Yeasts as Production Hosts for Biocatalysts

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Kamila Morka (geb. Rzeszutko)
aus Grodków, Polen

Düsseldorf, 2015

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Jörg Pietruszka
Korreferent: Prof. Dr. Karl-Erich Jaeger

Tag der mündlichen Prüfung: 04.05.2015

Parts of this thesis have already been published or presented before:

Publications:

K. Morka, J. Pietruszka, S. Meyer zu Berstenhorst, „Comparative expression of lipase CAL-A in the yeasts *Saccharomyces cerevisiae*, *Kluyveromyces lactis* and *Hansenula polymorpha* to investigate a possible host influence", *J. Biotech*., **2014,** *191, 176–186*.

Poster presentations:

2nd CLIB-GC Retreat 07.-08.04.2011, Bonn; K. Morka, S. Meyer zu Berstenhorst, J. Pietruszka *„Comparative expression analysis in yeast systems*".

3rd CLIB-GC Retreat 22-24.02.2012, Bergisch Gladbach; K. Morka, S. Meyer zu Berstenhorst, J. Pietruszka *„Comparative expression analysis in yeast systems: CAL-A as the first challenge*" (poster and talk).

Biocat 2012, 02-06.09.2012, Hamburg; K. Morka, S. Meyer zu Berstenhorst, J. Pietruszka "*Does the host matter? Comparative expression study of lipase CAL-A in yeast systems*".

4th CLIB-GC Retreat 13-15.02.2013, Lünen; K. Morka, S. Meyer zu Berstenhorst, J. Pietruszka *„Does the host matter? Comparative expression study of lipase CAL-A in yeast systems*".

VAAM 2013, 10-13.03.2013, Bremen; K. Morka, S. Meyer zu Berstenhorst, J. Pietruszka *„Does the host matter? Comparative expression study of lipase CAL-A in yeast systems*".

5th CLIB-GC Retreat 12-15.02.2014, Lünen; K. Morka, S. Meyer zu Berstenhorst, J. Pietruszka *„Does the host matter? Comparative expression study of lipase CAL-A in yeast systems*".

To my mother and father, Katarzyna & Waldemar,
who have blessed me with an admiration of nature.
And to my husband, Mariusz,
whose love and confidence is a constant source of inspiration and encouragement.

I am among those who think that science has great beauty. A scientist in his laboratory is not only a technician: he is also a child placed before natural phenomena which impress him like a fairy tale.

Maria Skłodowska-Curie

I. Abbreviations used in this work

Abbreviation	Full name
BSA	bovine serum albumin
Å	angstrom
c	concentration
CAL-A	lipase A from *Candida antarctica*
dH_2O	distilled water
DMSO	dimethyl sulfoxide
dNTP	deoxynucleotide triphosphates
DSMZ	German Collection of Microorganisms and Cell Cultures
DTT	ditiotreitol
DWP	deep well plate
e.e.	enantiomeric excess
EDTA	ethylenediaminetetraacetic acid
Euroscarf	European *Saccharomyces cerevisiae* Archive for Functional Analysis
g	gram or gravity
GH	glycoside hydrolase
GlcNAc	N-Acetylglucosamine
GS	glycosynthase
GT	glycosyl transferase
Hp	*Hansenula polymorpha* (recently reclassified as *Ogataea angusta*)
Hp CAL-A T	lipase A from *Candida antarctica* secreted from Hp in presence of tunicamycin
IMAC	immobilised metal ion affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
k_{cat}	catalytic constant
kD	kilodaltons
Kl	*Kluyveromyces lactis*
K_M	Michaelis constant
KP_i buffer	potassium phosphate buffer
l	length
L	litre
MOPS	3-(N-morpholino)propanesulfonic acid
MTP	microtiter plate
NC	nitrocellulose
NEB	New England Biolabs
ng	nanogram
Ni-NTA	nickel-nitrilotriacetic acid
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol

pM	picomolar
*p*NP	*p*-nitrophenol
Rha	rhamnosidase
Rha_{Ba}	rhamnosidase from *Bacillus* sp.
Rha_{Hp}	rhamnosidase from *Hansenula polymorpha*
Rha_{La}	rhamnosidase from *Lactobacillus acidophilus*
RT	room temperature
SAD	single-wavelength anomalous dispersion
Sc	*Saccharomyces cerevisiae*
SDS	sodium dodecyl sulphate
TEMED	tetramethylethylenediamine
TLC	thin layer chromatography
T_M	melting temperature
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
U	unit
V	Volt
V (L)	volume
v/v	volume per volume
vs.	versus
w/v	weight per volume
WT	wild type
YNB	yeast nitrogen base

Table 1: List of standard amino acid abbreviations.

Amino acid	3-letter code	1-letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Table 2: Standard genetic code. The stop codons are marked with asterisks.

Nucleotide position in codon					
first	**second**				**third**
	U	**C**	**A**	**G**	
U	UUU - Phe	UCU - Ser	UAU - Tyr	UGU - Cys	**U**
	UUC - Phe	UCC - Ser	UAC - Tyr	UGC - Cys	**C**
	UUA - Leu	UCA - Ser	UAA - *	UGA - *	**A**
	UUG - Leu	UCG - Ser	UAG - *	UGG - Trp	**G**
C	CUU - Leu	CCU - Pro	CAU - His	CGU - Arg	**U**
	CUC - Leu	CCC - Pro	CAC - His	CGC - Arg	**C**
	CUA - Leu	CCA - Pro	CAA - Gln	CGA - Arg	**A**
	CUG - Leu	CCG - Pro	CAG - Gln	CGG - Arg	**G**
A	AUU - Ile	ACU - Thr	AAU - Asn	AGU - Ser	**U**
	AUC - Ile	ACC - Thr	AAC - Asn	AGC - Ser	**C**
	AUA - Ile	ACA - Thr	AAA - Lys	AGA - Arg	**A**
	AUG - Met	ACG - Thr	AAG - Lys	AGG - Arg	**G**
G	GUU - Val	GCU - Ala	GAU - Asp	GGU - Gly	**U**
	GUC - Val	GCC - Ala	GAC - Asp	GGC - Gly	**C**
	GUA - Val	GCA - Ala	GAA - Glu	GGA - Gly	**A**
	GUG - Val	GCG - Ala	GAG - Glu	GGG - Gly	**G**

1. Abstracts

1.1 Zusammenfassung

Enzyme als natürliche Katalysatoren finden in der organischen Chemie mehr und mehr Anwendungen vor allem aufgrund ihrer Selektivität. Darüber hinaus wurde es durch Fortschritte in der Proteintechnologie möglich, wesentliche Enzymeigenschaften zu beeinflussen. Effiziente Expression des Enzyms in einem heterologen Wirt ist eine Grundvoraussetzung zur erfolgreichen Anwendung in der Biokatalyse. In dieser Arbeit wurden die eukaryotischen Hefeexpressionssysteme *Saccharomyces cerevisiae*, *Kluyveromyces lactis* und *Hansenula polymorpha* verglichen und nicht nur die Produktivität, sondern auch die Eigenschaften des hergestellten Enzyms analysiert, um den möglichen Einfluss der Wirts-spezifischen Änderungen zu untersuchen. Im zweiten Teil dieser Arbeit wurde die Eignung der Hefen zur effizienten Produktion einer weniger bekannten Familie der Glycosylhydrolasen, Rhamnosidasen, untersucht, um die potentielle Anwendbarkeit dieser Enzyme für die Biokatalyse zu erhöhen.

Als erstes Modellenzym wurde die bekannte Lipase A aus *Candida antarctica* (CAL-A) ausgewählt und erstmalig die Codonverwendung desselben Gens zur parallen Expression in drei Organismen optimiert. Alle Hefen waren in der Lage aktive CAL-A zu produzieren, jedoch mit signifikanten Unterschieden in der Ausbeute: *H. polymorpha* zeigte eine zehnfache Steigerung der Produktivität im Vergleich zu *S. cerevisiae*. Auch die Untersuchungen zur thermischen Stabilität und Aktivität der gereinigten Enzyme gegenüber verschiedenen Substraten ergaben eine signifikante Wirkung des Wirts auf die biochemischen Eigenschaften des hergestellten Proteins. So behielt die CAL-A aus *K. lactis* 70% ihrer Aktivität nach Inkubation bei 60 °C im Vergleich zu 45% Restaktivität aus *S. cerevisiae* und bei Tests mit verschiedenen Substraten wurde eine Vervierfachung der Aktivität zwischen den Enzymen aus H. polymorpha und S. cerevisiae gefunden. Insgesamt zeigen die Ergebnisse beispielhaft, dass die Auswahl des Expressionswirts selbst innerhalb einer taxonomischen Familie (*Saccharomycetaceae*) die Eigenschaften des produzierten Enzyms erheblich beeinflusst und zukünftig mehr Beachtung verdient.

Als weniger bekannte Enzymfamilie wurden die Rhamnosidasen der GH Familie 78 ausgewählt, um ihre mögliche Anwendung in der chemischen Synthese zu untersuchen. Dazu wurden drei veröffentlichte Rhamnosidasen verwendet: Rha_{Ba} aus *Bacillus* sp., Rha_{La} aus *Lactobacillus acidophilus* und RhaHp aus *Hansenula polymorpha*. Auch mit den drei eukaryotischen Testwirten konnte keine der getesteten Rhamnosidasen in aktiver Form sekretiert werden. Die Rhamnosidase Rha_{Ba} wurde allerdings erfolgreich in *E. coli* hergestellt und daher für Mutageneseversuche zu einer Glycosynthase angewendet. Die erzeugten Varianten wurden mit verschiedenen Donoren und Akzeptoren getestet und obwohl das gewünschte Produkt nicht entdeckt werden konnte, können einige wertvolle Beobachtungen aus diesen Versuchen entnommen werden. Als Vorteile zur industriellen Anwendung zeigten sich neben der guten Herstellung mit *E. coli* auch eine lange Lagerbarkeit des Lyophilisates sowie die geringe Wirkung von bis zu 20% DMSO auf die Enzymaktivität. Das Enzym scheint selektiv für Rhamnose als Glycosylgruppe zu sein, da das Standardsubstrat *p*NP-α-Rhamnopyranosid mit hoher Spezifität hydrolysiert und *p*NP-β-D-Glucopyranosid nicht als Substrat akzeptiert wurde. Es ist möglich, dass Rha_{Ba} nicht in eine Glycosynthase modifiziert werden kann, wie bereits für andere Enzyme berichtet wurde, allerdings wären zunächst weitere Versuche mit stabileren Glycosylgruppe-Donoren, z.B. Aziden, zu empfehlen. Bei der Rhamnosidase Rha_{Hp} wurde das bisher unbekannte Gen erfolgreich identifiziert und aktives intrazelluläres Protein homolog mit H. polymorpha produziert. Das Kristallisationsscreening führte zur Bildung von mehreren kleinen plattenförmigen Kristallen, deren anisotrope Diffraktion leider nicht ausreichte zur Klärung einer weiteren der nahezu noch unveröffentlichten Rhamnosidasestrukturen aber dennoch Hoffnung macht.

1.2 Abstract

Enzymes, as natural catalysts, are becoming more widely applied in organic chemistry due to their selectivity and availability. Moreover, with advances in protein engineering it has become possible to influence crucial enzyme properties, such as thermostability or selectivity. High yield expression of a desired enzyme in a heterologous host is a prerequisite for its successful application in biocatalysis. With diverse expression systems available, it is important to select the most suitable one. In this work, yeasts as model eukaryotic expression systems were compared based on productivity and the properties of the produced enzyme were analysed to investigate a potential influence of host-specific modifications.

As a model enzyme, the well-studied lipase A from *Candida antarctica* (CAL-A) was selected. The codon usage of the gene was optimised for expression in three industrially relevant hosts: *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, and *Hansenula polymorpha*. All the hosts produced the active enzyme; however, significant differences in the obtained yield were observed. *H. polymorpha* emerged as the most beneficial host with a ten-fold higher amount of the expressed protein per culture volume in comparison to *S. cerevisiae*. Analysis of both thermostability and activity of the purified enzyme towards various substrates showed a significant impact of the host on these biochemical properties of the produced lipase isoforms. The most thermostable CAL-A from *K. lactis* retained 70% of its activity after incubation at 60 °C, in comparison to 45% remaining activity for the enzyme purified from *S. cerevisiae*. A difference in activity, as obtained for the standard substrate *p*NP octanoate (**1**) was also observed when bulky substrates were applied for the screening.

Especially noteworthy disparities were obtained for indan-2-carboxylate. Its *p*NP ester (*S*)-**3** was hydrolysed four times more efficiently by the *H. polymorpha* variant than both other enzyme variants. Interestingly its enantiomer (*R*)-**3** was not accepted as a substrate by CAL-A produced in *S. cerevisiae*, while both other preparations showed similar activity for the substrate (Table 3).

To investigate the impact of glycosylation on the enzyme properties, expression with glycosylation inhibitor tunicamycin was performed and two variants of the lipase, glycosylated and under-glycosylated, obtained by secretion from *H. polymorpha* were compared.

Table 3: Selected results obtained in this study for the lipase CAL-A produced in three yeast hosts. The data for the optical density and cell number refer to the stationary phase after 72 h of growth (30 °C, 120 rpm) when the cultures (100 mL in 500 mL shake flasks) were harvested. Activity per volume was measured in crude extracts, whereas purified enzymes were used for other measurements. Substrates used for the specific activity measurements: **1** - *p*NP octanoate, *(R)*-**3** - *p*NP indan-2-carboxylate. The presented data are an average from three independent measurements under standard conditions (pH=8.0, 30 °C).

Sample	Final OD_{600}	Cell number (cells mL^{-1})	Activity per volume (U mL^{-1})	Thermostability* after incubation at 60 °C for 30 min	Specific activity **1** (U mg^{-1})	Specific activity *(R)*-**3** (U mg^{-1})
Sc CAL-A	139	7.7×10^6	0.04 ± 0.006	45 ± 6	180 ± 3	below detection
Kl CAL-A	158	5.2×10^6	0.58 ± 0.022	69 ± 7	301 ± 9	4.04 ± 0.3
Hp CAL-A	62	21.1×10^6	0.85 ± 0.058	67 ± 6	434 ± 12	4.88 ± 0.1

*% as compared to unheated sample

The under-glycosylated variant was two times less active towards the *p*NP octanoate **1** than the glycosylated enzyme. Additionally, thermostability of the under-glycosylated variant was significantly reduced in temperatures ranging from 30-50 °C, whereas the glycosylated variant showed only a small loss of activity. In summary, it could be shown that for the tested lipase the hosts had an influence on the properties of the expressed enzyme, with a significant part of this influence being the result of various glycosylation. In the future, further investigation of this topic by comparative expression of other enzymes, preferably not closely related to each other, would allow making more general conclusions on this subject.

The second part of this thesis investigates rhamnosidases. These enzymes catalyse the cleavage of terminal α-L-rhamnose from a wide range of natural products. In spite of their biotechnological relevance, this class of enzymes is still far from being well characterised, especially as the heterologous expression is often problematic. Thus, apart from comparative expression, it was among the main goals of this study to gain more insight into the structure of rhamnosidases and their mechanism of action, as well as to modify the enzyme activity to the rhamnosynthase (Figure 1).

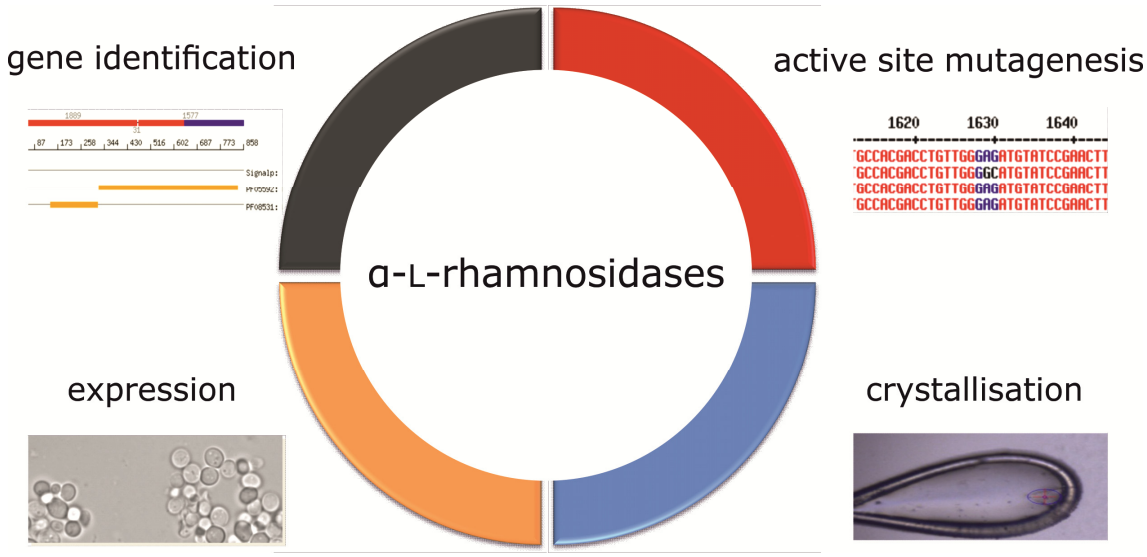


Figure 1: Schematic representation of the main goals targeted in the rhamnosidases investigation.

Initially, the gene for a new rhamnosidase from *Hansenula polymorpha,* for which encouraging characteristics have been previously published, was identified and expressed. Although many hosts were tested, in the end only intracellular homologous expression in *H. polymorpha* resulted in production of the active enzyme, albeit with a low yield of 1 mg L^{-1}. The enzyme was crystallised and plate-like crystals were obtained in several tested conditions; however, those resulted in poor resolution preventing complete structural analysis. In future experiments, these obstacles may be overcome by finding more suitable crystallisation conditions, where less anisotropic crystals would be obtained.

In light of these results, a rhamnosidase from *Bacillus* sp. was selected as the model enzyme to be modified to the glycosynthase. The crucial amino acids that are responsible for the hydrolytic activity of the enzyme were modified and a total of seven enzyme variants were created. Expression in *Escherichia coli* resulted in a good yield of 326 mg L^{-1} regardless of the introduced mutations. The enzyme was stable in up to 20% DMSO in reaction solution with only minor loss of activity. However, in spite of the various donors and acceptors tested, no transglycosylation product was observed; also optimisation of reaction conditions (pH, temperature, and buffer) did not bring any improvement. Further experiments are necessary to verify, if instability of the applied glycosyl group donors, fluorides, prevented the enzyme from efficient transglycosylation, thus leading to lack of product in spite of the variant's ability to transfer the glycosyl group. More stable donors (e.g., azides) should be selected for future glycosynthase approaches.

2. Scope and aims of this thesis

2.1 Introduction

The word 'enzyme' was first used by Wilhelm Kühne in 1877, but it was Eduard Büchner in 1897 who showed that a cell-free extract of yeast cells is able to ferment sugars, subsequently proving that enzymes are catalysing the reaction.[1] This discovery opened a new era in biochemistry and Büchner was thus awarded the Nobel Prize in chemistry in 1907. However, it took many more years of development in the field of molecular biology to grasp the full potential of enzymes as catalysts as many obstacles needed to be overcome. Firstly, potentially interesting enzymes must be identified from their host organisms. This challenge is within the realm of metagenomics and numerous protocols have been established and new tools are being constantly developed.[2] Furthermore, development of advanced sequencing tools combined with rapid decrease of sequencing prices has resulted in the publication of whole genomes of many microorganisms. As a result, thousands of sequences of potentially interesting enzymes can be easily compared and analysed *via* online tools (e.g., SIB ExPASy Bioformatics Resources Portal,[3] BLAST[4]). However, it is estimated that only 1% of all available microorganisms are accessible through cultivation conditions[5] and as such their potentially attractive enzymes need to be heterologously produced in established expression platforms.

A wide selection of expression platforms designed for different applications is available nowadays. *E. coli* and other bacteria are well-suited for enzyme production for technical purposes, whereas eukaryotic organisms are preferred when specific modifications are required and safety and authenticity concerns apply.[6] Numerous comparative expression studies have been previously performed, including both prokaryotic and eukaryotic host organisms.[7, 8] However, the main focus was primarily on the differences in productivity. Eukaryotic hosts are able to perform a variety of post-translational modifications making it possible to alter the characteristics of the enzyme to a certain extent. In fact, some reports have tackled this issue, but the published results are contentious. For example, the elastase from *Pseudomonas aeruginosa*, expressed in *Pichia pastoris*, was significantly less stable during reaction in both aqueous medium and with 50% (v/v) organic solvent in nonglycosylated form than the glycosylated form.[9] The cellobiohydrolase I from the filamentous fungus *Trichoderma reesei*, when heterologously expressed in *Pichia pastoris,* showed a significant folding difference and an activity that was 17-times lower when

compared to the original host's enzyme.[10] On the other hand, glucoamylase from the dimorphic yeast *Arxula adeninivorans* was heterologously expressed in *S. cerevisiae* and no difference in basic enzyme properties was observed when compared to the enzyme produced in *A. adeninivorans*.[11] In summary, although a lot of interesting remarks on the potential host influence were published, the systematic study, including more hosts and detailed characterisation of the produced enzyme, was missing to obtain the whole picture. It is also important to mention that most of the comparative studies discussed thus far directly used the gene isolated from the host, therefore using the original codon usage of the source microorganism. As a consequence, rare codons could be a reason for low yields of enzyme obtained. This factor should be eliminated for the comparative study in order to make the results as comparable as possible.

High yielding expression of the desired enzyme is a prerequisite for application in organic synthesis; however, it may still be just the first step towards successful application of the enzyme. When the use of enzymes as catalysts was first established, crude extracts of the cells were used and there was no possibility to influence the enzyme selectivity. As the discipline of molecular biology progressed, the sequences and structures of enzymes became known and protein engineering enabled expansion of the substrate range of the enzymes, including also non-natural substrates. In the third wave of biocatalysis, as summarised by Bornscheuer *et al.*,[12] it became possible to match the enzyme to the reaction; specific parameters such as stereoselectivity, activity, stability could be selectively targeted and improved. One prominent example is an (*R*)-selective transaminase ATA-117, a close homologue of the wild-type enzyme. It had no detectable activity on the desired substrate prositagliptin and was submitted to many rounds of protein engineering, using a substrate walking, modelling, and mutation approach. The first variant had a very low activity (0.2% conversion of 2 g L^{-1} substrate using 10 g L^{-1} enzyme) towards prositagliptin, whereas the best obtained variant converted 200 g L^{-1} prositagliptin ketone to sitagliptin of >99.95% e.e. (enantiomeric excess) with a 92% yield at the end of reaction.[13] As a result of the progress in the enzyme modifications, more and more reactions that are problematic with classical chemical synthesis methods are performed with the usage of specialised enzymes.

One field of organic chemistry where the application of enzymes is especially practical is carbohydrate chemistry. There are three major challenges to overcome during the synthesis of oligosaccharides:

- stability of the leaving group on the monosaccharide serving as a donor
- regioselectivity towards the selected hydroxyl group on the monosaccharide serving as an acceptor
- stereoselectivity in forming just one anomer of the final product

As a result, multiple protection and deprotection steps are necessary, leading to complicated procedures and low yields. Application of selective enzymes is therefore a promising alternative. Two types of enzymes can be used for synthesis of glycosidic bonds: glycosyl hydrolases and glycosyltransferases. Glycosyl hydrolases can promote oligosaccharide synthesis in two ways: either by shifting the reaction equilibrium *via* large excess of the acceptor or by using activated glycosyl donors. In both approaches, low to moderate yields can be obtained. Glycosyltransferases, on the other hand, offer high selectivity and excellent yields, but are limited by their poor availability and the high cost of the substrates.

The third and preferred route for oligosaccharide synthesis is application of engineered glycosyl hydrolases, glycosynthases. This method was first introduced by Withers *et al.* in 1998 wherein they created a mutant β-glycosidase with inactivated hydrolytic activity yet preserved glycosyl group transfer activity. Application of specific glycosidic group donors resulted in oligosaccharide production with an 80% yield.[14] Since then, more than 40 glycosynthases have been reported and a variety of glycosyl groups can be transferred. However, there are still glycosyl hydrolase families which are only poorly studied. To date, no glycosynthase capable of transferring a rhamnosyl group has been reported, although this group is widespread among natural products, such as naringin, rutin, quercitrin, hesperidin, diosgene and terpenyl glycosides. Rhamnosidases, although biotechnologically important, have never been applied to protein engineering; no attempts to improve their catalytic activity have yet been undertaken and there are only three crystal structures available. There is a scientific need to investigate these enzymes in more detail, as their full potential is still far from being discovered.[15] To fill in this gap and

gain a more thorough insight into the rhamnosidases, three enzymes from this family were selected for investigation.

2.2 Aims of this thesis

Within this thesis, two main topics are addressed. Firstly, it was aimed to answer an important question: whether the microbial expression host has an influence on the properties of the produced enzyme. For this purpose *Candida antarctica* lipase A (CAL-A) was selected as a model biocatalyst for the comparative expression which was to be performed as presented in Figure 2.

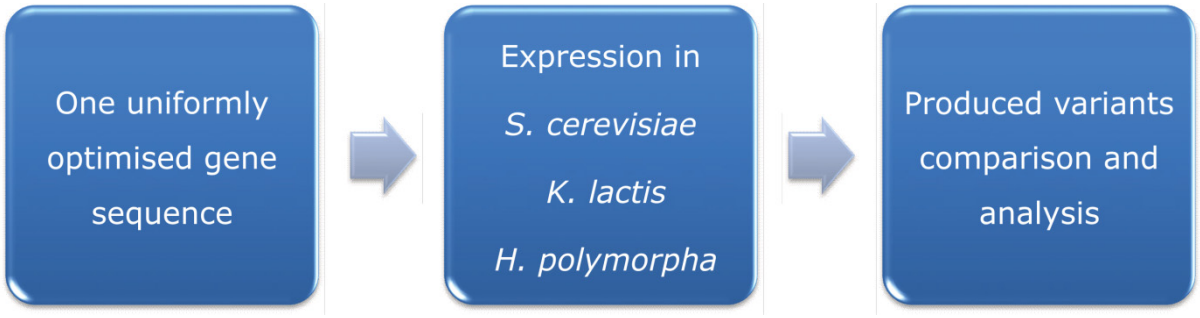


Figure 2: A workflow of the comparative expression analysis in yeast systems.

To exclude the codon usage effect on the produced enzyme quantity, the original CAL-A sequence was first codon optimised, taking into account preferred codon usage of all three hosts. The gene was then cloned and heterologously expressed in three industrially relevant hosts: *S. cerevisiae*, *K. lactis* and *H. polymorpha*. Efforts were undertaken to keep the experiments as comparable as possible. Not only the productivity was compared, but the enzyme variants were purified and further characterised. The key points were thermostability, kinetic parameters and selectivity of the enzyme. In addition, more detailed analysis of the glycosylation as a possible factor in influencing these parameters was performed. It was important to elucidate if the glycosylation impact on the enzyme properties is caused solely by the carbohydrate chains' presence on the enzyme surface. Alternatively, the translation process itself can also have an impact on the enzyme *via* subtle changes in enzyme folding, caused by co-translationally occurring glycosylation. To investigate this matter, expression in both presence and absence of the glycosylation inhibitor tunicamycin was performed and resulting enzymes compared.

The second topic addressed during this thesis was into a lesser known class of enzymes - rhamnosidases [E. C. 3.2.1.40], with the ultimate goal being the creation of a new type of glycosynthase *via* modification of the rhamnosidase. Application of yeast host systems for the rhamnosidase expression was supposed to facilitate the expression of the enzyme in desired quantity, as heterologous expression of the rhamnosidase in standard host *E. coli* was shown to be problematic in certain cases.[16] Three rhamnosidases are selected for this investigation. The rhamnosidase from *Bacillus* sp., which is the best characterised rhamnosidase to date, was modified *via* the Quikchange PCR to create variants which are unable to hydrolyse the substrate, yet preserve the ability to transfer carbohydrate groups onto acceptor substrates. The enzyme variants were then tested for potential rhamnosynthase activity (Figure 3).

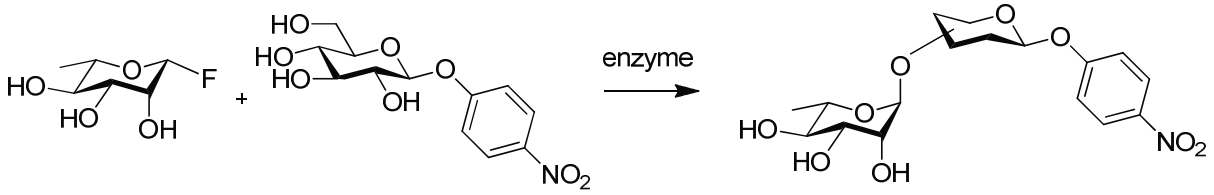


Figure 3: Example of potential biocatalytic usage of the rhamnosynthase. The rhamnosyl group could be transferred from the donor β-D-rhamnosyl fluoride onto *p*NP β-D-glucopyranoside *via* an enzyme, resulting in disaccharide product.

Since this type of glycosidase has, to the best of my knowledge, not previously been applied in a glycosynthase approach, many parameters including optimal amino acid modification, donor, acceptor, and reaction conditions had to be established. Since not every modified enzyme results in glycosynthase activity, two lesser known rhamnosidases were also included in the experiments. The rhamnosidase from *Lactobacillus acidophilus* was heterologously expressed, but no active enzyme could be obtained, thus the enzyme could not be applied for mutagenesis.

For the rhamnosidase from *H. polymorpha*, gene identification was performed. The enzyme was successfully produced in *H. polymorpha* and the protocols for expression and purification of this enzyme were established. The protein was applied for crystallisation trials and several conditions for crystal formation were identified. Consequently, an important step towards gaining more information about this enzyme class with only three available structures was made. These results are an encouraging basis for further experiments leading to the structure solution the enzyme.

3. Comparative expression in yeast systems

3.1 Introduction

In the first part of this thesis, the potential influence of eukaryotic host on the produced enzyme's properties was investigated. Three industrially relevant yeast strains were chosen for comparison: *S. cerevisiae*, *K. lactis* and *H. polymorpha*. To combine all conceivable host effects, one model protein was produced in all three yeasts and not only the amount of produced protein was compared, but also other biotechnologically important properties, including the activity towards various substrates. Thereby, the main goal of this study was to verify, if the differences in processing between the yeasts could result in altered characteristics of the enzyme. As a new approach to minimise effects originating from the original heterologous sequence, I used one synthetic gene with uniformly optimised codon usage for all three hosts.

Lipase A from *Candida antarctica* (CAL-A) was chosen as a model enzyme for its interesting biocatalytic properties and the easy-to-handle size (45 kD). In addition, more detailed characterisation of CAL-A revealed high thermostability and high activity towards sterically hindered alcohols. It was therefore interesting subject to verify how different expression hosts can affect the properties of this biocatalyst.

In the following subsections, current state of the knowledge is summarised, whereas the results and discussion are presented in section 3.2.

3.1.1 Enzymes in catalysis

The importance of enantiomerically pure organic compounds is steadily growing in many branches of industry; the global 'chiral market' was worth nearly $5.3 billion in 2011 and is expected to grow to $7.2 billion by 2016, according to market research company BCC Research.[17] In this, application of enzymes as chiral catalysts has gained increasing attention with the focus on safe, resource efficient and environment-friendly production procedures.[18] Though more problematic when compared to traditional catalysts, they offer several advantages that have made them more and more popular in recent years, both in the research and on the industrial scale. Firstly, they usually operate at mild conditions (pH 5.0 to 8.0, temperatures 20 to 40 °C), mostly in aqueous systems, although many enzymes are able to tolerate organic solvents. From a chemical point of view, high selectivity for the catalysed reactions

combined with high degrees of diastereo-, regio- and enantioselectivity, make these proteins highly desirable alternatives to the traditional chemical catalysts. These exceptional features make the enzymatic synthesis more profitable, even when taking into account the high cost of enzyme production and purification, as well as limited stability.[19]

Since the first usage of the word 'enzyme' in 1876, thousands of proteins with catalytic activity have been discovered and described. To organise the data in a more transparent fashion, in 1966 the Enzyme Commission published the enzyme nomenclature scheme, which divided enzymes into six classes based on the type of the catalysed reaction (Table 4). Since then, each enzyme has been defined by a four-number code. For example, rhamnosidase is identified with the code EC 3.2.1.40 where EC is short for Enzyme Commission. The first number indicates the type of catalysed reaction (3 - hydrolase) and the second indicates the nature of the chemical bond hydrolysed (2 - glycoside). The third number specifies the nature of the substrate (1 - *O*-glycoside) and finally the fourth one is a serial number of the enzyme in its sub-subclass.

Table 4: Enzyme classification and the catalysed reaction types.

Class	Name	Catalysed reaction	Examples of enzymes
EC 1	Oxidoreductases	oxidation/reduction reactions	Dehydrogenase, oxidase
EC 2	Transferases	transfer of a functional group from one substance to another	Transaminase, kinase
EC 3	Hydrolases	hydrolysis reaction	Lipase, protease
EC 4	Lyases	non-hydrolytic addition or removal of groups from substrates	Aldolase, decarboxylase
EC 5	Isomerases	intramolecular rearrangement, i.e. isomerisation changes within a single molecule	Racemase, phosphoglucomutase
EC 6	Ligases	condensation reaction of two molecules with simultaneous breakdown of ATP	Synthetase

Even though more and more enzymes were characterised, further progress in molecular biology was necessary to make them accessible in required amount. It

is estimated that only 1% of all available microorganisms are accessible through cultivation conditions.[5] Therefore, the identified enzymes need to be heterologously produced in established expression platforms. The development of the PCR method[20] and rapidly decreasing costs of gene synthesis has made heterologous expression possible. Modern recombinant DNA techniques, development of new expression hosts, and an increasing database of characterised enzymes [e. g., www.brenda-enzymes.org] increase the cost-effectiveness of biocatalytic processes. Furthermore, the third wave of biocatalysis offers new advanced tools, like protein engineering, gene synthesis, sequence analysis, bioinformatics tools, and computer modelling, making it possible to design the enzyme exactly for the desired application.[12]

3.1.2 Recombinant protein production in yeast

Yeasts are among the best-studied hosts for heterologous gene expression. They can be easily grown to high density in shake flasks or fermenters as is done with *E. coli*, but in addition possess the ability to perform posttranslational modifications, which may be crucial for the expression of eukaryotic proteins.[21] Many eukaryotic proteins require these modifications as a prerequisite for correct folding and activity. Among the posttranslational modifications, glycosylation is of great importance, as added sugar chains can constitute a significant part of the total protein and thus have an effect on enzyme properties.[22, 23] Each yeast has its own glycosylation pattern and additionally incorporates various amount of the carbohydrate residues into the protein, leading to the mixture of glycoforms of the same enzyme.[24] In spite of this fact, most comparative expression studies have focused on the productivity of various organisms.[7] Although an impact of posttranslational modifications, e.g., glycosylation on temperature stability of the enzyme was previously reported,[25, 26] no systematic comparison of the glycosylated protein properties from three eukaryotic hosts is to my knowledge available to date.

In a comparative expression study of glucoamylase from *Arxula adeninivorans*, the original gene was used for expression in *S. cerevisiae* and no difference in thermostability of enzyme produced in different hosts could be observed.[27] However, as summarised by Skropeta[28] examples of both increases and decreases in stability and activity of enzyme due to glycosylation

can be found. Additionally, in the heterologous expression study of cellobiohydrolase I from T*richoderma reesei* in *P. pastoris*, a significant folding difference between the proteins produced in both organisms was found, which resulted in a 17-times lower activity when compared to the enzyme purified from the native host.[29] These differences were also reported for closely-related yeast species. When a thermophilic esterase was expressed in *Kluyveromyces lactis* and *Kluyveromyces marxianus*, it was similarly glycosylated in both organisms as judged from SDS-PAGE analysis, yet differences in thermostability were found.[30]

For this study, three industrially relevant yeast strains were chosen: *S. cerevisiae*, *K. lactis* and *H. polymorpha*. Their key features, as well as examples of yield obtained, are summarised in Table 5. Detailed description of each strain can be found in the following sections.

Table 5: Yeast strains used in this study and their key features. Promoters selected for this study are marked in bold.

Name	Preferred carbon sources	GC content	Established promoters	Examples of the produced proteins and yields obtained	
Saccharomyces cerevisiae	Glucose Galactose	38%	***GAL1*** *ADH2* *CUP1* *GAP* *PGK*	scFv insulin precursor	20 mg L^{-1} [31] 75 mg L^{-1} [32]
Kluyveromyces lactis	Glucose Lactose Galactose	40%	***LAC4*** *PHO5*	α-galactosidase lysozyme	250 mg L^{-1} [33] 100 mg L^{-1} [34]
Ogataea angusta (*Hansenula polymorpha*)	Glucose Glycerol Methanol	48%	***FMD*** *MOX*	glucoamylase phytase	1.4 g L^{-1} [35] 13.5 g L^{-1} [36]

3.1.2.1 Saccharomyces cerevisiae

S. cerevisiae is budding yeast that has been used for ages in biotechnological applications. Although it is commonly associated with the brewing and baking industries due to its ability to produce ethanol and carbon dioxide, it is also a model eukaryotic host and is widely applied for the production of proteins even on an industrial scale. In culture, it has a relatively short generation time, doubling its cell density approximately every 1.5 – 2.5 h (when growing on glucose) at its preferred growth temperature of 30 °C. The genome

of *S. cerevisiae* was published in 1996[37] and it was the first completely sequenced genome from a eukaryote. Nowadays, there is a wide selection of genetic tools for this host available. A strong *GAL* promoter from the galactose utilisation pathway is one of the most often used promoters for this organism. The major disadvantages of this system are possible hyperglycosylation of produced proteins, as well as the retention of the product in periplasmic space, resulting in its partial degradation.[38] Nonetheless, *S. cerevisiae* still remains the main eukaryotic host for the production of recombinant therapeutics approved by the Food and Drug Agency (FDA).[39]

3.1.2.2 Kluyveromyces lactis

The budding yeast *K. lactis* is closely related to *S. cerevisiae* and well established as a producer yeast strain. It offers a powerful promoter *LAC4* from the lactose utilisation pathway which is repressed by glucose in low amounts;[40] stable transformants can be obtained both by episomal plasmids or by integration of a gene of interest into the genome;[41] in both cases, transformants can be obtained without the time-consuming passaging procedure, which is required by *H. polymorpha*. As with *S. cerevisiae*, its preferred growth temperature is 30 °C. *K. lactis* has been used in the food industry for the production of lactase (β-galactosidase) since 1950s, as well as for heterologous expression of bovine chymosin (rennin).[42] The big advantages of this host are the ability to secrete high molecular weight proteins[43] and a completely known genome, which was published in 2004.[44] In addition, a complete secreted proteome was analysed in detail in 2008,[45] followed by the analysis of the carbon source impact on the secretome in 2009.[46] When compared to other popular yeast strains, protein expression in *K. lactis* proceeds without application of toxic methanol, which is necessary for *P. pastoris* expression under commonly used *AOX* promoter.

3.1.2.3 Hansenula polymorpha

Hansenula polymorpha, recently reclassified as *Ogataea angusta*,[47] is one of the few yeasts able to utilise methanol as a carbon source. It is a thermotolerant yeast and some strains have been reported to survive even 56.5 °C heat shock, due to their ability to accumulate trehalose in their

cytosol.[48] During the growth on methanol, some enzymes regulated on the transcriptional level are present in high amounts, especially methanol oxidase (*MOX*), formate dehydrogenase (*FMD*) and dihydroxyacetone synthase (*DHAS*).[35] These promoters are regulated by carbon sources; repressed by glucose, derepressed by glycerol and induced by methanol. High yield expression of a protein of interest by *H. polymorpha* is possible only after multiple plasmid integration into the genome after a passaging procedure[49] and can result in up to 100 copies per haploid genome after supertransformation.[50] The combination of high copy number and efficient promoters can result in high yield of the produced proteins (e.g., 13.5 g L^{-1} for secreted phytase [36]).

3.1.3 Gene optimisation

The expression of functional proteins in heterologous hosts in reasonable yield is the basis for their biotechnological application. Unfortunately, some proteins are problematic in expression, resulting either in low yields, and/or in an inactive enzyme. There are several possible explanations to these difficulties: e.g., the genes of the problematic proteins might contain codons which are rarely used in the selected host, originate from organisms that use a non-canonical code, or contain expression-limiting regulatory elements within their coding sequence.[51] To analyse this phenomenon, codon usage databases have been created, collecting data on the frequencies with which different codons are used in the organism (e. g., http://www.kazusa.or.jp/codon). In general, the more codons a gene contains that are rarely used in the expression host, the less likely it is that the heterologous protein will be expressed at reasonable levels. In particular, rare codons appearing in clusters or at the N-terminal end of the protein can lead to low expression levels.

Thus, a possible solution to improve expression is to exchange the rare codons in the target gene so that they more closely reflect the codon usage of the host, without modifying the amino acid sequence of the encoded protein (silent modification). Numerous studies have shown that there are other factors involved, such as the GC content and repeated sequences;[52] furthermore, the rare codons in the native gene may play a significant role during protein folding by regulating the rate of protein synthesis and thus allowing for required secondary and tertiary structure formation by the nascent polypeptide.[53] Therefore, the method of choice would be to optimise the codon usage of the

natural host by application of the codons in comparable usage frequency for the heterologous host. However, this is impossible to do when more than one host is used for expression of the same protein.

3.1.4 Posttranslational modifications

Protein posttranslational modification (PTM) is a cell's tool to modulate the characteristics of the produced protein, leading to – among other things - increased functional diversity of this protein. Covalent addition of a modifying group, which is one of the most common PTM, influences the basic properties of a protein, such as activity, localisation, stability and interaction with other proteins. There is a wide selection of posttranslational modifications which can be found in Fungi, among others phosphorylation, glycosylation, acetylation, methylation, ubiquitination, and sumoylation (covalent addition of **S**mall **U**biquitin-like **Mo**difier proteins); together with tight gene regulation, PTMs allow the microorganisms to successfully cope with the constantly changing environment they occupy.[54] For secreted proteins, glycosylation is a major modification, as large carbohydrate chains are added in a host-specific manner.

Protein glycosylation of secretory and membrane-bound proteins is an essential and the most abundant protein modification in lower and higher eukaryotes; additionally, glycosylation occurrence in Bacteria and Archea was also recently reported.[55] Fungi and animals share three essential types of protein glycosylation that are initiated in the endoplasmic reticulum (ER): *N*-glycosylation, *O*-glycosylation and glypiation. It was shown in numerous studies that a variety of protein-linked glycan structures play crucial roles in many cellular processes including cell-cell recognition, signal transduction, and ER protein quality control.[56, 57]

***N*-glycosylation** is the most complex covalent modification, and many proteins require this kind of modification to be able to fulfil their functions. This modification is species-specific and typically very heterogeneous, thus very challenging to analyse. It proceeds by a pathway that has been highly conserved during evolution, which suggests that the saccharide moieties have an important function.[57] Carbohydrates can make up to 86% of the total mass of the glycoprotein, thus having a significant impact on its properties.[58] Addition of such a huge oligosaccharide chain cannot help but influence/impact the enzyme.

In fact, it has been shown for numerous proteins that carbohydrates can stabilise the protein[59] as well as influence enzyme characteristics.[60]

Glycoproteins occur as mainly extracellular, vacuolar or membrane bound enzymes and cell wall structural components. *N*-glycosylation occurs co- and posttranslational on the specific recognition sequence N-X-S/O, where X can be any of the amino acids with the exception of proline. First, the core 14-saccharide structure is added at the recognition sequence in the ER. The core structure is then further modified in the Golgi apparatus in an organism-dependent way (Figure 4).

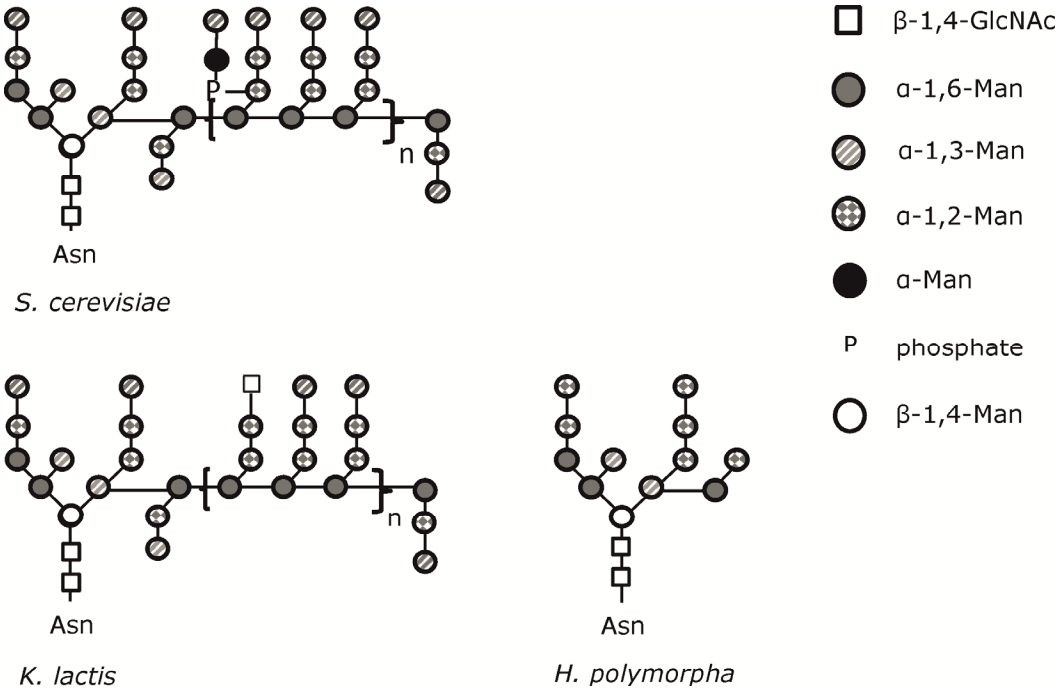


Figure 4: ***N*-glycosylation structures found in different yeast species used in this study.** The information on structures is from [61] and [62]. Adapted from Gellissen, 2002.[50] Abbreviations: GlcNAc – *N*-Acetylglucosamine; Man – mannose; Asn – asparagine, the amino acid where the sugar chain is added *via* β-linkage.

Terminal mannose residues in *N*-linked glycans are added in *S. cerevisiae* by an α-1,3 bond which is considered to be allergenic (they can cause immunogenic reactions in humans), whereas in *H. polymorpha* and *K. lactis* a non-allergenic α-1,2 bond is present.[63] Asparagine-linked oligosaccharides in yeasts are mainly of the high mannose type, whereas in higher eukaryotes they can be very complex. In addition, *S. cerevisiae* has the most extensive hyper-glycosylation, typically adding 50–150 mannose residues, while in *H. polymorpha and K. lactis* hyperglycosylation is less pronounced.[64]

***O*-glycosylation** in yeasts occurs *via* the addition of a short oligomannose chain. It forms a glycosidic bond in an alpha anomeric configuration with the hydroxyl group of a serine or threonine residue. So far, no specific amino acid recognition sequence or structural features have been identified, which would be a prerequisite for *O*-glycosylation to occur.[65] Nevertheless, *O*-glycosylation can be predicted to certain extent using advanced analytics tools, e.g., the YinOYang 1.2 server.[66] The addition of the mannose takes place in the ER and is catalysed by a conserved family of protein *O*-mannosyltransferases (PMTs). It is an essential modification as knockout of a single PMT can affect various cellular processes like mating, filamentation, or cell wall integrity. Knockout of certain combinations of PMTs has been demonstrated to be lethal for the cell.[67]

Glypiation is the covalent attachment of a glycosylphosphatidylinositol (GPI) anchor to the nascent protein in the ER lumen, which directs proteins to the cell membranes. GPI assembly takes place entirely on the cytoplasmic side of ER, followed by translocation to the luminal side, where it is attached to the protein.[56] Many biologically important cell surface proteins are GPI-linked, and it is implicated that this modification provides stable membrane anchorage, as well as is involved in signal transduction events.[68]

3.1.5 Model enzyme – lipase CAL-A

Lipase A from *Candida antarctica* (CAL-A) was chosen as a model enzyme for this study, as it is a relatively small protein (45 kD), was already successfully expressed in a eukaryotic host[69] and it has been shown to possess interesting biocatalytic properties.[70] The sequence of this lipase was reported in 1996[71]together with a lipase B from the same organism (CAL-B), which gathered much more attention at the beginning due to its early recognised stereoselectivity. More detailed characterisation of CAL-A revealed high thermostability and exceptionally high activity toward sterically hindered alcohols.[70] The crystal structure of CAL-A was solved in 2008[72] (Figure 5, PDB code 2VEO); it is reported to be typical for hydrolases α/β domain with a well-defined lid.

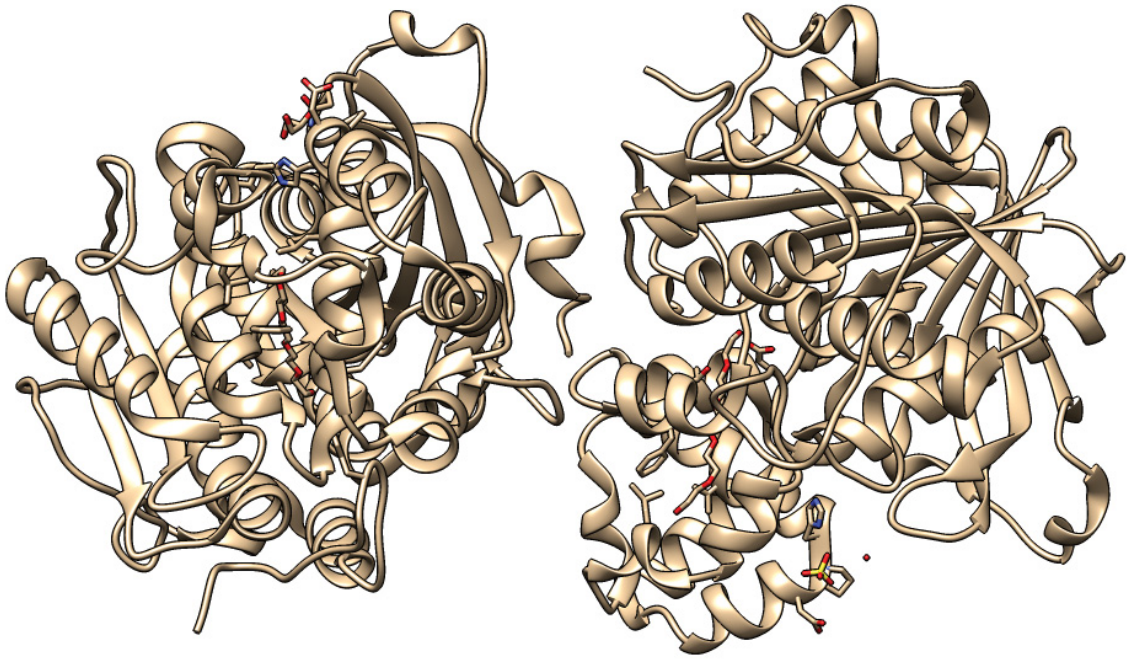


Figure 5: Crystal structure of lipase A from *C. antarctica*. The structure is deposited at PDB under code 2VEO. The asymmetric unit was reported to consist of two monomers, although they do not make extensive contact with each other.[72]

These data allowed better understanding of the exceptional activity of the enzyme, as well as rational design of the improved enzyme variants. Moreover, protein engineering of CAL-A enabled astounding improvement of both enantioselectivity and activity towards broad scope of α-substituted esters.

Several rounds of directed evolution resulted in CAL-A variant with the E-value (the enantioselectivity value - the parameter commonly used in characterising the enantioselectivity of reaction) for the model substrate *(S)*-4-nitrophenyl 2-methylheptanoate increased from 5.1 to 52.[73] Consequently, semirational mutagenesis approach has resulted in an excellent E-value of 100 obtained for the *(S)*-4-nitrophenyl ibuprofen ester, while the wild type was reported to accept this substrate with only low enantioselectivity (E-value 3.4).[74] In summary, CAL-A is nowadays recognised as a powerful biocatalyst, which has recently found an application in acylation of phytosterols[75] and kinetic resolution of 1,2-diarylethanols and 1,2-diarylethanamines.[76]

3.2 Results and discussion

3.2.1 Sequence optimisation

The primary aim of this study was the investigation of the influence of the host on the produced protein. For this purpose, a single optimised gene was cloned into three different plasmids for expression, keeping the constructs as identical as possible. Lipase CAL-A originates from the extremophilic microorganism *Candida antarctica*, and the original gene sequence was characterised by a high GC content of 62%, whereas for the investigated yeasts the GC value on average was reported to be much lower, between 38-48%.[77] Numerous studies were conducted on codon optimisation influencing the level of gene expression; [52, 78] but no universal optimising strategy to assure the highest possible expression yields has been established so far. A number of variables affects the final obtained yield in often independent manner, including relative frequencies of different codons in the sequence, GC content, the presence of cryptic transcriptional terminators as well as the tendency of the 5'-end of the mRNA to fold into stable secondary structures.[78] Following a published approach,[51] the CAL-A sequence was first codon-optimised using *S. cerevisiae* codon usage table. In this method, rare codons, defined as occurring with frequency per thousand lower than 5, were replaced by more common ones, as presented on the arginine example, where codons CGC, CGA, CGG were replaced by clearly preferred in *S. cerevisiae* codon AGA (Figure 6).

A

[triplet] [amino acid] [fraction] [frequency: **per thousand**]

CGU	**R**	0.14	6.4
CGC	**R**	0.06	2.6
CGA	**R**	0.07	3.0
CGG	**R**	0.04	1.7
AGA	**R**	0.48	21.3
AGG	**R**	0.21	9.2

B

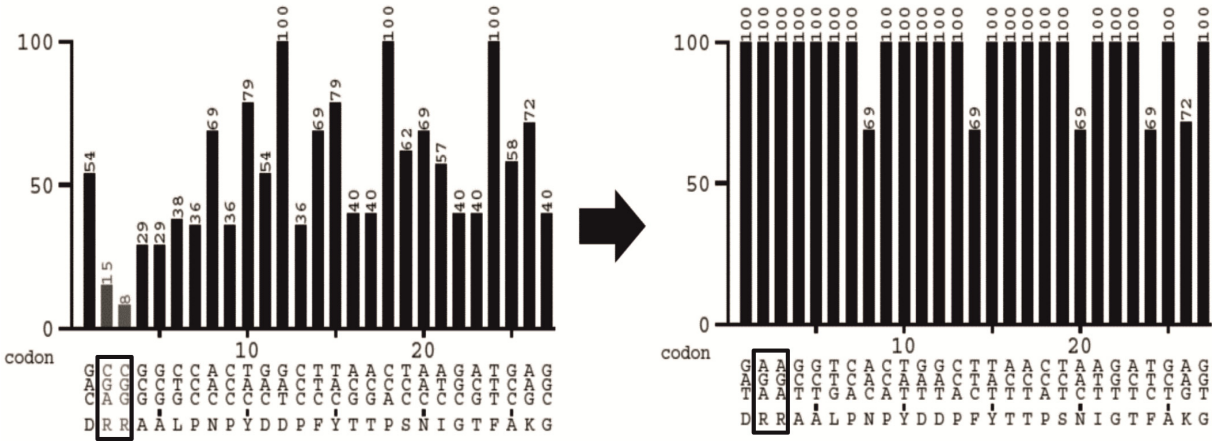


Figure 6: Example of codon usage optimisation of original CAL-A sequence for *S. cerevisiae*. A – published codon usage frequencies in *S. cerevisiae* for the codons representing arginine, with rarely used codons marked in red and the preferred codon in blue (Codon Usage Database), B – fragment of original CAL-A sequence before (left graph) and after (right graph) the codon usage optimisation for *S. cerevisiae*. The graphs represent the relative adaptiveness values for each codon, which were calculated according to published codon usage data (NCBI GenBank®) on a scale from 0 to 100, with 100 being the optimal codon for the amino acid. The original sequence contained two rarely used codons for the arginine, marked with black box (left graph), which were then replaced by preferred one for this organism, AGA (right graph). The remaining part of the sequence was then optimised accordingly.

The optimised sequence was then further adjusted for two other yeasts *via* the exchange of rare codons (defined as less than 10% of relative adaptiveness value) to more common ones, followed by adjusting the GC content of the gene to 40%. It is worth mentioning that there were only minor differences in preferred codons between *S. cerevisiae* and *K. lactis*, as this two species are closely related (see the used codon usage tables in section 6.2.1). A comparison of the parameters for the original and optimised sequence is shown in Figure 7, in which only a short fragment of the original 462 codon long ORF is presented for clarity; calculated parameters for full sequences before and after codon optimisation can be found in section 6.2.

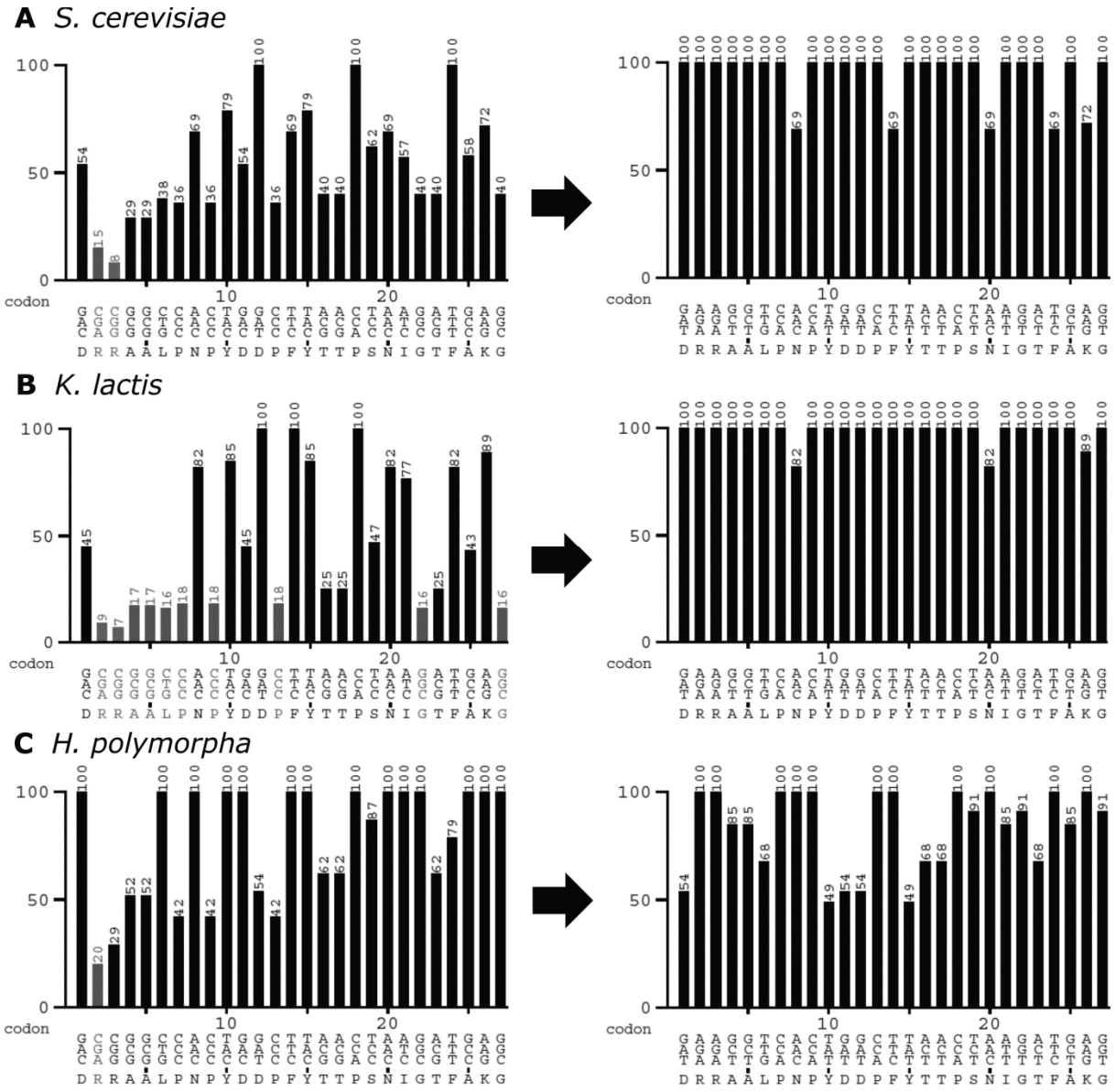


Figure 7: Codon usage optimisation of CAL-A. A – graphs for *S. cerevisiae*, B - graphs for *K. lactis*, C - graphs for *H. polymorpha*. The graphs represent the relative adaptiveness values for each codon, which were calculated according to published codon usage data (NCBI GenBank®) on a scale from 0 to 100, with 100 being the optimal codon for the amino acid: graphs on the left side – original *C. antarctica* sequence; graphs on the right side – final optimised ORF. Values lower than 20% are marked in light grey. One optimised sequence was used for the expression of the protein in three hosts. For clarity, only a 20 amino acids fragment of the CAL-A sequence is shown; the whole sequences are available in section 6. All graphs were prepared with the Graphical Codon Usage Analyser tool, version 2.0.[79]

For each host the relative adaptiveness values per codon of the original sequence are presented on the left graph. Low values mark the unusual codons for the host and thus potential bottlenecks, which have been removed *via* manual exchange of codons, as shown on the graphs on the right site. It can be concluded that relative adaptiveness values of the final optimised CAL-A gene

sequence were usually above 50% with a slightly better adjustment to the taxonomically more closely related *S. cerevisiae* and *K. lactis* than to *H. polymorpha.*

3.2.2 Cloning and selection of the best expressing transformants

The standard yeast plasmids (pFPMT121 for *H. polymorpha* and pKLAC2 for *K. lactis*) were modified *via* insertion of a new common multiple cloning site as described in detail section 5.3. The restriction sites *AvrII* and *Pfl23II* were then used to insert CAL-A, obtained as a codon-optimised gene from GenScript in standard pUC57 plasmid, into the modified plasmids. For *S. cerevisiae*, two constructs from the basic pYES263 plasmid were prepared: both with a new multiple cloning site and one of them including an MFα secretion signal added for secretion of the lipase.

