Challenges in Biocatalysis –
Case Studies on Protein Expression, Engineering and Application

Herausforderungen in der Biokatalyse –
Fallstudien zur Expression, Optimierung und Anwendung von Proteinen

an approved thesis presented to the
Faculty of Energy Technology, Process Engineering and
Biological Engineering of the University of Stuttgart
in fulfillment of the requirements for the
Degree of Doctor in Natural Sciences (Dr. rer. nat.)

submitted by
Christian Kazenwadel
from
Schorndorf, Germany

Main examiner: Prof. Dr. Bernhard Hauer
Co-examiner: Prof. Dr. Andreas Schaller
Thesis defense date: 14.11.2013

Institute of Technical Biochemistry, University of Stuttgart
2013

I. Declaration of Authorship

I hereby declare that the present thesis entitled "Challenges in Biocatalysis – Case Studies on Protein Expression, Engineering and Application" is the result of my own work, that all sources used or quoted have been indicated, and that I have not used any illegitimate means. I further declare that I have not submitted this thesis for a degree in some form or another.

Erklärung über die Eigenständigkeit der Dissertation

Ich versichere, dass ich die vorliegende Arbeit mit dem Titel „Herausforderungen in der Biokatalyse – Fallstudien zur Expression, Optimierung und Anwendung von Proteinen" selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Aus fremden Quellen entnommene Passagen und Gedanken sind als solche kenntlich gemacht. Des Weiteren bestätige ich ausdrücklich, dass die hier vorgelegte Dissertation nicht in gleicher oder ähnlicher Form bei einer anderen Institution zur Erlangung eines akademischen Grades eingereicht wurde

Name/name: Christian Kazenwadel

Unterschrift/signature: ______________________________________

Ort und Datum/place and date: ____________________________________

II. Publications

Part of this work has already been published. All results that have already been published are indicated according to the following references:

I. [215] Kazenwadel, C. *et al.* (2012) Optimized expression of the dirigent protein *At*DIR6 in *Pichia pastoris* and impact of glycosylation on protein structure and function. *Applied Microbiology and Biotechnology*. doi: 10.1007/s00253-012-4579-x

II. [219] Kazenwadel, C. *et al.* (2012) Thiol-functionalization of acrylic ester monomers catalyzed by immobilized *Humicola insolens* cutinase. *Enzyme and Microbial Technology*. 51, 9–15

III. Acknowledgments

This research project was carried out from February 2010 to March 2013 at the Institute of Technical Biochemistry at the University of Stuttgart. The individual parts of the project were funded by the BASF SE (Germany), Royal DSM (The Netherlands), and the Deutsche Forschungsgemeinschaft (Germany, Sonderforschungsbereich 706, Teilprojekt B4).

First, I would like to thank my Ph.D. supervisor Prof. Dr. Bernhard Hauer for the opportunity to work on this thesis at the Institute of Technical Biochemistry at the University of Stuttgart. I am especially thankful for the freedom to pursue many different challenges in the scope of this thesis and the continued support and encouragement throughout the entire time.

I am very grateful to Prof. Dr. Andreas Schaller for being the co-examiner of my thesis as well as for his support and helpful discussions during our academic cooperation.

Special thanks are in order for my two supervisors, Dr. Sandra Facey and Dr. Janosch Klebensberger. Thank you so much for all the time you invested not only towards the end while critically reading my thesis, but also for all your support, helpful discussions, and inspiration during these three years!

In this regard, I also want to thank Dr. Katja Koschorreck as my supervisor during my diploma thesis and for our cooperation regarding the cutinase project.

I would like to thank Dr. Bernd Nebel for invaluable practical help with many analytical challenges. I also owe many thanks to M. Eng. Sven Richter for all his hard work and support with fermentations and protein purifications. You guys helped a lot, thanks so much!

While not directly involved in my projects, I would also like to express my feline thanks to Dr. Bettina M. Nestl for enlightening me in so many discussions about biocatalysis and enzyme kinetics.

Many thanks as well to Dr. Jürgen Pleiss and his Bioinformatics team for the common work on the ribokinase project.

