered according to the loop length from the shortest to the longest loop and entitled Loop A, Loop B and Loop C (table 3.4). The length of Loop A differs between the selected ene reductases from 13 amino acids at two *thermophilic-like* reductases, over 16 amino acids for the bacterial members NCR and MR as well as the two plant enzymes, to 19 amino acids for the two yeast enzymes. Due to its limited size Loop A possesses no secondary structure element and is the most related loop within the three superimposed reductases with regard to the crystal structure (figure 3.6). The second defined loop region, Loop B, lasts from 18 aa for the bacterial MR to 27 aa of the plant enzyme OPR3 and is clearly more diverse within the aligned, as well as superimposed family members (figures 3.5, 3.6). In this loop region the bacterial enzymes NCR (with 19 aa) and MR (with 18 aa), as well as the plant reductase OPR1 (with 19 aa), are shorter in length than the *thermophilic-like* subclass representatives YqjM and GkOYE (with 22 aa).

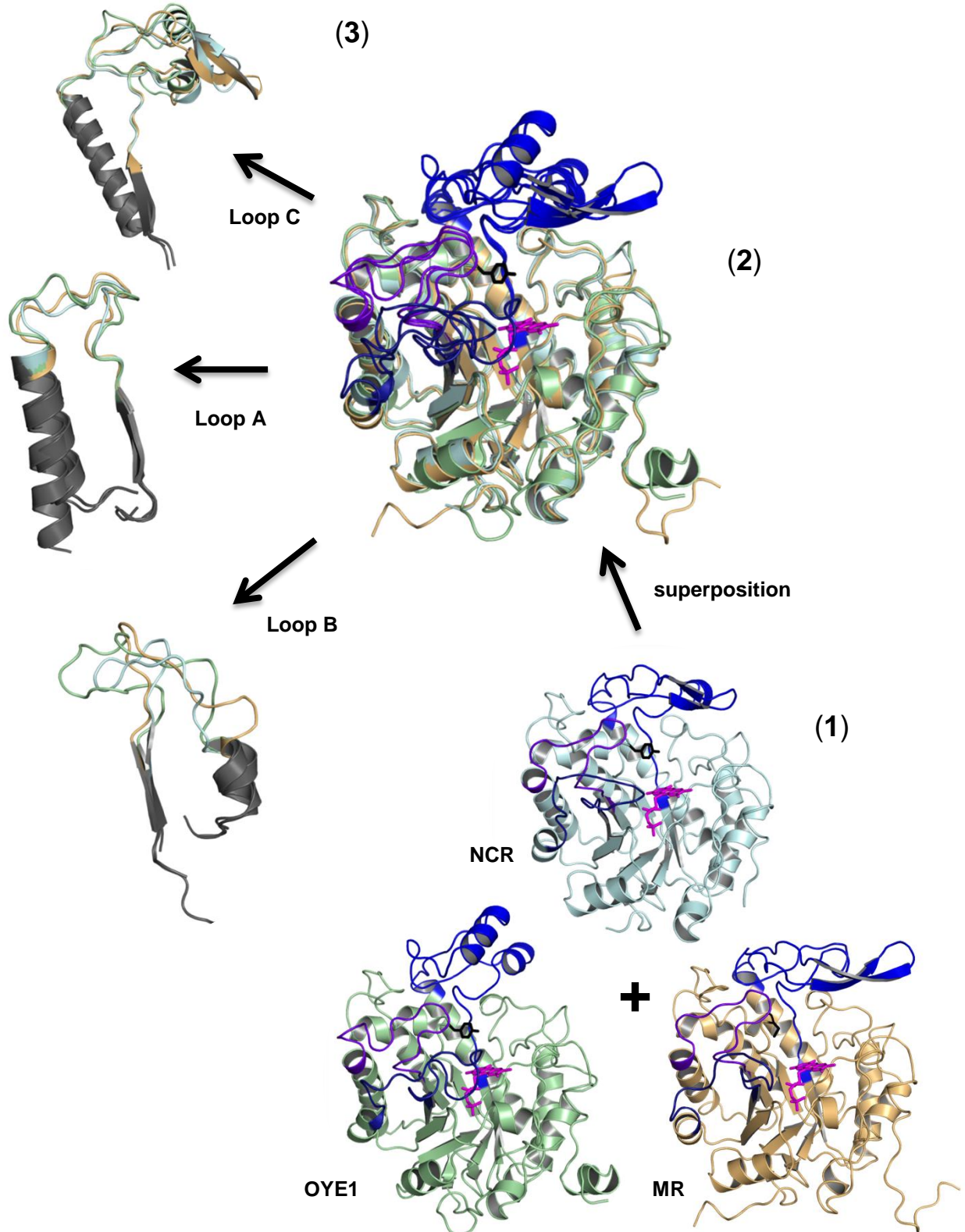


Figure 3.6: Structural representation of the three defined β/α surface loop regions based on the superimposition of three ene reductases (NCR colored in cyan, OYE1 colored in green and MR colored in yellow). (1) Cartoon representation of NCR (pdb file: 4a3u), OYE1 (pdb file: 1oya) and MR (pdb file: 1gwj) with the three defined loop regions. Loop A is highlighted in purple, Loop B in dark blue and Loop C in blue. (2) Superposition of NCR, OYE1 and MR with the three defined loop regions. (3) Excerpt of the superposition of the three ene reductases containing the three defined β/α loops with the previous and following secondary structure element colored in gray.

While the two yeast enzymes, OYE1 and OYE2, possess the same length (23 aa) as well as nearly the same amino acid sequence (only one mutation in OYE2 compared to OYE1, E306N), the two plant reductases OPR1 and OPR3 differ considerably. OPR1 has a rather small Loop B with just 19 aa in length while OPR3 possesses the largest loop with 27 aa. In terms of the structural superposition of NCR, OYE1 and MR it is apparent that this loop region still possesses no additional secondary structure element, but is more diverse in its loop orientation than Loop A. Loop C is within all aligned family members clearly noticeable the largest of the three defined loop regions. Again, the two thermophilic ene reductases YqjM and GkOYE have the smallest loop with 37 aa in length, followed by 49 aa of NCR and 52 for OYE1, OYE2 and OPR1. The Loop C of the second plant reductase OPR3 is 54 amino acids long; however, the largest loop is presented by the bacterial MR with 57 aa. Figure 3.6 illustrates the large differences in Loop C regarding loop structure, as well as the included secondary structure elements. NCR additionally carries in the Loop C region two small β -sheets, while OYE1 has an α -helix. MR encompasses, like NCR, two further β -sheets; however, they are clearly increased in size.

To sum up, the β/α loop regions building the catalytic interface of the TIM barrel based OYE family differ clearly in amino acid composition, loop size as well as loop structure. For the answering of the central question of the present thesis on the role of β/α surface loops within the enzyme two of the three defined loop regions, Loop A and Loop B, were selected as starting point due to their limited size in comparison with Loop C.

Table 3.4: Compilation of the defined three loop regions in NCR, MR and OYE1 with the corresponding loop length

Loop region	Amino acid position and loop length of the defined loop regions					
	NCR		MR		OYE1	
	aa position	loop length	aa position	loop length	aa position	loop length
Loop A	L ₂₂₅ – E ₂₄₀	16 aa	L ₂₃₉ – E ₂₅₄	16 aa	L ₂₄₅ – V ₂₆₃	19 aa
Loop B	M ₂₆₀ – S ₂₇₈	19 aa	F ₂₇₄ – R ₂₉₁	18 aa	L ₂₈₈ – N ₃₁₀	23 aa
Loop C	Q ₉₈ – E ₁₄₆	49 aa	Q ₁₀₄ – G ₁₆₀	57 aa	Q ₁₁₄ – E ₁₆₅	52 aa

As model enzyme for the rational analysis of β/α loop regions the bacterial 2-cyclohexen-1-one reductase NCR from *Zymomonas mobilis* was chosen, because

of the fact that NCR represents a recently described Old Yellow Enzyme, belonging to the *classical* subfamily.⁸ Additionally, it acts as an active monomer and exhibits a different enantioselectivity for several substrates, like the aliphatic isomer mixture citral compared to other *classical* OYE⁶⁰. As comparison partners for the influence of loop composition, length and loop structure on ene reductase properties, two already introduced *classical* subfamily members were chosen:

(1) Old Yellow Enzyme 1 (OYE1) from *Saccharomyces pastorianus*, because it is by far the best described ene reductase^{8,28,29,37,38,43,55,56,60,69,72,119,121} differing in the activity and enantioselectivity of several substrates compared to NCR and possesses an average sequence similarity of 29.4 % with the model enzyme.

(2) The morphinone reductase (MR) from *Pseudomonas putida* M10 which is one of the few bacterial ene reductases with known physiological function. It is also the only described member that has a catalytic active cysteine instead of a tyrosine.^{34,62,63,66,122–124} The sequence identity of MR with NCR is 42.4 %.¹¹²

By initially first comparing the defined Loop A region of NCR, OYE1 and MR concerning loop length and amino acid composition, it is obvious that both bacterial reductases have a loop of 16 amino acids while OYE1 demonstrates a by three amino acids (aa) elongated loop area. The sequence identity of the Loop A region between NCR/MR (43.8 %) and NCR/OYE1 (25.0 %) corresponds exactly with overall sequence identity (NCR/MR 42.4 % and NCR/OYE1 29.4 %) for them. All three loops possess a negatively charge character; OYE1 exhibits one glutamic acid, NCR has in total three negatively charged amino acids, two glutamic acids and one aspartic acid, while the Loop A of MR is the most acidic region with five negatively charged residues especially at the C-terminal end of the loop (table 3.5).

Table 3.5: Amino acid composition of Loop A and Loop B in NCR, MR and OYE1. Highlighted in bold are the amino acids located in the previous and subsequent secondary structure element.

Enzyme	Amino acids loop sequence	
	Loop A	Loop B
NCR	²²⁵ L S P N G E I Q G T V D S H P E ₂₄₀	²⁶⁰ M R E G A V D G T F G K T D Q P K L S ₂₇₈
MR	²³⁹ L T P F L E L F L G L T D D E P E ₂₅₄	²⁷⁴ F N E P D W I G G D I T E P E G F R ₂₉₁
OYE1	²⁴⁵ L S P Y G V F N S M S G G A E T G I V ₂₆₃	²⁸⁸ L V E P R V T N P F L T E G E G E Y E G G S N ₃₁₀

Loop B demonstrates a clearly different behavior. By comparing the three ene reductases NCR, MR and OYE1 it is observable that their loops have different lengths as well as a completely different amino acid composition, which is noticeable in the low sequence identity of just 2.9 % for both NCR and MR, as well as NCR and OYE1. Therefore, one could conclude based on those observations, that there exists no correlation between different β/α loop regions of Old Yellow Enzymes. Would a correlating behavior exist between different β/α loop regions, Loop B would demonstrate the same behavior like Loop A with respect to the connection of loop sequence identity and overall sequence identity between two compared family members.

3.2.2 Systematic and bioinformatic analysis of the OYE family and the selected β/α loop regions

This section will address the issue of a more detailed investigation of the familiar relationship within the superfamily of the Old Yellow Enzymes with regard on overall sequence comparison, as well as the two loop region A and B. For this purpose four main approaches were determined, which were investigated in cooperation with Dr. Michael Widmann from the bioinformatic group:

- (1) Creation of a phylogenetic tree of the superfamily of the Old Yellow Enzymes with all published ene reductase sequences.
- (2) Generation of Hidden Markov profiles for the two loop regions A and B of the three chosen ene reductases NCR, MR and OYE1.
- (3) Generation of a phylogenetic tree of the homologous family (HF) 1 and the distribution of Loop A and B within the phylogenetic tree.
- (4) Alanine scanning mutagenesis of Loop A of NCR in order to determine catalytic important amino acids.

Creation of a phylogenetic tree of the superfamily of the Old Yellow Enzymes: By today, a total amount of 4483 protein sequences of ene reductases are stored in databases building the superfamily of the Old Yellow Enzymes. Based on a multiple sequence alignment and the overall sequence identity of these 4483 members, a phylogenetic tree for the Old Yellow Enzymes was created with the PHYLIP package (figure 3.7).¹¹⁶ Each amino acid sequence entrance obtained an internal ID number

Thermus scotoductus SA-01.^{41,42} Interestingly, the representatives from the *thermophilic-like* subclass³⁰ are divided into two homologous families, HF4 and HF5. While most of the described thermophilic ene reductases, like YqjM from *B. subtilis*,^{35,36} GkOYE from *G. kaustophilus*⁴⁰ and TOYE from *T. pseudethanolicus*¹³² are present in the small homologous family HF5 (271 members and 20 solved crystal structures), some of them are also part of the HF4. By comparison of the amino acid sequences of the known and described HF4 subfamily members XenA and CrS with the representatives YqjM and GkOYE from HF5 it is apparent that both homologous families are quite related to each other. XenA possesses in a pairwise sequence alignment a sequence identity of 40.3 % with YqjM and 41.4 % with GkOYE, respectively, using Needleman-Wunsch with standard parameters,¹¹⁵ while the also thermophilic CrS even demonstrates a higher similarity with 49.0 % for YqjM, 53.5 % for GkOYE, respectively. However, the members of the HF5 are even more related with sequence similarities up to 66.5 % (YqjM with GkOYE) which can also be seen in their close distance within the phylogenetic tree (marked in red). The last two homologous families, HF2 and HF3, are two rather small ones with 101 and 307, respectively, family members. For both homologous families not even one biochemically characterized or described ene reductase is known so far, and no crystal structure is solved. Both, HF2 and HF3 are built by sequences which are mainly obtained by genome sequencing projects. However, it would be interesting to investigate some of their members with respect to activity, substrate specificity, enantioselectivity and thermo-, as well as solvent stability to compare them with already described OYE. Such comparisons would uncover whether the differences in amino acid composition are also present in enzyme properties.

Generation of Hidden Markov profiles for Loop A and Loop B: Due to the fact that the main interest of this thesis is focused on β/α surface loop regions of ene reductases, it should be determined, if these loops are related between different family members possessing conserved and functionally important amino acids or if their amino acid composition and length is just coincidence. In order to obtain an insight into the amino acid composition, as well as an understanding of the relationship of β/α loop regions a general strategy for the generation of Hidden Markov loop profiles was developed (figure 2.1). Hidden Markov profiles indicate the frequency of an amino acid at a certain position within a set of sequences representing the level of conservation and relationship, respectively. For the generation of the Hidden Markov

profiles secondary structure predictions using the psipred method¹³³ were performed in order to define the β/α loop region solely depending on sequence data. Based on the prediction, the starting amino acid for each loop region was determined as the last amino acid of the leading β -sheet. Additionally, the final amino acid for the loop profile was defined as the last amino acid of the flexible coil region (figure 7.5). As reference sequences the already introduced NCR from *Z. mobilis*, MR from *P. putida* M10 and OYE1 from *S. pastorianus*, all belonging to homologous family HF1, were used. Essentially, the developed strategy was based on three steps: (1) the first step was a BLAST search within the superfamily of the ene reductases for each complete amino acid sequence of NCR, MR and OYE1. From the first 500 hits of each BLAST search all sequences with more than 30 % gaps within the defined loop regions A or B were discarded. (2) With the remaining BLAST hits pairwise sequence alignments in comparison with the corresponding reference sequence were performed. (3) For the generation of the first Hidden Markov loop profile only the sequences possessing > 40 % identity with the reference sequence in the loop region were used to ensure a relationship. To increase the amount of sequences building a Hidden Markov profile, a profile search against the developed and established OYE superfamily database was performed. All received sequence hits were used to generate the final Hidden Markov loop profiles. In total six loop profiles were generated three for the Loop A region (figure 3.8) and three for the Loop B region (figure 3.9).

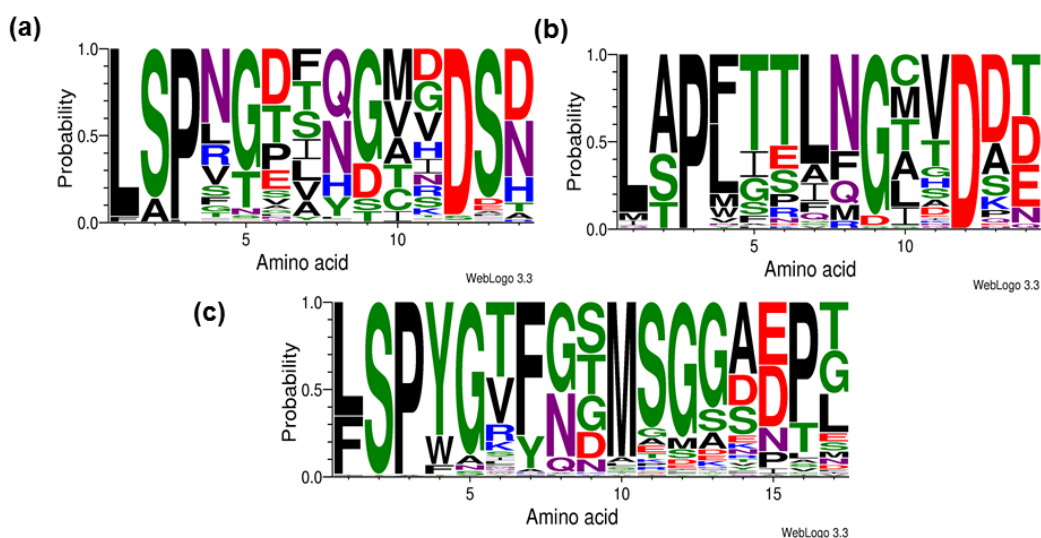


Figure 3.8: Graphical representation of the generated Hidden Markov profiles for Loop A. The size of the amino acid letter corresponds with its frequency at the relevant position within the loop region. The Hidden Markov profiles are visualized with WebLogo 3.3 and the amino acids are colored on basis of their chemistry. (a) Hidden Markov profile of Loop A from NCR with a loop length of 14 aa and made up by 342 sequences. (b) Hidden Markov profile of Loop A from MR with a loop length of 14 aa and made up by 253 sequences. (c) Hidden Markov profile of Loop A from OYE1 with a loop length of 17 aa made up by 270 sequences.

Based on the psipred secondary structure prediction loop lengths of 14 amino acids (aa) for the Loop A of NCR and MR could be identified (NCR: from L225 to H238; MR: from L239 to E252) and a length of 17 aa for OYE1 (from L245 to G261). The three generated loop profiles were composed of altogether 865 ene reductases which are all located in the HF1 subfamily. Thereby 1/3 of the HF1 sequences are represented by these three loop profiles, with the NCR profile being the largest one consisting of 342, followed by the OYE1 profile with 270 and finally the MR profile representing 253 sequences. By comparing the three Loop A profiles, it could be detected that all profiles possess a common, highly conserved N-terminal hinge region consisting of the three amino acids L-S/A-P, as well as at least one negatively charged conserved amino acid D/E at the C-terminal loop end. The middle part of the loop profiles is more flexible in terms of amino acid composition. For example, the 342 sequences forming the NCR profile possess two quite conserved glycines which are bordering a polar common asparagine/glutamine. The residual part of the loop is quite flexible in the amino acid composition. The second conserved glycine, as well as the preceding asparagine are also present in the MR profile, while the first common glycine is missing there. Instead of the first glycine the MR profile exhibits a more conserved middle part build up by two quite conserved threonines and one leucine in front of an asparagine. The two Hidden Markov profiles of NCR and MR are quite related to each other, even sharing 109 sequences. The OYE1 based profile, however, demonstrates a completely different middle part of the Loop A region. Here, a highly conserved methionine is present, surrounded by four quite conserved glycine residues leading to the consensus sequence Y-G-X-F-G-X-M-S-G-G, whereas the X stands for a variable amino acid (figure 3.8c). Based on the three Hidden Markov profiles of Loop A it can be assumed, that this loop motif is a frequently appearing region within the HF1 subfamily with some necessary, conserved amino acids and also parts that allow a diversity in the amino acid composition. These three profiles represent a broad spectrum of ene reductases.

In contrast, the Loop B region demonstrates a different behavior. Assessing the psipred secondary structure prediction Loop B consists of 18 amino acids (from R261 to S278) for NCR, 13 aa (from N275 to P287) for MR and 22 aa (from V289 to S309) for OYE1. The sequence-based loop definition differed from the structural loop definition, especially for the Loop B of MR. The flexible coil region of the structural loop definition based on the crystal structure of MR (pdb file 1GWJ), is three amino

acids longer (E288, G289 and F290) in size than the one obtained by the psipred prediction. These different results indicate that the exact definition of the end of a secondary structure element and the beginning of the following loop region is a flexible transition state. The three Loop B profiles included in total 229 ene reductases, which are all also part of the HF1 subfamily. Altogether, these 229 sequences represent 1/4 of sequences forming the Loop A profiles. Therefore, the Loop B region seems to be a more specialized loop region within a specific enzyme. This assumption is also supported by the fact that with the same strategy like for Loop A only 40 sequences with NCR, 116 with MR and 73 with the OYE1 sequence were found. This lower amount of sequences forming a profile led to a higher degree of conservation within each profile (figure 3.9).

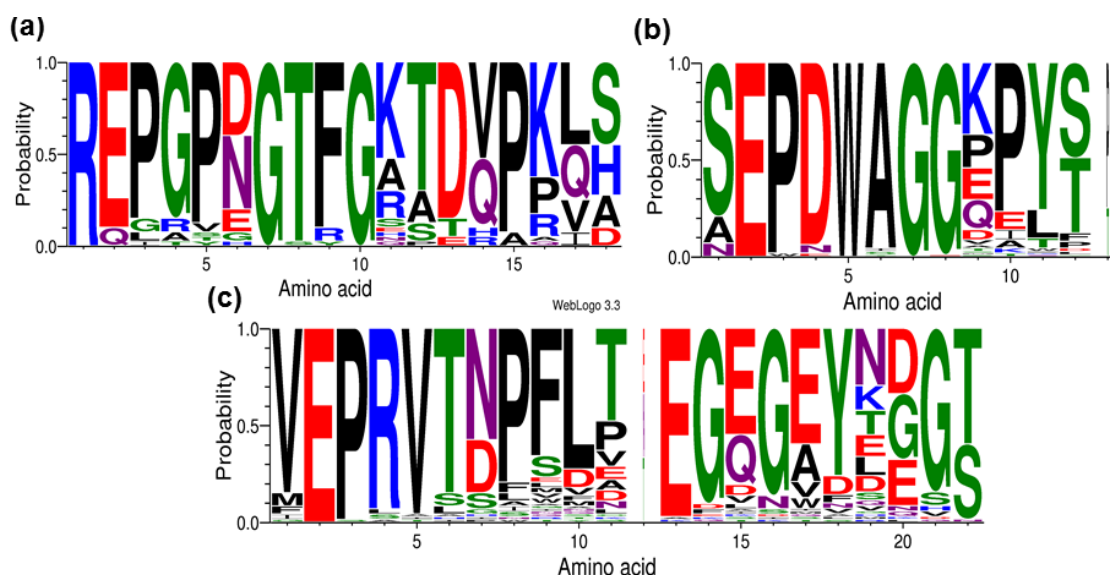


Figure 3.9: Graphical representation of the generated Hidden Markov profiles for Loop B. The size of the amino acid letter corresponds with its frequency at the relevant position within the loop region. The Hidden Markov profiles are visualized with WebLogo 3.3 and the amino acids are colored on basis of their chemistry. (a) Hidden Markov profile of Loop B from NCR with a loop length of 18 aa and made up by 40 sequences. (b) Hidden Markov profile of Loop B from MR with a loop length of 13 aa and made up by 116 sequences. (c) Hidden Markov profile of Loop B from OYE1 with a loop length of 22 aa made up by 73 sequences.

Therefore, for example, the NCR profile demonstrates the highly conserved consensus sequence R-E-P-G-P-D/N-G-T-F-G-X-T-D-V/Q-P-X-X-X, possessing only four variable amino acid positions (indicated by X), mainly at the C-terminal end of the loop. The other 14 amino acids are highly conserved within the 40 sequences. This loop schema of a long, highly conserved N-terminal as well as middle part, and a more flexible C-terminal part is also present in the other two Loop B profiles,

independent from the loop length. In contrast to Loop A, it was not possible to determine an N- or C-terminal common hinge region. These facts lead to the assumption that each loop region within one enzyme possesses a different behavior regarding its development. If the development of the β/α loop region would be correlated, all sequences containing one loop should also contain the other one. However, this is clearly not the case for the occurrence of Loop A and Loop B. Loop A represents a more general region, being present in a large number of sequences while Loop B seems to be a more specialized loop region within one enzyme, not present in many other relative sequences. Obviously, it would now be interesting to determine, where within a phylogenetic tree the sequences building one loop profile are located and if the overall sequence identity is correlated with the loop sequence identity. Due to the fact that the sequences forming all six loop profiles are located within the HF1, a phylogenetic analysis of this subfamily is sufficient to address that question.

Generation of a phylogenetic tree of HF1: For the investigation of the distribution of the loop profile forming sequences a phylogenetic tree using the 2558 sequences representing the HF1 subfamily was made. Similar to the phylogenetic tree of the OYE superfamily, this one was also based on a multiple sequence alignment and the overall sequence identity by using the PHYLIP package (figure 3.10).¹¹⁶ The positions of the selected reference sequences NCR, MR and OYE1 within the phylogenetic tree are highlighted in figure 3.10a. The most relevant observations and results of this phylogenetic analysis of the HF1 can be summarized as followed:

(1) NCR, MR and OYE1 represent a broad sequence spectrum of the HF1 superfamily. The three reference sequences are widely distributed over the whole sequence space of the HF1 subfamily, due to the fact that their sequence similarities are ranging from 29.4 % (NCR/OYE1) to 42.4 % (NCR/MR) (figure 3.10a). Therefore, they are ideal reference sequences for covering a large sequence space within the HF1 subfamily.

(2) No correlation exists between the overall sequence identity and the occurrence of one specific loop motif. This finding can be deduced by the fact that, for example, the 270 sequences forming the OYE1 Loop A profile are located in three separate branches of the phylogenetic tree. Two of them are situated in close proximity to each other, while the third branch is located at an opposite site of the

phylogenetic tree (figure 3.10.b, highlighted in blue) indicating a lower overall sequence identity with the other two branches. This finding is also supported by the behavior of the sequences forming the Loop A profiles of NCR and MR, although both demonstrate a different distribution compared to the OYE1 profile forming sequences. The NCR and MR profile forming sequences are mixed in their distribution over the phylogenetic tree and not allocated with particular branches. Both profiles even possess 109 sequences in common (figure 3.10b, colored cyan) due to their rather high sequence identity of 42.4 %. If the occurrence of one specific loop motif would be correlated with the overall sequence identity, the sequences forming a loop profile would be restricted to closely related enzymes. This is not the case; therefore a loop region can be evolved independent from the overall sequence.

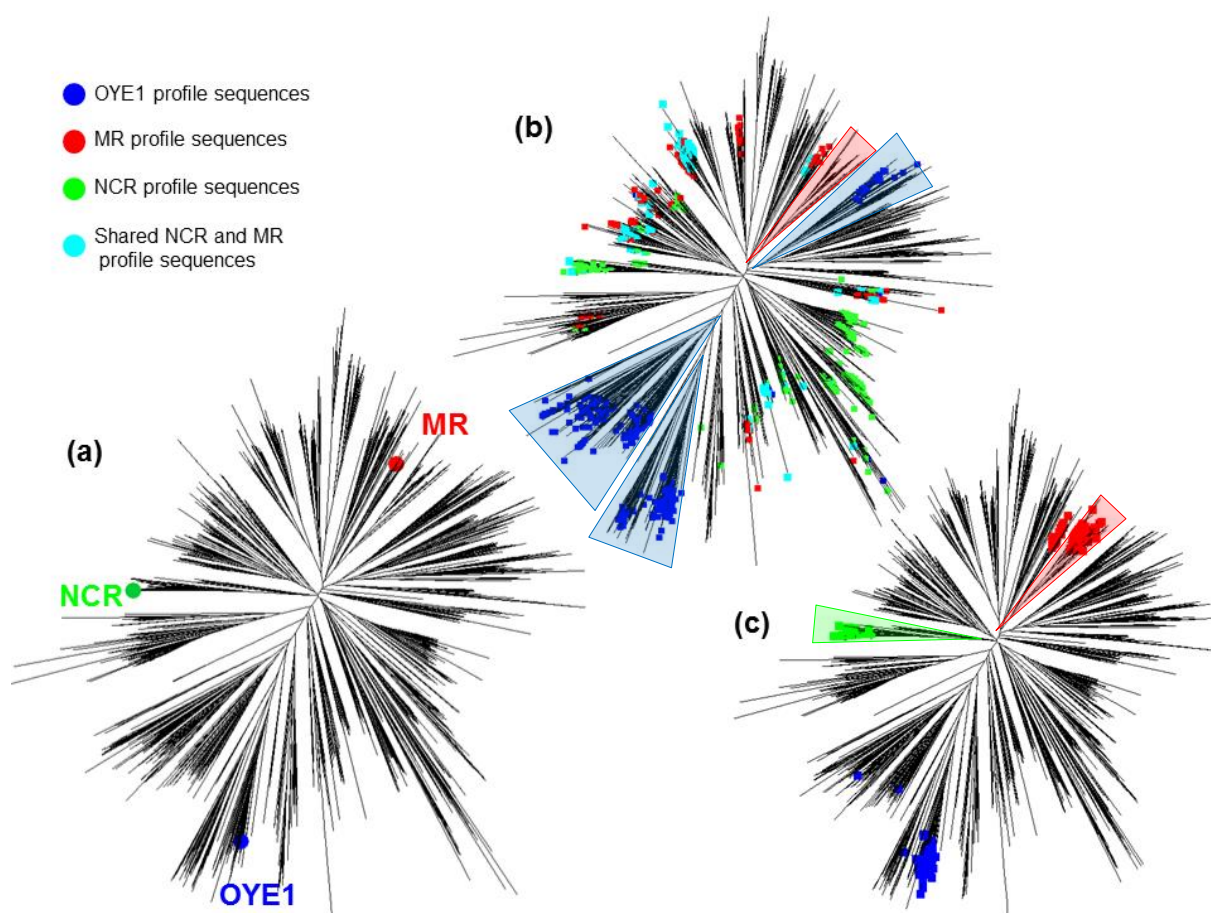


Figure 3.10: Phylogenetic tree of 2258 ene reductases of the HF1 subfamily of the Old Yellow Enzymes. Each sequence is represented by one line. (a) Localization of the three reference sequences NCR (green), MR (red) and OYE1 (blue) within the phylogenetic tree. (b) Localization of all 865 sequences building the Loop A profiles. (c) Localization of all 229 sequences building the Loop B profiles. Color code: NCR profile sequences green, MR profile sequences red, sequences present in the NCR and MRE profile cyan and OYE1 profile sequences blue.