Two of the hosts used - *K. lactis* and *H. polymorpha* - integrate the expression construct, which may lead to significant variation in the protein expression levels between the transformants due to variation in copy number. To identify the best expressing transformants, initial screening of a number of transformants was performed. For *K. lactis*, 18 transformants were first tested for the correct integration of the gene into the genome using the standard colony PCR method, which was optimised in this study for application in yeasts as described in section 5.6.7.2, and five hits were identified. For *H. polymorpha*, 36 transformants were used for the passaging as described in section 5.5.8. The small scale expression in the DWP format, followed by SDS-PAGE and activity analysis (5.7.7) were used to identify the most efficient transformants from the *K. lactis* hits, the stabilised *H. polymorpha* transformants and six *S. cerevisiae* transformants. In case of *S. cerevisiae*, no significant differences in intracellular expression levels between the tested transformants could be observed. Unfortunately, for the construct with secretion signal no detectable presence of the lipase could ever be confirmed, either on SDS-PAGE or with the activity tests. Therefore, the transformants producing CAL-A intracellularly were used for further experiments. In case of both secreting yeasts the most effective producer was chosen for further experiments. The production of CAL-A was further tested on tributyrin agar plates (Figure 8).

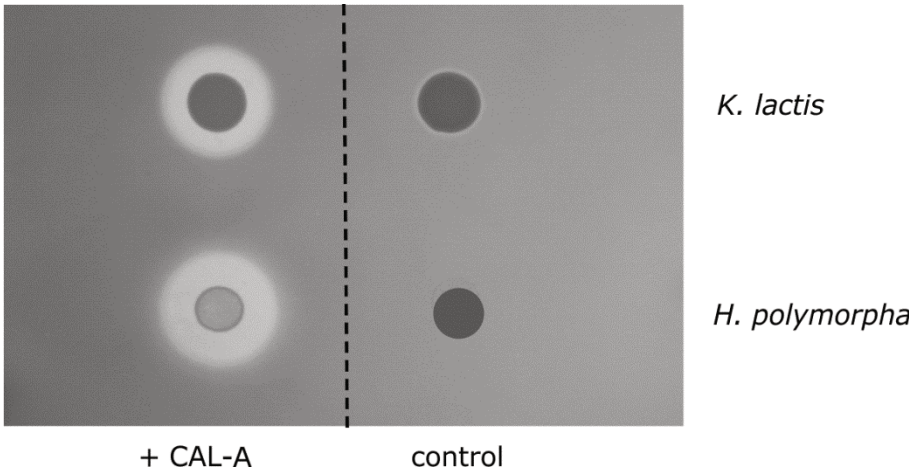


Figure 8: Tributyrin plate assay to verify secreted lipase activity. 3 µL of each culture was incubated on an agar plate containing 2% tributyrin at 30 °C for 48 h. On the left (+ CAL-A), clear halo formation around transformants secreting the lipase is visible, which is missing on the right (control) where transformants with the respective empty plasmids were used.

In contrast to the control colonies containing only empty plasmid, the lipase was successfully secreted by transformed *K. lactis* and *H. polymorpha*, as confirmed by formation of clear halos around the transformants. The transformants were then used for comparative expression in shake flask scale.

3.2.3 Media screening & purification from various yeasts

Initial expression trials were performed in standard media: YPGal (2% peptone, 1% yeast extract, 2% galactose) for *S. cerevisiae* and *K. lactis* and YNB (0.67% YNB; 1% glycerol; pH 6.0) for *H. polymorpha*. For both *S. cerevisiae* and *K. lactis*, using more concentrated 2×YPGal medium resulted in higher optical density values and increased amount of the produced protein, thus this was the medium of choice for the further experiments. Purification from the three hosts was performed with the usage of His_6Tag, added at the C-terminal end of the protein, using either an automated ÄKTA system (section 5.7.3) or hand-packed small columns (section 5.7.4). Purification from both *S. cerevisiae* and *K. lactis* could be performed with good yields; however, the initial purification from *H. polymorpha* failed, although high activity of the secreted lipase could be observed in the collected supernatant. A possible reason for this problem could be overglycosylation of the protein, which resulted in hiding of the His_6Tag, or expression of a truncated form of the protein with the part of C-terminal end missing.[80] Since the composition of media can have an influence on the degree of glycosylation, screening was performed using different composition of media

(Table 6), which were chosen from most popular media reported in the literature to be successful for high yield protein production in *H. polymorpha.* The selection enables to follow the influence of almost each ingredient on the overall protein yield.

After 72 h of expression, the supernatants were collected and used for protein purification as described in section 5.5.9. The activities in starting sample and flowthrough were determined *via* the standard assay (see section 5.7.7 for details) and compared. The results are presented in Table 6 and Figure 9.

Table 6: Composition of media used for the *H. polymorpha* expression optimisation. The improvement in obtained CAL-A activity in medium after 72 h of expression is presented as compared to the activity obtained with the standard medium (nr 8), taken as 100%. The activity was measured using activity assay with standard substrate *p*NP octanoate (see section 5.7.7 for details). YNB - yeast nitrogen base without amino acids, with ammonium sulphate unless otherwise specified.

Nr	Medium composition	Activity in medium (%)
1	0.67% YNB; 2% glycerol; pH 3.0	162
2	1.34% YNB; 1% glycerol; pH 6.0	178
3	1.34% YNB; 1% casamino acids; 1% glycerol; pH 6.0	1049
4	1.34% YNB; 1% casamino acids; 0.1% yeast extract; 1% glycerol; pH 6.0	1217
5	0.2% peptone; 0.1 % yeast extract; 0.3% casamino acids; 0.17% YNB w/o $(NH_4)_2SO_4$; 0.06% KH_2PO_4; 1% glycerol; pH 6.0	213
6	2% peptone; 1% yeast extract; 2% methanol	0
7	2% peptone; 1% yeast extract; 2% methanol, 2% sorbitol	0
8	**0.67% YNB; 1% glycerol; pH 6.0**	**100**

Out of 8 media tested, no growth could be observed in media nr 6 and 7, so they have been excluded from further investigations (Table 6). For all the others, a significant improvement in CAL-A activity in medium, more than 10-times in both cases, was obtained in media 3 and 4 when compared to the standard medium.

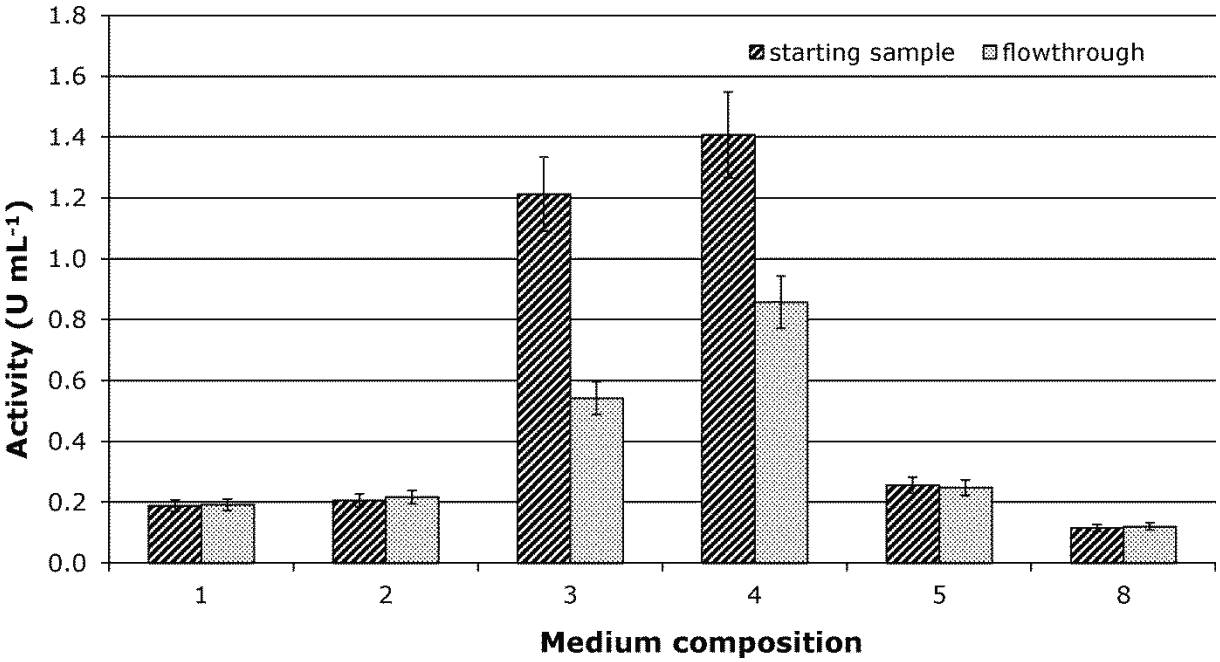


Figure 9: Binding analysis of CAL-A produced in *H. polymorpha* in the different media. After 72 h of expression, the spent culture medium from each culture was collected and the activity was measured in the starting sample and in the flowthrough after loading on the Ni-NTA column using the standard assay. The presented data are the average of three measurements; standard deviation values were within 10%. See Table 6 for composition of the tested media.

In addition, improved binding to the column could be observed for both media 3 and 4 (Figure 9). When the lipase was expressed in media 1, 2, 5 and 8, almost all activity of the initial sample was found in the flowthrough, indicating that no sufficient binding of the protein to the column was achieved. In contrast, when the supernatant from the cultures expressed in media 3 and 4 was applied on the columns, a significant decrease in activity between the initial sample and the flowthrough was found. For medium 3, 55% of the protein was bound to the column, and for medium 4 - 39%. Since the highest amount of protein per culture volume could be obtained from medium 4, this composition was selected as the optimal *H. polymorpha* medium for lipase CAL-A expression.

3.2.4 Comparative expression in shake flask scale

First, the production of the lipase CAL-A was confirmed by SDS-PAGE analysis as compared to the transformants containing respective plasmids without the lipase gene. Obtained bands in the expected size were then cut out from gel and used for MALDI TOF analysis, which confirmed production of the desired enzyme. The mass analysis was performed at the Center for

Biotechnology, Bielefeld University. After tryptic digestion of in-gel proteins, based on the m/z values, 12 peptides were matched for the CAL-A from *H. polymorpha* (46% sequence coverage), 9 peptides were matched for the CAL-A from *S. cerevisiae* (27% sequence coverage) and 11 peptides were matched for the CAL-A from *K. lactis* (45% sequence coverage).

For quantification of the CAL-A expression capacity of all three yeasts, comparative expression in 100 mL media in baffled shake flasks was performed. For all three hosts highest lipase activity was observed in the stationary growth phase as measured with the standard lipase assay (see section 5.7.7 for details) and this time point was thus selected for the following experiments. The results are summarised in Figure 10, where the produced enzyme's activity values are at first presented as activity per culture volume; additionally the corresponding CAL-A activities referring to the cell numbers are shown, as significant differences in the cultures' final optical densities were observed. Typical OD_{600} values at the time of harvesting were about 139 for *S. cerevisiae*, around 158 for *K. lactis*, and ca. 62 for *H. polymorpha*. The OD_{600} value, although widely applied to follow the cell growth in liquid cultures, does not take into account possible variations between the cultures, such as different cell sizes or media turbidity. To cope with the apparent growth disparities of the hosts, comparison of the produced CAL-A activities per cell instead of per OD_{600} units was used. The cells were counted as described in section 5.5.12. The obtained numbers of cells per mL culture diluted to OD_{600}=1 were 7.7×10^6 cells mL^{-1} for *S. cerevisiae*, 5.2×10^6 for *K. lactis,* and 21.1×10^6 for *H. polymorpha*.

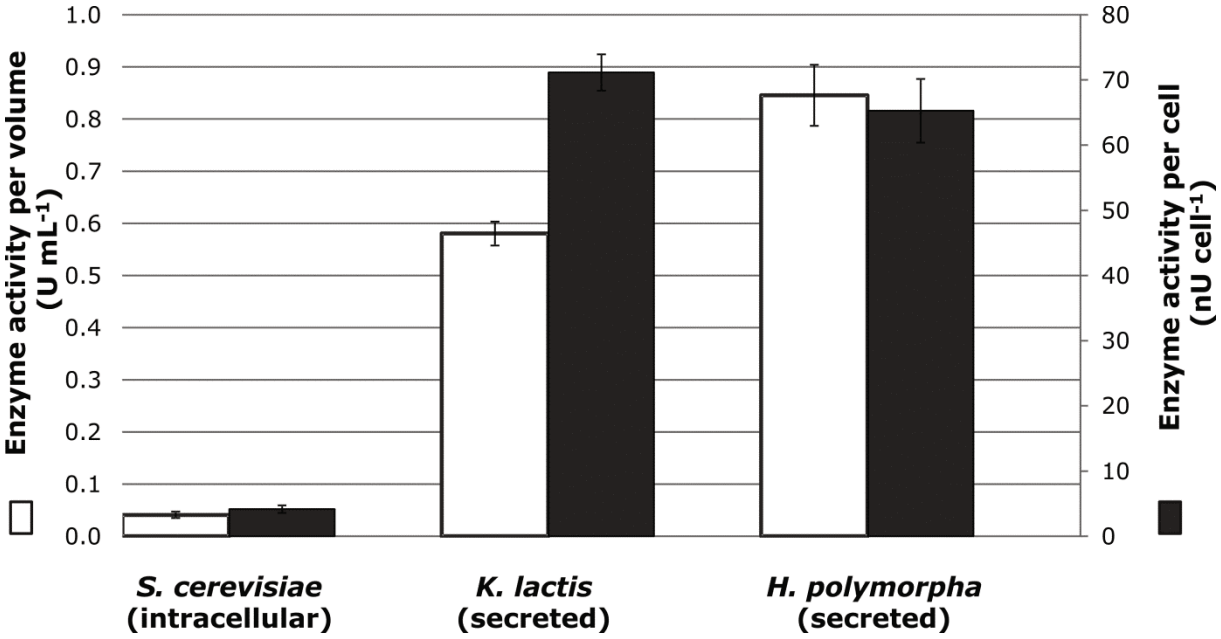


Figure 10: Lipase CAL-A production capacity of the different strains. Enzyme activity per original culture volume (empty bars) and activity per cell number of the respective culture (filled bars) was measured in crude *S. cerevisiae* extract or in fivefold diluted spent culture medium of the other yeasts. All samples were taken at the stationary growth phase, when usually maximal CAL-A activity was obtained. In each comparative set of results, the highest activity value measured was set to 100%.

The comparison of the CAL-A activities in Figure 10 shows significant differences in productivity between the strains. *S. cerevisiae* produced only 6-7% of respective enzyme activities obtained for both secreting yeasts. This result is in agreement with previous studies,[7] where alternative hosts were also able to produce higher amounts of protein than *S. cerevisiae*; however, the low yield can also be a result of cytoplasmic expression where production of a protein is more limited than when protein is secreted.[81, 82] Unfortunately, no secreted lipase CAL-A could be obtained with the used plasmid when the standard secretion signal MFα for *S. cerevisiae* was applied. Both *K. lactis* and *H. polymorpha* reached high activities in the non-purified supernatants, with 30% higher volume activity measured for *H. polymorpha* (0.85 U mL^{-1}). However, *K. lactis* showed about 10% higher productivity per cell than *H. polymorpha* (71 nU $cell^{-1}$ compared to 65.3 nU $cell^{-1}$, respectively).

3.2.5 Glycosylation analysis

Since CAL-A originates from a eukaryotic organism it was supposed that the protein sequence contains glycosylation sites. The applied analysis focused on *N*-glycosylation, as it is the major modification in comparison to several

glycosides added during *O*-glycosylation. The sequence of CAL-A was analysed using NGlycPred software [83] and two potential *N*-glycosylation sites were found (N291 and N357), of which the first one was predicted to be glycosylated with a high probability. In fact, this residue was found to be glycosylated in CAL-A produced by *P. pastoris*[72] and one of the residues was probably also glycosylated in the original CAL-A purified from *C. antarctica.*[71]

As mentioned in section 3.2.4, the purified lipase CAL-A from the three yeasts was used for MALDI TOF analysis. These analysis aimed not only to confirm the production of the CAL-A, but were also planned to get more insight into the differences in host-specific modification between the samples. Unfortunately, analysis of these complex modifications failed in spite of extensive attempts at two facilities. Consequently, other methods had to be used to estimate the glycosylation degree of the produced protein.

As a qualitative measure, the glycosylation of the enzyme produced by both secreting yeasts was analysed *via* SDS-PAGE (Figure 11, p. 45). The secretion of CAL-A from *S. cerevisiae* failed; thus the intracellularly produced Sc CAL-A was not included in glycosylation analysis, as it cannot be glycosylated. Yeasts as heterologous hosts tend to overglycosylate heterologous proteins resulting in multifarious protein species, which can show a dispersed electrophoretic migration. Accordingly, smeared bands were observed for the native glycosylated proteins secreted from *H. polymorpha* and *K. lactis*. As shown in Figure 11.A, after deglycosylation with PNGase F, the same samples resulted in sharp bands, confirming the presence of *N*-glycosylation. It is important to notice, that in both deglycosylated samples not one, but actually two distinct bands for CAL-A are present, which can most probably be explained by presence of two forms of CAL-A: native enzyme and a shorter, truncated isoform. The occurrence of two isoforms of CAL-A produced in various hosts has been reported before.[80] Comparison of an *N*-deglycosylated sample with the respective *O*-deglycosylated sample of the Hp CAL-A, which is presented in Figure 11.B, indicates that the majority of the glycosylation originates from the *N*-glycans, as their removal results in a sharp band in the range of the non-glycosylated, His-tagged CAL-A. When *O*-deglycosylation was performed, no significant difference in band migration was observed as compared to the native sample.

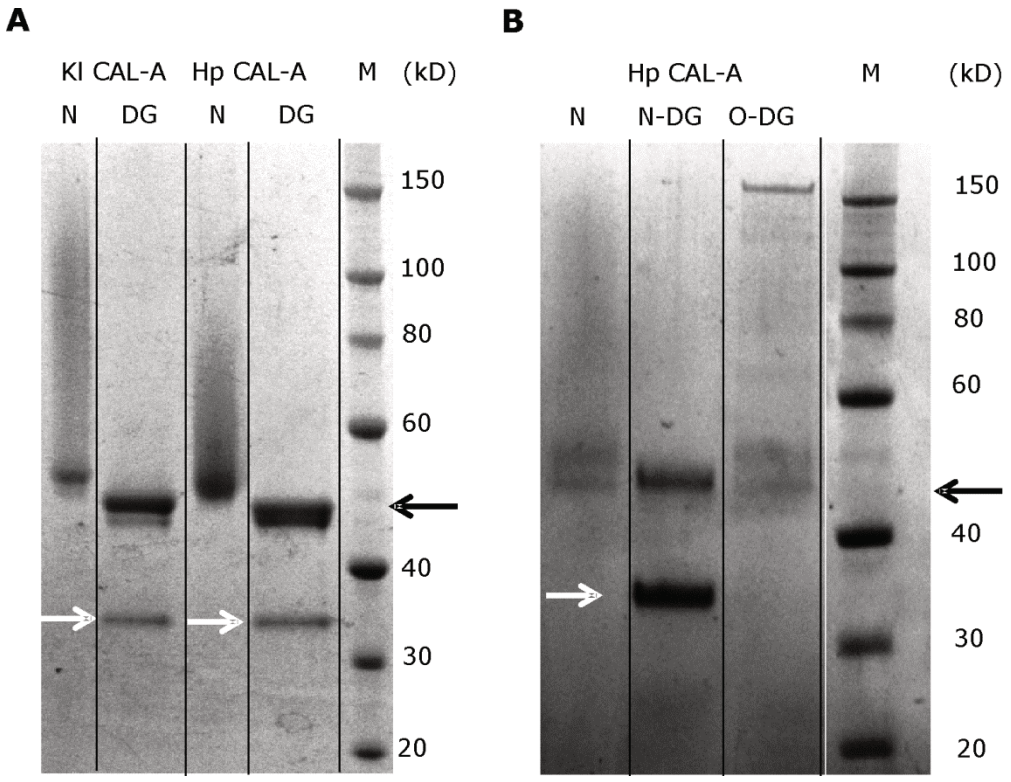


Figure 11: Glycosylation analysis of CAL-A purified from the secreting yeasts. A- per each secreting host, native samples are compared to samples deglycosylated with PNGase F according to the manual (New England Biolabs); B- comparison of native Hp CAL-A sample with *N*-deglycosylated (*N*-DG) and *O*-deglycosylated (*O*-DG) samples. The expected size of non-glycosylated CAL-A is indicated with black arrows. The published molecular weight of CAL-A is 45 kD,[70] whereas the calculated molecular weight including the His_6Tag is 48.8 kD (ProtParam tool[84]). M – marker RotiMark 10-150; Kl CAL-A – purified from *K. lactis* (3.2 µg per lane); Hp CAL-A – purified from *H. polymorpha* (3.4 µg per lane); N – native protein, N-DG/O-DG – the same sample after *N*-/*O*-deglycosylation. The additional band in the deglycosylated samples, indicated with white arrows, originates from PNGase F (36 kD).

3.2.5.1 Expression with tunicamycin

To further analyse the influence of glycosylation on the enzyme, expression with tunicamycin was performed. Tunicamycin belongs to a class of nucleoside antibiotics, was first identified in *Streptomyces lysosuperificus* and consists of an eleven-carbon aminodialdose moiety, known as tunicamine, present as a nucleoside of uracil, and a fatty acid of variable length, branching, and unsaturation (Figure 12). Tunicamycin inhibits the first enzyme of the *N*-glycosylation pathway, GlcNAc phosphototransferase, in eukaryotes by acting as a competitive inhibitor.[85]

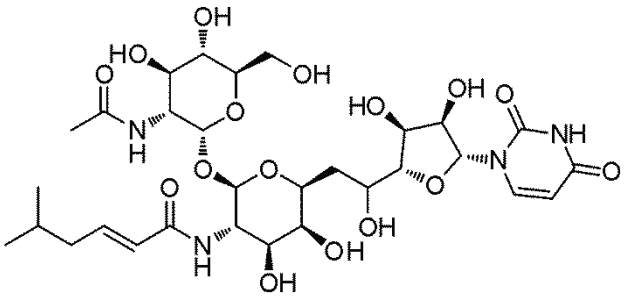


Figure 12: Structure of tunicamycin. The compound belongs to a family of related agents that vary in the length, branching and degree of unsaturation of the fatty acid amide linked to tunicamine.

The amount of tunicamycin typically needed for inhibition of glycosylation varies in different cells (0.1–10.0 µg mL^{-1}), presumably due to variable uptake and culture conditions, as well as differences in the level of expression of the phosphotransferase.[85] Since the optimal concentrations for both *K. lactis* and *H. polymorpha* were not published previously, two final concentrations of tunicamycin were used for the expression: 1 µg mL^{-1} and 5 µg mL^{-1}. As a control, a third culture without the addition of tunicamycin was used. The cultivation was performed as described in 5.5.10: 200 mL of the optimal medium for each yeast was used in 2 L baffled shake flasks.

Table 7: Measured growth rates in different yeast with and without addition of tunicamycin. Different concentrations of tunicamycin (0, 1 and 5 µg mL^{-1}) were used in the expression study and the optical density (OD_{600}) over the expression time was measured. The measurements were performed in duplicate. In both *S. cerevisiae* and *H. polymorpha*, addition of only 1 µg mL^{-1} had a tremendous effect on the initial growth rates, whereas in case of *K. lactis* only a slight decrease in cell density was observed even at 5 µg mL^{-1} of tunicamycin. The cultures were harvested after 65 h of cultivation.

Growth time (h)	***S. cerevisiae***			***K. lactis***			***H. polymorpha***			Yeast strain
	0	1	5	0	1	5	0	1	5	c tunicamycin (µg mL^{-1})
0	0.168	0.168	0.168	0.124	0.124	0.124	0.107	0.107	0.107	measured
2	0.188	0.179	0.185	0.155	0.146	0.148	0.234	0.185	0.177	OD_{600}
18	1.82	0.26	0.4	35.9	33.9	31.2	22.7	10.8	1.06	
65	196	64.5	22.5	73.5	32.5	15.7	31.5	42.6	27.3	

The growth rates were followed by optical density (OD_{600}) measurements (Table 7). The growth of each yeast was influenced by tunicamycin in a different way. After 18 h of expression *S. cerevisiae* was growing 7 times slower with 1 µg mL^{-1}

of tunicamycin than without the antibiotic. Addition of 5 µg mL^{-1} also severely impaired the growth of *S. cerevisiae* at this initial stage. *K. lactis* growth rates, on the other hand, were only slightly lower even in the presence of 5 µg mL^{-1} of tunicamycin. Initial growth of *H. polymorpha* after 18 h of expression was almost two times slower with 1 µg mL^{-1} of tunicamycin and more than 20 times slower with 5 µg mL^{-1} of tunicamycin.

In conclusion, directly after tunicamycin addition *S. cerevisiae* growth was the most sensitive to tunicamycin, even at the lowest concentration. The growth of *H. polymorpha* was significantly reduced with the addition of 5 µg mL^{-1} of tunicamycin, whereas the growth of *K. lactis* remained on a similar level even in the presence of a higher concentration of tunicamycin. These differences may be caused by variable uptake of tunicamycin or a different level of expression of the phosphotransferase. However, when the final optical density values at the time of harvest are compared, it can be concluded that for all tested yeasts a concentration of 5 µg mL^{-1} of tunicamycin caused significant growth impairment.
After 65 h of growth, cultures were harvested and activities were measured as described in section 5.7.7. Although the growth of *S. cerevisiae* was significantly slower with the addition of tunicamycin, the measured lipase activities in the crude extract were increased when 5 µg mL^{-1} of tunicamycin was applied (Figure 13). Although the impact on the growth of *K. lactis* was relatively low, even the lower concentration of tunicamycin, led to almost no secretion of CAL-A. The growth of *H. polymorpha* was gradually decreased with the increase of tunicamycin concentration, and a similar decrease of activity in the supernatant in the cultures was observed. The values for the comparative analysis were normalised to the OD_{600} and not to the cell number, as tunicamycin influences cell division,[86] therefore presumably altering the cell number per OD_{600} unit.

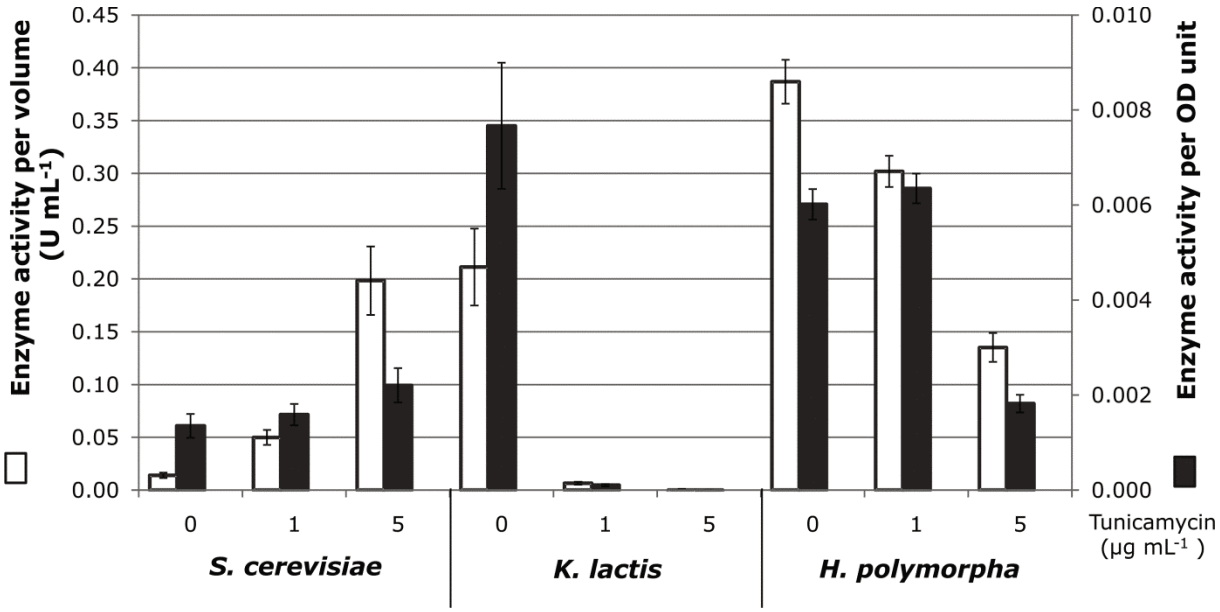


Figure 13: Influence of tunicamycin addition on the productivity of different strains. Enzyme activity per original culture volume (empty bars) and activity per OD_{600} unit of the respective culture (filled bars) were compared in crude *S. cerevisiae* extract or in spent culture medium for both secreting yeasts. All samples were taken at the stationary growth phase, when usually maximal CAL-A activity was obtained.

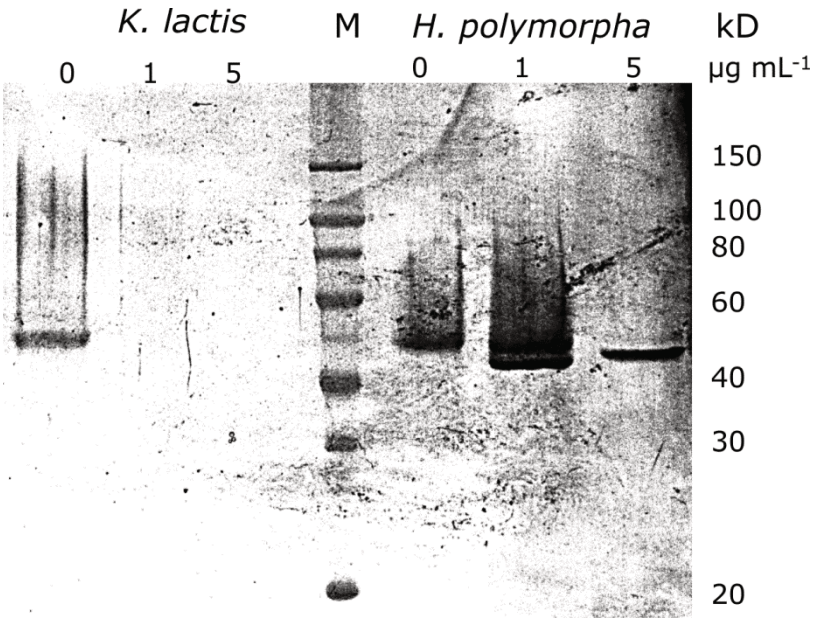


Figure 14: SDS-PAGE analysis of CAL-A obtained from both secreting yeasts in presence of various tunicamycin concentrations. Spent culture media were concentrated and used for IMAC purification; 10 µL of the obtained fractions were analysed on SDS-PAGE. For *K. lactis*, the purified protein concentration in samples with tunicamycin addition (1 and 5 µg mL^{-1}) was below detection. The protein was visualised with the silver staining method.

The samples were purified as described in section 5.7.3 and visualised on an SDS gel (Figure 14) as described in section 5.7.16. No protein could be obtained from

the *K. lactis* spent culture medium when tunicamycin was added; the measured protein concentration in all fractions after purification was below detection. For *H. polymorpha*, a gradual decrease of glycosylation with the increase of tunicamycin concentration was observed. In the control samples without tunicamycin, a band of glycosylated CAL-A could be seen. In the sample with 1 µg mL^{-1} of tunicamycin, two major bands can be seen, one for the glycosylated, and the other for the presumably non-glycosylated CAL-A. In the sample from the higher concentration of tunicamycin, only sharp bands of the non-glycosylated CAL-A was obtained.

Purification of glycosylated and under-glycosylated (tunicamycin treated) variants of the enzyme was successful only for *H. polymorpha* and thus these two samples were used for the investigation of the potential influence of glycosylation on the enzyme‘s properties. Both samples were therefore diluted to the same concentration and applied to thermostability and activity tests (see section 3.2.6).

3.2.5.2 Analysis of carbohydrate content

To get more insight into the differences between the glycoprotein secreted from the different yeasts, the total carbohydrate content was estimated by using the Glycoprotein Carbohydrate Estimation Kit (Pierce Biotechnology). The purified CAL-A obtained from the secreting yeasts was concentrated up to 0.25 mg mL^{-1} and then used for assay as described in section 5.7.12. In the assay, the carbohydrate groups were first oxidated for 10 minutes with sodium meta-periodate to produce detectable aldehyde groups. In the following 1 h, the reaction with the colorimetric Glycoprotein Detection Reagent occurred. Finally, the resulting purple reaction product was measured with a spectrophotometer at 550 nm and the values obtained for the tested proteins were compared to the standard proteins with known carbohydrate content: lysozyme bovine serum albumin, ovalbumin, apo-transferrin, fetuin, and α_1-acid glycoprotein. The final result was obtained through comparison to the proteins as presented in Figure 15.

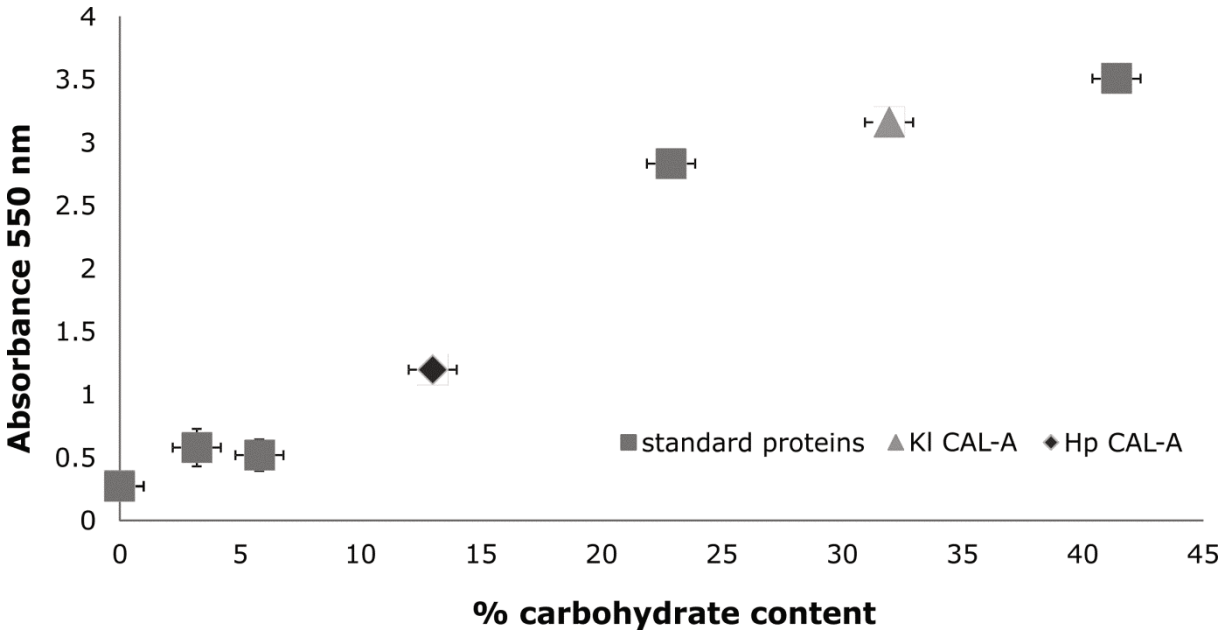


Figure 15: Carbohydrate content estimation of the secreted CAL-A. 0.25 mg mL^{-1} of the purified protein was used for the assay according to the manufacturer's protocol and the result was compared to the standard proteins with known carbohydrate content. When measured absorbance for a sample was higher than 1, the sample was measured again after 10×dilution to assure reliable results.

As a result, the carbohydrate content was estimated to be approximately 32% for Kl CAL-A and 13% for Hp CAL-A. This identified significant difference is, in fact, reflected in enzyme's characteristics, and will be presented in following sections. Unfortunately, no data on the carbohydrate content of the CAL-A produced in *P. pastoris* is available to date, therefore it cannot be compared.

3.2.6 Thermostability

CAL-A was purified as described in section 5.7.3 and tested by the standard photometric assay with *p*NP octanoate as the substrate (see section 5.7.7 for details). The analysis revealed a distinction in specific activity of the enzyme from different sources. The enzyme was either used directly for assay or stored in the 50 mM Sørensen buffer pH 8.0 at 4 °C; under these conditions it was stable for up to two months. For the assay, the samples from the three hosts were diluted to the same concentration (0.016 mg mL^{-1}) and each sample was then divided into 6 aliquots. Five of them were incubated at various temperatures for 30 minutes and the control one was stored at 4 °C during this time. The obtained results were then normalised to the activity of this control sample, which was taken as 100%.

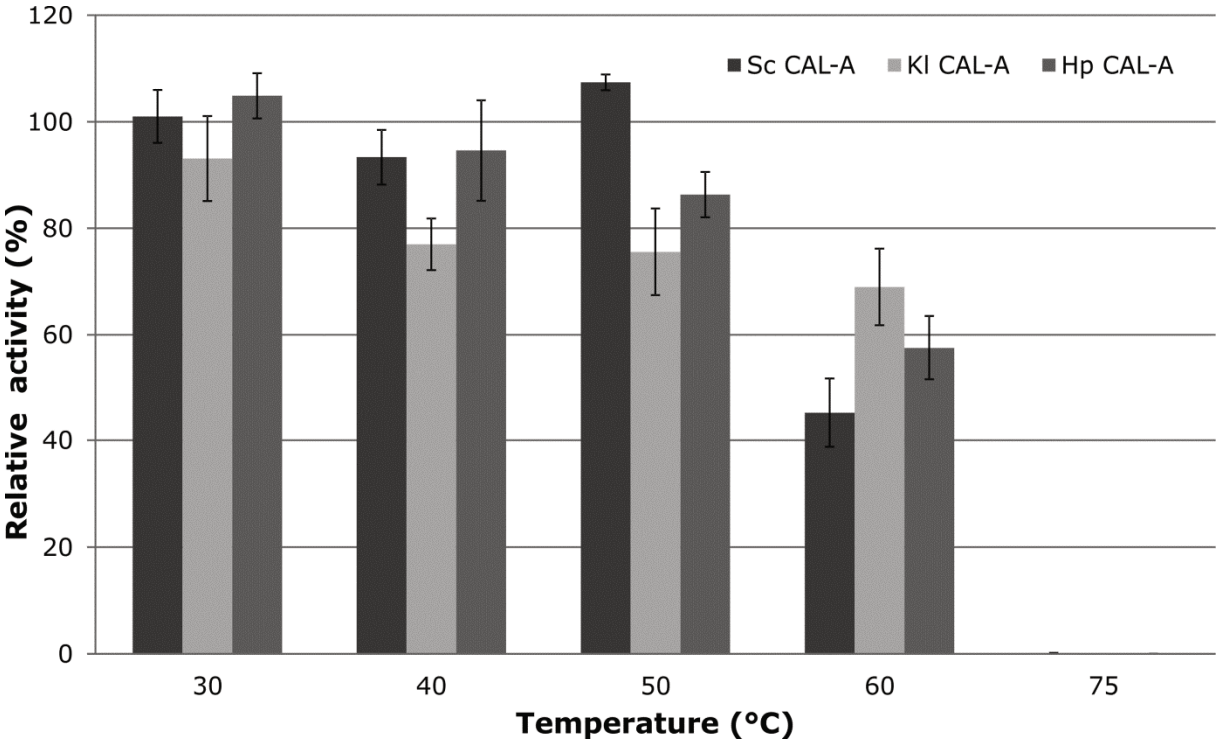


Figure 16: Thermostability of CAL-A purified from different hosts. Aliquots of enzyme solution (enzyme concentration 0.016 mg mL^{-1}) were incubated for 30 min at the respective temperatures and the residual activity was assayed by the standard assay. The measurement was done in triplicate and the average values are presented. The results were normalised to the activity of the unheated sample stored at 4 °C, which was taken as 100%. (0.1 U mL^{-1} for *S. cerevisiae*, 2.6 U mL^{-1} for *K. lactis*, 1.1 U mL^{-1} for *H. polymorpha*).

Interestingly, the temperature profiles (Figure 16) revealed significant differences in thermostability; although it is the same enzyme, stability in elevated temperatures differs between the samples. Sc CAL-A shows enhanced activity after incubation at 50 °C, but the values decrease very rapidly from this point on. On the other hand, Kl CAL-A shows a very slight loss of activity in the range between 40 °C and 60 °C. Hp CAL-A is stable at lower temperatures, and then loses its activity slower than *S. cerevisiae* but more rapidly than *K. lactis*. The differences for one temperature are as large as 30%. In comparison to the published results, the original CAL-A purified from *C. antarctica* (Ca CAL-A) retained 62% of its activity after incubation at 60 °C as measured in an olive oil-based assay[87] same as Hp CAL-A for *p*NP octanoate, whereas Kl CAL-A with 69% activity is more stable at this temperature. On the other hand, Ca CAL-A after incubation at 80 °C was still 29% stable, whereas all preparations obtained in this study were less than 1% active at 75 °C. The published data for CAL-A expressed in *P. pastoris* (Pp CAL-A) are not directly comparable as the activity in different temperatures was measured by another assay and without incubation

time.[69] However, under their conditions, Pp CAL-A showed a temperature optimum at 50 °C like Sc CAL-A; on the contrary, Pp CAL-A retained approx. 30% of its activity at 80 °C albeit without pre-incubation.

In conclusion, at temperatures higher than 60 °C CAL-A produced by *C. antarctica* appeared more stable than recombinant CAL-A; however, for applications up to 60 °C both *K. lactis* and *H. polymorpha* are suitable hosts.

The influence of glycosylation on the thermostability of the protein was then further investigated. As already presented, significant differences in the thermostability for the enzyme produced in different yeast were observed, and it was suggested that it could be a result of the differences in glycosylation. To test this further, two purified samples from *H. polymorpha*, from the culture without tunicamycin (marked as Hp CAL-A) and with 5 µg mL^{-1} tunicamycin (Hp CAL-A T), were used for the thermostability analysis.

The concentration of the protein sample has an influence on the thermostability of the enzyme; since the concentration from the previous experiment could not be achieved due to a limited amount of the purified protein, the comparison of Hp CAL-A and Hp CAL-A T was performed using exactly the same amount of the protein and both samples were harvested and purified on the same day to provide results as comparable as possible. The results are presented in Figure 17. Surprisingly, when the activities of the untreated samples were compared, which were then taken as 100%, the activity of Hp CAL-A T was only half of the activity of the glycosylated variant.

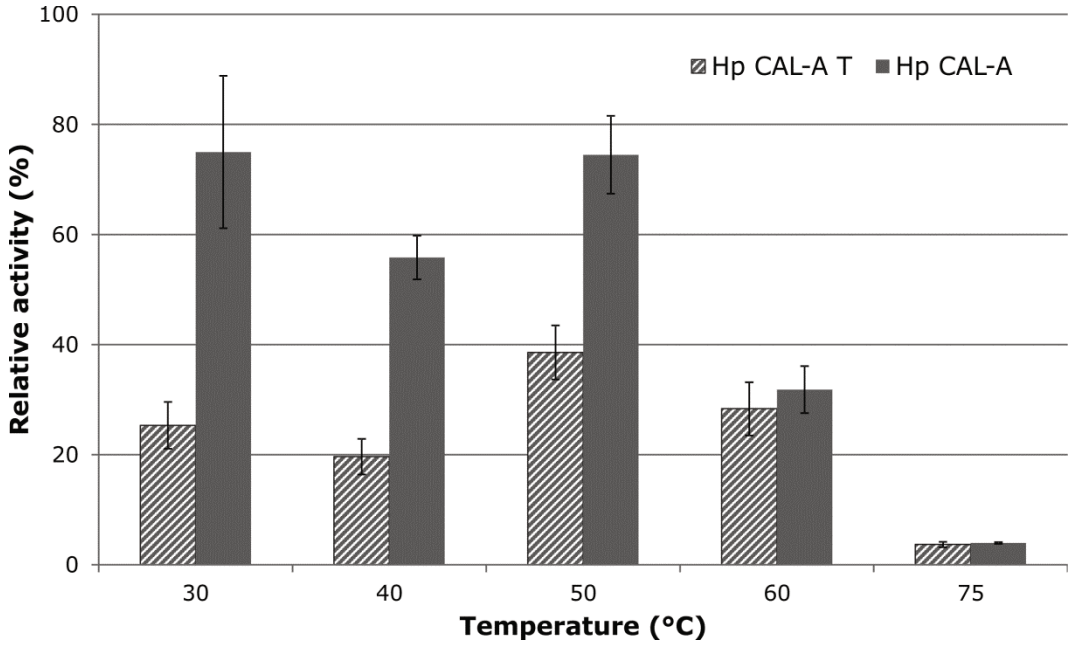


Figure 17: Thermostability of CAL-A produced in *H. polymorpha* in the presence (Hp CAL-A T, striped bars) or absence (Hp CAL-A, grey bars) of tunicamycin. 6.5 µg mL^{-1} of each enzyme sample (0.19 U mL^{-1} Hp CAL-A; 0.11 U mL^{-1} Hp CAL-A T) was incubated at various temperatures for 30 min and the residual activity was measured using the standard assay. The results were normalised to the activity of the unheated sample stored at 4 °C, which was taken as 100%. The measurement was done in triplicate and the average values are presented.

Significant differences in the thermostability can be seen already at the lowest temperature tested, 30 °C; the under-glycosylated variant of the enzyme (Hp CAL-A T) shows only 25% of the starting activity, whereas the glycosylated Hp CAL-A is still above 70%. The clear difference between the two samples can be observed up to 50 °C. At higher temperatures, where apparently denaturation of the Hp CAL-A increases, no further difference is visible. This result confirms the hypothesis that addition of sugar chains on the protein surface stabilises its structure, thus making it more resistant to higher temperatures. In case of CAL-A, this stabilising effect can be observed up to 50 °C.

3.2.7 Kinetic analysis

Kinetic parameters for the enzyme produced in different hosts were tested in a photometric assay with *p*NP octanoate **1** in various concentrations as the standard substrate. The kinetic constants were calculated from the Michaelis Menten equation with application of the least square fitting method.[88] In addition to the three samples from the tested hosts, a fourth sample of CAL-A (Hp CAL-A T), expressed in *H. polymorpha* with the presence of 5 µg mL^{-1}

tunicamycin and thus under-glycosylated, was included in the experiment to investigate the potential effects of *N*-glycosylation on the enzyme characteristics (Figure 18).

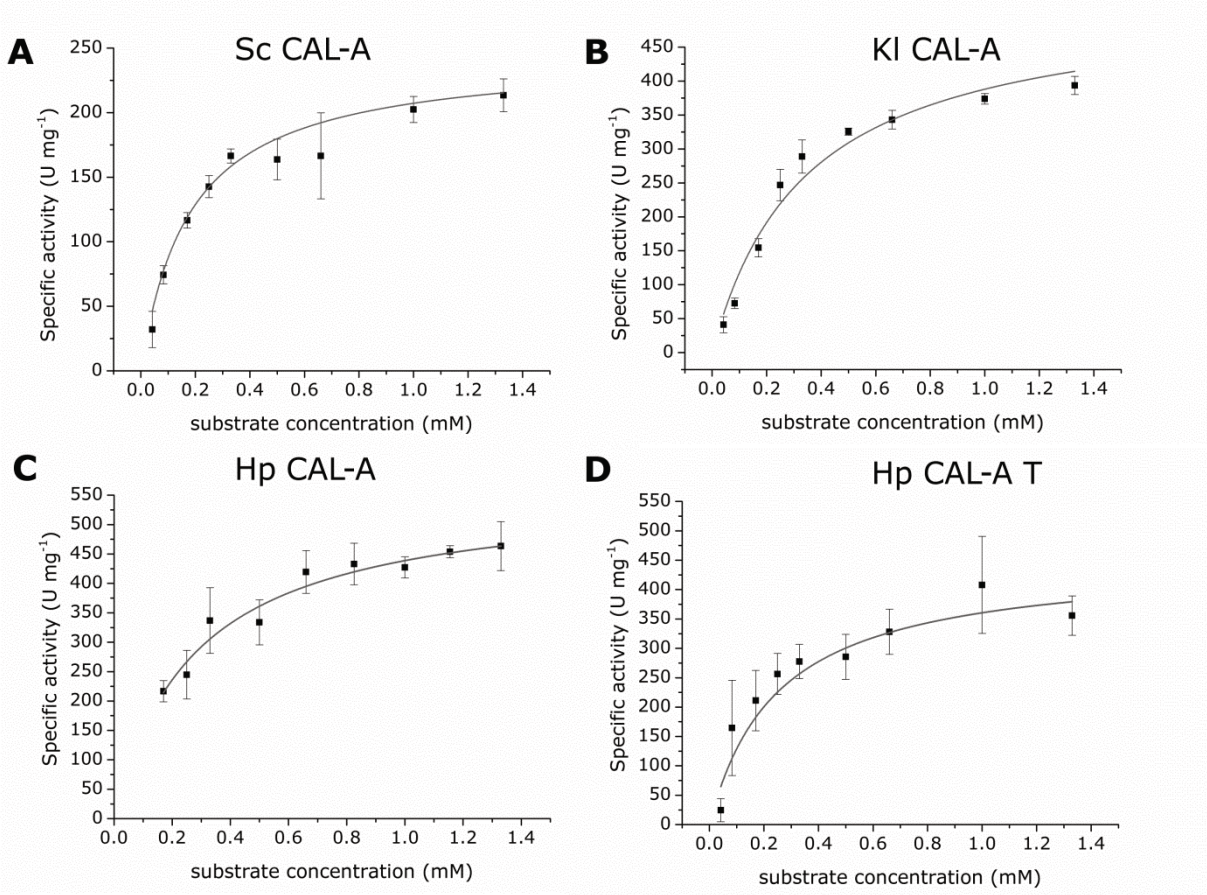


Figure 18: Steady state Michaelis Menten kinetic analysis of CAL-A expressed in various hosts. The activity of the enzyme was measured with different concentrations of the standard substrate *p*NP octanoate ester, in 50 mM Sørensen buffer pH 8.0. The calculated parameters are presented in Table 8.

Table 8: Comparison of the kinetic parameters for CAL-A lipase obtained with the *p*NP octanoate 1. The data were calculated using the least square fitting method from the obtained activities, using the Michaelis Menten equation. The concentration of the proteins for the k_{cat} value calculation was determined using the Bradford method with BSA as reference. Three independent experiments were carried out to measure the initial velocity of the reaction at various substrate concentrations. Reported values are the average of at least two independent determinations. All standard deviation values for k_{cat} and k_{cat}/ K_M were within 15%.

Sample	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/ K_M ($mM^{-1} s^{-1}$)
Sc CAL-A	0.18 ± 0,02	2.09×10^4	1.16×10^5
Kl CAL-A	0.34 ± 0,07	3.13×10^4	9.10×10^4
Hp CAL-A	0.27 ± 0,02	3.52×10^4	1.28×10^5
Hp CAL-A T	0.25 ± 0,08	2.54×10^4	1.02×10^5

As presented in Figure 18 and Table 8, small differences in K_M values were found. The lowest K_M constant was measured for Sc CAL-A and, interestingly, the highest for Kl CAL-A, however the differences are minimal. This is contrary to the results obtained for β-galactosidase BglS, where the variant produced in the presence of tunicamycin resulted in a lower K_M value than the glycosylated one.[60] Influence of glycosylation on the substrate affinity seems to be enzyme-dependent, since no similar effect was observed for the lipase CAL-B when produced in two hosts: *Pichia pastoris* and *Aspergillus niger*. Both variants were characterised with four *p*NP substrates and no significant differences for K_M and k_{cat} values were found.[89] Similarly, only minor differences for the k_{cat} parameter between the CAL-A variants produced in this thesis were observed.

The outcome of this kinetic characterisation shows that the host-dependent modification can have influence on the enzyme efficiency; however, the differences between the variants were rather small. Kl CAL-A showed a decreased catalytic efficiency with a k_{cat}/K_M value of 9.10 $\times 10^4$ mM^{-1} s^{-1} as compared to the other hosts: Sc CAL-A (1.16 $\times 10^5$ mM^{-1} s^{-1}) and Hp CAL-A (1.28 $\times 10^5$ mM^{-1} s^{-1}). There is no information on the kinetic constants for CAL-A produced in *C. antarctica* or *P. pastoris* available for the used substrate *p*NP octanoate. The kinetic constants of CAL-A produced in *P. pastoris* using tributyrin as a substrate were determined as K_M = 0.0045 mM, k_{cat} = 1.496×10^3 s^{-1} and k_{cat}/K_M = 3.34×10^5 mM^{-1} s^{-1}.[69]

3.2.8 Substrate spectra

Substrate scope and activity are important parameters for an enzyme's application in biocatalysis. The wild type CAL-A, produced in *P. pastoris,* has shown low activity and poor enantioselectivity towards bulky substrates like 2-(4-isobutylphenyl) propanoate ester (ibuprofen ester), which could be later overcome by reshaping of the substrate pocket.[74] Therefore, it was decided to test the activity of the CAL-A isoforms produced in the different hosts towards sterically demanding substrates to analyse a potential influence of the host.

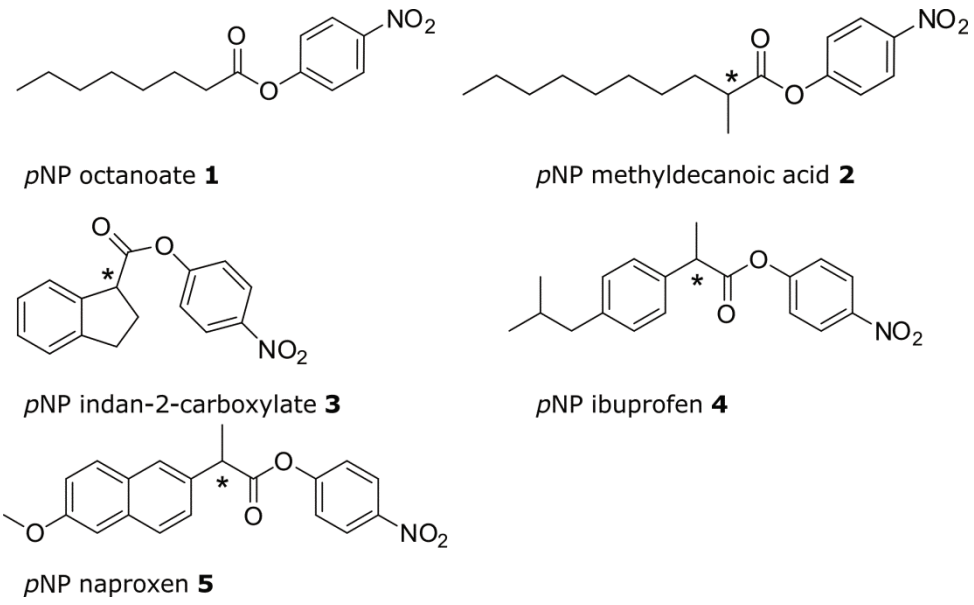


Figure 19: Substrates used for the substrate spectra screening of a produced lipase variants. The stereogenic centres are marked with asterisks.

As depicted in Figure 19, in addition to the standard substrate **1** the chiral substrates methyldecanoate **2,** indan-2-carboxylate **3**, ibuprofen ester **4,** and naproxen ester **5,** all coupled with 4-nitrophenol, were used for the spectrophotometric activity assay (Figure 20).

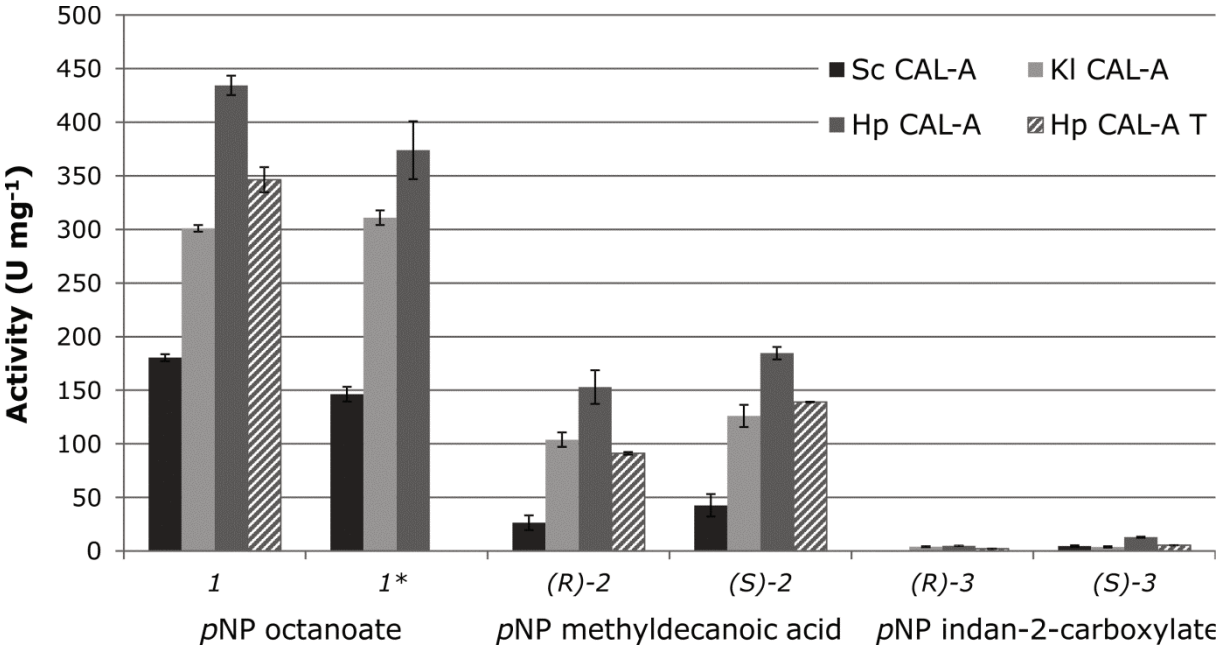


Figure 20: Specific activity of the CAL-A variants towards tested substrates. 30 µl of CAL-A solution purified from the different yeasts (i.e. ~3.3 mU Sc CAL-A, ~5.1 mU Kl CAL-A, ~6.9 mU Hp CAL-A, ~3.5 mU Hp CAL-A T) was tested with various substrates and the specific activity was measured. The measurement was done in triplicate and the average values are presented. * – assay with deglycosylated enzyme samples including PNGase F.

All CAL-A isoforms tested showed no detectable activity towards ibuprofen ester and naproxen ester, which is in agreement with the result published for unmodified Pp CAL-A;[74] these substrates are thus not included in the Figure 20. Significant differences between the enzyme isoforms purified from different hosts were found. The standard substrate **1** was converted with highest specific activity by Hp CAL-A (430 U mg^{-1}), and more than two times slower by Sc CAL-A (180 U mg^{-1}). Analysis towards the other substrates always indicated the highest activity for Hp CAL-A. Especially striking differences were observed for indan-2-carboxylate: Hp CAL-A (13 U mg^{-1}) was hydrolysing ester *(S)*-**3** four times more efficiently than both other enzyme variants, whereas its enantiomer *(R)*-**3** was not accepted as a substrate by Sc CAL-A at all, although both other preparations showed similar activity (~4.4 U mg^{-1}) towards it. A slight preference towards the *(S)*-enantiomer also appeared for the methyldecanoate esters, but in these cases both enantiomers (*R*)-**2** and (*S*)-**2** were accepted by all CAL-A isoforms. Compared to the published result of a methyldecanoate cleavage by Pp CAL-A (101 U mg^{-1}[69]) the obtained activity of Kl CAL-A was comparable, whereas Hp CAL-A was about 1.5-fold more efficient.

The significant differences observed in the specific CAL-A activities of the various hosts could be a result of host-specific effects. An alternative explanation would be the presumed co-occurrence of both a native and a truncated CAL-A isoform, produced in different ratios by various hosts. Both possibilities will be discussed in detail in the following section.

3.3 Summary and outlook

The lipase CAL-A was chosen as a model enzyme to establish three different yeast expression systems in the working group and to investigate their potential influence not only on the quantity, but also on the quality of the produced protein. It has already been proven in numerous comparative experiments that different microorganisms are able to produce the same protein with varying efficiency. Several studies indicate that glycosylation, which is a major covalent modification performed by yeasts, can contribute to several aspects of protein characteristics,[90-92] especially to thermostability.[93, 94] However, an influence of three industrially-relevant yeasts on one produced enzyme's properties, although possible, has to the best of my knowledge never been studied in a comparative way when based on one uniformly optimised synthetic gene. The presented results show significant differences not only in the production efficiency, but also in the thermostability and in the specific activity of the CAL-A isoforms purified from *S. cerevisiae*, *K. lactis*, or *H. polymorpha*.