My sincere thanks go to all our collaboration partners from the University of Hohenheim including Dr. Uwe Gerken, Dr. Domenico Lupo, and Dr. Jens Pfannstiel. Furthermore, I would

like to thank our industrial cooporation partners from DSM, Dr. Oliver May, Dr. René de Jong, and Paul de Waal.

I would like to thank all my colleagues and friends in the lab for the fantastic support, company, and fun time together. I will not write down the names, you guys know who I am speaking of. I really hope these friendships will last a lifetime!

Special thanks go to my family, especially my parents, who supported me in all things and particularly invested so much time (and money!) in my education over so many years.

Last but not least, I need to give so much credit and all my love to my wife and daughter. Kathrin, this would not have been possible without you and all your support! Leonie Jule, thank you for sleeping peacefully in my arms for so many hours while I was finishing up this thesis!

IV. Table of contents

I. Declaration of Authorship 2

II. Publications 3

III. Acknowledgments 4

IV. Table of contents 6

V. Abstract 12

VI. Zusammenfassung 14

VII. Abbreviations 17

1. Introduction 20

1.1 Challenges in biocatalysis 20

1.1.1 Industrial biocatalysis 20

1.1.2 The biocatalysis cycle 22

1.1.3 Biocatalyst production 24

1.1.4 Biocatalyst characterization and engineering 28

1.1.5 Biocatalyst application and downstream-processing 29

1.2 Dirigent proteins 31

1.2.1 Secondary plant metabolites 31

1.2.2 Lignans 31

1.2.3 Stereoselective control of lignan biosynthesis by dirigent proteins 35

1.2.4 Reaction mechanism of dirigent proteins 36

1.2.5 Biochemical characterization of dirigent proteins 37

1.2.6 Physiological function 39

1.3 Ribokinases 40

1.3.1 D-ribose transport and metabolism 40

1.3.2 The ribokinase family of proteins 42

1.3.3 Structure and reaction mechanism of the *Escherichia coli* ribokinase 43

1.3.4 Biochemical characterization of ribokinases 45

1.3.5 Biotechnological and medicinal applications 47

1.4 Cutinases 48

1.4.1 Structure and function 48

1.4.2 Biocatalytic applications 50

1.4.3 Functional monomers and polymers 51

1.5 Objectives of the thesis 53

1.5.1 Expression of the *Arabidopsis thaliana* dirigent protein *At*DIR6 53

1.5.2 Enzyme engineering of the *Saccharomyces cerevisiae* ribokinase RBK1 53

1.5.3 Biocatalytic application of the *Humicola insolens* cutinase HiC 54

2. Materials & Methods 55

2.1 Materials 55

2.1.1 Chemicals and enzymes 55

2.1.2 Media 55

2.1.3 Plasmids 57

2.1.4 Strains 58

2.1.5 PCR and sequencing primers 58

2.2 Molecular- and microbiological methods 59

2.2.1 Agarose gel electrophoresis 59

2.2.2 Plasmid DNA preparation 60

2.2.3 DNA restriction digests 60

2.2.4 DNA gel extraction and PCR product purification 60

2.2.5 DNA ligation 60

2.2.6 Polymerase chain reaction 61

2.2.7 Preparation of competent *Escherichia coli* cells 63

2.2.8 Heat-shock transformation of competent *Escherichia coli* cells 64

2.2.9 Transformation of competent *Pichia pastoris* cells by electroporation 64

2.2.10 Isolation of genomic DNA of *Pichia pastoris* 65

2.2.11 Sequencing of plasmid constructs 65

2.2.12 Storage of *Escherichia coli* and *Pichia pastoris* strains 66

2.2.13 Cloning 66

2.2.13.1 *Humicola insolens* cutinase 66