(3) β/α Loops are classified in loops with a generalist type motif and loops demonstrating a specialist motif. If one now regards the distribution of the 229 sequences forming the three Loop B profiles, it is obvious that they demonstrate another behavior than the Loop A profile building ones. The Loop B profiles are formed by sequences that are restricted to a very tight and closely related sequence space surrounding the corresponding reference sequence. For example, the 40 sequences of the NCR Loop B profile are all located at the same small branch as the NCR reference sequence (figure 3.10c, highlighted in green). This loop profile is therefore a specialist motif which is restricted to a very small part of the HF1 subfamily. All 40 sequences forming the Loop B motif of NCR are already included in the generalist type motif of the Loop A profile. The same outcome is obtained with the OYE1 Loop B sequences. The OYE1 Loop B forming sequences are also representing a specialist motif which is already present in the generalist motif of the corresponding Loop A profile. The MR Loop B forming sequences, however, possess a slightly modified behavior. All 116 sequences are also located next to the MR reference sequence representing a specialist motif, though not all of them are included in the generalist motif of the Loop A profile. A small branch next to the MR reference sequence consists of sequences included in the Loop B MR profile but not in the corresponding Loop A profile (figure 3.10 b/c, highlighted by red triangles). For the Loop B profile of MR one should, however, keep in mind that there are differences in the determination of the total loop length between the secondary structure prediction used for this analysis and the crystal structure loop prediction. Therefore, if you would elongate the Loop B profile by three amino acid positions of the crystal structure, you would obtain fewer sequences for the profile, which would probably be even closer to the MR reference sequence.

Summarizing this section, one can conclude that various β/α surface loops within one enzyme can have different loop motifs according to which they can be allocated to a generalist or a specialist motif. Now the question will be addressed whether some of the amino acids present in such loop regions are important for the activity of the enzyme.

Alanine scanning mutagenesis of Loop A of NCR: For the investigation of the functional importance of amino acids in β/α loop regions the Loop A of NCR was selected as studying object for performing an alanine scanning mutagenesis

(ASM).^{134–137} In addition to the 14 amino acids included in the Loop A profile the subsequent two amino acids P239 and E240 located in the following α -helix were also included. Therefore, the relevance of amino acids anchoring a flexible loop region in the leading and following secondary structure element were also taken into account. In total 16 alanine scanning mutagenesis variants were created and tested in the reduction of one standard substrate for ene reductases, ketoisophorone. The levodione product formation of the 16 ASM variants (L225A – E240A), as well as the wild type NCR is illustrated in figure 3.11a.

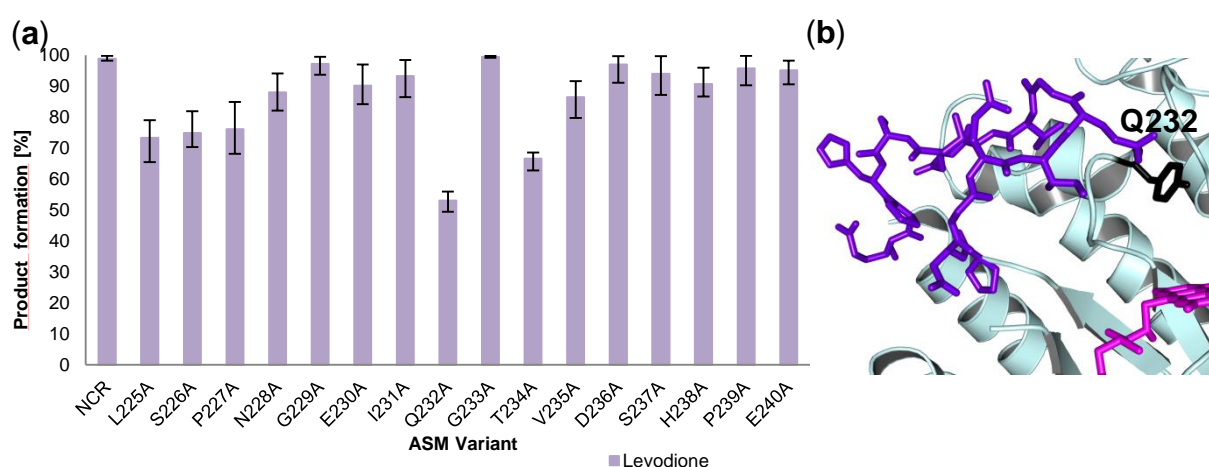


Figure 3.11: (a) Product formation in the reduction of the standard substrate ketoisophorone with the created ASM variants. Reactions were performed in triplicates with 100 μ g purified protein (protein purity > 90 %), 10 mM ketoisophorone and 10 mM NADH in a final volume of 1 ml Tris-HCl pH 7.5 and run at 30°C with 180 rpm for 2.5 h. (b) Representation of the Loop A region of NCR in a stick representation colored in purple/blue with the catalytic influencing Q232. The catalytic active tyrosine is colored in black and the FMN in magenta.

As expected, the exchange of each of the three conserved N-terminal hinge amino acids L225A, S226A and P227A led to a decrease in the reduction activity of about 20 %. Interestingly, the largest effect of the alanine exchange was demonstrated by two amino acids located in the flexible part of the loop region, Q232 and T234. When Q232 was exchanged against an alanine, the enzyme activity was decreased by 50 %. This result could be explained by taking a deeper look into the crystal structure of the ene reductase NCR (figure 3.11b), where the Q232 is located in close proximity to the catalytic active and important tyrosine 177 of the enzyme. The exchange Q232A created more space next to the active center, which influences the enzyme activity. The same assumption could also be stated for T234, which is located in the same loop area as Q232. The exchange G233A in-between Q232 and

T234 did not demonstrate the same behavior due to the fact that the exchange glycine against an alanine is leading to a barely steric difference. Mutations in the C-terminal part of the loop region showed the lowest influence on the activity of the enzyme. Furthermore, the mutation D236A representing the C-terminal conserved aspartic acid demonstrated no activity decrease. Therefore, mutations in this part of the loop were more tolerated. Based on the alanine scanning mutagenesis data, one can conclude that N-terminal conserved hinge residues, as well as residues located in the middle part of the loop and therefore in close neighborhood to the catalytic important amino acids, are important for the enzyme activity. The amino acid composition of β/α surface loop regions possesses an influence on the overall enzyme activity.

3.2.3 Design of rational loop variants

In the analysis of β/α surface loop regions of ene reductases, it was so far possible to show that these loops differ in the amino acid composition, in loop length, as well as in their secondary structure arrangement. In addition, they can be classified in different loop profile motifs and consist of flexible and catalytic influencing amino acids. Now the impact of β/α surface loop regions on enzyme properties will be investigated in a more global approach addressing the questions, whether the loop length or the amino acid composition demonstrate a major role on enzymatic properties and whether it is possible to transfer properties between different ene reductases by the grafting of their loop regions. Therefore, two different approaches for the design of rational NCR loop variants have been chosen:

- (1) The creation of rational loop length variants of intrinsic NCR loops by being elongated or shortened in the overall loop size based on multiple sequence and structural alignments in order to determine the influence of the loop length on enzyme properties.
- (2) The creation of loop grafting variants based on the exchange of the defined loop regions A and B from NCR against the corresponding loop regions of OYE1 or MR in order to specify the impact of the amino acid composition, and the question whether enzyme properties can be transferred between different family members by loop grafting.

In total, seven loop variants were created, four variants of the Loop A region and three variants of the Loop B region consisting of three loop length and four loop grafting variants. The four variants of the Loop A region were composed of two length and two grafting variants (table 3.6).

Table 3.6: Amino acid composition and loop length of the Loop A region in the three wild type enzymes NCR, MR and OYE1, as well as in the four designed Loop A variants.

Wild type enzymes & Loop A variants		Loop A amino acid sequence	Loop length
Wild type enzymes	<i>NCR</i>	²²⁵ LSPNGEIQGTVD ²⁴⁰ SHPE	16 aa
	<i>MR</i>	²³⁹ LTPFLELFGLTD ²⁵⁴ DEPE	16 aa
	<i>OYE1</i>	²⁴⁵ LSPYGVFNSMSGGAETGIV ²⁶³	19 aa
NCR loop length variants	<i>Loop A_Short</i>	²²⁵ LSPNGSHPE ²³³	9 aa
	<i>Loop A_Long</i>	²²⁵ LSPNGEIQAAAGTV ²⁴² DSHPE	19 aa
NCR loop shuffling variants	<i>Loop A_MR</i>	²²⁵ LSPFLELFGLTD ²⁴⁰ DEPE	16 aa
	<i>Loop A_OYE1</i>	²²⁵ LSPYGVFNSMSGGAETGIV ²⁴²	19 aa

First, based on the crystal structure of NCR wild type a variant was created which demonstrated a shortened Loop A compared to the wild type (Loop A_Short). Therefore, the seven amino acids located closest to the catalytic active tyrosine were deleted leading to a shortened loop of just nine amino acids generating conspicuously more space at the entrance of the active site (figure 3.12a and b). Taken the fact into account that yeast ene reductases like OYE1 and OYE2 possess a loop A region of 19 amino acids in length, a loop length extension variant of NCR was created, demonstrating also a length of 19 amino acids by insertion of three alanines in-between the catalytic influencing Q232 and the neighboring G233 (figure 3.12c).

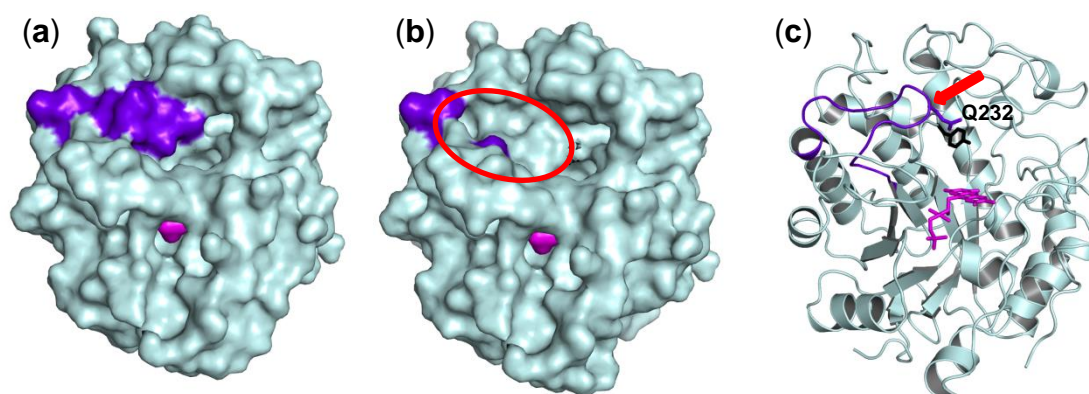


Figure 3.12: Representation of NCR wild type colored in pale cyan (pdb file 4a3u). Loop A is colored in purpleblue, FMN in magenta and the catalytic active tyrosine in black. **(a)** Illustration of the surface of NCR. **(b)** Illustration of the surface of loop length variant Loop A_Short. Highlighted in red is the area where more space is generated. **(c)** Cartoon representation of NCR, including Q232 in a stick representation. The red arrow indicates the position for the insertion of three alanine residues for the creation of the variant Loop A_Long.

Next to the two rational loop length variants also two loop grafting variants for the Loop A region were designed. Therefore, the defined Loop A of NCR was exchanged against the corresponding Loop A of MR leading to the variant Loop A_MR possessing the same loop length as NCR wild type but a different amino acid composition. Loop A of MR exhibits almost the same loop backbone structure as the corresponding NCR region, however it consists of the clearly different amino acid composition (figure 3.13a). So is, for example, the catalytic influencing Q232 of NCR exchanged against an aromatic phenylalanine (F246).

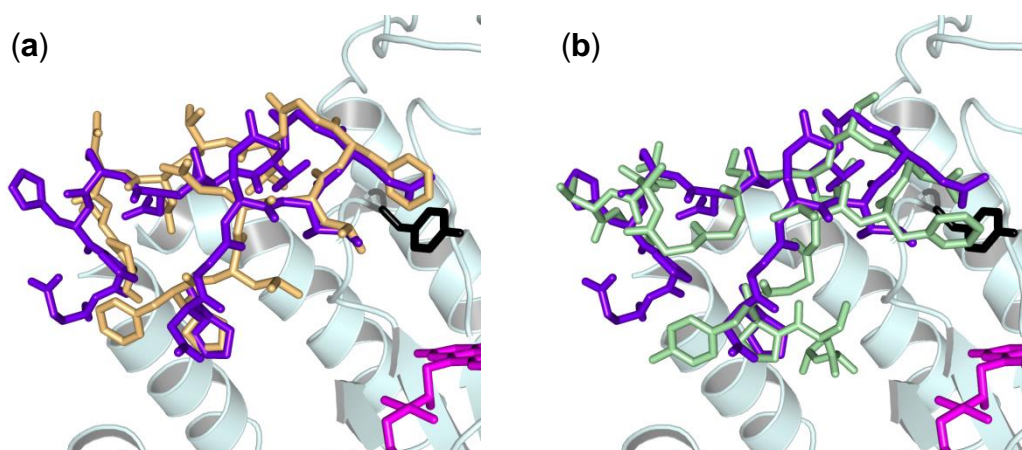


Figure 3.13: Structural superposition of NCR Loop A region (purpleblue) in a stick representation with the corresponding loop regions of **(a)** MR (colored in light orange) **(b)** OYE1 (colored in pale green). The prosthetic FMN is colored in magenta; the catalytic active Y177 is colored in black.

This loop grafting led additionally to a different charge, due to the two further negatively charged acidic amino acids in Loop A of MR, which are localized at the C-terminal end of the loop region. As second loop grafting variant Loop A_OYE1 was created by exchanging the 16 amino acid NCR loop against the corresponding 19 amino acid long OYE1 Loop A. This variant combines the two applied rational loop design approaches by elongating the loop length about three amino acids, as well as changing the amino acid composition (figure 3.13b). The variant Loop A_OYE1 also demonstrated an altered charge, in this case a decreased one, compared to wild type NCR. It possesses in total just one negatively charged amino acid (E259), while NCR has three (E230, D236, E240). None of the three Loop A regions possesses a positively charged amino acid like lysine or arginine. Only Loop A of NCR includes one histidine.

Table 3.7: Amino acid composition and loop length of the Loop B region of the three wild type enzymes NCR, MR and OYE1 as well as the three designed Loop B variants.

Wild type enzymes & Loop B variants		Loop B amino acid sequence	Loop length
Wild type enzymes	NCR	²⁶⁰ MREGAVDGTFGKTDQPKLS ²⁷⁸	19 aa
	MR	²⁷⁴ FNEPDWIGGDITYPEGFR ²⁹¹	18 aa
	OYE1	²⁸⁸ LVEPRVTNPFLTEGEGEYEGGSN ³¹⁰	23 aa
NCR loop length variants	Loop B_Short	²⁶⁰ MREGAVDKTDQPKLS ²⁷⁴	15 aa
NCR loop shuffling variants	Loop B_MR	²⁶⁰ FNEPDWIGGDITYPEGFR ²⁷⁷	18 aa
	Loop B_OYE1	²⁶⁰ LVEPRVTNPFLTEGEGEYEGGSN ²⁸²	23 aa

For the investigation of the Loop B region a total of three loop variants were created, one loop length and two loop grafting variants (table 3.7). Since the Loop B region represents a more diverse loop region in-between different ene reductase family members, an additional characteristic of this loop region is taken in account for the

generation of a loop length variant, its high B-factor values. Apart from the C-terminus of the enzyme, the Loop B region demonstrates the highest B-factor values of NCR, representing the four most flexible amino acids (G267, T268, F269 and G270) of the entire enzyme (figure 3.14c). For the loop length reduction variant Loop B_Short solely these four amino acids were cut out, leading to a more ridged enzyme and also more space next to the prosthetic FMN (figure 3.14b).

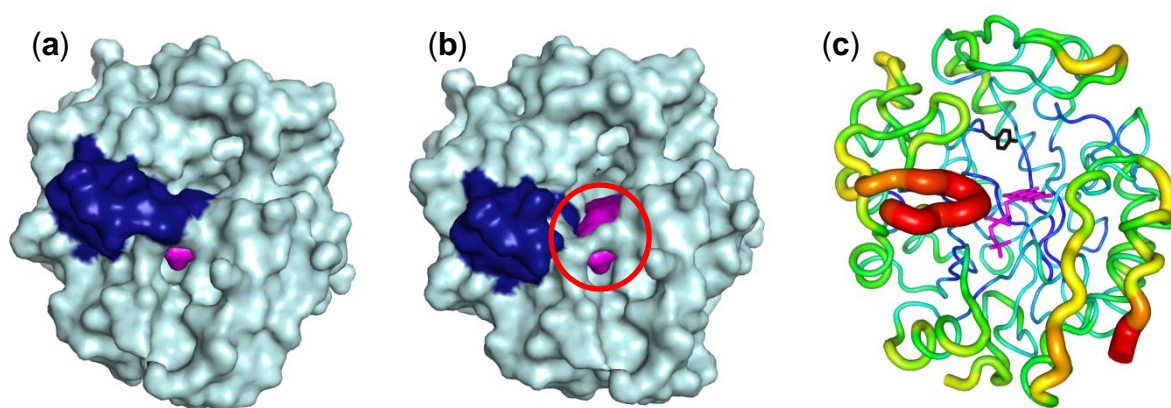


Figure 3.14: Representation of NCR wild type colored in pale cyan (pdb file 4a3u). Loop B is colored in dark blue, FMN in magenta and the catalytic active tyrosine in black. (a) Illustration of the surface of NCR. (b) Illustration of the surface of loop length variant Loop B_Short. Highlighted in red is the area where more space is generated. (c) B-factor representation of NCR. Loop B represents the part of the enzyme possessing the largest B-factors indicated by thickness of the structure and the red color.

For Loop B so far no loop variant with an extended length was created. As for Loop A, next to the rational loop length variants also two loop grafting variants for the Loop B region were designed. Therefore, again the defined Loop B of NCR was exchanged against the corresponding loop of MR and OYE1, respectively. In contrast to Loop A, where NCR and MR have the same loop length, the Loop B region of MR is one amino acid shorter than the one from NCR. Once more, OYE1 demonstrates the largest loop with 23 amino acids representing an elongation of four amino acids for the variant Loop B_OYE1 compared to wild type NCR. If one now compares the amino acid composition of the two inserted loops from OYE1 and MR with the loop of NCR, it attracts attention that both demonstrate a total different amino acid composition, as well as structural orientation (figure 3.15). Loop B of NCR possesses three acidic amino acids (E262, D266, D273), as well as three basic ones (R261, K271, K276). In contrast to that, MR includes four negatively charged acidic amino acids (E276, D278, D283, E288), but only one positively charged one (R291) leading

to an overall more acidic loop region. Additionally, the MR Loop B contains the rarest amino acid within an enzyme, a tryptophan (W279), which is located closest to the prosthetic FMN (figure 3.15a). Looking now at the amino acid composition of Loop B from OYE1, it is apparent that it contains a quite acidic posterior loop part with four glutamic acids in close proximity to each other (E300, E302, E304, E306) plus one more acidic (E290), as well as one basic (R292) amino acid. The closest amino acid for the prosthetic FMN is, like in the case of NCR (F269), a phenylalanine (F296 of OYE1), though in a different structural orientation (figure 3.15b). As a result of that amino acid composition, the loop variant Loop B_OYE1 exhibits a longer and more acidic Loop B than the one of wild type NCR.

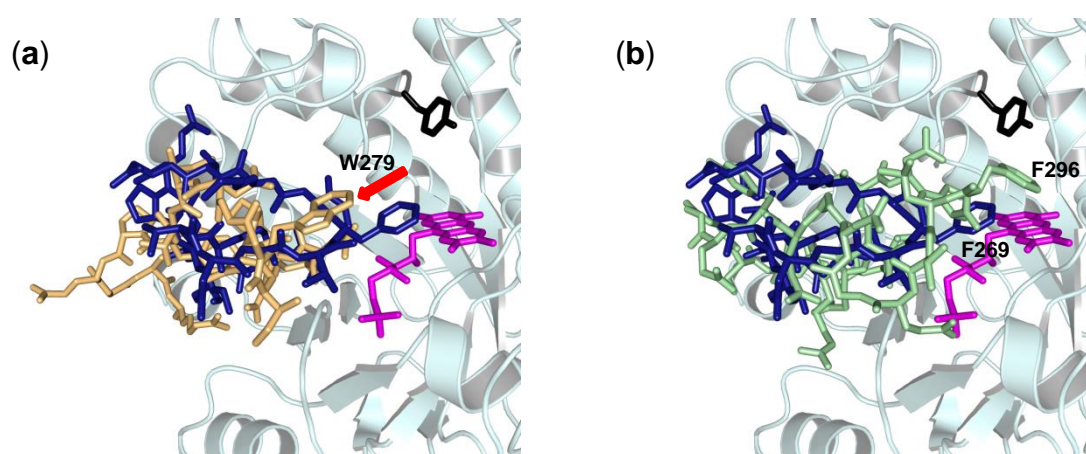


Figure 3.15: Structural superposition of NCR Loop B region (dark blue) in a stick representation with the corresponding loop regions of (a) MR (colored in light orange). Marked with a red arrow is the rare amino acid tryptophan W279 (b) OYE1 (colored in pale green). The prosthetic FMN is colored in magenta; the catalytic active Y177 is colored in black.

Beside the differences in the so far discussed loop parameters such as loop length, amino acid composition and charge between NCR and the created rational loop variants within the new loop regions, also structural alterations of the enzyme variants are of interest. To address that question, the crystal structure of the loop grafting variant Loop A_OYE1 was solved together with the cooperation partner from the Structural Analysis of Biopolymers group at the Institute of Bioanalytic Chemistry from the University of Leipzig. This variant was chosen for crystallization due to the fact that it combines the two rational design approaches of loop length variation and loop amino acid composition alteration. After expression in *E.coli* BL21 (DE3) and a two-step purification the highly pure enzyme solution was passed to the cooperation partner at the University of Leipzig which solved the crystal structure with a resolution

of 1.8 Å. For detailed crystallographic information, as well as data collection and refinement statistics see supplementary material section 7.3.1 and table 7.9. The loop grafting variant still possesses a TIM barrel as structural scaffold, which is expected since the changes are restricted to one surface loop region. Notwithstanding, there are several differences present in the crystallography of the loop grafting variant Loop A_OYE1. First of all, the variant forms crystals at different conditions compared to wild type NCR.⁹ Secondly, the TIM barrel itself demonstrates an even higher rigidity than the wild type enzyme, which can be seen in the lower B-factor values (supplementary material figure 7.6). And finally several remarkable differences in the catalytic interface of the ene reductase variant concerning all three defined β/α surface loop regions are observable resulting in the following three statements graphically displayed in figure 3.16:

(1) By the grafting of Loop A all three defined β/α surface loop regions (A, B and C) become more flexible. This statement can be proven by the fact that due to the high flexibility of the loop building amino acids, it is not possible to solve the crystal structure of the three β/α loops completely. Thus, there are four areas within the catalytic interface without a detectable electron density. Affected are the following amino acid (aa) positions: aa 63-64 (figure 3.16 (1), green), aa 108-140 (light blue), aa 185-192 (blue) and aa 234-241 (red). The last area (aa 234-241) is located within the grafted loop region. Interestingly, not only the grafted loop region alone becomes more flexible, also the large Loop C region is affected, even in two areas. Based on that result one can learn that changes in one specific β/α surface loop region can lead to more flexibility within the complete catalytic interface.

(2) The grafting of Loop A leads to a different orientation of the catalytic active tyrosine. By the exchange of the Loop A region of NCR against the corresponding one from OYE1 the entire loop region becomes more flexible. Furthermore, the catalytic influencing Q232, which is closely located to the catalytically important tyrosine 177, is replaced by an aromatic phenylalanine. The inserted phenylalanine is just 3.4 Å away from the catalytic Y177. This amino acid substitution leads to a change in the orientation of the catalytic tyrosine compared to wild type NCR (figure 3.16 (2)). In the grafting variant Loop A_OYE1 the hydroxyl group of Y177 points away from the prosthetic FMN, which should result in a decreased efficiency of the, for the reduction activity necessary, proton transfer from the tyrosine onto the C_α of the carbon-carbon double bond.

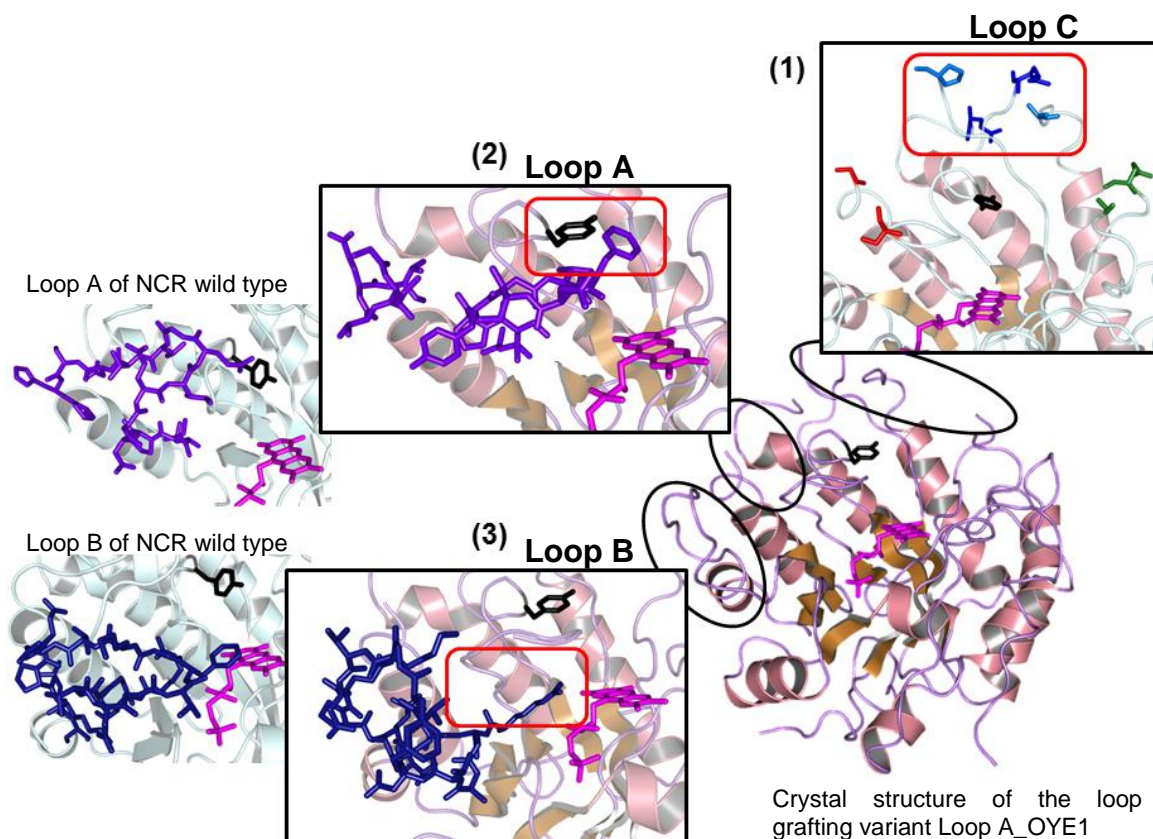


Figure 3.16: Visualization of the crystal structure of the grafting variant Loop A_OYE1 with accentuation of the three main differences within the catalytic interface of the enzyme. (1) Highlighting of the four unsolved areas indicated by the four different colors. Each amino acids pair in the same colored stands for the beginning and the end of one unsolved area. The two largest of them are located within the Loop C. (2) Different orientation of Y177 compared to wild type NCR colored in pale cyan. (3) Altered localization of the loop B region compared to wild type NCR. Loop A is colored in purpleblue; Loop B is colored in dark blue; FMN is colored in magenta, Y177 is colored in black. The differences are marked by a red frame.

(3) The grafting of Loop A leads to an altered localization of the Loop B region within the enzyme variant. Due to the larger size and the altered orientation of the inserted OYE1 Loop A in the grafting variant, also the adjacent Loop B region demonstrates a different localization even though it consists of the same amino acids. In wild type NCR Loop B is folded towards the prosthetic FMN with F269 being situated closest to it. By contrast, the variant Loop A_OYE1 exhibits a Loop B which is turned away from the FMN, leading to the fact that the last amino acid of the previous β -sheet, R264, is the closest amino acid to the active center. Interestingly, the B-factor values are modified as well. Like already explained, the Loop B of wild type NCR possesses the four amino acids with the highest flexibility within the whole

enzyme, whereas the Loop B of the variant exhibits clearly lower B-factor values indicating more rigidity.

To sum up this section, a total amount of seven rational loop variants was designed by two different approaches in order to investigate the influence of loop length, amino acid composition and structural orientation on several enzyme properties. Additionally, it was possible to show that the grafting of one β/α surface loop region resulted in large structural alterations of the entire catalytic interface. All created variants were cloned in pET28a(+) and expressed in TB media with *E. coli* BL 21 (DE3).

3.2.4 Biotransformation with the designed loop variants

The three wild type enzymes NCR from *Z. mobilis*, OYE1 from *S. pastorianus* and MR from *P. putida* M10, as well as the seven created rational loop variants were expressed in soluble form in *E. coli* BL21 (DE3) and purified via an immobilized metal affinity chromatography yielding protein concentrations between 3.4 and 25.9 mg/ml protein solution with protein purities > 90 % (for detailed information see supplementary material table 7.10 & figure 7.7). The different levels of protein concentrations were on the one hand caused by different protein solution volumes after the concentration via ultrafiltration and on the other hand by different expression volumes. For all biotransformation reactions NADH was used as cofactor without a cofactor regeneration system. The reactions were analyzed via GC FID and GC MS.

3.2.4.1 Reduction activity towards α,β -unsaturated substrates of the designed loop variants

In a first step the question was addressed whether the designed rational loop variants were still active in the reduction of activated carbon-carbon double bonds. Therefore, altogether six standard substrates were tested with the seven variants and the three wild type enzymes: Two cyclic ones, comprised of ketoisophorone and cinnamaldehyde, as well as four aliphatic substrates varying in size from the small

2-methyl-2-pentenal over 2,4-heptandienal towards the challenging substrate citral, which consists of the two isomers neral and geranial.

Based on the results shown in table 3.8 it is possible to confirm that all created variants remain active in the reduction of α,β -unsaturated substrates without any byproduct formation, however with partly reduced activity.

Table 3.8: Product formation in the reduction of known α,β -unsaturated substrates catalyzed by wild type enzymes, as well as the created loop variants.

Substrate/ Enzyme	Product formation in [%]						
	Ketoiso- phorone	2-Methyl- 2-pentenal	Cinnam- aldehyde	2,4-Heptan- dienal	Neral	Geranial	
Wild type enzymes	<i>NCR</i>	79.9±2.6	14.8±7.9	67.6±1.9	27.9±2.5	25.9±8.9	32.5±6.7
	<i>OYE1</i>	19.3±0.8	10.8±2.6	7.7±0.9	4.6±0.8	0.4±0.1	4.2±1.2
	<i>MR</i>	29.3±3.3	30.7±4.0	14.7±4.5	0.4±0.02	7.1±0.8	3.5±0.3
Loop A variants	<i>Loop A_Short</i>	68.3±6.4	21.1±6.1	34.4±3.0	3.7±0.5	29.6±1.0	23.9±3.2
	<i>Loop A_Long</i>	43.2±2.9	23.6±0.2	11.4±1.9	0.9±0.3	9.8±4.2	5.6±0.5
	<i>Loop A_OYE1</i>	35.5±0.1	13.3±1.1	9.2±2.5	0.3±0.04	3.0±0.7	1.7±0.3
	<i>Loop A_MR</i>	71.6±3.3	17.0±3.0	10.0±2.7	0.7±0.4	20.4±1.4	4.7±0.9
Loop B variants	<i>Loop B_Short</i>	52.8±1.4	17.0±1.0	20.0±2.5	7.8±0.7	39.9±1.8	21.2±1.3
	<i>Loop B_OYE1</i>	87.9±3.6	29.9±3.5	16.9±2.8	1.1±0.1	11.8±3.1	10.3±3.4
	<i>Loop B_MR</i>	33.1±4.7	37.2±3.2	12.5±0.7	0.9±0.6	8.6±2.6	6.3±2.6

Reactions were performed in triplicates in a final volume of 1 ml 50 mM Tris-HCl pH 7.5 and run at 30°C with 180 rpm for 2.5 h.