Lipase CAL-A could not be obtained as a secreted enzyme from *S. cerevisiae*, whereas intracellular production resulted in active enzyme, albeit the obtained yield was low when compared to both other yeasts tested. The highest productivity within 72 h of cultivation was observed for *K. lactis*; however, taking into account cell number, *H. polymorpha* was slightly more effective. Obtained yields of the purified protein produced in *K. lactis* were comparable to the published values for *P. pastoris*. In case of *H. polymorpha*, purification difficulties led to a lower yield; media screening was applied to overcome this obstacle. As a result, a 10-fold improvement in the amount of the produced protein was observed as compared to the standard YNB medium. Additionally, the amount of the protein bound to the column during IMAC purification increased up to 55%. These results emphasise the significance of careful selection of the parameters to exploit the maximal potential of the host, which is especially important for a comparative study. Moreover, it is in agreement with published results for *H. polymorpha*, where the carbon source, as well as the cell culture stage, were shown to have an impact on the secreted protein glycosylation.[95]

When the enzyme variants were purified and applied for characterisation, the Kl CAL-A variant showed increased thermostability at 60 °C (69%) as compared to Hp CAL-A as well as to the published values for native Ca CAL-A

(~60%). The increased thermostability of the glycosylated variants corresponds well to the estimated carbohydrate content: the most thermostable variant Kl CAL-A is also the most glycosylated one with 32% estimated carbohydrate content. An effect of the glycosylation on the thermostability was further emphasised when glycosylated and under-glycosylated Hp CAL-A were compared. The under-glycosylated variant of the lipase has shown below 40% of the starting activity after incubation in temperatures as low as 30 °C, whereas in the glycosylated variant the addition of carbohydrates seemingly stabilises the enzyme structure, resulting in only minor loss of activity after incubation in temperatures up to 50 °C. Determination of kinetic constants revealed a slightly elevated K_M value for Kl CAL-A as compared to the other variants; however, only minor differences between the samples were found. Especially noteworthy was the result of substrate spectra screening: indan-2-carboxylate (*R*)-**3** was not accepted at all as a substrate by one of the variants (Sc CAL-A), while the other two were able to hydrolyse it. In addition, Hp CAL-A has shown the highest specific activity towards all substrates tested. Both *K. lactis* and *H. polymorpha* are thus suitable hosts for CAL-A expression, whereas *S. cerevisiae* is not recommended due to low yields.

There are two possible explanations for the observed differences between the same enzyme produced in different yeasts and the final result is most probably a cumulative effect of both of them. As already mentioned, one reason for the differences found between the CAL-A isoforms could be that the complex translation process introduces various posttranslational modifications like phosphorylation, amino acid modifications, and in particular different glycosylation patterns in a host-specific manner, thus influencing the enzyme's properties. Apart from the *N*-glycosylation, which was investigated in more detail in this study, there are 18 predicted potential phosphorylation sites and 12 potential *O*-glycosylation sites in the CAL-A protein sequence (as predicted using the NetPhosYeast tool[96] and the YinOYang 1.2 tool[66]), all together resulting in a plethora of possible produced variants. Obtained results on CAL-A thermostability show especially significant differences between glycosylated and under-glycosylated enzyme. Deglycosylation of the already produced enzyme variants resulted in only slight decrease of the activity. On the other hand, the specific activity of the under-glycosylated variant Hp CAL-A T was for some substrates significantly decreased, with the biggest difference being only half of the activity

obtained for Hp CAL-A in case of the ester *(S)*-**3**. These observations could be explained by the hypothesis that the carbohydrates added co- and post-translationally lead to formation of slightly different CAL-A isoforms with increased stability, even after deglycosylation of the active enzyme. Experiments with treatment of deglycosylated CAL-A with foldases[97] to obtain CAL-A variants without the influence of a glycan could be the next step to further investigate this theory. Alternatively, another enzyme could also be applied for the same procedure to confirm that the observed effect was not only protein-specific.

Another interesting explanation for the observed differences in the specific activities is based on the findings published by Liljeblad *et al.*[80] In this publication, commercially available CAL-A preparations were tested for their ability to catalyse the hydrolysis of amide bonds in certain compounds and, surprisingly, significant disparities in substrate preferences between them were found. The analysis on SDS-PAGE gel, followed by LC-MS/MS characterisation, revealed the presence of several impurities, but also of two variants of CAL-A: the native enzyme and a C-terminally truncated isoform, in which only the latter one hydrolysed one of the tested substrates. Analysis of the already published crystal structure revealed that the C-terminal region constitutes a motile antiparallel β-sheet structure which may function as an "active-site flap" and has been suggested to affect the substrate binding properties of the enzyme. The authors suggested that both isoforms were co-produced and – as a consequence – co-purified from the hosts tested. It is amenable that in the experiments of this thesis similar co-purification of the lipase from the tested hosts occurred. CAL-A was crystallised as unit of two monomers, which do not make extensive contact with each other.[72] However, the interaction is apparently strong enough to co-purify the truncated CAL-A form without the C-terminal His_6Tag. In fact, on the SDS-PAGE gel from the purified and deglycosylated CAL-A (Figure 11, p. 46), two bands for the protein can be observed in both Kl CAL-A and Hp CAL-A samples. As judged from the gel and the weak binding of active Hp CAL-A during IMAC purification, the amount of truncated CAL-A significantly increased from Kl CAL-A to Hp CAL-A – and these ratios are exactly reflected by the significant differences in specific activity towards various substrates (Figure 20, p. 57).[98]

In summary, the chosen eukaryotic heterologous expression host can indeed influence biochemical properties of the produced enzyme, even beyond typical posttranslational modifications like glycosylation. Since reproducibility of

results is a major issue in enzyme studies, now considering the host as another factor, variances in published values for specific activity - in spite of exactly the same assay and test substrate used - can be explained (an excellent summary of varying activity values measured for the same enzymes produced in different expression studies can be found in recent publication from Mellitzer *et al.*). [99] Especially eukaryotic proteins are often characterised by a number of co-existing isoforms due to factors including host-specific modifications, e.g., splicing, glycosylation, phosphorylation, or truncation. Since a fourfold increase in specific activity of the model enzyme CAL-A expressed in different yeasts was observed, changing the production host to improve the enzyme's properties is worth consideration and may be a simple first step for a time and cost-effective optimisation towards biocatalytic applications and processes. In certain cases it may be even possible to find a host expressing an enzyme isoform with slightly improved chemical activity, as it was shown in this study when Hp CAL-A was used instead of Sc CAL-A for ester (*R*)-**3**.

As an outlook, it is proposed that the productivity should not be considered as the only criterion for the selection of a suitable expression host; a comparison of several expression hosts can result in not only higher productivity, but also in improvement of the enzyme's characteristics. It is worth consideration especially for enzymes designed for industrial applications, where even a small improvement in the enzyme's properties, e.g., thermostability or activity, can have a huge impact on the whole process cost-effectiveness.

4. Glycosidases & glycosynthases

4.1 Introduction

This part of the thesis focused on the application of the yeast expression systems for expression and modification of glycosyl hydrolases and glycosynthases as prerequisite for their potential application in the synthesis of glycoside containing natural products. A lesser known class of enzymes — rhamnosidases — was targeted, as this family of glycoside hydrolases has to the best of my knowledge never been applied for any kind of protein engineering, yet selective addition of rhamnose would be important for synthesis of numerous natural compounds. Since difficulties with effective expression of such enzymes were already a known problem in the working group, application of the eukaryotic yeast hosts was planned to overcome this obstacle and obtain the enzyme in good yield for further experiments. In addition, the ultimate goal was to investigate the possibility to modify a rhamnosidase into a rhamnosynthase, thus allowing the application of the enzyme in biocatalysis. Three published rhamnosidases, originating from prokaryotic and eukaryotic sources, were selected for this investigation. As a model enzyme, rhamnosidase from *Bacillus* sp. (Rha_{Ba}) was chosen, as this is the best characterised rhamnosidase currently available; in addition, it can be overexpressed in large quantities in *E. coli*. Two other rhamnosidases selected were less studied, but therefore interesting subjects of investigation; a rhamnosidase from *H. polymorpha*, Rha_{Hp}, for which an encouraging characteristic was published (high activity and selectivity towards the α-L-rhamnopyranoside configuration), yet no sequence was available, and a rhamnosidase from *L. acidophilus*, Rha_{La}, which has shown high activity with a range of flavonoid substrates. For both of them, expression in eukaryotic hosts should result in improved yield. Rha_{Hp} was purified in active form from the original host; therefore it should presumably be produced in *H. polymorpha* with good yields. Rha_{La} was reported to be produced in *E. coli*, albeit with low yield and purification difficulties, which was planned to be circumvented by expression in yeasts.

In the following subsections, the current state of knowledge is summarised, whereas the results and discussion are presented in section 4.2.

4.1.1 Glycosynthases in biocatalysis

Glycosides are common building blocks in nature; they can serve as structural biopolymers in cellulose, but can also bear the information in biological processes such as cell-cell recognition, cell signalling or ligand-receptor binding. Clinical administration of these compounds has been shown to have potential to mitigate some symptoms in the treatment of cancer,[100] Alzheimer's[101] and Parkinson's diseases.[102] In fact, most of the bioactive natural products contain glycosyl groups, which are often crucial for their activity. However, the major obstacle in applying these compounds is a low yield in isolation of naturally occurring substances. Classical chemical synthesis offers no easy alternative, as additional protection and deprotections steps are required, which — together with the demand for high regio- and stereoselective formation of the glycosidic bond — can lead to low yields.

A solution for this problem may be the application of enzymes for glycosidic bond formation. A significant development in this field has occurred in recent years,[103, 104] nonetheless there is still room for improvements. As an example, L-menthyl-α-glucopyranoside is a desirable derivative of L-menthol, with possible applications in mouthwash production, as it is slowly hydrolysed in the mouth, resulting in a prolonged sensation of freshness.[105] Application of lyophilised cells containing α-glycosidase allowed a one-step anomer-selective synthesis reaction of the compound with 99% yield, such that 202 mg of the product were obtained after 48 h of reaction (Figure 21). [106]

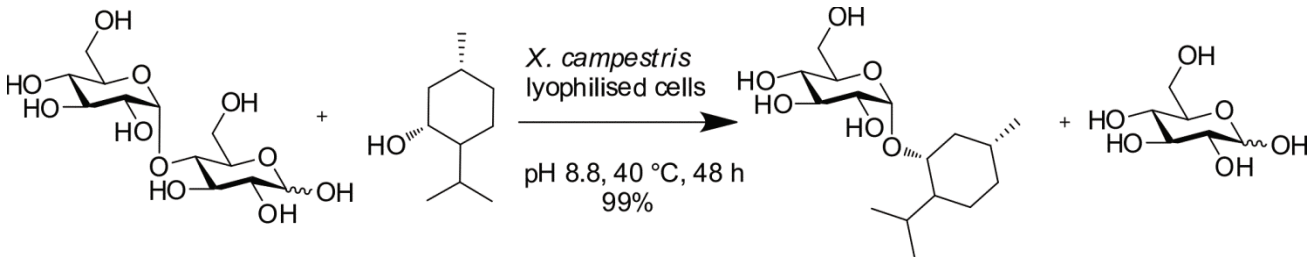


Figure 21: α-L-glycosylation of menthol catalysed by α-glycosidase from lyophilised *Xanthomonas campestris* cells. An excellent yield of 99% was obtained under optimal conditions.[106] The reaction was proven to occur as a result of transglycosylation in the presence of 1.0 M maltose per 100 mg of L-menthol.

There are two types of enzymes capable of glycosidic bond formation: glycosyl transferases (GT) [EC 2.4.-.-] and glycosyl hydrolases (GH) [EC 3.2.1.-]. Unfortunately, application of GTs is limited by poor availability, as well as by very expensive and often unstable sugar donor substrates required for their activity.

Therefore, this thesis focused on GHs, which are abundant in nature and are usually applied for the hydrolysis of *exo*- or *endo*-sugar moieties from a wide range of natural products. Analyses based on data from genome projects have shown that approximately 1-3 % of an organism's genes are dedicated to carbohydrate degradation, thus indicating the importance of glycosyl hydrolases.[107] Interestingly, the reverse reaction performed under kinetic control and in the presence of high concentrations of acceptors different from water allows obtaining regiomers in moderate yields (30-40%) using simple glycosides as donors.[108, 109]

To overcome this limitation, the glycosynthase approach was introduced in 1998 by Withers and colleagues,[14] which allows obtaining glycosides in quantitative yields.[103] It is achieved by redesign of glycosidases and blocking their ability to hydrolyse the substrates. As an effect, the enzyme becomes a glycosynthase (GS): it is no longer able to hydrolyse the substrate, but it preserves the ability to transfer sugars from activated donors.[110] In this way, good to excellent yields can be obtained and often the product is synthesised with high enantio- and regioselectivity. As an example, β-galactosidase BgaC from *Bacillus circulans* was successfully modified to a galactosynthase, showing transglycosylation activity with high regioselectivity for the formation of β-(1,3)-galactosidic bonds and excellent yields up to 98%.[111]

Glycosynthases have thus enormous potential for the production of glycan analogs of natural products. They have been already applied for the synthesis of as challenging compounds as glycosphingolipids, a class of therapeutically valuable compounds that have been extremely difficult to access on a larger scale through chemical and chemoenzymatic methods, by the group of Whiters. In the first step of their approach, an *endo*-glycoceramidase from *Rhodococcus* sp. was modified into a GS by introduction of single mutation E351S. The modified enzyme was then used to synthesise D-*erythro*-sphingosine derivatives and excellent >90% yields were obtained; in addition, it also accepted several donors (e.g., lactoside, cellobioside, 3'-sialyllactoside) as well as acceptors (various sphingosine analogues), proving its flexibility in the preparation of both natural and unnatural glycosphingolipids.[112]

4.1.1.1 Mechanism of glycosidase engineering

Carbohydrate active enzymes are classified in the online database, CAZy (www.cazy.org[113]) into families based on the amino acid sequence similarities. Glycoside hydrolases (GH) are divided into 133 families, which are then further grouped in clans (GH-A till GH-F), based on the conservation of the protein fold. They can be also divided into two classes, inverting and retaining enzymes, based on the changes of the anomeric configuration of the substrate/product during the reaction (Figure 22).

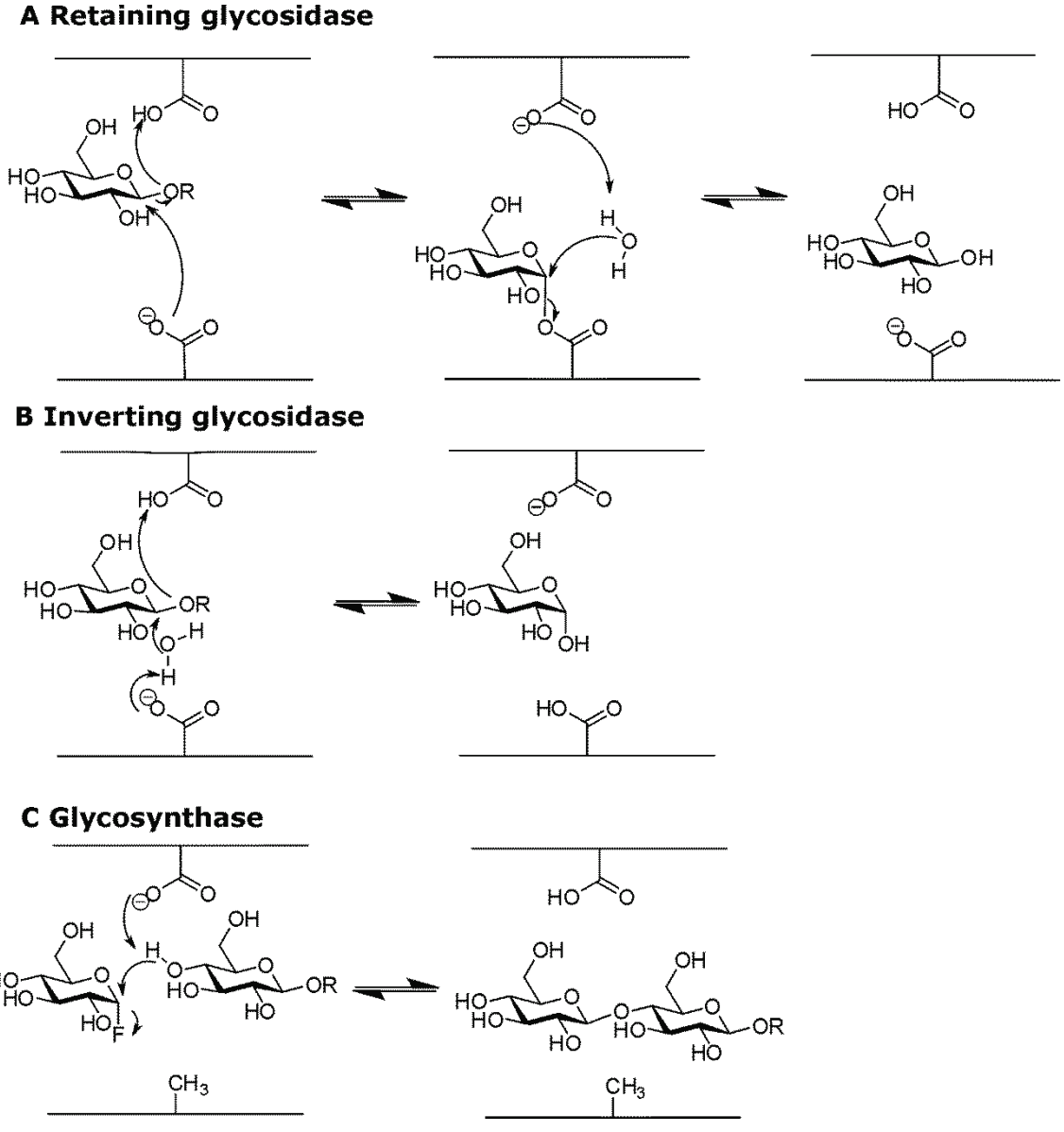


Figure 22: An illustration of the catalytic mechanisms for (a) retaining glycosidases, (b) inverting glycosidases, (c) glycosynthases. A glycosynthase, in which the catalytic nucleophile was mutated to Ala, uses fluoride of the opposite anomeric configuration as a donor. Abbreviation: R- aglycon (e.g., *p*-nitrophenyl). Adapted from Hancock *et al.*, 2006.[110]

In retaining GHs, which follow a reaction mechanism that was first proposed by Koshland in 1953,[114] one residue acts as a nucleophile and the other as an acid/base catalyst. The reaction follows the double displacement mechanism and the covalent substrate-enzyme intermediate is formed. In the next step, hydrolysis of the intermediate occurs. The configuration of the anomeric carbon is thus retained in the product as compared to the substrate. In inverting GHs, the reaction takes place *via* direct attack of nucleophilic water, activated by the general base residue, on the glycosidic bond; the anomeric configuration is therefore inverted. Glycosynthases (GS) are mutants of GHs, in which the catalytic nucleophile (Asp or Glu) at the active site is replaced by a non-nucleophilic residue (usually Ala, Ser or Gly). Through this mutation, the enzymes have lost their hydrolytic activity, but are still able to transfer the glycosyl group on specific acceptors. The first published example of a glycosynthase was β-glucosidase from *Agrobacterium* sp. (Abg1) in which the single introduced mutation Glu358Ala let to a catalytically inactive mutant with preserved glycan transfer ability.[14] Since then, numerous modified enzymes have been published as summarised in Table 9.

As presented there, some of the glycosidases are able to use thioacceptors after mutation introduction, thus bearing thioglycosynthase activity. It allows production of thioglycosides, in which the glycosidic oxygen atom has been replaced by a sulphur atom, which are thus resistant towards enzymatic hydrolysis. These compounds have proven to be useful as competitive glycosyl hydrolase inhibitors and therefore potential therapeutics. Thioglycosides have already found application in X-ray crystallography, where they form stable complexes with the investigated glycosidases.[115] There are only few naturally occurring thioglycosides in plants and their application for the synthesis of thioglycosidic linkages in oligosaccharides *in vitro* is still limited. Therefore, application of the glycosynthase approach to obtain thioglycosides is of great importance for further investigation and application of these compounds.[116]

Table 9: Currently available glycosynthases.*

Nr	Parent glycosidase	Organism	GH family	Catalytic activity	Altered residue	Linkage synthesised	Ref.
1	β-glucosidase (Abg1)	*Agrobacterium* sp.	GH1	Glycosynthase	E359	β-(1,3/4)	[14]
				Thioglycoligase	E171	β-(1,4)	[116]
				Thioglycosynthase	E359 E171	β-(1,4)	[117]
2	β-glucosidase	*Sulfolobus solfataricus*	GH1	Glycosynthase	E387	β-(1,3/4/6)	[118]
3	β-glycosidase	*Thermosphaera aggregans*	GH1	Glycosynthase	E134	β-(1,3/4/6)	[119]
4	β-glucosidase (CelB)	*Pyrococcus furiosus*	GH1	Glycosynthase	E372	β-(1,3)	[119]
5	β-glycosidase	*Thermus thermophilus*	GH1	Glycosynthase	E338	β-(1,3)	[120]
6	β-glycosidase	*Oryza sativa*	GH1	Glycosynthase	E414	β-(1,4)	[121]
7	β-glycosidase	*Pyrococcus horikoshii*	GH1	Glycosynthase	E324	β-(1,3)	[122]
8	β-mannosidase (Man2A)	*Cellulomonas fimi*	GH2	Glycosynthase	E519	β-(1,3/4)	[123]
				Thioglycoligase	E429	β-(1,4)	[116]
9	β-galactosidase (LacZ)	*Escherichia coli*	GH2	Glycosynthase	E537	β-(1,6)	[124]
10	β-glucuronidase	*Escherichia coli*	GH2	Glycosynthase	E504	β-(1,1)	[125]
11	β-glucuronidase	*Thermotoga maritima*	GH2	Glycosynthase	E476	β-(1,4)	[126]
				Thioglycoligase	E383		[126]
12	endoglyco-ceramidase II	*Rhodococcus* sp.	GH5	Glycosynthase	E351	β-(1,1)	[112]
13	endo-β-1,4-glucanase	*Clostriduim cellulolyticum*	GH5	Glycosynthase	E307	β-(1,3/4)	[127]
14	xyloglucanase	*Paenibalillus pabuli*	GH5	Glycosynthase	E323	β-(1,4)	[128]
15	cellulase	*Humicola insolens*	GH7	Glycosynthase	E197	β-(1,4)	[1]
16	oligoxylanase	*Bacillus halodurans*	GH8	Glycosynthase	D263	β-(1,4)	[129]
17	endo-1,4-β-xylanase	*Cellulomonas fimi*	GH10	Glycosynthase	E235	β-(1,4)	[130]
18	xylanase	*Thermotoga maritima*	GH10	Glycosynthase	E259	β-(1,4)	[131]
19	xylanase	*Bacillus halodurans*	GH10	Glycosynthase	E301	β-(1,4)	[90]
20	xylanase	*Clostridium sterorarium*	GH10	Glycosynthase	E293	β-(1,4)	[90]
21	xylanase	*Cellulomonas fimi*	GH10	Glycosynthase	E233	β-(1,4)	[90]
22	xylanase (Bcx)	*Bacillus circulans*	GH11	Thioglycoligase	E172	β-(1,4)	[132]
23	xyloglucanase	*Bacillus licheniformis*	GH12	Glycosynthase	E155	not reported	[133]

Table 9: Currently available glycosynthases.* (continued)

24	1,3-1,4-β-glucanase	*Bacillus licheniformis*	GH16	Glycosynthase	E143	β-(1,3/4)	[134]
25	laminarinase (LamA)	*Pyrococcus furiosus*	GH16	Glycosynthase	E170	β-(1,3/4)	[135]
26	xyloglucan endotransferase 1-xyloglucanase	*Tropaeolum majus*	GH16	Glycosynthase	E94	β-(1,4)	[133]
27	xyloglucan endotrans-glycosylase	*Populus tremula x tremuloides*	GH16	Glycosynthase	E85	β-(1,4)	[136], [137]
28	laminarinase (Lam16 A)	*Phanerochaete chrysosporium*	GH16	Glycosynthase	E115	cyclic β-(1,3)	[138]
29	glucan endo-1,3-β-D-glucosidase	*Hordeum vulgare*	GH17	Glycosynthase	E231	β-(1,3)	[139]
30	mannanase (Man26A)	*Cellvibrio japonicus*	GH26	Glycosynthase	E320	β-(1,4)	[140]
31	α-L-fucosidase	*Sulfolobus solfataricus*	GH29	Glycosynthase	D242	α-(1,3)	[141]
32	α-L-fucosidase	*Thermotoga maritima*	GH29	Glycosynthase	D224	α-(1,3)	[141]
33	α-glucosidase	*Schizosaccharomyces pombe*	GH31	Glycosynthase	D481	α-(1,4)	[142]
34	α-xylosidase (YicI)	*Escherichia coli*	GH31	Thioglycoligase	D482	α-(1,4/6)	[143]
35	α-glucosidase	*Sulfolobus solfataricus*	GH31	Thioglycoligase	D416	α-(1,4)	[143]
36	α-mannosidase	*Sulfolobus solfataricus*	GH31	Glycosynthase	D320	α-(1,3)	[144]
37	β-galactosidase (BgaX)	*Xanthomonas manihotis*	GH35	Thioglycoligase	G184	β-(1,3/4)	[145]
38	β-galactosidase (BgaC)	*Bacillus circulans*	GH35	Glycosynthase	E233	β-(1,3)	[111]
39	α-galactosidase	*Thermotoga maritima*	GH36	Glycosynthase	D327	α-(1,2/3/4)	[146]
40	β-xylosidase (XynB2)	*Geobacillus stearothermophilus*	GH52	Glycosynthase	E335	α-(1,3/4)	[147]
41	endo-β-N-acetyl glucosaminidase (Endo-A)	*Arthrobacter protophormie*	GH85	Glycosynthase	N171	β-(1,4)	[148]
42	endo-β-N-acetyl glucosaminidase (Endo-M)	*Mucor hiemalis*	GH85	Glycosynthase	N175	β-(1,4)	[148]
43	1,2-α-l-fucosidase (AfcA)	*Bifidobacterium bifidum*	GH95	Glycosynthase	D766	α-(1,2)	[149]

GH- glycosyl hydrolase;*- the presented data cover the period until June 2014

As it can be concluded from Table 9, different types of glycosynthases have been described in several CAZy GH families (GH1, GH2, GH5, GH7, GH8, GH10, GH11, GH12, GH16, GH17, GH26, GH29, GH31, GH35, GH36, GH52, GH85, GH95), which still represent a minority of the 133 families listed so far (June 2014). The vast majority of the published glycosynthases are β-glycosynthases, which are able to form the glycosidic bond from a wide selection of glycosyl donors. On the contrary, α-glycosynthases reported so far are restricted to glucose (GH31), fucose (GH29 and GH95), mannose (GH31), xylose (GH52) and galactose (GH36) transfer.

Until now, mostly retaining enzymes have been used to create glycosynthases; replacement of the nucleophile residue to the inert one has been the most often used modification. Application of a glycosyl fluoride of the opposite anomeric configuration to the natural substrate mimics the intermediate step of the reaction. The first GS from an inverting GH was GH8 *exo*-oligoxylanase from *Bacillus halodurans*, reported in 2006.[129] Saturation mutagenesis was performed on the nucleophilic residue D263 and nine amino acids mutants resulted in transglycosylation activity, with the D263C and D263N variants resulting in highest productivity. In fact, the other variants were also able to perform transglycosylation, but the product was directly hydrolysed. As concluded in this publication, the mechanisms of acquisition of glycosynthase activity from the retaining and inverting enzymes are different, as for the retaining enzymes the modification of the catalytic base completely deactivates the hydrolytic activity of the enzyme. In the case of the inverting enzymes, the water molecule, which is activated by the base residue, attacks the C1 of the glycoside; therefore, the mutation of the base residue cannot completely remove the hydrolytic activity, as the water molecule can still access the substrate even without the aid of the catalytic residue. Therefore, when modifying an inverting enzyme, it is important to minimise the hydrolytic activity of the new variants.

New approaches are constantly being developed, including more stable substrates, e. g., azides,[141] and optimisation of the screening procedure,[150] which is one of the major drawbacks. An alternative strategy for generating a glycosynthase from an inverting enzyme involves the replacement of a residue that holds a catalytic water molecule with a non-polar residue.[151]

However, some questions still remain unanswered. It is for instance impossible to predict which amino acid will be the best replacement for the catalytic

nucleophile and some enzymes failed to be a successful glycosynthase for yet unknown reasons.[152]

4.1.1.2 Substrates and products

The published glycosynthases use different types of donor substrates, with glycosyl fluorides being the most popular. Their main advantage is wide applicability; however, they are quite unstable and usually not commercially available. Recently, application of glycosyl azides has been proposed as an interesting alternative for α-GS, since the β-glycosyl fluorides have a short half-life time and thus are problematic substrates for slowly-reacting α-glycosynthases.[141] In general, the specificity for the donor reflects that of the parental GH. There exist GS with wide substrate specificity (e.g., GH family 1, 2, 7) as well as GS which are very specific for the donor. Acceptor specificity is less predictable and must be investigated for each enzyme separately. It was also shown that for certain enzymes directed evolution approaches can expand the specificity toward the acceptor.[153]

As can be deduced from Table 9, most currently available GS possess high regiospecificity, whereas some of them are able to synthesise several types of linkages. Surprisingly, in certain cases the acceptor molecule was shown to alter the regiospecificity of the enzyme. For example, for the GS from GH1, the small change from a 2- to 4-nitropenol group in the glycoside acceptor resulted in altered specificity from β-(1,3) to β-(1,4).[154]

4.1.2 Rhamnosidases

α-L-Rhamnosidase [E. C. 3.2.1.40] is an *exo*-glycosidase that cleaves terminal α-L-rhamnose from a wide range of natural products, including hespedrin, naringin, diosgene and terpenyl glycosides. It occurs naturally in bacteria, fungi and plants, as well as in animal tissues.[15] Rhamnosidases are structurally classified into GH28, GH78 and GH106 families according to the CAZY database[113] and thus belong to the inverting enzymes. The rhamnosidases have found applications in industry mostly in debittering of citrus fruit juices, enhancement of wine aroma and to some extent in rhamnose production.[15] They were also demonstrated to be useful in preparation of the anti-

inflammatory flavonoid prunin, as well as in prodrug therapy, where the drugs of interest are capped with rhamnose and activated by rhamnosidase.[155]

Table 10: Overview of the known rhamnosidases* and their selected properties.

Nr	Organism	Method of protein production	UniProt accesion no.	pH opt.	T opt. (°C)	Ref.
1	*Acrostalagmus luteoalbus*	isolated from organism	-	8.0	55	[156]
2	*Aspergillus aculeatus*	isolated from organism	Q9HFW5	4.5-5.0	-	[157]
3	*Aspergillus flavus*	isolated from organism	-	6.5	50	[158]
4	*Aspergillus kawachii*	isolated from organism	A9ZT55	4.0	50	[159]
5	*Aspergillus nidulans*	isolated from organism	-	4.5-6.0	60	[160]
6	*Aspergillus niger*	purified from commertial preparation	-	4.5	65	[161]
7	*Aspergillus terreus*	isolated from organism	-	5.5	60	[162]
8	*Bacillus sp.GL1*	produced in *E. coli*, crystallised	Q93RE7	7.0	50	[163]
9	*Bacteroides JY-6*	isolated from organism	-	7.0	-	[164]
10	*Clostridium stercorarium*	produced in *E. coli*	Q9S3L0	7.5	60	[165]
11	*Corticium rolfsii*	isolated from organism	-	2.0	-	[166]
12	*Emericella nidulans*	isolated from organism	-	4.5-6.0	60	[160]
13	*Fagopyrum esculentum*	isolated from organism	-	-	-	[167]
14	*Fusarium sambucinum*	isolated from organism	-	5.5-6.5	55-60	[158]
15	*Fusarium solani*	isolated from organism	-	6.5	-	[158]
16	*Fusobacterium sp.*	isolated from organism	-	5.5	-	[168]
17	*Lactobacillus acidophilus*	produced in *E. coli*	C2HQP9	6.0	-	[16]
18	*Lactobacillus plantarum*	produced in *E. coli*	C4PG47	7.0	50	[169]
19	*Lactobacillus plantarum*	produced in *E. coli*	C4PG45	5.0	60	[169]

Table 10: Overview of the known rhamnosidases* and their selected properties (continued)

20	*Mucor racemosus*	isolated from organism	-	5.5-6.5	55-60	[158]
21	*Pediococcus acidilactici*	produced in *E. coli*	E0NEV1	5.5	50	[170]
22	*Pediococcus acidilactici*	produced in *E. coli*	E0NEK0	4.5	70	[170]
23	*Penicillium aureatiogriseum*	isolated from organism	-	-	60	[158]
24	*Pichia angusta*	isolated from organism	-	6.0	40	[171]
25	*Pichia guilliermondii*	isolated from organism	-	6.0	-	[172]
26	*Pseudoalteromonas sp.*	isolated from organism	-	6.0	40	[173]
27	*Pseudomonas paucimobilis*	isolated from organism	-	7.8	45	[174]
28	*Ralstonia pickettii*	isolated from organism	-	-	-	[173]
29	*Rhizopus nigricans*	isolated from organism	-	6.5	60-80	[175]
30	*Sphingomonas paucimobilis*	produced in *E. coli*	Q76LC4	-	-	[176]
31	*Streptomyces avermitilis*	produced in *E. coli*, crystallised	Q82PP4	6.0	50	[177]
32	*Sus scrofa f. domestica*	isolated from organism	-	7.0	42	[178]
33	*Thermomicrobia bacterium PRI-1686*	produced in *E. coli*	Q6RCI9	7.9	70	[179]
34	*Thermomicrobia bacterium PRI-1686*	produced in *E. coli*	Q6RCI8	6.2	70	[179]
35	*Trichoderma longibrachiatum*	isolated from organism	-	4.5-5.5	60	[158]
36	*Turbo cornutus*	isolated from organism	-	2.8	-	[180]

* the presented data cover the period until June 2014

The initial difficulty with assaying the enzyme activity was overcome in 1985 by application of the commercially available substrate *p*NP rhamnopyranoside,[181] which enables a one-step photometric assay. The specificities of the enzymes were studied and mostly α-1,2 (naringin) and α-1,6 (hesperidin, rutin) types of linkages are hydrolysed, although some enzymes have been reported to also hydrolyse α-1,3 (gellan) or α-1,4 (saikosaponin C) linkages.[15] Interestingly, two rhamnosidases from *Pediococcus acidilactici* (Ram and Ram2) were unable to

hydrolyse naringin, while Ram2 displayed high specificity toward hesperidin and rutinose.[170] On the other hand, the rhamnosidase from *Aspergillus kawachii* was able to efficiently hydrolyse both hespedrin and naringin.[159] Thus, more intensive investigations of the reaction mechanism are required.

The rhamnosidases are still far from being well studied; to date (June 2014), 36 rhamnosidases have been characterised, albeit mostly enzymes purified from culture filtrates were used for this purpose (Table 10). Only thirteen rhamnosidase protein sequences are published and three crystal structures, all from bacterial sources, are available to date (including one yet unpublished structure of a putative rhamnosidase).[182]

In conclusion, there is a scientific need to crystallise α-L-rhamnosidase from other sources as well as to further investigate rhamnosidases for new applications in biotech industry.[155] These aims were addressed in this study. In particular selective addition of rhamnose would be important for synthesis of numerous natural compounds. Thus the ultimate goal was to investigate the possibility to modify a rhamnosidase into a rhamnosynthase, enabling the new application of the enzyme in biocatalysis. Three published rhamnosidases, originating from prokaryotic and eukaryotic sources, were selected for this investigation. Three published rhamnosidases, originating from prokaryotic and eukaryotic sources, were selected for this investigation. As a model enzyme, rhamnosidase from *Bacillus* sp. (Rha_{Ba}) was chosen, as this is the best characterised rhamnosidase currently available; in addition, it can be overexpressed in large quantities in *E. coli*. Two other selected rhamnosidases were less studied, but therefore interesting subjects of investigation; a rhamnosidase from *H. polymorpha*, Rha_{Hp}, for which an encouraging characteristic was published (high activity and selectivity towards the α-L-rhamnopyranoside configuration), yet no sequence was available, and a rhamnosidase from *L. acidophilus*, Rha_{La}, which has shown high activity with a range of flavonoid substrates. Most important features of the three selected rhamnosidases are summarised in the following subsections.

4.1.2.1 Rhamnosidase Rha_{Ba}

This rhamnosidase was first reported in 2003; the gene was isolated from the bacterium *Bacillus* sp. strain GL1.[183] This strain was known for its ability to assimilate gellan, the heteropolysaccharide, which is depolymerised by

successive reactions of gellan lyase, glucuronyl hydrolase, β-D-glucosidase and α-L-rhamnosidase. The identified rhamnosidase gene was isolated, cloned and overexpressed in *E. coli,* resulting in an excellent protein yield of 335 mg L^{-1}. The molecular mass of Rha_{Ba} was determined to be 106 kD; the enzyme was most active at a pH range of 6.5-7.0 in KP_i buffer and 40 °C.[183] The rhamnosidase has shown high specificity for rhamnosyl saccharides; only minor activities were observed for *p*NP-β-D-fucopyranoside (2.9%) and *p*NP β-D-galactopyranoside (1.6%) when the activity towards *p*NP α-L-rhamnopyranoside was taken as 100%. The enzyme also released rhamnose from both naringin and gellan disaccharide.

The rhamnosidase was the first enzyme from GH78 family with a known crystal structure, which was published in 2007 (PDB code 2OKX) (Figure 23).[163]

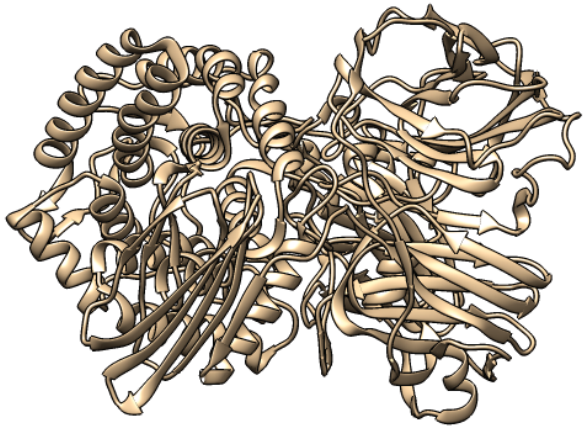


Figure 23: Crystal structure of the rhamnosidase from *Bacillus* sp. The structure is deposited at PDB under code 2OKX.

The structure of this enzyme was solved at 1.9 Å resolution using single-wavelength anomalous diffraction (SAD) analysis with incorporated selenomethionine. Additionally, the structure of rhamnose bound Rha_{Ba} was solved at 2.1 Å resolution. The overall structure of Rha_{Ba} is divided into five distinct domains, of which four (N, D1, D2, C) are β-strand abundant domains and one (A) is purely α-helical. Interestingly, two bound calcium ions were found in the structure; they were located far from rhamnose-binding site. Since it was already shown that activity of that enzyme is independent of any divalent cations,[183] the authors concluded that these ions are probably essential for the protein crystallisation.

4.1.2.2 Rhamnosidase Rha_{La}

A putative rhamnosidase gene $RamA_{La}$ from *Lactobacillus acidophilus* was identified and heterologously expressed in *E. coli* in 2009 by Beekwilder *et al.*[16] The sequence analysis revealed that this rhamnosidase is an ortholog of rhamnosidase Rha_{Ba}. In consequence, the mutations identified to be crucial for altering Rha_{Ba} activity would presumably have similar effect on Rha_{La} activity.

The protein was produced as fusion to an N-terminal His_6tag, but the IMAC purification failed for unknown reasons, thus the enzyme was applied for characterisation as a crude extract *E. coli* preparation. Unfortunately, no information on expression efficiency or on protein's molecular weight is given in the publication; calculated molecular mass is 107.75 kD, which is very similar to Rha_{Ba} (106 kD).

The enzyme was reported to be stable at pH range 5.0-7.0 and has shown activity with a range of flavonoid substrates: rutin (quercetin-3-glucoside-6-β-1-rhamnoside), nicotiflorin (kaemferol-3-glucoside-6-β-1-rhamnoside), narirutin (naringenin-7-glucoside-6-β-1-rhamnoside) and naringin (naringenin-7-glucoside-2-β-1-rhamnoside), yielding at least 80% conversion for each of them.[16] Thus, the rhamnosidase hydrolysed both α-1,2 and α-1,6 types of linkages.

4.1.2.3 Rhamnosidase Rha_{Hp}

The rhamnosidase Rha_{Hp} was first reported in 2000 by Yanai and Sato[171] as an intracellular enzyme, which, after purification from the native host *H. polymorpha*, was shown to possess high activity towards several tested substrates of industrial interest. The molecular mass of the native α-L-rhamnosidase was estimated to be around 90 kD; the enzyme was stable from pH 5.0 to 7.0 with optimal activity at pH 6.0 and 40 °C. The enzyme was tested towards a wide range of *p*-nitrophenyl glycosides and has shown high selectivity towards the α-L-rhamnopyranoside configuration,[171] liberating rhamnose from naringin, rutin, hespedrin and 3-quercitrin.

Since the authors considered this rhamnosidase potentially useful for application in wine aroma improvement, its inhibition by enological relevant compounds was analysed. The enzyme had 76% of its activity in the presence of 500 mM glucose, which is a typical concentration for grape juice. Moreover, at the typical

12% (v/v) ethanol concentration used in wine it showed only 20% inhibition. In addition, the enzyme efficiently released monoterpenols such as linalool and geraniol from an aroma precursor extracted from Muscat grape juice. Thus, the authors concluded that this rhamnosidase may be useful in many industrial applications.
In the original publication, no enzyme sequence was elucidated, only the first 17 amino acids from the N-terminal end of the protein were determined. Therefore, one of the first tasks of this study was to identify and clone the gene.

4.1.3 Site-directed mutagenesis of the enzymes

As already explained in previous sections, introduction of predefined modifications in the enzyme gene sequence is one of the basic tools of molecular biology, essential in functional study, genetic engineering, biochemistry and protein engineering. It was also applied in this study to create rhamnosidase mutants with altered activity. Mutagenesis of a protein at a specific position was first introduced in a report in 1978;[3] since then numerous approaches have been developed, which are discussed in more detail below. The basic procedure is based on the polymerase chain reaction (PCR) method and requires the synthesis of a short DNA primer which contains the desired mutation and is complementary to the template DNA around the mutation site. As a result, it can hybridise with the template DNA. In the next step, the primer is extended using a DNA polymerase, which completes the rest of the gene. Thus, the new version of the gene contains the mutated site. Finally, the product is analysed *via* DNA sequencing to check if the desired mutation was properly introduced. The mutation may be a single base change (a point mutation), multiple base changes, deletion or insertion. In this study, both point mutations and deletions of a part of the gene (intron in Rha_{Hp} gene) were performed using the described methods. Detailed descriptions of the used methods can be found in section 5.6.7.

In the **overlap extension PCR** method, large insertions or deletions are possible. The protocol consists of three PCR reactions; in the first PCR both fragments of the gene are amplified separately with the primers containing complementary ends. As a result, an overlap fragment between both amplicons is created, which is then used in the PCR 2 reaction to assemble them. In a PCR 3, the new created DNA with the introduced change is further amplified using the

outer primers from the PCR 1. The efficiency of this method depends on the sequence of the primers, which is strictly defined by the gene sequence. High AT content or repetitions may be a reason for failure. The accuracy of the method also depends on the polymerase error rate.

Cassette mutagenesis is an efficient method for the introduction of multiple modifications in a defined DNA fragment without using PCR. In this method restriction sites flanking the regions of interest are required. The flanking restriction sites in the plasmid are the digested with the cognate restriction enzymes and short synthetic pairs of complementary oligonucleotides, containing the mutation in the gene of interest, are inserted and ligated.[184] This method can generate mutants with high efficiency and is PCR independent; it may also be used for big fragments. On the other hand, cassette mutagenesis is limited by the availability of suitable restriction sites flanking the site of the interest, as their introduction *via* silent mutagenesis is not always possible.

QuikChange Site-Directed Mutagenesis is an alternative mutagenesis method, especially suitable for introduction of point mutations, which was developed by the company Stratagene (La Jolla, CA) and was published in 1996.[185] In this approach, a pair of complementary mutagenic primers is used to amplify the entire plasmid in the thermocycling reaction using a high-fidelity non-strand-displacing DNA polymerase (e.g., *Pfu* or *Phusion* polymerase). The reaction generates a nicked, circular DNA. In contrast to the other methods, in this approach only the template DNA is amplified, resulting in lower yields of the product. The template DNA is eliminated in the following step by the enzymatic digestion with the restriction enzyme *DpnI*, which is specific for methylated template DNA. As a result, only PCR-amplified plasmid with the introduced mutation is obtained.[185] The main advantages of this method are simplicity and high efficiency. Additionally, usage of the online tools dedicated for the design of the primer pair for this site-directed mutagenesis (e. g., PrimerX[186]) further facilitates this method. However, this method is restricted to primer pairs of 25–45 bases in length with melting temperature (T_M) ≥ 78 °C. When this requirement is not fulfilled, primer dimer formation will eventually become more favourable compared to the primer-template annealing.

As a modification of the Quikchange method, the **Megaprimer method** for introduction of large insertions or deletions was developed in 2001 by Geiser *et al.*[187] In this approach, the fragment of DNA to be inserted is amplified with

overhangs that are complementary to the template DNA. After purification, the fragment is used as a megaprimer in the PCR reaction together with the high-fidelity DNA polymerase and the template DNA. As in the standard Quikchange protocol, the template DNA is eliminated in the following step by enzymatic digestion with *DpnI*. As reported by Geiser *et al.*, insertions of fragments up to 1 kb are possible with the method.[187]

In general, the selection of the most suitable method depends on many factors, including gene sequence, available restriction sites close to the targeted region, size of the plasmid as well as size of the fragment to be inserted/deleted. In addition, usually a selection of conditions (e.g., T_M, various polymerases, buffer) needs to be tested prior to obtaining satisfactory results.

4.2 Results and discussion

4.2.1 Rhamnosidase from Bacillus sp.

4.2.1.1 Design of potential rhamnosynthase mutants

One of the main goals of this study was to modify the activity of a rhamnosidase to a glycosynthase. As introduced in section 4.1.1.1, the classical mutation of a glycosidase to a glycosynthase is performed *via* replacement of the catalytic nucleophile (Asp or Glu) at the active site by a non-nucleophilic residue (usually Ala, Ser or Gly). Unfortunately, in the published structure of the rhamnosidase Rha_{Ba},[163] no clear pair of catalytic residues could be identified, although the structure was solved in the presence and the absence of rhamnose in the catalytic pocket. Usually for the inverting enzymes the proper distance between donor and acceptor is ca. 10 Å but there are no amino acid residues in Rha_{Ba} which would fulfil this requirement (Figure 24).

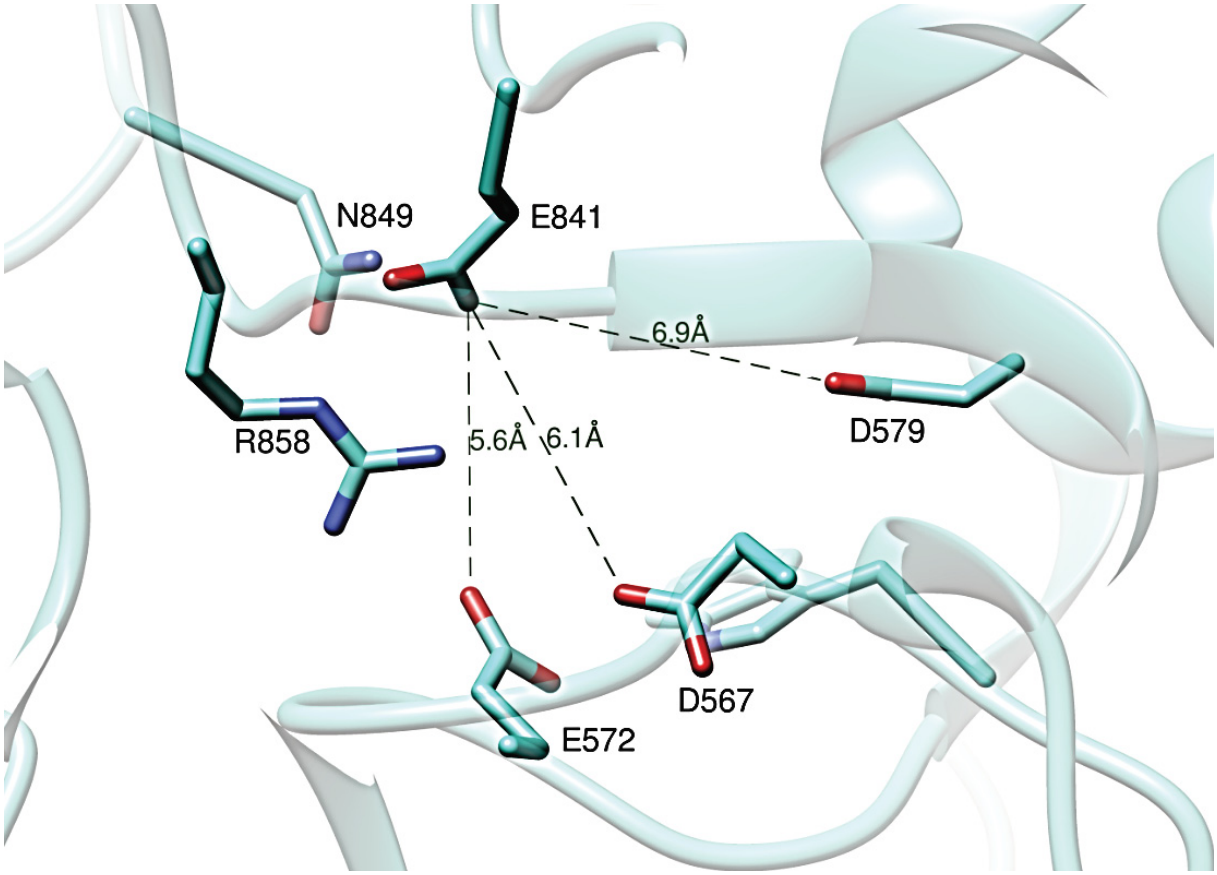


Figure 24: Active cleft of rhamnosidase Rha_{Ba} and the distances between potential catalytic amino acid residues. Although in the inverting enzymes usually two catalytic residues are located ca. 10 Å apart, the distances obtained from the crystal structure are smaller. The structure is deposited in PDB under code 2OKX.

When the crystal structure of rhamnosidase Rha_{Ba} was reported in 2007,[163] E841 was identified as a potential general base through an alignment with other

published GH, but the other presumably catalytic amino acid residues - D567, E572 - are located on the same side of the active pocket and the distances between them are short (3.2-6.9 Å). As concluded in the publication, this may suggest a new inverting mechanism, especially since the acidic residues are arranged in a conservative way among GH78 enzymes, as elucidated from the alignment with rhamnosidases from *Clostridium stercorarium* and *Thermomicrobia bacterium*. This hypothesis was further confirmed *via* a mutagenesis experiment: when the mutants Rha_{Ba} D567N, E572Q and E841Q were tested, a dramatic decrease of the hydrolytic activity toward *p*NP rhamnopyranoside could be observed.[163]

As presented in Figure 25, when the crystal structure of the rhamnosidase from *Streptomyces avermitilis* was solved in 2013,[188] structural similarities between the catalytic domains in both enzymes classified into the GH78 family could be identified. In both structures, the catalytic domain is located in the central part of the protein and surrounded by several additional domains - five in Rha_{Sa} and four in Rha_{Ba}. The amino acid residues in the active pocket are organised in analogous order and corresponding residues for the three catalytic amino acids identified in Rha_{Ba} were found in Rha_{Sa} (Rha_{Ba} - Rha_{Sa} corresponding amino acid residues: D567 - D630; E572 - E636; E841 - E895, see the superimposition of the residues in Figure 25 B). Consequently, it is possible that a mutation inverting the activity towards GS in one of the enzymes would result in the same effect for the other one.

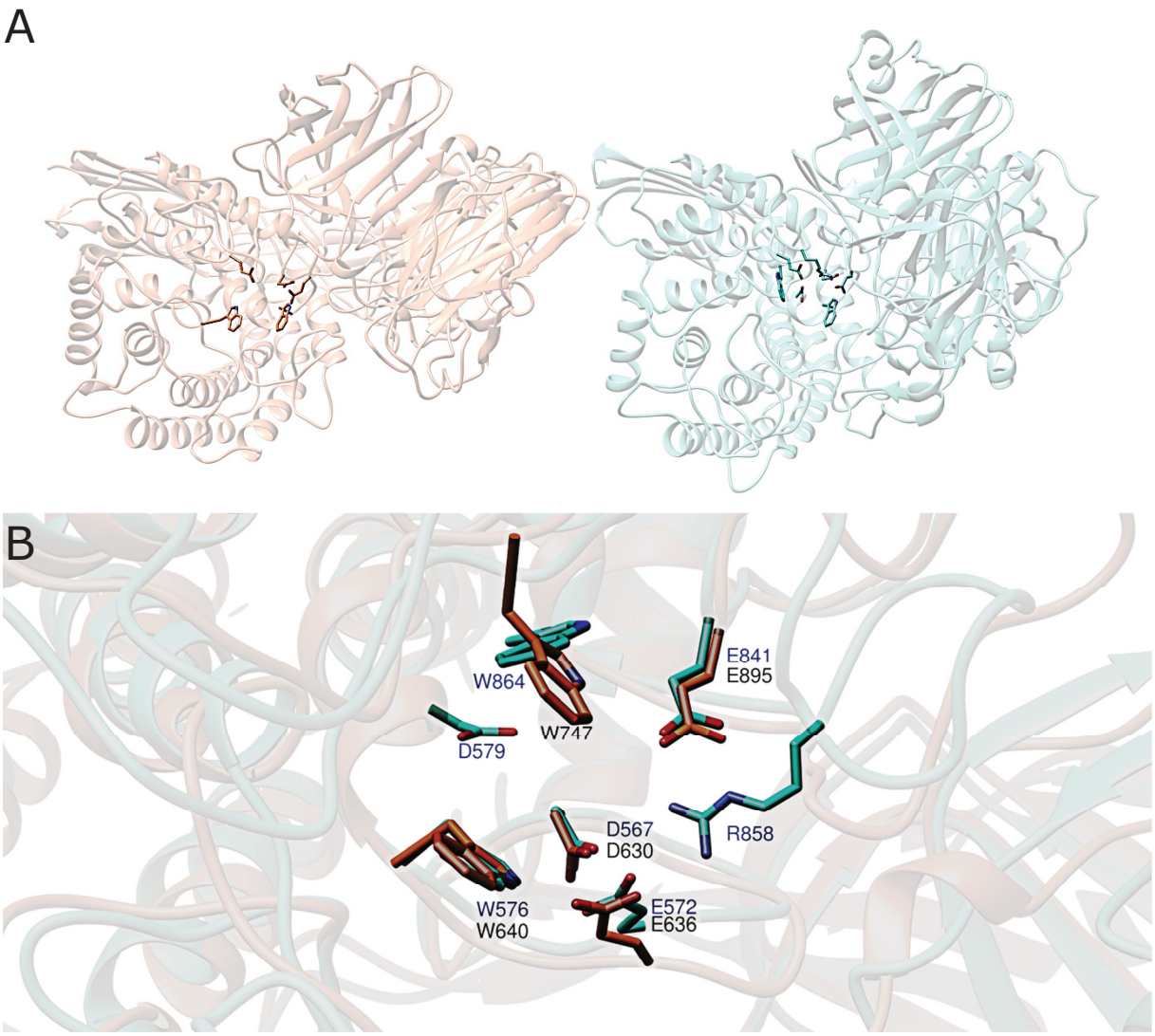


Figure 25: Comparison of the two published rhamnosidases structures- Rha_{Sa} from *S. avermitilis* (coral, PDB code 3W5M) and Rha_{Ba} from *Bacillus* sp. GL1 (turquoise, PDB code 2OKX). A- Localisation of the catalytic domain in the overall enzyme structure, B- superimposition of the identified amino acids located in the substrate pocket (blue identifier- Rha_{Ba}, black identifier- Rha_{Sa}).

These promising observations were the basis for the selection of the rhamnosidase Rha_{Ba} for the mutagenesis experiment. As already mentioned in the introduction, due to the available structural information it was the model enzyme to be mutated into the rhamnosidase.

Since three GH78 rhamnosidases were selected for the study, as explained in detail in section 4.1, an alignment of their sequences was performed to verify if the amino acid residues identified in the active pocket of Rha_{Ba} are also conserved in the two others (Figure 26). As it was presented at the beginning of this section, significant structural similarities in the active pocket organisation were found between two GH78 rhamnosidases with published crystal structures.

However, no crystal structures for Rha_{La} and Rha_{Hp} are available to date; in fact, the sequence of the rhamnosidase Rha_{Hp} was first identified in this study. Therefore, it was important to analyse, if the similarities on the sequence level can be identified, which would be a prerequisite for the application of both Rha_{La} and Rha_{Hp} for the mutagenesis. As can be seen in Figure 26, the crucial amino acid residues, as well as the surrounding regions, are indeed quite well conserved for the rhamnosidases used for the alignment. Consequently, these three residues were chosen to be mutated.

```
                                                          567   572
RhaBa       PVKLHEIYIRQSTYPVAEQGSFRCSDALLNATWEISRHTTRLCMEDTFVDCPSYE-QVFW     576
RhaLa       VNK-DDFESWVLYSDLKQTGFIKTNNDKVNRLFKNVIWGQKSNFMDVPTDCPQRDERLGW     448
RhaHp       EFSPNNIEFVVIGNLMETTGGFECSDKMLNRLHQNVVHSMRGNFLAIPMDCPQRDERLGW     474
              . .::        :   * :. .:  :*   :      :  :     ***. : :: *

RhaBa       VGDSRNEALVNYYVFGETEIVERCLNLVP-GSADETPLYLDQVPSAWSSVIPNWTFFWIL     635
RhaLa       TGDAEIFAPTASFNMNTYEFYKKYAKDMLVEQEDNKGM----LPIIVPSLKQK------S     498
RhaHp       TGDIAIFCPTALYLCDCYSFLQSWLKDLKLEQKEKGGA----PAVVVPDIIHAFDTFWNG     530
            .**    . .  :     .: :   : :   . ::            .:

RhaBa       AC-------------REYAAHTGNEAFAARIWPAVKHTLTHYLEHIDDSGLLNMAGWNLL     682
RhaLa       TGMAIWSDAATIIPWVTYR-FFDDLGVLKQNYSQM-KNWVDWITQNTKTKYLWIGQMQLG     556
RhaHp       KIAAIWQDASVIVPYELYR-SSGNKHILQAQYESM-VSWIDYIPKIEGKVRWNKIELQLG     588
                       *     :  .    :  :  .  .:: :   .        :*

RhaBa       DWAPIDQPNEGIVT-HQNLFLVKALRDSRALAAAAGATEEADAFAARADLLAETINAVL-     740
RhaLa       DWLSLDNGAN-PQGKTNEDYI-ASIYYFVSAS----IV------SKAARLLHYDMESDYY     604
RhaHp       DWLDPSAPPENPLLALTDAYLVADAFLFKILT----MT------SETAEILN--------     630
            **   .   :       : ::          :     .      :  * :*

RhaBa       --WDEEKRA-YI-DCIHADGRRSDVYSMQTQV-VAYLCGVAQG-----------------     778
RhaLa       ENLARNIKTNILNEFVTEKGRIA--IDTQTALVLALHFGLVHDYQKSQV-VADLVKKVKD     661
RhaHp       --WLPKCRSDFNEAYISSSGKLT--SNTQTAYALAICFGLYKNDDQISYAGNQLSELVRQ     686
                 : ::      :  .*: :   . **   :*   *: :

RhaBa       EREAVIEGYLSSPPPAFVQIGSPFMSFFYYEALEKAGRQTLMLD----DIRRNYGQMLRY     834
RhaLa       DNKHLQTGF----------VGTPFL----LSVLSNNNQHHLAMDIFMQEDCPSWLYEVNM     707
RhaHp       SDYKISTGF----------AGTPFV----TEALTITGHLEDAYKMILQKECPSWLYPVSV     732
            .   :  *:           *:**:     ..*    :     .    .   .:   :
                  841
RhaBa       DATTCWEMYPNFA-ENRSNPDMLTRSHCHAWSAAPGYFLGSSILGVKRGADGWRTVDIAP     893
RhaLa       GATTIWERWNSVLPDGKMNPEGMNSLNHYSF-GAVMMWMYQCVVGLNQFDAGFKEIYFAP     766
RhaHp       GATTVWERWDSMLPTGHINPGEMTSFNHYAL-GSIANWMHERMGGLKLKEPGWREFYLRP     791
             *** ** : ..    : **  :.  . ::  .:   :: . : *::    *:: . : *
```

Figure 26: Sequence alignment of the rhamnosidases selected for the study. Only fragments of the whole sequences are presented for clarity reasons (alignment of the whole sequences can be found in section 6.3). The sequence of rhamnosidase Rha_{Hp} was identified in this study for the first time (see section 4.2.3), while the two other sequences were already available. Asterisks - positions with fully conserved residue; colon - conserved groups with strongly similar properties; full stop - conserved groups with weakly similar properties. The amino acid residues which were elucidated to be crucial for the Rha_{Ba} activity are marked with red boxes, indicating that these regions are indeed conservative in GH family 78. The sequences were aligned using the CLUSTAL Omega tool.[189]

It is well known that non-nucleophilic residues such as Gly, Ala, and Ser are among the best acting amino acids in the preparation of glycosynthases. In particular, the polar character of the serine side chain can form hydrogen bonds to the external nucleophile and thus promote the nucleophilic attack to the anomeric centre of the substrate.[103] It is impossible to predict which amino acid residue should replace the catalytic one to obtain the active glycosynthase, thus it was decided to use two amino acids as a replacement for each potential catalytic residue - serine and glycine, as these residues are statistically the most successful ones in glycosynthase creation[103] (Table 11).

Table 11: Variants of Rha_{Ba} created in this study and the introduced mutations.