2.2.13.2 *Arabidopsis thaliana* dirigent protein 66

2.2.13.3 *Saccharomyces cerevisiae* ribokinase 67

2.3 Protein expression and purification 67

2.3.1 Expression of *At*DIR6 with *Escherichia coli* 67

2.3.2 Expression of HiC and *At*DIR6 with *Pichia pastoris* 68

2.3.3 Fed-batch fermentation of *Pichia pastoris* expressing *At*DIR6 68

2.3.4 Expression of RBK1 with *Escherichia coli* 70

2.3.5 Protein purification of *At*DIR6 70

2.3.6 Protein purification of RBK1 71

2.3.7 Protein quantification 71

2.3.8 SDS-PAGE analysis 71

2.3.9 Western blot analysis 72

2.4 Biochemical characterization of *At*DIR6 72

2.4.1 Deglycosylation of *At*DIR6 72

2.4.2 Circular dichroism 73

2.4.3 Mass spectrometric analysis and glycan structure 74

2.4.4 Dirigent activity assay 74

2.4.5 HPLC analytics 75

2.4.6 Analytical gel filtration 76

2.5 Biochemical characterization and structure elucidation of RBK1 ... 76

2.5.1 Kinase activity assay ... 76

2.5.2 Biotransformations with RBK1 ... 79

2.5.3 HPLC and LC/MS analytics ... 80

2.5.4 Size exclusion chromatography ... 81

2.5.5 Crystallization of RBK1 ... 82

2.6 Biochemical characterization and biotechnological use of *Humicola insolens* cutinase ... 83

2.6.1 Cutinase activity assay ... 83

2.6.2 Immobilization of HiC ... 84

2.6.3 Transesterification reactions with HiC ... 84

2.6.4 GC/MS analysis ... 85

2.6.5 Product purification and identification by NMR and FT-IR ... 85

3. Results ... 86

3.1 Optimized expression of the dirigent protein *At*DIR6 in *Pichia pastoris* and impact of glycosylation on protein structure and function ... 86

3.1.1 Heterologous expression of *At*DIR6 with *Escherichia coli* ... 86

3.1.2 High yield fed-batch fermentation of *At*DIR6 with *Pichia pastoris* ... 90

3.1.3 Characterization of the glycan structure and enzymatic deglycosylation ... 92

3.1.4 Dirigent activity of *At*DIR6 ... 95

3.1.5 Dirigent activity of deglycosylated *At*DIR6 ... 98

3.1.6 Changes in protein structure induced by deglycosylation of *At*DIR6 ... 99

3.2 Enzyme engineering of *Saccharomyces cerevisiae* ribokinase ... 101

3.2.1 A novel D-xylose metabolizing pathway ... 101

3.2.2 Homology modelling and smart library design ... 102

3.2.3 Expression and purification of RBK1 ... 105

3.2.4 Biochemical characterization of RBK1 111

3.2.4.1 Activity and stability 111

3.2.4.2 Optimization of ion concentrations and pH 112

3.2.4.3 Biotransformations and analytics via HPLC 113

3.2.5 Evaluation of the focused library 118

3.2.5.1 Initial screening 118

3.2.5.2 Phosphate dependency 120

3.2.5.3 Substrate spectrum of the focused library variants 121

3.2.6 Kinetic data of RBK1 125

3.2.7 Crystallization and x-ray diffraction of RBK1 127

3.3 Thiol-functionalization of acrylic ester monomers catalyzed by immobilized *Humicola insolens* cutinase 129

3.3.1 Preparation of immobilized *Humicola insolens* cutinase 130

3.3.2 Kinetic data of *Humicola insolens* cutinase 130

3.3.3 Transesterification reactions and substrate specificity 131

3.3.4 Influence of radical inhibitor BHT on transesterifications 132

3.3.5 Product purification and identification by NMR and FT-IR 132

4. Discussion 134

4.1 Optimized expression of the dirigent protein *At*DIR6 in *Pichia pastoris* and impact of glycosylation on protein structure and function 134