From the three tested wild type enzymes NCR showed the broadest substrate spectrum being active towards all substrates with a preference for the two cyclic substrates. For the reduction of cinnamaldehyde NCR is even the best of all tested biocatalysts. However, for the also cyclic ketoisophorone the shuffling variant

Loop B_OYE1 demonstrated a slightly increased activity compared to NCR. If one now looks at the aliphatic substrates, starting with the shortest 2-methyl-2-pentenal you can see that in particular the two Loop B grafting variants demonstrated a considerable increased activity compared to NCR wild type. The best catalyst for this substrate was the variant Loop B_MR, which possesses a 2.5 fold increase of activity in comparison with NCR. Interestingly, MR wild type had also an enhanced product formation in the reduction of 2-methyl-2-pentenal, indicating that by the grafting of the Loop B region it was possible to transfer that ability in the NCR scaffold. For the second aliphatic substrate, 2,4-heptandienal, the wild type NCR demonstrated once more the highest product yield. All variations in both loop regions lead to a drastic decrease in the enzyme activity. The last aliphatic substrate citral, consisting of the two isomers neral and geranial, however, indicated some interesting influence of the loop regions on the performance of the enzyme. First of all, it could be shown that NCR is able to reduce both isomers, with a slight preference for the *trans* isomer, geranial. The other two wild type enzymes were less active; however MR demonstrated a preference for the *cis* isomer neral. If one now looks at the conversions of the designed loop variants, two of them attracted attention, Loop A_MR and Loop B_Short. The Loop B length reduction variant exhibited a reduction of citral in the same range as NCR wild type; however it demonstrated an increase in the activity towards *cis* citral and a decrease for the *trans* citral reduction leading to a twofold *cis* preference of the enzymes. The change of the substrate specificity from *trans* towards *cis* was even higher with the grafting variant Loop A_MR. This variant demonstrated a fivefold *cis* over *trans* selectivity obtained by a dramatically decrease of the reduction of geranial. Based on these results, one can assume that for the *cis/trans* substrate specificity of the enzyme in the case of citral, the amino acid composition of Loop A, as well as the length of the Loop B play a critical role.

Next to the reduction capability also the enantioselectivity of the formed reaction products was of interest, in particular in the case of ketoisophorone and 2-methyl-pentenal (for detailed information see supplementary material table 7.11). For both substrates all three wild type enzymes possessed the same enantiopreference: ketoisophorone is reduced to (*R*)-levodione and 2-methyl-2-pentenal to (*S*)-2-methylpentanal. However, the three wild types exhibited different enantiomeric excesses (*ee*): NCR and OYE1 produces (*R*)-levodione with a high enantiomeric

excess of 96 % and 95 %, respectively, while MR showed only a slight (*R*)-preference with an enantiomeric excess of 19 %. In the case of 2-methyl-2-pentenal, OYE1 demonstrated the best enantiomeric excess of 82 %, while NCR and MR are moderate (*S*)-selective with an *ee* of 31 % and 49 %, respectively. If one now looks at the enantiomeric excesses of the Loop A variants, only one of them showed an improving influence. The loop grafting variant Loop A_OYE1 had an improved *ee*-value of 67 % in the formation of (*S*)-2-methylpentanal compared to wild type NCR. The greater influence of the enantioselectivity can be detected in the rational variants of the Loop B region. Thus, in the reduction of ketoisophorone the grafting variant Loop B_MR demonstrated a diminished (*R*)-selectivity with an *ee* of only 13 %, which is in the range of the MR wild type. The grafting of the Loop B region from MR to NCR led to a transfer of the enantiomeric excess in the reduction of ketoisophorone. For the small substrate 2-methyl-2-pentenal it could be shown that the grafting of the Loop B of OYE1 into NCR resulted in the increase of the enantioselectivity. The variant Loop B_OYE1 produced exclusively the (*S*)-enantiomer of 2-methylpentanal. The other two Loop B variants, both possessing a reduced loop length compared to NCR, exhibited a clearly decreased enantioselectivity. Loop B_Short even demonstrated a minimal (*R*)-preference (*ee*-value 2.6 for (*R*)-2-methylpentanal). Thus, it seems reasonable to assume that the length of the Loop B is important for the enantioselectivity in the reduction of the small substrate 2-methyl-2-pentenal. An increased loop length, in the case of Loop B_OYE1 resulted in a considerably improved *ee*-value while the reduction of the length led to a decrease.

In conclusion, it was possible to show that variations in β/α loop regions of ene reductases resulted in enzyme variants, which are still active in the reduction of α,β -unsaturated substrates. Furthermore, it could have been demonstrated that the Loop A region, which is located at the entrance of the active site, plays a role in the *cis/trans* specificity of the enzyme. Additionally, the Loop B region possesses, depending on the substrate, an influence on the activity, as well as the enantioselectivity of the enzyme.

3.2.4.2 Solvent tolerance of the designed loop variants

Next to the reduction activity and enantioselectivity, the stability of an enzyme towards organic solvent and increased reaction temperatures is a desired property of a biocatalyst in terms of industrial application. Therefore, first the solvent tolerance of the rationally designed loop variants was investigated with four different organic solvents (acetone, isopropanol, ethylacetate and tetrahydrofuran) at four concentrations (5, 10, 20 and 30 % v/v) in the reduction of the standard substrate ketoisophorone. Due to the fact that for all organic solvents tested related behaviors could be observed, figure 3.17 visualizes exemplarily the results for the conversion rates with acetone and isopropanol (for ethylacetate and tetrahydrofuran see supplementary material figure 7.8).

If one looks at the three wild type enzymes it is possible to see that NCR demonstrated for all solvents tested the best solvent tolerance. At low solvent concentrations (until 10 % v/v) the conversion of ketoisophorone even increased. This can be explained by the better solubility of the hydrophobic substrate and the stability of NCR in the reduced medium polarity. The best conversion rates for NCR were obtained with 10 % v/v acetone or THF. The other two wild type enzymes demonstrated a different behavior. Both of them showed compared to NCR a considerably diminished activity towards ketoisophorone even without any solvent present. However, MR exhibited a remarkable tolerance towards isopropanol with a decrease in activity of only 15 % at 30 % v/v solvent compared to a reaction without any solvent present. Therefore, MR represented the best tolerance towards one of the tested organic solvents. In contrast, OYE1 did not exhibit any solvent tolerance. The best conversion with this wild type was achieved without any solvent present in the reaction mixture. Based on these results one can conclude that the solvent tolerance is depending on the used enzyme as well as the applied solvent. Each enzyme possesses a different optimal solvent and concomitant solvent concentration.

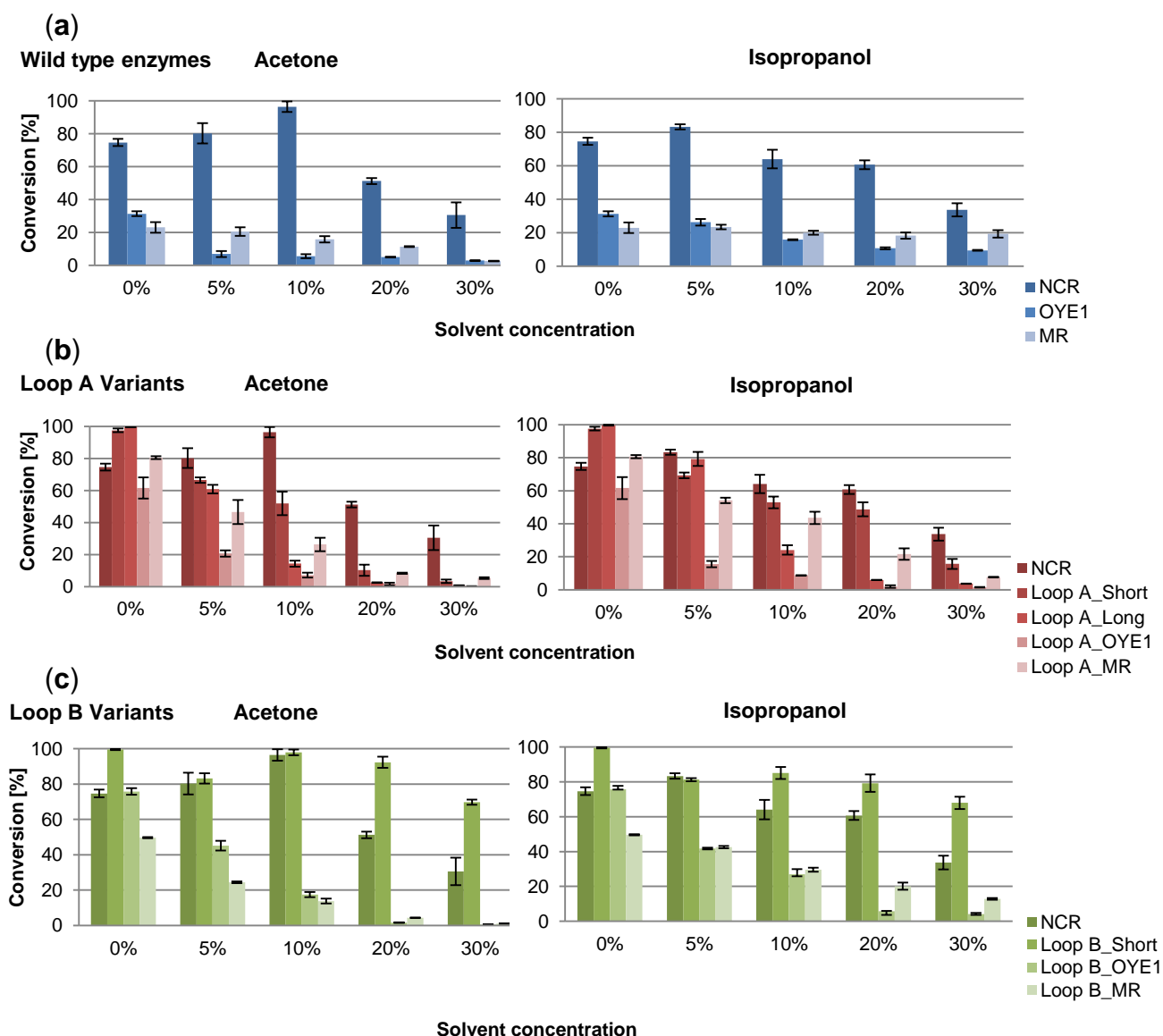


Figure 3.17: Conversion of ketoisophorone at different solvent concentrations ranging from 0 – 30 % v/v organic solvent. Shown are the results with two solvents acetone (0.12) and isopropanol (0.19) (a) The three wild type enzymes NCR, MR and OYE1 colored in blue (b) The four created Loop A variants in comparison with NCR colored in red (c) The three Loop B variants compared to NCR colored in green. The reactions were performed in triplicates at 30°C with 180 rpm for 4 h in 1 ml 50 mM Tris-HCl pH 7.5. LogP values are indicated in brackets.

If now the influence of the rationally designed loop variations on the solvent tolerance of the enzyme is examined, one can observe that all Loop A variants demonstrated a clearly enhanced solvent sensitivity compared to wild type NCR. Already 10 % v/v organic solvent present in the reaction mixture resulted for these four variants in at least 45 % activity loss. Thus, changes whether in loop length or amino acid composition in the Loop A region result in less solvent tolerant enzyme variants. For the grafting variants of the Loop B regions, Loop B_OYE1 and Loop B_MR, the same behavior could be observed. However, the loop length variant

Loop B_Short demonstrated an entirely different performance towards all organic solvents tested. Both, at low or high solvent concentrations, Loop B_Short exhibited a considerable increase of the solvent tolerance compared to wild type NCR. This variant was able to convert 60 % of ketoisophorone in the presence of 30 % v/v acetone, isopropanol, as well as ethylacetate.¹⁰⁸

These results led to the statement that the length of the Loop B region plays an important role in the solvent tolerance of the enzyme. The deletion of the four amino acids possessing the highest flexibility located in that loop region result in a more ridged protein that exhibits a considerable enhanced tolerance towards all tested organic solvents.

3.2.4.3 Thermostability of the designed loop variants

The second stability aspect, which is addressed within this thesis, is the influence of the rational designed loop variants on the thermostability of the enzyme. Therefore, two approaches haven been chosen: (1) The determination of the melting temperature with the ThermoFAD method¹¹⁷ and (2) the performance of biotransformation reactions at elevated temperatures. The ThermoFAD method is a fast and reliable method depending on the intrinsic fluorescence of the flavin cofactor. The FMN flavin cofactor possesses excitation maxima between 373-375 and 445-450 nm and an emission maximum at 535 nm. In a properly folded enzyme the fluorescence of the flavin is quenched by the protein environment. However, the raising of the temperature leads to the unfolding of the protein and therefore to an exposure of the flavin to water resulting in an increase in the fluorescence signal. As a result, by plotting the fluorescence signal against the temperature, a sigmoidal curve is obtained, including the unfolding temperature of the protein as the maximum derivation, or in other words, as the inflection point of the curve. Figure 3.18 illustrates the sigmoidal curves of the three wild type enzymes NCR, MR and OYE1 for the determination of their unfolding temperatures. The highest thermostability was presented by MR with 70.8°C, followed by NCR 60.5°C and finally OYE1 54.5°C.

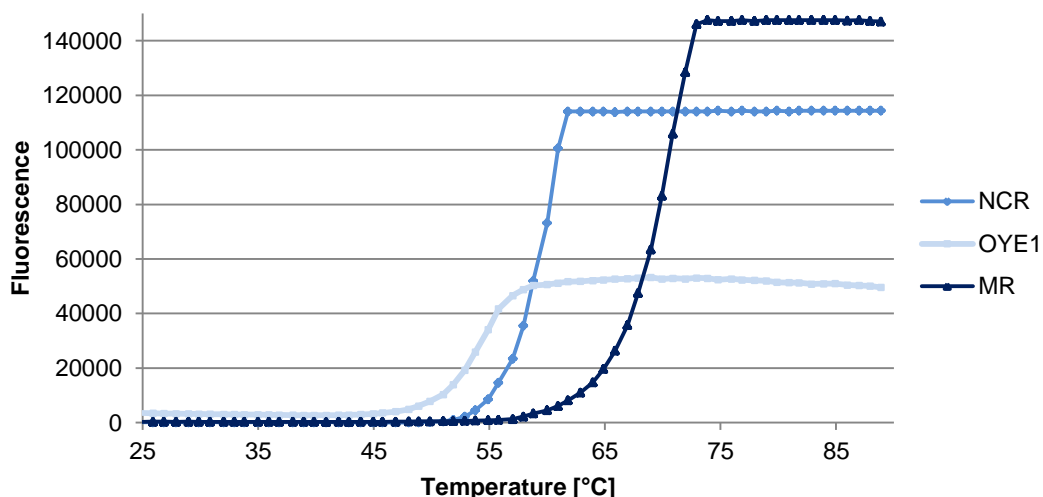


Figure 3.18: Visualization of the FMN fluorescence emission based on the ThermoFAD method depending on the temperature. The fluorescence signal is plotted against the temperature. The point of inflection represents the melting point of the respective enzyme. The amount of the fluorescence signal depends on the used protein content.

If one now looks at the sigmoidal curves and unfolding temperatures (supplementary material table 7.12 and figure 7.9) of the rational designed loop variants it awakens interest that the two most stable ones were the grafting variants which contain the Loop A and Loop B of MR. So it was possible to increase the melting temperature about 6°C by grafting the β/α Loop A or B into the NCR scaffold. This leads to the assumption that such loop regions plays an important role in the thermostability of the ene reductases. All other variants, except Loop B_OYE1 which demonstrated a by 10°C decreased unfolding temperature, were similar to NCR wild type with Loop A_OYE1 being the most stable one (63.4°C).

In a next step it should be clarified if the differences in the unfolding temperature were influencing the performance of biotransformation at elevated reaction temperatures. Therefore, the conversion of ketoisophorone was determined at seven different reaction temperatures (supplementary material table 7.13) and the percentage alteration in product yield between 30°C and an elevated temperature (table 3.9) was calculated. If now firstly the optimal reaction temperature of the three wild type enzymes is defined, it is apparent that the performance of OYE1 was in accordance with the unfolding data possessing the best conversion at 30°C. The medium stable NCR exhibited not one specific optimal reaction temperature, but rather an optimal temperature range from 30-40°C for the reduction of ketoisophorone. The most stable wild type enzyme MR had the best conversion rate

at 37°C reaction temperature; however, the activity remained more or less the same until 50°C. Notwithstanding of an unfolding temperature of over 70°C, MR was not able to reduce ketoisophorone with high yields greater than 55°C reaction temperature.

By considering now the optimal reaction temperature of the two MR loop grafting variants possessing elevated unfolding temperatures, one could observe that Loop B_MR demonstrated exactly the same behavior as wild type MR, with an optimal reaction temperature of 37°C and a quite stable product yield until 50°C. The Loop A grafting variant of MR, however, performed, despite its also increased unfolding temperature, differently. Loop A_MR demonstrated an optimal reaction temperature of 30°C and exhibits already a decrease in product yield about 80 % at 45°C. The other three Loop A variants were also quite sensitive towards increased reaction temperatures resulting in the assumption that the Loop A region is a critical part of the enzyme with regard to elevations of the reaction temperatures.

All applied alterations in Loop A region led to less active variants in reduction of ketoisophorone at higher temperatures despite partially increased unfolding temperatures. Therefore, an elevated unfolding temperature is not directly correlated with an increase of the optimal reaction temperature. This assumption is also supported by the optimal reaction temperatures of the Loop B variants. So Loop B_OYE1 demonstrated the best substrate conversion with an increase of over 120 % in product yield at the elevated reaction temperature of 37°C. Furthermore, Loop_B Short, which exhibited almost the same unfolding temperature as NCR, had the best conversion at 45°C with an increase of over 150 % product yield compared to 30°C reaction temperature.

Table 3.9: Percentage alteration in the product yield in the conversion of ketoisophorone between two different reaction temperatures with the three wild type enzymes as well as the seven designed rational loop variants.

<i>Reaction temperature/ Enzyme</i>	Alteration in product yield between two reaction temperatures in [%]						
	30/37°C	30/40°C	30/45°C	30/50°C	30/55°C	30/60°C	
Wild type enzymes	<i>NCR</i>	-0.2	+0.2	-45.9	-67.5	-79.1	-87.2
	<i>OYE1</i>	-10.6	-56.9	-83.9	-92.5	-95.1	-96.2
	<i>MR</i>	+53.7	+22.7	-7.8	-30.1	-78.8	-88.0
Loop A variants	<i>Loop A_Short</i>	-33.7	-43.1	-84.1	-99.3	-99.4	-99.5
	<i>Loop A_Long</i>	-31.2	-53.4	-80.0	-98.8	-99.2	-99.3
	<i>Loop A_OYE1</i>	+6.7	-9.2	-40.8	-97.0	-97.8	-98.3
	<i>Loop A_MR</i>	-15.4	-29.3	-83.0	-90.9	-95.8	-96.0
Loop B variants	<i>Loop B_Short</i>	+101.3	+108.9	+153.6	+2.4	-41.8	-51.6
	<i>Loop B_OYE1</i>	+124.3	-65.9	-63.2	-91.3	-86.9	-92.5
	<i>Loop B_MR</i>	+68.5	+30.0	+23.1	-3.1	-77.5	-89.0

Reactions were performed in triplicates in a final volume of 1 ml 50 mM Tris-HCl pH 7.5 and run at 30°C with 180 rpm for 4 h.

If one now looks at the stability of the enzyme as a whole, combining solvent tolerance and thermostability, it is apparent that variations in the Loop A region are correlated with a decrease in the overall enzyme stability while alterations in Loop B can result in improved stability properties.

3.2.4.4 Reduction of allyl alcohols with the designed loop variants

So far it could be successfully demonstrated that β/α surface loops have an impact on the enzyme properties activity, *cis/trans* substrate specificity, enantioselectivity as

well as stability. As a next step, the conversion of alternative, less activated substrates were of great interest, due to the fact that recently the thermophilic CrS from *Thermus scotoductus* SA-01¹³⁸, as well as whole cell baker's yeast¹³⁹, were described to be able to reduce allyl alcohols in a bienzymatic cascade reaction together with an alcohol dehydrogenase. The proposed reaction mechanism consists of three steps: (1) The oxidation of the alcohol group to the associated aldehyde with an alcohol dehydrogenase by the use of the cofactor NAD^+ followed by (2) subsequent reduction of the obtained α,β -unsaturated aldehyde with the applied ene reductase and NADH as cofactor and (3) finally the reverse reduction of the aldehyde to the corresponding alcohol with the same alcohol dehydrogenase using NADH as cofactor (figure 3.19 and supplementary material figure 7.10).

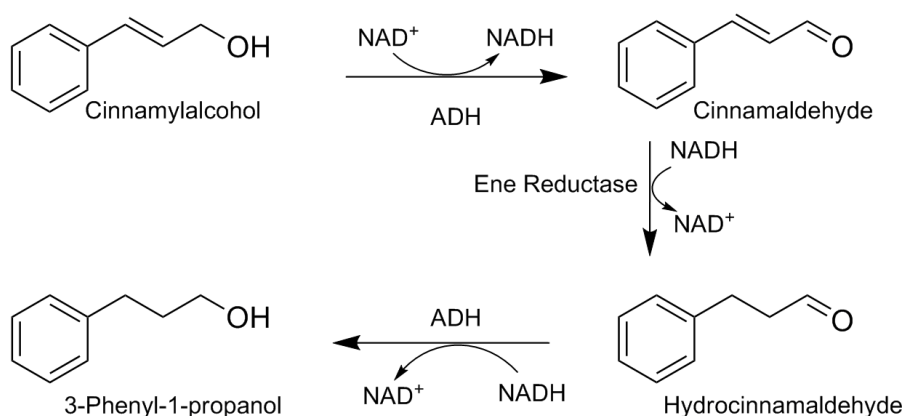


Figure 3.19: Proposed reaction mechanism for the reduction of allyl alcohols with a two enzyme approach based on an ene reductase and an alcohol dehydrogenase (ADH).^{138,139}

Therefore, the question arose whether the wild type enzymes NCR, OYE1 and MR were able to perform the bienzymatic reduction of allyl alcohols in combination with a commercially available alcohol dehydrogenase (ADH equine ≥ 10 U/ml from Sigma Aldrich, Taufkirchen, Germany), which can perform both necessary reactions steps within the cascade reaction. As substrates, three different allyl alcohols were selected: cinnamyl alcohol, geraniol and perillyl alcohol.

It was possible to show that the three wild type enzymes demonstrated a different behavior in the bienzymatic reduction of allyl alcohols. While NCR was not able to perform the reaction with noteworthy conversion rates, the other two wild type enzymes were active in the bienzymatic cascade reaction, however with different conversion rates (table 3.10). OYE1 was the best wild type enzyme tested with almost complete conversion of the allyl alcohols to the corresponding reduced alcohol

for all three substrates. MR demonstrated a high conversion of cinnamyl alcohol to 3-phenyl-1-propanol, but just a moderate formation of the reduced alcohol citronellol in the cascade reaction of geraniol. For this reaction one should, however, keep in mind that the formed intermediate geranial is a poor substrate for MR with only 3.5 % conversion. That fact is supported by the quite high accumulation of geranial intermediate in the cascade reaction (15.8 %) compared to the other wild type enzymes. The most remarkable difference between OYE1 and MR in the conversion of allyl alcohols was present in the conversion of perillyl alcohol. The reduction of perillyl alcohol can result in the formation of two different diastereomers of the alcohol shisool. While OYE1 produced just one of the two shisool diastereomers, further on named as diastereomer 1 (retention time 7.0 min; see supplementary material figure 7.11), MR formed both diastereomers, however, with a considerable preference for the opposite shisool diastereomer, named diastereomer 2, than OYE1 (retention time 7.2 min).

By examining the designed Loop A variants in the bienzymatic cascade reaction one could notice that the reduction of the length in that loop region did not result in a considerably improved biocatalyst. However, the loop extension variant demonstrated a medium activity ranging from 36 % to 45 % conversion of the corresponding reduced alcohol (table 3.10). Therefore, it is reasonable to conclude that the length of the loop plays a role in the allyl alcohol reduction ability of the ene reductases. This fact is supported by the conversion rates of the loop grafting variant Loop A_OYE1, which next to the different amino acid composition also possesses the same increased loop length like Loop A_Long. Loop A_OYE1 demonstrated almost complete conversion of perillyl alcohol to the diastereomer 1 of shisool as well as moderate conversion of cinnamyl alcohol. Only in the reduction of geraniol a low conversion rate was observed. The altered amino acid composition of Loop A_OYE1 compared to Loop A_Long resulted in an increase of the reduction activity of cinnamyl alcohol and perillyl alcohol. Thus, it is reasonable to conclude that next to the loop length also the amino acid composition of the Loop A region is influencing the bienzymatic allyl alcohol reduction. This conclusion can again be supported by the second loop grafting variant Loop A_MR. The product formation of this variant in the geraniol reduction was similar to that of wild type MR. However, Loop_A MR demonstrated a complete allyl alcohol reduction for cinnamyl alcohol to 3-phenyl-1-propanol, as well as for perillyl alcohol to shisool. In contrast to wild type

MR, which is favoring the formation of diastereomer 2, the grafting variant Loop A_MR produced almost pure diastereomer 1. Therefore, the Loop A region did not possess an influence on the stereomeric outcome of the reaction.

Table 3.10: Conversion of allyl alcohols in a bienzymatic cascade reaction combining an ene reductase and an alcohol dehydrogenase. The final products are highlighted in bold and the intermediates in italic.

Substrate		Product formation in [%]								
		Cinnamyl alcohol			Geraniol			Perillyl alcohol		
		<i>Hydrocinnamaldehyde</i>	3-Phenyl-1-propanol	<i>Cinnamaldehyde</i>	<i>Citronellal</i>	Citronellol	<i>Geranial</i>	<i>cis/trans Shisool</i>	<i>Perillylaldehyde</i>	<i>cis/trans Shisool</i>
Wild type enzymes	NCR	2.2±0.1	4.3±0.4	0.2±0.1	4.2±0.4	3.3±0.6	-	1.3±0.04	0.3±0.1	0.1±0.1
	OYE1	0.3±0.1	97.5±0.6	0.1±0.01	-	99.1±0.1	0.2±0.1	95.4±0.1	4.1±0.1	0.5±0.1
	MR	2.3±0.1	85.9±0.4	8.4±1.1	-	25.9±1.0	15.8±2.7	12.0±5.3	6.7±6.7	71.4±2.4
Loop A variants	Loop A Short	0.6±0.1	13.6±2.9	0.9±0.5	1.9±0.5	5.1±1.7	-	5.7±3.7	0.3±0.1	0.3±0.1
	Loop A Long	0.1±0.1	36.6±4.9	1.3±0.3	0.1±0.1	43.1±9.1	7.9±2.5	45.3±0.4	19.1±4.6	1.4±0.1
	Loop A OYE1	0.5±0.1	57.6±8.5	0.70±0.1	0.2±0.1	10.3±1.0	8.2±0.4	94.2±0.2	0.9±0.1	3.4±0.1
	Loop A MR	-	97.6±0.7	1.0±0.2	-	20.9±4.1	10.06±1.1	93.4±0.4	2.2±0.3	3.37±0.7
Loop B variants	Loop B Short	0.2±0.1	98.3±1.2	0.1±0.1	6.3±1.8	90.2±0.7	0.7±0.2	95.4±0.1	1.1±0.4	3.4±0.5
	Loop B OYE1	-	94.4±1.1	0.1±0.1	-	42.0±4.8	3.4±0.2	82.9±1.4	3.2±0.4	3.8±0.7
	Loop B MR	0.9±0.5	30.7±1.3	31.3±0.8	-	45.4±1.2	21.0±10.4	6.2±1.0	5.9±0.9	84.2±0.9

Reactions were performed in triplicates in a final volume of 1 ml 50 mM Tris-HCl pH 7.5 and run at 30°C with 180 rpm for 24 h.

If one now looks at the bienzymatic reduction ability of the designed Loop B variants, it is observable that they demonstrated in parts a considerably altered behavior. All three designed Loop B variants were able to perform the bienzymatic cascade reaction for the allyl alcohol reduction. The loop length reduction variant Loop B_Short possessed, in contrast to Loop A_Short, a large impact on the allyl alcohol reduction. For all three substrates tested, Loop B_Short exhibited a formation of the desired reduction product of over 90 %, representing therefore the best biocatalyst for the allyl alcohol reduction. Loop B_Short, as well as the loop grafting variant Loop B_OYE1, formed the diastereomer 1 in the perillyl alcohol reaction. Loop B_OYE1 exhibited clearly increased conversion rates for the reduction of cinnamyl alcohol and geraniol, while the product formation in the perillyl alcohol reduction was slightly decreased compared to the Loop A_OYE1 grafting variant. However, the most significant influence on the reduction behavior within the Loop B variants was demonstrated by the Loop B_MR variant. Next to its moderate conversion rates in the cinnamyl alcohol and geraniol reduction, it exhibited a high conversion of perillyl alcohol, however to the opposite diastereomer 2. Based on these results one can conclude that the grafting of Loop B of MR in the NCR scaffold transferred the ability to produce the opposite diastereomer in the perillyl alcohol reduction. Loop B, in contrast to Loop A, plays therefore an important role in the control of the stereomeric outcome of the allyl alcohol reduction. This result goes in hand with the previous reported finding that Loop B also can influence the enantioselectivity of the enzyme.

To sum up this chapter, it could be successfully shown that variations, both in loop length and amino acid composition, within β/α loop regions of ene reductases resulted in enzyme variants possessing considerable increased conversion rates in the bienzymatic cascade reduction of allyl alcohols. Furthermore, it was possible to demonstrate that the Loop B region, which is located near the prosthetic FMN, plays a role in the stereomeric outcome of the reduction reaction. This raises the bigger question of the origin of such differences in the allyl alcohol reduction. The alteration in-between the different wild type enzymes, as well as the loop variants, can just partly be explained by the different activities of the ene reductases towards the formed activated reaction intermediates like cinnamaldehyde, geraniol and perillylaldehyde. Next to the reduction activity there has to be an additional distinction between the tested enzymes, most probable in the handling of the cofactor NADH, which is next to the used enzymes and substrates the only other possibility.

3.2.4.5 Cofactor handling of the designed loop variants

During the standard reduction, as well as the bienzymatic cascade reaction, the required nicotinamide cofactor in the reduced form (NADH) was consumed and in parallel the oxidized form (NAD⁺) was accumulated. Now the question should be addressed, whether the accumulation of the oxidized cofactor had an influence in the reduction activity of the tested enzymes. As model substrate the intermediate of the bienzymatic cinnamyl alcohol reduction, cinnamaldehyde, was chosen resulting in the reduced product hydrocinnamaldehyde. Additionally, it was noticed that the presence of the reduced cofactor NADH alone, without an alcohol dehydrogenase, can further reduce the obtained hydrocinnamaldehyde to the corresponding alcohol 3-phenyl-1-propanol (unpublished data). Therefore, the reduction of cinnamaldehyde resulted in two reaction products, which were combined in the term reduced products (figure 3.20).