Nr	Introduced mutation	Modification in the gene sequence
1	D567G	GAC→GGC
2	D567S	GAC→AGC
3	E572G	GAA→GGC
4	E572S	GAA→AGC
5	E841G	GAG→GGC
6	E841S	GAG→AGC
7	D567G; E572G	GAC→GGC; GAA→GGC

4.2.1.2 Cloning and expression in *E. coli*

A synthetic Rha_{Ba} gene (UniProt accession no. Q93RE7) was ordered from GenScript with a His_6Tag at the C-terminal end of the sequence to facilitate the purification process. The point mutations (Table 11) were introduced *via* the QuikChange method as described in section 5.6.7.6, using *Pfu* polymerase and Rha_{Ba} in pUC57 vector. After successful introduction of the mutations, each plasmid was digested with *HindIII*, *NcoI* and *NdeI* and the 2870 bps fragment containing the gene of interest was ligated into pET28a(+) plasmids, resulting in the pET28a(+)::Rha_{Ba} vectors. After introduction of the correct mutation was verified *via* sequencing, the newly-created plasmids were used for *E. coli* BL21 strain transformation and the strains were then used for expression. Test expression of the created mutants confirmed efficient production of the protein (Figure 27).

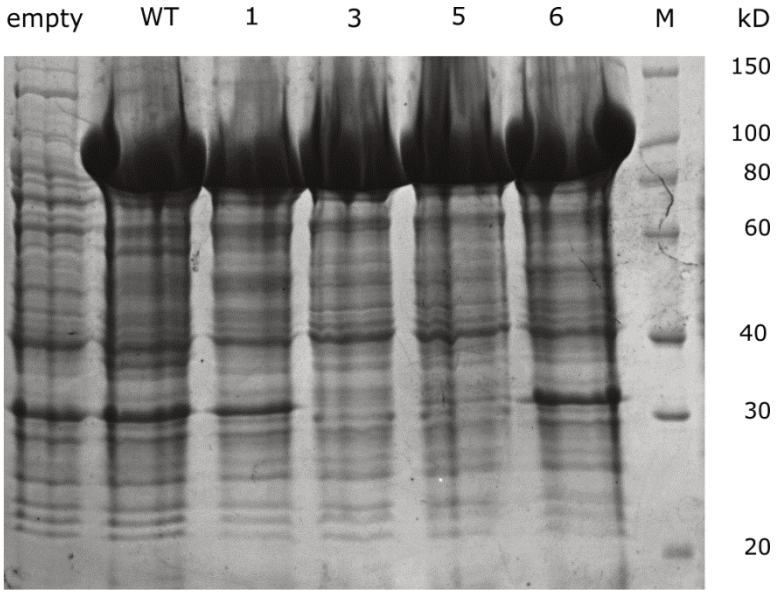


Figure 27: Test expression of the rhamnosidase Rha_{Ba} in *E. coli.* Strains transformed with empty pET28a(+) plasmid, pET28a(+) containing wild type (WT) rhamnosidase gene or the rhamnosidase gene with introduced mutations (1, 3, 5, 6) were grown in 100 mL LB medium, 16 °C for 31 h. The cells were then harvested, disrupted with glass beads (section 5.7.1) and the obtained crude extracts were diluted tenfold and analysed on SDS-PAGE. Clear bands in expected size (106 kD) were observed in the samples where rhamnosidase gene was introduced, in contrast to the control sample with the empty plasmid. The proteins were visualised with the Coomassie staining method.

The expression was performed essentially as described in the original publication by Hashimoto *et al*.[183] (see section 0 for details). After centrifugation, the pellets were disrupted *via* sonication as described in section 5.7.1, centrifuged at 4 °C, 12000 g and the crude extract was used for IMAC purification as presented in Figure 28. The initially achieved yields in LB medium of 7-10 mg L^{-1} were lower than the reported 335 mg L^{-1} of produced rhamnosidase which were obtained in BL21(DE3)pLysS *E. coli* strain.[183] However, when the expression medium was changed from LB to the more complex TB medium, excellent yields up to 326 mg L^{-1} of the purified Rha_{Ba} could be obtained. No impact of the introduced mutations on the obtained protein yield could be observed.

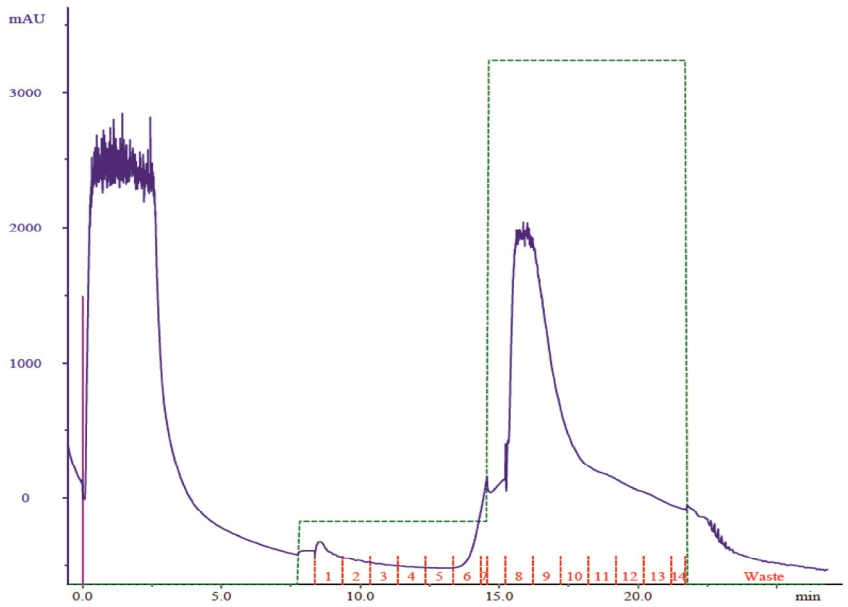


Figure 28: Typical profile of the Rha_{Ba} purification with the ÄKTA system. The blue line indicates the protein concentration measured spectrophotometrically at 280 nm. After the crude extract was applied on the Ni-NTA column, the concentration of imidazole (dashed green line) was increased to 30 mM to elute impurities, and finally the rhamnosidase was eluted in fractions 8 and 9 with 250 mM imidazole in the elution buffer.

The wild type enzyme was then tested for the best storage method. For this purpose, the purified enzyme sample was first desalted (see section 5.7.21 for details) and then either lyophilised (section 5.7.22) or stored in the fridge. After four weeks of storage, the activities of enzyme stored in both ways were compared using the activity assay with *p*NP α-rhamnopyranoside as described (section 5.7.13). The lyophilised sample showed 20% higher activity when compared to the sample stored in the fridge. Therefore, lyophilisation was selected as the preferred long-term storage method.

4.2.1.3 Application for chemical synthesis

Production of the enzyme variants was the first step towards the application in chemical synthesis. First, it had to be evaluated whether the introduced mutations had an impact on the enzyme hydrolytic activity as it was planned. Six single mutants were tested with the standard rhamnosidase substrate *p*NP α-rhamnopyranoside. In addition, the rhamnosidase activity towards *p*NP β-glucopyranoside was tested. This compound is similar to the natural product of naringin hydrolysis, prunin (Figure 29), as in both structures glucose is bound in β-configuration to an aryl group; it could therefore be a

potential rhamnosyl group acceptor. In addition, this acceptor would allow ready reaction monitoring. Therefore, it should preferably not be hydrolysed by the rhamnosidase.

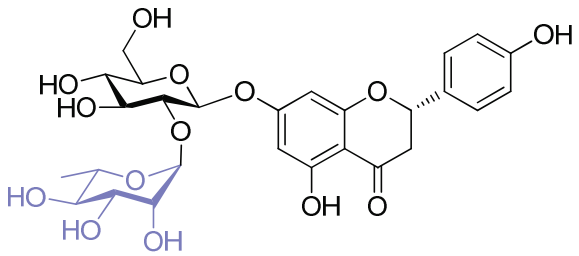


Figure 29: Structure of the naringin, one of the natural substrates for a rhamnosidase. The terminal rhamnose, marked in blue, is cleaved by a rhamnosidase to form a prunin.

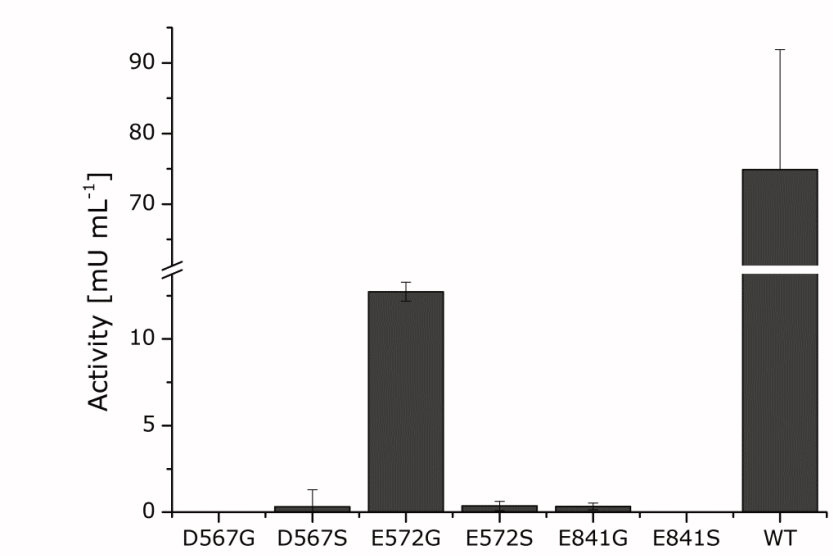


Figure 30: Hydrolytic activity of the wild type rhamnosidase Rha_{Ba} as compared to the mutants. 0.5 mg mL^{-1} of each enzyme variant preparation was incubated with the standard substrate *p*NP α-rhamnopyranoside for 20 minutes and its hydrolysis was followed at 410 nm for 15 minutes at 30 °C. The presented values are the average from three measurements.

As presented in Figure 30, a significant loss of hydrolytic activity was observed for almost all the mutants when compared to the wild type (WT) enzyme; only variant 3 (E572G) was still able to hydrolyse the *p*NP α-rhamnopyranoside substrate, although with decreased activity. Therefore, it was decided to create the additional double mutant 7 (D567G; E572G), in which both potential catalytic residues were replaced with glycine. This mutant was not able to hydrolyse the substrate. No hydrolytic activity towards *p*NP β-glucopyranoside was observed (data not shown), thus it could indeed be further used as a potential acceptor. In conclusion, both initial assumptions could be verified for the created mutants:

loss of hydrolytic activity towards rhamnose group and lack of hydrolytic activity towards the selected potential rhamnosyl group acceptor.

As published before, the addition of organic solvent may have a beneficial effect on the reaction outcome, therefore the stability of the Rha_{Ba} in the presence of various concentrations of DMSO was analysed (Figure 31). 0.5 mg mL^{-1} of enzyme sample was incubated with a given concentration of DMSO for 1 h at 30 °C and the activity after incubation was assayed by standard assay (see section 5.7.13 for details). The enzyme was relatively stable; after 1 h incubation at 30 °C in presence of 20% DMSO, more than 80% of the starting activity was still remaining. Therefore, this solvent was suitable for application in the reaction.

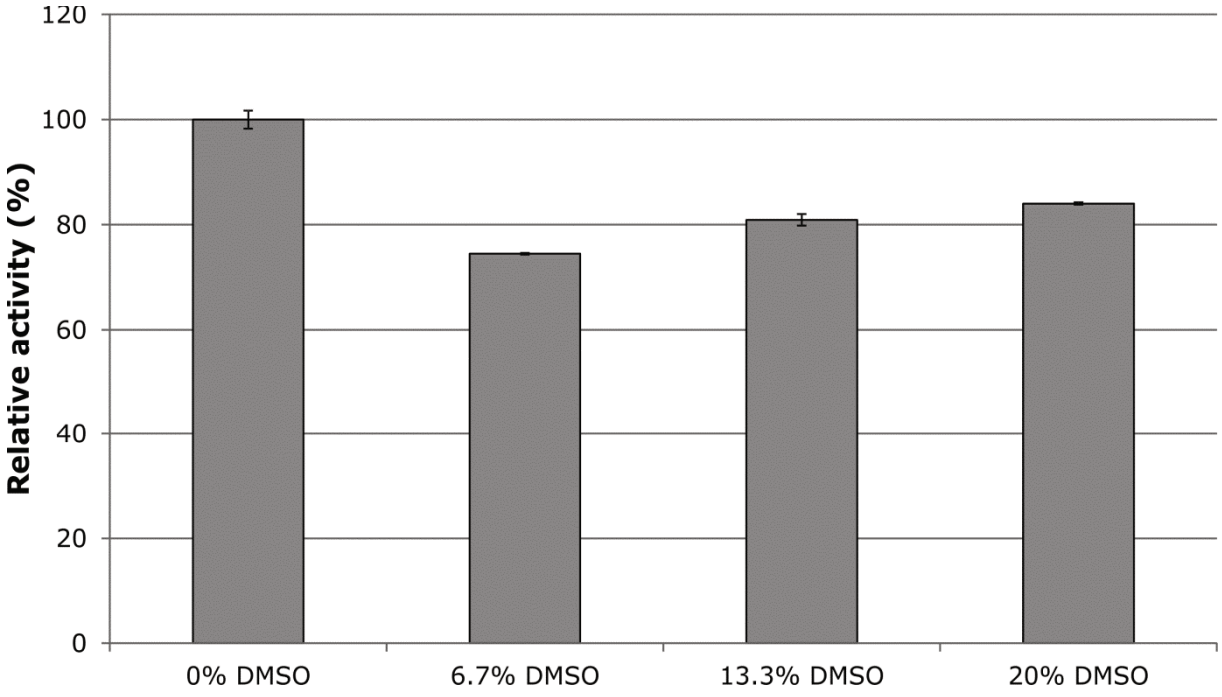


Figure 31: Stability of wild-type rhamnosidase Rha_{Ba} in presence of DMSO. 0.5 mg mL^{-1} of enzyme sample was incubated with a given concentration of DMSO for 1 h at 30 °C and the activity after incubation was assayed by standard assay. The obtained values were then normalised to the control enzyme sample activity without DMSO (84 mU mL^{-1}), which was taken as 100%. The presented values are the average from three measurements.

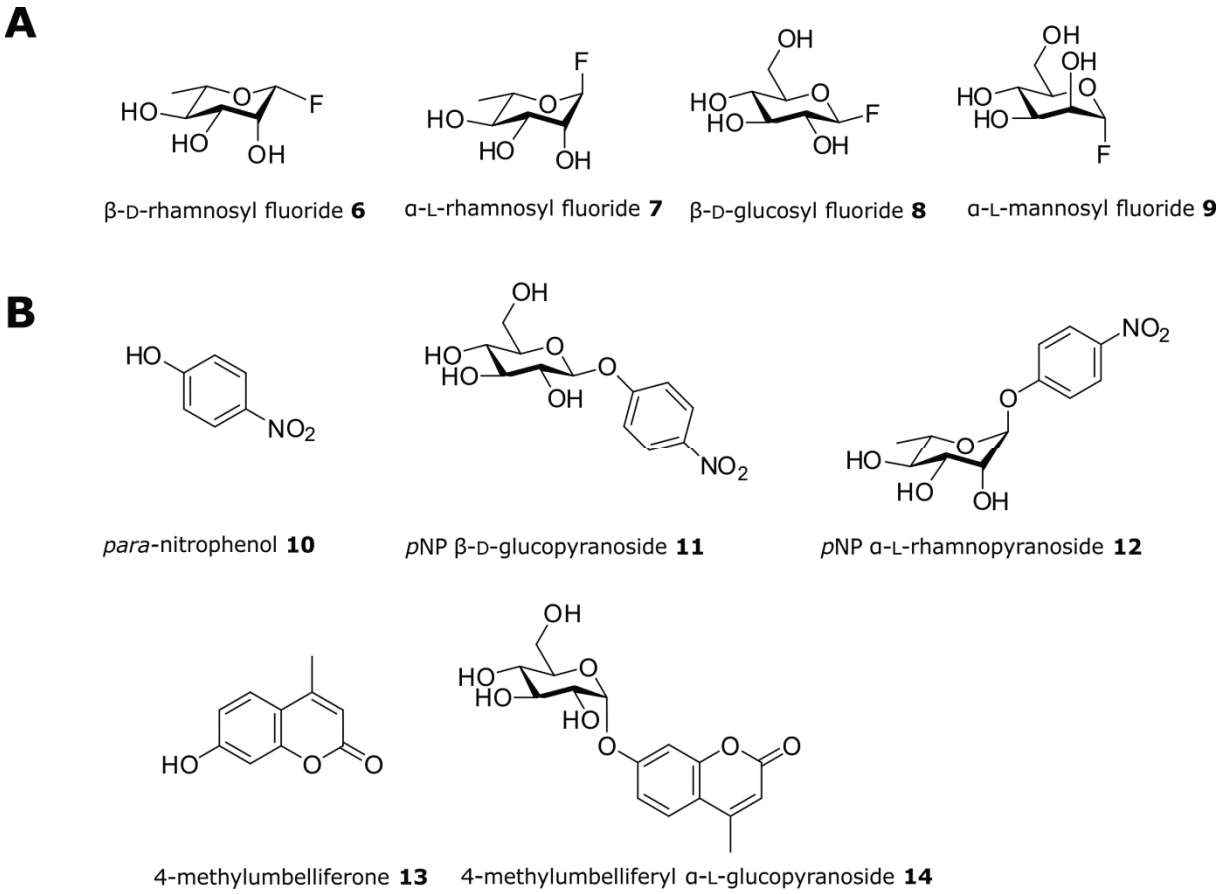


Figure 32: Glycosyl donors (A, 6-9) and acceptors (B, 10-14) used in this study.

As the basic parameters were established, the produced enzyme variants were applied for the first test reactions. All tested donors and acceptors are presented in Figure 32. The synthesis of the donors was performed by Natalie Hoffmann, who cooperated in this part of the project; all obtained results were analysed and discussed in cooperation. The acceptors were obtained from commercial sources.

Finding optimal reaction conditions is of great importance for efficient transglycosylation. As pointed out in a recent review by Cobucci-Ponzano *et al*.[103] already as a basic factor as the donor: acceptor stoichiometric ratio can significantly influence the yield of the reaction. While in general excess of donor facilitates the transfer of the glycosyl moiety onto the acceptor molecule, it may also have a disadvantageous effect on the reaction, as the acceptor might compete in the -1 donor binding site. Another important factor is the pH of the reaction; in case of *Bacillus circulans* xylanase GS variant, the best yields of product were obtained at pH 8.5, although the optimal pH for the native enzyme was 6.0.[132] Possible explanations for this effect may be faster hydrolysis of a newly formed product at basic pH or the requirement of a protonated residue of higher pKa to assist in carboxylic acid departure. Taking this information into

account, I decided to test several conditions to determine which factor is of the greatest importance for the rhamnosynthase activity.

As an easy-to-follow initial reaction, the potential rhamnosylation of 4-nitrophenol **10** was chosen (Figure 33). Since the educt 4-nitrophenol ions give absorbance at 410 nm, whereas the glycosylated 4-nitrophenol product does not, a successful reaction progress should be easy to follow *via* a decrease of the absorbance over the time.

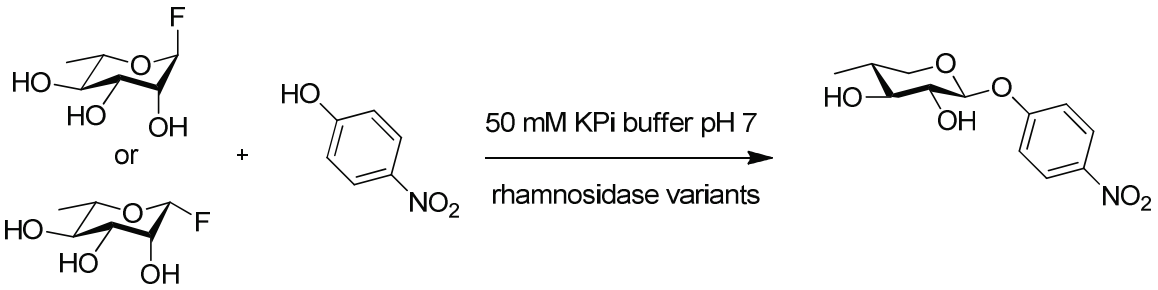


Figure 33: Rhamnosylation of 4-nitrophenol. Two donors (**6** and **7**) were tested in the reaction with the acceptor 10 at pH 7.0 in microplate format. The progress of the reaction was followed *via* absorbance measurement at 410 nm.

The reactions were performed in microplate format and the following conditions were tested: 0.5 mg mL^{-1} enzyme (wild type and the 6 mutants), 0.2 mM to 1.2 mM donor, 0.2 mM 4-nitrophenol, 50 mM KP_i pH 7, 30 °C. Two donors (rhamnosyl fluorides **6** and **7**) were tested in various ratios to the acceptor to identify the most suitable conditions. The reaction progress was monitored for one hour.

Unfortunately, the values obtained for most of the mutants were comparable to the controls, indicating no real product formation during the incubation time. However, for the variant 1 (D567G) the measured change of the absorbance in time was slightly higher than for other variants. The small scale of the reaction prevented from detailed analysis of the product; therefore this variant was selected for more detailed experiments in 1 mL total reaction volume carried out in 1.5 mL Eppendorf tubes. The tested conditions are summarised in Table 12.

Table 12: Overview of the tested reaction conditions for potential rhamnosynthase activity. Entries 5-13 (marked in red) were tested only with variant 1 of the enzyme, since the preliminary screening results were most promising for this variant, whereas all other conditions were tested with all created mutants and the wild type as a control. All reactions were performed in a total volume of 1 mL. Abbreviations: D - donor, A – acceptor, see Figure 32, p. 90 for names and structures.

Entry	Donor	Acceptor	Ratio (D:A)	Buffer	Temperature	Time
1	7	11	1 : 1	50 mM KPi pH 7	30 °C	1 h
2	8	11	1 : 1	50 mM KPi pH 7	30 °C	1 h
3	7	10	6 : 1	50 mM KPi pH 7	30 °C	2 h
4	6	10	6 : 1	50 mM KPi pH 7	30 °C	2 h
5	7	12	2 : 1	50 mM KPi pH 7	4 °C	72 h
6	7	12	2 : 1	50 mM KPi pH 7	15 °C	72 h
7	7	12	2 : 1	50 mM KPi pH 7	30 °C	72 h
8	6	12	2 : 1	50 mM KPi pH 7	4 °C	72 h
9	6	12	2 : 1	50 mM KPi pH 7	15 °C	72 h
10	6	12	2 : 1	50 mM KPi pH 7	30 °C	72 h
11	7	10	6 : 1	50 mM KPi pH 7	4 °C	72 h
12	7	10	6 : 1	50 mM KPi pH 7	15 °C	72 h
13	7	10	6 : 1	50 mM KPi pH 7	30 °C	72 h
14	7	10	1 : 1	50 mM KPi pH 7	4 °C	72 h
15	7	10	1 : 1	50 mM KPi pH 7	15 °C	72 h
16	6	10	1 : 1	50 mM KPi pH 7	4 °C	72 h
17	6	10	1 : 1	50 mM KPi pH 7	15 °C	72 h
18	9	10	1 : 2	50 mM Sørensen pH 8	30 °C	24 h
19	9	11	1 : 2	50 mM Sørensen pH 8	30 °C	24 h
20	7	10	1 : 2	50 mM Sørensen pH 8	30 °C	24 h
21	7	11	1 : 2	50 mM Sørensen pH 8	30 °C	24 h
22	8	11	1 : 3	50 mM Sørensen pH 8	30 °C	2 h
23	8	11	1 : 3	50 mM KPi pH 6	30 °C	2 h
24	8	14	10 : 1	50 mM KPi pH 7	30 °C	2 h
25	8	14	2 : 1	50 mM KPi pH 7	30 °C	2 h
26	8	14	5 : 1	50 mM Sørensen pH 8	30 °C	2 h
27	8	13	5 : 1	50 mM Sørensen pH 8	30 °C	2 h

Entries 1 and 2 were performed as test reactions. After 1 h, the reaction progress was analysed *via* TLC and the degradation product of the donor was visible. Therefore, in the following reactions, lower temperatures like 4 °C and 15 °C, as well different ratios of donor and acceptor were tested and the reaction solutions were analysed by HPLC after the reaction time given in Table 12. Only traces of the desired product could be found, which were also present in the control samples with the wild type enzyme. NMR analysis revealed only the starting substances and the degradation product from the fluorides. No positive effect of the lower temperatures could therefore be found. A lower pH of 6.0 and higher pH 8.0 were also tested with no positive effect. In the final attempt,

4-methylumbelliferyl α-L-glucopyranoside **14** was tested as an acceptor, since it possess structural similarities to the natural product of the derhamnosylation, prunin; however, no expected product could be obtained as verified by NMR analysis.

It can be assumed that only donors with a rhamnosyl moiety are accepted, as the glycosynthases are usually donor group-specific, as a result of the active site space limitation. This is supported by the observation that the wild type Rha_{Ba} was unable to hydrolyse the *p*NP β-D-glucopyranoside **11**; presumably the glycosyl group could not fit into the active site.

In summary, despite the screening of a wide range of conditions, no expected product could be obtained. It is possible that this particular enzyme cannot be modified into a glycosynthase, as it was reported previously for several enzymes. For example, *Thermosporum saccharolyticum* β-xylosidase and *Bacillus circulans* β-galactosidase converted to alanine or serine nucleophile mutants did not show glycosynthase activity.[152] It can be assumed that many other enzymes failed to be modified onto glycosynthase, since unsuccessful approaches are often not published at all.

On the other hand, it can be speculated that the instability of the β-glycosyl fluorides was a significant obstacle and could also be a main reason for the failure of the used approach. In each case, degradation products of applied donors were already observed after 1 h of reaction. Similar difficulties were reported for *Sulfolobus solfataricus* α-L-fucosidase; in spite of introducing mutations and screening of variety of conditions, no oligosaccharide formation was observed.[190] However, when β-glycosyl azides were used as glycosyl donors, yields up to 91% were obtained for the same mutants.[141] In case of *Thermotoga maritima* α-D-galactosidase, application of significantly more stable β-glycosyl azides as donors allowed to modify the enzyme to an efficient glycosynthase with good yields and high regioselectivity.[146] In conclusion, the use of more stable donors, e.g., azides, could be a possible solution to overcome the donor instability problem.

4.2.2 Rhamnosidase from L. acidophilus

The rhamnosidase Rha_{La} was selected for this study, as it is an ortholog to Rha_{Ba}, which was reported to be a versatile catalyst as described in detail in section 4.1.2.2. The rhamnosidase gene was obtained from Sebastian Schumacher, who isolated it from *L. acidophilus* cells. It was cloned into a pET28a(+) plasmid both as a full length and as a short inactive version (missing part of the active centre) to investigate the potential influence of the enzyme activity on the expression level. Plasmids containing both versions were used for *E. coli* BL21 strain transformation. The expression was performed as described in section 5.5.4. The harvested cells were disrupted by sonication (section 5.7.1) and the crude extracts were analysed *via* SDS-PAGE. As it can be seen in Figure 34, a protein in the expected size range for the Rha_{La} short version could be produced in *E. coli*, although it was mostly found in the pellet (lanes marked P from induced and uninduced cultures).

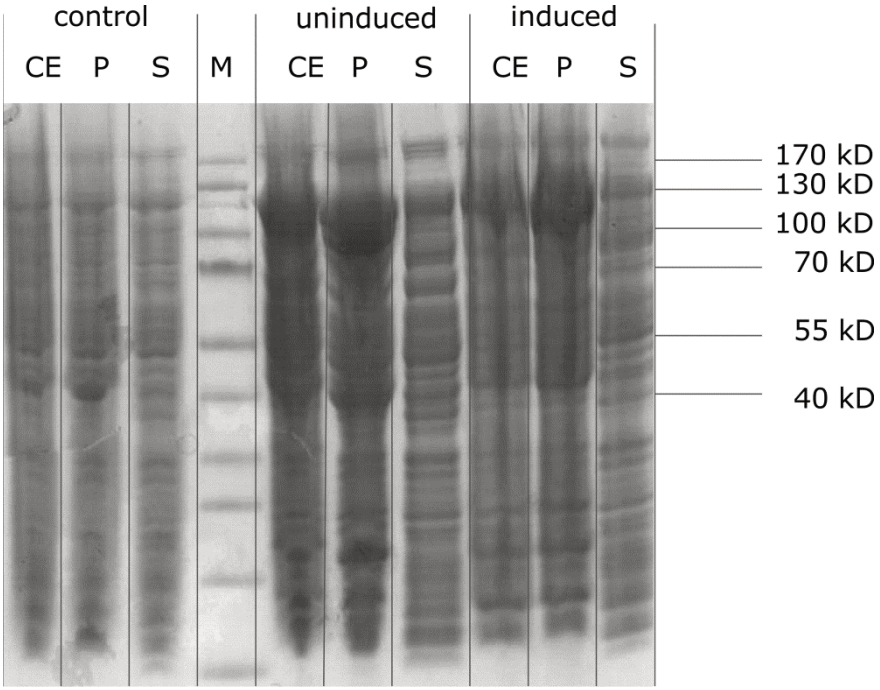


Figure 34: SDS-PAGE analysis of the expression of the short inactive version of Rha_{La}. 10 µL of *E. coli* crude extracts were loaded on the gel, separated and stained with Coomassie Blue. Samples: control - *E. coli* transformed with empty plasmid), uninduced - uninduced *E. coli* cells containing $sRha_{La}$ gene, induced - induced *E. coli* cells containing $sRha_{La}$ gene. The expected size of the product is 99.3 kD. Abbreviations: CE - crude extract; P - pellet; S – supernatant; M – marker - GeneRuler Prestained Protein Ladder (Fermentas GmbH).

On the other hand, the expression of the whole gene in *E. coli* failed as no additional bands on the SDS-PAGE in the expected size could be observed and no activity in the crude extracts could be detected with the standard assay (data not

shown). It can be therefore assumed that the activity of the enzyme was problematic for the cells, and thus the expression was blocked. This result is contrary to the published data,[16] where active rhamnosidase Rha_{La} was reported to be expressed in *E. coli*. These difficulties were further discussed with the author of the publication and some new details were revealed; although they could measure the rhamnosidase activity in the *E. coli* crude extract, as well as apply it in this form for reactions, no distinct rhamnosidase band on the SDS-PAGE was ever observed by the authors; in addition, IMAC purification failed for unknown reason (Beekwilder, personal communication[191]). In the light of this information it becomes clear, that *E. coli* is not an effective host for Rha_{La} expression. Although the same expression strain as published was used in my experiments, a different plasmid was applied for the expression; moreover, the His_6Tag of the published construct was located on the N-terminal end. These small differences could decrease the already poor rhamnosidase yield to undetectable level, which would explain the lack of observed activity in *E. coli* crude extract.

To overcome the clearly problematic expression of Rha_{La} in *E. coli*, three yeast expression systems, introduced in section 3.1.2, were applied for the expression of the rhamnosidase. Both intracellular expression (*S. cerevisiae*, *K. lactis*) as well as secretion (*K. lactis*, *H. polymorpha*) did not result in production of the active enzyme, in spite of intensive optimisation attempts of the expression conditions (media, temperature, pH, induction). Therefore, the Rha_{La} was excluded from further investigations.

4.2.3 Rhamnosidase from H. polymorpha

4.2.3.1 Gene isolation and removal of intron

The rhamnosidase Rha_{Hp} from *H. polymoprha* was selected for this study as it is an eukaryotic enzyme and no structural information on an eukaryotic rhamnosidase is available to date. Moreover, it was shown to have potential for industrial applications. As explained in the introduction, in the original report on the Rha_{Hp} rhamnosidase the enzyme was purified from *H. polymorpha* without the information on the protein or gene sequence.[171] Only the first 17 amino acids, VEVREVRLESNALGPTN, were identified. This information was used to find the gene in the genome of *H. polymorpha* NCYC495 strain, which was sequenced by the US Department of Energy Joint Genome Institute.[192] Application of the

BLAST tool with the known rhamnosidase protein sequence resulted in only one match and the first amino acids of the potential rhamnosidase were exactly the same as the published ones. The identification of the rhamnosidase gene was performed by Dr. Adelheid Degelmann from the industrial partner of this project, ARTES Biotechnology GmbH.

To amplify the gene from the host organism, rather long primers based on the genome sequence were designed, as genomic DNA from another strain than NCYC495 was used as the template. First, the genomic DNA was extracted from the *H. polymorpha* cells as described in section 5.6.2 and used as a template for the gene amplification. Unfortunately, despite extensive optimisation and a variety of polymerases tested, no product could be seen on agarose gels. As an alternative approach, whole *H. polymorpha* cells after thermal disruption were used as the template and the desired product could be obtained, when the *Taq* polymerase was applied. The amplified gene was ligated into a pCR2.1 plasmid (TOPO® TA Cloning® Kit, Invitrogen).

In the gene sequence, a 31 bps intron was identified, which had to be removed prior to expression as it was causing a frame shift and would presumably not be processed correctly during heterologous expression. At first, the overlap extension PCR method was applied, but only the first part of the gene could be obtained as a final product, despite protocol optimisation. Consequently, the QuikChange approach was tested, using the second part of the gene (after the intron) as a megaprimer as described in detail in section 5.6.7.6. The amplified mega primer had a size of 795 bp, with 32 bps overhang at the 5’ end and 91 bps overhang at the 3’ end of the sequence. *E. coli* MACH1 competent cells were transformed with the obtained construct, five colonies grown overnight on the selective plates were selected and colony PCR was performed to verify the size of the insert in the plasmid (Figure 35).

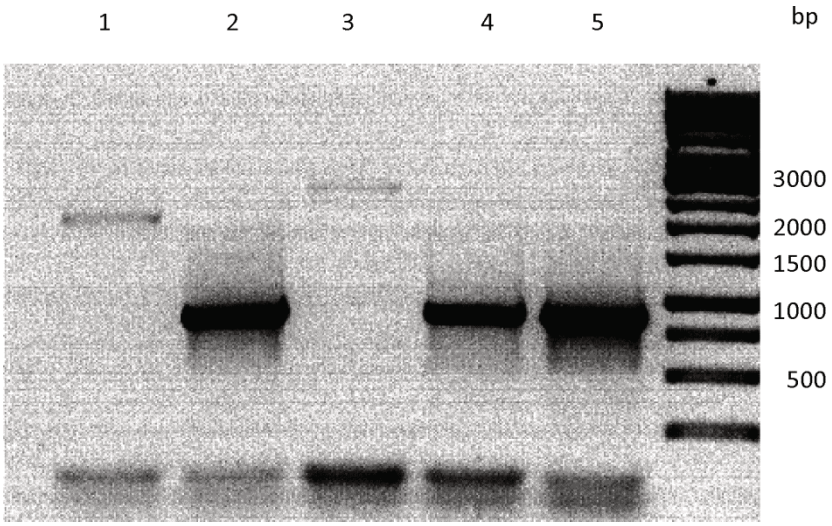


Figure 35: Colony PCR to verify intron removal from the rhamnosidase Ram_{Hp} gene. Five colonies, marked as 1-5, were selected for the procedure, amplifying the insert region in pCR2.1. The expected size of the whole gene was 2571 bps, and in lanes 1 and 3 the bands around this size can be observed, although with an apparent difference in size.

In samples 1 and 3 amplicons in the range of the expected size were observed and although it appears as if there is a difference in size between them, when the plasmid was isolated, sequencing confirmed correct removal of the intron in both samples as presented in Figure 36 (only the result for one of the samples is presented for clarity reasons).

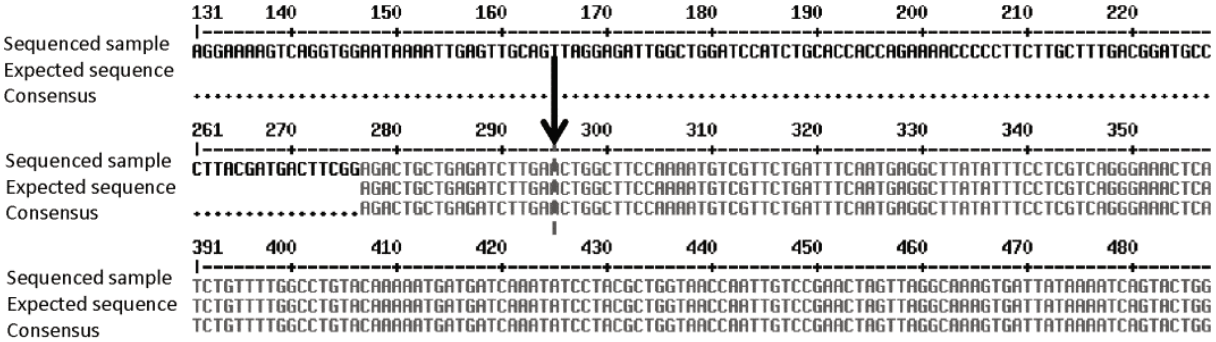


Figure 36: Sequence verification of the cloned rhamnosidase Ram_{Hp} after intron removal *via* QuikChange. The obtained sequence was aligned to the expected product after the intron removal. The arrow indicates the sequence spot from where the intron was successfully removed. Apart from the fragment presented, the whole gene was sequenced to exclude potential mutations.

4.2.3.2 Expression in three yeasts

The amplified rhamnosidase gene was cloned into modified plasmids piSC, pKL, pHP for intracellular expression (*S. cerevisiae*) or secretion (*K. lactis*, *H. polymorpha*), using *AvrII* and *Pfl23II* restriction sites and an additional C-terminal His_6Tag. Test expression in both *S. cerevisiae* and *K. lactis* did not yield

any detectable production of the protein, whereas in samples from *H. polymorpha* small amounts of the tagged protein could be detected by Western Blot. To further investigate the optimal conditions for the homologous rhamnosidase Rha_{Hp} production, the Colony Lift Assay was performed as described in detail in section 5.7.19 (Figure 37).

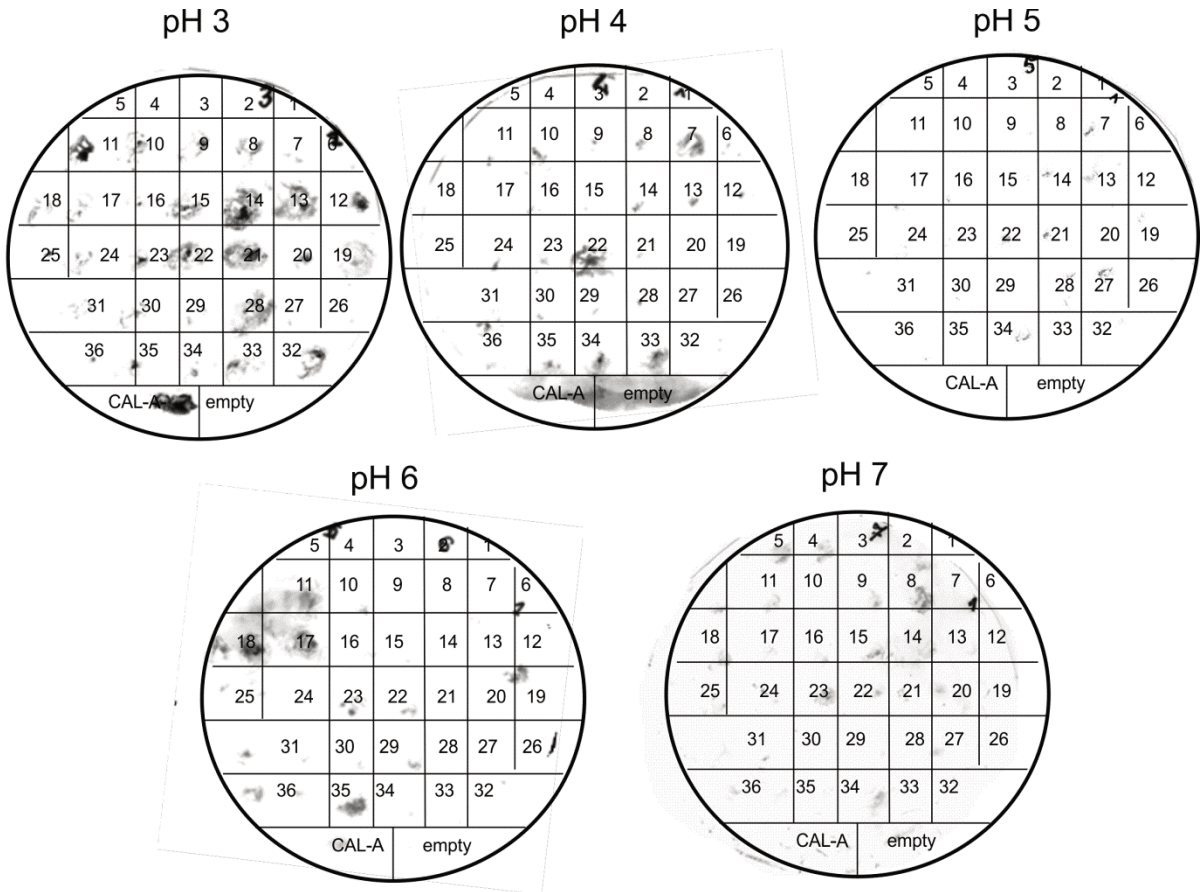


Figure 37: Colony Lift Assay. The nitrocellulose membranes, after incubation on YNB agar plates of various pH containing 38 *H. polymorpha* transformants were developed according to standard Western Blot protocol (section 5.7.18) to visualise the secreted His_6tagged rhamnosidase. As a positive control, a transformant secreting lipase CAL-A (bottom left) was used, and as a negative control a *H. polymorpha* transformant containing the empty pHP plasmid (bottom right, marked as 'empty') was applied.

Briefly, 36 *H. polymorpha* transformants after the passaging (section 5.5.7) and two controls (a transformant containing the empty plasmid and a transformant secreting a high amount of His_6tagged CAL-A) were transferred onto five agar plates containing YNB medium at various pH values, ranging from 3.0 to 7.0. After the colonies were formed, the nitrocellulose membranes were put on the plates and incubated for 24 h. After this time, the membranes were washed and developed by the standard Western Blot method (see section 5.7.18

for details) to visualise the protein, which was detectable through the C-terminal His_6Tag.

As presented in Figure 37, apparently several of the transformants (e.g., 2, 22, 34) were able to secrete the rhamnosidase, especially on plates where medium of a low pH of 3 or 4 was used. However, no activity of the produced protein could be detected with the standard assay (section 5.7.12). In conclusion it is assumed that the rhamnosidase Rha_{Hp}, which is naturally located intracellularly, cannot be secreted in active form by the yeast cells.

Consequently, the plasmids pKL and pHP were modified to remove the MFα secretion signal. The rhamnosidase gene was then cloned into the piKL and piHP plasmids and the constructs were used for the *K. lactis* and *H. polymorpha* competent cells transformation. When small scale expression was performed, rhamnosidase activity could be detected only in *H. polymorpha* crude extracts. Out of 36 passaged transformants, 6 were able to produce the rhamnosidase in detectable amounts. No rhamnosidase activity in the cells transformed with an empty plasmid was found, which is in agreement with the published observation that the expression of the rhamnosidase Rha_{Hp} in *H. polymorpha* is not constitutive and occurs only upon growth on rhamnose as a carbon source.[171] Transformant nr 24, as the most effective producer (data not shown), was selected for further experiments.

4.2.3.3 Purification and crystallisation attempts

The expression was performed as described in section 5.5.11 in 100 mL scale. After centrifugation, the pellets were disrupted *via* sonication as described in section 5.7.2, centrifuged at 4 °C, 12000 g and the crude extract was used for IMAC purification. The purified protein was desalted (see section 5.7.21 for details) and the buffer was exchanged to 50 mM MOPS pH 7.0 using PD-10 columns. The enzyme was then concentrated down to the 1 mL volume using a Vivaspin 20 device (Sartorius AG) with a 10 kD membrane (see section 5.7.20 for details).

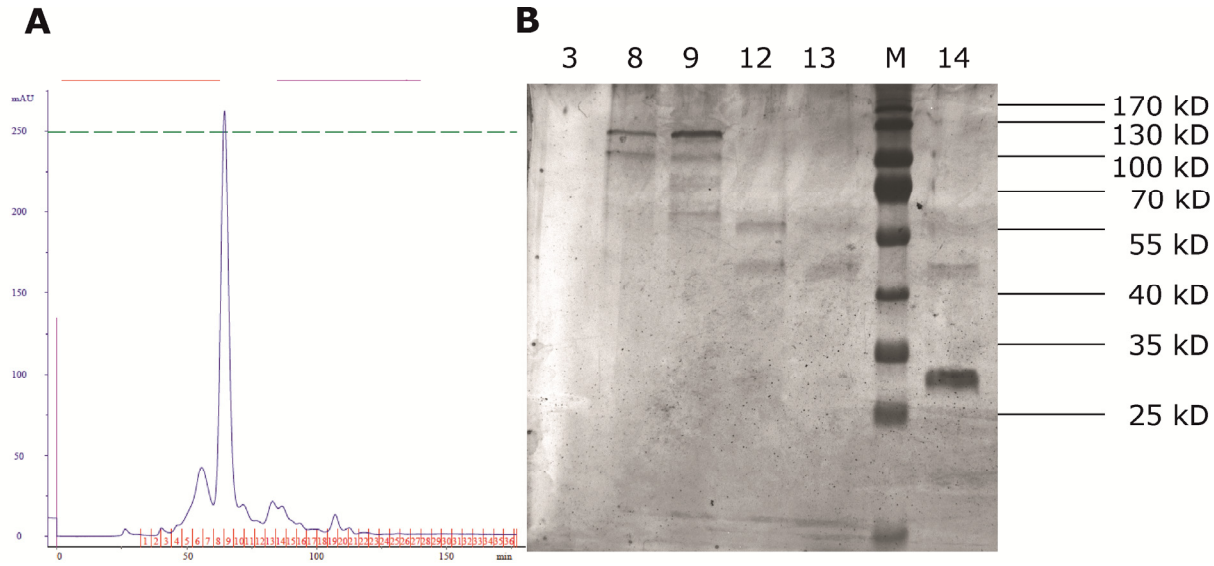


Figure 38: Typical run for the Rha_{Hp} SEC purification with the ÄKTA system (A) and analysis of purified fractions on SDS-PAGE (B). Calculated expected size of the protein is 96.2 kD. (A) The column (HiLoad 16/600 Superdex 200 prep grade) was first equilibrated with the buffer (20 mM MOPS, pH 7.0), then the IMAC purified rhamnosidase sample was injected. The protein concentration was measured spectrophotometrically at 280 nm (blue line). Aliquots from protein containing fractions were taken and analysed on a silver stained SDS gel (B). The rhamnosidase was eluted in fractions 8 and 9.

To remove the remaining impurities Size Exclusion Chromatography (SEC) was performed as described in section 5.7.5 and the rhamnosidase was finally eluted with 20 mM MOPS, pH 7.0 (Figure 38). Typically 1 mg of purified, active rhamnosidase (17.6 U mg^{-1}) could be obtained from 1 L *H. polymorpha* culture. It was then concentrated to 5 mg mL^{-1}, centrifuged 5 min, 12000 g, and applied to crystallisation trials.

This part of the project was performed in cooperation with Dr. Oliver H. Weiergräber (Institute of Complex Systems, ICS-6: Structural Biochemistry, Forschungszentrum Jülich); he measured the crystals and performed their analysis. Several conditions for crystallisation were tested and the sitting-drop crystallisation method was used. The plates were incubated at RT for about two weeks until the small plate-like crystals were formed (Figure 39). The crystals were found in conditions which contained around 20% PEG and additional salts (JCSG Core I: 37- 0.2 M sodium iodide, 20%(w/v) PEG 3350; JCSG Core II: 1 - 0.2 M sodium chloride, 0.1 M CAPS pH 10.5, 20% (w/v) PEG 8000; JCSG Core II: 38 – 0.2 M sodium nitrate, 20% (w/v) PEG 3350; JCSG Core IV: 27 – 0.17 M sodium acetate, 0.085 M Tris HCl, pH 8.5, 25.5% PEG 4000, 15% glycerol; the Core Suits are products of Qiagen). These crystals were measured at the Synchrotron (Grenoble, France), but unfortunately the diffraction was moderate

and very anisotropic due to the crystal's shape. Additional screening with rhamnose as inhibitor did not result in any crystal. The data for the best crystal identified so far, catalogued as Rh12, were collected at 3.2 Å resolution. It suggest the monocyclic crystal system (space group P2(1)) and the following lattice constants: a = 122.18 Å; b = 171.15 Å; c = 137.37 Å; alpha = 90.0°; beta = 102.0°; gamma = 90.0°, with probably six molecules per asymmetric unit.

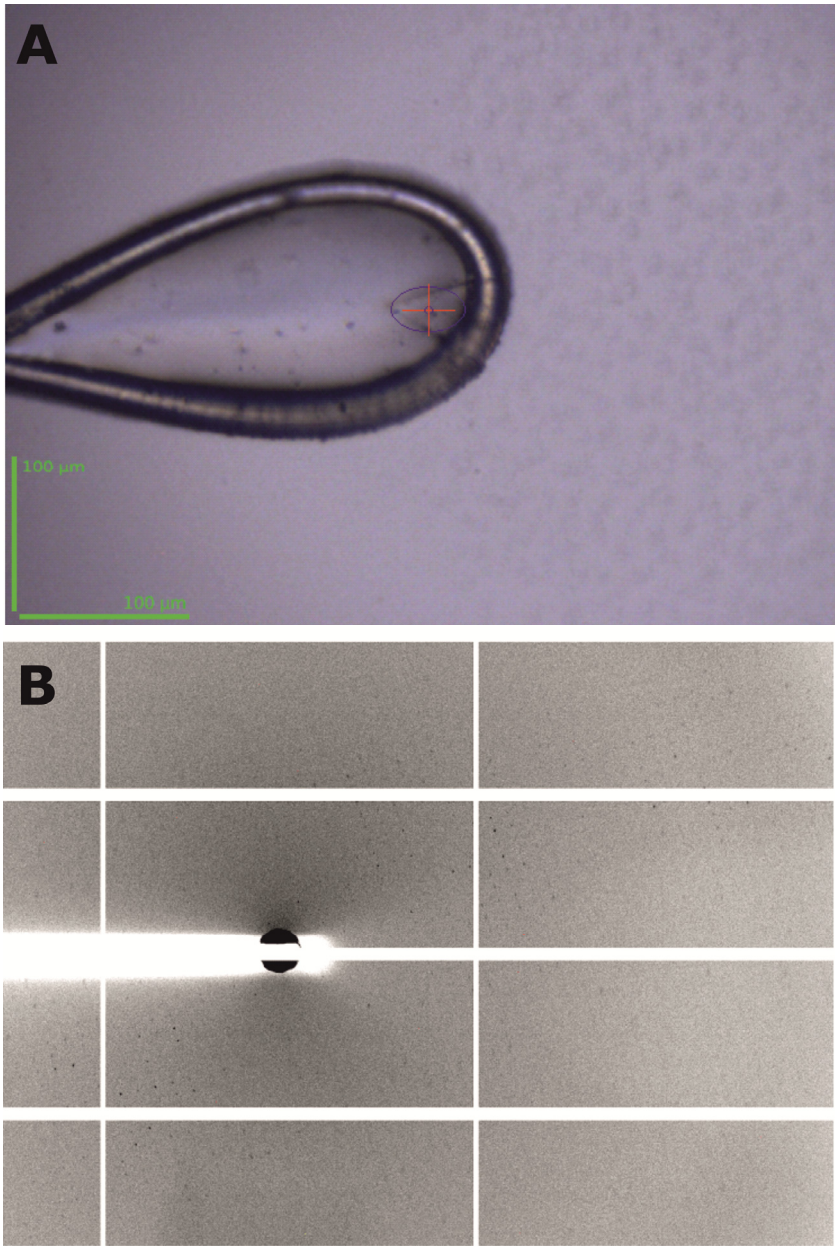


Figure 39: The small crystals obtained in the crystallisation trial and used for the diffraction measurement (A) and a fragment of diffraction image obtained from this crystal (B). The photos were taken and kindly provided by Dr Oliver H. Weiergräber.

An additional difficulty in resolving the crystal structure arises from the fact, that to date only three crystal structures of the rhamnosidases are accessible. Moreover, although similarities in the catalytic domains can be found, there are significant differences in the overall structures of the proteins when compared, thus the molecular replacement has failed. Out of the three structures presented in Figure 40, detailed characterisations were published only for two of them. The structure of Rha_{Ba} is reported to be a homodimer and consists of five domains: one $(\alpha/\alpha)_6$-barrel structure (catalytic domain A) and four β-sandwich structures named N, D1, D2 and C. Two calcium ions were found in domains D2 and C and although the experiments have shown that the activity of the enzyme is independent of any divalent ions, they were reported to be essential for the crystallisation process.[163] Interestingly, a calcium ion was identified also in the crystal structure of the rhamnosidase Rha_{Sa}. Moreover, the respective $(\alpha/\alpha)_6$-barrel structure, also fulfilling catalytic function, and five β-sandwich domains (N, D, E, F, C) were identified in this structure.[188]

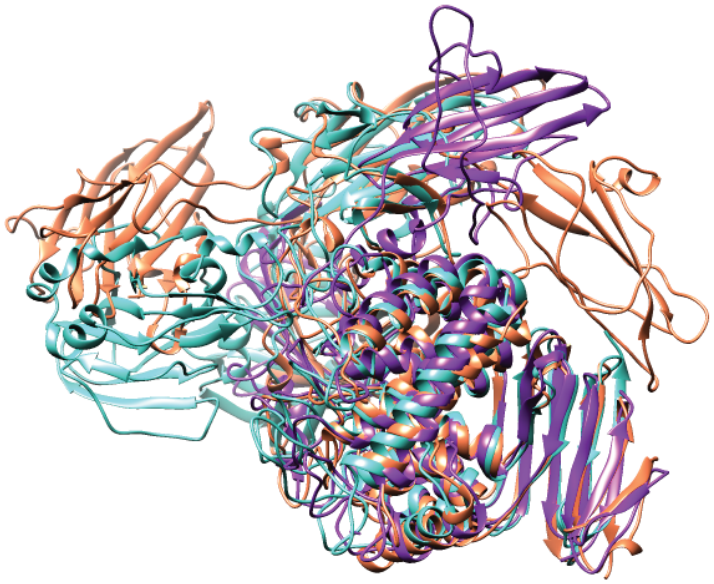


Figure 40: Superimposition of overall structures of rhamnosidases: Rha_{Sa} from *S. avermitilis* (coral, PDB: 3W5M), Rha_{Ba} from *Bacillus* sp. GL1 (turquoise, PDB: 2OKX) and Rha_{Bt} from *Bacteroides thetaiotaomicron* (purple, PDB: 3CIH). Rha_{Sa} consists of six, whereas Rha_{Ba} of five domains. No characterisation for Rha_{Bt} was published to date.

Although both catalytic domains from the two enzymes display structural similarities to each other, which were discussed in more detail at the beginning of this chapter, superimposition of the overall structures reveals differences in

the domain arrangements between the proteins. For example, the domain N from Rha_{Sa} has no equivalent domain in Rha_{Ba}.
In spite of structural similarities, the total amino acid identity is lower than 15% when the Rha_{Sa} is compared to both earlier reported rhamnosidases.[188] This data emphasise the need for further study on rhamnosidase structures in order to investigate how conserved the different domains are.

In summary, the data collected on the unpublished Rha_{Hp} crystals obtained within this thesis, although promising, were not sufficient to completely solve the structure. Two major obstacles were identified: a high anisotropy of the crystals and a lack of homologous protein structures. A possible solution could be screening for conditions resulting in crystals of better quality. It might be achieved by lowering the temperature, so that the crystals are formed slower and therefore with a higher degree of organisation. The ration between the protein and test buffer may be also altered to change the concentrations of additives.
Another interesting idea to elucidate the protein structure would be the application of single-wavelength anomalous dispersion (SAD) or multi-wavelength anomalous dispersion with selenomethionine-incorporated Rha_{Hp}. The major difficulty with this method is the fact that out of three yeast strains tested, the rhamnosidase could be produced in active form only in *H. polymorpha*, for which no strain auxotrophic for methionine is available. As a possible solution, expression in a non-auxotrophic strain in media containing selenomethionine would be an interesting approach. A modified variant of *Penicillium minioluteum* dextranase could be produced with this method in *P. pastoris* and approximately 50% selenomethionine incorporation was reported, which was enough to solve the structure.[193] However, it is difficult to predict the influence of selenomethionine, which is known to be cytotoxic for yeast,[194] on the growth of *H. polymorpha*; various concentrations of the compound should be tested to identify the best ratio between selenomethionine incorporation and rhamnosidase yield.

4.3 Summary and outlook

The rhamnosidases from GH family 78, as a poorly studied family of enzymes, were selected to investigate their potential application in chemical synthesis. These enzymes have a great potential in biotech industry. Beside of the already established function in wine making and debittering of citrus juices, new applications are constantly being developed. Rhamnose plays a role as a chiral intermediate in the organic synthesis of pharmaceutically important agents, thus rhamnosidases could be applied in the preparation of many drug precursors. For example, rhamnosidase hydrolyses diosgene to diosgenin, a compound used in the synthesis of steroid drugs, e.g., progesterone.[15] Another promising and important rhamnosidase application is a prodrug therapy named LEAPT: Lectin-directed enzyme-activated prodrug therapy.[195] It is a targeted drug delivery approach, in which two components are used: synthetically glycosylated α-L-rhamnosidase and a rhamnoside-capped prodrug. The enzyme is delivered to specific cell types, followed by a prodrug, which can be processed exclusively by the α-L-rhamnosidase. In addition, modification of rhamnosidase into rhamnosynthase would open a whole new field for the enzyme usage in biocatalysis. Three published rhamnosidases were targeted during this study and four major aspects: gene identification, expression, creation of mutants with altered activity and crystallisation were investigated.

At the beginning, the well-characterised rhamnosidase Rha_{Ba} from *Bacillus* sp. was used to create seven variants with assumed decreased hydrolytic activities, which were then tested for the potential glycosynthase activity. In this, the selection of the modified amino acids was based on their published influence on the enzyme's activity; however, no catalytic acid/base pair typical for inverting enzymes was found. It is possible that the enzyme follows an unusual reaction mechanism, similar to the one already demonstrated for the 1,2-α-L-fucosidase AfcA from *Bifidobacterium bifidum.* Crystallographic and mutational studies revealed that in the active site of this enzyme, N423 is activated by a neighbouring D766 and acts as a base while N421 makes a hydrogen bond to an attacking water molecule and positions the side-chain of an acid residue, E566, for catalysis.[196] When the mutant derivatives N421G, N423G and D766G were tested for the potential fucosynthase activity, the product fucosyllactose was detected with each of them, and the highest yield of 6% was obtained for the D766G mutant.[149] Similar data on more than two residues involved in the

reaction mechanism were obtained in the original publication for Rha_{Ba}, therefore it could be assumed that the mutagenic approach used in this study could also reverse the activity of the enzyme. Investigated reaction variables included pH, temperature, enzyme quantity, buffer and donor/acceptor ratio. Screening of a variety of glycosyl group donors and acceptors did not result in detectable glycosynthase activity of any of the variants tested. A potential reason for this failure might be the instability of the β-glycosyl fluorides, which should be the preferred donors for the enzyme. This thesis is supported by the fact that degradation products could already be detected after 1 h of the reaction. A possible solution to overcome the problem would be an application of more stable donors, e. g., glycosyl azides or (hetero)aryl- glycosides. Although the usage of lower reaction temperature increases the stability of donors, it also decreases the activity of the enzyme in a significant way; therefore the optimal temperature of 30 °C should be applied as default for future reactions.

Secondly, the rhamnosidase Rha_{La}, which was also selected for this study, could not be produced in the active form in *E. coli*, which is contrary to the published results.[16] A personal communication with the author of the original publication revealed, that also in their case this rhamnosidase was produced with poor yield, as no distinct band could be observed on SDS-PAGE gel and moreover IMAC purification also failed.[191] It is possible that this low expression level was further diminished by the usage of different plasmids with various selection markers (pET28 with kanamycin selection in this study vs. pACYC-DUET1 with chloramphenicol selection in the original publication), resulting in protein expression below the detection level. Alternatively, the localisation of the His_6Tag could also have an influence on the expression level, as in the publication it was located N-terminally, whereas in the performed experiments the tag was fused at the C-terminal end of the used construct. Interestingly, when the short inactive version of the enzyme was used for expression, clear bands of the expected size were visible on SDS-PAGE. This result suggests that the activity of the enzyme was problematic for the cells; as a result, the expression was hindered. Application of the eukaryotic hosts did not result in any improvement. Therefore, this enzyme could not be applied for the glycosynthase approach.

Thirdly, the hitherto unknown gene for the published rhamnosidase Rha_{Hp}[171] was successfully identified and a predicted intron fragment was successfully removed from it. When the gene was expressed in *S. cerevisiae* and

K. lactis, no protein production was detected, either in intra- or in extracellular approaches. Only homologous intracellular expression in *H. polymorpha* resulted in production of the active enzyme, albeit with a low yield of 1 mg L^{-1}. It is possible that the intron removal decreased the efficiency of the rhamnosidase production in *H. polymorpha*; therefore, application of the wild type gene for the expression could be a further step in improving the expression efficiency. Although the crystallisation trials resulted in formation of several small plate-like crystals, anisotropic diffraction prevented acquisition of the crystal structure. Further screening and application of inhibitors, which could stabilise the structure, would be a possible approach to improve the crystals' quality. So far, only rhamnose was tested as an inhibitor, but no crystals were obtained in this screening. Since there are only three crystal structures of rhamnosidases available to date and there are significant differences between them, replacement of methionine by selenomethionine for anomalous dispersion phasing would also be a valuable help to completely solve the structure.

In summary, attempts have been made to get more insight into some rhamnosidases belonging to the GH78 family. With all the possible applications, there is a scientific need to increase the knowledge about the rhamnosidases, as well as make them more accessible for potential use. Identification of genes, new crystal structures, enzyme mutants with altered activity, all the issues addressed during this study, are required to fully unravel the potential of these enzymes.