4.1.1 Recombinant protein production of *At*DIR6 134

4.1.2 Characterization of *At*DIR6 137

4.1.3 Outlook 142

4.2 Enzyme engineering of *Saccharomyces cerevisiae* ribokinase 143

4.2.1 Recombinant protein production of RBK1 143

4.2.2 Characterization of RBK1 144

4.2.3 Enzyme engineering and evaluation of the focused library 146

4.2.4 Crystallization of RBK1 150

4.2.5 Outlook 151

4.3 Thiol-functionalization of acrylic ester monomers catalyzed by immobilized *Humicola insolens* cutinase 152

4.3.1 Characterization of *Humicola insolens* cutinase 152

4.3.2 Transesterification of acrylic monomers 153

4.3.3 Outlook 156

4.4 Conclusion 156

5. References 158

6. List of figures 175

7. List of supplementary figures 178

8. List of tables 179

9. Appendix 180

9.1 Supplementary figures 180

9.2 Nucleotide sequences 187

9.2.1 *At*DIR6 codon optimized for *Escherichia coli* 187

9.2.2 *At*DIR6 codon optimized for *Pichia pastoris* 188

9.2.3 Ribokinase RBK1 from *Saccharomyces cerevisiae* 189

9.2.4 Cutinase HiC from *Humicola insolens* 190

10. Curriculum vitae 191

V. Abstract

Biotechnology is emerging as a key technology in the 21st century. In the past years, scientific and technological advances have established the field of biocatalysis as a competitive alternative to traditional metallo- and organocatalysis in chemical synthesis [1]. Yet many, often basic challenges remain in the key steps of biocatalytic processes, which include the biocatalyst production, characterization, engineering, and application as well as down-stream processing [2]. In the scope of this thesis, three case studies were performed to illustrate solutions for selected, critical challenges in these key steps.

The first case study featured dirigent proteins, which are involved in the stereoselective control of the first step of lignan biosynthesis, the phenoxyl radical coupling of coniferyl alcohol radicals to yield optically pure (+)- or (-)-pinoresinol. Up to now, extraction of dirigent proteins from plant material or expression in plant and insect cell culture systems resulted in low yields. For biotechnological applications and the solution of a structure, which might help in elucidation of the exact reaction mechanism, a heterologous expression platform allowing the cost-effective, fast, and robust expression in high yields is necessary. A comprehensive study with the heterologous expression host *Escherichia coli* revealed severe complications in the high yield expression of dirigent proteins. The metylotrophic yeast *Pichia pastoris* was subsequently evaluated as a possible host for recombinant protein expression. A reliable high yield fed-batch fermentation process with *Pichia pastoris* resulting in 47 mg mL^{-1} of the dirigent protein *At*DIR6 representing a more than 250-fold increase compared to other sources was established. Biochemical characterization of *At*DIR6 produced with *Pichia pastoris* showed an overall agreement in protein structure, *N*-glycosylation sites, and dirigent activity compared to *At*DIR6 produced by plant cell cultures of *Solanum peruvianum*, validating *Pichia pastoris* as a suitable expression system for dirigent proteins. Enzymatic deglycosylation of the protein induced conformational changes leading to the loss of function and subsequent protein aggregation. This demonstrated that the glycan structures of *At*DIR6 are essential for structure, solubility, and function of the protein.

The second case study focused on a ribokinase from *Saccharomyces cerevisiae*, which was engineered for an altered substrate spectrum by rational protein design. The enzyme was produced in high yields in a heterologous expression system using *Escherichia coli* as host and biochemically characterized for the first time. Optimum reaction conditions were investigated regarding the dependency on mono-, di-, and pentavalent ions and the pH. Conditions featuring 100 mM potassium phosphate at pH 6.0 with 2.5 mM magnesium chloride were found to offer the highest catalytic activity. Based on a family alignment and rational design using a homology model of the protein, a small focused library of the ribokinase featuring 72 variants was established, including combinations of single, double, triple, and quadruple mutations at four residues in the substrate binding area. The library was evaluated for alterations in the substrate spectrum with a focus on D-xylose, which is an epimer to the natural substrate D-ribose. More radical alterations such as triple and quadruple mutations abolished catalytic activity towards D-ribose and D-xylose almost completely. Several variants were identified broadening the substrate spectrum to other tetroses, pentoses, and hexoses, albeit with low catalytic efficiency. A single variant exhibited about twice the activity of the wild-type enzyme towards D-xylose (0.77 s^{-1} versus 0.38 s^{-1}), which can serve as a basis for further studies. In this regard, an attempt to solve a crystal structure for more information about D-xylose binding in the active site was started. While several well diffracting crystals were obtained, molecular replacement did not lead to a solution of the structure so far.