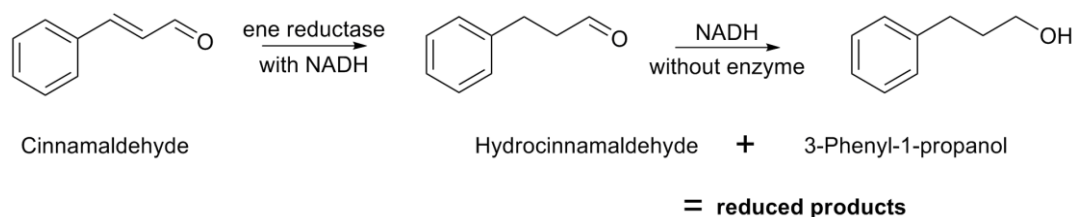


Figure 3.20: Reaction mechanism of the reduction of cinnamaldehyde with an ene reductase and the subsequent over reduction of the aldehyde group using NADH.

For the determination of the influence of the oxidized cofactor NAD⁺ on the activity of the enzyme, biotransformation reactions with different cofactor ratios at the reaction beginning were performed. Therefore, reactions with the following three NADH/NAD⁺ cofactor ratios were set up: 1/0, 4/1 and 1/1. In all biotransformation reactions sufficient NADH cofactor was present to fully reduce the used amount of substrate (table 3.11).

Table 3.11: Product yields in the conversion of cinnamaldehyde at different cofactor ratios. For all performed reactions there was an excess in the concentration of NADH over the cinnamaldehyde concentration to ensure a complete conversion of the substrate. The remained substrate is highlighted in italic, the obtained products conversion yields in bold. For detailed information and deviations see supplementary material table 7.14.

Enzyme	Product distribution	Product yield in [%]			
		Cofactor ratio NADH / NAD ⁺			
		1 / 0	4 / 1	1 / 1	
Wild type enzymes	<i>NCR</i>	<i>Cinnamaldehyd</i>	1.4	61.9	81.0
		reduced products	98.4	37.9	19.0
	<i>OYE1</i>	<i>Cinnamaldehyd</i>	31.2	57.4	74.1
		reduced products	59.4	37.9	22.0
	<i>MR</i>	<i>Cinnamaldehyd</i>	65.1	78.4	81.6
		reduced products	34.9	21.6	18.4
Loop A variants	<i>Loop A_Short</i>	<i>Cinnamaldehyd</i>	0.6	0.6	67.5
		reduced products	95.2	99.3	32.4
	<i>Loop A_Long</i>	<i>Cinnamaldehyd</i>	0.2	4.3	58.1
		reduced products	97.2	95.3	41.8
<i>Loop A_OYE1</i>	<i>Cinnamaldehyd</i>	0.6	35.1	55.5	
	reduced products	92.6	59.8	41.5	
<i>Loop A_MR</i>	<i>Cinnamaldehyd</i>	6.4	8.9	60.0	
	reduced products	89.0	89.2	38.8	
Loop B variants	<i>Loop B_Short</i>	<i>Cinnamaldehyd</i>	1.8	21.8	72.6
		reduced products	98.2	77.4	27.2
	<i>Loop B_OYE1</i>	<i>Cinnamaldehyd</i>	43.4	56.3	67.7
reduced products		56.6	43.7	32.3	
<i>Loop B_MR</i>	<i>Cinnamaldehyd</i>	60.9	61.7	87.1	
	reduced products	39.1	38.3	12.9	

Reactions were performed in triplicates in a final volume of 1 ml 50 mM Tris-HCl pH 7.5 and run at 30°C with 180 rpm for 24 h.

If one now firstly looks at the three wild type enzymes NCR, OYE1 and MR it was conspicuous that in the case of NCR the reduction activity was drastically reduced when the oxidized cofactor NAD⁺ was present in the reaction. The yield of the reduced products dropped about 80 % comparing the NADH/NAD⁺ cofactor ratio of 1/0 with 1/1. In view of this result, it is explained why NCR is not able to perform the bienzymatic cascade allyl alcohol reduction in high yields despite its good conversion

rates for the formed intermediate aldehydes cinnamaldehyde and geranial. The accumulation of the oxidized cofactor, which was produced by the ene reductase as well as the applied alcohol dehydrogenase resulted in the low conversion yields. The other two wild type enzymes, OYE1 and MR, however, possessed a considerable lower influence of the NAD^+ cofactor. For them the decrease of product yield was also detectable, but less severely pronounced with a decline of 60 % (OYE1), respectively 48 % (MR), at the cofactor ratio 1/1 compared to no NAD^+ present.

If one subsequently looks at the cofactor interaction of the designed Loop A variants in the reduction of cinnamaldehyde, it is apparent that all of them possessed an increased NAD^+ tolerance compared to NCR wild type. In particular, a lower concentration of NAD^+ in the reaction mixture (ratio NADH/NAD^+ 4/1) did not show any influence on the activity at all, except variant Loop A_OYE1 possessing a decrease in the product yield at the ratio 4/1, similar to the one from wild type OYE1. By enhancing the NAD^+ concentration further the product yield decreased for the Loop A variants, however, all of them remained more active than wild type NCR. This behavior led to the suggestion that alterations within the Loop A region, whether the loop length or the amino acid composition is addressed, resulted in a different handling of the cofactor.

Next to the Loop A region, also variations in the Loop B lead to an influence in the cofactor interaction. The loop length reduction variant Loop B_Short, as well as the loop grafting variant Loop B_OYE1, exhibited already a decline of product yield around 20 % at low concentrations of NAD^+ in the reaction mixture. Nevertheless, they still possessed an enhanced product yield compared to NCR wild type. The second loop grafting variant Loop B_MR demonstrated an activity which remained constant even at low amount of NAD^+ in the reaction.

In summary, it seems reasonable to conclude that both β/α surface loop regions are involved in the interaction with the reduced, as well as the oxidized form of the cofactor nicotinic adenine dinucleotide. Alterations within these loops, both the loop length and the amino acid composition, resulted in different product yields in the reduction of activated substrates. However, the exact type of interaction between the β/α loop regions and the NAD cofactor remains unclear.

4 Discussion

4.1 Golden Gate Shuffling-valuable tool for semi-rational loop exchange

The Golden Gate Shuffling is a robust, simple and efficient DNA shuffling technique depending on restriction enzymes type IIS which was developed in 2009 in order to enhance the trypsin activity in plants. Therefore, three highly related trypsinogen genes, with amino acid sequence identities ranging from 74-78 % were selected and based on a sequence alignment each divided into nine fragments. One round of Golden Gate Shuffling resulted in a fourfold increased trypsin activity compared to the wild type.¹⁰³ Therefore, the Golden Gate Shuffling represents a suitable method for the improvement of enzymatic activity by shuffling of defined fragments. Now the question arises whether it is possible to influence other enzyme properties such as selectivity and stability using this technique. As an object of study, TIM barrel enzymes are particularly well suited due to their structural division in a catalytic interface, built by variable β/α surface loops, and a stability interface made up of the TIM barrel itself as well as the connecting α/β loops. In particular, the catalytic interface forming β/α surface loops represent a great target for the application of the Golden Gate Shuffling in order to elucidate their influence on enzyme properties.

The Golden Gate Shuffling method offers several advantages, which are beneficial for an application in the targeted semi-rational exchange of loop regions. Firstly, the exact definition of fragments concerning start and end point and, additionally, the guarantee of the same fragment sequence in all obtained variants. These two features ensure the shuffling of entire loop regions in the proper order. However, there are also some method adjustments necessary in order to apply the Golden Gate Shuffling for the targeted shuffling of loop regions between selected members of the Old Yellow Enzyme family: (i) To target specifically the loop regions within a protein for the shuffling, it is indispensable to possess next to sequence data, also structural information about the selected target proteins. Therefore, the selection of the shuffling partners is restricted to enzymes with available structural information, namely solved crystal structures. (ii) For the comparative investigation of the influence of loop regions on enzyme activity, it is necessary to maintain in all created variants the same solid structural backbone, as well as a common active site. Based

on comparison with the corresponding wild-type enzyme, these two conditions will allow a statement about the influence of the loop regions on activity. Therefore, that part of the defined fragments ensuring the TIM barrel structure, as well as containing the catalytically important amino acids needs to be derived from the same parental protein. As wild type comparing partner the 2-cyclohexen-1-one reductase NCR from *Zymomonas mobilis* was chosen due to the fact that it possesses a monomeric structure avoiding the occurrence of oligomerization problems within the shuffling. In total four shuffling partners for NCR were selected based on literature and structural data covering both subfamilies of the Old Yellow Enzyme family, as well as a broad range of substrates, reactions and stabilities: OYE1 from *Saccharomyces pastorianus* and PETNR from *Enterobacter cloacae*, representing the *classical* subfamily, as well as YqjM from *Bacillus subtilis* and GkOYE from *Geobacillus kaustophilus*, representing the *thermophilic-like* subfamily. However, it should be mentioned that the amino acid sequence identities of the selected shuffling partners, ranging from 25 to 40 %, are considerably lower than the ones in the original study from Engler *et al.* (2009) ranging from 74-78 %.¹⁰³ These lower sequence identities cause two main problems in the fragment definition:

- (1) The β/α surface loops of the five selected ene reductases are characterized by a large diversity within the small loop regions including differences in the amino acid chain length, as well as differences in the amino acid composition in each shuffling partner. Therefore, it is inevitable to include in some cases the previous β -sheet, respectively subsequent α -helical secondary structure element of a loop region in the fragment definition for the Golden Gate Shuffling in order to find a conserved amino acid position for the four nucleotide recombination site, as well as provide at least a fragment size of approximately 100 bp.
- (2) The presence of conspicuous sequential amino acid differences in the C-terminal part of the five selected ene reductases result in the insertion of altogether three point mutations in three different fragment transitions in order to provide the four nucleotide long recombination site at the beginning and end of each fragment set.

On the basis of structural and sequential alignments the five selected ene reductases were divided into a total of seven fragments. For providing the same solid structural backbone and a common active site in all created variants, three of the seven

fragments remain from NCR, namely the N- (F1) and C-terminus (F7) of the enzyme, as well as the middle fragment (F4) containing the catalytic important tyrosine. The remaining four fragments contain each one β/α loop region which will be shuffled during the Golden Gate Shuffling. Due to the small size as well as the low sequence identity of the four β/α loop regions it is necessary to include always at least the subsequent α -helix in the fragment. In addition, for fragment 2 it is inevitable to involve next to the subsequent α -helix also the following TIM barrel forming β -sheet in the fragment definition in order to provide a conserved recombination site. Most of the additional secondary structure elements are, however, necessary for the definition of fragment 6, since this part of the shuffling partners exhibits the greatest diversity in terms of amino acid composition. This diversity is also demonstrated by the fact that all three necessary point mutations in the recombination sites are located directly in front of or behind of fragment 6. So in total, three secondary structure elements are present in fragment 6, the previous β -sheet of the loop region, as well as the subsequent α -helix and the following β -sheet. Figure 4.1 illustrates the fragment definition at NCR, with the three constant NCR fragments colored in black possessing five of the eight TIM barrel forming β -sheets and the four shuffling fragments highlighted in several colors containing three TIM barrel forming β -sheets.

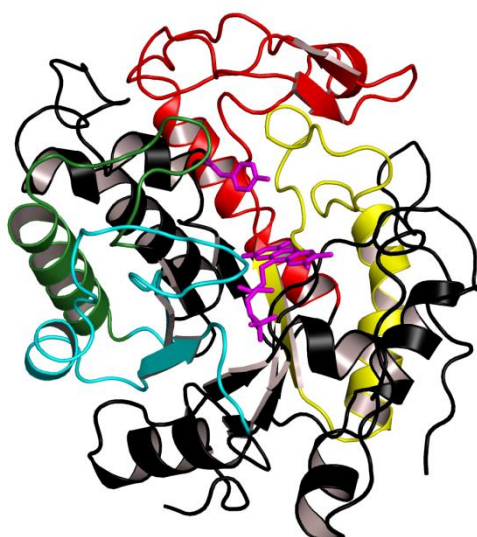


Figure 4.1: Cartoon representation of NCR with the seven defined fragments for the Golden Gate Shuffling. The fragments representing the four shuffled loop regions are highlighted in yellow (F2), red (F3), green (F5) and cyan (F6). The three fragments from the NCR backbone (F1, F4 and F7) are colored in black. The catalytic active tyrosine, located in F4, as well as the prosthetic FMN is shown in stick representation and colored in magenta.

The set up of three constant and four shuffling fragments resulted in a total of 1024 possible enzyme variants which needed to be screened with an appropriate assay for the desired property. However, during the development of an appropriate screening assay, one should always have in mind the first law of all semi-rational and directed evolution methods which is “you get what you want to screen for”.⁸³ In the Old Yellow Enzyme Golden Gate Shuffling approach the main interest is dealing with the question whether the obtained variants are still reduction active enzymes or just unsolvable protein waste due to the large alterations in amino acid composition and structure. In principle a solid protein expression level, as well as a proper structural folding are indispensable for the activity of an enzyme and all other related properties. Therefore, in order to investigate the reduction activity of Golden Gate Shuffling variants a fast photometric screening assay was developed based on the consumption of NAD(P)H during the reaction measured by the associate change in absorbance at 340 nm. A lysate-based NAD(P)H consumption photometric activity assay always has some restrictions in application:

- (1) The investigated reaction has to be always fast, due to the fact that NAD(P)H is not permanently stable over a longer period resulting in absorption changes. Therefore, the detection of enzyme variants which are active, however, with a rather slow reduction activity is not possible with such an assay system.
- (2) In addition to the GGS reductase variants many other enzymes within the applied cell lysate are NAD(P)H-dependent and thus, promote a false positive decrease in absorption, even without an existing reductase activity. As a consequence it is always necessary to perform both, positive as well as negative control reactions.
- (3) The selected substrate must not be reduced by NAD(P)H alone nor representing a suitable substrate for any other enzyme present in the cell lysate.

Haven taken these three limitations into account, a total of 1100 GGS variants (by 1024 theoretically possible different variants due to the fragment setup) were screened with the developed NAD(P)H assay for the ability to reduce the standard substrate ketoisophorone resulting in 18 variants demonstrating an at least slightly increased activity compared to lysate background. This small number of active enzymes elucidate, however, that most of the obtained GGS variants exhibit no measurable activity. Thus, the control of the expression level of non-active enzyme

variants showed a clearly detectable protein expression. Therefore, there are two reasons/explanations possible for the negative result of the NAD(P)H assay: (i) firstly, the obtained GGS variants were too slow in the reduction of ketoisophorone to be detected in the screened reaction period of 1500 sec, or (ii) secondly, the variants possessed no proper structural folding due to the large inserted alterations. Biotransformation reactions performed with several of the non-active variants of the NAD(P)H assay over a reaction time of 24 h suggest that rather the first possible explanation of a considerably slower reaction rate is the case. Therefore, it would be probably better to develop an alternative screening system which directly measures the product formation or design a viability assay in which the cells just are able to grow when the desired reduction activity is available.

Looking now at the substrate spectrum of the five selected GGS variants, consisting of the four in the NAD(P)H assay most active ones, plus one which is located in the range of the negative lysate control reaction, it is noticeable that all variants tested are active in the reduction of α/β -unsaturated substrates, however, neither can reduce sterically difficult nor less activated substrates. Regarding now the fragment composition of the two most active GGS variants (GGS 225 and 229), which possess almost always higher substrate conversions than wild type NCR it is apparent that both are just built by fragments from NCR, YqjM and PETNR. While the variant GGS 229 possesses twice two consecutive fragments, firstly from PETNR (F2 and F3) and secondly from YqjM (F5 and F6), variant GGS 225 has just one fragment from PETNR (F2) and one from YqjM (F6). The other two fragments (F3 and F5) are derived from NCR. Though, the presence of the two additional fragments from the scaffold forming NCR wild type in the variant GGS 225 led to a decrease in reductase activity towards the two aliphatic substrates 2,4-heptandienal and geranial compared to variant GGS 229 resulting in a product yield of GGS 225 similar to wild type NCR. This result leads to the assumption of the existence of beneficial cooperative effects between two neighboring fragments. If the adjacent fragments F2 and F3 originate from the same enzyme, as well as F5 and F6, the reduction capability of the obtained variant increased. Due to the structural alignment of the five parental shuffling ene reductases (figure 3.2), in which the also in the shuffling included secondary structure elements hardly demonstrate any considerable differences, it seems likely that the positive cooperative effects in the GGS variants arise from the shuffled β/α loop regions. Furthermore, it would now be interesting to elucidate which of the two

possible cooperative effects (F2/F3 or F5/F6) is enhancing the reduction activity of variant GGS 229 or whether both are important. Therefore, the creation of two additional variants would be necessary, one with fragment F2 and F3 from PETNR and one with fragment F5 and F6 from YqjM. All the other fragments should be from NCR.

Interestingly, when all four shuffling fragments originate from one parental enzyme, like in variant GGS 174 from PETNR, or just one for the four fragments is shuffled, like in variant GGS 10 fragment F3 from OYE1, the reduction activity is diminished compared to variant GGS 229. This may suggest that the common occurrence of fragment F2 and F3, or F5 and F6, is important for a beneficial catalytic activity. This statement could be tested in a second Golden Gate Shuffling approach by only permitting the shuffling of fragments F2 and F3, respectively F5 and F6. A further advantage of the Golden Gate Shuffling method is that once all defined fragments are available in a cloning vector the rational assembling of these fragments is possible based on obtained information. Furthermore, it is now in progress to investigate the created GGS variants towards other novel enzyme properties like stability or enantioselectivity with an appropriate screening assay. Going even further, one could additionally test the obtained GGS variants towards new catalytic reactions or substrate promiscuity, like disproportion or isomerization reactions.

Finally it can be concluded that the adapted Golden Gate Shuffling represents a valuable method for the investigation of loop exchanges and cooperative loop effects within the family of the Old Yellow Enzymes.

4.2 Surface loops as small elements in the evolution of enzymes

Concerning the role and distribution of β/α surface loop regions within the evolution of the Old Yellow Enzyme family, it is firstly advisable to examine the evolutionary connection, as well as the relationship of all family forming enzymes among each other. For that purpose the phylogenetic tree analysis represents one of the best sequenced based techniques. For a small subpart of the Old Yellow Enzyme family, namely for 74 12-oxophytodienoate reductases (OPR) originating from plants, already in 2009 a phylogenetic analysis was performed in order to investigate their overall phylogenetic relationship, structural evolution, as well as function divergence

within the plantae kingdom. The study revealed that the plant originating OPRs can be subdivided into seven well conserved subfamilies based on overall sequence identity which are all originating from one ancestor enzyme.¹⁴⁰

By creating an un-rooted maximum likelihood phylogenetic tree of all so far in databases stored ene reductases, namely 4483 enzymes, it strikes out that the entire family of the Old Yellow Enzymes can be divided into five homologous subfamilies of different size (figure 3.7). In order to examine now in detail the composition, as well as the distribution of individual β/α surface loops within the Old Yellow Enzyme family, a total of three reference enzyme (NCR, OYE1 and MR), all pertaining to the homologous family HF1, were selected and in each two loop regions (Loop A and B) have been defined based on secondary structure prediction. Using the within this work developed strategy for the creation of Hidden Markov loop profiles based on the before defined loop regions, six loop profiles were created demonstrating that the β/α surface loops are consisting of a sequence of conserved and variable amino acids. Furthermore, it was shown that two types of loop profiles exist, namely loops demonstrating a generalist motif, which could be found in many sequences, like Loop A, and loops possessing a specialist motif, which is restricted to a small amount of closely related family members, like Loop B. The further discussion of the role of β/α surface loops within the Old Yellow Enzyme family will be focused on the generalist type loop, Loop A.

Looking at the already created three Loop A profiles of the HF1 subfamily members NCR, OYE1 and MR it attracts attention that all of the in total 865 profile forming sequences possess a highly conserved N-terminal hinge region consisting of three amino acids, L-S/A-P, as well as a conserved negatively charged amino acid at the C-terminal end D/E (figure 4.2b). In order to prove the finding of these two conserved elements within the Loop A regions of ene reductases belonging to the HF1 subfamily an additional loop profile was generated. Therefore, the plant reductase OPR1 from *Arabidopsis thaliana* was selected as reference sequence, due to the fact that all plant reductases also belong to the HF1 and are not present in the three already created profiles. Figure 4.2a illustrates the obtained OPR1 Loop A profile colored after the chemistry of the included amino acids, consisting of in total 162 sequences and also exhibiting the N-terminal conserved hinge region, as well as the C-terminal acidic amino acid.

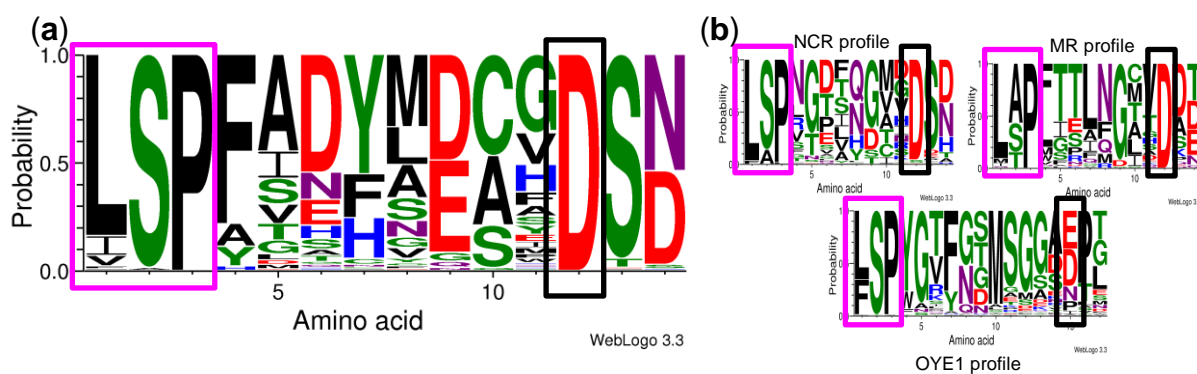


Figure 4.2: Graphical representation of in total four Loop A Hidden Markov profiles from the homologous subfamily HF1. The size of the amino acid letter corresponds with its frequency at the relevant position within the loop region. The Hidden Markov profiles are visualized with WebLogo 3.3 and the amino acids are colored on basis of their chemistry. (a) Hidden Markov profile of Loop A from OPR1 with a loop length of 14 aa made up by 162 sequences. (b) Arrangement of the three already described Loop A Hidden Markov profiles of NCR, MR and OYE1. Highlighted in pink is the N-terminal conserved hinge region and in black the C-terminal conserved acidic amino acid.

Therefore, one can proceed with the assumption that the N-terminal L-S/A-P hinge region as well as the acidic D/E at the C-terminal end are representing conserved features of the Loop A region. The question now arises is, whether these two features are representing a common characteristic of all Loop A regions within all five homologous subfamilies, or if they are unique for sequences belonging to the HF1 subfamily. To answer this question, also Hidden Markov loop profiles for the other homologous subfamilies were created with, on the basis of literature data, selected reference sequences. However, all so far described ene reductase family members are spread over only three of the five subfamilies, HF1, HF4 and HF5. The two homologous subfamilies HF2 and HF3 are formed solely of ene reductase sequences obtained by genome sequencing projects. Therefore, only additional reference sequences for the Hidden Markov profile generation of the HF4 and HF5 subfamily were chosen. While HF5 is a quite small subfamily built by just 279 closely related sequences, all belonging to the *thermophilic-like* subclass of Old Yellow Enzymes and therefore could be represented by one reference sequence (YqjM from *Bacillus subtilis*), HF4 is a rather large subfamily with 1246 members. If you have a deeper look in the sequences building the HF4 subfamily, one can notice that it also includes several *thermophilic-like* subclass members like CrS from *Thermus scodoductus* SA-01, which is selected as reference sequence to represent that part of the subfamily. The second part of the HF4 subfamily is represented by the quite well described xenobiotic reductase XenA from *Pseudomonas putida* 86 as reference

sequence. These both HF4 members possess a sequence similarity of 44.5 %, however demonstrate considerable differences in the β/α surface loop regions.

Looking now at the three newly created loop profiles (figure 4.3) it is striking that all three of them also possess a conserved N-terminal hinge region, which is however consisting of only two amino acids. For the two profiles representing the *thermophilic-like* subclass of ene reductases, the Crs and YqjM profile, these two N-terminal conserved hinge amino acids are a serine followed by an alanine. Within the 125 sequences forming the XenA motif all possess as glycine followed by a valine as N-terminal hinge. One can therefore conclude that the presence of the N-terminal hinge region is a general characteristic of this specific β/α surface loop within the family of the Old Yellow Enzymes, which, however, differs in length, as well as amino acid composition among the different homologous subfamilies.

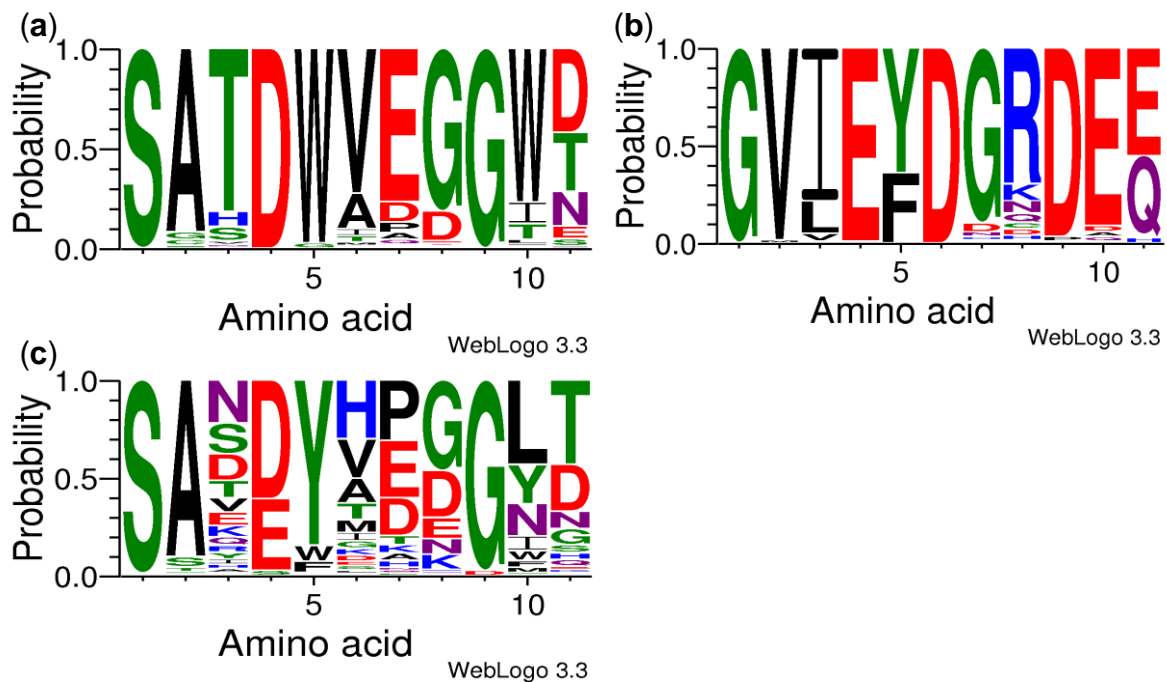


Figure 4.3: Graphical representation of the three additionally created Loop A Hidden Markov profiles from the homologous subfamilies HF4 and HF5. The size of the amino acid letter corresponds with its frequency at the relevant position within the loop region. The Hidden Markov profiles are visualized with WebLogo 3.3 and the amino acids are colored after their chemistry. (a) Hidden Markov profile of Loop A from CrS of the homologue family HF4 with a loop length of 11 aa made up by 511 sequences. (b) Hidden Markov profile of Loop A from XenA of the homologous family HF4 with a loop length of 11 aa made up by 125 sequences. (c) Hidden Markov profile of Loop A from YqjM of the homologous family HF5 with a loop length of 11 aa made up by 259 sequences.

Considering now the presence of the second feature within the three new loop profiles, the presence of a conserved acidic amino acid in the C-terminal part of the loop, one can see that all of them have also a conserved amino acid at the same position, however with a different character. Again, the two profiles representing the *thermophilic-like* subclass possess the same amino acid, in particular a glycine. The XenA profile contains not one, but two conserved acidic amino acids at the C-terminal loop end. Remarkably, the entire XenA profile exhibits a rather acidic character with a total of five acidic amino acids by a loop length of just eleven residues (figure 4.3b). Therefore, one can say that in addition to the N-terminal hinge region also the presence of a conserved amino acid at the C-terminal end is a common feature of the entire Loop region A within the family of ene reductases. However the character of the amino acid at the conserved positions is varying between the different subfamilies.

Now the question arises how well the created loop profiles represent the respective homologous subfamilies, since not all sequences of a subfamily are included in the created loop profiles. Considering the three investigated homologous subfamilies, in HF1 only 41 % of the 2558 subfamily forming sequences are included in the four created loop profiles NCR, MR, OYE1 and OPR1, namely in total 1027. In HF4 a total of 636 are represented by the two Loop A profiles CrS and XenA representing 51 % of the 1246 subfamily forming sequences. The YqjM Hidden Markov profile representing the small subfamily HF5 is formed of 259 sequences representing 96 % of the 271 subfamily forming sequences. Since the four Loop A profiles comprising of NCR, MR, OYE1 and OPR1 include the lowest amount of subfamily forming sequences, the HF1 subfamily was chosen for further studies. On the basis of the subfamily HF1 it is investigated whether one can make general statements about the Loop A region of the entire subfamily by using the created Hidden Markov loop profiles or not. Therefore, figure 4.4 illustrates the labeling of all Loop A profile forming sequences in the phylogenetic tree of the HF1 subfamily. One can clearly see that the four reference sequences are located at four different branches within the tree representing already a broad sequence space. If one now looks at the distribution of the profile forming sequences it is quite apparent that the 1027 sequences are spread over the whole phylogenetic tree. There is just one branch present within the phylogenetic tree of which no sequences are included in the four Loop A profiles (figure 4.4b highlighted in yellow). But this does not

automatically imply that these sequences do not also exhibit a related loop profile. The sequences located at this branch could maybe possess a considerably different loop length than NCR, MR or OYE1, which would lead to elimination in the Hidden Markov profile generation due to the exclusion criterion of less than 30 % gaps for a profile forming sequence. One can therefore assume that the four created loop profiles represent well the entirety of the sequences forming the HF1 subfamily. Hence, it is therefore possible to obtain general valid conclusions for the entire Old Yellow Enzyme subfamily based on these loop profiles.