Low yields or inactive enzyme were the main obstacles that prevented two out of three rhamnosidases used in this study from being applied for the glycosynthase approach. The experiment with improved expression of the shorter and therefore inactive version of the Rha_{La} when compared to expression of the full length protein suggests that in some cases the activity of the enzyme may be an obstacle in protein expression. On the other hand, both rhamnosidases with reported crystal structures were produced with good yields in *E. coli.*[163, 188] Therefore, there must be an additional factor(s) influencing the expression level, that could not be identified in this study. For future experiments it would be interesting to investigate a possible influence of a gene codon usage optimisation on the rhamnosidase expression level, as – to the best of my knowledge - this kind of modification was never tested with these enzymes.

None of the tested rhamnosidases could be secreted in the active form, neither when the three yeasts as eukaryotic hosts were applied. The rhamnosidase Rha_{Ba} was applied for mutagenesis towards a glycosynthase and the produced variants were tested with various donors and acceptors. The *p*NP β-D-glucopyranoside was not accepted as a substrate, indicating that the enzyme may be selective for rhamnose as a donor group, as the *p*NP α-rhamnopyranoside was hydrolysed with high specificity. Although the desired product could not be detected, some valuable observations can be taken from these attempts. The rhamnosidase could be easily produced in *E. coli* and stored as a lyophilisate for a long time. In addition, application of up to 20% of DMSO had only minor effect on the enzyme activity. The Rha_{Ba} is thus a promising enzyme for application in biocatalysis. As an outlook, application of more stable glycosyl group donors, e.g., azides, would be recommended to further investigate potential glycosynthase activity of the created mutants.

The rhamnosidase Rha_{Hp} was successfully produced in *H. polymorpha* and the protocols for expression and purification of this enzyme were established. The protein was applied for crystallisation trials and several conditions for crystal formation were identified. Consequently, an important step towards gaining more information about this enzyme class with only three available structures was made. These results are an encouraging basis for further experiments leading to the structure solution of Rha_{Hp}.

5. Materials and methods

5.1 Chemicals, used materials and software

All enzymes, materials and chemicals used in this study, unless specified, were commercially obtained from Alfa Aesar GmbH, Applichem, Carl Roth GmbH, Fermentas GmbH, Sigma Aldrich, Fluka, New England Biolabs, Analytik Jena, Life Technologies or Qiagen. The data were analysed and plotted with Microsoft Office 2010 tools and OriginPro 9.0 software; the graphs were prepared with CorelDRAW X6 software. The cloning and mutagenesis of the genes were designed with CloneManager9 and PrimerX[186] software. The protein structures were drawn with UCSF Chimera[197] software. The sequence alignments were performed with MultAlign[198] or Clustal Omega[189] online tools. The codon usage was analysed and optimised with Graphical Codon Usage Analyser[79] based on the information collected in Kazusa codon usage database.[199] The chemical structures were drawn using ChemBioDraw Ultra 12.0 software.

5.2 Strains

All strains used in this study are summarised in Table 13.

Table 13: Strains used in this study.

Strain		Genotype	Source	Ref.
E. coli	*DH5α*	F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17(r_K^- m_K^+), λ–	DSMZ, Braunschweig	[200]
E. coli	Mach1	*ΔrecA1398 endA1 tonA Φ80ΔlacM15 ΔlacX74 hsdR(r_K^- m_K^+)*	DSMZ, Braunschweig	[201]
E. coli	BL21(DE3)	*F^- ompT gal dcm lon hsdS_B(r_B^- m_B^-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*	DSMZ, Braunschweig	[202]
S. cerevisiae	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	EUROSCARF, Frankfurt am Main	[203]
K. lactis	GG799	MATα	New England Biolabs Ipswich, MA, USA	[204]
H. polymorpha	KLA 8-1	Ura3; leu2	ARTES Biotechnology GmbH, Langenfeld	[205]

5.3 Plasmids and primers

All the used primers were synthesised by Sigma Aldrich and are listed in Table 14.

Table 14: Primers used for this study.

Nr	Primer	Sequence (5' to 3')
Introduction of new common multicloning site (MCS) into the yeast plasmids		
1	Kl MCS fwd	TCGAGAAGAGAGAAGCTGACCTAGGCGCGCGCGCGTACGCGCC GGCGGCATCACCATCACCATCACTAGG
2	Kl MCS rev	GATCCCTAGTGATGGTGATGGTGATGCCGCCGGCGCGTACGCG CGCGCGCCTAGGTCAGCTTCTCTCTTC
3	Sc pYES fwd	AGCTTATGAAGAGAGAAGCTGACCTAGGCGCGCGCGTACGCGC CGGCGGCATCACCATCACCATCACTAGC
4	Sc pYES MCS rev	CATGGCTAGTGATGGTGATGGTGATGCCGCCGGCGCGTACGCG CGCGCCTAGGTCAGCTTCTCTCTTCATA
5	Hp MCS fwd	AGCTTGGATAAAAGAGAAGCTGACCTAGGCGCGCGCGCGTACG CGCCGGCGGCATCACCATCACCATCACTAGG
6	Hp MCS rev	GATCCCTAGTGATGGTGATGGTGATGCCGCCGGCGCGTACGCG CGCGCGCCTAGGTCAGCTTCTCTTTTATCCA
Sequencing of the plasmids		
7	pFPMT MFa seq fwd	CCGGCATTTTGAAGGTGAC
8	pFPMT MFa rev	TCTCTCAGAGAGGGGAATGG
9	piKL fwd	GTGAGCGGATAACAAGCTCAAC
10	piKL seq rev	CAAGACAATCGGCACTAATAACC
11	pSC seq fwd	CGTCAAGGAGAAAAAACCCCGGATCG
12	pSC seq rev	CACTGGCGGCCGTTACTAGTGG
13	pYES263 seq fwd	CGTCAAGGAGAAAAAACCCCGGA
14	pYES263 seq rev	CACTGGCGGCCGTTACTAGTGG
15	M13 fwd	GTAAAACGACGGCCAG
16	M13 reverse	CAGGAAACAGCTATGAC
17	pKL seq fwd	CATTGACTTAACCGGGGATG
18	pKL seq rev	GGAACGTCATTATCGCACAAG
Isolation of Rha_{Hp} gene and intron removal		
19	Ram1 fwd Hp	CGCCCTAGGCATGGTTGAGGTTAGAGGAGTTCGACTAGAG
20	Ram1 rev Hp	CGACATTTTGGAAGCCAGTTCAAGATCTCAGCAGTCTCCGAAG
21	Ram2 fwd Hp	GCTGAGATCTTGAACTGGCTTCCAAAATGTCGTTCTG
22	Ram2 rev Hp	CGGCGTACGAGCTGAGCACGTAAGAGACC
Introduction/removal of secretion signal in the shuttle vectors		
23	pKL_spacer_fwd	AGCTTCCATGGGCCGCC
24	pKL_spacer_rev	CTAGGGCGGCCCATGGA
25	pHP_ Mfa_fwd	AATTCGCCATGACTGCA
26	pHP_Mfa_rev	AGCTTGCAGTCATGGCG
27	QC_HindIIIfromMFa_fwd	CTGCTAAAGAAGAAGGGGTATCTTTGGATAAAAGAGAAGCTGAC
28	QC_HindIIIfromMFa_rev	GTCAGCTTCTCTTTTATCCAAAGATACCCCTTCTTCTTTAGCAG
29	piSC_ Mfa_fwd	GCAAGCTTATGAGATTTCCTTCAATTTTTACTGC
30	piSC_ Mfa_rev	CCATGCCTAGGTCAGCTTC

Table 14: Primers used for this study (continued).

Amplification of Rha_{La} and addition of the restriction sites for cloning		
31	Ram_La_fwd	CGCCCTAGGCATGGGTATGAAAATTACAAATATTTTAGTTAACC
32	Ram_La_rev	CGCCGTACGAAAGAGATTCAACTTTATAAAAATTGTTTTATG
Creation of Rha_{Ba} mutants		
33	RhaB_D567G_fwd	GACACGTTCGTCGGCTGCCCGTCCTATG
34	RhaB_D567G_rev	CATAGGACGGGCAGCCGACGAACGTGTC
35	RhaB_D567S_fwd	GGACACGTTCGTCAGCTGCCCGTCCTATG
36	RhaB_D567S_rev	CATAGGACGGGCAGCTGACGAACGTGTCC
37	RhaB_E572G_fwd	GCCCGTCCTATGGCCAGGTGTTCTGGG
38	RhaB_E572G_rev	CCCAGAACACCTGGCCATAGGACGGGC
39	RhaB_E572S_fwd	GACTGCCCGTCCTATAGCCAGGTGTTCTGGGTG
40	RhaB_E572S_rev	CACCCAGAACACCTGGCTATAGGACGGGCAGTC
41	RhaB_E841G_fwd	CCACGACCTGTTGGGGCATGTATCCGAACTTTG
42	RhaB_E841G_rev	CAAAGTTCGGATACATGCCCCAACAGGTCGTGG
43	RhaB_E841S_fwd	GATGCCACGACCTGTTGGAGCATGTATCCGAACTTTGC
44	RhaB_E841S_rev	GCAAAGTTCGGATACATGCTCCAACAGGTCGTGGCATC

The yeast plasmids modified for this study and their key features are listed in Table 15.

Table 15: Shuttle plasmids used for protein expression in yeast and introduced modifications.

Name	Host	Size (kb)	Homologous promoter	Selection marker in yeast	Modification	Original plasmid and source
piSC	*S. cerevisiae*	5.9	*GAL1*	*URA3*	New MCS / His_6Tag	pYES263 / EUROSCARF
pSC	*S. cerevisiae*	6.1	*GAL1*	*URA3*	MFα secretion signal	piSC
pKL	*K. lactis*	9.1	*LAC4*	amdS	New MCS / His_6Tag	pKLAC2 / NEB
piKL	*K. lactis*	8.9	*LAC4*	amdS	without MFα secretion signal	pKL
pHP	*H. polymorpha*	7.3	*FMD*	*URA3*	New MCS / His_6Tag	pFPMTMFa / ARTES Biotechnology GmbH
piHP	*H. polymorpha*	7.1	*FMD*	*URA3*	without MFα secretion signal	pHP

Plasmid pFPMTMFa, which was kindly provided by ARTES Biotechnology GmbH, contains all features of the vector pFPMT121[49] and a secretion signal- the MFα

prepro sequence with one introduced amino acid exchange D83E. Standard plasmids yeast vectors were modified by insertion of one common multicloning site (MCS): *AAG AGA GAA GCT* GACCTAGGCGCGCGCGTACGCGCCGGCGG **CAT CAC CAT CAC CAT CAC** TAG. It consists essentially of three elements: a cleavage site for the KEX2 protease (italics) followed by restriction sites for *AvrII* and *Pfl23II* (underlined), which were used to clone the CAL-A gene into each plasmid, and six histidine codons (in bold) plus stop codon, to facilitate the purification procedure *via* the resulting C-terminal His_6Tag. To insert this new MCS, additional flanking restriction sites were used: *HindIII* and *NcoI* for plasmid piSC, *XhoI* and *BamHI* for pKL, and *HindIII* and *BamHI* for pHP[98] All used plasmids are summarised in Table 16.

Table 16: Plasmids used in this study.

Nr.	Plasmid name	Properties	Source
empty plasmids			
1	pET28a+	PT7, lacZ', KanR	Novagen (USA)
2	piSC	GAL1, URA3, AmpR	this work
3	pSC	GAL1, URA3, AmpR	this work
4	pKL	LAC4, amdS, AmpR	this work
5	piKL	LAC4, amdS, AmpR	this work
6	pHP	FMD, URA3, AmpR	this work
7	piHP	FMD, URA3, AmpR	this work
source plasmids for the enzymes			
8	pUC57 RhaBa	pUC57 plasmid with rhamnosidase RhaBa gene	GenScript (USA)
9	pUC57 CalA	pUC57 plasmid with codon optimised CAL-A gene	GenScript (USA)
10	pHT::RhaLa	pHT plasmid with rhamnosidase RhaLa gene	Sebastian Schumacher
11	pCR Ram1	pCR plasmid with the first part of the RhaHp gene (till intron)	this work
12	pCR Ram2	pCR plasmid with the second part of the RhaHp gene (from after intron)	this work
13	pCR RhaHp	pCR plasmid with rhamnosidase RhaHp gene amplified from the *H. polymorpha* genome	this work
lipase CAL-A			
13	piSCL	piSC plasmid with ligated CAL-A gene	this work
14	pSCL	pSC plasmid with ligated CAL-A gene	this work
15	pKLL	pKL plasmid with ligated CAL-A gene	this work
16	pHPL	pHP plasmid with ligated CAL-A gene	this work
rhamnosidase Rha_{La}			
17	pET28a+::RhaLa	pET28a+ with ligated RhaLa gene	this work
18	pET28a+::sRhaLa	pET28a+ with ligated short version of RhaLa gene	this work
19	piSC RhaLa	piSC with ligated RhaLa gene	this work
20	piKL RhaLa	piKL with ligated RhaLa gene	this work
21	pHP RhaLa	pHP with ligated RhaLa gene	this work
22	piHP RhaLa	piHP with ligated RhaLa gene	this work
rhamnosidase Rha_{Ba}			
23	pET28a+::RhaBa WT	pET28a+ with ligated wild type RhaLa gene	this work
24	pET28a+::QC 1 RhaBa	pET28a+::RhaBa WT with introduced D567G mutation	this work
25	pET28a+::QC 2 RhaBa	pET28a+::RhaBa WT with introduced D567S mutation	this work
26	pET28a+::QC 3 RhaBa	pET28a+::RhaBa WT with introduced E572G mutation	this work
27	pET28a+::QC 4 RhaBa	pET28a+::RhaBa WT with introduced E572S mutation	this work
28	pET28a+::QC 5 RhaBa	pET28a+::RhaBa WT with introduced E841G mutation	this work
29	pET28a+::QC 6 RhaBa	pET28a+::RhaBa WT with introduced E841S mutation	this work
30	pET28a+::RhaBa DM	pET28a+::RhaBa WT with introduced D567G and E572G mutations	this work

Table 17: Plasmids used in this study (continued).

rhamnosidase Rha_{Hp}			
31	pET28a+::RhaHp	pET28a+ with ligated RhaHp gene	this work
32	piSC RhaHp	piSC with ligated RhaHp gene	this work
33	pKL RhaHp	pKL with ligated RhaHp gene	this work
34	piKL RhaHp	piKL with ligated RhaHp gene	this work
35	piHP RhaHp	pHP with ligated RhaHp gene	this work
36	pHP RhaHp	piHP with ligated RhaHp gene	this work

5.4 Media and solutions

All the media and stock solutions were prepared with deionised water unless otherwise specified.

5.4.1 Stock solutions and media for E. coli

Stock solutions

Kanamycin stock solution (1000×)

50 mg mL^{-1} kanamycin sulphate

Filter sterilised; stored at -20 °C.

Ampicillin stock solution (1000×)

100 mg mL^{-1} ampicillin sulphate

Filter sterilised; stored at -20 °C.

Media for *E. coli*

LB medium

10 g L^{-1} peptone

10 g L^{-1} NaCl

5 g L^{-1} yeast extract

20 g L^{-1} agar

TB medium

Solution A		Solution B	
12 g L^{-1}	peptone	2.31 g L^{-1}	KH_2PO_4
24 g L^{-1}	yeast extract	12.54 g L^{-1}	K_2HPO_4
4 g L^{-1}	glycerol		
800 mL	dH_2O	200 mL	dH_2O

Both solutions were autoclaved separately and mixed afterwards.

5.4.2 Media for yeasts

YPD complex medium

20 g L^{-1}	peptone
10 g L^{-1}	yeast extract
20 g L^{-1}	glucose
20 g L^{-1}	agar for plates

2×YPGal medium for expression

40 g L^{-1}	peptone
20 g L^{-1}	yeast extract
40 g L^{-1}	galactose

100×aminoacid stock solutions

2 mg mL^{-1}	histidine
6 mg mL^{-1}	leucine
2 mg mL^{-1}	methionine

Filter sterilised; stored at -20 °C.

SD minimal medium

6.7 g L^{-1}	yeast nitrogen base without amino acids, with NH_4
10 mL L^{-1}	100×AA stock solution necessary (addition after autoclaving)
20 g L^{-1}	glucose
20 g L^{-1}	agar for plates

pH was adjusted to 6.0 with NaOH. For *S. cerevisiae* methionine, histidine and leucine stock solutions were added (final concentrations: 20 µg mL^{-1} Met, 20 µg

mL^{-1} His, 60 µg mL^{-1} Leu), whereas for *H. polymorpha* only leucine addition (final concentration: 60 µg mL^{-1} Leu) was necessary.

Selective plates for *K. lactis* (YCB Agar Medium with 5 mM acetamide)

11.7 g L^{-1}	yeast carbon base medium powder
30 mL L^{-1}	1 M Tris/HCl pH 7.0
20 g L^{-1}	agar
10 mL L^{-1}	100×acetamide stock solution provided with the competent cells (addition after autoclaving)

Derepression medium for *H. polymorpha*

6.7 g L^{-1}	yeast nitrogen base without amino acids, with NH_4
10 mL L^{-1}	100×leucine stock solution (addition after autoclaving)
10 mL	glycerol

pH was adjusted to 6.0 with NaOH.

YNB-glucose medium for *H. polymorpha*

6.7 g L^{-1}	yeast nitrogen base without amino acids, with NH_4
10 ml L^{-1}	100×leucine stock solution (addition after autoclaving)
20 g L^{-1}	glucose

pH was adjusted to 6.0 with NaOH.

2×YNB expression medium for *H. polymorpha*

13.4 g L^{-1}	yeast nitrogen base without amino acids, with NH_4
10 mL L^{-1}	100×leucine stock solution (addition after autoclaving)
1 g L^{-1}	yeast extract
10 g L^{-1}	casamino acids
10 mL L^{-1}	glycerol

pH was adjusted to 6.0 with NaOH.

5.5 Microbiology methods

5.5.1 Preparation of chemically competent E. coli cells

The method was based on established protocol from Sambrook and Fritsch.[206] In this protocol, the cells' competency is induced chemically using rubidium and manganese cations, which increase the permeability of the membranes, allowing DNA to cross the cell envelope. Briefly, 100 mL of *E. coli* culture in LB medium of OD_{600}= 0.4-0.5 was centrifuged (10 min, 3000 g, 4 °C). The pellet was resuspended in 30 mL ice-cold TFBI solution and incubated on ice for 15 minutes, then centrifuged again (10 min, 3000 g, 4 °C). The pellet was gently resuspended in 2 mL of TFBII solution and aliquoted 50 µL per 1.5 mL vial. The aliquots were shock frozen in liquid nitrogen and stored at -80 °C.

TFBI		**TFBII**	
30 mM	KOAc	10 mM	MOPS
50 mM	$MnCl_2$	10 mM	RbCl
100 mM	RbCl	75 mM	$CaCl_2$
10 mM	$CaCl_2$	15 % (v/v)	glycerol
15% (v/v)	glycerol		
pH 5.8 with acetic acid; filter sterilised		pH 7.5 with NaOH; filter sterilised	

5.5.2 Transformation of competent E. coli cells

The competent cells, prepared as described in section 5.5.1, were thawed on ice; up to 8 µL of sample DNA (1 µL of Miniprep DNA, 5 µL of the ligation mixture or 8 µL of sample for QuikChange procedure) was added and incubated 25 min on ice. The heat shock (90 sec at 42 °C) was followed by cooling on ice for 2 min. 400 µL of LB medium was added to the tube and it was incubated at 37 °C for 1 h on a rotary shaker. 100 µL of the transformation mixture was plated on a selective LB agar plate with the appropriate antibiotic, the rest was centrifuged and the pellet was plated on another selective LB agar plate. The plates were then incubated overnight at 37 °C. As a control, one vial with sterile water added instead of DNA was prepared accordingly and plated onto a selective plate. When any colonies were obtained on the control plate, the whole procedure would be repeated with a fresh batch of competent cells.

5.5.3 Expression of rhamnosidase Rha_{Ba} in E. coli

The overnight preculture was inoculated from the *E. coli* BL21 strain containing pET-28a Rha_{Ba} plasmid in 5 mL LB medium supplemented with kanamycin. The main cultures were inoculated in 400 mL LB or TB medium in 3 L Fernbach shake flasks to starting OD_{600}=0.1 and incubated at 28 °C, 120 rpm. When the turbidity (OD_{600}) reached 0.5, the culture was supplemented with 0.1 mM isopropyl b-D-thiogalactopyranoside (IPTG), followed by incubation at 16 °C for 38 h. The cultures were then harvested *via* centrifugation (4 °C, 6000 g, 20 min) and the pellets were either directly used or stored at -20 °C.

5.5.4 Expression of rhamnosidase Rha_{La} in E. coli

The overnight culture was inoculated from the *E. coli* BL21 strain containing pET-28a Ram_{La} plasmid in 5 mL LB medium supplemented with kanamycin. The main cultures were inoculated in 100 mL TB medium in shake flasks to starting OD_{600}=0.1 and incubated at 30 °C, 120 rpm. When the turbidity (OD_{600}) reached 0.5, the culture was supplemented with 0.1 mM IPTG and further incubated till the cultures reached the stationary phase. The cells were then harvested *via* centrifugation (4 °C, 6000 g, 20 min) and the pellets were either used immediately or stored at -20 °C.

5.5.5 Preparation of chemically competent yeast cells

The method for preparation of competent cells and their transformation was modified from Dohmen *et al.*[207] In this protocol, exposure of intact yeast cells to polyethylene glycol (PEG 1000) allows the cells to uptake the DNA. In the following step, added single stranded DNA binds to the yeast cell wall, preventing plasmid DNA from doing so; it is therefore available for transformation.
Briefly, 50 mL of yeast culture in YPD medium of OD_{600}= 0.6-0.7 was centrifuged for 5 min, 2000 g, RT. The pellet was resuspended in 25 mL of solution A and centrifuged for 5 min, 2000 g, RT. Then it was resuspended again in 1 mL solution A and aliquoted 200 µL per 2 mL vial. The cells were frozen at -80 °C for at least 1 h before use.

Solution A

10 mM	bicine
1 M	sorbitol
5% (v/v)	DMSO
3% (v/v)	ethylene glycol

pH was adjusted to 8.35 with NaOH

5.5.6 Transformation of yeasts

The method for preparation of competent cells and their transformation was modified from Dohmen *et al.*[207] 5 µL of 10 µg µL^{-1} herring sperm DNA (Sigma Aldrich) and 0.1-5 µg sample DNA (max. 20 µL) were added to an un-thawed vial of competent cells prepared as described in section 5.5.5 and directly put on a shaker at 37 °C, 5 min, 500 g. 1.4 mL of solution B was then added and incubated for 1 h at 30 °C. The cells were centrifuged 5 min, 3000 g, RT and the supernatant was discarded. The pellet was resuspended in 1 mL of solution C and centrifuged 5 min, 3000 g, RT. After the supernatant was discarded, the pellet was resuspended in 100 µL of solution C and the suspension was transferred to a 15 mL Falcon tube containing 5 mL YPD. It was incubated at 30 °C for 5-6 h. After incubation, the cells were centrifuged at 3000 g, 5 min, RT. The pellet was resuspended in 100 µL of sterile water and plated onto selective plates. The plates were incubated at 30 °C (*S. cerevisiae, K. lactis*) or 37 °C (*H. polymorpha*) for 3-4 days until colonies were visible.

Solution	**B**	**Solution**	**C**
40%	PEG 100 (Carl Roth)	10 mM	bicine
0.2 M	bicine	0.15 M	NaCl
pH was adjusted to 8.35 with NaOH		pH was adjusted to 8.35 with NaOH	

5.5.7 Passaging and stabilisation of H. polymorpha transformants

The passaging and stabilisation protocol was provided by the ARTES GmbH company, based on a published method.[50] Passaging of *H. polymorpha* transformants is a procedure that allows gradual integration of the plasmid into the genomic DNA of the host. Up to 50 copies of plasmid per cell can be incorporated as a result. The optimal copy number, leading to the highest

expression levels, is protein dependent and must be tested separately for each enzyme. Therefore, usually a set of 36 transformants is used for the passaging and stabilisation procedure and tested for the protein production afterwards.

First, a set of 36 culture tubes containing 3 mL YNB-glucose each was inoculated with the colonies grown on the agar plate after transformation. The culture tubes were inoculated at 37 °C, 120 rpm for 48 h (passage 1). Every 48 h, 20 µL of culture from each culture tube was transferred to a fresh culture tube with the same medium. In total 8 rounds of passaging were performed, and afterwards stabilisation started. For this purpose, the same amount of culture was transferred to fresh culture tubes containing non-selective YPD medium (passage 9). After 48 h, this step was repeated (passage 10) and afterwards the final selection was inoculated again in the selective YNB-glucose medium (passage 11). When the passage 11 had grown, 3 µL of each culture were transferred and stored on a selective agar plate. Small scale expression in MTP format was performed to select the best-expressing clones. Alternatively, the stabilisation steps could be performed already after 3 steps of passaging to give the first hint on possible expression levels of the protein tested.

5.5.8 Enzyme expression in yeast in DWP format

The small scale expression of protein in yeast was performed in 96 deep well polypropylene plates, 2 mL, with flat bottom (Greiner Bio-one). The flat bottom and maximal medium volume up to 500 µL were essential parameters to provide enough oxygen supply. The plates were filled with 500 µL of medium and inoculated directly from the agar plates using pipette tips. The plates were then sealed with sterile Rotilabo® lids (Carl Roth) and incubated at optimal temperature - 30 °C (*S. cerevisiae, K. lactis*) or 37 °C (*H. polymorpha)*, 600 rpm for 72 h.

5.5.9 Media screening for H. polymorpha

Since the problem with purification of the produced protein using His_6Tag was identified (section 3.2.3), media screening was performed to overcome this obstacle. A possible explanation was overglycosylation of secreted protein, which can result in partially inaccessible His_6Tag. It was already reported in literature that media composition as well as culture stage has an influence on the produced

protein glycosylation;[95] hence media screening was performed to verify that thesis. 50 mL of each media (see Table 6, p. 40 for used media composition) in 250 mL baffled flasks were inoculated from the same precultures to the starting OD_{600}=0.05 and expression was performed in at 37 °C with 120 rpm for 72 h. After 72 h of expression, cultures were harvested *via* centrifugation (12000 g, 4 °C, 10 min) and the supernatants were stored on ice. Polypropylene columns (Bio-Rad) were packed with 1 mL Ni-NTA resin (Qiagen). The columns were equilibrated with LE buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0). 10 mL of sample was loaded on the column and after 1 mL went through the column, the flowthrough sample was collected. The activities in the starting sample and the flowthrough were then measured *via* the standard assay (section 5.7.7) and compared.

5.5.10 Expression of lipase CAL-A in yeasts

Each overnight culture was inoculated from the respective yeast strain containing in 5 mL YPD medium. The main cultures were inoculated in 250 mL of the optimal media (2×YPGal for *S. cerevisiae* and *K. lactis*, 2×YNB for *H. polymorpha*) in baffled shake flasks to a starting OD_{600}= 0.1 and incubated at 30 °C, 130 rpm until stationary phase of growth was reached. The cultures were then harvested *via* centrifugation (4 °C, 6000 g, 20 min) and the spent culture medium from the secreting yeasts was stored at 4 °C or used directly. CAL-A in spent culture medium was stable under these conditions for up to one week. The pellets from *S. cerevisiae* were either directly used or stored at -20 °C.

5.5.11 Intracellular expression of rhamnosidase Ram_{Hp} in H. polymorpha

An overnight culture was inoculated from the transformant grown on an agar plate in 5 mL YPD medium. The main cultures were inoculated in 300 mL of the 2×YNB medium in 2 L baffled shake flasks to starting OD_{600}=0.1 and incubated at 30 °C, 130 rpm for 72 h. The cultures were induced with 1% methanol after 24 and 48 h. The cultures were then harvested *via* centrifugation (4 °C, 6000 g, 20 min) and the pellets were either stored at -20 °C or directly used.

5.5.12 Estimation of the cell number in the culture

The size of the cell varies during its growth, depending on several factors, including media composition, culture stage and protein production. It was important to keep the study as comparative as possible, therefore in addition to the OD_{600} measurement, which is a standard culture observation tool, but does not reflect variation in cell size between the strains, the measurement of the relationship between the OD_{600} measurement and the exact cell number at the stationary growth phase of the yeast cultures was performed. The measurement was carried out with the help of a Multisizer™ 3 COULTER COUNTER® (Beckman Coulter Inc., USA) automated system. The device uses the so-called Electrical Sensing Zone Method to provide number, volume, mass and surface area size distributions of the cells in the sample. The cells are diluted in weak electrolyte solution and drawn through a small aperture which separates two electrodes that have an electronic current flowing between them. When a cell passes through the aperture, it displaces its own volume of electrolyte, which then increases the impedance of the aperture. This change generates a pulse, which is directly proportional to the volume of the cell.[208] This method allows obtaining a size distribution from the sample and, as a consequence, the total cell number.

Three samples were taken from the three CAL-A producing yeast cultures in stationary growth phase and diluted to OD_{600}=0.1 in the respective optimal media. The probes were then 20×diluted in CASY buffer (commercially available buffer used for cell number analysis) and the particles were calculated in the diameter ranging from 115 till 901 µm. As a control, used media without the cells were also diluted in the same way and measured to exclude the false positive signals; however no signal was obtained for them. A size distribution was obtained for the yeasts. Most of the *S. cerevisiae* cells had a diameter in the range of 345-408 µm; for *K. lactis* it was in the range of 326-345 µm, whereas for *H. polymorpha* wide range of cell diameters between 190 and 280 µm was observed. The obtained number of cells per mL culture diluted to OD_{600}=1 were 7.7×10^6 cells mL^{-1} for *S. cerevisiae*, 5.2×10^6 for *K. lactis,* and 21.1×10^6 for *H. polymorpha*.

5.6 Molecular biology methods

5.6.1 Plasmid isolation from E. coli

Plasmid DNA isolation from *E. coli* was performed with a commercially available kit [innuPREP DNA Mini Kit (Analytik Jena)] according to the producer's protocol. In this method, alkaline lysis of cells is followed by precipitation of debris. Plasmid DNA, which was released to the solution in the process, binds to the silica resin in presence of high ionic strength buffer. The impurities are then removed with ethanol, and pure DNA is released from the column by decreasing ionic strength of the elution buffer.

5-10 mL of overnight *E. coli* precultures in LB medium were used for plasmid isolation procedure; the obtained DNA was eluted in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0) and stored at -20 °C.

5.6.2 Genomic DNA isolation from yeasts

The yeasts were grown on respective agar plates. The cells were scratched off from the agar plate, using a pipette tip, to form a small ball, which was resuspended in 200 µL of solution A. After addition of glass beads (ø 0.5 mm), the tube was vortexed for 2 min to disrupt the cells. 100 additional µL of solution A was added and the DNA was extracted *via* Roti Phenol twice, followed by chloroform/isoamyl alcohol extraction (24:1). Finally, the DNA was precipitated with 3 M sodium acetate pH 5.0 /ethanol absolute (25:1) at -20 °C overnight. The solution was then centrifuged (15 min, 4 °C, 15000 g), the pellet was washed with 70% ethanol and dried at RT. The dry pellet was resuspended in 40 µL of RNAse in TE buffer for at least 30 minutes at RT. The isolated genomic DNA was stored at 4 °C.

Solution A		**RNAse in TE buffer**	
100 mM	NaCl	10 mM	TRIS/HCl buffer pH 7.5
10 mM	Tris/HCl buffer pH 8.0	1 mM	EDTA
1 mM	EDTA pH 8.0	0.33 mg mL^{-1}	RNAse A
0.1%	SDS		

5.6.3 Determination of DNA concentration

The concentration of the DNA samples was determined spectrophotometrically at 260 nm, using NanoDrop 2000c. 1 µL of the sample was used for this purpose.

5.6.4 Agarose gel electrophoresis

For the analysis of the DNA samples or the separation of different DNA fragments, agarose gel electrophoresis was applied. 0.8% or 1% agarose in TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8.0), supplied with 0.01% GelRed™ (Biotium Inc., Hayward, USA), was used for the gel preparation. Samples, prepared with addition of 5× Loading Dye (100 mM EDTA, 43% glycerol, 0.05% bromophenol blue), were applied on the gel and the electrophoresis was performed at 180 V for 25 min. The bands were visualised with UV light and the size was estimated in comparison to the marker (**GeneRuler™ 1 kb DNA ladder** (Fermentas): 250, 500, 750, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 5000, 6000, 8000, 10000 bp).

5.6.5 DNA extraction from agarose gel

Extraction of DNA samples from agarose gel was performed with a commercially available kit (Gel extraction kit, Qiagen GmbH) according to the producer's protocol.

5.6.6 Enzymatic digestion and ligation of DNA fragments

The enzymatic digestion of DNA was performed with restriction endonucleases from commercial sources (Fermentas GmbH or New England Biolabs) according to the producer's protocol; for double digestion purposes, the supplier's online tools were used to identify the most suitable conditions.

After digestion, the samples were separated on an agarose gel and the desired fragments were cut out from the gel and purified as described in 5.6.5. The ratios of the volumes required for the ligation reaction were calculated on the basis of the obtained concentrations and length of the used fragments with the formula:

$$V_{\text{insert}} = 3\frac{V_{\text{vector}} * c_{\text{ vector}} * l_{\text{ insert}}}{c_{\text{ insert}} * l_{\text{ vector}}}$$

The final reaction was prepared in ligase buffer (Fermentas GmbH) and 1 µL of T4 DNA ligase was added. Ligation was performed for 1 h at RT; in challenging ligations, prolonged incubation at 4 °C overnight was applied.

5.6.7 Polymerase chain reaction PCR

Polymerase chain reaction (PCR) was applied for the amplification of the genes from different sources, as well as for the introduction of mutations. This technique applies thermostable, high-fidelity polymerases to selectively amplify the target DNA. Typical PCR reaction consists of several steps. In the first step, the two strands of the DNA double helix are physically separated at a high temperature. In the second step, the temperature is lowered, allowing added primers to bind to the DNA; the DNA strand becomes a template for DNA polymerase. The amount of target DNA is then amplified in repeated cycles of heating and cooling.

5.6.7.1 Colony PCR from *E. coli* cells

Colony PCR is a quick and reliable method, which allows to quickly screen for plasmids and their inserts directly from *E. coli* colonies. Usually after transformation, the plates were incubated overnight at 37 °C. On the next day, three to five colonies were selected and marked on the plate. Colonies were sampled with a sterile pipette tip in 15 µL of sterile water by pipetting up and down several times, and a small quantity of cells was transferred into a PCR mix, which was prepared as presented in Table 18. To release the DNA from the cells, the PCR was started with an extended heating time of 12 min and 30 sec at 95 °C (Table 19). After the reaction was completed, aliquots of 10 µL were analysed on an agarose gel together with marker to check the presence of amplicon of the expected size. Identified hit colonies were then used to amplify and isolate the plasmid, which was afterwards sequenced to verify the correct sequence.

Table 18: PCR mix composition for colony PCR. For bigger amount of tested colonies, a Mastermix can be prepared with amplified first four compounds pipetted together and then 9.6 µL of it aliquoted into reaction tubes. In this case, the cells suspension from each colony is added at the end separately.

PCR mix for colony PCR (volumes for one colony)	
5 µL	REDTaq® ReadyMix™ (Sigma Aldrich)
0.3 µL	100 µM forward primer
0.3 µL	100 µM reverse primer
4 µL	sterile H_2O
0.4 µL	cell suspension

Table 19: Temperature profile of the colony PCR method.

Step	**Temperature [°C]**	**Length**	**Cycles**
Cell disruption & denaturation	95	12 min 30 sec	
Denaturation	95	30 sec	×35
Annealing	55	30 sec	
Extension	72	60 sec/kb	
Final elongation	65	260 sec	
Hold	12		

5.6.7.2 Colony PCR from yeast cells

Colony PCR can be also applied to verify the presence of the gene of interest in yeast after transformation. In fact, in case of *K. lactis*, it is also possible to verify, if the gene was integrated into the genome at the desired position; this allows quick screening for the transformants, which should be able to produce the protein, without the need of expression. First attempts to use the same protocol as for *E. coli* cells failed, also prolonged heat disruption of cells did not result in any amplicon. Surprisingly, the solution for this problem was to decrease the initial thermal denaturation step. As a result, a new protocol could be established and successfully applied for all yeast strains used. The PCR mix used was as presented in Table 18 and the program used is summarised in Table 20. After the reaction was completed, aliquots of 10 µL were analysed on an agarose gel together with the marker to check for the presence of amplicon of the expected size.

Table 20: Colony PCR program adapted for the yeast cells.

Step	Temperature [°C]	Length	Cycles
Cell disruption & denaturation	95	5 min	
Denaturation	95	30 sec	×30
Annealing	55	30 sec	
Extension	72	60 sec/kb	
Final elongation	72	5 min	
Hold	12		

This method was also used to verify the integration of the gene of interest in the desired position into the genome in *K. lactis*. For this purpose the primers 1 and 2 from the commercial expression kit (*K. lactis* Protein Expression Kit Instruction Manual, version 1.0, 5/09, New England Biolabs) were used.

5.6.7.3 Overlap extension PCR

The overlap extension PCR method enables large insertions or deletions in the plasmid of interest.[209] The potential difficulty arises from the fact, that the efficiency of this method depends on the sequence of the primers, which is strictly defined by the gene sequence. The protocol consists of three PCR steps (Figure 41).

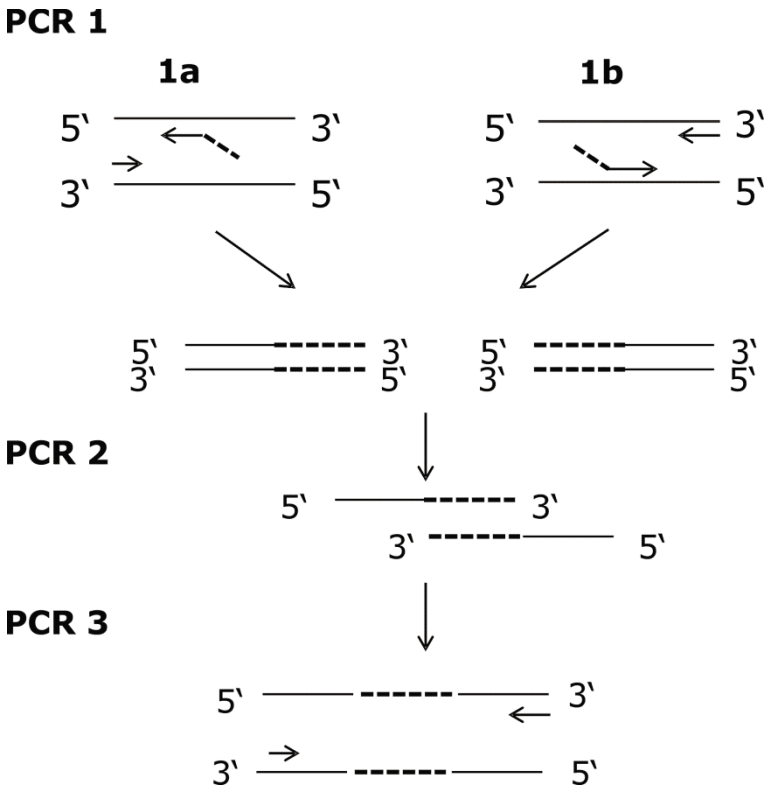


Figure 41: Schematic illustration of the overlap extension PCR method. The introduced mutation is indicated by dashed line and the primers are represented by arrows. The figure was based on Heckman *et al.*[210]

In the first step (PCR 1) two parts of the Rha_{Hp} gene were amplified in two separated PCR reactions: from the beginning to the intron (1a) and from after the intron till the end of the gene (1b), with the primer-introduced complementary fragment common to both of them, which is indicated by dashed line in Figure 41. The method was tested for intron removal in rhamnosidase Rha_{Hp} gene.

Table 21: PCR composition for the oePCR- PCR 1a and 1b.

oePCR:PCR 1a and 1b	
25 µL	REDTaq® ReadyMix™ (Sigma Aldrich)
1 µL	100 µM forward primer (see Tab. 16)
1 µL	100 µM reverse primer (see Tab. 16)
1 µL	DNA template (pCR Ram1 or Ram2)
22 µL	dd H_2O

Table 22: Temperature profile of the oePCR method, PCR 1.

Step	**Temperature [°C]**	**Length**	**Cycles**
Initial denaturation	95	5 min	
Denaturation	95	1 min	
Annealing	50-55	45 sec	×30
Extension	72	2 min	
Final elongation	72	5 min	
Hold	12		

Both fragments were then isolated using agarose gel electrophorese, digested with *DpnI* to avoid potential template contamination, and annealed in the subsequent PCR 2 reaction.

Table 23: PCR composition for the oePCR- PCR 2.

oePCR:PCR 2	
30.5 µL	dd H_2O
5 µL	*Pfu* polymerase reaction buffer
1 µL	dNTPs
2.5 µL	DMSO
5 µL	PCR 1a
5 µL	PCR 1b
1 µL	*Pfu* polymerase

Table 24: Temperature profile of the oePCR method, PCR 2.

Step	**Temperature [°C]**	**Length**	**Cycles**
Initial denaturation	94	2 min	
Denaturation	94	45 sec	
Annealing	50	30 sec	×10
Extension	72	3 min	
Final elongation	72	5 min	
Hold	12		

The product of the PCR 2 reaction was then directly used for PCR 3, in which the whole annealed gene was further amplified.

Table 25: PCR composition for the oePCR- PCR 3.

oePCR:PCR 3	
24.5 µL	dd H_2O
5 µL	*Pfu* polymerase reaction buffer
1 µL	dNTPs
2.5 µL	DMSO
10 µL	PCR 2 product
2 µL	PCR 1a product
2 µL	PCR 1b product
1 µL	Ram1 fwd Hp primer (see Tab. 16)
1 µL	Ram2 rev Hp primer (see Tab. 16)
1 µL	*Pfu* polymerase

Table 26: Temperature profile of the oePCR method, PCR 2.

Step	**Temperature [°C]**	**Length**	**Cycles**
Initial denaturation	94	2 min	
Denaturation	94	45 sec	×30
Annealing	50	45 sec	
Extension	72	3 min	
Final elongation	72	5 min	
Hold	12		

The PCR 3 reaction product was then ligated into pCR2.1 plasmid (TOPO® TA Cloning® Kit, Life Technologies, USA) according to manufacturer's protocol and sequenced, but in spite of optimisation only the first part of the whole gene was always obtained.

5.6.7.4 Mutagenesis *via* Megaprimer method

Megaprimer method is a modification of the QuikChange mutagenesis, which is an alternative to the oePCR method, similarly allowing for large insertions or deletions in the gene sequence.[187] Since the intron removal with oePCR did not work, it was tested for this purpose. As presented in Figure 42, the method consists of two PCR reactions and subsequent removal of the template. In the first PCR reaction, second part of the RhaHp gene was amplified with 5' overhangs complementary to the end of the first part (780 bp product).

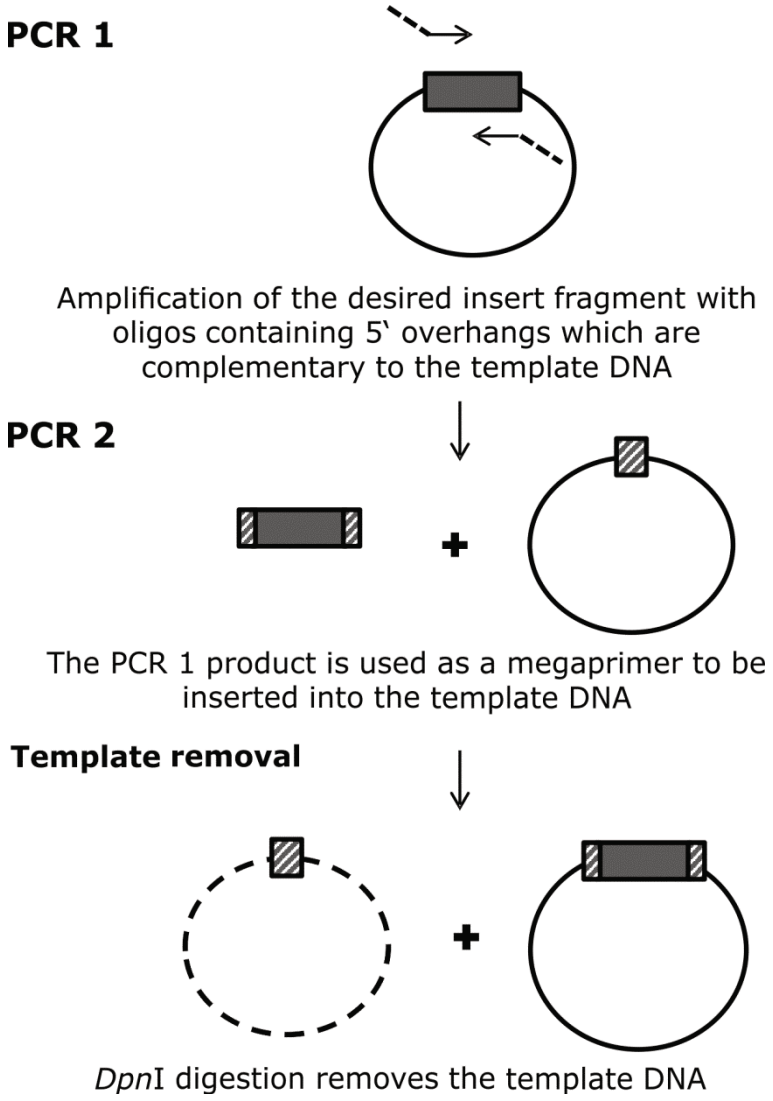


Figure 42: Schematic illustration of the Megaprimer mutagenesis PCR method. The intact plasmid is represented by the circle, whereas the dashed line indicates *DpnI* digested plasmid. Figure modified from Application Note from Agilent Technologies.[187]

Table 27: PCR composition for the Megaprimer PCR 1.

PCR 1	
25 µL	REDTaq® ReadyMix™ (Sigma Aldrich)
1 µL	100 µM Ram2fwd primer (see Tab. 16)
1 µL	100 µM M13 rev primer (see Tab. 16)
1 µL	DNA template (pCR Ram2)
22 µL	dd H_2O

Table 28: Temperature profile of the Megaprimer PCR method, PCR 1.

Step	Temperature [°C]	Length	Cycles
Initial denaturation	95	10 min	
Denaturation	95	45 sec	×30
Annealing	50-55	45 sec	
Extension	72	1 min	
Final elongation	72	5 min	
Hold	12		

The product was then isolated using agarose gel electrophorese, extracted from gel, eluted in dd H_2O and used for the second PCR reaction.

Table 29: PCR composition for the Megaprimer PCR 2.

PCR 1	
5 µL	10×*Pfu* polymerase reaction buffer
1 µL	dNTPs
3 µL	PCR1 product
1 µL	DNA template (pCR Ram1)
39 µL	dd H_2O
1 µL	*Pfu* polymerase

Table 30: Temperature profile of the Megaprimer PCR method, PCR 2.

Step	Temperature [°C]	Length	Cycles
Initial denaturation	95	30 sec	
Denaturation	95	30 sec	×5
Annealing	52	45 sec	
Extension	68	8 min	
Denaturation	95	30 sec	×13
Annealing	55	1 min	
Extension	68	8 min	
Hold	12		

The reaction product was then expected to be a mixture of the template DNA and amplified plasmid containing the insert. *DpnI* digestion was performed to remove the template DNA; after heat inactivation of the enzyme (80 °C, 20 min) 3 µL and 8 µL of the reaction were used for *E. coli* transformation. Colonies obtained on the next day were then tested for the presence of the insert in the plasmid.

5.6.7.5 Cassette mutagenesis

The method of cassette mutagenesis was used to modify the existing plasmids *via* the introduction of a new multicloning site between two existing restriction sites. To prepare the insert, two complementary oligonucleotides were designed, which were annealed to create the cassette with sticky ends. Both oligonucleotides were diluted 10× in 1×T4 ligase buffer and equal volumes of

10 µL each were mixed in a PCR tube. The sample was then incubated in the thermal cycler and the program was set as presented in Table 31.

Table 31: Temperature profile applied for the cassette mutagenesis method.

Temperature [°C]	Length
90	5 min
T_M	20 min
T_M -5	120 min
24	hold

The plasmid was cut with the respective restriction enzymes according to the manufacturer's protocol, and the cut fragment was analysed on an agarose gel and purified. The ligation of the plasmid and the cassette was performed as described in section 5.6.6 with an insert: plasmid ratio of 5 pM: 0.2 pM in which 25 to 50 ng of the plasmid were used.

5.6.7.6 Mutagenesis *via* QuikChange method

As already introduced in the section 4.1.3 (p.79) QuikChange PCR (QC PCR) was the preferred method for the introduction of point mutations or restriction sites into the gene of interest. A pair of complementary primers, containing the desired mutation, was designed using the PrimerX online tool. The T_M of the primers was calculated using Clone Manager software. The primers were then annealed to the original plasmid and elongated using a high-fidelity polymerase (*Phusion* or *Pfu* polymerases were used as an alternative, since in certain cases the desired product could be obtained just with one of them, while the other did not result in any amplicon; optimised protocols applied for each of them are presented separately). As a result, the whole plasmid was amplified with the introduced mutation. The template plasmid was then digested with the enzyme *DpnI*; after this step, only modified plasmid should be present in the sample, which was then used for *E. coli* MACH1 competent cells transformation to obtain it in high concentration.

QC PCR with *Phusion* polymerase

3 sample reactions with different concentrations of the salts and DMSO were prepared to identify the optimal conditions for each reaction (Table 32) and the cycling parameters presented in

Table **33** were used.

Table 32: PCR composition used for the QuikChange with *Phusion* polymerase.

Sample reaction component	1	2	3	
dNTP mix	1	1	1	µL
5×HF buffer	5	5	10	µL
~ 50 ng template plasmid	x	x	x	µL
10×diluted forward primer (100 µM)	1	1	1	µL
10×diluted reverse primer (100 µM)	1	1	1	µL
DMSO	-	1	-	µL
Phusion® polymerase	0.5	0.5	0.5	µL
ddH_2O	to the final volume of 50			µL

Table 33: Temperature profile applied for the QuikChange with *Phusion* polymerase.

Step	Temperature [°C]	Length	Cycles
Initial denaturation	98	30 sec	
Denaturation	98	10 sec	×17
Annealing	T_M -5	30 sec	
Extension	72	60 sec/kb	
Final elongation	72	260 sec	
Hold	12		

Each reaction sample was added 1 µL of *DpnI* and incubated at 37 °C for 2 h. The *DpnI* was then inactivated via incubation at 80 °C for 20 min. 3 and 8 µL of the reaction solution was then used for transformation of the *E. coli* competent cells.

QC PCR with *Pfu* polymerase

If the QC PCR reaction with *Phusion* polymerase had failed, the *Pfu* polymerase was tested using reaction composition as presented in Table 34 and the cycling parameters as in Table 35.

Table 34: PCR composition used for the QC reaction with Pfu polymerase.

Sample reaction component		
dNTP mix	1	µL
10×Pfu reaction buffer	5	µL
~ 50 ng template plasmid	x	µL
10×diluted forward primer (100 µM)	1	µL
10×diluted reverse primer (100 µM)	1	µL
Pfu polymerase	1	µL
ddH_2O	to the final volume of 50	µL

Table 35: Temperature profile applied for the QuikChange with *Pfu* polymerase.

Step	Temperature [°C]	Length	Cycles
Initial denaturation	95	30 sec	
Denaturation	95	30 sec	
Annealing	T_M -5	60 sec	×17
Extension	68	24 min	
Final elongation	68	40 min	
Hold	12		

To each reaction sample 1 µL of *DpnI* was added and incubated at 37 °C for 2 h. The *DpnI* was then inactivated *via* incubation at 80 °C for 20 min. 3 and 8 µL of the reaction solution was then used for transformation of the *E. coli* competent cells.

5.7 Biochemical methods

5.7.1 E. coli cell disruption

E. coli pellets were disrupted *via* pulsed sonication (Ultrasonic Desintegrator Sonoplus HP 2070, Bandelin, Berlin). Cells were first resuspended as 20% (w/v) solution in Lysis buffer (100 mM TRIS/HCl pH 7.5, 10 mM β-mercaptoethanol) or in WB buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0), if the produced enzyme was stable under these conditions. The sonication was performed on ice for 3×30 seconds, 50 % pulsation cycles, 40% power with 3 min intervals to cool the sample down. It was then centrifuged for 15 min, 4 °C, 12000 g and the supernatant was collected for purification or activity assay.

5.7.2 Yeast cell disruption

Yeast pellets were disrupted *via* pulsed sonication (Ultrasonic Desintegrator Sonoplus HP 2070, Bandelin, Berlin). Cells were first resuspended as 20% (w/v) solution in Lysis buffer (100 mM TRIS/HCl pH 7.5, 10 mM β-mercaptoethanol) or directly in WB buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0), if the produced enzyme was stable under these conditions. The sonication was performed on ice for 5× 4 min, 50 % pulsation cycles, 40% power with 3 min intervals to cool the sample down, as observation under microscope confirmed that the cells were effectively disrupted after this time. The sonified sample was then centrifuged for 15 min, 4 °C, 12000 g and the supernatant was collected for purification or activity assay.

5.7.3 Protein purification with ÄKTA

The expression cultures were harvested at stationary phase *via* centrifugation at 4 °C, 12000 g for 20 min. For *S. cerevisiae* the pellets were then disrupted as a 20% (w/v) suspension in buffer WB (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) using either glass beads and vortexing or pulsed sonication (Ultrasonic Desintegrator Sonoplus HP 2070, Bandelin, Berlin) as described in section 5.7.2. After centrifugation, the crude cell extract was filtered with a 0.45 µM filter (Sartorius Stedim Biotech S.A., France) and applied to a prepacked 5 mL Ni-NTA column (Protino®, Macherey Nagel)

connected to an Äkta™ purifier device (GE Healthcare). Elution was performed with buffer EB (50 mM NaH_2PO_4, 300 mM NaCl, 500 mM imidazole, pH 8.0). The collected fractions containing the protein of interest were then desalted over a PD-10 column (Sephadex™G-25M, GE Healthcare) and eluted with Sørensen buffer[211] pH 8.0 to remove imidazole. For *K. lactis* or *H. polymorpha* the collected spent culture medium was first concentrated using a Vivacell 250 mL device with a 10 kD exclusion membrane (Sartorius Stedim Biotech S.A., France), then 50 mL of the retentate was replenished by 200 mL buffer WB and concentrated again. Subsequently, 50 mL of the final sample was filtered (0.45 µM pore size) and applied to a 5 mL Ni-NTA column followed by purification as described above. Finally, the obtained enzyme sample was analysed for purity *via* a silver-stained SDS-PAGE gel.

5.7.4 Batch purification of the protein

Batch purification with the fresh Ni-NTA slurry was used for test purifications of the protein from smaller culture volumes as well as to assure the highest purity of the purified protein. 1 mL of the 50% Ni-NTA slurry (Qiagen) was loaded onto an empty column (GE Healthcare) with capped bottom outlet. After the slurry settled, the cap was removed and the column was equilibrated with 2×5 mL WB buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0). The protein sample, prepared as for purification with ÄKTA (section 5.7.3) was then applied on the column. In subsequent steps, the column was washed with 2×5 mL WB buffer, then 2×5 mL WB 30 buffer (50 mM NaH_2PO_4, 300 mM NaCl, 30 mM imidazole, pH 8.0) and the protein was eluted with 3 mL of buffer EB (50 mM NaH_2PO_4, 300 mM NaCl, 500 mM imidazole, pH 8.0). The eluate was then desalted over a PD-10 column (Sephadex™G-25M, GE Healthcare) and eluted with respective optimal buffer. Finally, the obtained enzyme sample was analysed for purity *via* a silver-stained SDS-PAGE gel.

5.7.5 Size Exclusion Chromatography

Size Exclusion Chromatography was used to achieve the required purity of the rhamnosidase Rha_{Hp} sample, which was the used for crystallisation trials, as described in section 4.2.3.3. In this chromatographic method, proteins applied

onto a column are separated by their size, which allows analysing the sample's content and removing the impurities.

The protein obtained after IMAC purification was concentrated down to 1 mL and applied onto prepacked Superdex™ 200 10/300 GL column (GE Healthcare) connected to an Äkta™ purifier device (GE Healthcare). The column was preequilibrated with 20 mM MOPS buffer, pH 7.0 and the separated protein fractions (4 mL) were then eluted in this buffer with constant 1 mL min^{-1} flowrate. The highest protein peak was observed in fractions 8 and 9; aliquots from the fractions containing protein peaks were then analysed *via* a silver-stained SDS-PAGE gel and an activity test, which confirmed presence of an active rhamnosidase in fractions 8 and 9. The fractions were then used directly for the crystallisation.

5.7.6 Measurement of protein concentration

The protein concentration was measured either spectrophotometrically at 280 nm as described for CAL-A[212] or by the Bradford procedure[213] using bovine serum albumin as the calibration standard.

5.7.7 Lipase activity assay with standard substrate

The activity of CAL-A was measured with a photometric assay at 410 nm absorbance and 30 °C in 150 µL final volume in microtiter plates (GENios Microplate Reader, Tecan, Switzerland) using *para*-nitrophenol (*p*NP) octanoate (**1**, Figure 43, p. 139) as the standard substrate in a final concentration of 0.33 mM. During the reaction, the enzyme releases chromogenic product *p*-nitrophenol with absorbance at 410 nm. Accurate measurement of enzyme activity requires that the 4-nitrophenol product is fully deprotonated, therefore basic pH 8.0 was used. The reaction mix contained 50 mM Sørensen buffer pH 8.0, 5 mM sodium deoxycholate, and 0.1% gum arabic.[214] Assay reactions were initiated by the addition of 30 µL CAL-A solution (purified or the spent culture medium). In the standard assay, reaction measurements were taken every 30 sec for 10 min. One unit of CAL-A activity was defined as the amount of enzyme that releases 1 µmol *para*-nitrophenolate per minute under the given conditions. All samples were assayed in triplicates.

5.7.8 Thermostability analysis of lipase CAL-A

CAL-A was purified as described in section 5.7.3. For the thermostability analysis, the samples from the three hosts were diluted to the same concentration (0.016 mg mL^{-1}) and each sample was then divided into 6 aliquots. Five of them were incubated at various temperatures for 30 minutes, and the control one was stored at 4 °C during this time. After incubation time the samples were tested in a photometric assay with *p*NP octanoate as the standard substrate (see section 5.7.7 for details). The obtained results were then normalised to the activity of this control sample, which was taken as 100%.

5.7.9 Kinetic characterisation of lipase CAL-A

Kinetic constants were determined by measuring the activity of lipase CAL-A at different substrate concentrations of *p*NP octanoate and fitting the obtained data to the Michaelis-Menten equation, with application of the least square fitting method, using OriginPro 9.0 software. All measurements were performed in the following conditions: Sørensen buffer pH 8.0, 5 mM sodium deoxycholate, and 0.1% gum arabic[214] at 30 °C. All samples were assayed in triplicates.

5.7.10 Lipase activity assay with bulky substrates

CAL-A activity on chiral esters was tested with the *para*-nitrophenolates of (*S*)-/(*R*)-2-methyldecanoic acid (**2**, Figure 43)[215], (*S*)-/(*R*)-2,3-dihydro-1*H*-indan-2-carboxylate (i.e. indan-2-carboxylate, **3**)[216], (*S*)-/(R)-ibuprofen (**4**), and (*S*)-/(*R*)-naproxen **5**. [74, 217]

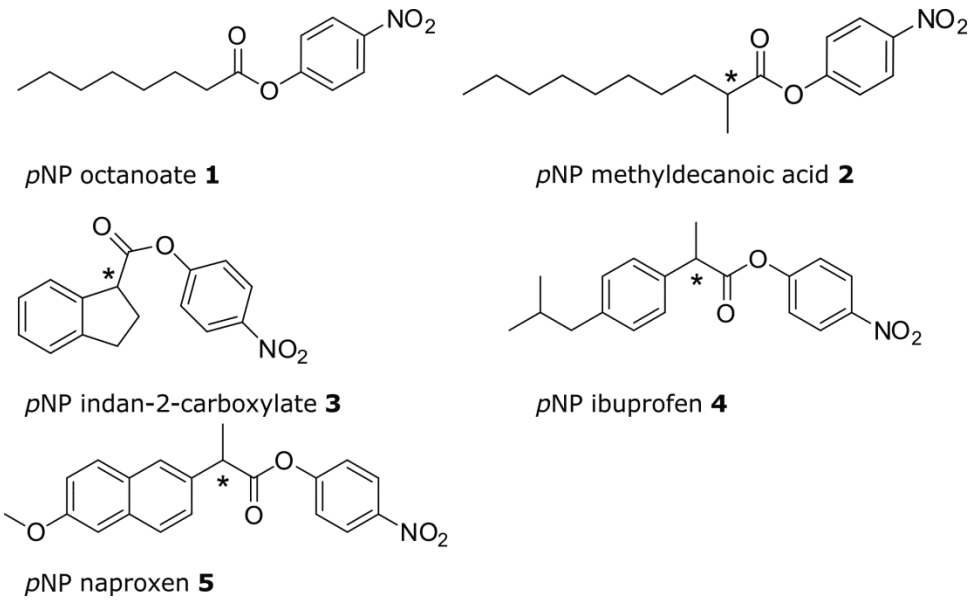


Figure 43: *Para*-nitrophenol esters used for the substrate spectra screening.

The respective enantiomers of the indicated esters were used instead of the standard substrate in the same concentration. The measurements with these substrates were conducted for 1 h and carried out in duplicates.

5.7.11 Glycosylation analysis on SDS-PAGE

The glycosylation level of CAL-A was analysed using SDS-PAGE. Deglycosylation was performed with enzymes: PNGase F from *Flavobacterium meningosepticum* (*N*-glycosylation) or *O*-glycosidase from *Enterococcus faecalis* (*O*-glycosylation), commercially available from New England Biolabs. The deglycosylation reactions were performed according to the manufacturer's protocol. For the non-denaturing deglycosylation of the active CAL-A variants, the samples were also prepared according to the New England Biolabs protocol and incubated for 18 h at 37 °C before analysis on SDS-PAGE (see section 5.7.15 for details).

5.7.12 Glycoprotein estimation

The carbohydrate content of lipase CAL-A secreted from *K. lactis* and *H. polymorpha* was analysed by using a Glycoprotein Carbohydrate Estimation Kit (Pierce Biotechnology). The assay was performed in microplate format according to the manufacturer's protocol. 50 µL of purified protein (used concentration 0.25 mg mL^{-1}) was transferred in the wells and 25 µl of 10 mM

sodium meta-periodate was added. In the assay, the carbohydrate groups were first oxidated for 10 minutes with sodium meta-periodate to produce detectable aldehyde groups. After 10 min incubation, 150 µl of 0.5% Glycoprotein Detection Reagent were added to the wells and incubated for 1 h at RT. Although no information about this reagent is provided by the supplier, it can be assumed that it is a Schiff-type reagent, commonly used in a qualitative test for aldehydes. In this test, the unknown sample is added to the decolourised (sulfonated) Schiff reagent; when aldehyde is present, the free, uncharged aromatic amine groups of the reagent react with the aldehyde being tested to form aldimine groups, which then react with bisulfite to form the product. The solutions appear coloured due to the visible wavelength absorbance of product's central quinoid structure.

The resulting purple reaction product was measured with a spectrophotometer at 550 nm and the values obtained for the tested proteins were compared to the standard proteins with known carbohydrate content: lysozyme bovine serum albumin, ovalbumin, apo-transferrin, fetuin, and α_1-acid glycoprotein.

5.7.13 Rhamnosidase activity assay with standard substrate

Rhamnosidase activity was measured using the commercially available substrate *para*-nitrophenyl rhamnopyranoside (Sigma Aldrich), in which the *p*-nitrophenyl group is attached to the 1 position of rhamnose, as reported by Romero *et al.*[181] During the reaction, the enzyme releases chromogenic product *p*-nitrophenol with absorbance at 410 nm. Briefly, 30 µL of the tested protein sample (crude extract or purified rhamnosidase) was mixed with 50 mM MOPS buffer, pH 7.0 and 1.7 mM substrate (final concentration) in 150 µL final volume at 410 nm absorbance and 30 °C in microtiter plates (GENios Microplate Reader, Tecan, Switzerland). In the standard assay, reaction measurements were taken every 30 sec for 15 min. One unit (U) was defined as the amount of enzyme that liberates 1 µmol nitrophenol min^{-1}. All samples were assayed in triplicates.

5.7.14 Rhamnosynthase activity assay

A potential glycosynthase activity of the rhamnosidase mutants was tested using glycosyl fluorides as donors, as they are most widely used for this purpose.

All reactions were performed in 1 mL total volume. 200 µL of 100 mM donor [α-L-rhamnosyl fluoride (7) or β-D-glucosyl fluoride (8), see Figure 32, p. 89 for the structures] and 200 µL of 100 mM acceptor [*p*NP β-D-glucopyranoside 11] were added to 600 µL 50 mM KP_i buffer pH 7.0 and 0.5 mg mL^{-1} of lyophilised enzyme variant and incubated at 30 °C, 300 rpm. After 1 h, the reaction progress was analysed *via* TLC (EE:MetOH = 9:1).

5.7.15 SDS-PAGE

Polyacrylamide gel electrophoresis (PAGE) is a technique to separate macromolecules according to their electrophoretic mobility. A strong detergent, SDS, is added during sample preparation as component of Loading Dye and binds to polypeptides in a constant weight ratio of 1.4 g SDS per g of polypeptide, contributing negative charge. The electrophoretic mobilities of these proteins are therefore a linear function of the logarithms of their molecular weights.

The gels were usually prepared directly before use. If not immediately used, they could be stored at 4 °C, covered with wet paper towel to avoid drying out, for up to two weeks.

Table 36: Composition of the working solutions used for the SDS-PAGE method. The solutions were stored at RT.

1×SDS buffer	**Stacking gel solution**		**12% running gel solution**	
25 mM Tris/HCl pH 8.8	125.5 mL	ddH_2O	143.7 mL	glycerol
192 mM glycine	52.6 mL	0.5 M Tris/HCl pH 6.8	102.2 mL	1.5 M Tris/HCl pH 8.8
0.1% (w/v) SDS	21.9 mL	10% (w/v) SDS	4.14 mL	10% (w/v) SDS

Table 37: Solutions composition for the preparation of one SDS gel.

Running gel		**Stacking gel**	
1.99 mL	acrylamide/bisacrylamide	0.3 mL	acrylamide/bisacrylamide
2.25 mL	12% running gel solution	1.71 mL	stacking gel solution
7.3 µL	TEMED	5 µL	TEMED
18.4 µL	10% (w/v) ammonium peroxodisulphate	20.5 µL	10% (w/v) ammonium peroxodisulphate

The protein samples were prepared with 5× Loading Dye (10% SDS, 30% saccharose, 0.1% Bromophenol blue, 0.5 M Tris/HCl, pH 6.8, 50 mM DTT) and heated for 10 minutes at 95 °C. The samples were loaded on the gel and 100 V was applied for concentration in the stacking gel, and the separation was

performed at 180 V. The size of the protein bands was estimated as compared to the protein marker Roti-Mark 10-150 protein ladder (Carl Roth); 10,20, 30, 40, 60, 80, 100, 150 (kD) or PageRuler Prestained Protein Ladder (Fermentas); 10, 15, 25, 35, 40, 55, 70, 100, 130, 170 (kD).

5.7.16 Silver staining

For visualisation of the protein bands separated with the SDS-PAGE method, when a low amount of the protein was expected or to analyse the purity of purified protein, the silver staining method was applied. All following incubations were performed on an orbital shaker. The gel was first incubated in the Fixer for 10 min, washed with water and incubated in Washer for 10 min. During this time, 0.2% (v/v) formaldehyde was added to the Silver nitrate solution and the Developer solution. The Washer solution was discarded and the gel was incubated in Silver nitrate solution for 10 minutes.
Finally, Silver nitrate solution was discarded, the gel was washed with water and the Developer solution was added. When the desired intensity of the staining was obtained, 10 mL of the Stop solution were added to stop the reaction.

Table 38: Solutions used for the silver staining of the SDS-PAGE gels.

Fixer		**Washer**	
30% (v/v)	ethanol	30% (v/v)	ethanol
30% (v/v)	acetic acid	0.5% (w/v)	sodium acetate
		0.2% (w/v)	$Na_2S_2O_3$

Silver nitrate		**Developer**		**Stop solution**	
0.1% (w/v)	silver nitrate	2.5% (w/v)	Na_2CO_3	2.3 M	citric acid

5.7.17 Coomassie staining

Coomassie staining method was used for visualisation of the protein bands separated with the SDS-PAGE method, when a higher concentration of protein was applied. All the following incubations were performed on an orbital shaker. The gel was first incubated in used destaining solution for 10 min, then in staining solution for 10 min. In the following step, the gel was incubated again in the used destaining solution to remove remaining staining solution. Finally, the

gel was incubated for 10 minutes in fresh destaining solution (which was then stored for further use as used destaining solution). The destaining procedure was then completed in water after overnight incubation.

Table 39: Solutions used for the Coomassie staining of the SDS-PAGE gels.

Staining solution		Destaining solution	
50% (v/v)	methanol	50% (v/v)	methanol
10% (v/v)	acetic acid	10% (v/v)	acetic acid
		0.1% (w/v)	Coomassie Brilliant Blue

5.7.18 Western Blot

Western Blot method was used to detect even small quantities of His_6tagged protein in the samples. After the SDS-PAGE, the gel was disassembled and carefully transferred onto a PVDF membrane that had been previously soaked in transfer buffer. The blotting was performed for 1 h at 30 V on ice. The membrane was then blocked with 5% low fat milk powder in TBST solution for 20 min and afterwards incubated with an Anti-His antibody coupled to horseradish peroxidase (Anti-His HRP, Carl Roth GmbH) diluted 1:500 in TBST solution for 1 h. After washing with TBST (3× 5 min), the membrane was added 2 mL of the chromogenic substrate for the peroxidase, BM blue POD precipitating substrate (Roche) and developed in the dark until the bands were visible. The size of the protein bands was estimated as compared to the protein marker Roti-Mark 10-150 protein ladder (Carl Roth); 10, 20, 30, 40, 60, 80, 100, 150 (kD).

Table 40: Solutions used for the Western Blot.

Transfer buffer	TBST
25 mM Tris	10 mM Tris/HCl pH 7.5
192 mM glycine	154 mM NaCl
10% methanol	0.1% Tween 20

5.7.19 Colony Lift Assay

Colony Lift Assay was performed in order to test the ability of the *H. polymorpha* transformants to secrete the Rha_{Hp} rhamnosidase as explained in section 4.2.3.2. 5 agar plates with YNB medium supplemented with 2% (w/v) agar and various pHs ranging from 3.0 to 7.0 were prepared. 36 transformants of *H. polymorpha* after the passaging procedure were transferred to the plates using pipette tips. As controls, a transformant with an empty plasmid and a

transformant expressing a high level of His-tagged protein (CAL-A) were used. The plates were incubated overnight at 37 °C. On the next day, a circular piece of nitrocellulose (NC) filter in the size of the petri dish was carefully placed on each of them and incubated for 24 h at RT. After incubation, the membranes were removed and the remains of the cells were washed away with water. The membranes were then further processed in the same way as standard membranes in Western Blot (see section 5.7.18 for details).

5.7.20 Concentration of the protein samples

For the concentration of the protein samples in small volumes up to 20 mL, a Vivaspin 20 centrifugal concentrator (Sartorius AG) with 10 kD membrane was used. The sample was centrifuged at 8000 g, 4 °C, until the desired protein concentration was obtained.
For the bigger volumes from 50 to 250 mL, a Vivacell 250 device (Sartorius AG) with 10 kD membrane was applied. A pressure of 3 bar was applied and the sample was concentrated on an orbital shaker at 4 °C until the desired protein concentration was obtained.

5.7.21 Desalting of the protein samples

For the removal of imidazole after IMAC purification or for the buffer exchange, PD-10 columns (matrix: Sephandex G-25, GE Healthcare) were applied. The column was first equilibrated with 3×3.5 mL of the target buffer, then 2.5 mL of the sample was loaded on the column and the protein was eluted with 3.5 mL of the target buffer.

5.7.22 Lyophilisation of the protein samples

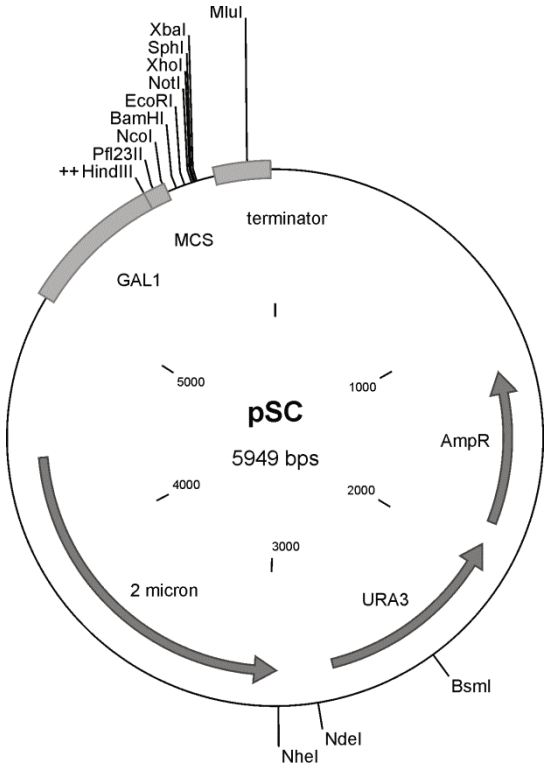
Lyophilisation was performed for long time storage of proteins without significant loss of activity. This method of storage is not optimal for all the proteins and therefore required activity tests for each protein to elucidate whether it remained active after the procedure.
The protein sample was first frozen in liquid nitrogen for 5 min and then left in the Zirbus VaCo 2 freeze-drier (ZIRBUS technology GmbH, Bad Grund/Harz) at RT, 3 mbar for 3 days to assure complete drying. The dried protein powder was then stored at -20 °C.

6. Additional figures and tables

6.1 Genes & plasmid maps

The standard yeast plasmids (pYES263 plasmid for *S. cerevisiae,* pKLAC2 for *K. lactis* and pFPMT121 for *H. polymorpha*) were modified by insertion of one common multicloning site (MCS): AAG AGA GAA GCT GA *CCTAGG* CGCGCG*C GTACG* CGCCGGCGG CAT CAC CAT CAC CAT CAC **TAG**. It consists essentially of three elements: a cleavage site for the KEX2 protease followed by restriction sites for *AvrII* and *Pfl23II* (italics), which were used to clone the gene of interest into each plasmid, and six histidine codons plus stop codon (in bold), to facilitate the purification procedure *via* the resulting C-terminal His_6Tag. To insert this new MCS, additional flanking restriction sites were used: *HindIII* and *NcoI* for plasmid piSC, *XhoI* and *BamHI* for pKL, and *HindIII* and *BamHI* for pHP.

6.1.1 Plasmid for intracellular expression in S. cerevisiae



> Introduced modification to the original pYES263 plasmid: new MCS sequence

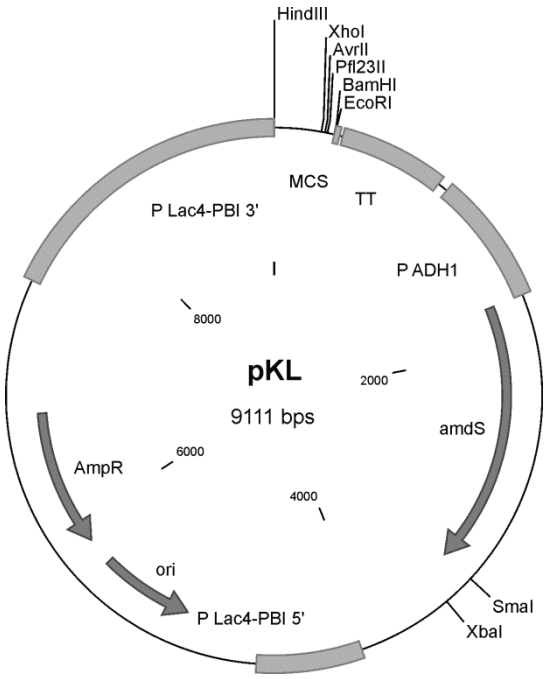
HindIII M K R E *AvrII* MauBI *Pfl23II* *MreI*

AAGCTT ATG AAG AGA GAA GCT GA*CCTAGG*CGCGCG*CG TAC G* CG CCG GCG

His_6Tag **STOP** *NcoI*

G CAT CAC CAT CAC CAT CAC **TAG** *CCA TGG*

6.1.2 Plasmid for extracellular expression in K. lactis



> Introduced modification to the original pKLAC2 plasmid: new MCS sequence

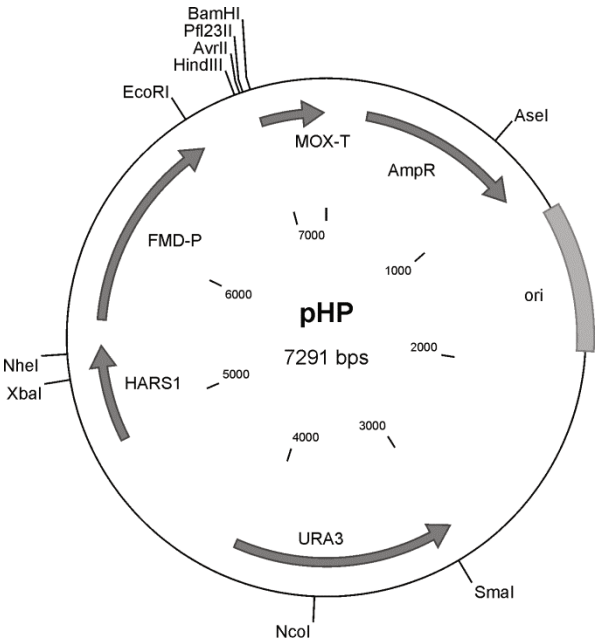
XhoI K R E A *AvrII* MauBI *Pfl23II* MreI

CTC GAG AAG AGA GAA GCT GA*C CTA GG*C GCG CG*C G TAC G* CG CCG GCG

His$_6$Tag **STOP** *BamHI*

G CAT CAC CAT CAC CAT CAC **TAG** *GGA TCC*

6.1.3 Plasmid for extracellular expression in H. polymorpha



> Introduced modification to the original pFPMT121 plasmid: new MCS sequence

HindIII D K R E A *AvrII* MauBI *Pfl23II* *MreI*

AAGCTT GGAT AAG AGA GAA GCT GA*C CTA GG*C GCG CG*C G TAC G CG CCG GCG*

His$_6$Tag **STOP** *BamHI*

G CAT CAC CAT CAC CAT CAC **TAG** *GGA TCC*

6.1.4 Codon optimised CAL-A gene sequence

> Synthetic CALA gene sequence (1359 bps)

ATGGCTCCAGCTACTGAAACTTTGGATAGAAGAGCTGCTTTGCCAAACCCATATGATGA
TCCATTCTATACTACTCCATCTAACATTGGTACTTTCGCTAAGGGTCAAGTTATTCAATCC
AGAAAGGTTCCAACTGATATTGGTAACGCTAACAACGCTGCTTCTTTCCAATTGCAATAT
AGAACTACTAACACTCAAAACGAAGCTGTTGCTGATGTTGCTACTGTTTGGATTCCAGCT
AAGCCAGCTTCTCCACCAAAGATTTTCTCTTATCAAGTTTATGAAGATGCTACTGCTTTG
GATTGTGCTCCATCTTATTCTTATTTGACTGGTTTGGATCAACCAAACAAGGTTACTGCT
GTTTTGGATACTCCAATTATTATTGGTTGGGCTTTGCAACAAGGTTATTATGTTGTGTCTT
CTGATCACGAAGGTTTCAAGGCTGCTTTCATTGCTGGTTATGAAGAAGGTATGGCTATTT
TGGATGGAATTCGAGCTTTGAAGAACTATCAAAACTTGCCATCTGATTCTAAGGTTGCTC

TCGAGGGTTATTCTGGTGGAGCTCACGCTACTGTTTGGGCGACGTCTTTGGCTGATTCT
TATGCTCCAGAATTGAACATTGTTGGTGCTTCTCACGGTGGTACACCGGTTTCTGCTAAG
GATACTTTCACTTTCTTGAACGGTGGTCCATTCGCTGGTTTCGCTTTGGCTGGTGTTTCT
GGTTTGTCGCTAGCTCACCCAGATATGGAATCTTTCATTGAAGCTAGATTGAACGCTAA
GGGTCAACAAACTTTGAAGCAAATTAGAGGTAGAGGTTTCTGTTTGCCACAAGTTGTTTT
GACTTATCCATTCTTGAACGTTTTCTCTTTGGTTAACGATACTAACTTGTTGAACGAAGCT
CCAATTGCTGGTATTTTGAAGCAAGAAACTGTTGTTCAAGCTGAGGCTTCTTATACCGTC
TCTGTTCCAAAGTTCCCAAGATTCATTTGGCACGCTATTCCAGATGAAATTGTTCCATAT
CAACCAGCTGCTACTTATGTTAAGGAACAATGTGCTAAGGGTGCTAACATTAACTTCTCT
CCATATCCAATTGCTGAACACTTGACTGCTGAAATTTTCGGACTAGTTCCATCTTTGTGG
TTCATTAAGCAAGCCTTCGATGGTACTACTCCAAAGGTTATTTGTGGTACTCCAATTCCA
GCTATTGCTGGTATTACTACTCCATCTGCTGATCAAGTTTTGGGTTCTGATTTGGCTAAC
CAATTGAGATCTTTGAACGGTAAGCAATCTGCTTTCGGTAAGCCATTCGGTCCAATTACT
CCACCACGTACGCGCCGGCGGCATCACCATCACCATCAC

> CAL-A protein sequence (453 aas)

MAPATETLDRRAALPNPYDDPFYTTPSNIGTFAKGQVIQSRKVPTDIGNANNAASFQLQYR
TTNTQNEAVADVATVWIPAKPASPPKIFSYQVYEDATALDCAPSYSYLTGLDQPNKVTAVL
DTPIIIGWALQQGYYVVSSDHEGFKAAFIAGYEEGMAILDGIRALKNYQNLPSDSKVALEGY
SGGAHATVWATSLADSYAPELNIVGASHGGTPVSAKDTFTFLNGGPFAGFALAGVSGLSLA
HPDMESFIEARLNAKGQQTLKQIRGRGFCLPQVVLTYPFLNVFSLVNDTNLLNEAPIAGILK
QETVVQAEASYTVSVPKFPRFIWHAIPDEIVPYQPAATYVKEQCAKGANINFSPYPIAEHLTA
EIFGLVPSLWFIKQAFDGTTPKVICGTPIPAIAGITTPSADQVLGSDLANQLRSLNGKQSAF
GKPFGPITPPRTRRRHHHHHH

6.1.5 Rhamnosidase Rha_{Ba} sequence

> Rha_{Ba} gene sequence (2868 bps)

ATGGCAGGCAGGAATTGGAACGCTTCATGGATTTGGGGAGGACAAGAGGAGAGTCCGC
GCAACGAGTGGCGGTGCTTCCGGGGCAGCTTCGACGCGCCTGCGTCGGTCGAGGGAC
CGGCCATGCTTCATATAACGGCGGATTCGCGATACGTACTGTTCGTGAACGGCGAGCAA
GTGGGGAGAGGCCCCGTGCGCTCCTGGCCGAAGGAGCAGTTTTACGATTCGTACGACA
TCGGCGGGCAGCTGCGCCCGGGCGTCCGCAATACGATCGCGGTGCTGGTGCTTCATTT
CGGCGTGTCGAACTTTTATTACTTGCGCGGACGCGGCGGGCTGATCGCCGAGATCGAA
GCCGATGGCCGCACGCTTGCGGCGACGGATGCCGCATGGCGGACGGAGCGGCTGGGC
GGACAGCGTTCCAATTCCCCTCGGATGGCCTGCCAGCAGGGATTCGGGGAAGTCATCG
ACGCGCGCGAGCTGGCGGAAGACTGGGCCCTTCCGGCGTTCGACGACGGCGGCTGGG
CGCAAGCCCGATCGATCGGACCCGCAGGCACGGCGCCCTGGACCTCGCTCGTTCCGCG
CGATATTCCTTTTTTGACGGAAGAAAAGCTGTATCCCGCCTCGATCCAGTCGCTTAGCC
GGGTGAAGGCGCCCAAGTACGCAGCCGCGCTGGATCTGCGCAATCAAATGGTGCCCGA
GAGCGTTAACCATGCGAACCCCGTCTCCTACTGCGGTTATGTGGCGACGATCCTCACGC
TCGAGACAAGCGGCGTCGTCACGCTCGGATTCCCGACCGGCGTACGGGGGAGCGGCG
TATGGGTCGACGGCGTCCTGCAAACCGAGTGGACGGGCGTGCAGCCGGAGCGATATTA
CAGCTTGAACCTCGCTGCAGGCGAGCATCTGGTTCTTGTCGATATTACGAGCTCCGATC
ATGGCGGCAGCAGCCATTTTGCCATCGACAGCGAAGCGGCGTTCACGCTGCGCTCGCC
GGCCGGCGACAATGGCGTGCCGCTGGCGACGATCGGTACGTTCGACCAGTCCGAATAC
ATCGATCACCGCCCGGGCAGACGGATGCAGACGGACCATCCGGATTATCGGGCGCTGC
CGGAAGCCGCGCCTACCGCCGCCGCGCTTGAAGCGTTCGCTTCCTGGGTCAAGCCGTT
CGAGCCTTCGCTCTATACGGAGGAAAACGTGTTCGGATCTAACGTATGGCGAACGCTCG
CCGAACGCAGGGCGGTGCCGAGATCCGTTCTGAACGCGATATTGCCGGTTCCCGAGCC
GGGCGTCCTGCCCGTATTCGAGGACGGAGACTGCGAGCTCGTCATCGATCTGGGCGCG
GAGCGCTCCGGGTTTATCGGCTTCGAGCTCGAAGCGCCCGCCGGTACGATCATCGATG
CCTATGGCGTCGAATATATGAGAGAAGGCTACACGCAGCACACGTACGGGCTCGACAA
CACGTTCCGCTATATTTGCCGCGAAGGCAGACAATCGTACGTGTCCCCCGTGCGCCGC
GGTTTCCGGTACTTGTTCCTGACCGTCAGAGGCAATAGCGCGCCGGTGAAGCTGCACG
AGATTTATATCCGCCAGAGCACGTACCCGGTCGCGGAACAAGGCAGCTTCCGCTGCTC
CGACGCGCTGCTGAACGCAACGTGGGAGATCAGCAGGCACACCACCAGATTGTGCATG
GAGGACACGTTCGTCGACTGCCCGTCCTATGAACAGGTGTTCTGGGTGGGCGACAGCC
GCAACGAGGCGCTGGTCAACTATTACGTTTTCGGCGAGACCGAGATCGTGGAGCGCTG
CCTGAATCTCGTGCCGGGCTCGGCGGACGAGACGCCGCTTTATCTCGACCAGGTGCCG
AGCGCATGGAGCAGCGTCATCCCGAACTGGACGTTCTTCTGGATACTCGCTTGCCGGG

AATATGCCGCGCATACGGGCAACGAAGCCTTCGCCGCCCGCATCTGGCCTGCGGTGAA
GCACACGCTGACGCATTATTTAGAACACATCGACGACAGCGGCCTGCTGAACATGGCG
GGTTGGAATCTGCTGGACTGGGCGCCGATCGACCAGCCGAACGAAGGCATCGTCACCC
ACCAGAACCTGTTCCTCGTCAAGGCGCTTCGGGATTCGCGGGCGCTTGCCGCCGCGGC
CGGCGCAACCGAAGAGGCAGACGCGTTCGCGGCGCGCGCCGACCTTCTGGCCGAGAC
GATCAACGCGGTATTGTGGGACGAGGAAAAGCGCGCTTATATCGATTGCATCCACGCG
GACGGGCGCCGTTCGGACGTATACAGCATGCAGACGCAGGTCGTCGCTTATCTGTGCG
GGGTTGCGCAGGGCGAACGCGAAGCCGTCATCGAAGGCTACCTGTCGTCCCCGCCGC
CGGCTTTCGTACAGATCGGCAGCCCGTTTATGTCGTTTTTCTACTACGAGGCGCTCGAG
AAGGCCGGCCGTCAAACGCTGATGCTCGACGACATCCGCCGCAATTACGGCCAGATGC
TGCGCTACGATGCCACGACCTGTTGGGAGATGTATCCGAACTTTGCGGAAAATCGCAGC
AACCCGGACATGCTGACCCGCAGCCATTGCCATGCATGGTCGGCGGCGCCGGGCTATT
TCCTGGGCTCGAGCATTCTCGGCGTCAAGCGGGGAGCTGACGGGTGGCGAACCGTCG
ATATCGCGCCGCAGCCTTGCGATCTAACCTGGGCCGAAGGCGTCGTGCCGCTGCCGCA
GGGCGGTCACATCGCGGTGAGCTGGGAGTTCGTATCCGCCGGCAAGCTGAAGCTGAG
AATCGAGGCGCCGGAGGATATCGAGGTGAACGTGACGCTGCCCGAAGGAATAGAAGG
CGAAGTGACGCAGGTTAAGTATATGAGC

> Rha_{Ba} protein sequence (956 aas)

MAGRNWNASWIWGGQEESPRNEWRCFRGSFDAPASVEGPAMLHITADSRYVLFVNGEQ
VGRGPVRSWPKEQFYDSYDIGGQLRPGVRNTIAVLVLHFGVSNFYYLRGRGGLIAEIEADG
RTLAATDAAWRTERLGGQRSNSPRMACQQGFGEVIDARELAEDWALPAFDDGGWAQAR
SIGPAGTAPWTSLVPRDIPFLTEEKLYPASIQSLSRVKAPKYAAALDLRNQMVPESVNHANP
VSYCGYVATILTLETSGVVTLGFPTGVRGSGVWVDGVLQTEWTGVQPERYYSLNLAAGEHL
VLVDITSSDHGGSSHFAIDSEAAFTLRSPAGDNGVPLATIGTFDQSEYIDHRPGRRMQTDH
PDYRALPEAAPTAAALEAFASWVKPFEPSLYTEENVFGSNVWRTLAERRAVPRSVLNAILPV
PEPGVLPVFEDGDCELVIDLGAERSGFIGFELEAPAGTIIDAYGVEYMREGYTQHTYGLDNT
FRYICREGRQSYVSPVRRGFRYLFLTVRGNSAPVKLHEIYIRQSTYPVAEQGSFRCSDALLN
ATWEISRHTTRLCMEDTFVDCPSYEQVFWVGDSRNEALVNYYVFGETEIVERCLNLVPGSA
DETPLYLDQVPSAWSSVIPNWTFFWILACREYAAHTGNEAFAARIWPAVKHTLTHYLEHID
DSGLLNMAGWNLLDWAPIDQPNEGIVTHQNLFLVKALRDSRALAAAAGATEEADAFAARA
DLLAETINAVLWDEEKRAYIDCIHADGRRSDVYSMQTQVVAYLCGVAQGEREAVIEGYLSS
PPPAFVQIGSPFMSFFYYEALEKAGRQTLMLDDIRRNYGQMLRYDATTCWEMYPNFAENRS
NPDMLTRSHCHAWSAAPGYFLGSSILGVKRGADGWRTVDIAPQPCDLTWAEGVVPLPQG
GHIAVSWEFVSAGKLKLRIEAPEDIEVNVTLPEGIEGEVTQVKYMS

6.1.6 Rhamnosidase Rha_{La} sequence

> **Rha_{La} gene sequence** (2793 bps) (*HindIII* restriction site, where the shorter version was cut, is marked in grey)

ATGGGTATGAAAATTACAAATATTTTAGTTAACCAAATGGAACATCCTCTAGGCTTTGAT
TTGAGTAATTTAAGGATTACTTTTGAATTGACAGAGATGGAAAATATAATAGGAAACGTT
TATAAGAATATATCTGTCGGTAAAGTTGAATCTGAACAGCCGATTTATTTTGAGCCTGAT
GAATTATACGAGAATAACGCTTTTAAAATAAACATGGAGTTAGAGCCTCGTACTAAGTAT
TGGGTAAAAATCGGAGTAAGAAATGATAACGAAGTTACGTCTAGTAATACTTGGTTTGA
AACAGGGAAAATGGATGAAAAGTTTTATGGAAAGTGGATAACCAATAAAAAAGACGTAG
AAAATACATTATTTAAAAAAGATTTTGAGTTAGCAAATAAGCAGATAAAAAGTGCAAGAT
TATATTCAACCACATTAGGGGTGTATGAAGTTGATTTAAATGGAGTAAAAGTTGGAAATG
AATTCTTGGCGCCTGGTTTTACTAATTATGACAAAATCGTACAATTGCAAACTTATGATG
TTACAAAATTAGTAACAAAAAATAGTAATAATGAATTGGTATTTTCTGTTGGAGATGGTT
GGTATAAAGGAAATCTAGGCTTTGATGGCGGCCAAACTAATATCTATGGTGATAAAAAA
TCAATTTTAGCAGAACTTCATGTTACGTATACTGATAATAGTGAACAAGTCATTAGTACA
GATTCAAGTTGGTTAACAACTGAAGGAAAAATAATTAAATCATCAATTTATTATGGTGAG
GATATTGATGATACAAAAGATATTTTAGACTGGTCATCGGTAGTCATTCTTAATAAATCT
ACTAGTATTGTTAGAGATCGCTTAAGTTTACCAATAATGAAAAAAGAGGTATTGAAGGTT
AAAGAAATCATTCATACTCCTAAGAATGAAATAGTATTAGATTTTGGCCAAAATCATGCT
GGATGGCCTGTATTTATTAACCGACTTGCACGTGGAAAGAAAATTACGTTACAAATGGG
GGAAATACTACAAGATGGTAATTTTTATAACAAAAATCTTAGGTTAGCTCGTGCAGCATT
TACTTATATTTCAGATGGTGAAGAAAAATTAATACGACCACATTTTACGTATTTTGGATTT
AGGTATGTAAAAATCTCTGGAGTAACTGATGTGAATAAGGATGATTTTGAATCGTGGGT
ATTATATTCTGATCTAAAACAAACTGGATTTATTAAAACTAATAATGATAAAGTAAATCGT
CTTTTTAAGAATGTAATTTGGGGACAAAAAAGTAATTTCATGGATGTGCCAACTGATTGT
CCGCAAAGGGATGAGCGGCTGGGGTGGACAGGAGACGCTGAAATATTTGCACCAACTG
CCAGTTTCAATATGAATACTTATGAGTTTTACAAGAAGTATGCTAAAGATATGCTTGTTG
AACAAGAAGATAACAAAGGAATGCTTCCTATTATTGTACCTAGTTTAAAACAAAAAAGTA
CTGGAATGGCAATTTGGAGTGATGCTGCTACCATCATTCCTTGGGTTACTTATCGATTTT
TTGACGATTTAGGAGTTTTGAAGCAAAATTATTCTCAAATGAAAAATTGGGTTGATTGGA
TTACCCAAAATACTAAAACCAAATACTTATGGATTGGACAAATGCAATTAGGAGATTGGC
TTTCCTTGGATAATGGGGCTAATCCTCAGGGCAAAACTAACGAAGATTATATTGCATCTA
TTTATTATTTTGTTTCTGCAAGCATAGTATCTAAAGCAGCTAGATTGCTTCATTACGATAT
GGAGTCAGACTATTATGAAAATTTGGCTAGAAATATAAAGACAAATATTTTAAATGAATT
TGTTACTGAAAAAGGTAGGATAGCCATTGATACTCAGACAGCATTGGTTTTAGCGTTACA

TTTCGGATTAGTGCATGATTATCAAAAGTCACAGGTTGTAGCTGATTTAGTGAAAAAAGT
AAAAGATGACAATAAGCATTTACAGACTGGATTTGTTGGGACACCATTTTTATTGTCGGT
ATTATCAAATAATAATCAGCATCATTTGGCTATGGATATTTTTATGCAAGAAGATTGCCC
AAGTTGGCTTTATGAAGTAAATATGGGAGCAACCACAATATGGGAACGTTGGAATTCGG
TATTACCTGATGGTAAGATGAATCCTGAAGGGATGAATTCATTAAATCATTATAGTTTTG
GCGCTGTGATGATGTGGATGTATCAATGTGTAGTAGGTTTAAATCAATTTGATGCTGGA
TTTAAAGAAATTTACTTTGCACCTAAATTTGATTGTCGATTAAAAGATATATATTCAGAGT
TTGATTCAACATATGGGAAGATTAAAGTGGAATATCATTTGGAGACCAATGAAAAGCATT
TGATAAGAATGAATTTAGTCATACCATTTGGAGTGAAAATGAAGGTTAAGTTGCCTAGAT
CTGCTAAATATTTGATTAACGGAAAAGAAAAGATAGGTATTGTTAAATTAGAATATGGGA
AATACGATATTAGTTATATTCCAACTAAGAGCTACTTAAATTATTATGATTTAAATAGTAA
GCTTGTTGATATTTTAGATAATAACGATCTTGTGAAAAGAATTGATCAAATTGATGAAAA
GATCTTGCAGAAAGTAAAAAGAATGGGAAATACCAGAAGTATATTTATTAATAAAAAAAT
AGATGAATTGCTTGATTTCGAAGAAATTTCACAGGAAGAAAAAAATCAGCTTGTTGATAT
ATTACATAAAACAATTTTTATAAAGTTGAATCTCTTT

> **Rha_{La} protein sequence** (931 aas) (*HindIII* restriction site, where the shorter version was cut, is marked in grey)

MGMKITNILVNQMEHPLGFDLSNLRITFELTEMENIIGNVYKNISVGKVESEQPIYFEPDELY
ENNAFKINMELEPRTKYWVKIGVRNDNEVTSSNTWFETGKMDEKFYGKWITNKKDVENTL
FKKDFELANKQIKSARLYSTTLGVYEVDLNGVKVGNEFLAPGFTNYDKIVQLQTYDVTKLVT
KNSNNELVFSVGDGWYKGNLGFDGGQTNIYGDKKSILAELHVTYTDNSEQVISTDSSWLT
TEGKIIKSSIYYGEDIDDTKDILDWSSVVILNKSTSIVRDRLSLPIMKKEVLKVKEIIHTPKNE
IVLDFGQNHAGWPVFINRLARGKKITLQMGEILQDGNFYNKNLRLARAAFTYISDGEEKLIR
PHFTYFGFRYVKISGVTDVNKDDFESWVLYSDLKQTGFIKTNNDKVNRLFKNVIWGQKSN
FMDVPTDCPQRDERLGWTGDAEIFAPTASFNMNTYEFYKKYAKDMLVEQEDNKGMLPIIVP
SLKQKSTGMAIWSDAATIIPWVTYRFFDDLGVLKQNYSQMKNWVDWITQNTKTKYLWIG
QMQLGDWLSLDNGANPQGKTNEDYIASIYYFVSASIVSKAARLLHYDMESDYYENLARNIK
TNILNEFVTEKGRIAIDTQTALVLALHFGLVHDYQKSQVVADLVKKVKDDNKHLQTGFVGT
PFLLSVLSNNNQHHLAMDIFMQEDCPSWLYEVNMGATTIWERWNSVLPDGKMNPEGMNS
LNHYSFGAVMMWMYQCVVGLNQFDAGFKEIYFAPKFDCRLKDIYSEFDSTYGKIKVEYHLE
TNEKHLIRMNLVIPFGVKMKVKLPRSAKYLINGKEKIGIVKLEYGKYDISYIPTKSYLNYYDL
NSKLVDILDNNDLVKRIDQIDEKILQKVKRMGNTRSIFINKKIDELLDFEEISQEEKNQLVDI
LHKTIFIKLNLF

6.1.7 Rhamnosidase Rha_{Hp} sequence

> **Rha_{Hp} with intron (in grey)** (2602 bps)

ATGGTTGAGGTTAGAGGAGTTCGACTAGAGTCTAACGCTCTCGGACCAACCAATGGGTT
AATCGCAACTAAAAGACCAGTCATTTCCTGGGAGATTTTTGGTAACGAGAAGAACTGGT
ATCAAGGTGCTTATCAAATCAAGATTAAATACGGGGACCAACCTTGCAGCACTTACGAT
GTTGTCAGCTCTGACAATAATCTGTTTGTGGAATGGCCAGGAAGAGACTTGAAGTCAAG
AGAAGTAATAGCTATTTCGATTCGTGTTGCACCAAAATCTGATATTTCTTCTTTTTCTGAA
TGGAGTGCGCCCGCAGTCGCTCAAGTTGGAATTCTGGATAACTCCTGGAGGGCTCCATT
TATTTCTATGGTTGGGCAGCCCAAGGATGAAAACGTCTCACCAGAAATACTATTCAGGA
AGAATTTCGATTTATCGGGTAACAAAATCAAGTCTGCTAAGATCTATAGCACAGCTTTGG
GGGTTTACGAAATTGAAGTCAATGGAAGGAAGGTTGGAAAAGACTACTTAGCACCTGGT
TGGACAAGCTATGATCATCGTTTGCTCCATCAATTCTACGATATAACAGACTTGCTTGCA
CCGCATACTAATGCTATTGGTGCTAGAGTAGGGTCTGGATGGTATAGTGGAATGCTTGG
TTTTGACGGTGGGGCCAGAAATATTTACGGAGACCGAAGAGCCATTAGTATCGAGATAG
AGGTGGAATACGAAAATGGTGAGAAAGCTGTGATTCTCTCTGATGAAGCTTGGAAATCA
AATTATGGCCCCATTAAAGATGCACAGTTGTACAATGGTGAGGTTTATGATGCAACTTGC
GAGTTTCCAGGATGGTCTCTCCCGACCTTTGAAGATAAATCGTGGAACCCAGTTGAAAT
TATTCAATGCAGCTCCAAATCAATTGAGCCGCAAGCTTTTGGTTACATATGTGAAATTGA
CTCTCTGACACCAATCAATATTATCGAGACACCAAAAAAAAAGTTGATTGTTGATTTCGG
TCAAAATGCGGTTGGTTTCGCTCGGATCAATAATGCCAAAGCTCCAAAAGGGCACACTA
TTACTTTAAGATTTGCTGAAGTTTTGGAGAATGGGGAGCTCGGCACTCGTCCCCTGAGG
CTGGCAAAAGCTACAGACCAGTACACTTTCAAAGGTGATATAGATGGTGAGACATACGC
TCCTAGATTTACATTTCATGGAGCAAGATACTGTCAGATTGACAACTGGTATGGAGAGTT
TAGCCCAAACAATATTGAATTCGTTGTCATTGGAAATCTAATGGAGACAACTGGTGGTTT
TGAGTGTTCGGACAAAATGCTCAACAGGCTCCACCAGAATGTTGTCCATAGCATGAGAG
GCAATTTCCTAGCCATACCGATGGATTGCCCACAAAGAGATGAACGTCTCGGTTGGACA
GGTGACATTGCCATATTCTGTCCAACAGCACTCTACTTATGTGATTGTTACTCGTTCCTG
CAAAGCTGGCTAAAGGATTTGAAACTAGAACAGAAAGAGAAAGGTGGAGCTCCTGCTG
TTGTGGTCCCAGATATTATACATGCTTTTGATACATTTTGGAATGGTAAGATAGCCGCTA
TTTGGCAGGATGCCTCGGTTATTGTACCTTATGAGCTTTACAGATCATCTGGAAACAAGC
ATATTTTACAAGCCCAATACGAAAGTATGGTTTCATGGATTGACTACATACCTAAAATTG
AAGGAAAAGTCAGGTGGAATAAAATTGAGTTGCAGTTAGGAGATTGGCTGGATCCATCT
GCACCACCAGAAAACCCCCTTCTTGCTTTGACGGATGCCTATCTGGTTGCTGATGCTTTT
CTTTTCAAAATACTTACGATGACTTCGGAGACTGCTGAGATCTTGAAGTTGGAAAAAGAT
GCGGCTTATTACGCTGAGCTGGCTTCCAAAATGTCGTTCTGATTTCAATGAGGCTTATAT

TTCCTCGTCAGGGAAACTCACATCAAACACACAAACGGCATATGCTTTGGCTATCTGTTT
TGGCCTGTACAAAAATGATGATCAAATATCCTACGCTGGTAACCAATTGTCCGAACTAGT
TAGGCAAAGTGATTATAAAATCAGTACTGGATTTGCTGGAACTCCATTTGTCACTGAAGC
TTTGACTATTACAGGACATCTAGAAGACGCCTACAAAATGATACTACAAAAAGAATGCCC
TTCATGGCTCTATCCAGTTTCTGTGGGAGCAACAACAGTCTGGGAACGATGGGATAGTA
TGTTGCCTACAGGGCACATCAATCCTGGTGAAATGACCTCTTTCAATCATTACGCCTTGG
GTTCAATTGCTAACTGGATGCATGAAAGAATGGGTGGTCTAAAACTGAAGGAACCAGGC
TGGCGAGAGTTTTACCTTAGGCCTATGCCTGGATCTAACATCAGCTATTGTGAGACTTTC
CACAAATCGCCTAGTGGCTTGATCAAGTCTGAATGGAAATTAGAAGCTGGTAAGTTTGT
CTACAATGTGACTGTTCCCTTGAATTCAACTGCCCACATCACACTTCCTGACGGAACCAC
TCATTCTGTGGGCTCAGGATCGTGGTCTCTTACGTGCTCAGCT

> **Rha$_{Hp}$ protein sequence** (857 aas)

MVEVRGVRLESNALGPTNGLIATKRPVISWEIFGNEKNWYQGAYQIKIKYGDQPCSTYDVV
SSDNNLFVEWPGRDLKSREVIAISIRVAPKSDISSFSEWSAPAVAQVGILDNSWRAPFISM
VGQPKDENVSPEILFRKNFDLSGNKIKSAKIYSTALGVYEIEVNGRKVGKDYLAPGWTSYD
HRLLHQFYDITDLLAPHTNAIGARVGSGWYSGMLGFDGGARNIYGDRRAISIEIEVEYENG
EKAVILSDEAWKSNYGPIKDAQLYNGEVYDATCEFPGWSLPTFEDKSWNPVEIIQCSSKSI
EPQAFGYICEIDSLTPINIIETPKKKLIVDFGQNAVGFARINNAKAPKGHTITLRFAEVLENGE
LGTRPLRLAKATDQYTFKGDIDGETYAPRFTFHGARYCQIDNWYGEFSPNNIEFVVIGNLME
TTGGFECSDKMLNRLHQNVVHSMRGNFLAIPMDCPQRDERLGWTGDIAIFCPTALYLCDC
YSFLQSWLKDLKLEQKEKGGAPAVVVPDIIHAFDTFWNGKIAAIWQDASVIVPYELYRSSG
NKHILQAQYESMVSWIDYIPKIEGKVRWNKIELQLGDWLDPSAPPENPLLALTDAYLVADA
FLFKILTMTSETAEILNWLPKCRSDFNEAYISSSGKLTSNTQTAYALAICFGLYKNDDQISYA
GNQLSELVRQSDYKISTGFAGTPFVTEALTITGHLEDAYKMILQKECPSWLYPVSVGATTV
WERWDSMLPTGHINPGEMTSFNHYALGSIANWMHERMGGLKLKEPGWREFYLRPMPGS
NISYCETFHKSPSGLIKSEWKLEAGKFVYNVTVPLNSTAHITLPDGTTHSVGSGSWSLTCS
A

6.2 CAL-A optimisation codon bias analysis of the whole sequences

The codon usage of the sequences was analysed and is presented using Graphical Codon Usage Analyser (http://gcua.schoedl.de) tool, version 2.0.[79] As an indicator for the codon usage, the relative adaptiveness value was chosen, as it reflects the number of codons which code for the respective amino acid. The relative adaptiveness value is derived from the codon usage frequency, but then the most popular codon for the analysed amino acid has this value set to 100% and the other codons for it are scaled accordingly. As a result, the comparison between the amino acids with different numbers of codons becomes more transparent.[218] Codons with the relative adaptiveness value lower then 20% are shown in grey and the codons with the relative adaptiveness value lower then 10% are shown in red.
Starting from the section 6.2.2, respective fragments of the lipase CAL-A before (left site) and after codon optimisation (right site) for each yeast are compiled to enable direct comparision of the sequences.

6.2.1 Codon usage tables used for the codon bias analysis and optimisation

The codon usage tables used in this study were taken from the Codon Usage Database (http://www.kazusa.or.jp/codon/), which is based on NCBI GenBank.[199]

***Saccharomyces cerevisiae* [gbpln]: 14411 CDS's (6534504 codons)**
fields: [triplet] [amino acid] [fraction] [frequency: **per thousand**] [number]

UUU	**F**	0.59	26.1	170666	UCU	**S**	0.26	23.5	153557	UAU	**Y**	0.56	18.8	122728	UGU	**C**	0.63	8.1	52903
UUC	**F**	0.41	18.4	120510	UCC	**S**	0.16	14.2	92923	UAC	**Y**	0.44	14.8	96596	UGC	**C**	0.37	4.8	31095
UUA	**L**	0.28	26.2	170884	UCA	**S**	0.21	18.7	122028	UAA	*****	0.47	1.1	6913	UGA	*****	0.30	0.7	4447
UUG	**L**	0.29	27.2	177573	UCG	**S**	0.10	8.6	55951	UAG	*****	0.23	0.5	3312	UGG	**W**	1.00	10.4	67789
CUU	**L**	0.13	12.3	80076	CCU	**P**	0.31	13.5	88263	CAU	**H**	0.64	13.6	89007	CGU	**R**	0.14	6.4	41791
CUC	**L**	0.06	5.4	35545	CCC	**P**	0.15	6.8	44309	CAC	**H**	0.36	7.8	50785	CGC	**R**	0.06	2.6	16993
CUA	**L**	0.14	13.4	87619	CCA	**P**	0.42	18.3	119641	CAA	**Q**	0.69	27.3	178251	CGA	**R**	0.07	3.0	19562
CUG	**L**	0.11	10.5	68494	CCG	**P**	0.12	5.3	34597	CAG	**Q**	0.31	12.1	79121	CGG	**R**	0.04	1.7	11351
AUU	**I**	0.46	30.1	196893	ACU	**T**	0.35	20.3	132522	AAU	**N**	0.59	35.7	233124	AGU	**S**	0.16	14.2	92466
AUC	**I**	0.26	17.2	112176	ACC	**T**	0.22	12.7	83207	AAC	**N**	0.41	24.8	162199	AGC	**S**	0.11	9.8	63726
AUA	**I**	0.27	17.8	116254	ACA	**T**	0.30	17.8	116084	AAA	**K**	0.58	41.9	273618	AGA	**R**	0.48	21.3	139081
AUG	**M**	1.00	20.9	136805	ACG	**T**	0.14	8.0	52045	AAG	**K**	0.42	30.8	201361	AGG	**R**	0.21	9.2	60289
GUU	**V**	0.39	22.1	144243	GCU	**A**	0.38	21.2	138358	GAU	**D**	0.65	37.6	245641	GGU	**G**	0.47	23.9	156109
GUC	**V**	0.21	11.8	76947	GCC	**A**	0.22	12.6	82357	GAC	**D**	0.35	20.2	132048	GGC	**G**	0.19	9.8	63903
GUA	**V**	0.21	11.8	76927	GCA	**A**	0.29	16.2	105910	GAA	**E**	0.70	45.6	297944	GGA	**G**	0.22	10.9	71216
GUG	**V**	0.19	10.8	70337	GCG	**A**	0.11	6.2	40358	GAG	**E**	0.30	19.2	125717	GGG	**G**	0.12	6.0	39359

Coding GC 39.77% 1st letter GC 44.58% 2nd letter GC 36.64% 3rd letter GC 38.10%

***Kluyveromyces lactis* [gbpln]: 301 CDS's (150121 codons)**
fields: [triplet] [amino acid] [fraction] [frequency: **per thousand**] [number]

UUU	**F**	0.45	19.5	2929	UCU	**S**	0.32	26.7	4012	UAU	**Y**	0.54	20.2	3033	UGU	**C**	0.75	9.5	1433
UUC	**F**	0.55	23.5	3534	UCC	**S**	0.15	12.7	1901	UAC	**Y**	0.46	17.0	2557	UGC	**C**	0.25	3.2	483
UUA	**L**	0.26	23.5	3534	UCA	**S**	0.20	16.3	2445	UAA	*****	0.54	1.1	163	UGA	*****	0.28	0.6	83
UUG	**L**	0.37	33.9	5083	UCG	**S**	0.09	7.7	1150	UAG	*****	0.18	0.4	55	UGG	**W**	1.00	11.3	1697
CUU	**L**	0.13	11.9	1779	CCU	**P**	0.31	13.5	2020	CAU	**H**	0.67	14.2	2130	CGU	**R**	0.16	6.7	1001
CUC	**L**	0.05	4.3	649	CCC	**P**	0.09	3.8	573	CAC	**H**	0.33	6.9	1043	CGC	**R**	0.04	1.5	228
CUA	**L**	0.13	11.8	1766	CCA	**P**	0.50	21.3	3201	CAA	**Q**	0.71	29.4	4411	CGA	**R**	0.05	2.1	318
CUG	**L**	0.06	5.1	770	CCG	**P**	0.10	4.2	633	CAG	**Q**	0.29	11.8	1769	CGG	**R**	0.04	1.5	224
AUU	**I**	0.43	27.5	4123	ACU	**T**	0.40	22.9	3444	AAU	**N**	0.55	31.5	4735	AGU	**S**	0.15	12.8	1917
AUC	**I**	0.33	20.9	3138	ACC	**T**	0.23	12.8	1923	AAC	**N**	0.45	25.5	3829	AGC	**S**	0.08	6.3	953
AUA	**I**	0.25	15.8	2368	ACA	**T**	0.27	15.2	2282	AAA	**K**	0.53	37.5	5629	AGA	**R**	0.58	24.7	3707
AUG	**M**	1.00	21.0	3158	ACG	**T**	0.10	5.8	874	AAG	**K**	0.47	33.8	5070	AGG	**R**	0.14	6.0	902
GUU	**V**	0.42	25.9	3893	GCU	**A**	0.47	28.1	4217	GAU	**D**	0.69	40.7	6116	GGU	**G**	0.61	35.5	5335
GUC	**V**	0.23	14.2	2138	GCC	**A**	0.20	11.8	1778	GAC	**D**	0.31	18.4	2762	GGC	**G**	0.10	5.6	845
GUA	**V**	0.18	10.9	1642	GCA	**A**	0.26	15.5	2334	GAA	**E**	0.75	47.5	7124	GGA	**G**	0.20	11.5	1727
GUG	**V**	0.18	10.9	1636	GCG	**A**	0.08	4.9	734	GAG	**E**	0.25	15.9	2393	GGG	**G**	0.09	5.2	788

Coding GC 39.61% 1st letter GC 45.28% 2nd letter GC 36.75% 3rd letter GC 36.82%

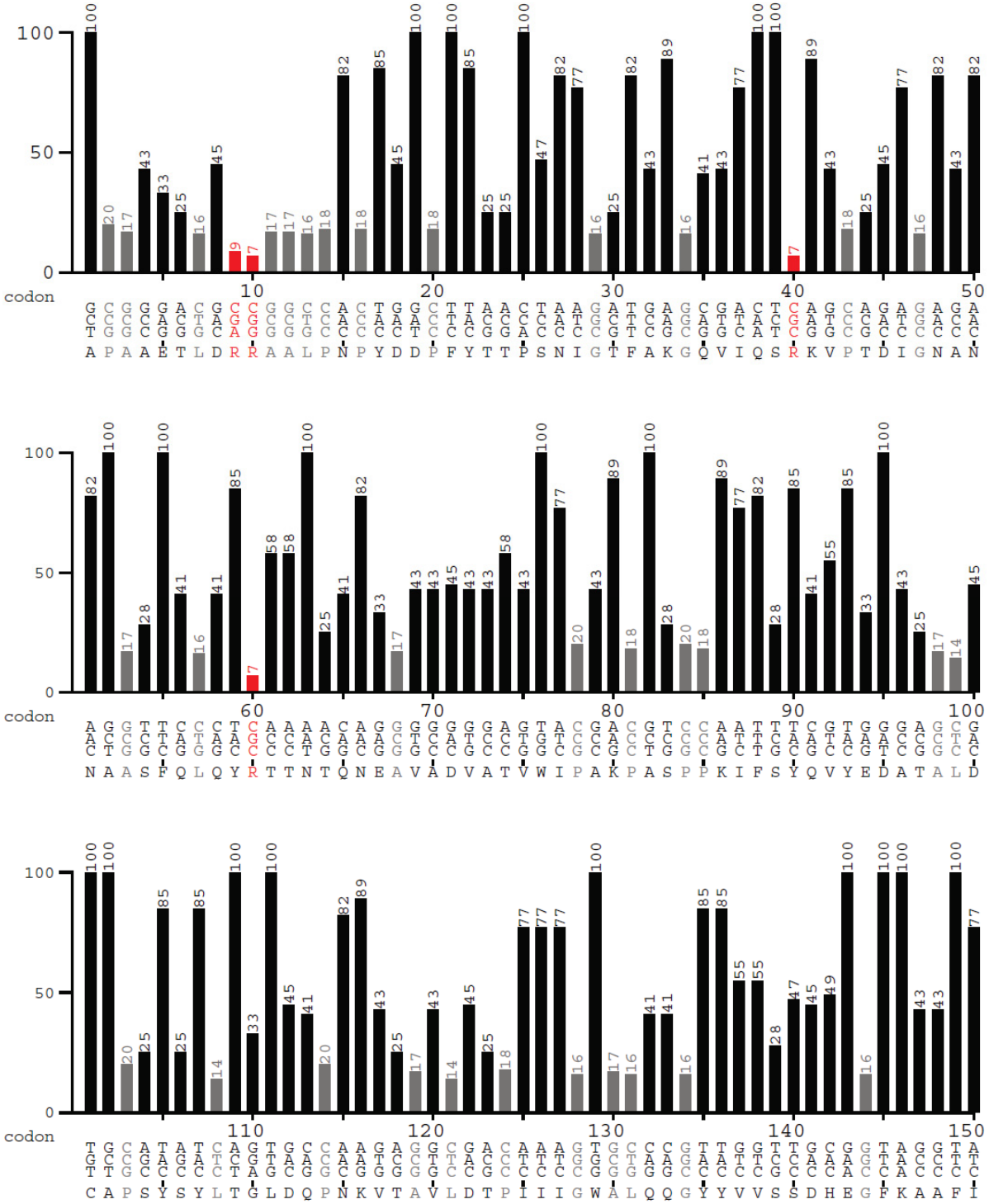
***Pichia angusta* [gbpln]: 198 CDS's (99894 codons)**
fields: [triplet] [amino acid] [fraction] [frequency: **per thousand**] [number]

UUU	**F**	0.44	19.6	1957	UCU	**S**	0.21	17.0	1701	UAU	**Y**	0.33	11.5	1146	UGU	**C**	0.47	6.2	622
UUC	**F**	0.56	24.9	2491	UCC	**S**	0.20	16.3	1624	UAC	**Y**	0.67	23.2	2313	UGC	**C**	0.53	7.0	700
UUA	**L**	0.06	5.6	562	UCA	**S**	0.11	8.7	867	UAA	*****	0.43	0.9	86	UGA	*****	0.26	0.5	51
UUG	**L**	0.21	21.1	2108	UCG	**S**	0.23	18.0	1797	UAG	*****	0.31	0.6	61	UGG	**W**	1.00	12.3	1225
CUU	**L**	0.17	16.9	1691	CCU	**P**	0.29	13.5	1349	CAU	**H**	0.37	8.0	803	CGU	**R**	0.10	4.8	475
CUC	**L**	0.18	17.4	1737	CCC	**P**	0.14	6.4	637	CAC	**H**	0.63	13.8	1380	CGC	**R**	0.15	7.2	715
CUA	**L**	0.07	7.3	726	CCA	**P**	0.33	15.1	1506	CAA	**Q**	0.37	14.0	1396	CGA	**R**	0.08	3.8	378
CUG	**L**	0.31	30.5	3042	CCG	**P**	0.24	11.2	1118	CAG	**Q**	0.63	24.1	2405	CGG	**R**	0.12	5.6	562
AUU	**I**	0.40	22.8	2282	ACU	**T**	0.23	12.8	1283	AAU	**N**	0.35	16.3	1627	AGU	**S**	0.08	6.7	673
AUC	**I**	0.47	27.0	2702	ACC	**T**	0.34	19.1	1903	AAC	**N**	0.65	30.5	3042	AGC	**S**	0.16	13.2	1314
AUA	**I**	0.13	7.7	769	ACA	**T**	0.22	12.2	1218	AAA	**K**	0.41	25.8	2577	AGA	**R**	0.41	19.1	1909
AUG	**M**	1.00	20.7	2064	ACG	**T**	0.21	12.0	1203	AAG	**K**	0.59	37.5	3746	AGG	**R**	0.13	6.1	614
GUU	**V**	0.29	19.6	1959	GCU	**A**	0.28	19.3	1927	GAU	**D**	0.35	21.0	2096	GGU	**G**	0.29	16.6	1658
GUC	**V**	0.28	18.5	1852	GCC	**A**	0.33	22.8	2278	GAC	**D**	0.65	38.3	3829	GGC	**G**	0.32	18.7	1864
GUA	**V**	0.05	3.4	338	GCA	**A**	0.22	15.0	1500	GAA	**E**	0.36	23.5	2345	GGA	**G**	0.29	17.0	1702
GUG	**V**	0.38	25.2	2518	GCG	**A**	0.17	11.7	1164	GAG	**E**	0.64	41.3	4125	GGG	**G**	0.10	5.8	582