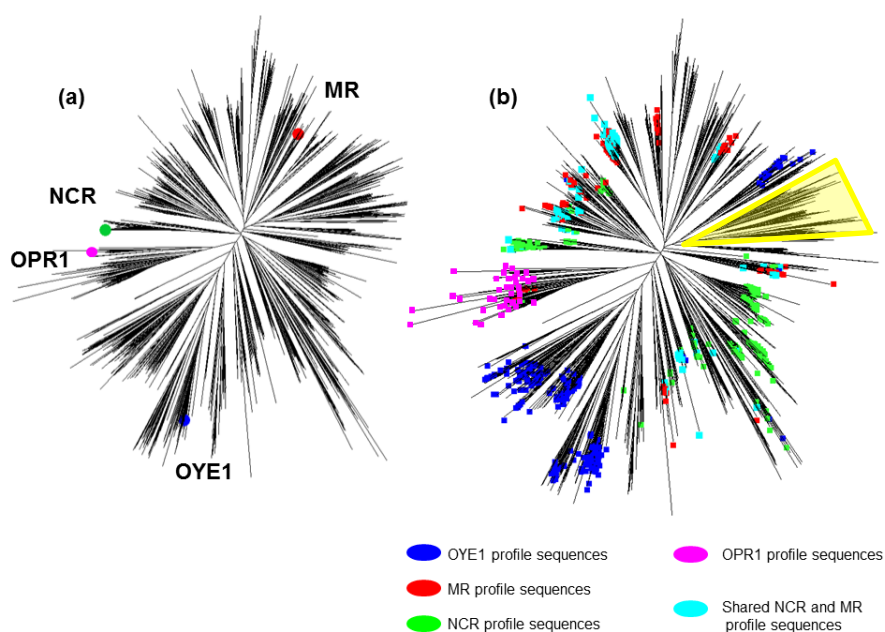


Figure 4.4: Phylogenetic tree of the HF1 subfamily containing 2558 Old Yellow Enzymes. Each line represents one sequence. (a) Localization of the four reference sequences NCR (green), MR (red), OYE1 (blue) and OPR1 (pink). (b) Distribution of all 1027 Loop A profile forming sequences in the corresponding color of their reference sequence. Additionally, sequences being present in the NCR as well as MR profile are colored in cyan. Highlighted in yellow is the part of the phylogenetic tree, which is not included in the four Loop A profiles.

Due to the fact that the three Loop A profiles representing the homologous subfamilies HF4 and HF5 even include percentage more subfamily forming sequences, one can assume that the findings for HF1 concerning the validity of loop-based conclusions can be taken over for them. Looking now at the seven created Loop A profiles, one can deduce that each of the on the overall sequence identity obtained homologous subfamilies possess their own unique composition of the two conserved Loop A features. That fact allows the classification of ene reductase sequences in homologous families only by looking at the amino acid composition of the conserved features of the Loop A region (figure 4.5). You can

therefore create a guideline for the classification of Old Yellow Enzymes in homologous subfamilies only based on the composition of their intrinsic Loop A region:

(1) Consideration of length and composition of the N-terminal hinge region and the amino acid character of the C-terminal conserved amino acid:

(a) L-S/A-P + D/E → subfamily HF1

(b) G-V + D-E → subfamily HF4 XenA like enzyme

(c) S-A + G → *thermophilic-like* subclass either HF5 or HF4

(2) For the exact assignment of the *thermophilic-like* subclass members in the two homologous subfamilies HF5 or HF4, there is an additional consideration necessary, namely the type of a highly conserved aromatic residue in the middle part of the loop.

(d) S-A + G + Y → subfamily HF5

(e) S-A + G + W → subfamily HF4

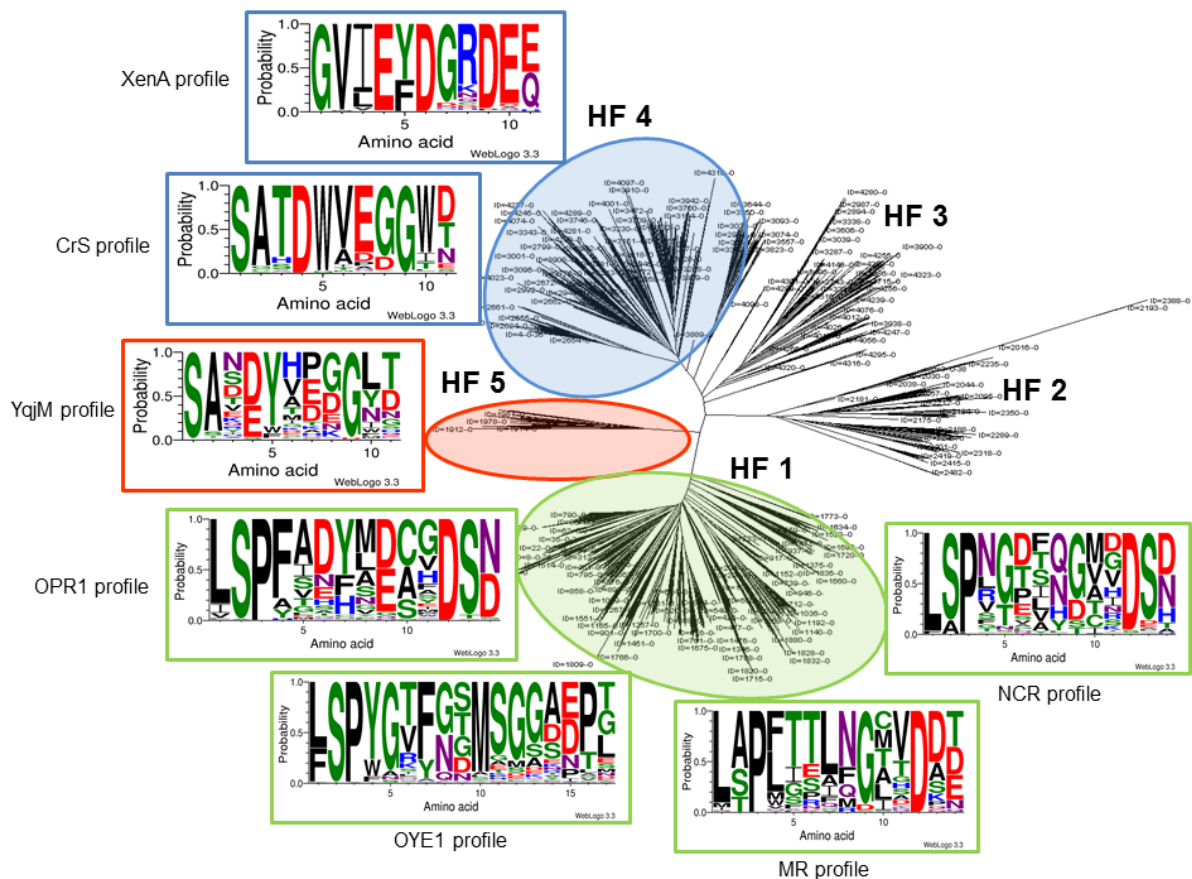


Figure 4.5: Graphical representation of the allocation of Hidden Markov loop profiles to the corresponding homologous subfamily. The amino acid composition of the Loop A region allows the classification of the entire sequence within the corresponding homologous subclass. The seven created loop profiles are highlighted in the color of the corresponding homologous subfamily.

In summary, one can conclude that on the basis of the amino acid composition of the β/α surface region Loop A it is possible to assign the Old Yellow Enzyme family members in the same homologous subfamilies which are obtained by the creation of a phylogenetic family tree based on the overall protein sequence identity. Therefore, Loop A represents a small element within a reductase that reflects the relationship of the entire enzyme within the family of the Old Yellow Enzymes.

4.3 Influence of surface loop regions on enzyme properties

Regarding the question of the influence of loop regions on the properties of an enzyme it strikes out, that so far not too much is reported about that purpose in literature. If at all, almost exclusively active site loops have been investigated towards their contribution on enzyme properties. For example, in 2006 Park *et al.* demonstrated by simultaneous substitution of several pre-designed active site forming loops that it was possible to introduce a new catalytic activity in an existing protein scaffold.¹⁴¹ Furthermore, in 2008 Boersma *et al.* showed that the enantioselectivity of lipase A from *Bacillus subtilis* could be altered by active site loop grafting.¹⁴² If one wants to consider now the influence of surface loop regions on enzyme properties, there are even fewer examples known. In 2010 Prokop *et al.* published the modulation of the enantioselectivity of a haloalkane dehalogenase through the presence or absence of a surface loop region located in-between the core of the α/β hydrolase and its cap domain.¹⁴³

Turning now to the Old Yellow Enzyme family, then so far almost nothing is known about the influence of the β/α surface loop regions on enzyme properties. However, recently Daugherty *et al.* succeeded in enhancing the reduction activity of the ene reductase OYE1 from *S. pastorianus* by the generation of new protein termini in the three, within this thesis defined β/α surface loop regions A,B and C of the catalytic interface by circular permutation.¹⁴⁴ In order to answer the question which β/α surface loop within an Old Yellow Enzyme possesses an influence on which specific enzyme property, imparted through which intrinsic loop feature (loop length or amino acid composition), a total amount of seven rational loop variants of the previous two defined Loop A and B were created. In total three loop length variants (two of Loop A, one of Loop B) were rationally designed based on structural and sequential

alignments in order to address the impact of the loop length on the enzyme properties such as activity, substrate specificity, enantioselectivity, stability, selectivity and cofactor interaction. Additionally, four loop grafting variants were created by grafting the Loop A, respectively Loop B, region between NCR and OYE1, as well as MR in order to determine whether it is possible to transfer enzyme properties from one enzyme to another by loop grafting. However, it should be mentioned that always only one loop region was altered allowing no cooperative effect examination.

Due to the fact that the activity, as well as all other properties of an enzyme are already determined by its spatial arrangement the crystal structure of one variant was solved in order to elucidate structural changes within the created variant compared to wild type NCR and therefore to be able to explain potential property alterations. One of the loop grafting variants, namely Loop A_OYE1, was chosen for crystallography due to the fact that it combines both rational design approaches by demonstrating a by three amino acids increased loop length, as well as an altered amino acid composition. The crystal structure of the variant Loop A_OYE1 was solved with a resolution of 1.8 Å demonstrating a proper TIM barrel folding, thus possessing considerable alterations in the structure and flexibility of all three defined surface loop regions. This result suggests that there are existing cooperative effects between the β/α loop regions of Old Yellow Enzymes, which was already be assumed by the results obtained in the Golden Gate Shuffling. The Golden Gate Shuffling results pointed out that cooperative effects probably occur between two adjacent loop regions, like Loop A and Loop B. This assumption is supported by the crystal structure of the variant Loop A_OYE1, where the exchange of Loop A results in a completely different structural orientation of the adjacent, intrinsic NCR Loop B region. Therefore, it would be interesting to create double variants possessing two grafted loop regions in order to investigate the cooperative effects between them. In addition to the different structural orientation of the Loop B region, the grafting of Loop A led to an increased flexibility of all β/α surface loops (except Loop B), which could be seen in their low detectable electron density. A maximum flexibility of the loop regions may, however, demonstrates a negative impact on the stability of the enzyme. And indeed, one can see an increased solvent, as well as thermo sensitivity of the loop grafting variant Loop A_OYE1 compared to wild type NCR in biotransformation reactions at different organic solvent concentrations and/or reaction temperatures. Thus, the higher flexibility of the surface loop regions induced by the

grafting of Loop A promoted to the expected decrease in the stability of the enzyme. Furthermore, it can be seen in the crystal structure of the variant Loop A_OYE1 that the grafting of Loop A had an influence on the architecture of the active site. Firstly the catalytic active tyrosine is located in a different orientation pointing away from the prosthetic FMN and also surrounded by several additional amino acids. Secondly the altered orientation of the Loop B region results in a wider active site with a FMN exposed to the solvent (figure 4.6). These two structural features should complicate the proper proton transfer from the catalytic tyrosine onto the C α of the substrate's carbon-carbon double bond resulting in a decreased reduction activity compared to wild type NCR. And indeed, by looking at the conversion rates for the reduction of the six selected activated α,β -unsaturated substrates it is obvious, that the activity of the loop grafting variant Loop A_OYE1 considerably decreases compared to wild type NCR.

One can therefore conclude that the grafting of a single loop region results in significant structural alterations which are directly linked to changes in enzyme properties like stability and activity.

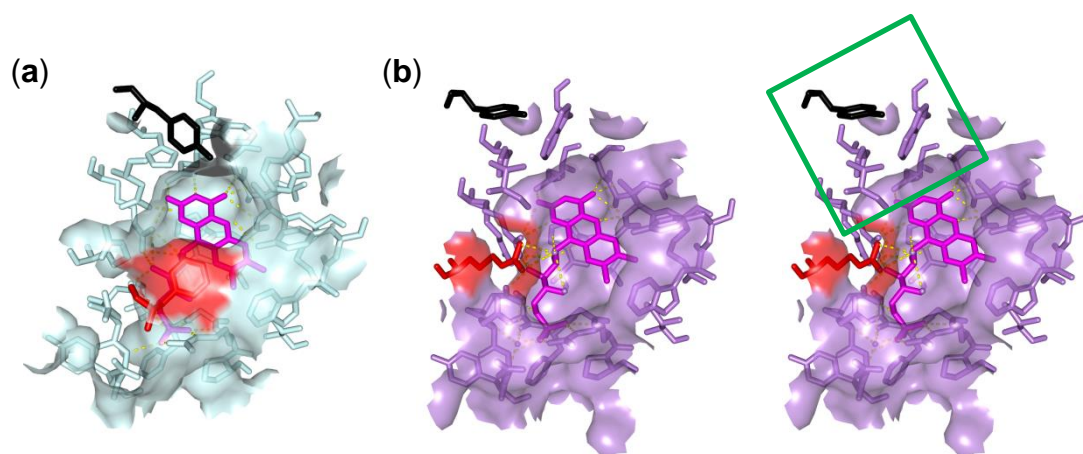


Figure 4.6: Graphical representation of the active site of (a) NCR wild type colored in pale cyan and (b) the loop grafting variant Loop A_OYE1 colored in light purple. Highlighted in red are the two amino acids of the Loop B region which are located closest to the prosthetic FMN, F269 in NCR closing the active site and R264 in Loop A_OYE1 opening the active site of the variant. The catalytic active tyrosine is colored in black. The different orientation of the catalytic tyrosine in the loop grafting variant is marked in green.

Thus, one can assume that in the other six rational designed loop variants also structural alterations are present which result in changes in the enzymes properties.

Now the question arises whether one can identify which enzyme property is affected by which loop region, and which loop alterations can result in an improved biocatalyst. Looking now first at the reductase activity, it is noticeable that all created Loop A variants either have an activity in the same range like wild type NCR or possess a decrease in the conversion rates of all tested substrates. This finding could be explained by the close proximity of the Loop A region to the catalytic active tyrosine. Alterations in the Loop A region of the enzyme led probably to an altered orientation of the catalytic active tyrosine and therefore, to a decline in activity like already shown for the variant Loop A_OYE1. However, the loop grafting variant Loop A_MR possesses an interesting feature; it transfers the substrate preference for the *cis*-isomer of citral, neral, from the wild type MR to the geranial preferring NCR scaffold leading to a variant with altered *cis/trans* substrate specificity. Therefore, one can conclude that the Loop A region has an influence on the substrate recognition and therefore the substrate specificity of the enzyme. If one now looks at the reduction activity of the designed Loop B variants it is interesting to observe that they are entirely more active in the reduction of standard substrates than the Loop A variants. For small substrates, like 2-methyl-2-pentenal, all of them are even more active than NCR wild type. Thus, for improving the reduction activity of the Old Yellow Enzyme the Loop B regions seems to be the more promising candidate with regard to the respective substrate.

A similar picture emerges when one looks at the influence of the two loop regions A and B on the enantioselectivity of the enzyme. Alterations in Loop B, both in chain length as well as amino acid composition, result in varied enantiomeric excesses of the formed products which stands in contrast to alterations in the Loop A region. Firstly, the shortening of the length of Loop B resulted in a deterioration of the enantioselectivity in the reduction of the small substrate 2-methyl-2-pentenal, which could most likely be explained with more space near the catalytically important elements in the active site, and therefore poor substrate localization. Looking now at the two grafting variants it can be seen that by the shuffling of the Loop B the enantiopreference of the parental enzyme, OYE1 respectively MR, was transferred into NCR for both substrates tested, 2-methyl-2-pentenal as well as ketoisophorone. Therefore, one can assume that for enantioselectivity the amino acid composition of

the Loop B region represents an important factor and that it is possible to transfer properties from one enzyme to another by the grafting of this loop.

Looking next at the influence of length and amino acid composition alterations in the two β/α surface loop regions A and B on the solvent, as well as thermal stability, it is striking that for both stabilities the same behavior could be observed. All variations in the Loop A resulted in a decreased stability. Therefore, when the stability is the desired enzymatic property addressed in enzyme engineering approaches of Old Yellow Enzymes, alterations in this loop area should be avoided. However, alterations in Loop B region, which possesses the four amino acids with the highest B-factors, standing for high flexibility within the entire NCR reductase, show a quite different behavior. The loop length reduction variant Loop B_Short, which is shortened by the four highly flexible amino acids, possesses a considerably increased solvent, as well as thermostability. This result can be explained by the fact that the deletion of the four most flexible amino acids, indicated by the high B-factor values, results in more rigidity and therefore a higher thermal and solvent stability without almost a considerable loss of reduction activity.

Finally, now considering the performance of the created β/α surface loop variants in the bienzymatic reduction of allyl alcohols it is apparent that both loop regions possess an influence on this property. While the NCR wild type is not at all capable of catalyzing this cascade reaction the other two wild type enzymes, MR and OYE1, are rather acceptable biocatalysts for it. By grafting the Loop A or the Loop B region from MR and/or OYE1 in NCR, it is possible to transfer the activity towards allyl alcohols in the NCR scaffold. In addition to the grafting variants also the loop length variants, Loop A_Long and Loop B_Short, possess a high activity in the allyl alcohol reduction towards all tested substrates. Therefore, one can conclude that the loop length, as well as the amino acid composition of both loop regions is important for the ability to reduce allyl alcohols. Asking now the question where the differences in allyl alcohol reduction between NCR and the other wild types enzymes, as well as the created variants are coming from, it could be shown that NCR possesses a rather sensitive behavior towards the accumulating oxidized cofactor NAD^+ . When NAD^+ is present in the reaction mixture the reduction activity of NCR wild type considerably decreased, while the activity of the other two wild types, as well as the created variants stay on a constant level. Perhaps the, during the reaction obtained, oxidized NAD^+ acts as a

kind of constitutive inhibitor within the reaction. One can therefore infer that the two β/α surface loop regions are interacting in some manner with the cofactor NAD(P)H.

By grafting the Loop B region of MR in the NCR scaffold a second enzyme property is transferred next to the ability to reduce allyl alcohol, namely an altered product stereoselectivity. While the wild type enzymes NCR and OYE1, as well as all created loop variants, except the grafting variant Loop B_MR, produce diastereomer 1 of shisool (retention time 7.0 min) in the reduction of perillyl alcohol, MR wild type and variant Loop B_MR form the other possible diastereomer (retention time 7.2 min). Therefore, in addition to the allyl alcohol reduction activity also the stereo preference of the product formation is transmitted with the shuffling of Loop B.

Finally, thus one can summarize that β/α surface loop regions of Old Yellow Enzymes possess an influence on several enzyme properties. Especially rational alterations within the specialist type Loop B region regarding loop length and amino acid composition can lead to improved properties, such as increased stability and enantioselectivity as well as altered product specificity. Furthermore, it is possible to transfer an enzyme property and even a new catalytic activity from one enzyme to another by the grafting of β/α surface loop regions of Old Yellow Enzymes.

5 Conclusion and Outlook

Within present thesis β/α surface loop regions of Old Yellow Enzymes were investigated with semi-rational, bioinformatic and rational design methods in order to understand their role and influence on properties within a single enzyme, as well as in the entire Old Yellow Enzyme family.

Throughout the semi-rational Golden Gate Shuffling four β/α surface loop regions of five different Old Yellow Enzyme family members were shuffled resulting in 18 variants demonstrating at least a slightly increased reduction activity in the developed NAD(P)H-based photometric screening assay compared to the corresponding wild type enzyme NCR. Based on the loop composition of the two best created GGS variants (GGS 229 and GGS 225), one can conclude that two adjacent loop regions possess beneficial cooperative effects which can result in an increased

activity. The Golden Gate Shuffling represents a valuable tool for the simultaneous and semi-rational exchange of β/α surface loop regions between different Old Yellow Enzyme family members in order to create a defined number of variants. However, care is required for a proper and careful definition of the shuffling fragments to avoid the production of too many insoluble variants. Additionally, it is necessary to possess or develop a suitable screening assay for the desired enzyme property.

By means of computational and bioinformatic methods like the phylogenetic tree analysis, as well as the developed strategy for the creation of Hidden Markov loop profiles of previously defined loop regions, it could be shown that β/α surface loop regions of Old Yellow Enzymes could be divided into loops consisting of a more generalist motif, as well as loops possessing a specialist motif present in only a small number of closely related sequences. Furthermore, it was possible to show successfully that the generalist Loop A region within the entire superfamily of the Old Yellow Enzymes possesses an N-terminal conserved hinge region and a C-terminal conserved amino acid. Based on the length and character of the conserved amino acids of the N-terminal hinge region, as well as the nature of the C-terminal conserved amino acid, the entire reductase can be categorized into the same homologous subfamilies which are obtained by a phylogenetic tree analysis based on the overall sequence identity. The generalist type loop, Loop A, represents a small element within the enzyme which represents the evolutionary relationship of the entire enzyme within the family of the Old Yellow Enzymes. The common conserved residues within a homologous subfamily at the N- and C-terminal end of a loop region may stand for possible recombination site in the natural evolution of this enzyme family.

By the rational design of the loop length of β/α surface loop regions of Old Yellow Enzymes, as well as the rational loop grafting between two family members, it could be successfully demonstrated that such surface loop regions possess a considerable influence on enzymes properties. Therefore, the understanding of the nature of the influence is important to design rationally enzyme properties based on β/α surface loop regions. Within this work it was possible to show that mainly the specialist type loop region B possesses beneficial effects on enzyme activity, enantio- and stereo selectivity as well as stability. Loop A turned out be responsible for the *cis/trans* substrate specificity and represents a critical part within an enzyme concerning

stability issues. Furthermore, it was demonstrated successfully that it is possible to transfer one enzyme property to another by grafting of one loop region.

If one now combines all the knowledge obtained within this thesis over the role of each β/α surface loop region within an Old Yellow Enzyme, one should be able to create a tailor made biocatalysts with desired properties. A challenge would be, for example, the creation of a biocatalyst which converts both isomers of citral with absolute enantioselectivity to the industrially important (*R*)-citronellal. This task would require the combination of the properties of two enzymes: (1) the ability of NCR to accept both isomers of citral, neral as well as geranial, as substrate in order to reduce them both with absolute enantioselectivity to one product (in the case of NCR (*S*)-citronellal) (2) the absolute (*R*)-enantioselectivity of OYE1 and OYE2 in the reduction of the *trans* citral geranial.⁸ Based on the results obtained within this work, it would now be necessary to exchange the Loop A region, which was shown to demonstrate an influence on the *cis/trans* substrate specificity, of OYE1 against the corresponding region of NCR in order to transfer the NCR ability to accept both citral isomers as substrates in the OYE1 backbone. For the fine tuning of the enantioselectivity as well as, if desired, an improved thermo or solvent stability it would be worthwhile to alter rationally the loop length of the Loop B region. Ideally, these two alterations within the OYE1 enzyme should result in a variant which is able to accept both isomers of citral and reduce them to the (*R*)-enantiomeric product with high specificity in an organic solvent system at elevated temperatures. Further experiment will be necessary to prove this assumption. Overall, the rational loop design represents a valuable tool in protein engineering in order to influence enzyme properties.

6 References

1. Tawfik, D. S. Protein Dynamism and Evolvability. *Science*. **324**, 203–207 (2009).
2. Berg, J.M. Tymoczko, J.L., Stryer, L. *Biochemie* 45–80 (2003).
3. Faber, K. *Biotransformations in Organic Chemistry*. 3–9 (2004).
4. Berg, J.M., Tymoczko, J.L., Stryer, L. *Biochemie* 190–207 (2003).
5. Patthy, L. *Protein Evolution* 170–171 (2008).
6. Schmid, A., Dordick, J.S., Hauer, B., Kiener, A., Wubbolts, M., Witholt, B. Industrial biocatalysis today and tomorrow. *Nature* **409**, 258–68 (2001).
7. Wohlgemuth, R. Biocatalysis - key to sustainable industrial chemistry. *Curr. Opin. Biotechnol.* **21**, 713–24 (2010).
8. Müller, A., Hauer, B., Rosche, B. Asymmetric Alkene Reduction by Yeast Old Yellow Enzymes and by a Novel *Zymomonas mobilis* Reductase. *Biotechnology* **98**, 22–29 (2007).
9. Reich, S., Hoeffken, H.W., Rosche, B., Nestl, B.M., Hauer, B. Crystal Structure Determination and Mutagenesis Analysis of the Ene Reductase NCR. *ChemBiochem* 1–9 (2012). doi:10.1002/cbic.201200404
10. Stuermer, R., Hauer, B., Hall, M., Faber, K. Asymmetric bioreduction of activated C=C bonds using enoate reductases from the old yellow enzyme family. *Curr. Opin. Chem. Biol.* **11**, 203–13 (2007).
11. Wong, C.-H., Whitesides, G.M. Enzyme-Catalyzed Organic Synthesis: NAD(P)H Cofactor Regeneration by using Glucose 6-Phosphate and the Glucose-6-phosphate Dehydrogenase from *Leuconostoc mesenteroides*. *JACS* **103**, 4890–4899 (1981).
12. Massey, V. Introduction: Flavoprotein structure and mechanism. *FASEB J.* **9**, 473–475 (1995).
13. Fraaije, M.W., Mattevi, A. Flavoenzymes: diverse catalysts with recurrent features. *Trends Biochem. Sci.* **25**, 126–32 (2000).
14. De Colibus, L., Mattevi, A. New frontiers in structural flavoenzymology. *Curr. Opin. Struct. Biol.* **16**, 722–8 (2006).
15. Joosten, V., van Berkel, W.J.H. Flavoenzymes. *Curr. Opin. Chem. Biol.* **11**, 195–202 (2007).
16. Hefti, M.H., Vervoort, J., van Berkel, W.J.H. Deflavination and reconstitution of flavoproteins. *Eur. J. Biochem.* **270**, 4227–4242 (2003).

17. Van Berkel, W.J.H., Kamerbeek, N.M., Fraaije, M.W. Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J. Biotechnol.* **124**, 670–89 (2006).
18. Malito, E., Alfieri, A., Fraaije, M.W., Mattevi, A. Crystal structure of a Baeyer-Villiger monooxygenase. *Proc. Natl. Acad. Sci.* **101**, 13157–62 (2004).
19. Montellano, P.R.O. *Cytochrome P450: structure, mechanism and biochemistry*. (Plenum Publisher, 2005).
20. Warburg, O., Christian, W. Über das neue Oxydationsferment. *Naturwissenschaften* 980–981 (1932).
21. Williams, R.E., Bruce, N.C. “ New uses for an Old Enzyme ” – the Old Yellow Enzyme family of flavoenzymes. *Microbiology* **148**, 1607–1614 (2002).
22. Hall, M., Stueckler, C., Ehammer, H., Pointner, E., Oberdorfer, G., Gruber, K., Hauer, B., Stuermer, R., Kroutil, W., Macheroux, P., Faber, K. Asymmetric Bioreduction of C=C Bonds using Enoate Reductases OPR1, OPR3 and YqjM: Enzyme-Based Stereocontrol. *Adv. Synth. Catal.* **350**, 411–418 (2008).
23. Pietruszka, J., Schölzel, M. Ene Reductase-Catalysed Synthesis of (R)-Profen Derivatives. *Adv. Synth. Catal.* **354**, 751–756 (2012).
24. Stueckler, C., Hall, M., Ehammer, H., Pointner, E., Kroutil, W., Macheroux, P., Faber, K. Stereocomplementary Bioreduction of α,β Unsaturated Dicarboxylic Acids and Dimethyl Esters using Enoate Reductases: Enzyme- and Substrate-Based Stereocontrol. *Org. Lett.* **9**, 5409–5411 (2007).
25. Stueckler, C., Mueller, N.J., Winkler, C.K., Glueck, S.M., Gruber, K., Steinkellner, G., Faber, K. Bioreduction of α methylcinnamaldehyde derivatives: chemo-enzymatic asymmetric synthesis of Lilial TM and Helional TM †. *Dalt. Trans.* **39**, 8472–8476 (2010).
26. Stueckler, C., Winkler, C.K., Bonnekessel, M., Faber, K. Asymmetric Synthesis of (R)-3-Hydroxy-2-methylpropanoate (“Roche Ester”) and Derivatives via Biocatalytic C=C-Bond Reduction. *Adv. Synth. Catal.* **352**, 2663–2666 (2010).
27. Stueckler, C., Winler, C.K., Hall, M., Hauer, B., Bonnekessel, M., Zangger, K.F.K. Stereo-Controlled Asymmetric Bioreduction of α,β -Dehydroamino Acid Derivatives. *Adv. Synth. Catal.* **353**, 1169–1173 (2011).
28. Swiderska, M.A., Stewart, J.D. Stereoselective enone reductions by *Saccharomyces carlsbergensis* old yellow enzyme. *J. Mol. Catal. B Enzym.* **42**, 52–54 (2006).
29. Swiderska, M.A., Stewart, J.D. Asymmetric bioreductions of beta-nitro acrylates as a route to chiral beta2-amino acids. *Org. Lett.* **8**, 6131–6133 (2006).

30. Toogood, H.S., Gardiner, J.M., Scrutton, N.S. Biocatalytic Reductions and Chemical Versatility of the Old Yellow Enzyme Family of Flavoprotein Oxidoreductases. *ChemCatChem* **2**, 892–914 (2010).
31. Williams, R.E., Rathbone, D.A., Scrutton, N.S., Bruce, N.C. Biotransformation of Explosives by the Old Yellow Enzyme Family of Flavoproteins. *Appl. Environ. Microbiol.* **70**, 3566–3574 (2004).
32. Breithaupt, C., Kurzbauer, R., Schaller, F., Stintzi, A., Schaller, A., Huber, R., Macheroux, P., Clausen, T. Structural basis of substrate specificity of plant 12-oxophytodienoate reductases. *J. Mol. Biol.* **392**, 1266–77 (2009).
33. Breithaupt, C., Strassner, J., Breiting, U., Huber, R., Macheroux, P., Schaller, A., Clausen, T. X-ray structure of 12-oxophytodienoate reductase 1 provides structural insight into substrate binding and specificity within the family of OYE. *Structure* **9**, 419–29 (2001).
34. Barna, T., Messiha, H.L., Petosa, C., Bruce, N.C., Scrutton, N.S., Moody, P.C. Crystal structure of bacterial morphinone reductase and properties of the C191A mutant enzyme. *J. Biol. Chem.* **277**, 30976–83 (2002).
35. Fitzpatrick, T.B., Amrhein, N., Macheroux, P. Characterization of YqjM, an Old Yellow Enzyme homolog from *Bacillus subtilis* involved in the oxidative stress response. *J. Biol. Chem.* **278**, 19891–7 (2003).
36. Kitzing, K., Fitzpatrick, T.B., Wilken, C., Sawa, J., Bourenkov, G.P., Macheroux, P., Clausen, T. The 1.3 Å crystal structure of the flavoprotein YqjM reveals a novel class of Old Yellow Enzymes. *J. Biol. Chem.* **280**, 27904–13 (2005).
37. Saito, K., Thiele, D.J., Davio, M., Lockridge, O., Massey, V. The cloning and expression of a gene encoding Old Yellow Enzyme from *Saccharomyces carlsbergensis*. *J. Biol. Chem.* **266**, 20720–4 (1991).
38. Fox, K.M., Karplus, P.A. Old yellow enzyme at 2 Å resolution: overall structure, ligand binding, and comparison with related flavoproteins. *Structure* **2**, 1089–105 (1994).
39. Schaller, F., Hennig, P., Weiler, E. 12-Oxophytodienoate-10,11-reductase: occurrence of two isoenzymes of different specificity against stereoisomers of 12-oxophytodienoic acid. *Plant Physiol.* **118**, 1345–51 (1998).
40. Schittmeyer, M., Glieder, A., Uhl, M.K., Winkler, A., Zach, S., Schrittwieser, J.H., Kroutil, W., Macheroux, P., Gruber, K., Kambourakis, S., Rozzell, J.D., Winkler, M. Old Yellow Enzyme-Catalyzed Dehydrogenation of Saturated Ketones. *Adv. Synth. Catal.* **353**, 268–274 (2011).
41. Opperman, D.J., Sewell, B.T., Litthauer, D., Isupov, M.N., Littlechild, J.A., van Heerden, E. Crystal structure of a thermostable old yellow enzyme from *Thermus scotoductus* SA-01. *Biochem. Biophys. Res. Commun.* **393**, 426–31 (2010).