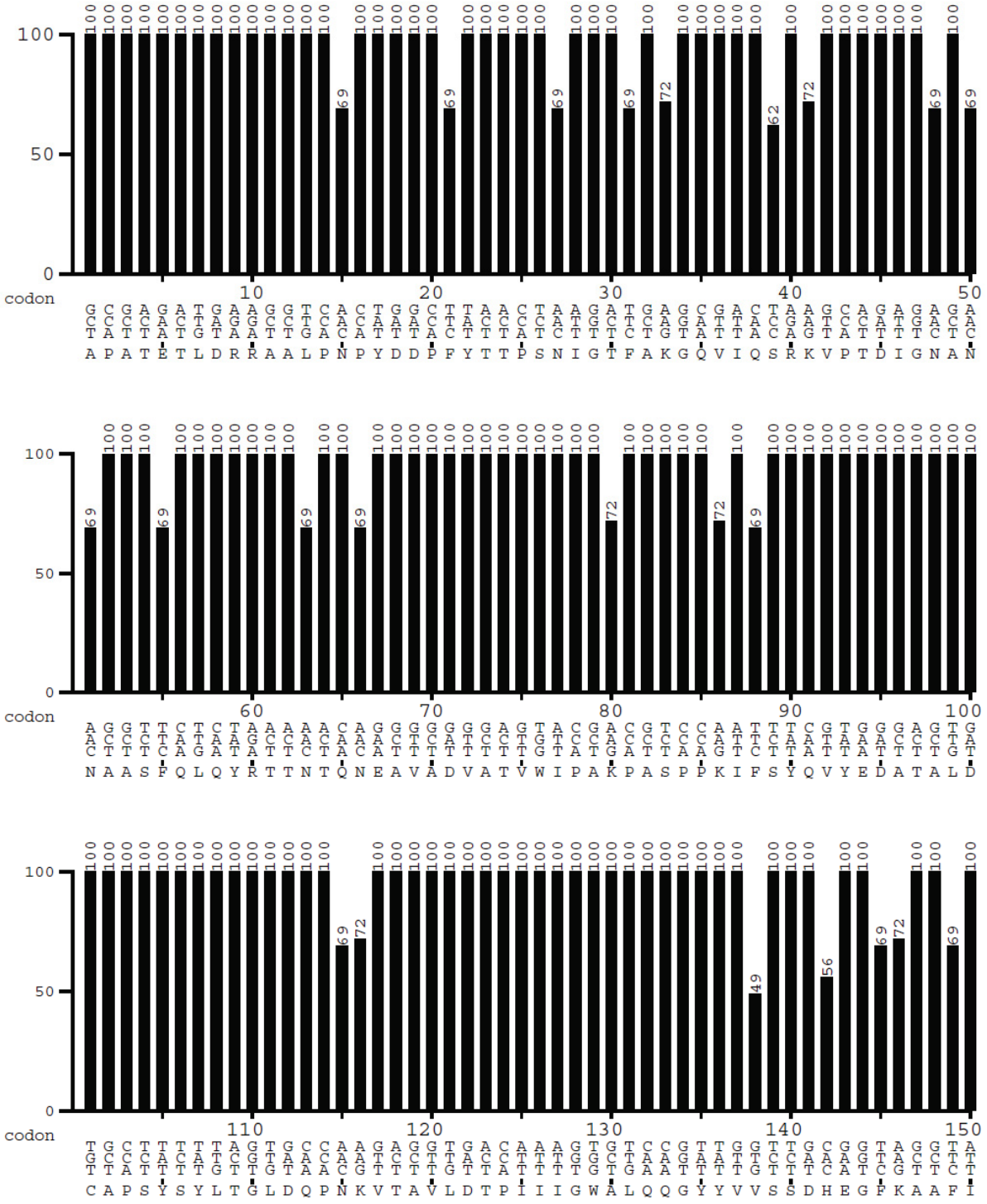
Coding GC 49.55% 1st letter GC 51.71% 2nd letter GC 38.16% 3rd letter GC 58.78%

6.2.2 CAL-A codon optimisation for S. cerevisiae.

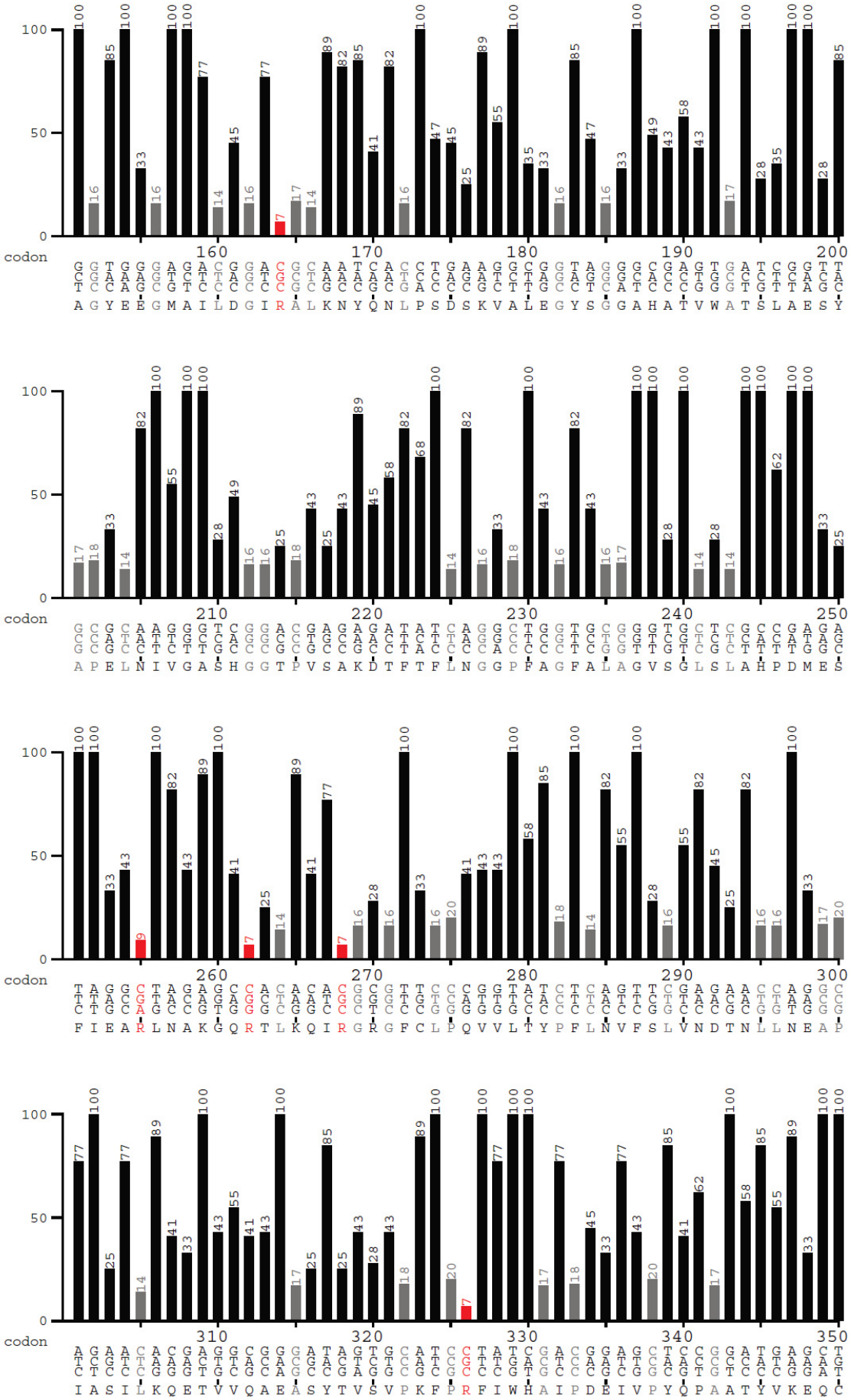
original sequence (1-150 aas)



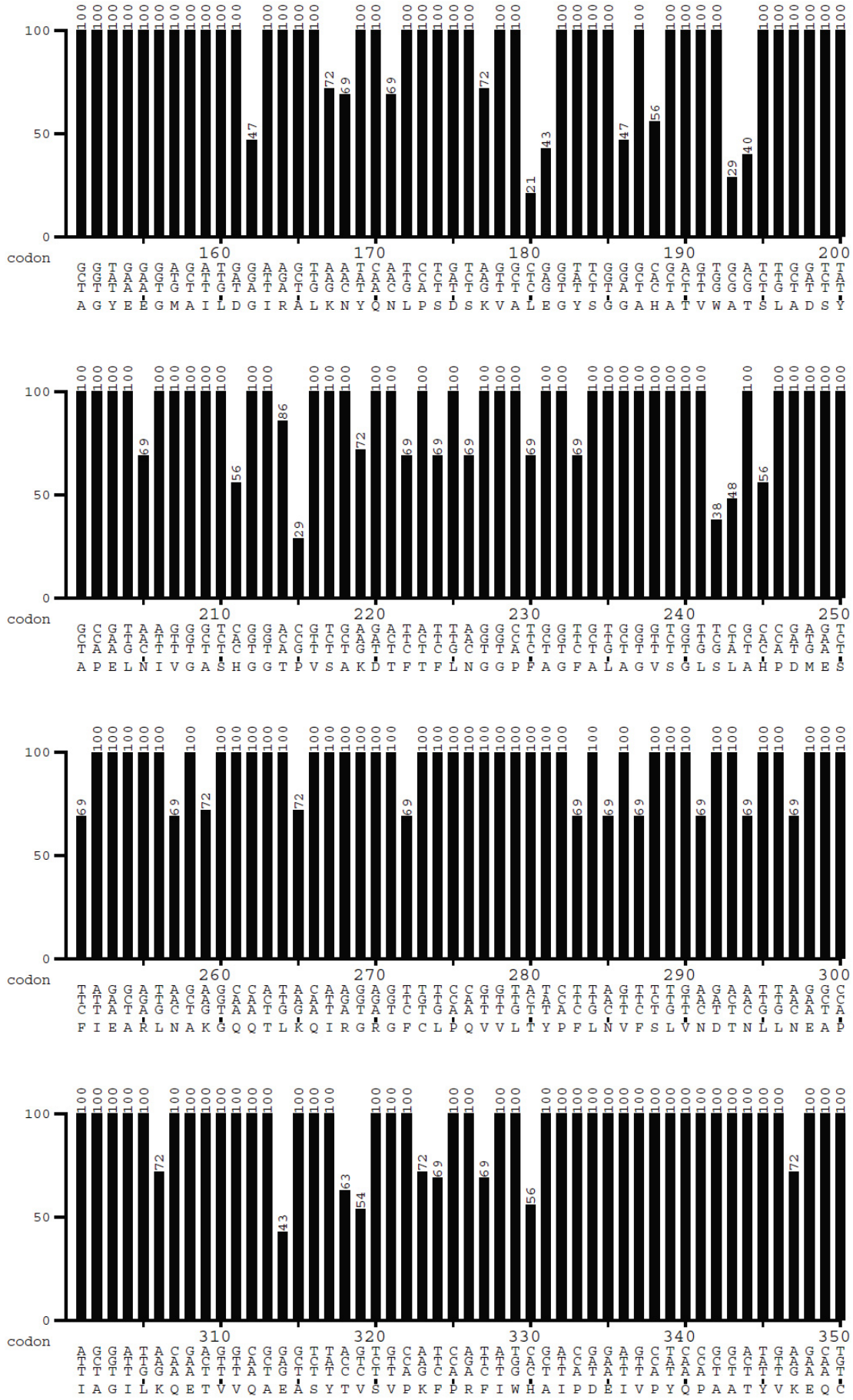
optimised sequence (1-150 aas)



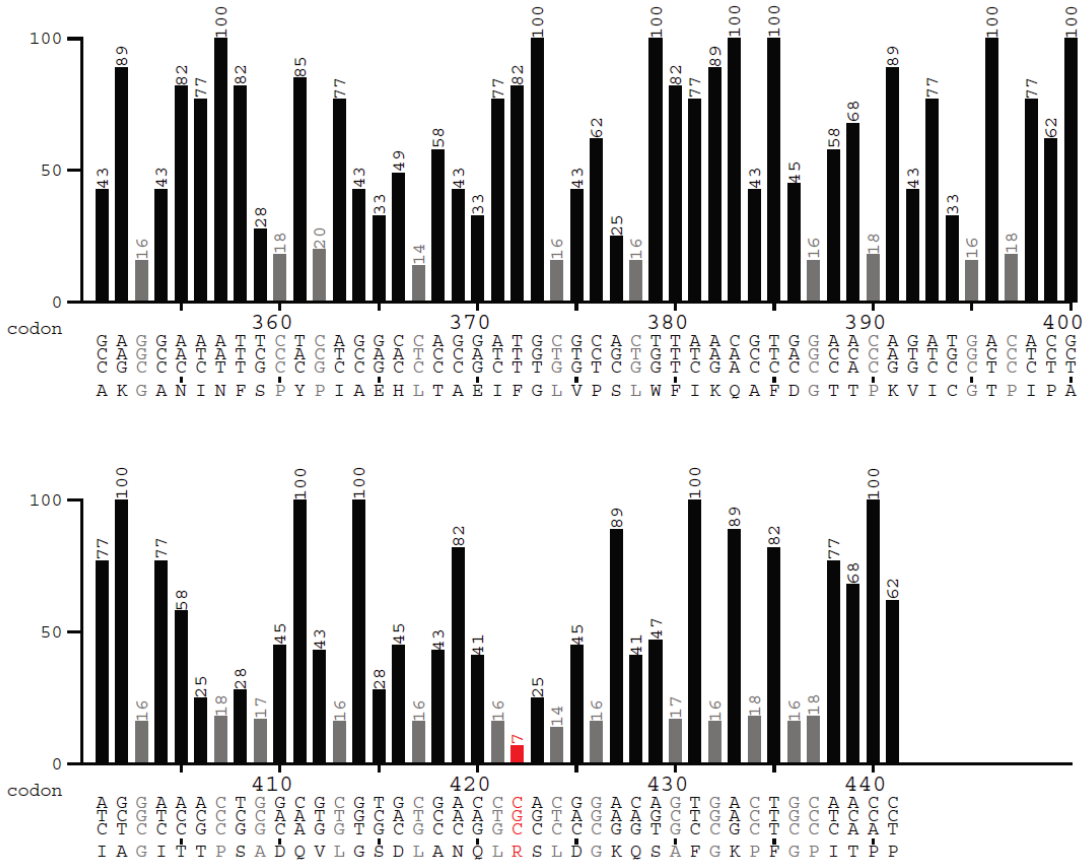
original sequence (151-350 aas)



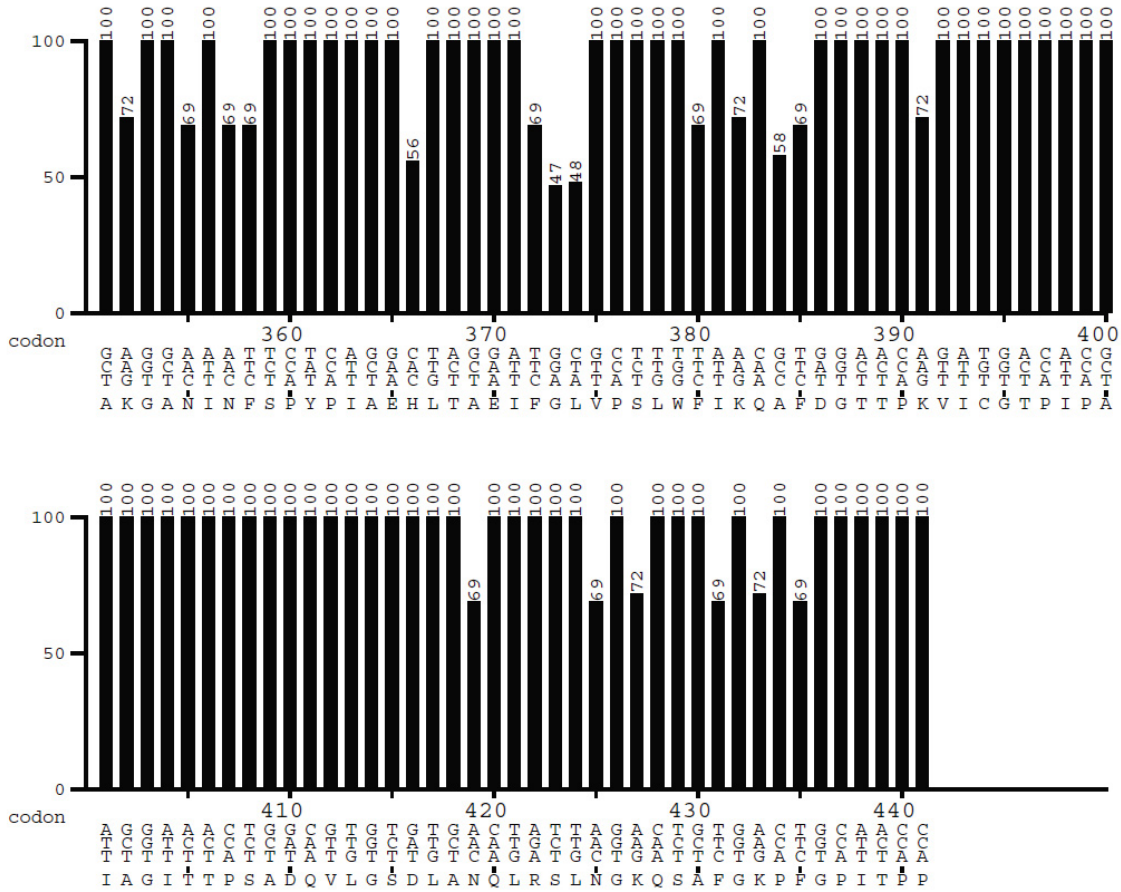
optimised sequence (151-350 aas)



original sequence (351-442 aas)

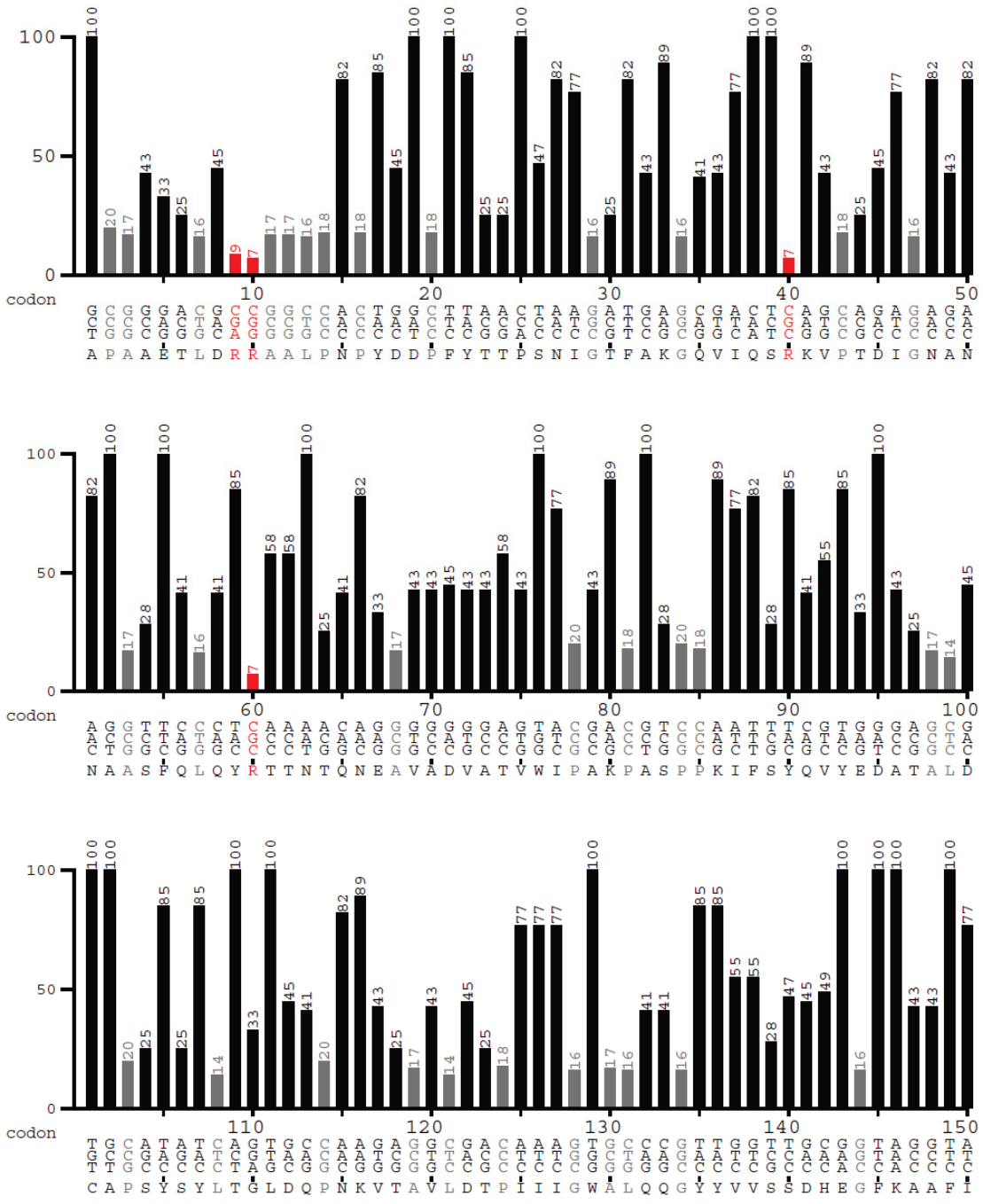


optimised sequence (351-442 aas)

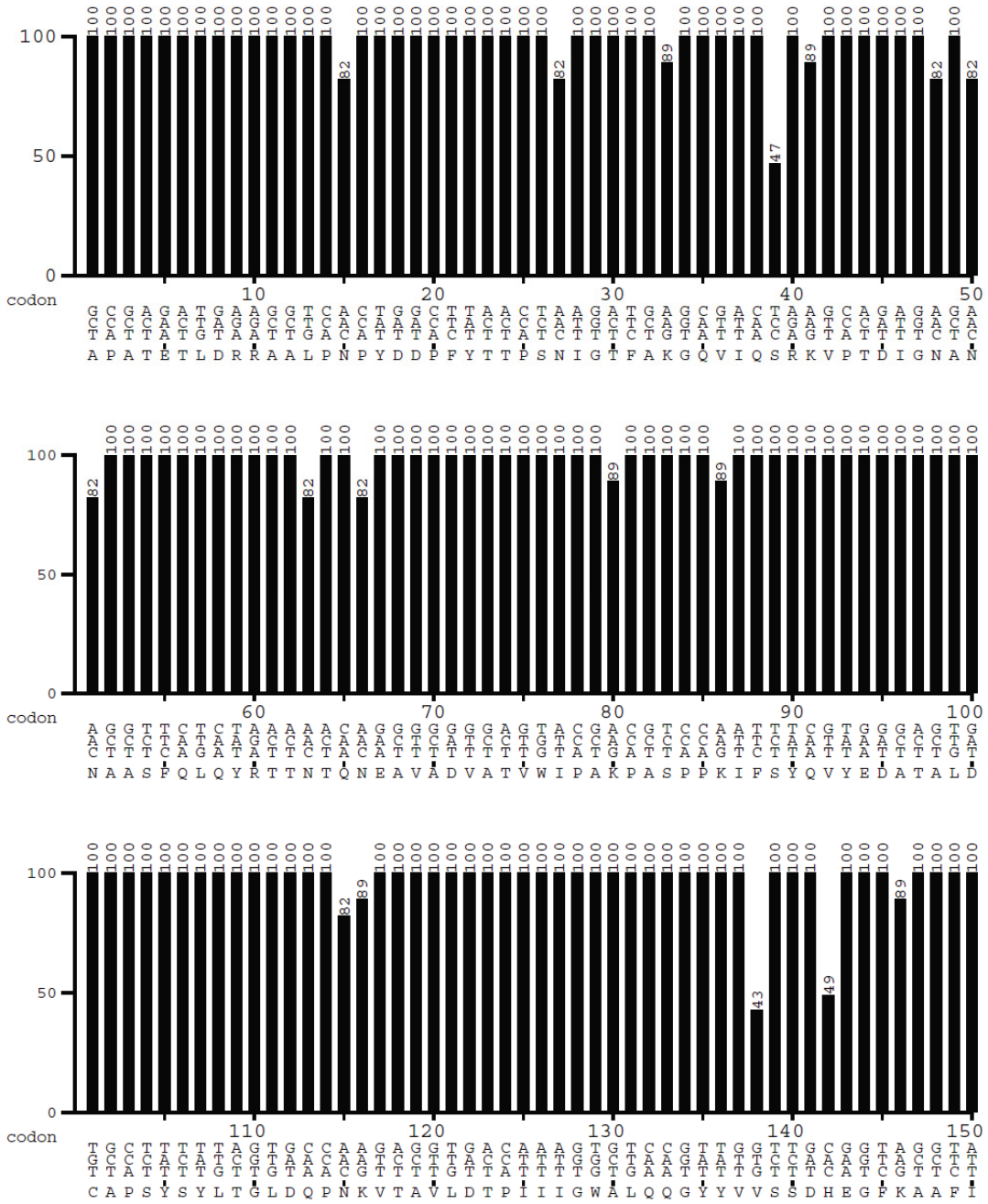


6.2.3 CAL-A codon optimisation for K. lactis.

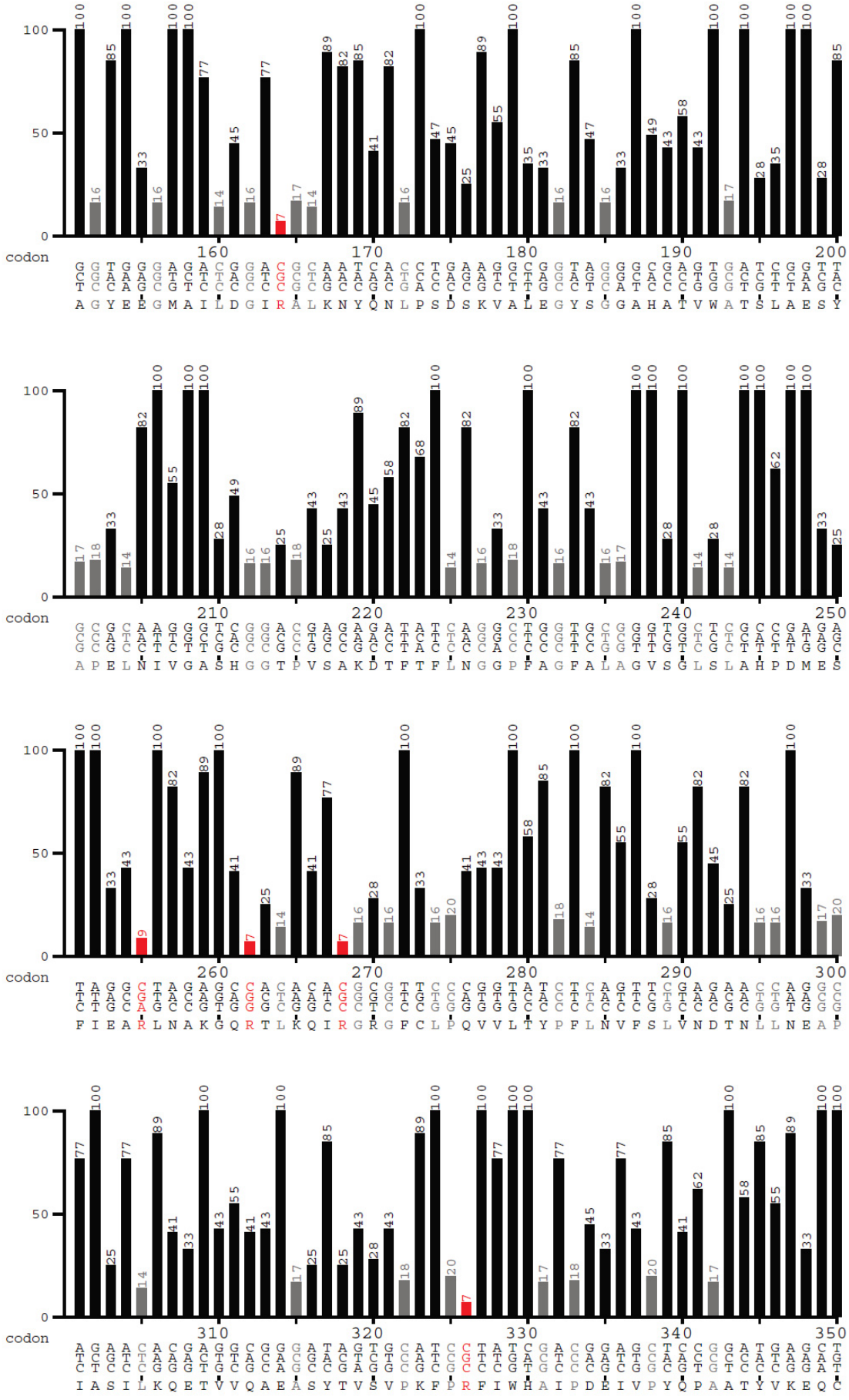
original sequence (1-150 aas)



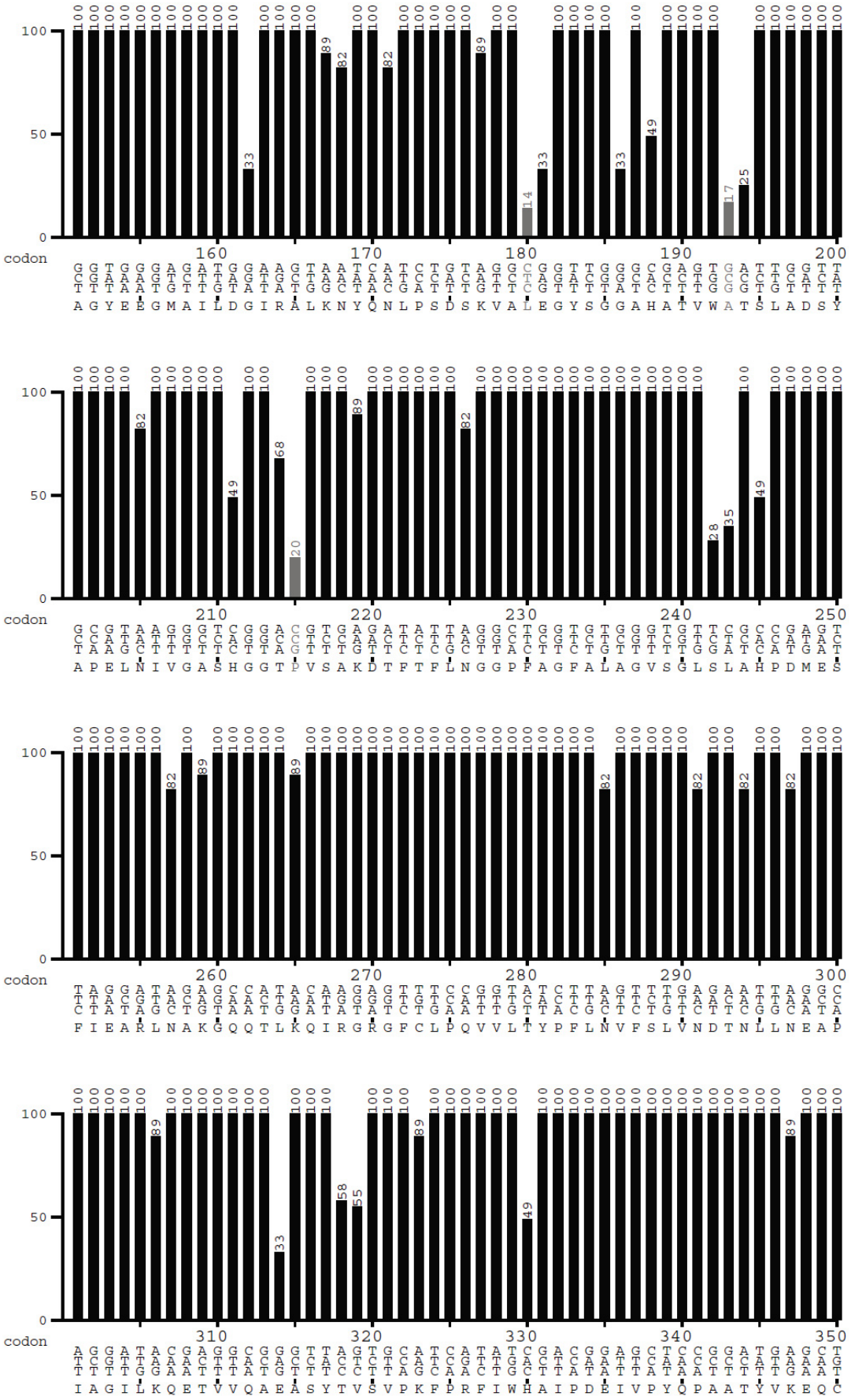
optimised sequence (1-150 aas)



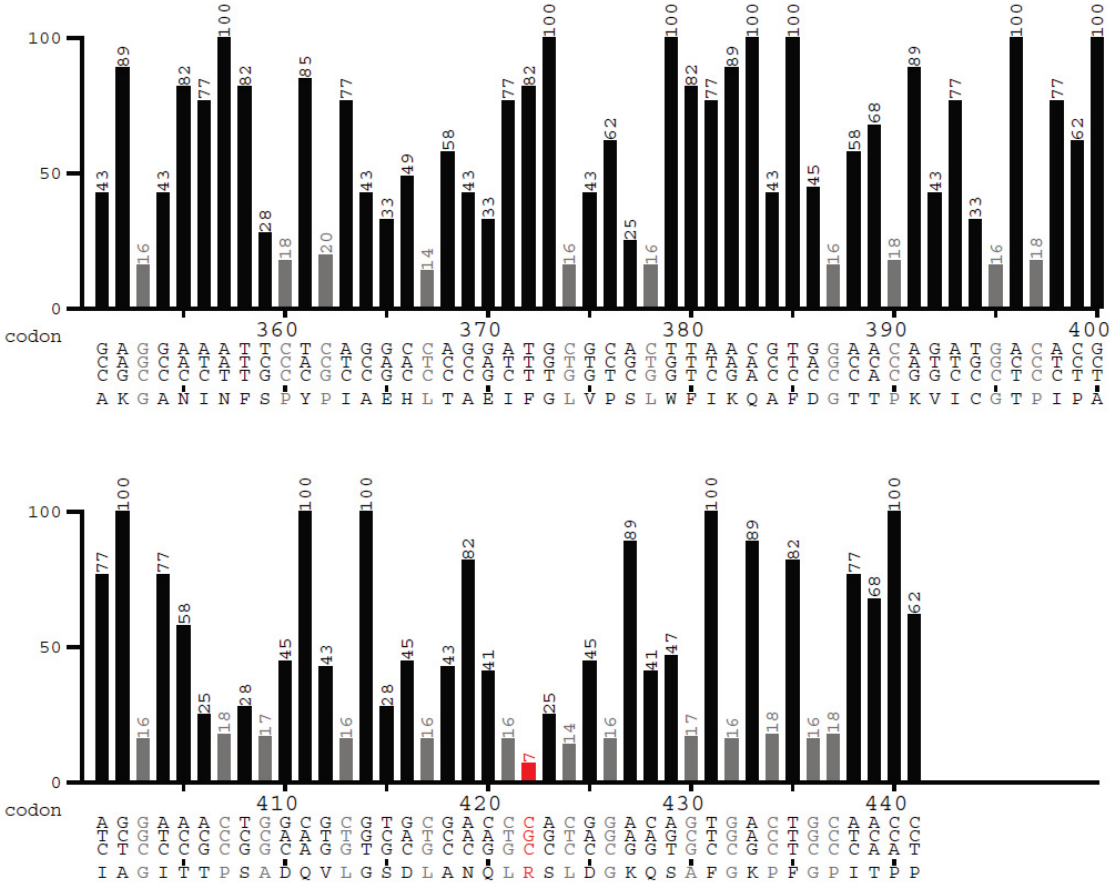
original sequence (151-350 aas)



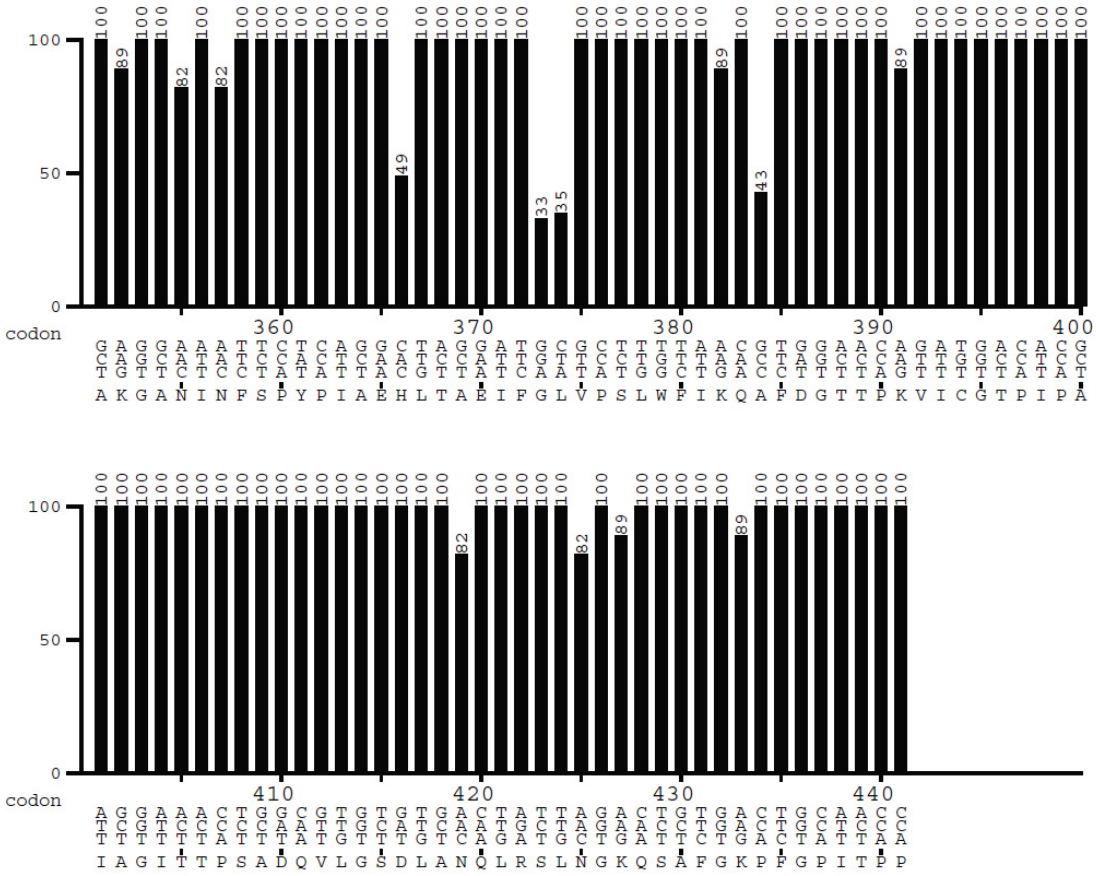
optimised sequence (151-350 aas)



original sequence (351-442 aas)

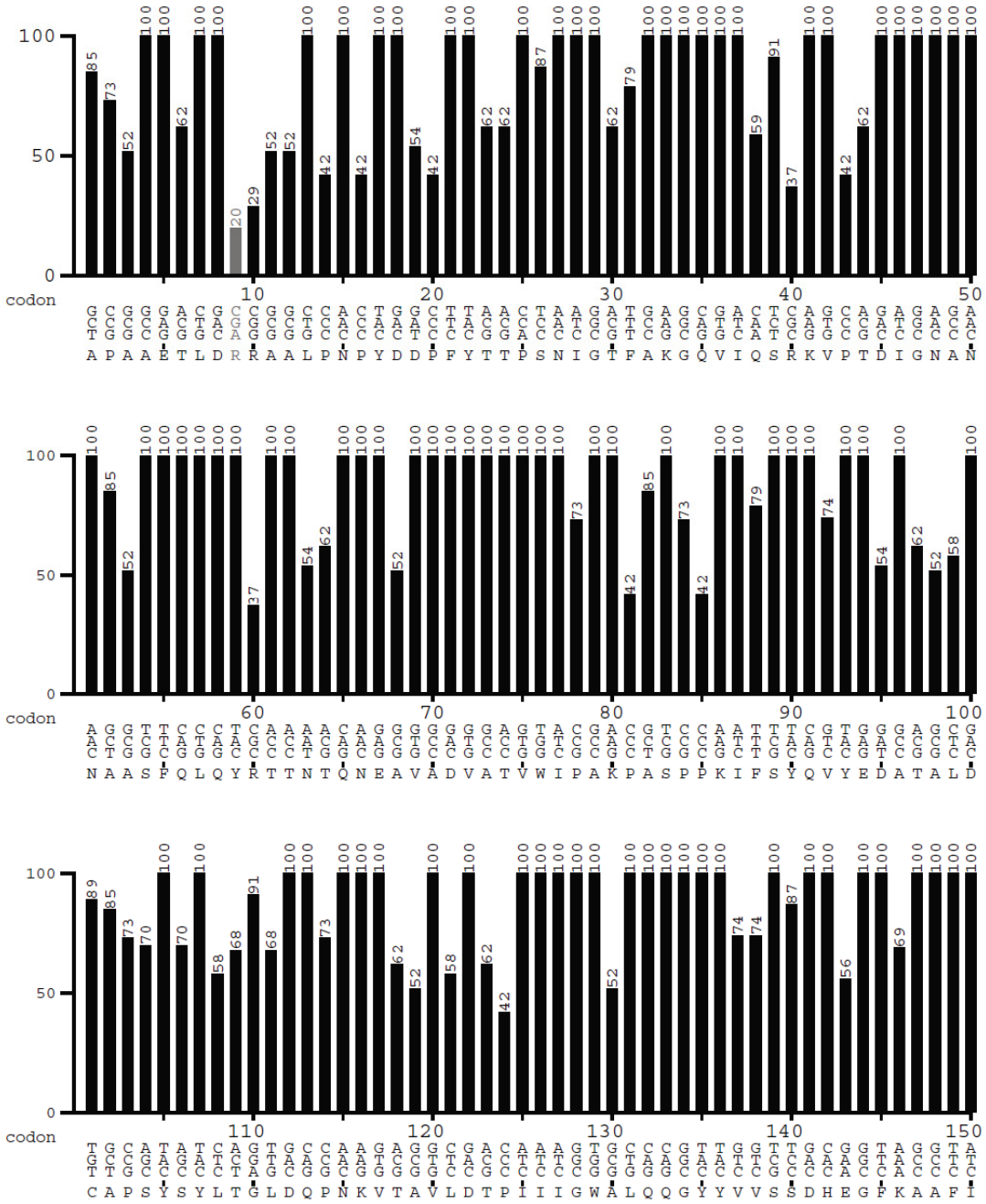


optimised sequence (351-442 aas)

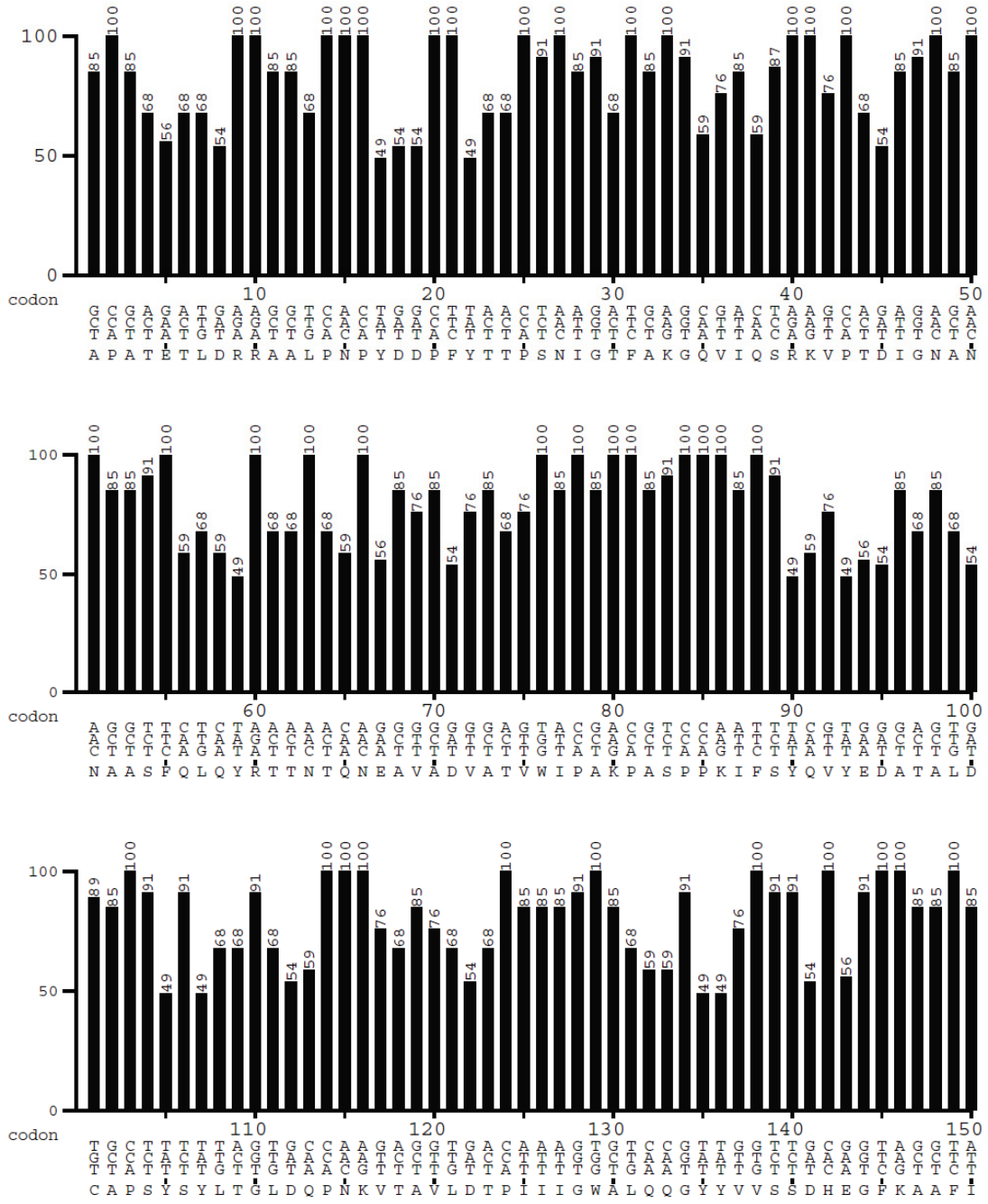


6.2.4 CAL-A codon optimisation for H. polymorpha.

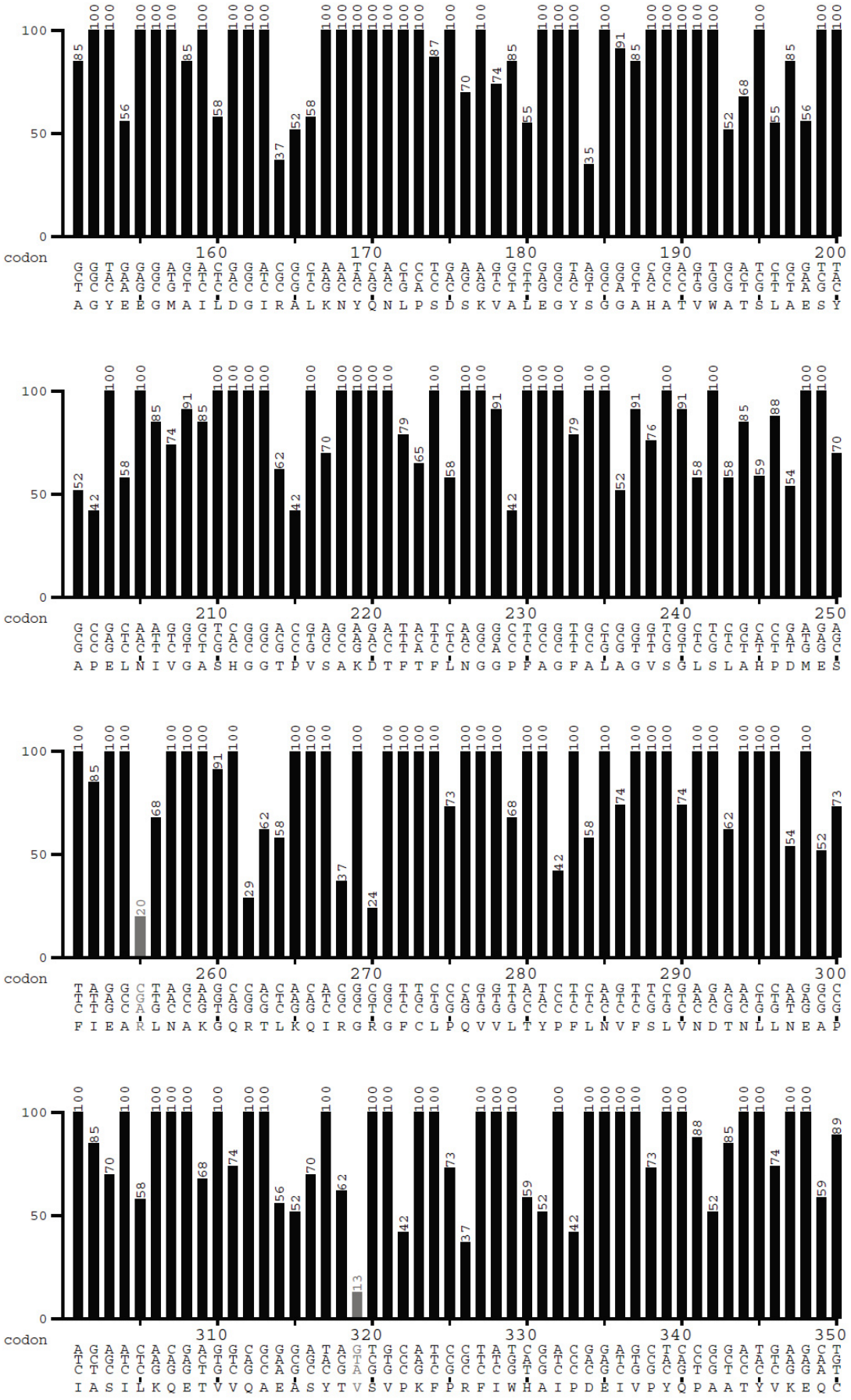
original sequence (1-150 aas)



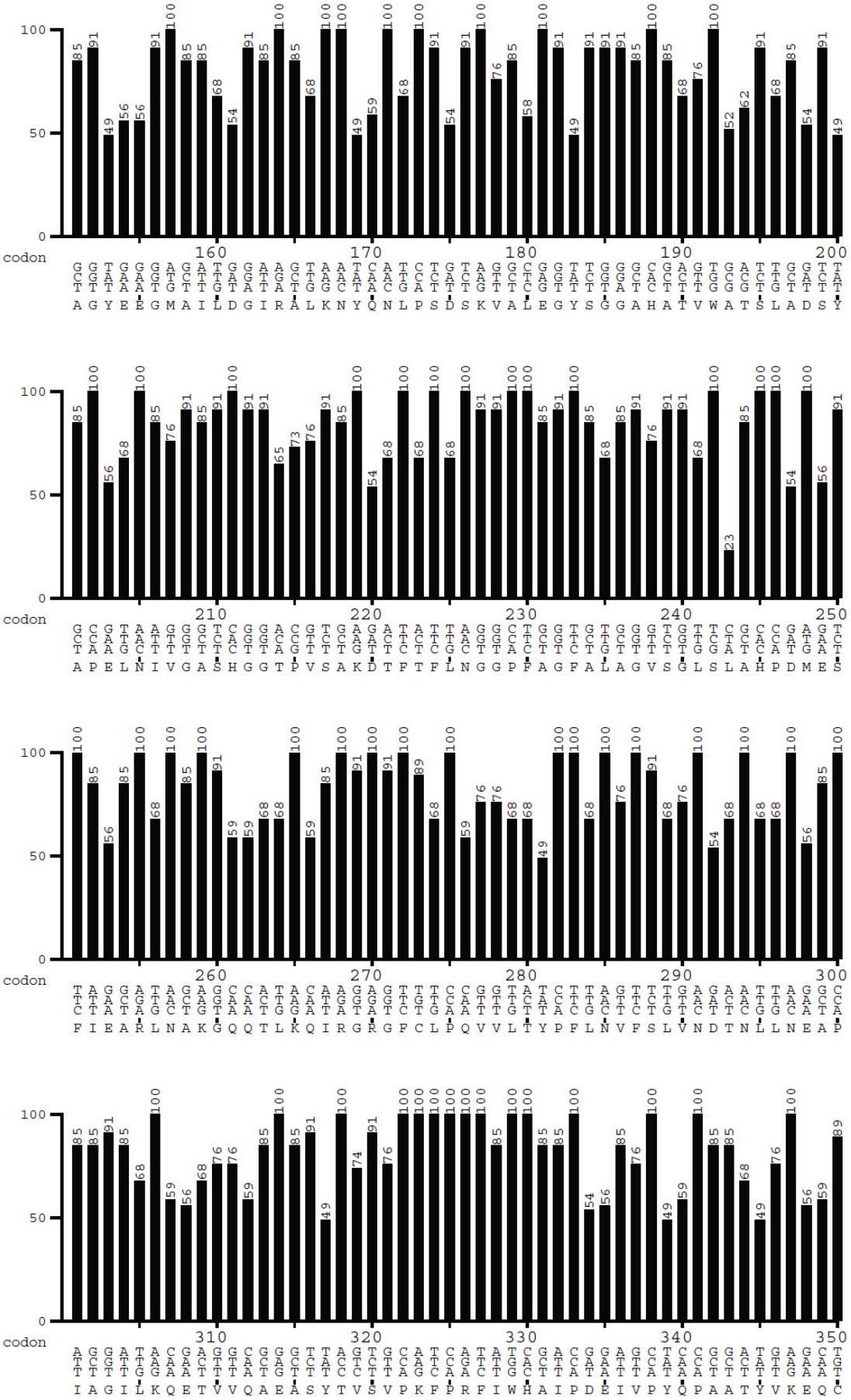
optimised sequence (1-150 aas)



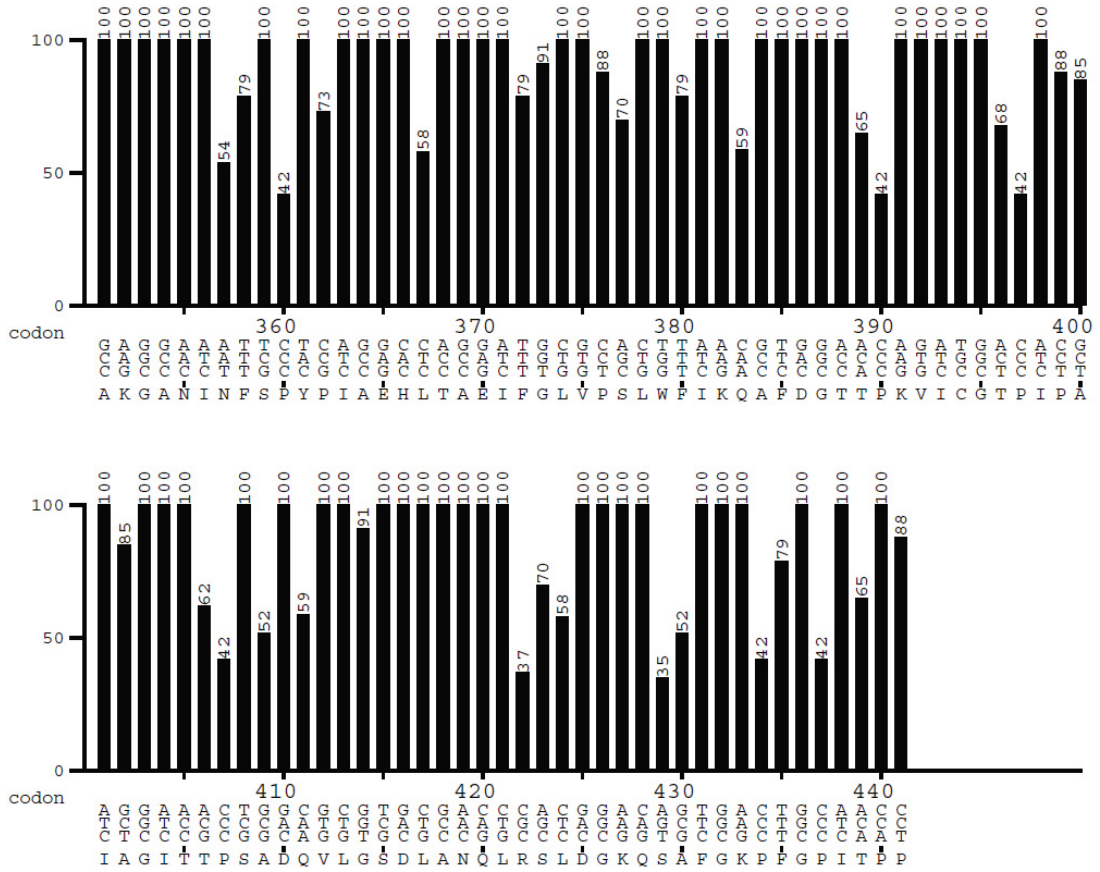
original sequence (151-350 aas)



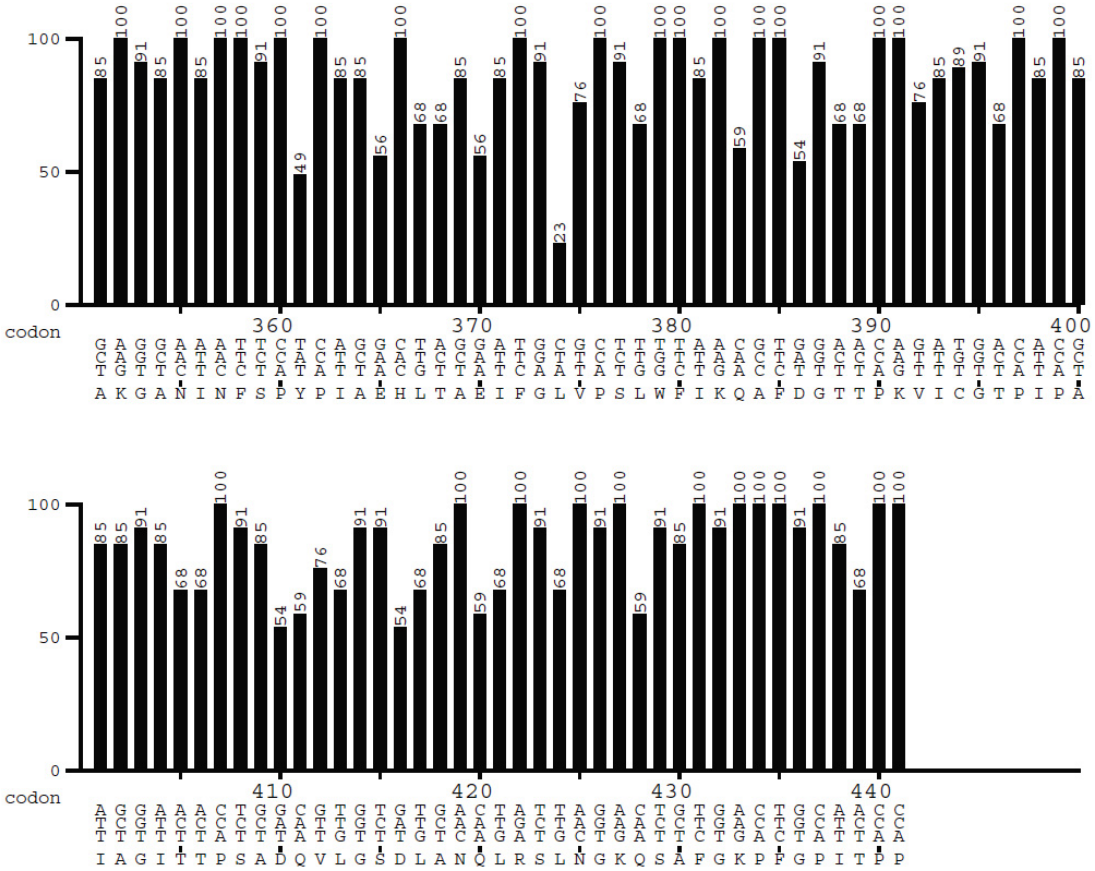
optimised sequence (151-350 aas)



original sequence (351-442 aas)



optimised sequence (351-442 aas)



6.3 Sequence alignment of the rhamnosidases used for this study.

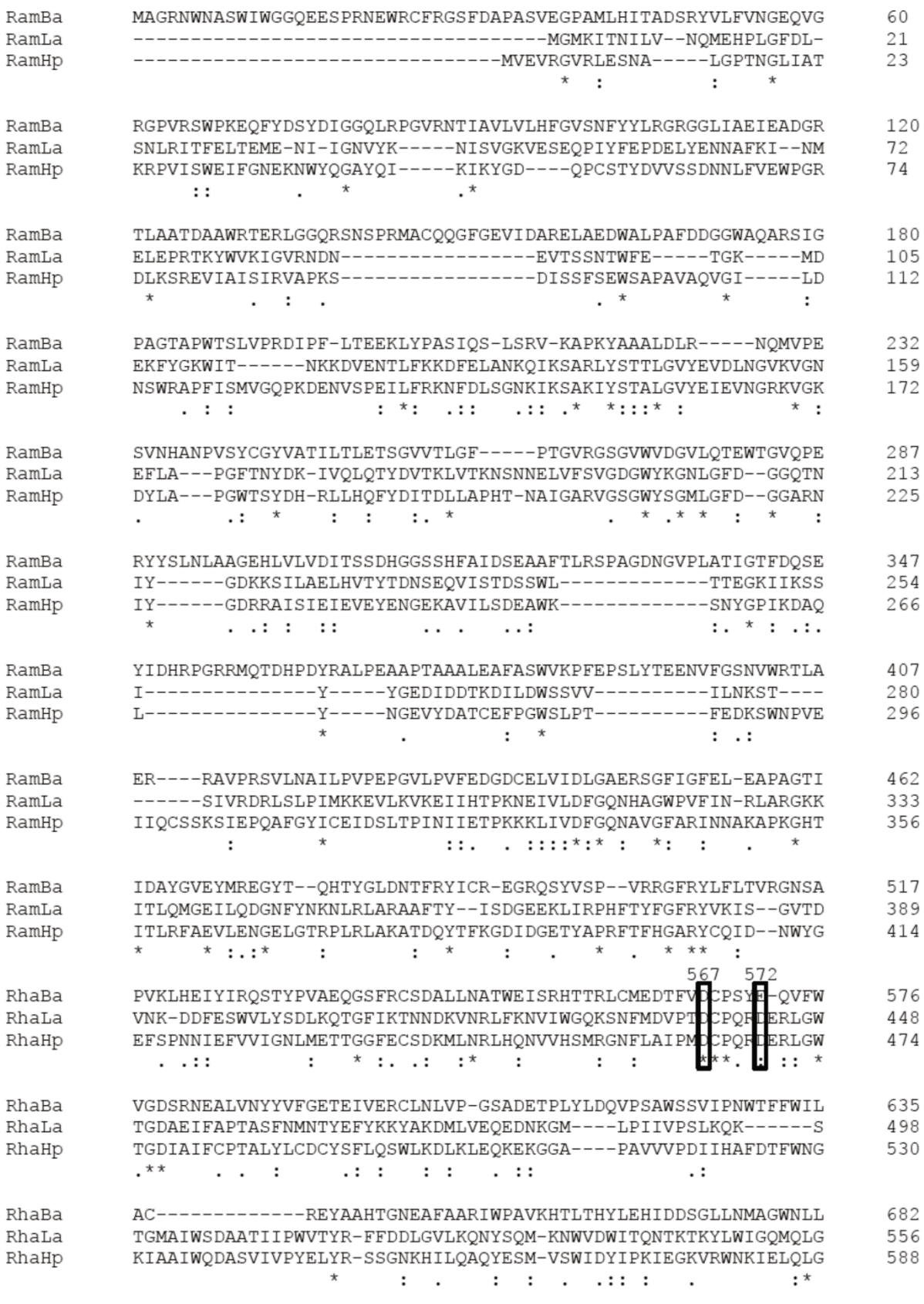


Figure 44: Sequence alignment of the rhamnosidases used for this study. The sequence of rhamnosidase Rha_{Hp} was identified in this study for the first time (see section 4.2.3 for details), while the two other sequences were already available. Asterisks - positions with fully conserved residue; colon - conserved groups with strongly similar properties; full stop - conserved groups with weakly similar properties. The amino acid residues which were elucidated to be crucial for the Rha_{Ba} activity are marked with black boxes, indicating that these regions are indeed conservative in GH family 78. The sequences were aligned using the CLUSTAL Omega tool.[189]

7. Literature

1. Fort, S., Boyer, V., Greffe, L., Davies, G.J., Moroz, O., Christiansen, L., Schulein, M., Cottaz, S., Driguez, H., *J. Am. Chem. Soc.* **2000**, *122*, 5429-5437. '*Highly Efficient Synthesis of β(1→4)-Oligo- and -Polysaccharides Using a Mutant Cellulase.*'
2. Baldrian, P. Lopez-Mondejar, R., *Appl. Microbiol. Biotechnol.* **2014**, *98*, 1531-1537. '*Microbial genomics, transcriptomics and proteomics: new discoveries in decomposition research using complementary methods.*'
3. Hutchison, C.A. 3rd, Phillips, S., Edgell, M.H., Gillam, S., Jahnke, P., Smith, M., *J. Biol. Chem.* **1978**, *253*, 6551-6560. '*Mutagenesis at a specific position in a DNA sequence.*'
4. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., *J. Mol. Biol.* **1990**, *215*, 403-10. '*Basic local alignment search tool.*'
5. Schmeisser, C., Steele, H., Streit, W.R., *Appl. Microbiol. Biotechnol.* **2007**, *75*, 955-962. '*Metagenomics, biotechnology with non-culturable microbes.*'
6. Gellissen, G. **2006**: John Wiley & Sons,'*Production of recombinant proteins: Novel microbial and eukaryotic expression systems.*'
7. Müller, S., Sandal, T., Kamp-Hansen, P., Dalbøge, H., *Yeast* **1998**, *14*, 1267–1283. '*Comparison of Expression Systems in the Yeasts Saccharomyces cerevisiae, Hansenula polymorpha, Klyveromyces lactis, Schizosaccharomyces pombe and Yarrowia lipolytica. Cloning of Two Novel Promoters from Yarrowia lipolytica.*'
8. Mack, M., Wannemacher, M., Hobl, B., Pietschmann, P., Hock, B., *Protein Expression Purif.* **2009**, *66*, 165-171. '*Comparison of two expression platforms in respect to protein yield and quality: Pichia pastoris versus Pichia angusta.*'
9. Han, M., Wang, X., Ding, H., Jin, M., Yu, L., Wang, J., Yu, X., *Enzyme Microb. Technol.* **2014**, *54*, 32-37. '*The role of N-glycosylation sites in the activity, stability, and expression of the recombinant elastase expressed by Pichia pastoris.*'
10. Godbole, S., Decker, S.R., Nieves, R.A., Adney, W.S., Vinzant, T.B., Baker, J.O., Thomas, S.R., Himmel, M.E., *Biotechnol. Prog.* **1999**, *15*, 828-833. '*Cloning and Expression of Trichoderma reesei Cellobiohydrolase I in Pichia pastoris.*'
11. Bui, D.M., Kunze, I., Förster, S., Wartmann, T., Horstmann, C., Manteuffel, R., Kunze, G., *Appl. Microbiol. Biotechnol.* **1996**, *44*, 610-619. '*Cloning and expression of an Arxula adeninivorans glucoamylase gene in Saccharomyces cerevisiae.*'
12. Bornscheuer, U.T., Huisman, G.W., Kazlauskas, R.J., Lutz, S., Moore, J.C., Robins, K., *Nature* **2012**, *485*, 185-194. '*Engineering the third wave of biocatalysis*'
13. Savile, C.K., Janey, J.M., Mundorff, E.C., Moore, J.C., Tam, S., Jarvis, W.R., Colbeck, J.C., Krebber, A., Fleitz, F.J., Brands, J., Devine, P.N., Huisman, G.W., Hughes, G.J., *Science* **2010**, *329*, 305-309. '*Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture.*'
14. Mackenzie, L.F., Wang, Q., Warren, R.A.J., Withers, S.G., *J. Am. Chem. Soc.* **1998**, *120*, 5583-5584. '*Glycosynthases: Mutant Glycosidases for Oligosaccharide Synthesis.*'
15. Yadav, V., Yadav, P.K., Yadav, S., Yadav, K.D.S., *Process Biochem.* **2010**, *45*, 1226-1235. '*α-l-Rhamnosidase: A review.*'
16. Beekwilder, J., Marcozzi, D., Vecchi, S., de Vos, R., Janssen, P., Francke, C., van Hylckama Vlieg, J., Hall, R.D., *Appl. Environ. Microbiol.* **2009**, *75*, 3447-54. '*Characterization of Rhamnosidases from Lactobacillus plantarum and Lactobacillus acidophilus.*'

17. Dewan, S. S., *BCCResearch* **2012**; Available from: www.bccresearch.com. '*Global Markets for Chiral Technology*'
18. Wohlgemuth, R., *Curr. Opin. Biotechnol.* **2010**, *21*, 713-724. '*Biocatalysis — key to sustainable industrial chemistry.*'
19. Aehle, W. **2007**: John Wiley & Sons, '*Enzymes in industry: production and applications.*'
20. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Erlich, H., *Cold Spring Harbor Symp. Quant. Biol.* **1986**, *51*, 263-273. '*Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction.*'
21. Gellissen, G., Hollenberg, C.P., *Gene* **1997**, *190*, 87-97. '*Application of yeasts in gene expression studies: a comparison of Saccharomyces cerevisiae, Hansenula polymorpha and Kluyveromyces lactis - a review.*'
22. Helenius, A., Aebi, M., *Annu. Rev. Biochem.* **2004**, *73*, 1019-1049. '*Roles of N-linked glycans in the endoplasmic reticulum.*'
23. Imperiali, B., O'Connor, S.E., *Curr. Opin. Chem. Biol.* **1999**, *3*, 643–649. '*Effect of N-linked glycosylation on glycopeptide and glycoprotein structure.*'
24. Rudd, P.M., Dwek, R.A., *Crit. Rev. Biochem. Mol. Biol.* **1997**, *32*, 1-100. '*Glycosylation Heterogeneity and the 3D structure of proteins.*'
25. Mitra, N., Sinha, S., Ramya, T.N., Surolia, A., *Trends Biochem. Sci.* **2006**, *31*, 156-63. '*N-linked oligosaccharides as outfitters for glycoprotein folding, form and function*'
26. Han, M., Wang, X., Ding, H., Jin, M., Yu, L., Wang, J., Yu, X., *Enzyme and Microbial Technology* **2014**, *54*, 32-37. '*The role of N-glycosylation sites in the activity, stability, and expression of the recombinant elastase expressed by Pichia pastoris*'
27. Bui, D.M., Kunze, I., Förster, S., Wartmann, T., Horstmann, C., Manteuffel, R., and Kunze, G., *Applied Microbiology and Biotechnology* **1996**, *44*, 610-619. '*Cloning and expression of an Arxula adeninivorans glucoamylase gene in Saccharomyces cerevisiae*'.
28. Skropeta, D., *Bioorg. Med. Chem.* **2009**, *17*, 2645-2653. '*The effect of individual N-glycans on enzyme activity*'.
29. Godbole, S., Decker, S.R., Nieves, R.A., Adney, W.S., Vinzant, T.B., Baker, J.O., Thomas, S.R., and Himmel, M.E., *Biotechnology Progress* **1999**, *15*, 828-833. '*Cloning and Expression of Trichoderma reesei Cellobiohydrolase I in Pichia pastoris*'.
30. Rocha, S.N., Abrahao-Neto, J., Cerdan, M.E., Gombert, A.K., and Gonzalez-Siso, M.I., *Appl. Microbiol. Biotechnol.* **2011**, *89*, 375-385. '*Heterologous expression of a thermophilic esterase in Kluyveromyces yeasts.*'.
31. Shusta, E.V., Raines, R.T., Pluckthun, A., and Wittrup, K.D., *Nat. Biotechnol.* **1998**, *16*, 773-777. '*Increasing the secretory capacity of Saccharomyces cerevisiae for production of single-chain antibody fragments.*'.
32. Kjeldsen, T., *Appl. Microbiol. Biotechnol.* **2000**, *54*, 277-286. '*Yeast secretory expression of insulin precursors.*'.
33. Bergkamp, R.J., Kool, I.M., Geerse, R.H., and Planta, R.J., *Curr. Genet.* **1992**, *21*, 365-70. '*Multiple-copy integration of the alpha-galactosidase gene from Cyamopsis tetragonoloba into the ribosomal DNA of Kluyveromyces lactis.*'.
34. Iwata, T., Tanaka, R., Suetsugu, M., Ishibashi, M., Tokunaga, H., Kikuchi, M., and Tokunaga, M., *Biotech. Lett.* **2004**, *26*, 1803-1808. '*Efficient secretion of human lysozyme from the yeast Kluyveromyces lactis.*'.

35. Gellissen, G., Janowicz, Z.A., Weydemann, U., Melber, K., Strasser, A.W.M., and Hollenberg, C.P., *Biotech. Adv.* **1992**, *10*, 179-189. '*High-level expression of foreign genes in Hansenula polymorpha.*'.
36. Mayer, A.F., Hellmuth, K., Schlieker, H., Lopez-Ulibarri, R., Oertel, S., Dahlems, U., Strasser, A.W.M., and van Loon, A.P.G.M., *Biotechnol. Bioeng.* **1999**, *63*, 373-381. '*An expression system matures: A highly efficient and cost-effective process for phytase production by recombinant strains of Hansenula polymorpha.*'.
37. Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S.G., *Science* **1996**, *274*, 546-567. '*Life with 6000 Genes*'.
38. Mattanovich, D., Branduardi, P., Dato, L., Gasser, B., Sauer, M., and Porro, D., *Methods Mol. Biol.* **2012**, *824*, 329-358. '*Recombinant protein production in yeasts.*'.
39. Huang, C., Jr., Lowe, A., and Batt, C., *Appl. Microbiol. Biotechnol.* **2010**, *87*, 401-410. '*Recombinant immunotherapeutics: current state and perspectives regarding the feasibility and market.*'.
40. Pannala, V.R., Bhartiya, S., and Venkatesh, K.V., *FEBS J.* **2010**, *277*, 2987-3002. '*Experimental and steady-state analysis of the GAL regulatory system in Kluyveromyces lactis.*'.
41. van Ooyen, A.J., Dekker, P., Huang, M., Olsthoorn, M.M., Jacobs, D.I., Colussi, P.A., and Taron, C.H., *FEMS Yeast Res.* **2006**, *6*, 381-392. '*Heterologous protein production in the yeast Kluyveromyces lactis.*'.
42. Van den Berg, J.A., Van der Laken, K.J., Van Ooyen, A.J.J., Renniers, T.C.H.M., Rietveld, K., Schaap, A., Brake, A.J., Bishop, R.J., Schultz, K., Moyer, D., Richman, M., and Shuster, J.R., *Nat. Biotechnol.* **1990**, *8*, 135-139. '*Kluyveromyces as a host for heterologous gene expression: Expression and secretion of prochymosin.*'.
43. Wésolowski-Louvel, M., Breunig, K.D., and Fukuhara, H. **1996**, Springer Berlin Heidelberg. p. 139-201. '*Kluyveromyces lactis*, in *Nonconventional Yeasts in Biotechnology*'
44. Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuveglise, C., Talla, E., Goffard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.M., Beyne, E., Bleykasten, C., Boisrame, A., Boyer, J., Cattolico, L., Confanioleri, F., De Daruvar, A., Despons, L., Fabre, E., Fairhead, C., Ferry-Dumazet, H., Groppi, A., Hantraye, F., Hennequin, C., Jauniaux, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R., Lemaire, M., Lesur, I., Ma, L., Muller, H., Nicaud, J.M., Nikolski, M., Oztas, S., Ozier-Kalogeropoulos, O., Pellenz, S., Potier, S., Richard, G.F., Straub, M.L., Suleau, A., Swennen, D., Tekaia, F., Wesolowski-Louvel, M., Westhof, E., Wirth, B., Zeniou-Meyer, M., Zivanovic, I., Bolotin-Fukuhara, M., Thierry, A., Bouchier, C., Caudron, B., Scarpelli, C., Gaillardin, C., Weissenbach, J., Wincker, P., and Souciet, J.L., *Nature* **2004**, *430*, 35-44. '*Genome evolution in yeasts*'.
45. Swaim, C.L., Anton, B.P., Sharma, S.S., Taron, C.H., and Benner, J.S., *Proteomics* **2008**, *8*, 2714-23. '*Physical and computational analysis of the yeast Kluyveromyces lactis secreted proteome*'.
46. Madinger, C.L., Sharma, S.S., Anton, B.P., Fields, L.G., Cushing, M.L., Canovas, J., Taron, C.H., and Benner, J.S., *Proteomics* **2009**, *9*, 4744-54. '*The effect of carbon source on the secretome of Kluyveromyces lactis*'.

47. Suh, S.O. and Zhou, J.L.J., *FEMS Yeast Res.* **2010**, *10*, 631-638. '*Methylotrophic yeast near Ogataea (Hansenula) polymorpha: a proposal of Ogataea angusta comb. nov and Candida parapolymorpha sp nov.*'.
48. Reinders, A., Romano, I., Wiemken, A., and De Virgilio, C., *J. Bacteriol.* **1999**, *181*, 4665-4668. '*The thermophilic yeast Hansenula polymorpha does not require trehalose synthesis for growth at high temperatures but does for normal acquisition of thermotolerance.*'.
49. Degelmann, A., Muller, F., Sieber, H., Jenzelewski, V., Suckow, M., Strasser, A.W.M., and Gellissen, G., *FEMS Yeast Res.* **2002**, *2*, 349-361. '*Strain and process development for the production of human cytokines in Hansenula polymorpha.*'.
50. Gellissen, G. **2002**, Weinheim: Wiley-VCH,'*Hansenula polymorpha: biology and applications.*'
51. Gustafsson, C., Govindarajan, S., and Minshull, J., *Trends Biotechnol.* **2004**, *22*, 346-353. '*Codon bias and heterologous protein expression.*'.
52. Lanza, A.M., Curran, K.A., Rey, L.G., and Alper, H.S., *BMC Syst. Biol.* **2014**, *8*, 1-10. '*A condition-specific codon optimization approach for improved heterologous gene expression in Saccharomyces cerevisiae.*'.
53. Angov, E., *Biotechnol. J.* **2011**, *6*, 650-659. '*Codon usage: Nature's roadmap to expression and folding of proteins.*'.
54. Leach, M.D. and Brown, A.J.P., *Eukaryotic Cell* **2012**, *11*, 98-108. '*Posttranslational Modifications of Proteins in the Pathobiology of Medically Relevant Fungi.*'.
55. Abu-Qarn, M., Eichler, J., and Sharon, N., *Curr. Opin. Struct. Biol.* **2008**, *18*, 544-550. '*Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea.*'.
56. Spiro, R.G., *Glycobiology* **2002**, *12*, 43R-56R. '*Protein glycosylation nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds*'.
57. Helenius, A. and Aebi, M., *Annu. Rev. Biochem.* **2004**, *73*, 1019-1049. '*Roles of N-linked glycans in the endoplasmic reticulum.*'.
58. Tanner, W. and Lehle, L., *Biochim. Biophys. Acta* **1987**, *906*, 81-99. '*Protein glycosylation in yeast.*'.
59. Wang, C.Q., Eufemi, M., Turano, C., and Giartosio, A., *Biochemistry* **1996**, *35*, 7299-7307. '*Influence of the carbohydrate moiety on the stability of glycoproteins.*'.
60. Wei, W., Chen, L., Zou, G., Wang, Q.F., Yan, X., Zhang, J., Wang, C.S., and Zhou, Z.H., *Biotechnol. Bioeng.* **2013**, *110*, 3075-3084. '*N-Glycosylation Affects the Proper Folding, Enzymatic Characteristics and Production of a Fungal beta-Glucosidase.*'.
61. Gemmill, T.R. and Trimble, R.B., *Biochim. Biophys. Acta* **1999**, *1426*, 227-237. '*Overview of N- and O-linked oligosaccharide structures found in various yeast species.*'.
62. Kang, H.A. and Gellissen, G., G. Gellissen, Editor. **2005**, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, FRG. p. 111-142.'*Hansenula polymorpha*, in *Production of Recombinant Proteins: Novel Microbial and Eukaryotic Expression Systems.'*
63. Jigami, Y. and Odani, T., *Biochim. Biophys. Acta* **1999**, *1426*, 335-345. '*Mannosylphosphate transfer to yeast mannan.*'.
64. De Pourcq, K., De Schutter, K., and Callewaert, N., *Appl. Microbiol. Biotechnol.* **2010**, *87*, 1617-31. '*Engineering of glycosylation in yeast and other fungi: current state and perspectives.*'.

65. Gentzsch, M. and Tanner, W., *Glycobiology* **1997**, *7*, 481-486. '*Protein-O-glycosylation in yeast: Protein-specific mannosyltransferases*'.
66. Gupta, R. **2001**, Technical University of Denmark.'*Prediction of glycosylation sites in proteomes: from post-translational modifications to protein function.*'
67. Immervoll, T., Gentzsch, M., and Tanner, W., *Yeast* **1995**, *11*, 1345-1351. '*PMT3 and PMT4, two new members of the protein-O-mannosyltransferase gene family of Saccharomyces cerevisiae*'.
68. Ferguson, M.A., *J Cell Sci.* **1999**, *112 (Pt 17)*, 2799-2809. '*The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research.*'.
69. Pfeffer, J., Richter, S., Nieveler, J., Hansen, C.E., Rhlid, R.B., Schmid, R.D., and Rusnak, M., *Appl. Microbiol. Biotechnol.* **2006**, *72*, 931-8. '*High yield expression of lipase A from Candida antarctica in the methylotrophic yeast Pichia pastoris and its purification and characterisation.*'.
70. Kirk, O. and Würtz Christensen, M., *Org. Process Res. Dev.* **2002**, *6*, 446-451. '*Lipases from Candida antarctica: Unique Biocatalysts from a Unique Origin.*'.
71. Hoegh, I., Patkar, S., Halkier, T., and Hansen, M.T., *Can. J. Bot.* **1996**, *73*, S869-S875. '*Two lipases from Candida antarctica cloning and expression in Aspergillus oryzae.*'.
72. Ericsson, D.J., Kasrayan, A., Johansson, P., Bergfors, T., Sandstrom, A.G., Backvall, J.E., and Mowbray, S.L., *J. Mol. Biol.* **2008**, *376*, 109-19. '*X-ray structure of Candida antarctica lipase A shows a novel lid structure and a likely mode of interfacial activation.*'.
73. Sandström, A.G., Engström, K., Nyhlén, J., Kasrayan, A., and Bäckvall, J.E., *Protein Eng., Des. Sel.* **2009**, *22*, 413-20. '*Directed evolution of Candida antarctica lipase A using an episomaly replicating yeast plasmid.*'.
74. Sandström, A.G., Wikmark, Y., Engström, K., Nyhlén, J., and Bäckvall, J.E., *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 78-83. '*Combinatorial reshaping of the Candida antarctica lipase A substrate pocket for enantioselectivity using an extremely condensed library.*'.
75. Panpipat, W., Xu, X., and Guo, Z., *Biochem. Eng. J.* **2013**, *70*, 55-62. '*Improved acylation of phytosterols catalyzed by Candida antarctica lipase A with superior catalytic activity.*'.
76. Kim, S., Choi, Y.K., Hong, J., Park, J.S., and Kim, M.-J., *Tetrahedron Lett.* **2013**, *54*, 1185-1188. '*Candida antarctica lipase A and Pseudomonas stutzeri lipase as a pair of stereocomplementary enzymes for the resolution of 1,2-diarylethanols and 1,2-diarylethanamines.*'.
77. Ramezani-Rad, M., Hollenberg, C.P., Lauber, J., Wedler, H., Griess, E., Wagner, C., Albermann, K., Hani, J., Piontek, M., Dahlems, U., and Gellissen, G., *FEMS Yeast Res.* **2003**, *4*, 207-215. '*The Hansenula polymorpha (strain CBS4732) genome sequencing and analysis.*'.
78. Gustafsson, C., Minshull, J., Govindarajan, S., Ness, J., Villalobos, A., and Welch, M., *Protein Expression Purif.* **2012**, *83*, 37-46. '*Engineering genes for predictable protein expression.*'.
79. Fuhrmann, M., Hausherr, A., Ferbitz, L., Schödl, T., Heitzer, M., and Hegemann, P., *Plant Mol. Biol.* **2004**, *55*, 869-881. '*Monitoring dynamic expression of nuclear genes in Chlamydomonas reinhardtii by using a synthetic luciferase reporter gene.*'.

80. Liljeblad, A., Kallio, P., Vainio, M., Niemi, J., and Kanerva, L.T., *Org. Biomol. Chem.* **2010**, *8*, 886-95. '*Formation and hydrolysis of amide bonds by lipase A from Candida antarctica; exceptional features.*'.
81. Fleer, R., *Curr. Opin. Biotechnol.* **1992**, *3*, 486-496. '*Engineering yeast for high level expression.*'.
82. Schuster, M., Einhauer, A., Wasserbauer, E., Süßenbacher, F., Ortner, C., Paumann, M., Werner, G., and Jungbauer, A., *J. Biotechnol.* **2000**, *84*, 237-248. '*Protein expression in yeast; comparison of two expression strategies regarding protein maturation.*'.
83. Chuang, G.Y., Boyington, J.C., Joyce, M.G., Zhu, J., Nabel, G.J., Kwong, P.D., and Georgiev, I., *Bioinformatics* **2012**, *28*, 2249-2255. '*Computational prediction of N-linked glycosylation incorporating structural properties and patterns.*'.
84. Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., and Stockinger, H. **2012**. p. W597-W603.'*ExPASy: SIB bioinformatics resource portal.*'
85. Esko, J., A. Varki, et al., Editors. **1999**, old Spring Harbor Laboratory Press: Cold Spring Harbor (NY).'*Natural and Synthetic Inhibitors of Glycosylation.*, in *Essentials of Glycobiology.'*
86. Levin, D.E., *Genetics* **2011**, *189*, 1145-1175. '*Regulation of cell wall biogenesis in Saccharomyces cerevisiae: the cell wall integrity signaling pathway.*'.
87. Michiyo, I. **1988**.'*Positionally non-specific lipase from Candida spp. A method for producing it, its use and a recombinant DNA process for producing it.*'
88. Ritchie, R.J. and Prvan, T., *Biochemical Education* **1996**, *24*, 196-206. '*Current statistical methods for estimating the Km and Vmax of Michaelis-Menten kinetics*'.
89. Eom, G.T., Lee, S.H., Song, B.K., Chung, K.-W., Kim, Y.-W., and Song, J.K., *J. Biosci. Bioeng.* **2013**, *116*, 165-170. '*High-level extracellular production and characterization of Candida antarctica lipase B in Pichia pastoris.*'.
90. Mitra, N., Sinha, S., Ramya, T.N., and Surolia, A., *Trends Biochem. Sci.* **2006**, *31*, 156-63. '*N-linked oligosaccharides as outfitters for glycoprotein folding, form and function.*'.
91. Solá, R.J. and Griebenow, K., *J. Pharm. Sci.* **2009**, *98*, 1223-45. '*Effects of glycosylation on the stability of protein pharmaceuticals.*'.
92. Skropeta, D., *Bioorg. Med. Chem.* **2009**, *17*, 2645-2653. '*The effect of individual N-glycans on enzyme activity.*'.
93. Zou, S., Xie, L., Liu, Y., Kaleem, I., Zhang, G., and Li, C., *J. Biotechnol.* **2012**, *157*, 399-404. '*N-linked glycosylation influences on the catalytic and biochemical properties of Penicillium purpurogenum beta-D-glucuronidase.*'.
94. Meldgaard, M. and Svendsen, I., *Microbiology* **1994**, *140 (Pt 1)*, 159-166. '*Different effects of N-glycosylation on the thermostability of highly homologous bacterial (1,3-1,4)-beta-glucanases secreted from yeast.*'.
95. Kim, S.Y., Sohn, J.H., Pyun, Y.R., and Choi, E.S., *J Microbiol. Biotechnol.* **2007**, *17*, 1949-54. '*Variations in protein glycosylation in Hansenula polymorpha depending on cell culture stage.*'.
96. Ingrell, C.R., Miller, M.L., Jensen, O.N., and Blom, N., *Bioinformatics* **2007**, *23*, 895-7. '*NetPhosYeast: prediction of protein phosphorylation sites in yeast.*'.
97. Jaenicke, R., *Phil. Trans. R. Soc. B* **1993**, *339*, 287-294. '*What does protein refolding in vitro tell us about protein folding in the cell?*'.
98. Morka, K., Pietruszka, J., and Meyer zu Berstenhorst, S., *J. Biotech.* **2014**, *191*, 176-186. '*Comparative expression of lipase CAL-A in the yeasts Saccharomyces*

cerevisiae, Kluyveromyces lactis and Hansenula polymorpha to investigate a possible host influence.'.

99. Mellitzer, A., Weis, R., Glieder, A., and Flicker, K., *Microb. Cell Fact.* **2012**, *11*, 1-11. '*Expression of lignocellulolytic enzymes in Pichia pastoris.*'.
100. Wang, Z.N., Wang, M.Y., Mei, W.L., Han, Z., and Dai, H.F., *Molecules* **2008**, *13*, 3033-3039. '*A new cytotoxic pregnanone from Calotropis gigantea.*'.
101. Svennerholm, L., Brane, G., Karlsson, I., Lekman, A., Ramstrom, I., and Wikkelso, C., *Dement. Geriatr. Cogn. Disord.* **2002**, *14*, 128-136. '*Alzheimer disease - Effect of continuous intracerebroventricular treatment with GM1 ganglioside and a systematic activation programme*'.
102. Song, J.-X., Sze, S.C.-W., Ng, T.-B., Lee, C.K.-F., Leung, G.P.H., Shaw, P.-C., Tong, Y., and Zhang, Y.-B., *J. Ethnopharmacol.* **2012**, *139*, 698-711. '*Anti-Parkinsonian drug discovery from herbal medicines: What have we got from neurotoxic models?*'.
103. Cobucci-Ponzano, B., Strazzulli, A., Rossi, M., and Moracci, M., *Adv. Synth. Catal.* **2011**, *353*, 2284-2300. '*Glycosynthases in Biocatalysis.*'.
104. Shaikh, F.A. and Withers, S.G., *Biochem. Cell Biol.* **2008**, *86*, 169-177. '*Teaching old enzymes new tricks: engineering and evolution of glycosidases and glycosyl transferases for improved glycoside synthesis.*'.
105. Desmet, T., Soetaert, W., Bojarova, P., Kren, V., Dijkhuizen, L., Eastwick-Field, V., and Schiller, A., *Chemistry* **2012**, *18*, 10786-801. '*Enzymatic glycosylation of small molecules: challenging substrates require tailored catalysts*'.
106. Nakagawa, H., Dobashi, Y., Sato, T., Yoshida, K., Tsugane, T., Shimura, S., Kirimura, K., Kino, K., and Usami, S., *J. Biosci. Bioeng.* **2000**, *89*, 138-144. '*Alpha-anomer-selective glucosylation of menthol with high yield through a crystal accumulation reaction using lyophilized cells of Xanthomonas campestris WU-9701.*'.
107. Kötzler, M.P., Hancock, S.M., and Withers, S.G. **2014**, John Wiley & Sons, Ltd.'*Glycosidases: Functions, Families and Folds.*, in *eLS*'
108. Giordano, A., Andreotti, G., Tramice, A., and Trincone, A., *Biotechnol. J.* **2006**, *1*, 511-530. '*Marine glycosyl hydrolases in the hydrolysis and synthesis of oligosaccharides.*'.
109. Trincone, A. and Giordano, A., *Curr. Org. Chem.* **2006**, *10*, 1163-1193. '*Glycosyl Hydrolases and Glycosyltransferases in the Synthesis of Oligosaccharides.*'.
110. Hancock, S.M., Vaughan, M.D., and Withers, S.G., *Curr. Opin. Chem. Biol.* **2006**, *10*, 509-519. '*Engineering of glycosidases and glycosyltransferases.*'.
111. Li, C. and Kim, Y.-W., *ChemBioChem* **2014**, *15*, 522-526. '*Characterization of a Galactosynthase Derived from Bacillus circulans β-Galactosidase: Facile Synthesis of D-Lacto- and D-Galacto-N-bioside.*'.
112. Vaughan, M., Johnson, K., DeFrees, S., Tang, X., Warren, R.A., and Withers, S., *J. Am. Chem. Soc.* **2006**, *128*, 6300-6301. '*Glycosynthase-Mediated Synthesis of Glycosphingolipids.*'.
113. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B., *Nucleic Acids Res.* **2014**, *42*, D490-D495. '*The carbohydrate-active enzymes database (CAZy) in 2013.*'.
114. Koshland, D.E., *Biol. Rev.* **1953**, *28*, 416-436. '*Stereochemistry and the mechanism of enzymatic reactions.*'.
115. Sulzenbacher, G., Driguez, H., Henrissat, B., Schulein, M., and Davies, G.J., *Biochemistry* **1996**, *35*, 15280-15287. '*Structure of the Fusarium oxysporum*

endoglucanase I with a nonhydrolyzable substrate analogue: substrate distortion gives rise to the preferred axial orientation for the leaving group.'.

116. Jahn, M., Marles, J., Warren, R.A., and Withers, S.G., *Angew. Chem., Int. Ed. Engl.* **2003**, *42*, 352-354. '*Thioglycoligases: mutant glycosidases for thioglycoside synthesis.*'.
117. Jahn, M., Chen, H., Mullegger, J., Marles, J., Warren, R.A., and Withers, S.G., *Chem. Commun. (Camb.)* **2004**, 2 7 4 – 2 7 5. '*Thioglycosynthases: double mutant glycosidases that serve as scaffolds for thioglycoside synthesis.*'.
118. Moracci, M., Trincone, A., Perugino, G., Ciaramella, M., and Rossi, M., *Biochemistry* **1998**, *37*, 17262-17270. '*Restoration of the Activity of Active-Site Mutants of the Hyperthermophilic β-Glycosidase from Sulfolobus solfataricus: Dependence of the Mechanism on the Action of External Nucleophiles.*'.
119. Perugino, G., Trincone, A., Giordano, A., Van Der Oost, J., Kaper, T., Rossi, M., and Moracci, M., *Biochemistry* **2003**, *42*, 8484-8493. '*Activity of Hyperthermophilic Glycosynthases Is Significantly Enhanced at Acidic pH.*'.
120. Drone, J., Feng, H., Tellier, C., Hoffmann, L., Tran, V., Rabiller, C., and Dion, M., *Eur. J. Org. Chem.* **2005**, *2005*, 1977-1983. '*Thermus thermophilus Glycosynthases for the Efficient Synthesis of Galactosyl and Glucosyl β-(1→3)-Glycosides.*'.
121. Hommalai, G., Withers, S.G., Chuenchor, W., Cairns, J.R., and Svasti, J., *Glycobiology* **2007**, *17*, 744-53. '*Enzymatic synthesis of cello-oligosaccharides by rice BGlu1 {beta}-glucosidase glycosynthase mutants*'.
122. Perugino, G., Falcicchio, P., Michela Corsaro, M., Matsui, I., Parrilli, M., Rossi, M., and Moracci, M., *Biocatal. Biotransform.* **2006**, *24*, 23-29. '*Preparation of a glycosynthase from the β-glycosidase of the Archaeon Pyrococcus horikoshii.*'.
123. Nashiru, O., Zechel, D.L., Stoll, D., Mohammadzadeh, T., Warren, R.A., and Withers, S.G., *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 417-420. '*β-Mannosynthase: Synthesis of β-Mannosides with a Mutant β-Mannosidase.*'.
124. Jakeman, D.L. and Withers, S.G., *Can. J. Chem.* **2002**, *80*, 866-870. '*On expanding the repertoire of glycosynthases: Mutant β-galactosidases forming β-(1,6)-linkages.*'.
125. Wilkinson, S.M., Liew, C.W., Mackay, J.P., Salleh, H.M., Withers, S., and McLeod, M.D., *Org. Lett.* **2008**, *10*, 1585-1588. '*Escherichia coli Glucuronylsynthase: An Engineered Enzyme for the Synthesis of α-Glucuronides.*'.
126. Mullegger, J., Chen, H.M., Chan, W.Y., Reid, S.P., Jahn, M., Warren, R.A., Salleh, H.M., and Withers, S.G., *Chem**B**ioChem* **2006**, *7*, 1028-1030. '*Thermostable glycosynthases and thioglycoligases derived from Thermotoga maritima beta-glucuronidase.*'.
127. Tao, H., Peralta-Yahya, P., Decatur, J., and Cornish, V.W., *ChemBioChem* **2008**, *9*, 681-684. '*Characterization of a new glycosynthase cloned by using chemical complementation.*'.
128. Spadiut, O., Ibatullin, F.M., Peart, J., Gullfot, F., Martinez-Fleites, C., Ruda, M., Xu, C., Sundqvist, G., Davies, G.J., and Brumer, H., *J. Am. Chem. Soc.* **2011**, *133*, 10892-10900. '*Building custom polysaccharides in vitro with an efficient, broad-specificity xyloglucan glycosynthase and a fucosyltransferase.*'.
129. Honda, Y. and Kitaoka, M., *J. Biol. Chem.* **2006**, *281*, 1426-1431. '*The first glycosynthase derived from an inverting glycoside hydrolase.*'.
130. Kim, Y.W., Fox, D.T., Hekmat, O., Kantner, T., McIntosh, L.P., Warren, R.A., and Withers, S.G., *Org. Biomol. Chem.* **2006**, *4*, 2025-2032. '*Glycosynthase-based synthesis of xylo-oligosaccharides using an engineered retaining xylanase from Cellulomonas fimi.*'.

131. Sugimura, M., Nishimoto, M., and Kitaoka, M., *Biosci. Biotechnol. Biochem.* **2006**, *70*, 1210-7. '*Characterization of glycosynthase mutants derived from glycoside hydrolase family 10 xylanases.*'.
132. Armstrong, Z., Reitinger, S., Kantner, T., and Withers, S.G., *ChemBioChem* **2010**, *11*, 533-8. '*Enzymatic thioxyloside synthesis: characterization of thioglycoligase variants identified from a site-saturation mutagenesis library of Bacillus circulans xylanase.*'.
133. Gullfot, F., Ibatullin, F.M., Sundqvist, G., Davies, G.J., and Brumer, H., *Biomacromolecules* **2009**, *10*, 1782–1788. '*Functional Characterization of Xyloglucan Glycosynthases from GH7, GH12, and GH16 Scaffolds.*'.
134. Malet, C. and Planas, A., *FEBS Lett.* **1998**, *440*, 208-212. '*From beta-glucanase to beta-glucansynthase: glycosyl transfer to alpha-glycosyl fluorides catalyzed by a mutant endoglucanase lacking its catalytic nucleophile.*'.
135. Van Lieshout, J., Faijes, M., Nieto, J., Van Der Oost, J., and Planas, A., *Archea* **2004**, *1*, 285-292. '*Hydrolase and glycosynthase activity of endo-1,3- -glucanase from the thermophile Pyrococcus furiosus.*'.
136. Piens, K., Henriksson, A.M., Gullfot, F., Lopez, M., Faure, R., Ibatullin, F.M., Teeri, T.T., Driguez, H., and Brumer, H., *Org. Biomol. Chem.* **2007**, *5*, 3971-3978. '*Glycosynthase activity of hybrid aspen xyloglucan endo-transglycosylase PttXET16-34 nucleophile mutants.*'.
137. Saura-Valls, M., Faure, R., Brumer, H., Teeri, T.T., Cottaz, S., Driguez, H., and Planas, A., *J. Biol. Chem.* **2008**, *283*, 21853-21863. '*Active-site mapping of a Populus xyloglucan endo-transglycosylase with a library of xylogluco-oligosaccharides.*'.
138. Vasur, J., Kawai, R., Jonsson, K.H.M., Widmalm, G., Engström, Å., Frank, M., Andersson, E., Hansson, H., Forsberg, Z., Igarashi, K., Samejima, M., Sandgren, M., and Ståhlberg, J., *J. Am. Chem. Soc.* **2010**, *132*, 1724–1730. '*Synthesis of Cyclic β-Glucan Using Laminarinase 16A Glycosynthase Mutant from the Basidiomycete Phanerochaete chrysosporium.*'.
139. Hrmova, M., Imai, T., Rutten, S.J., Fairweather, J.K., Pelosi, L., Bulone, V., Driguez, H., and Fincher, G.B., *J. Biol. Chem.* **2002**, *277*, 30102-30111. '*Mutated barley (1,3)-beta-D-glucan endohydrolases synthesize crystalline (1,3)-beta-D-glucans.*'.
140. Jahn, M., Stoll, D., Warren, R.A.J., Szabó, L., Singh, P., Gilbert, H.J., Ducros, V.M.A., Davies, G.J., and Withers, S.G., *Chem. Commun.* **2003**, 1327–1329. '*Expansion of the glycosynthase repertoire to produce defined manno-oligosaccharides.*'.
141. Cobucci-Ponzano, B., Conte, F., Bedini, E., Corsaro, M.M., Parrilli, M., Sulzenbacher, G., Lipski, A., Dal Piaz, F., Lepore, L., Rossi, M., and Moracci, M., *Chem Biol.* **2009**, *16*, 1097-108. '*beta-Glycosyl azides as substrates for alpha-glycosynthases: preparation of efficient alpha-L-fucosynthases.*'.
142. Okuyama, M., Mori, H., Watanabe, K., Kimura, A., and Chiba, S., *Biosci. Biotechnol. Biochem.* **2002**, *66*, 928-933. '*Alpha-glucosidase mutant catalyzes "alpha-glycosynthase"-type reaction.*'.
143. Kim, Y.W., Lovering, A., Chen, H., Kantner, T., McIntosh, L.P., Strynadka, N.C.J., and Withers, S.G., *J. Am. Chem. Soc.* **2006**, *128*, 2202-2203. '*Expanding the Thioglycoligase Strategy to the Synthesis of α -Linked Thioglycosides Allows Structural Investigation of the Parent Enzyme/Substrate Complex.*'.
144. Yamamoto, K. and Davis, B.G., *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 7449-7453. '*Creation of an alpha-mannosynthase from a broad glycosidase scaffold.*'.

145. Kim, Y.W., Chen, H.M., Kim, J.H., Mullegger, J., Mahuran, D., and Withers, S.G., *ChemBioChem* **2007**, *8*, 1495-1499. '*Thioglycoligase-based assembly of thiodisaccharides: screening as beta-galactosidase inhibitors.*'.
146. Cobucci-Ponzano, B., Zorzetti, C., Strazzulli, A., Carillo, S., Bedini, E., Corsaro, M.M., Comfort, D.A., Kelly, R.M., Rossi, M., and Moracci, M., *Glycobiology* **2011**, *21*, 448-456. '*A novel alpha-D-galactosynthase from Thermotoga maritima converts beta-D-galactopyranosyl azide to alpha-galacto-oligosaccharides.*'.
147. Ben-David, A., Bravman, T., Balazs, Y.S., Czjzek, M., Schomburg, D., Shoham, G., and Shoham, Y., *ChemBioChem* **2007**, *8*, 2145-2151. '*Glycosynthase activity of Geobacillus stearothermophilus GH52 beta-xylosidase: efficient synthesis of xylooligosaccharides from alpha-D-xylopyranosyl fluoride through a conjugated reaction.*'.
148. Umekawa, M., Li, C., Higashiyama, T., Huang, W., Ashida, H., Yamamoto, K., and Wang, L.X., *J. Biol. Chem.* **2010**, *285*, 511-521. '*Efficient glycosynthase mutant derived from Mucor hiemalis endo-beta-N-acetylglucosaminidase capable of transferring oligosaccharide from both sugar oxazoline and natural N-glycan.*'.
149. Wada, J., Honda, Y., Nagae, M., Kato, R., Wakatsuki, S., Katayama, T., Taniguchi, H., Kumagai, H., Kitaoka, M., and Yamamoto, K., *FEBS Lett.* **2008**, *582*, 3739-3743. '*1,2-alpha-l-Fucosynthase: a glycosynthase derived from an inverting alpha-glycosidase with an unusual reaction mechanism.*'.
150. Ben-David, A., Shoham, G., and Shoham, Y., *Chem. Biol.* **2008**, *15*, 546-51. '*A universal screening assay for glycosynthases: directed evolution of glycosynthase XynB2(E335G) suggests a general path to enhance activity.*'.
151. Honda, Y., Fushinobu, S., Hidaka, M., Wakagi, T., Shoun, H., Taniguchi, H., and Kitaoka, M., *Glycobiology* **2008**, *18*, 325-30. '*Alternative strategy for converting an inverting glycoside hydrolase into a glycosynthase*'.
152. Ducros, V.M.A., Tarling, C.A., Zechel, D.L., Brzozowski, A.M., Frandsen, T.P., von Ossowski, I., Schülein, M., Withers, S.G., and Davies, G.J., *Chem Biol.* **2003**, *10*, 619-628. '*Anatomy of Glycosynthesis: Structure and Kinetics of the Humicola insolens Cel7B E197A and E197S Glycosynthase Mutants.*'.
153. Hancock, S.M., Rich, J.R., Caines, M.E.C., Strynadka, N.C.J., and Withers, S.G., *Nat. Chem. Biol.* **2009**, *5*, 508-514. '*Designer enzymes for glycosphingolipid synthesis by directed evolution.*'.
154. Trincone, A., Giordano, A., Perugino, G., Rossi, M., and Moracci, M., *ChemBioChem* **2005**, *6*, 1431-1437. '*Highly productive autocondensation and transglycosylation reactions with Sulfolobus solfataricus glycosynthase.*'.
155. Puri, M., *Appl. Microbiol. Biotechnol.* **2012**, *93*, 49-60. '*Updates on naringinase: structural and biotechnological aspects.*'.
156. Rojas, N.L., Voget, C.E., Hours, R.A., and Cavalitto, S.F., *J. Ind. Microbiol. Biotechnol.* **2011**, *38*, 1515-1522. '*Purification and characterization of a novel alkaline alpha-L-rhamnosidase produced by Acrostalagmus luteo albus.*'.
157. Manzanares, P., van den Broeck, H.C., de Graaff, L.H., and Visser, J., *Appl. Environ. Microbiol.* **2001**, *67*, 2230-2234. '*Purification and characterization of two different alpha-L-rhamnosidases, RhaA and RhaB, from Aspergillus aculeatus.*'.
158. Scaroni, E., Cuevas, C., Carrillo, L., and Ellenrieder, G., *Lett. Appl. Microbiol.* **2002**, *34*, 461-465. '*Hydrolytic properties of crude alpha-L-rhamnosidases produced by several wild strains of mesophilic fungi.*'.
159. Koseki, T., Mese, Y., Nishibori, N., Masaki, K., Fujii, T., Handa, T., Yamane, Y., Shiono, Y., Murayama, T., and Iefuji, H., *Appl. Microbiol. Biotechnol.* **2008**, *80*, 1007-1013. '*Characterization of an alpha-L-rhamnosidase from Aspergillus kawachii and its gene.*'.

160. Manzanares, P., Orejas, M., Ibanez, E., Valles, S., and Ramon, D., *Lett Appl. Microbiol.* **2000**, *31*, 198-202. '*Purification and characterization of an alpha-L-rhamnosidase from Aspergillus nidulans.*'.
161. Manzanares, P., de Graaff, L.H., and Visser, J., *FEMS Microbiol. Lett.* **1997**, *157*, 279-283. '*Purification and characterization of an alpha-L-rhamnosidase from Aspergillus niger.*'.
162. Gallego, M.V., Piñaga, F., Ramón, D., and Vallés, S., *J. Food Sci.* **2001**, *66*, 204-209. '*Purification and Characterization of an α-L-Rhamnosidase from Aspergillus terreus of Interest in Winemaking.*'.
163. Cui, Z., Maruyama, Y., Mikami, B., Hashimoto, W., and Murata, K., *J. Mol. Biol.* **2007**, *374*, 384-398. '*Crystal structure of glycoside hydrolase family 78 alpha-L-Rhamnosidase from Bacillus sp. GL1.*'.
164. Jang, I.S. and Kim, D.H., *Biol. Pharm. Bull.* **1996**, *19*, 1546-1549. '*Purification and characterization of alpha-L-rhamnosidase from Bacteroides JY-6, a human intestinal bacterium.*'.
165. Puri, M., Kaur, A., Singh, R.S., Schwarz, W.H., and Kaur, A., *Process Biochem.* **2010**, *45*, 451-456. '*One-step purification and immobilization of His-tagged rhamnosidase for naringin hydrolysis.*'.
166. Kaji, A. and Ichimi, T., *Agric. Biol. Chem.* **1973**, *37*, 431-432. '*α-Rhamnosidase Activity in Culture Filtrate of Corticium-Rolfsii Enzymatic Activity at Low pH.*'.
167. Bourbouze, R., Percheron, F., and Courtois, J.-E., *Eur. J. Biochem.* **1976**, *63*, 331-337. '*α-L-Rhamnosidase de Fagopyrum esculentum.*'.
168. Park, S.Y., Kim, J.H., and Kim, D.H., *J. Microbiol. Biotechnol.* **2005**, *15*, 519-524. '*Purification and characterization of quercitrin-hydrolyzing alpha-L-rhamnosidase from Fusobacterium K-60, a human intestinal bacterium.*'.
169. Avila, M., Jaquet, M., Moine, D., Requena, T., Pelaez, C., Arigoni, F., and Jankovic, I., *Microbiology* **2009**, *155*, 2739-2749. '*Physiological and biochemical characterization of the two alpha-L-rhamnosidases of Lactobacillus plantarum NCC245.*'.
170. Michlmayr, H., Brandes, W., Eder, R., Schumann, C., del Hierro, A.M., and Kulbe, K.D., *Appl. Environ. Microbiol.* **2011**, *77*, 6524-6530. '*Characterization of two distinct glycosyl hydrolase family 78 alpha-L-rhamnosidases from Pediococcus acidilactici.*'.
171. Yanai, T. and Sato, M., *Biosci. Biotechnol. Biochem.* **2000**, *64*, 2179. '*Purification and characterization of an alpha-L-rhamnosidase from Pichia angusta X349.*'.
172. Rodríguez, M.E., Lopes, C.A., Valles, S., and Caballero, A.C., *J. Appl. Microbiol.* **2010**, *109*, 2206-2213. '*Characterization of α-rhamnosidase activity from a Patagonian Pichia guilliermondii wine strain.*'.
173. Gastón Orrillo, A., Ledesma, P., Delgado, O.D., Spagna, G., and Breccia, J.D., *Enzyme Microb. Technol.* **2007**, *40*, 236-241. '*Cold-active α-l-rhamnosidase from psychrotolerant bacteria isolated from a sub-Antarctic ecosystem.*'.
174. Miake, F., Satho, T., Takesue, H., Yanagida, F., Kashige, N., and K., W., *Arch. Microbiol.* **2000**, *173*, 65-70. '*Purification and characterization of intracellular α-l -rhamnosidase from Pseudomonas paucimobilis FP2001.*'.
175. Shanmugam, V. and Yadav, K.D., *Indian J. Exp. Biol.* **1995**, *33*, 705-707. '*Extracellular production of alpha-rhamnosidase by Rhizopus nigricans.*'.
176. Miyata, T., Kashige, N., Satho, T., Yamaguchi, T., Aso, Y., and Miake, F., *Curr. Microbiol.* **2005**, *51*, 105-109. '*Cloning, Sequence Analysis, and Expression of the Gene Encoding Sphingomonas paucimobilis FP2001 α-L-Rhamnosidase.*'.

177. Ichinose, H., Fujimoto, Z., and Kaneko, S., *Biosci., Biotechnol., Biochem.* **2013**, *77*, 213-216. '*Characterization of an alpha-L-Rhamnosidase from Streptomyces avermitilis.*'.
178. Qian, S., Yu, H., Zhang, C., Lu, M., Wang, H., and Jin, F., *Chem. Pharm. Bull.* **2005**, *53*, 911-914. '*Purification and characterization of dioscin-alpha-L-rhamnosidase from pig liver.*'.
179. Birgisson, H., Hreggvidsson, G.O., Fridjónsson, O.H., Mort, A., Kristjánsson, J.K., and Mattiasson, B., *Enzyme Microb. Technol.* **2004**, *34*, 561-571. '*Two new thermostable α-l-rhamnosidases from a novel thermophilic bacterium.*'.
180. Kurosawa, Y., Ikeda, K., and Egami, F., *J. Biochem.* **1973**, *73*, 31-37. '*Alpha-L-Rhamnosidases of Liver of Turbo cornutus and Aspergillus niger.*'.
181. Romero, C., Manjón, A., Bastida, J., and Iborra, J., *Anal. Biochem.* **1985**, *149*, 566-571. '*A method for assaying the rhamnosidase activity of naringinase.*'.
182. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E., *Nucleic Acids Res.* **2000**, *28*, 235-242. '*The Protein Data Bank.*'.
183. Hashimoto, W., Miyake, O., Nankai, H., and Murata, K., *Arch. Biochem. Biophys.* **2003**, *415*, 235-244. '*Molecular identification of an α-l-rhamnosidase from Bacillus sp. strain GL1 as an enzyme involved in complete metabolism of gellan.*'.
184. Wells, J.A., Vasser, M., and Powers, D.B., *Gene* **1985**, *34*, 315-23. '*Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites*'.
185. Papworth, C., Bauer, J.C., Braman, J., and Wright, D.A., *Strategies* **1996**, *9*, 3-4. '*Site-directed mutagenesis in one day with >80% efficiency*.'.
186. Lapid, C. and Gao, Y. 2003; Available from: http://www.bioinformatics.org/primerx/index.htm.'*PrimerX- Automated design of mutagenic primers for site-directed mutagenesis.*'.
187. Geiser, M., Cebe, R., Drewello, D., and Schmitz, R., *Biotechniques* **2001**, *31*, 88-92. '*Integration of PCR fragments at any specific site within cloning vectors without the use of restriction enzymes and DNA ligase.*'.
188. Fujimoto, Z., Jackson, A., Michikawa, M., Maehara, T., Momma, M., Henrissat, B., Gilbert, H.J., and Kaneko, S., *J. Biol. Chem.* **2013**, *288*, 12376-12385. '*The Structure of a Streptomyces avermitilis α-l-Rhamnosidase Reveals a Novel Carbohydrate-binding Module CBM67 within the Six-domain Arrangement.*'.
189. McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley, A.P., and Lopez, R., *Nucleic Acids Res.* **2013**, *41*, W597-600. '*Analysis Tool Web Services from the EMBL-EBI.*'.
190. Cobucci-Ponzano, B., Conte, F., Mazzone, M., Bedini, E., Corsaro, M.M., Rossi, M., and Moracci, M., *Biocatal. Biotransform.* **2008**, *26*, 18-24. '*Design of new reaction conditions for characterization of a mutant thermophilic α-L-fucosidase.*'.
191. Beekwilder, J. **2010**: Plant Research International, P.O. Box 16, 6700 AA Wageningen, The Netherlands.'*Personal communication*'
192. Grigoriev, I.V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A., Minovitsky, S., Nikitin, R., Ohm, R.A., Otillar, R., Poliakov, A., Ratnere, I., Riley, R., Smirnova, T., Rokhsar, D., and Dubchak, I., *Nucleic Acids Res.* **2012**, *40*, D26-32. '*The genome portal of the Department of Energy Joint Genome Institute*'.
193. Larsson, A.M. and Jones, T.A., *Methods Mol. Biol.* **2007**, *389*, 165-174. '*Selenomethionine labeling of recombinant proteins.*'.
194. Kitajima, T., Jigami, Y., and Chiba, Y., *J. Biol. Chem.* **2012**, *287*, 10032-10038. '*Cytotoxic Mechanism of Selenomethionine in Yeast.*'.

195. Robinson, M.A., Charlton, S.T., Garnier, P., Wang, X.-T., Davis, S.S., Perkins, A.C., Frier, M., Duncan, R., Savage, T.J., Wyatt, D.A., Watson, S.A., and Davis, B.G., *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 14527-14532. '*LEAPT: Lectin-directed enzyme-activated prodrug therapy.*'.
196. Nagae, M., Tsuchiya, A., Katayama, T., Yamamoto, K., Wakatsuki, S., and Kato, R., *J. Biol. Chem.* **2007**, *282*, 18497-18509. '*Structural basis of the catalytic reaction mechanism of novel 1,2-alpha-L-fucosidase from Bifidobacterium bifidum.*'.
197. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E., *J. Comput. Chem.* **2004**, *25*, 1605-1612. '*UCSF Chimera--a visualization system for exploratory research and analysis.*'.
198. Corpet, F., *Nucl. Acids Res.* **1988**, *16*, 10881-10890. '*Multiple sequence alignment with hierarchical clustering.*'.
199. Nakamura, Y., Gojobori, T., and Ikemura, T., *Nucleic Acids Res.* **2000**, *28*, 292. '*Codon usage tabulated from international DNA sequence databases: status for the year 2000.*'.
200. Meselson, M. and Yuan, R., *Nature* **1968**, *217*, 1110-1114. '*DNA Restriction Enzyme from E. coli.*'.
201. Bloom, F.R., Schmidt, B.J., and Lin, J.J. **2010**, Google Patents.'*Rapid growing microorganisms for biotechnology applications*'
202. Studier, F.W. and Moffatt, B.A., *J. Mol. Biol.* **1986**, *189*, 113-130. '*Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-level Expression of Cloned Genes.*'.
203. Brachmann, C., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., and Boeke, J.D., *Yeast* **1998**, *14*, 115-132. '*Designer deletion strains derived from Saccharomyces cerevisiae S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications.*'.
204. Colussi, P.A. and Taron, C.H., *Appl. Environ. Microbiol.* **2005**, *71*, 7092-7098. '*Kluyveromyces lactis LAC4 promoter variants that lack function in bacteria but retain full function in K. lactis.*'.
205. Lahtchev, K.L., Semenova, V.D., Tolstorukov, II, van der Klei, I., and Veenhuis, M., *Arch. Microbiol.* **2002**, *177*, 150-158. '*Isolation and properties of genetically defined strains of the methylotrophic yeast Hansenula polymorpha CBS4732.*'.
206. Sambrook, J. and Fritsch, E.F.M., T. 2nd ed ed. **1989**, New York: Cold Spring Harbor Laboratory Press,'*Molecular cloning : a laboratory manual / J. Sambrook, E.F. Fritsch, T. Maniatis*'
207. Dohmen, R.J., Strasser, A.W.M., Höner, C.B., and Hollenberg, C.P., *Yeast* **1991**, *7*, 691-692. '*An efficient transformation procedure enabling long-term storage of competent cells of various yeast genera*'.
208. Coulter, H.W. **1953**, Google Patents.'*Means for counting particles suspended in a fluid.*'
209. Higuchi, R., Krummel, B., and Saiki, R.K., *Nucleic Acids Res.* **1988**, *16*, 7351-67. '*A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions*'.
210. Heckman, K.L. and Pease, L.R., *Nat. Protocols* **2007**, *2*, 924-932. '*Gene splicing and mutagenesis by PCR-driven overlap extension.*'.
211. Sørensen, S.P.L., *Biochem. Z.* **1909**, *21*, 131-304. '*Enzyme studies - Note II - The measurement and the significance of hydrogenic concentrate in enzymatic processes.*'.

212. Martinelle, M., Holmquist, M., and Hult, K., *Biochim. Biophys. Acta* **1995**, *1258*, 272-276. '*On the interfacial activation of Candida antarctica lipase A and B as compared with Humicola lanuginosa lipase.*'.
213. Bradford, M.M., *Anal. Biochem.* **1976**, *72*, 248-254. '*A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding*'.
214. Winkler, U.K. and Stuckmann, M., *J.Bacteriol.* **1979**, *138*, 663-670. '*Glycogen, hyaluronate, and some other polysaccharides greatly enhance the formation of exolipase by Serratia marcescens.*'.
215. Franken, B., Jaeger, K.E., and Pietruszka, J., K. Timmis, Editor. **2010**, Springer Berlin Heidelberg. p. 4581-4586.'*Protocols to Screen for Enantioselective Lipases.*, in *Handbook of Hydrocarbon and Lipid Microbiology*'
216. Pietruszka, J., Simon, R.C., Kruska, F., and Braun, M., *Eur. J. Org. Chem.* **2009**, *2009*, 6217-6224. '*Dynamic Enzymatic Kinetic Resolution of Methyl 2,3-Dihydro-1H-indene-1-carboxylate.*'.
217. Reetz, M.T., Prasad, S., Carballeira, J.D., Gumulya, Y., and Bocola, M., *J. Am. Chem. Soc.* **2010**, *132*, 9144-9152. '*Iterative Saturation Mutagenesis Accelerates Laboratory Evolution of Enzyme Stereoselectivity: Rigorous Comparison with Traditional Methods.*'.
218. Sharp, P.M. and Li, W.H., *Nucleic Acids Res.* **1987**, *15*, 1281-1295. '*The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications.*'.

8. Acknowledgements

Completion of this doctoral dissertation was only possible with the support of many people. I would like to express my sincere gratitude to all of them.

First, I would like to thank my supervisors for giving me the opportunity to complete my study at the Heinrich Heine University Düsseldorf. I thank Dr. Sonja Meyer zu Berstenhorst for the amazing job she did both as supervisor as well as a CLIB-Graduate Cluster coordinator. Her guidance helped me in all the time of research and writing of the thesis.

I owe my deepest gratitude to Prof. Dr. Jörg Pietruszka for his supervision during my time at Institute for Bioorganic Chemistry. I truly valued his advices and cooperation during the research. He has always made himself available to clarify my doubts despite his busy schedules and I truly appreciate that.

I would also like to thank Prof. Dr. Karl-Erich Jaeger for his supervision and friendly support during this time.

I am indebted to the CLIB-Graduate Cluster Industrial Biotechnology for the scholarship and all the additional oportunities to exchange the ideas with other scholarship holders. I thank Jessica Hilbig and Jennifer Pahlke for their help whenever it was necessary.

Special thanks for all present and former IBOC collegues, who have made this time unforgettable. There are two persons I need to thank especially, Katharina Neufeld and Benjamin Lauinger, for their enormous support during my first days in Germany, brilliant advices and friendship.

I was lucky to cooperate with some very smart people. I am very grateful to Dr. Oliver H. Weiergräber (Institute of Complex Systems, ICS-6) for our insightful and detailed discussions about protein crystallisation. I would also like to thank Natalie Hoffmann for the synthesis of the substrates and her valuable help during all the experiments.

For the amazing technical assistance, I am very grateful to Birgit Henßen, Bea Paschold, Monika Gehsing, Rainer Goldbaum, Vera Ophoven and Erik Kranz. My students - Shawn Whitfield, Dominik Schumacher and Ines Mancinella - I thank for the great time we had together in the lab.

I am also thankful to Shawn Whitfield, Marcus Dick and Ben Berntsson for doing the arduous work of reading and correcting the thesis.

I would like to thank my industrial partner, Artes Biotechnology GmbH, for providing the *H. polymorpha* strain and plasmids, as well as for sharing the methodology and extremely helpful Hands-On Training at the beginning of my work with the *H. polymorpha*.

During this study, I completed an industrial internship at BRAIN AG. It was very intensive time and I have learned a lot there. Special thanks to Dr. Klaus Liebeton for giving me this unique chance and for all his support during the internship. His wisdom, knowledge and commitment to the highest standards inspired and motivated me.

I would like to thank my amazing family for all their love and encouragement.
Finally, I owe much to my husband, without whose love and understanding I would not have completed this work.

Kochani Rodzice, drogi Mężu, dziękuję Wam z całego serca za Wasze wsparcie przez te wszystkie lata!

9. Statement

Hiermit erkläre ich gemaß § 3 Abs. 1 der Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fakultät der Heinrich-Heine-Universität Düssseldorf vom 13.10.2008, dass die vorliegende Arbeit selbstständig angefertigt wurde und ausschließlich die angegeben Quelle und Hilfsmittel verwendet wurden. Ich versichere, dass ich bisher keine Promotionsversuche unternommen habe und die vorliegende Arbeit bei keiner anderen Institution eingereicht wurde.

Düsseldorf, Datum

Kamila Morka