42. Opperman, D.J., Piater, L.A., van Heerden, E. A novel chromate reductase from *Thermus scotoductus* SA-01 related to old yellow enzyme. *J. Bacteriol.* **190**, 3076–82 (2008).
43. Xu, D., Kohli, R.M., Massey, V. The role of threonine 37 in flavin reactivity of the old yellow enzyme. *Proc. Natl. Acad. Sci.* **96**, 3556–61 (1999).
44. Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillios, D.C., Pogson, C.I., Wilson, I. A. Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 Å resolution. *Nature* **225**, 609–614 (1975).
45. Sterner, R., Höcker, B. Catalytic versatility, stability, and evolution of the (beta/alpha)₈-barrel enzyme fold. *Chem. Rev.* **105**, 4038–55 (2005).
46. Ochoa-Leyva, A., Soberón, F., Argüello, M., Montero-Morán, G., Saab-Rincón, G. Protein design through systematic catalytic loop exchange in the (beta/alpha)₈ fold. *J. Mol. Biol.* **387**, 949–64 (2009).
47. Ochoa-Leyva, A., Barona-Gómez, F., Saab-Rincón, G., Verdel-Aranda, K., Sánchez, F., Soberón, X. Exploring the Structure-Function Loop Adaptability of a (β/α)₈-Barrel Enzyme through Loop Swapping and Hinge Variability. *J. Mol. Biol.* **411**, 143–57 (2011).
48. Altamirano, M.M., Blackburn, J.M., Aguayo, C., Fersht, A.R. Directed evolution of new catalytic activity using the alpha/beta-barrel scaffold. *Nature* **403**, 617–22 (2000).
49. Cheon, Y.-H., Park, H.-S., Kim, J.-H., Kim, Y., Kim, H.-S. Manipulation of the active site loops of D-hydantoinase, a (beta/alpha)₈-barrel protein, for modulation of the substrate specificity. *Biochemistry* **43**, 7413–20 (2004).
50. Boersma, Y.L., Pijning, T., Bosma, M.S., van der Sloot, A.M., Godinho, L.F., Dröge, M.J., Winter, R.T., van Pouderoyen, G., Dijkstra, B.W., Quax, W.J. Loop grafting of *Bacillus subtilis* lipase A: inversion of enantioselectivity. *Chem. Biol.* **15**, 782–9 (2008).
51. Prokop, Z., Sato, Y., Brezovsky, J., Mozga, T., Chaloupkova, R., Koudelakova, T., Jerabek, P., Stephankova, V., Natsume, R., van Leeuwen, J.G.W., Janssen, D.B., Florian, J., Nagata, Y., Senda, T., Damborsky, J. Enantioselectivity of Haloalkane Dehalogenases and its Modulation by Surface Loop Engineering. *Angew. Chemie* **122**, 6247–6251 (2010).
52. Yedavalli, P., Madhusudhana Nalam., R. Engineering the loops in a lipase for stability in DMSO. *Protein Eng.* **26**, 317–324 (2013).
53. Farber, G. K. An alpha/beta barrel full of evolutionary trouble. *Curr. Opin. Struct. Biol.* **3**, 409–412 (1993).
54. Karplus, P.A., Fox, K.M., Massey, V. Structure-function relations for old yellow enzyme. *FASEB J.* **9**, 1518–1526 (1995).

55. Kohli, R.M., Massey, V. The Oxidative Half-reaction of Old Yellow Enzyme. *J. Biol. Chem.* **273**, 32763–32770 (1998).
56. Brown, B.J., Deng, Z., Karplus, P.A., Massey, V. On the Active Site of Old Yellow Enzyme. *Biochemistry* **273**, 32753–32762 (1998).
57. Hall, M., Bommarius, A.S. Enantioenriched compounds via enzyme-catalyzed redox reactions. *Chem. Rev.* **111**, 4088–110 (2011).
58. Vaz, A.D.N., Chakraborty, S., Massey, V. Old Yellow enzyme: aromatization of cyclic enones and the mechanism of a novel dismutation reaction. *Biochemistry* **34**, 4246–56 (1995).
59. Fryszkowska, A., Toogood, H.S., Sakuma, M., Gardiner, J.M., Stephens, G.M., Scrutton, N.S. Asymmetric Reduction of Activated Alkenes by Pentaerythritol Tetranitrate Reductase: Specificity and Control of Stereochemical Outcome by Reaction Optimisation. *Adv. Synth. Catal.* **351**, 2976–2990 (2009).
60. Hall, M., Stueckler, C., Hauer, B., Stuermer, R., Friedrich, T., Breuer, M., Kroutil, W., Faber, K. Asymmetric Bioreduction of Activated C=C Bonds Using *Zymomonas mobilis* NCR Enoate Reductase and Old Yellow Enzymes OYE 1–3 from Yeasts. *European J. Org. Chem.* **2008**, 1511–1516 (2008).
61. Hall, M., Stueckler, C., Kroutil, W., Macheroux, P., Faber, K. Asymmetric bioreduction of activated alkenes using cloned 12-oxophytodienoate reductase isoenzymes OPR-1 and OPR-3 from *Lycopersicon esculentum* (tomato): a striking change of stereoselectivity. *Angew. Chem. Int. Ed.* **46**, 3934–7 (2007).
62. Messiha, H.L., Munro, A.W., Bruce, N.C., Barsukov, I., Scrutton, N.S. Reaction of morphinone reductase with 2-cyclohexen-1-one and 1-nitrocyclohexene: proton donation, ligand binding, and the role of residues Histidine 186 and Asparagine 189. *J. Biol. Chem.* **280**, 10695–709 (2005).
63. Mueller, N.J., Stueckler, C., Hauer, B., Baudendistel, N., Housden, H., Bruce, N.C., Faber, K. The Substrate Spectra of Pentaerythritol Tetranitrate Reductase, Morphinone Reductase, N-Ethylmaleimide Reductase and Estrogen-Binding Protein in the Asymmetric Bioreduction of Activated Alkenes. *Adv. Synth. Catal.* **352**, 387–394 (2010).
64. Breithaupt, C., Strassner, J., Breiting, U., Huber, R., Macheroux, P., Schaller, A., Clausen, T. X-ray structure of 12-oxophytodienoate reductase 1 provides structural insight into substrate binding and specificity within the family of OYE. *Structure* **9**, 419–29 (2001).
65. Turner, J.G., Ellis, C., Devoto, A. The Jasmonate Signal Pathway. *Plant Cell* **14**, S153–S164 (2002).
66. Craig, D.H., Moody, P.C., Bruce, N.C., Scrutton, N.S. Reductive and oxidative half-reactions of morphinone reductase from *Pseudomonas putida* M10: a kinetic and thermodynamic analysis. *Biochemistry* **37**, 7598–607 (1998).

67. Leuenberger, H.G.W., Boguth, W., Widmer, E., Zell, R. Synthese von optisch aktiven, natürlichen Carotinoiden und strukturell verwandten Naturprodukten. *Helv. Chim. Acta* **59**, 1832–1849 (1976).
68. Durchschein, K., Hall, M., Faber, K. Unusual reactions mediated by FMN-dependent ene- and nitro-reductases. *Green Chem.* **15**, 1764 (2013).
69. Müller, A., Stürmer, R., Hauer, B., Rosche, B. Stereospecific alkyne reduction: novel activity of old yellow enzymes. *Angew. Chemie Int. Ed.* **46**, 3316–8 (2007).
70. Stueckler, C., Reiter, T.C., Baudendistel, N., Faber, K. Nicotinamide-independent asymmetric bioreduction of C=C-bonds via disproportionation of enones catalyzed by enoate reductases. *Tetrahedron* **66**, 663–667 (2010).
71. Barna, T.M., Khan, H., Bruce, N.C., Barsukov, I., Scrutton, N.S., Moody, P.C.E. Crystal structure of pentaerythritol tetranitrate reductase: “flipped” binding geometries for steroid substrates in different redox states of the enzyme. *J. Mol. Biol.* **310**, 433–47 (2001).
72. Meah, Y., Brown, B.J., Chakraborty, S., Massey, V. Old yellow enzyme: reduction of nitrate esters, glycerin trinitrate, and propylene 1,2-dinitrate. *Proc. Natl. Acad. Sci.* **98**, 8560–5 (2001).
73. Durchschein, K., Wallner, S., Macheroux, P., Zangger, K., Fabian, W.M.F., Faber, K. Unusual C=C Bond Isomerization of an α,β -Unsaturated γ -Butyrolactone Catalysed by Flavoproteins from the Old Yellow Enzyme Family. *ChemBioChem* **13**, 2346–51 (2012).
74. Winkler, C.K., Stueckler, C., Mueller, N.J., Pressnitz, D., Faber, K. Asymmetric Synthesis of O-Protected Acyloins Using Enoate Reductases: Stereochemical Control through Protecting Group Modification. *European J. Org. Chem.* **2010**, 6354–6358 (2010).
75. Winkler, C.K., Clay, D., Davies, S., O'Neill, P., McDaid, P., Debarge, S., Steflík, J., Karmilowicz, M., Wong, J.W., Faber, K. Chemoenzymatic Asymmetric Synthesis of Pregabalin-Precursors via Asymmetric Bioreduction of beta-Cyano-Acrylate Esters Using Ene-Reductases. *J. Org. Chem.* (2013). doi:10.1021/jo302484p
76. Bommarius, A.S., Blum, J.K., Abrahamson, M.J. Status of protein engineering for biocatalysts: how to design an industrially useful biocatalyst. *Curr. Opin. Chem. Biol.* **15**, 194–200 (2011).
77. Bornscheuer, U.T., Huisman, G.W., Kazlauskas, R.J., Lutz, S., Moore, J.C., Robins, K. Engineering the third wave of biocatalysis. *Nature* **485**, 185–194 (2012).
78. Arnold, F.H. Combinatorial and computational challenges for biocatalyst design. *Nature* **409**, 253–7 (2001).

79. Hogrefe, H.H., Cline, J., Youngblood, G.L., Allen, R.M. Creating Randomized Amino Acid Libraries with the QuikChange Multi Site-directed Mutagenesis Kit. *Biotechniques* **33**, 1158–1165 (2002).
80. Broun, P. Catalytic Plasticity of Fatty Acid Modification Enzymes Underlying Chemical Diversity of Plant Lipids. *Science*. **282**, 1315–1317 (1998).
81. Shanklin, J., Cahoon, E.B. Desaturation and related modification of fatty acids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 611–614 (1998).
82. Shanklin, J. Exploring the possibilities presented by protein engineering. *Curr. Opin. Plant Biol.* **3**, 243–8 (2000).
83. Arnold, F.H. Design by Directed Evolution. *Acc. Chem. Res.* **31**, 125–131 (1998).
84. Arnold, F.H. Directed evolution of enzyme catalysts. *Reviews* (1997).
85. Arnold, F.H. How proteins adapt: lessons from directed evolution. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 41–6 (2009).
86. Farinas, E.T., Bulter, T., Arnold, F.H. Directed enzyme evolution. *Curr. Opin. Biotechnol.* **12**, 545–51 (2001).
87. Cramer, A., Raillard, S.A., Bermudez, E., Stemmer, W.P.C. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288–91 (1998).
88. Stemmer, W.P.C. DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci.* **91**, 10747–10751 (1994).
89. Stemmer, W.P.C. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389–391 (1994).
90. Cramer, A., Dawes, G., Rodriguez Jr., E., Silver S., Stemmer, W.P.C. Molecular evolution of an arsenate detoxification pathway by DNA shuffling. *Nature* **15**, 436–438 (1997).
91. Bougioukou, D., Kille, S., Taglieber, A., Reetz, M.T. Directed Evolution of an Enantioselective Enoate-Reductase: Testing the Utility of Iterative Saturation Mutagenesis. *Adv. Synth. Catal.* **351**, 3287–3305 (2009).
92. Reetz, M.T., Carballeira, J.D., Vogel, A. Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angew. Chemie Int. Ed.* **45**, 7745–51 (2006).
93. Bornscheuer, U.T., Pohl, M. Improved biocatalysts by directed evolution and rational protein design. *Curr. Opin. Chem. Biol.* **5**, 137–43 (2001).

94. Cedrone, F., Ménez, A., Quéméneur, E. Tailoring new enzyme functions by rational redesign. *Curr. Opin. Struct. Biol.* **10**, 405–10 (2000).
95. Röthlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J.L., Althoff, E.A., Zanghellini, A., Dym, O., Albeck, S., Houk, K.N., Tawfik, D.S., Baker, D. Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–5 (2008).
96. Reetz, M.T. Laboratory evolution of stereoselective enzymes: a prolific source of catalysts for asymmetric reactions. *Angew. Chemie Int. Ed.* **50**, 138–74 (2011).
97. Chen, K., Arnold, F.H. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci.* **90**, 5618–22 (1993).
98. Reetz, M.T., Bocola, M., Carballeira, J.D., Zha, D., Vogel, A. Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. *Angew. Chemie Int. Ed.* **44**, 4192–6 (2005).
99. Reetz, M.T., Soni, P., Fernández, L., Gumulya, Y., Carballeira, J.D. Increasing the stability of an enzyme toward hostile organic solvents by directed evolution based on iterative saturation mutagenesis using the B-FIT method. *Chem. Commun.* **46**, 8657–8 (2010).
100. Carballeira, J.D., Krumlinde, P., Bocola, M., Vogel, A., Reetz, M.T., Bäckvall, J. E. Directed evolution and axial chirality: optimization of the enantioselectivity of *Pseudomonas aeruginosa* lipase towards the kinetic resolution of a racemic allene. *Chem. Commun.* 1913–1315 (2007). doi:10.1039/b700849j
101. Clouthier, C.M., Kayser, M.M., Reetz, M.T. Designing new Baeyer-Villiger monooxygenases using restricted CASTing. *J. Org. Chem.* **71**, 8431–7 (2006).
102. Bartsch, S., Kourist, R., Bornscheuer, U.T. Complete inversion of enantioselectivity towards acetylated tertiary alcohols by a double mutant of a *Bacillus subtilis* esterase. *Angew. Chemie Int. Ed.* **47**, 1508–11 (2008).
103. Engler, C., Gruetzner, R., Kandzia, R., Marillonnet, S. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One* **4**, e5553 (2009).
104. Engler, C., Kandzia, R., Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, e3647 (2008).
105. Roberts, R.J. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* **31**, 1805–1812 (2003).
106. Pingoud, A., Jeltsch, A. Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* **29**, 3705–3727 (2001).

107. Engler, C., Marillonnet, S. *cDNA Libraries*. **729**, 167–181 (Humana Press, 2011).
108. Reich, S., Kress, N., Nestl, B.M., Hauer, B. Variations in the stability of NCR ene reductase by rational enzyme loop modulation. *J. Struct. Biol.* (2013). doi:10.1016/j.jsb.2013.04.004
109. Reich, S. Site-directed mutagenesis and loop exchange at NCR, an enoate reductase from *Zymomonas mobilis*. (2010).
110. Morlock, L. Directed Evolution of the Ene Reductase NCR from *Zymomonas mobilis* by Golden Gate Shuffling. (2013).
111. HMMER. <http://hmmer.janelia.org> at <<http://hmmer.janelia.org>>
112. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
113. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W. GenBank. *Nucleic Acids Res.* **37**, D26–31 (2009).
114. Thompson, J.D., Higgins, D.G., Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–80 (1994).
115. Needleman, S.B., Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443–53 (1970).
116. Felsenstein, J. PHYLIP (Phylogeny Inference Package). (2005).
117. Forneris, F., Orru, R., Bonivento, D., Chiarelli, L.R., Mattevi, A. ThermoFAD, a Thermofluor-adapted flavin ad hoc detection system for protein folding and ligand binding. *FEBS J.* **276**, 2833–40 (2009).
118. Hall, M., Hauer, B., Stuermer, R., Kroutil, W., Faber, K. Asymmetric whole-cell bioreduction of an α,β -unsaturated aldehyde (citral): competing prim-alcohol dehydrogenase and C–C lyase activities. *Tetrahedron: Asymmetry* **17**, 3058–3062 (2006).
119. Walton, A.Z., Conerly, W.C., Pompeu, Y., Sullivan, B., Stewart, J.D. Biocatalytic Reductions of Baylis-Hillman Adducts. *ACS Catal.* **1**, 989–993 (2011).
120. Needleman, S.B., Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443–53 (1970).

121. Padhi, S.K., Bougioukou, D.J., Stewart, J.D. Site-saturation mutagenesis of tryptophan 116 of *Saccharomyces pastorianus* old yellow enzyme uncovers stereocomplementary variants. *J. Am. Chem. Soc.* **131**, 3271–80 (2009).
122. French, C.E., Bruce, N.C. Bacterial morphinone reductase is related to Old Yellow Enzyme. *Biochem. J.* **312**, 671–678 (1995).
123. French, C.E., Bruce, N.C. Purification and characterization of morphinone reductase from *Pseudomonas putida* M10. *Biochem. J.* **301**, 97–103 (1994).
124. Messiha, H.L. Role of Active Site Residues and Solvent in Proton Transfer and the Modulation of Flavin Reduction Potential in Bacterial Morphinone Reductase. *J. Biol. Chem.* **280**, 27103–27110 (2005).
125. Yanto, Y., Yu, H.-H., Hall, M., Bommarius, A.S. Characterization of xenobiotic reductase A (XenA): study of active site residues, substrate spectrum and stability. *Chem. Commun.* **46**, 8809–11 (2010).
126. Durchschein, K., Fabian, W.M.F., Macheroux, P., Zangger, K., Trimmel, G., Faber, K. Reductive biotransformation of nitroalkenes via nitroso-intermediates to oxazetes catalyzed by xenobiotic reductase A (XenA). *Org. Biomol. Chem.* **9**, 3364–9 (2011).
127. Spiegelhauer, O., Mende, S., Dickert, F., Knauer, S.H., Ullmann, G.M., Dobbek, H. Cysteine as a modulator residue in the active site of xenobiotic reductase A: a structural, thermodynamic and kinetic study. *J. Mol. Biol.* **398**, 66–82 (2010).
128. Spiegelhauer, O., Werther, T., Mende, S., Knauer, S.H., Dobbek, H. Determinants of substrate binding and protonation in the flavoenzyme xenobiotic reductase A. *J. Mol. Biol.* **403**, 286–98 (2010).
129. Griese, J.J., P Jakob, R., Schwarzinger, S., Dobbek, H. Xenobiotic reductase A in the degradation of quinoline by *Pseudomonas putida* 86: physiological function, structure and mechanism of 8-hydroxycoumarin reduction. *J. Mol. Biol.* **361**, 140–52 (2006).
130. Segura, A., Godoy, P., van Dillewijn, P., Hurtado, A., Arroyo, N., Santacruz, S., Ramos, J.-L. Proteomic Analysis Reveals the Participation of Energy- and Stress-Related Proteins in the Response of *Pseudomonas putida* DOT-T1E to Toluene Proteomic Analysis Reveals the Participation of Energy- and Stress-Related Proteins in the Response of *Pseudomonas*. *J. Bacteriol.* **187**, 5937–5945 (2005).
131. Blehert, D.S., Fox, B.G., Chambliss, G.H. Cloning and Sequence Analysis of Two *Pseudomonas* Flavoprotein Xenobiotic Reductases Cloning and Sequence Analysis of Two *Pseudomonas* Flavoprotein Xenobiotic Reductases. *J. Bacteriol.* **181**, (1999).
132. Adalbjörnsson, B.V., Toogood, H.S., Fryszkowska, A., Pudney, C.R., Jowitt, T.A., Leys, D., Scrutton, N.S. Biocatalysis with thermostable enzymes:

- structure and properties of a thermophilic “ene”-reductase related to old yellow enzyme. *ChemBiochem* **11**, 197–207 (2010).
133. Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).
134. Grodberg, J., Davis, K.L., Sykowski, A.J. Alanine scanning mutagenesis of human erythropoietin identifies four amino acids which are critical for biological activity. *Eur. J. Biochem.* **218**, 597–601 (1993).
135. Ashenazi, A., Presta, L.G., Marsters, S.A., Camerato, T.R., Rosenthal, K.A., Fendly, B.M., Capon, D.J. Mapping the CD4 binding site for human immunodeficiency virus by alanine-scanning mutagenesis. *Proc. Natl. Acad. Sci.* **87**, 7150–4 (1990).
136. Howlader, M.T.H., Kagawa, Y., Miyakawa, A., Yamamoto, A., Taniguchi, T., Hayakawa, T., Sakai, H. Alanine scanning analyses of the three major loops in domain II of *Bacillus thuringiensis* mosquitocidal toxin Cry4Aa. *Appl. Environ. Microbiol.* **76**, 860–5 (2010).
137. Gkountelias, K., Tselios, T., Venihaki, M., Deraos, G., Lazaridis, I., Rassouli, O., Gravanis, A., Liapakis, G. Alanine Scanning Mutagenesis of the Second Extracellular Loop of Type 1 Corticotropin-Releasing Factor Receptor Revealed Residues Critical for Peptide Binding. *Mol. Pharma.* **75**, 793–800 (2009).
138. Gargiulo, S., Opperman, D.J., Hanefeld, U., Arends, I.W.C.E., Hollmann, F. A biocatalytic redox isomerisation. *Chem. Commun.* **48**, 6630–6632 (2012).
139. Brenna, E., Gatti, F.G., Monti, D., Parmeggiani, F., Sacchetti, A. Cascade Coupling of Ene Reductases with Alcohol Dehydrogenases: Enantioselective Reduction of Prochiral Unsaturated Aldehydes. *ChemCatChem* **4**, 653–659 (2012).
140. Li, W., Liu, B., Yu, L., Feng, D., Wang, H., Wang, J. Phylogenetic analysis, structural evolution and functional divergence of the 12-oxo-phytodienoate acid reductase gene family in plants. *BMC Evol. Biol.* **9**, 90 (2009).
141. Park, H.-S., Nam, S.-H., Lee, J.K., Yoon, C.N., Mannervik, B., Benkovic, S.J., Kim, H.-S. Design and evolution of new catalytic activity with an existing protein scaffold. *Science* **311**, 535–8 (2006).
142. Boersma, Y.L., Pijning, T., Bosma, M.S., van der Sloot, A.M., Godinho, L.F., Dröge, M.J., Winter, R.T., van Pouderooyen, G., Dijkstra, B.W., Quax, W. J. Loop grafting of *Bacillus subtilis* lipase A: inversion of enantioselectivity. *Chem. Biol.* **15**, 782–9 (2008).
143. Prokop, Z., Sato, Y., Brezovsky, J., Mozga, T., Chaloupkova, R., Koudelakova, T., Jerabek, P., Stepankova, V., Natsume, R., van Leeuwen, J.G.E., Janssen, D.B., Florian, J., Nagata, Y., Senda, T., Damborsky, J. Enantioselectivity of Haloalkane Dehalogenases and its Modulation by Surface Loop Engineering. *Angew. Chemie* **122**, 6247–6251 (2010).

144. Daugherty, A.B., Govindarajan, S., Lutz, S. Improved biocatalysts from a synthetic circular permutation library of the flavin-dependent oxidoreductase old yellow enzyme. *J. Am. Chem. Soc.* **135**, 14425–32 (2013).
145. Kabsch, W. Xds. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–32 (2010).
146. Langer, G., Cohen, S.X., Lamzin, V.S., Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–9 (2008).
147. Project, C.C. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **50**, 760–3 (1994).
148. Cen, V.B., Arendall III, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 12–21 (2010).

7 Supplementary Material

7.1 Proteins, vectors, vector constructs, primers and strains

7.1.1 Proteins

NCR ene reductase from *Zymomonas mobilis* (UniProtKB Q5NLA1)

MPSLFDPIRFGAFTAKNRIWMAPLTRGRATRDHVPTEIMAEYYAQRASAGLIISEATGISQEGLGWPYAPGIWSDA
QVEAWLPITQAVHDAGGLIFAQLWHMGRMVPSNVSGMQPVAPSASQAPGLGHTYDGKKPYDVARALRLDEIPRL
LDDYEKAARHALKAGFDGVQIHAANGYLIDEFIRDSTNHRHDEYGGAVENRIRLLKDVTERVIATIGKERTAVRLSP
NGEIQGTVDSHPEQVFIPAAKMLSDDLIAFLGMREGAVDGTGFKTDQPKLSPEIRKVFKPPLVLNQDYTFETAQAA
LDSGVADAISFGRPFIGNPDLPRRFFEKAPLTKDVIETWYTQTPKGYTDYPLLGD

OYE1 ene reductase from *Saccharomyces pastorianus* (UniProtKB Q02899.3)

MSFVKDFKPQALGDTNLFKPIKIGNNELLHRAVIPPLTRMRALHPGNIPNRDWAWEYYTQRAQRPGTMIITEGAFIS
PQAGGYDNAPGVWSEEQMVEWTKIFNAIHEKKSFWVQLWVLGWAAFPDNLARDGLRYDSASDNVFMDAEQE
AKAKKANNPQHSLTKDEIKQYIKEYVQAAKNSIAAGADGVEIHSANGYLLNQFLDPHSNTRTDEYGGSIENRARFT
LEVVDALVEAIGHEKVGLRLSPYGVFNMSGGAETGIVAQYAYVAGELEKRAKAGKRLAFVHLVEPRVTNPFLTE
GEGEYEGGSNDFVYSIWKGPVIRAGNFALHPEVVREEVKDKRTLIGYGRFFISNPDLVDRLEKGLPLNKYDRDFTY
QMSAHGYIDYPTYEEALKLGWDDK

MR ene reductase from *Pseudomonas putida* M10 (UniProtKB Q51990)

MPDTSFSNPGLFTPLQLGSLSLPNRVIMAPLTRSRTPDPSVPGRLQIYYGQRASAGLIISEATNISPTARGVYVYTPG
IWTDAQEAGWKGVEAVHAKGGRIALQLWHVGRVSHELVQPDGQQPVAPSALKAEGAECFVEFEDGTAGLHPT
STPRALETDGIPGIVEDYRQAAQRAKRAFDMVEVHAANACLPNQFLATGTNRRTDQYGGSIENRARFPLEVVDA
VAEVFGPERVGIRLTPFLELFLGLTDDEPEAMAFYLAGELDRRGLAYLHFNEPDWIGGDITYPEGFREQMRQRFKG
GLIYCGNYDAGRAQARLDDNTADAVAFGRPFIANPDLPERFRLGAALNEPDPSTFYGGAEVGYTDYPFLDNGHD
RLG

7.1.2 Vectors

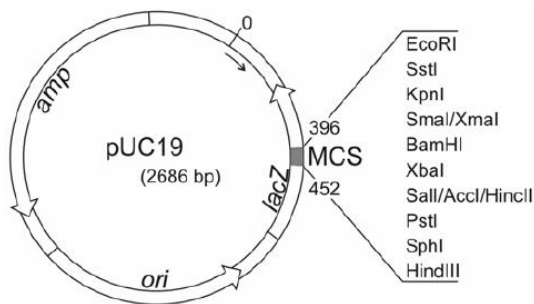


Figure 7.1: Cloning vector. pUC19 plasmid possessing an ampicillin resistance and a LacZ gene for the blue/white screening was used for the Golden Gate Shuffling.

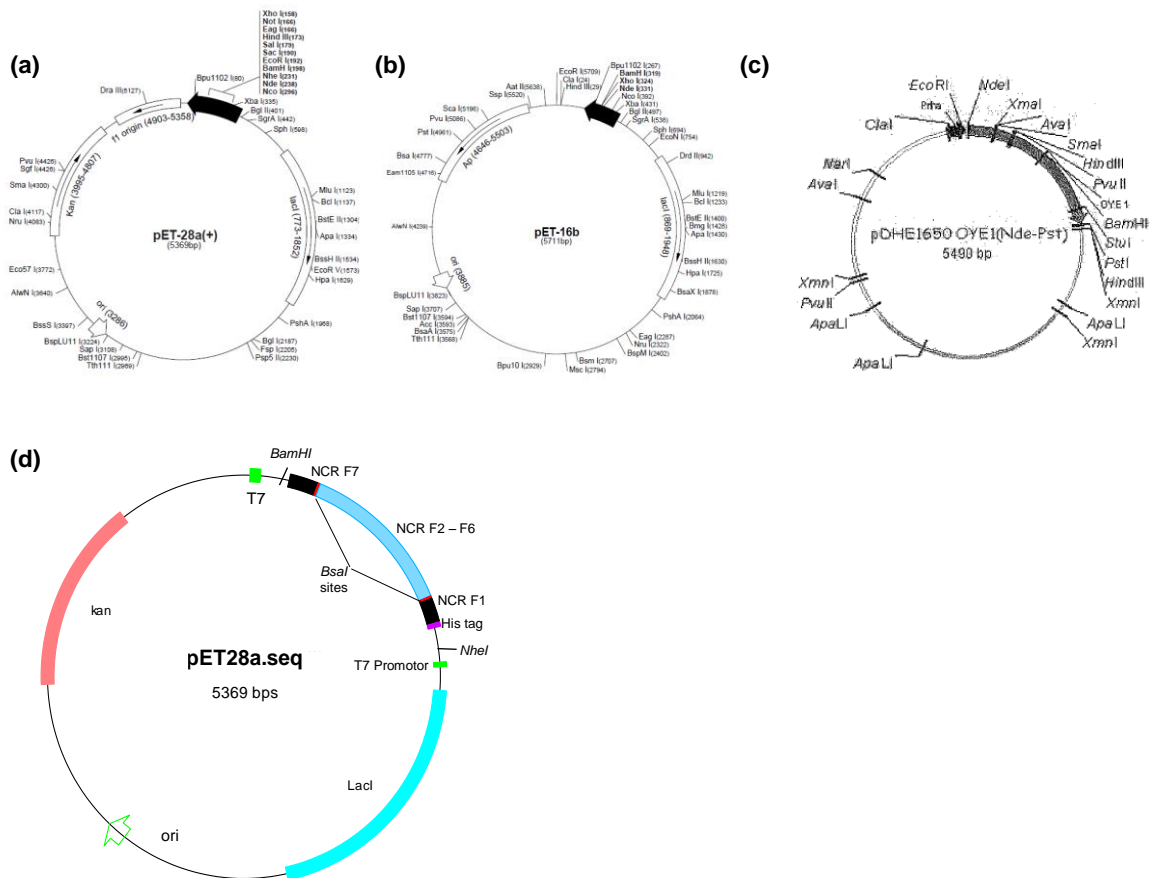


Figure 7.2: Used expression vectors. (a) pET-28a (+) vector (Novagen) was used for the expression of wild type NCR, as well as all designed loop and alanine scanning mutagenesis variants, with a N-terminal hexa-histidin tag. It was also used as expression vector for the Golden Gate Shuffling. (b) pET-16b vector (Novagen) was used for the expression of wild type MR with a N-terminal hexa-histidin tag. (c) pDHE1650 vector (BASF) was used for the expression of wild type OYE1 without any his tag. (d) Modified pET28a (+)_NCR_Bsal vector as expression vector for the Golden Gate Shuffling. In the NCR gene two *Bsal* restriction sites are inserted after fragment 1 and in front of fragment 7.

7.1.3 Vector constructs

Table 7.1: List of used and designed vector constructs. Unless marked otherwise pET28a (+) was used as vector.

Name	Gene Insert	Mutation	Available at ITB	ITB No.
NCR	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	-	yes	pITB770
OYE1 _{pDHE}	Old yellow enzyme 1 from <i>S. pastorianus</i>	-	yes	pITB772
MR _{pET16b}	Morphinone Reductase from <i>P. putida</i> M10	-	yes	pITB778
M2 Y177A	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	Y177A	yes	pITB787
L225A	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	L225A	yes	pITB791
S226A	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	S226A	yes	pITB792
P227A	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	P227A	yes	pITB793
N228A	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	N228A	yes	pITB794
G229A	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	G229A	yes	pITB795
E230A	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	E230A	yes	pITB796
I231A	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	I231A	yes	pITB797
Q232A	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	Q232A	yes	pITB798
G233A	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	G233A	yes	pITB799
T234A	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	T234A	yes	pITB800
V235A	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	V235A	yes	pITB801
D236A	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	D236A	yes	pITB802
S237A	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	S237A	yes	pITB803
H238A	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	H238A	yes	pITB804
P239A	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	P239A	yes	pITB805
E240A	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	E240A	yes	pITB806
GGs10	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	Combination: F2_NCR, F3_OYE1; F4_NCR; F5_NCR; F6_NCR	Yes	-
GGs174	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	Combination: F2_PETN, F3_PETN; F4_NCR; F5_PETN; F6_PETN	yes	pITB847
GGs222	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	Combination not determined	Yes	-
GGs225	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	Combination: F2_PETN; F3_NCR; F4_NCR; F5_NCR; F6_YqjM	yes	pITB848
GGs229	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	Combination: F2_PETN; F3_PETN; F4_NCR; F5_YqjM; F6_YqjM	yes	pITB849

Table 7.1: List of used and designed vector constructs (continued)

Name	Gene Insert	Mutation	Available at ITB	ITB No.
L3_Short	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	Deletion of 7 aa from E230 to D236	yes	plTB817
L3_Long	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	Insertion of 3 alanine between Q232 and G233	yes	plTB818
L3_OYE1	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	Exchange of aa P227-E240 of NCR against aa P247-V263 of OYE1	yes	plTB813
L3_MR	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	Exchange of aa P227-E240 of NCR against aa P241-E254 of MR	yes	plTB814
L4_Short	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	Deletion of 4 aa from G267 to G270	yes	plTB819
L4_OYE1	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	Exchange of aa M260-S278 of NCR against aa L288-N310 of OYE1	yes	plTB816
L4_MR	2-cyclohexen-1-one reductase from <i>Z. mobilis</i>	Exchange of aa M260-S278 of NCR against aa N275-R291 of MR	yes	plTB815

7.1.4 Primers

Table 7.2: Primers for Golden Gate Shuffling. (a) Overlapping extension PCR primers for the insertion of two *Bsal* restriction sites in the NCR gene cloned in pET28a(+). In the first PCR step three fragments were created with the six designed primers. In a second PCR step the three fragments were used as template with the F₁ forward and the R₃ reverse primer. Last step was a double digestion with *NheI* and *BamHI* and the subsequent ligation in the also digested pET28a(+). (b) Primers for the amplification of the defined fragments in the five selected ene reductases NCR, OYE1, PETN, YqjM and GkOYE with the insertion of a *Bsal* restriction site, as well as a blunt or sticky end restriction site.

(a) <i>Bsal</i> restriction site insertion		
	Primer	Sequence (5' – 3')
NCR <i>Bsal</i> _{Site1}	F ₁ R ₁	CCC CCC GCT AGC ATG CCT AGC TTG TTT GAT CCC ATC GAC TAA TGC CAG TCG GGT CTC A <u>CTTC</u> TGA AAT AAT CAG TCC GGC GCT
NCR <i>Bsal</i> _{Site1/2}	F ₂ R ₂	<u>GAAG</u> TGAG ACC CGA CTG GCA TTA GTC AGG AAG <u>AGTA</u> G GAGACC TAA TCC TGA TTA AGA ACC AAA GGC GGT TTG
NCR <i>Bsal</i> _{Site1}	F ₃ R ₃	GA TTA GGT CTC C <u>TACT</u> TTT GAA ACC GCG CAA GCT GC CCC GGA TCC TCA ATC CCC AAG CAA AGG ATA ATC
(b) Golden Gate Shuffling fragments		
	Primer	Sequence (5' – 3')
NCR F2	F R	GGTCTC A <u>GAAG</u> CGA CTG GCA TTA GTC AGG GGT CTC A <u>GC TG G</u> GCA AAG ATA AGA CCG CC
NCR F3	F R	CCC AAG CTT GGTCTC C <u>CAGC</u> TATGGCATATGGGACGTATGG CCC <u>TCT AGA</u> GGT CTC C <u>CGGC</u> C TTC AAG GCA TGA CG
NCR F4	F R	GGTCTC G <u>GCCG</u> GTT TTG ATG GCG TAC AGA GGT CTC G <u>TGAT</u> AAC CTA ACG GCC GTC CGC
NCR F5	F R	GGTCTC T <u>ATCA</u> CCGAATGGTGAATACAG GGG GGTCTC T <u>ACCC</u> AAA TCA GAC AAC ATT TTG GCC GCC
NCR F6	F R	GGTCTC T <u>GGGT</u> ATT GCC TTT TTA GGG ATG CGA GAA GGG GGTCTC A <u>AG TA</u> TAA TCC TGA TTA AGA ACC AAA GGC
OYE1 F2	F R	GGTCTC C <u>GAAG</u> GCGCGTTTATTAGCCCGC GGTCTC A <u>GCTG</u> CAC CCA CAC AAA GCT TTT TTT TTC

Table 7.2: Primers for Golden Gate Shuffling fragments with blunt and sticky end ligation (continued)

OYE1 F3	F R	CCC AAG CTT GGTCTC G <u>CAGC</u> TGT GGG TGC TGG GCT GGG CCC <u>TCT AGA GGTCTC</u> C <u>CGGC</u> CGC AAT GCT GTT TTT CGC CGC
OYE1 F5	F R	CCC AAG CTT GGTCTC T <u>ATCA</u> CCG TAT GGC GTG TTT AAC AGC ATG CCC <u>TCT AGA GGTCTC</u> T <u>ACCC</u> CGT TTT TCC AGT TCG CCC GCC ACA
OYE1 F6	F R	CCC AAG CTT GGTCTC G <u>GGGT</u> AAAGCGGGCAAACGCCTG CCC <u>TCT AGA GGTCTC</u> G <u>AGTA</u> AAG TTG CCC GCG CGA ATC ACC
PETN F2	F R	CCC AAG CTT GGTCTC C <u>GAAG</u> CGA CCC AGA TTA GCG C CCC <u>TCT AGA GGTCTC</u> A <u>GCTG</u> CAC CGC AAT GCG GCC
PETN F3	F R	CCC AAG CTT GGTCTC G <u>CAGC</u> TGT GGC ATA CCG GCC CCC <u>TCT AGA GGTCTC</u> C <u>CGGC</u> TTC GCG CGC GTT CGC
PETN F5	F R	CCC AAG CTT GGTCTC T <u>ATCA</u> CCG ATT GGC ACC TTT CAG AAC G CCC <u>TCT AGA GGTCTC</u> T <u>ACCC</u> CGT TTC GCC AGT TCT TCA ATC
PETN F6	F R	CCC AAG CTT GGTCTC G <u>GGGT</u> ATT GCG TAT CTG CAT ATG AGC CCC <u>TCT AGA GGTCTC</u> C <u>AGTA</u> TAC GCG CCC GCG CC
YqjM F2	F R	GGTCTC A <u>GAAG</u> CGT CAG CGG TTA ACC CTC AAG GGTCTC A <u>GCTG</u> AAT GCC GAT TTT TGA ACC TTG TTC TTT GAC
YqjM F3	F R	CCC AAG CTT GGTCTC T <u>CAGC</u> TTG CCC ATG CCG GAC CCC <u>TCT AGA GGTCTC</u> C <u>CGGC</u> TTC TTT TGC GCG GGC AGC
YqjM F5	F R	CCC AAG CTT GGTCTC T <u>ATCA</u> GCT TCT GAC TAC ACT GAT AAA GGC CCC <u>TCT AGA GGTCTC</u> C <u>ACCC</u> TGC TCC TTC ATC CAT TTT G
YqjM F6	F R	CCC AAG CTT GGTCTC A <u>GGGT</u> GTT GAC TTA ATT GAC TGC AGC CCC <u>TCT AGA GGTCTC</u> C <u>AGTA</u> ATC ATG CCG ACG GCA CCA GTA
GkOYE F2	F R	CCC AAG CTT GGTCTC T <u>GAAG</u> CGA CCG GCG TGA CGC CCC <u>TCT AGA GGTCTC</u> A <u>GCTG</u> GAT GCC GAT GGC CGC
GkOYE F3	F R	CCC AAG CTT GGTCTC C <u>CAGC</u> TTG CCC ATG CGG GGA G CCC <u>TCT AGA GGTCTC</u> C <u>CGGC</u> TTC CTT CGC GCG CCG
GkOYE F5	F R	CCC AAG CTT GGTCTC T <u>ATCA</u> GCG TCC GAC TAC CAT CCG CCC <u>TCT AGA GGTCTC</u> C <u>ACCC</u> TGT TCT TTC ATC CGC TTG
GkOYE F6	F R	CCC AAG CTT GGTCTC A <u>GGGT</u> GTC GAC CTC GTC GAT G CCC <u>TCT AGA GGTCTC</u> A <u>AGTA</u> ATG AGG CCG ACA GCG CCG

Highlighted in green the *NheI* restriction site; highlighted in pink the *BamHI* restriction site; highlighted in red the *HindIII* restriction site; highlighted in blue the *XbaI* restriction site; highlighted in bold the inserted *Bsal* restriction site; in italic spacer amino acid between *Bsal* restriction site and four nucleotide overlap region; underlined is the four nucleotide overhang of the *Bsal* restriction digest.

Table 7.3: Primers for the overlapping extension PCR for the creation of loop variants

Loop Variants	Primer	Sequence (5' – 3')
NCR WT	F R	CCC CCC GCT AGC ATG CCT AGC TTG TTT GAT CCC ATC CCC GGA TCC TCA ATC CCC AAG CAA AGG ATA ATC
Loop A_Short	F R	<u>GCC GTT AGG TTA TCA CCG AAT GGT GAT AGT CAT CCC GAA CAG GTT TTT</u> AAA AAC CTG TTC GGG ATG ACT ATC <u>ACC ATT CGG TGA TAA CCT AAC GGC</u>
Loop A_Long	F R	<u>GAA ATA CAG GCG GCG GCG</u> GGG ACG GTT GAT AGT CAT CCC GGG ATG ACT ATC AAC CGT CCC GCG GCG GCG CTG TAT TTC
Loop A_MR	F R	TTC CTG GAA CTG TTC GGA CTG ACT GAT GAT GAA CCC GAA CAG GTT TTT ATA CCG GCG TTC ATC ATC AGT CAG TCC GAA CAG TTC CAG GAA <u>CGG TGA TAA CCT AAC GGC CGT CCG</u>
Loop A_OYE1	F R	TAC GGT GTT TTC AAC AGT ATG TCT GGT GGT GCC GAG ACC GGC ATT GTT CAG GTT TTT ATA CCG GCG GCC AAC AAT GCC GGT CTC GGC ACC ACC AGA CAT ACT GTT GAA AAC ACC GTA <u>CGG TGA TAA CCT AAC GGC CGT</u>
Loop B_Short	F R	<u>G ATG CGA GAA GGG GCT GTT GAT</u> AAA ACA GAT CAG CCC AAA TTA TCG CC GGC GAT AAT TTG GGC TGA TCT GTT TTA TCA ACA GCC CCT TCT <u>CGC ATC</u>
Loop B_MR	F R	AAC GAA CCG GAT TGG ATT GGC GGC GAT ATT ACC TAT CCG GAA GGC TTT CGC CCT GAA ATC CGA AAA GTT TTC AAA CCG CCT TTG GCG AAA GCC TTC CGG ATA GGT AAT ATC GCC GCC AAT CCA ATC CGG TTC GTT <u>CAT CCC TAA AAA GGC AAT ATC CAA ATC AGA CAA</u>
Loop B_OYE1*	F _{NCR} R _{NCR}	CAT CCC TAA AAA GGC AAT ATC CAA ATC CCT GAA ATC CGA AAA GTT TTC AAA CCG
	F _{OYE1}	<u>GAT TTG GAT ATT GCC TTT TTA GGG ATG CTG GTG GAA CCG CGC</u> GTG ACC
	R _{OYE1}	CGG TTT GAA AAC TTT TCG GAT TTC AGG GTT GCT GCC GCC TTC ATA TTC GCC

Highlighted in green the *NheI* restriction site; highlighted in pink the *BamHI* restriction site; highlighted in bold is the inserted loop sequence; in italic is the nucleotide sequence after an altered loop part; underlined is the nucleotide sequence before an altered loop part.

* For the variant Loop B_OYE1 22 amino acids are inserted in the NCR scaffold in a different manner than for the other shuffling variants. Therefore the loop region was amplified in a first PCR step from the gene of OYE1 as template. For detailed information see Reich *et al.*, ChemBioChem, 2012

Table 7.4: Primers for the Alanine Scanning Mutagenesis of the Loop A region of NCR

Alanine Scanning Mutagenesis variants	Primer	Sequence (5' – 3')
L225A	F R	CGG ACG GCC GTT AGG GCG TCA CCG AAT GGT GAA ATA CAG GG CC CTG TAT TTC ACC ATT CGG TGA CGC CCT AAC GGC CGT CCG
S226A	F R	GG ACG GCC GTT AGG TTA GCG CCG AAT GGT GAA ATA CAG G C CTG TAT TTC ACC ATT CGG CGC TAA CCT AAC GGC CGT CC
P227A	F R	CG GCC GTT AGG TTA TCA GCG AAT GGT GAA ATA CAG GGG CCC CTG TAT TTC ACC ATT CGC TGA TAA CCT AAC GGC CG
N228A	F R	GCC GTT AGG TTA TCA CCG GCG GGT GAA ATA CAG GGG ACG CGT CCC CTG TAT TTC ACC CGC CGG TGA TAA CCT AAC GGC
G229A	F R	GTT AGG TTA TCA CCG AAT GCG GAA ATA CAG GGG ACG G C CGT CCC CTG TAT TTC CGC ATT CGG TGA TAA CCT AAC
E230A	F R	GG TTA TCA CCG AAT GGT GCG ATA CAG GGG ACG GTT G C AAC CGT CCC CTG TAT CGC ACC ATT CGG TGA TAA CC
I231A	F R	G TTA TCA CCG AAT GGT GAA GCG CAG GGG ACG GTT G C AAC CGT CCC CTG CGC TTC ACC ATT CGG TGA TAA C
Q232A	F R	CCG AAT GGT GAA ATA GCG GGG ACG GTT GAT AGT C G ACT ATC AAC CGT CCC CGC TAT TTC ACC ATT CGG
G233A	F R	G AAT GGT GAA ATA CAG GCG ACG GTT GAT AGT CAT CCC GGG ATG ACT ATC AAC CGT CGC CTG TAT TTC ACC ATT C
T234A	F R	GGT GAA ATA CAG GGG GCG GTT GAT AGT CAT CCC GGG ATG ACT ATC AAC CGC CCC CTG TAT TTC ACC
V235A	F R	GAA ATA CAG GGG ACG GCG GAT AGT CAT CCC GAA CAG CTG TTC GGG ATG ACT ATC CGC CGT CCC CTG TAT TTC
D236A	F R	CAG GGG ACG GTT GCG AGT CAT CCC GAA CAG G C CTG TTC GGG ATG ACT CGC AAC CGT CCC CTG
S237A	F R	CAG GGG ACG GTT GAT GCG CAT CCC GAA CAG G C CTG TTC GGG ATG CGC ATC AAC CGT CCC CTG
H238A	F R	G ACG GTT GAT AGT GCG CCC GAA CAG GTT TTT ATA CC GG TAT AAA AAC CTG TTC GGG CGC ACT ATC AAC CGT C
P239A	F R	CG GTT GAT AGT CAT GCG GAA CAG GTT TTT ATA CCG CGG TAT AAA AAC CTG TTC CGC ATG ACT ATC AAC CG
E240A	F R	CG GTT GAT AGT CAT CCC GCG CAG GTT TTT ATA CCG G C CGG TAT AAA AAC CTG CGC GGG ATG ACT ATC AAC CG

Highlighted in bold is the codon for the amino acid exchange against alanine.

7.1.5 Strains

As a cloning and plasmid amplification strain *E. coli* DH5 α (Clontech) was used for all enzymes and variants. For the blue/white screening *E. coli* XL-1 blue (Stratagene) was used. As an enzyme expression strain *E. coli* BL21 (*DE3*) (Novagen) was used for all enzymes and variants.

7.2 Golden Gate Shuffling

7.2.1 Multiple sequence alignment

		fragment 1	
classical subfamily	NCR	-----MPSLFDPIRFGAFTAKNRIWMAPLTRGRATR-DHVPT	36
	PETNR	-----MSAEKLFPLKVGAVTAPNRVFMAPLTRLRSLIEPDIPT	39
	OYE1	MSFVKDFKFPQALGDNTLNFKPIKIGNNELLHRAVIPPPLTRMRALHPGNI PN	50
thermophilic subfamily	GkOYE	-----MNTMLFSPYTRGLTLKNRIVMSPMCMYSCDTKDGAVR	38
	YqjM	-----MARKLFTPIITIKDMTLKNRIVMSPMCMYSSEHKDGKLT	38
		fragment 2	
classical subfamily	NCR	E-IMAEYYAQRASAG--LI ISE ATGISQEGLGWPYAPGIWSDAQVEAWLP	83
	PETNR	P-LMGEYYRQRASAG--LI ISE ATQISQAQAGYAGAPGLHSPEQIAAWKK	86
	OYE1	RDWAVEYYTQRAQRPGTMI ITE GAFISPGAGGYDNAPGVWSEEQMVVEWTK	100
thermophilic subfamily	GkOYE	T-WHKIHYPARAVGQVGLI IVE ATGVTPQGRISERDLGIWSDDDHIAGLRE	87
	YqjM	P-FHMAHYISRAIGQVGLI IVE ASAVNPQGRITDQDLGIWSDDEHIEGFAK	87
		fragment 3	
classical subfamily	NCR	ITQAVHDAGGLIFAQ LWHMGRMVPSNVS--GMQPVAPSASQAPGLGHTYD	131
	PETNR	ITAGVHAEDGRIAVQ LWHTGRISHSSIQGGQAPVSASALNANTRTSLRD	136
	OYE1	IFNAIHEKKSFFVWVQ LWVLGWAAPFDNLARDGLRYDSASDNVFMDAEQEA	150
thermophilic subfamily	GkOYE	LVGLVKEHGAATIGIQL AHAGRKS-----QVPGEIIAPSAVFPD	125
	YqjM	LTEQVKEQGSKIGIQL AHAGRKA-----ELEGDIFAPSAIAFD	125
classical subfamily	NCR	G-----KKPYDVARALRLDEI PRLLDDYEKAARHALK AG FDGVQIHAANG	176
	PETNR	ENGNAIRVDTTTTPRALELDEI PGIVNDFRQAVANARE AG FDLVELHSAHG	186
	OYE1	K-----AKKANNPQHS LTKDEIKQYIKEYVQAAKNSIA AG ADGVEIHSANG	196
thermophilic subfamily	GkOYE	D-----SSPTPKEMTKADIEETVQAFQNGARRAKE AGFDVIEIHAHAG	168
	YqjM	E-----QSATPVEMSAEKVKETVQEFKQAAARAKE AGFDVIEIHAHAG	168
		fragment 4	
classical subfamily	NCR	YLIDEFIRDSTNHRHDEYGGAVENRIRLLKDVTERVIATIGKERTAVRLS	226
	PETNR	YLLHQFLSPSSNQRTDQYGGSVENRARLVLEVVDVAVCNWSADRIGIRVS	236
	OYE1	YLLNQFLDPHSNTRTDEYGGSIENRARFTLEVVDALVEAIGHEKVGRLRS	246
thermophilic subfamily	GkOYE	YLINFLSPLSNRRQDEYGGSPENRYRFLGEVIDAVREVWDG-PLFVVRIS	217
	YqjM	YLIHEFLSPLSNHRTDEYGGSPENRYRFLREIIDEVKQVWDG-PLFVVRVS	217
		fragment 5	
classical subfamily	NCR	PNGEIQGTVDSDHP--- EQVFIPAAKMLSDLD ---- I AFLGMREGAVDGT	268
	PETNR	PIGTFTQN-VDNGPN-EEADALYLI EEELAKRG ---- I AYLHMSETDLAG-	278
	OYE1	PYGVFNMSGGAE TGIVAQYAYVAGELEKRA AKAGKRLAFVHLVEPRVTNP	296
thermophilic subfamily	GkOYE	ASDYHDPDGLTAKD----- YVPYAKRMKE QG---- V DLVDVSSGAIVPA	256
	YqjM	ASDYTDKGLDIAD----- HIGFAKWMKE QG---- V DLIDCSSGALVHA	256
		fragment 6	
classical subfamily	NCR	FGKTDQPKLS---- PEIRKVFKPPLVLNQDY TF ETAQAALDSG-VADAIS	313
	PETNR	-GKPYSEAFR---- QKVRERFHGVII GAGAYTA EKAEDLIGKG-LIDAVA	322
	OYE1	FLTEGEGEYEGGSND VFVSIWKG PVIRAGNFAL HPEVVREEVKDKRTLIG	346
thermophilic subfamily	GkOYE	RMNVYPGYQVP-FAELIR READIPTGAVGLI TS GWQAEIILQNGRADLVF	305
	YqjM	DINVFPGYQVS-FAEKIREQAD MATGAVGMI TD GSMABEILQNGRADLIF	305
		fragment 7	
classical subfamily	NCR	FGRFFIGNPDLPRRFFFEKAPLTKDVIETWYTQTPKGYTDYPLLGD-----	358
	PETNR	FGRDYIANPDLVARLQKKAELNPQRPESEFYGGGAEGYTDYPSL-----	365
	OYE1	YGRFFISNPDLVDRLEKGLPLNKYDRDTFYQMSAHGYIDYPTYEEALKLG	396
thermophilic subfamily	GkOYE	LGRELLRNYPWPYAAARELGAKISAPVQYERGWRF-----	340
	YqjM	IGRELLRDPFFARTAAKQLNTEIPAPVQYERGW-----	338
classical subfamily	NCR	----	
	PETNR	----	
	OYE1	WDKK 400	
thermophilic subfamily	GkOYE	----	
	YqjM	----	

Figure 7.3: Multiple sequence alignment of the five selected ene reductases NCR, PETN, OYE1, GkOYE and YqjM for the Golden Gate Shuffling. The five ene reductases were divided into seven fragments. Colored in black are the three fragments F1, F4 and F7 which are all taken from the NCR reductase. The four fragments being shuffled are colored in yellow (F2), red (F3), pink (F5) and cyan (F6). The two amino acids building the four nucleotide overhang created by the *Bsa*I digest are highlighted in bold. Underlined are the three defined loop regions A, B and C in the NCR sequence. Multiple sequence alignments were performed with ClustalW.

7.2.2 Supplementary tables

Table 7.5: Nucleotide overhangs in-between two defined fragments of the five selected ene reductases for the Golden Gate Shuffling consisting of nucleotide composition and position.

Enzyme	Position and composition of the four nucleotide overhang of a <i>BsaI</i> digest in between two serial fragments					
	<i>F1/F2</i> GAAG	<i>F2/F3</i> CAGC	<i>F3/F4</i> GCCG	<i>F4/F5</i> ATCA	<i>F5/F6</i> GGGT	<i>F6/F7</i> TACT
NCR	bp 163-166	292-295	490-493	675-678	759-762*	885-888
PETNR	bp 172-175	301-304	520-523	705-708	792-795	912-915
OYE1	bp 214-217	343-346	550-553	735-738	828-831*	981-984*
GkOYE	bp 175-178	304-307	466-469	648-651	723-726	858-861
YqjM	bp 175-178	304-307	466-469	648-651	723-726	858-861

*In total three point mutations were inserted to introduce the six four common nucleotide motifs in between the defined seven fragments in the five ene reductases. One point mutation was inserted in NCR D254G (nucleotide change GGAT → GGGT) and two point mutations were inserted in OYE1 A277G (CGCG → GGGT) and A328T (TGCG → TACT)

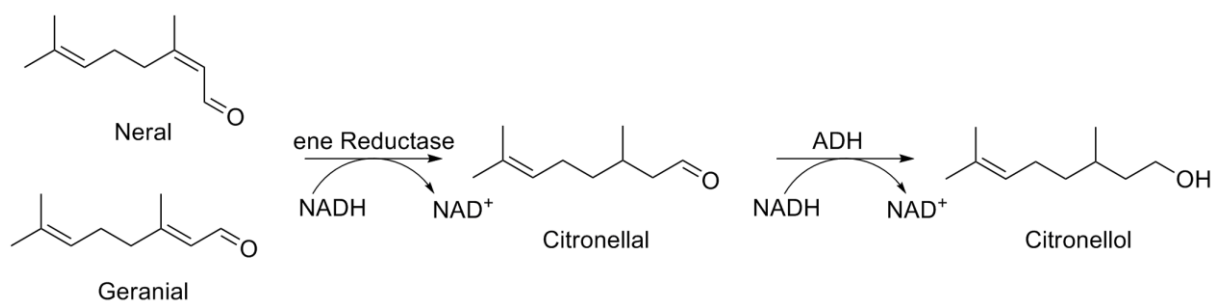
Table 7.6: Fragment composition of 22 different GGS variants determined by DNA sequencing. The first 19 generated colonies were picked randomly and sent for sequencing to verify the success of the Golden Gate Shuffling. Additionally the most active variants were sent to sequencing to determine their composition. Six of them (GSS 29, 72, 78, 649, 882 and 1034) were wild type NCR, one was not sequenced (GGS 222) and three (GGS 174, 225 and 229) possessed a different composition than the other sequenced variants.

GGS variants	Fragment 2	Fragment 3	Fragment 5	Fragment 6
1	NCR	PETNR	NCR	PETNR
2	NCR	PETNR	NCR	NCR
3	OYE1	PETNR	NCR	NCR
4	PETNR	OYE1	NCR	NCR
5	NCR	YqjM	PETNR	OYE1
6	OYE1	PETNR	NCR	OYE1
7	PETNR	YqjM	PETNR	PETNR
8	NCR	PETNR	YqjM	YqjM
9	PETNR	OYE1	YqjM	GkOYE
10	NCR	OYE1	NCR	NCR
11	NCR	YqjM	NCR	GkOYE
12	NCR	YqjM	YqjM	YqjM
13	NCR	PETNR	PETNR	GkOYE
14	PETNR	PETNR	PETNR	NCR
15	PETNR	PETNR	YqjM	GkOYE
16	PETNR	YqjM	NCR	PETNR
17	OYE1	NCR	NCR	GkOYE
18	YqjM	NCR	NCR	NCR
19	PETNR	NCR	PETNR	YqjM
174	PETNR	PETNR	PETNR	PETNR
225	PETNR	NCR	NCR	YqjM
229	PETNR	PETNR	YqjM	YqjM

Table 7.7: Product formation in the reduction of neral and geranial with the Golden Gate Shuffling variants

Products/ Enzymes	Production formation in [%] for the reduction of the two isomers of citral			
	neral reduction		geranial reduction	
	citronellal	citronellol	citronellal	citronellol
<i>E. coli</i> BL21 (<i>DE3</i>) lysate	0.3 ± 0.01	4.3 ± 0.01	0.6 ± 0.03	4.4 ± 0.01
NCR wild type	37.7 ± 1.6	-	43.9 ± 1.1	-
GGs 10	30.3 ± 0.8	0.9 ± 0.01	30.6 ± 6.3	2.8 ± 2.4
GGs 174	18.8 ± 2.6	10.8 ± 1.2	22.8 ± 3.5	11.6 ± 2.2
GGs 222	23.0 ± 5.7	6.1 ± 1.1	13.7 ± 5.2	3.1 ± 1.5
GGs 225	48.1 ± 0.1	0.2 ± 0.01	43.8 ± 0.2	0.6 ± 0.2
GGs 229	42.6 ± 1.2	1.5 ± 0.5	52.4 ± 4.0	1.8 ± 0.6

7.2.3 Supplementary figures

**Figure 7.4:** Reaction mechanism of the reduction of neral and geranial with an ene reductase by using NADH. The reduction product is further reduced by an alcohol dehydrogenase to the corresponding alcohol citronellol.

7.3 Rational loop modulation

7.3.1 Crystallization, data collection and structural determination

The crystallization of the shuffling variant Loop A_OYE1 was carried out by a cooperation partner at the University of Leipzig, Germany, with the following protocol and conditions: Initial crystallization conditions of the variant Loop A_OYE1 were determined by using the sitting-drop vapor diffusion technique in sparse-matrix screens based on commercially available crystallization screens from Hampton Research and Jena BioScience. A 1:1 mixture of 0.2 μ L protein solution and 0.2 μ L reservoir buffer was mixed and equilibrated against 90 μ L reservoir solution. Finally Loop A_OYE1 variant was crystallized using 0.1 M sodium citrate pH 5.2, 1.0 M lithium chloride and 21 % PEG 6 000 as reservoir solution. 1 μ L of a 1:1 mixture of protein and reservoir solutions was equilibrated by the hanging drop vapor diffusion method against the reservoir buffer. Crystals were obtained after a few weeks at 19°C. X-ray data were recorded at 100 K. Protein crystals were directly flash-frozen in liquid nitrogen without adding a cryoprotectant. The data set of the Loop A_OYE1 variant was collected at the BESSY beamline 14.1 and integrated using XDS¹⁴⁵. Details of data collection and refinement parameters are listed in table 7.9. The structure of the variant Loop A_OYE1 was solved by molecular replacement using the NCR wild type structure (PDB identifier 4A3U) without water molecules and ligands as the search model⁹. One molecule is present in the asymmetric unit. An initial model was built automatically by using ARP/wARP of the CCP4 package^{146,147}. The model was improved manually and the FMN cofactor was inserted by using coot. Refinement was carried out in refmac5 (CCP4 suite) until no further improvement was achieved. Based on the difference electron density ($> 3 \sigma$ ms), distance criteria ($2.5 \text{ \AA} < d < 3.6 \text{ \AA}$) as well as manual inspection of the environment and density water molecules were included. The final model of Loop A_OYE1 variant was evaluated with the program MolProbity¹⁴⁸. PyMOL (www.pymol.org) was used for visualisation of the protein structures. The atomic coordinates are available in the Protein Data Bank under the accession code specified in table 7.8.

Table 7.8: Data collection and refinement statistics. Highest resolution shell is shown in parenthesis.

Data collection	Loop grafting variant Loop A_OYE1
Space group	P3 ₁ 21
X-ray source	BL 14.1
Wavelength (Å)	0.918409
Unit cell a, b, c (Å)	121.24, 121.24, 58.65
α, β, γ (°)	90, 90, 120
Matthews coefficient (Å ³ /Da)	1.59
Solvent content (%)	22.59
Monomers per AU	1
Observed reflections	492512
Unique reflections	46091
R _{free} reflections (%)	1.0
Completeness (%)	99.9 (99.4)
Multiplicity	10.7 (10.2)
Mosaicity (°)	1.1
I/σ	21.52 (3.74)
Wilson B-factor (Å ²)	20.9
R _{sym} (%)	7.1 (66.7)
R _{meas} (%)	7.5 (70.3)
Refinement	
Resolution range (Å)	39.69 - 1.80 (1.91 - 1.80)
R _{work} (%)	16.53
R _{free} (%)	19.64
R.m.s derivations	
Bond lengths (Å)	0.0209
Bond angles (°)	2.0721
Ramachandran plot (%)	
Favoured	93.56
Allowed	5.76
Outliers	0.68

7.3.2 Supplementary tables

Table 7.9: GC-FID/GC-MS programs and columns for the analysis of biotransformation reactions

Compounds	GC programs	Column
Cyclic standard substrates	70°C, 15°C/min to 200°C, 30°C/min to 320 °C, hold 1 min; injector temperature 250°C, FID temperature 325°C	Agilent HP5 column, length 30 m, inner diameter 0,25 mm
Aliphatic standard substrates	3 min at 60°C, 10°C/min to 150°C, 50°C/min to 300°C, hold 1 min; injector temperature 250°C, FID temperature 325°C	Agilent HP5 column, length 30 m, inner diameter 0,25 mm
γ -Butyrolactones as substrates	80°C hold 2 min, 5°C/min to 95°C, hold 5 min, 2°C/min to 100°C, hold 5 min, 50°C/min to 300°C, hold 1 min; injector temperature 250°C, FID temperature 325°C	Agilent HP5 column, length 30 m, inner diameter 0,25 mm
Enantioselectivity cyclic substrates	100°C hold 1 min, 10°C/min to 200°C, hold 2 min; injector temperature 200°C, FID temperature 250°C	Machery-Nagel Hydrodex β -TBDAC column, length 25 m, inner diameter 0,25 mm
Enantioselectivity aliphatic substrates	80°C hold 1 min, 10°C/min to 100°C, 2°C/min to 120°C, 30°C/min to 200°C, hold 2 min; injector temperature 200°C, FID temperature 250°C	Machery-Nagel Hydrodex β -TBDAC column, length 25 m, inner diameter 0,25 mm
MS program cyclic substrates	70°C, 15°C/min to 210°C, hold 2 min, 30°C/min to 300°C hold 1 min; injector temperature 250°C, ion source temperature: 200°C, interface temperature 270°C, mass detection in scan mode 15 - 220 <i>m/z</i> end	Agilent DB-5MS column, length 30 m, inner diameter 0,25 mm

Table 7.10: Protein concentration of the three expressed wild type enzymes as well as the seven rational loop shuffling variants after the purification by immobilized metal affinity chromatography and the concentration via centrifugation.

Wild type enzymes	NCR	OYE1	MR
Protein concentration [mg/ml]	5.6	21.3	25.9

Rational loop variants	Loop A	Loop A	Loop A	Loop A	Loop B	Loop B	Loop B
	Short	Long	OYE1	MR	Short	OYE1	MR
Protein concentration [mg/ml]	7.6	3.4	5.6	7.0	5.5	3.6	6.8

Table 7.11: Enantiomeric distribution of the conversion of ketoisophorone and 2-methyl-2-pentenal with the wild type enzymes and seven created rational loop variants

Substrate		Enantiomeric distribution [%]					
		Ketoisophorone			2-Methyl-2-pentenal		
Products		R-Levodione	S-Levodione	ee-value	R-Methylpentanal	S-Methylpentanal	ee-value
Wild type enzymes	<i>NCR</i>	98.3	1.8	97	34.4	65.6	31
	<i>OYE1</i>	97.4	2.7	95	8.8	91.2	82
	<i>MR</i>	59.7	40.3	19	25.3	74.7	49
Loop A variants	<i>Loop A_Short</i>	98.3	1.8	97	35.7	64.4	29
	<i>Loop A_Long</i>	98.2	1.8	96	34.6	65.4	31
	<i>Loop A_OYE1</i>	98.3	1.7	97	16.5	83.5	67
	<i>Loop A_MR</i>	98.1	1.9	96	38.3	61.7	23
Loop B variants	<i>Loop B_Short</i>	97.9	2.1	96	51.3	48.7	-3
	<i>Loop B_OYE1</i>	98.1	1.9	96	0	100	100
	<i>Loop B_MR</i>	56.5	43.5	13	47.2	52.8	6

Table 7.12: Unfolding temperature of the three expressed wild type enzymes, as well as the seven rational loop shuffling variants determined with the ThermoFAD method.

Wild type enzymes	NCR	OYE1	MR
Unfolding temperature [°C]	60.5	54.5	70.8

Rational loop variants	Loop A	Loop A	Loop A	Loop A	Loop B	Loop B	Loop B
	Short	Long	OYE1	MR	Short	OYE1	MR
Unfolding temperature [°C]	61.3	60.9	63.4	65.9	61.8	50.5	66.5

Table 7.13: Product formation in the reduction of ketoisophorone at elevated reaction temperatures. Reaction conditions: 50 µg purified protein, 2 mM ketoisophorone, 2 mM NADH

<i>Reaction temperature/ Enzyme</i>	Product formation in the reduction of ketoisophorone in [%]							
	30°C	37°C	40°C	45°C	50°C	55°C	60°C	
Wild type enzymes	<i>NCR</i>	97.8±2.0	97.6±0.8	98.0±0.2	53.0±5.9	31.8±2.3	20.5±2.2	12.5±2.8
	<i>OYE1</i>	30.7±1.6	27.5±1.0	13.2±0.3	4.9±0.1	2.3±0.1	1.5±0.1	1.2±0.1
	<i>MR</i>	26.1±2.3	40.1±1.3	32.0±0.1	24.7±0.8	18.3±0.1	5.4±0.5	3.1±0.1
Loop A variants	<i>Loop A_Short</i>	95.2±2.0	63.1±5.6	54.3±5.6	15.2±0.1	0.6±0.1	0.5±0.1	0.5±0.1
	<i>Loop A_Long</i>	83.6±10.5	57.5±3.1	39.0±2.3	16.8±3.6	1.0±0.1	0.7±0.1	0.6±0.1
	<i>Loop A_OYE1</i>	64.0±4.0	68.3±4.5	58.1±1.1	37.9±1.4	1.9±0.1	1.4±0.2	1.1±0.1
	<i>Loop A_MR</i>	32.6±3.1	27.6±1.6	23.0±3.5	5.5±0.4	3.0±0.4	1.4±0.1	1.3±0.1
Loop B variants	<i>Loop B_Short</i>	38.5±0.8	77.6±3.0	80.5±0.6	97.7±2.2	39.5±2.3	22.4±1.4	18.6±1.9
	<i>Loop B_OYE1</i>	33.3±1.0	74.7±0.2	11.4±4.5	12.3±3.9	2.9±0.9	4.4±1.4	2.5±0.1
	<i>Loop B_MR</i>	24.6±1.0	41.5±0.5	32.0±4.6	30.3±3.7	23.9±0.4	5.5±0.1	2.7±0.1

Reactions were performed in triplicates in a final volume of 1 ml 50 mM Tris-HCl pH 7.5 and run at 30°C with 180 rpm for 24 h.

Table 7.14: Product yield in the conversion of cinnamaldehyde at different cofactor ratios. For all performed reactions an excess of NADH over the cinnamaldehyde concentration was used to ensure a possible complete conversion of the substrate. The remained substrate is highlighted in *italic*, the obtained reduction product hydrocinnamaldehyde in **bold**.

Enzyme variants	Product distribution	Product yield in [%]			
		Cofactor ratio NADH / NAD ⁺			
		1 / 0	4 / 1	1 / 1	
Wild type enzymes	NCR	<i>Cinnamaldehyd</i>	1.4±0.5	61.9±2.6	81.0±3.5
		Hydrocinnamaldehyde	83.2±2.7	36.5±2.5	18.3±3.2
		3-Phenyl-1-propanol	15.2±3.3	1.4±0.2	0.7±0.4
	OYE1	<i>Cinnamaldehyd</i>	31.2±1.1	57.4±3.4	74.1±0.5
		Hydrocinnamaldehyde	51.7±1.4	33.1±2.9	19.9±0.3
		3-Phenyl-1-propanol	7.7±0.2	4.8±0.9	2.0±0.3
	MR	<i>Cinnamaldehyd</i>	65.1±1.4	78.4±1.2	81.6±0.3
		Hydrocinnamaldehyde	34.8±1.3	21.4±1.3	17.9±0.3
		3-Phenyl-1-propanol	0.1±0.1	0.2±0.1	0.4±0.1
Loop A variants	Loop A_Short	<i>Cinnamaldehyd</i>	0.6±0.4	0.6±0.6	67.5±4.1
		Hydrocinnamaldehyde	80.7±1.5	52.3±10.9	30.9±3.5
		3-Phenyl-1-propanol	14.5±1.4	47.1±11.5	1.5±0.6
	Loop A_Long	<i>Cinnamaldehyd</i>	0.2±0.1	4.3±4.1	58.1±3.9
		Hydrocinnamaldehyde	75.8±2.4	83.3±0.9	40.0±30.2
		3-Phenyl-1-propanol	21.4±1.6	12.0±5.0	1.8±0.7
	Loop A_OYE1	<i>Cinnamaldehyd</i>	0.6±0.4	35.1±2.4	55.5±5.8
		Hydrocinnamaldehyde	77.5±4.6	55.6±0.4	38.8±4.6
		3-Phenyl-1-propanol	15.1±4.7	4.3±0.2	2.7±1.0
	Loop A_MR	<i>Cinnamaldehyd</i>	6.4±6.3	8.9±7.8	60.0±4.7
		Hydrocinnamaldehyde	71.0±5.9	73.7±0.1	36.0±3.4
		3-Phenyl-1-propanol	18.0±11.7	15.5±8.6	2.8±1.1
Loop B variants	Loop B_Short	<i>Cinnamaldehyd</i>	1.8±1.8	21.8±6.8	72.6±3.7
		Hydrocinnamaldehyde	67.8±4.4	67.3±4.7	26.1±3.8
		3-Phenyl-1-propanol	30.4±0.9	10.1±2.0	1.1±0.1
	Loop B_OYE1	<i>Cinnamaldehyd</i>	43.4±1.0	56.3±4.1	67.7±1.2
		Hydrocinnamaldehyde	56.4±0.4	42.9±3.9	32.1±1.2
		3-Phenyl-1-propanol	0.2±0.1	0.8±0.3	0.2±0.1
	Loop B_MR	<i>Cinnamaldehyd</i>	60.9±3.2	61.7±3.1	87.1±2.1
		Hydrocinnamaldehyde	37.0±3.7	36.5±3.0	12.8±2.0
		3-Phenyl-1-propanol	2.1±0.5	1.8±0.1	0.1±0.1

Reactions were performed in triplicates in a final volume of 1 ml 50 mM Tris-HCl pH 7.5 and run at 30°C with 180 rpm for 24 h.

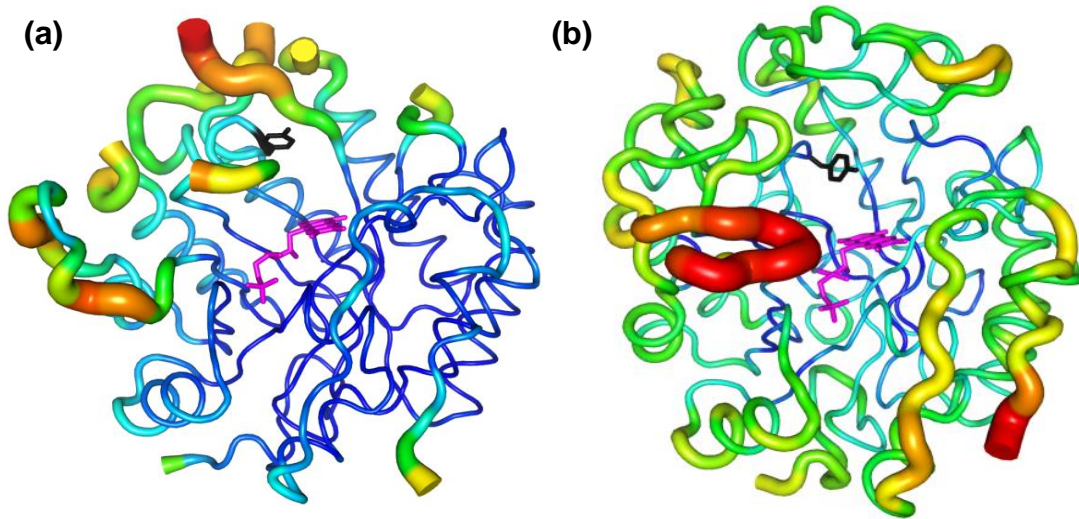


Figure 7.6: B-Factor representation of (a) the rational loop grafting variant Loop A_OYE1 and (b) NCR wild type. The higher the B-factor the thicker the structural scaffold. The areas with the highest are colored red, the one with the lowest blue.

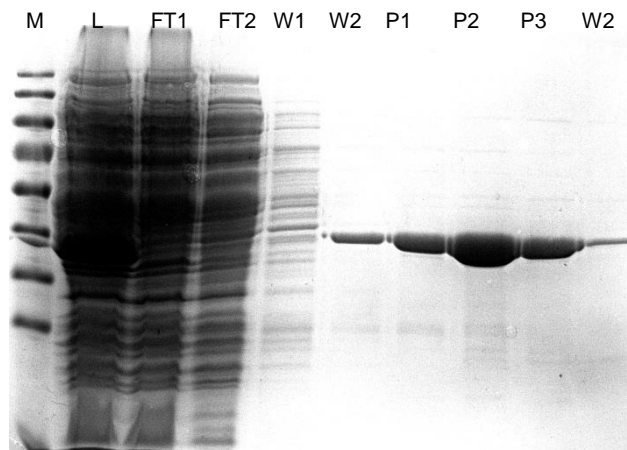


Figure 7.7: SDS PAGE for the his-tag purification of the rational variant Loop A_OYE1. M=unstained protein ladder (Fermentas), L=lysate, FT= flow through, W=waste, P=protein fraction. The three protein fractions were merged and concentrated via ultrafiltration with viaspin columns. Protein purity is > 90 %.

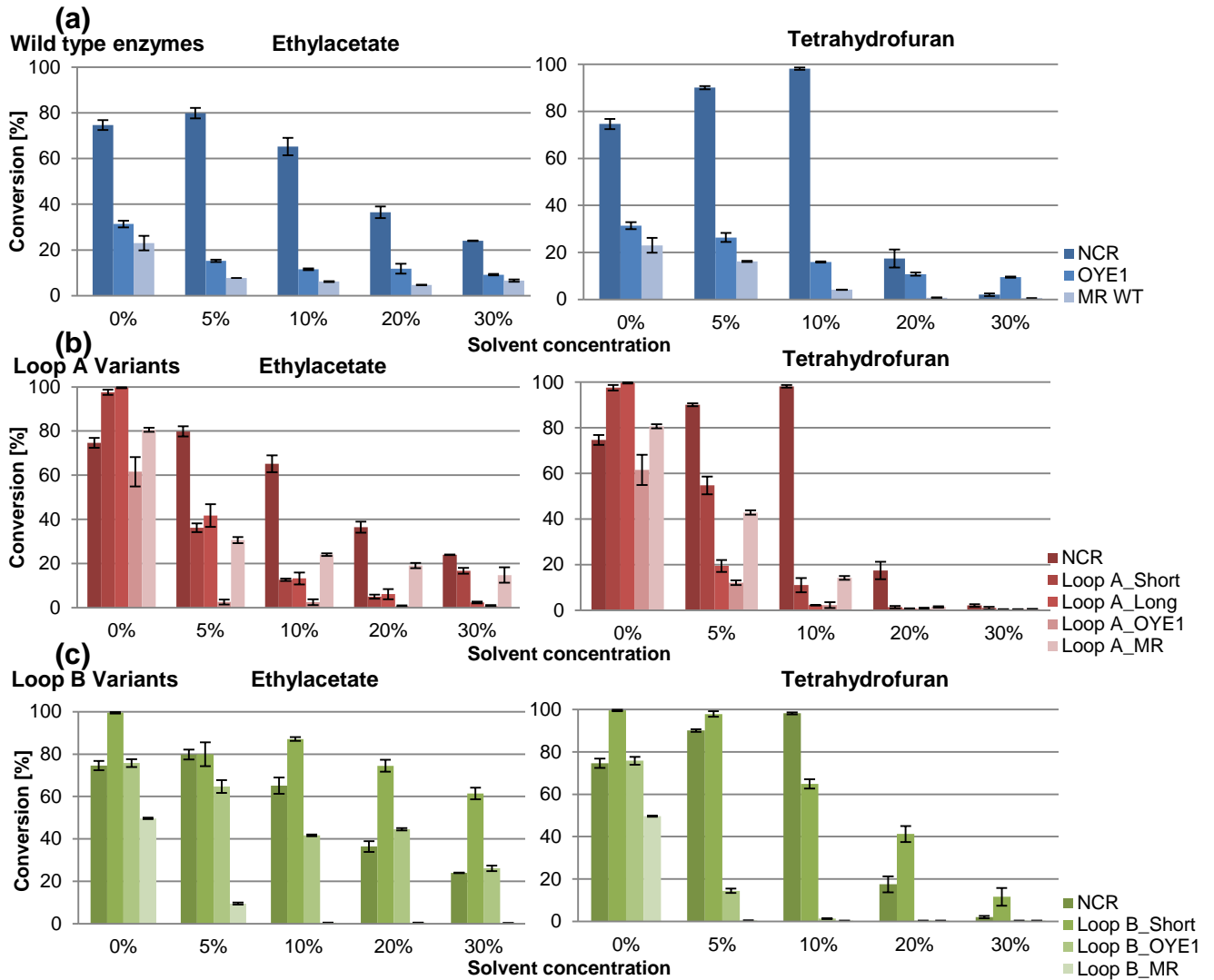


Figure 7.8: Conversion of ketosiphorone with different solvent concentrations ranging from 0 – 30 % v/v organic solvent. Shown are the results with two solvents ethylacetate (0.75) and tetrahydrofuran (0.53) (a) The three wild type enzymes NCR, MR and OYE1 colored in blue (b) The four created Loop A variants in comparison with NCR colored in red (c) The three Loop B variants compared to NCR colored in green. The reactions were run at 30°C with 180 rpm for 4h in 1 ml 50 mM Tris-HCl pH 7.5. LogP values are indicated in brackets.

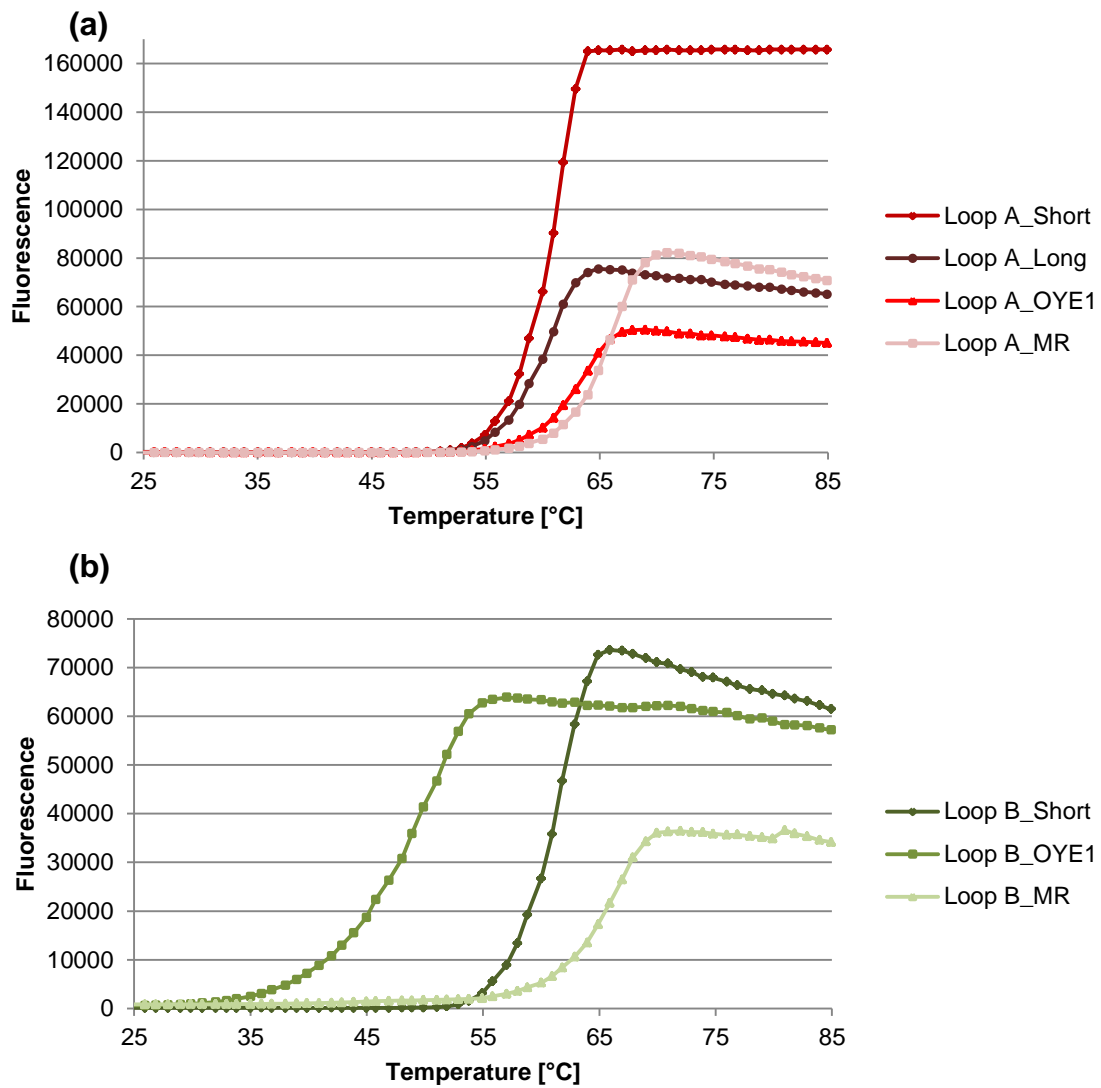


Figure 7.9: Visualization of the FMN fluorescence emission based on the ThermoFAD method in dependence of the temperature. The fluorescence signal is plotted against the temperature. The inflection point of each curve determines the melting point of the corresponding enzyme variant (a) Fluorescence curves of the four Loop A variants colored in shades of red (b) Fluorescence curves of the Loop B variants colored in shades of green.

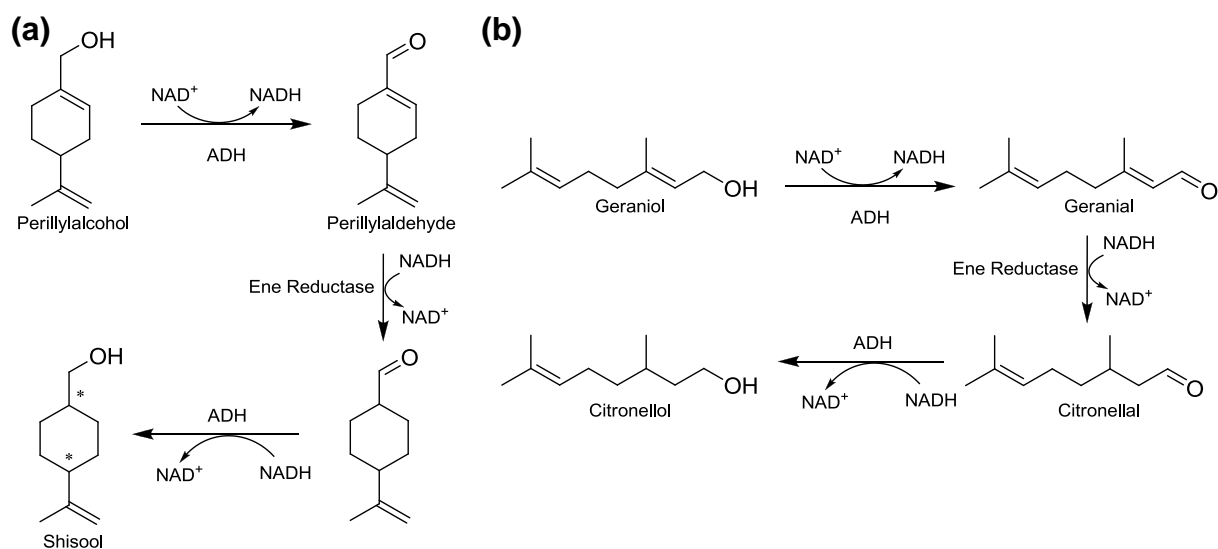


Figure 7.10: Proposed reaction mechanism of the reduction of allyl alcohols with a two enzyme approach based on an ene reductase and an alcohol dehydrogenase (ADH) **(a)** Reaction with perillyl alcohol as substrate containing the formed intermediates and the final reduction product shisool. **(b)** Reaction with geraniol as substrate containing the formed intermediates and the final reduction product citronellol

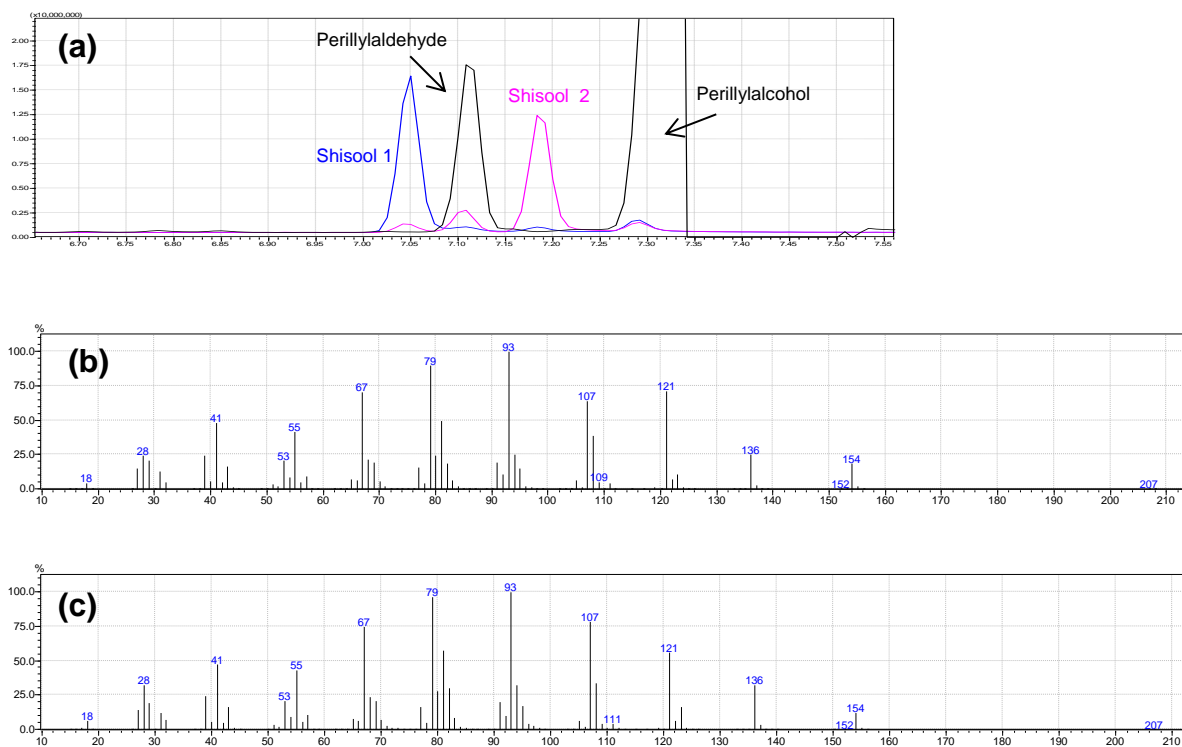


Figure 7.11: (a) Peak distribution of the reduction of perillyl alcohol. Diastereomer 1 of the reduced product shisool is colored in blue with a retention time of 7.0 min and diastereomer 2 is colored in pink with a retention time of 7.2 min. (b) GC/MS fragment pattern of the shisool 1 product peak with a retention time of 7.0 min (c) GC/MS fragment pattern of the shisool 2 product peak with a retention time of 7.2 min. Both peaks exhibit the same fragmentation pattern, indicating that the both products are diastereomers.

Curriculum vitae

Persönliche Daten

Name: Sabrina Reich
Geburtstag: 11. August 1984 in Calw
Nationalität: deutsch
Adresse: Bergstraße 37,
75382 Althengstett
Email: sab.reich@gmail.com



Hochschulbildung

11/2010 – 10/2013 Promotion bei Prof. Dr. Bernhard Hauer am Institut für Technische Biochemie der Universität Stuttgart

Titel: „Studien über variable Oberflächenloops der En-Reduktase NCR aus Zymomonas mobilis“

seit 10/2011 Stipendium der Landesgraduiertenförderung Baden-Württemberg

09/2010 Diplom Biologie technische orientiert, Universität Stuttgart
(Note: Mit Auszeichnung bestanden)

01/2010 – 09/2010 Anfertigung der Diplomarbeit am Institut für Technische Biochemie der Universität Stuttgart

Titel: „Ortsgerichtete Mutagenese und Loopaustausch bei NCR einer Enoat Reduktase aus Zymomonas mobilis“
Note: 1,0

09/2004 – 09/2010 Studium der Technischen Biologie an der Universität Stuttgart

01/2009 – 06/2009 Auslandssemester zur Anfertigung der Studienarbeit im Institut für Molekulare Systembiologie, ETH Zürich, Schweiz.

Titel: „Entwicklung eines FRET Sensors für die in vivo Echtzeitmessung der Citratkonzentration“
Note: 1,0

Schulbildung

07/2004 Allgemeine Hochschulreife, Maria-von-Linden Gymnasium, Calw-Stammheim, Note: 1,6

09/1995 – 07/2004 Maria-von-Linden Gymnasium Calw-Stammheim