The third case study featured a cutinase from the ascomycete *Humicola insolens*, which was applied as a novel biocatalyst for the synthesis of functionalized acrylic esters by transesterification. The transesterification of methyl acrylate with 6-mercapto-1-hexanol at a high molar ratio in a solvent free system was studied. The enzyme was employed in immobilized form on a microporous, polystyrenic resin to facilitate the usability and stability in a non-aqueous environment and allow simple recovery of the biocatalyst. Besides two minor by-products, 6-mercapothexyl acrylate ester was identified as the main product with the thiol as the functional end group. By optimization of the reaction conditions and critical water content, the transesterification yielded 95.4 ± 0.3% of 6-mercaptohexyl acrylate ester after 6 h at 40°C in the presence of 0.025% (w/w) water without formation of by-products in a solvent free system. When applying methyl methacrylate as an acyl acceptor, transesterification with 6-mercapto-1-hexanol was significantly lower (43.6 ± 0.1%).

VI. Zusammenfassung

Die Biotechnologie erweist sich als Schlüsseltechnologie des 21. Jahrhunderts. In den letzten Jahren konnte durch wissenschaftliche und technologische Fortschritte die industrielle Biokatalyse als kompetitive Alternative zu traditioneller Metall- bzw. Organokatalyse in der chemischen Synthese etabliert werden [1]. Allerdings verbleiben grundlegende Herausforderungen in der Etablierung von biokatalytischen Prozessen. Die grundlegenden Schritte eines biokatalytischen Prozesses umfassen sowohl die Produktion, die Charakterisierung, die Optimierung, die Anwendung sowie die Wiederverwertung von Biokatalysatoren als auch die Aufarbeitung der Produkte [2]. Im Rahmen der vorliegenden Arbeit wurden ausgewählte, kritische Herausforderungen in diesen Schlüsselschritten in drei Fallstudien näher untersucht.

Dirigierende Proteine sind in der stereoselektiven Kontrolle des ersten Schrittes in der Lignan Biosynthese beteiligt. In einem Radikalkupplungsschritt werden zwei Koniferylalkoholradikale unter Beteiligung von dirigierenden Proteinen so verknüpft, dass optisch reines (+)- bzw. (-)-Pinoresinol entsteht. Bisher wurden Dirigierende Proteine in geringen Ausbeuten durch Extraktion aus Pflanzenmaterial oder durch heterologe Expression in pflanzlichen Zellkulturen sowie Insektenzellkulturen gewonnen. Für eine biotechnologische Anwendung und um eine Strukturaufklärung voranzutreiben, welche in der Bestimmung des genauen Reaktionsmechanismus dirigierender Proteine nützlich wäre, sind allerdings höhere Proteinausbeuten notwendig. Dies könnte durch die Etablierung einer heterologen Expressionsplattform ermöglicht werden, welche eine kosteneffiziente, schnelle und robuste Expression von dirigierenden Proteinen erlaubt. Eine umfassende Studie zur heterologen Expression in *Escherichia coli* zeigte allerdings schwerwiegende Probleme für eine Expression von dirigierenden Proteinen mit hohen Ausbeuten. Die Hefe *Pichia pastoris* wurde deshalb anschließend als möglicher Expressionswirt für die rekombinante Proteinproduktion evaluiert. Mit *Pichia pastoris* konnte ein zuverlässiger Fed-Batch Fermentationsprozess etabliert werden, welcher in Ausbeuten von 47 mg mL^{-1} des dirigierenden Proteins *At*DIR6 resultierte. Dies entspricht einer ca. 250-fachen Steigerung der Ausbeute verglichen mit den bisherigen Systemen. Die biochemische Charakterisierung von *At*DIR6, welches mit *Pichia pastoris* hergestellt wurde, zeigte eine vergleichbare Proteinstruktur und dirigierende Aktivität sowie eine Glykosylierung an den gleichen

N-Glykosylierungsstellen im Vergleich zu *At*DIR6 aus einem heterologen Pflanzenzellsystem (*Solanum peruvianum*). Diese Ergebnisse bestätigten die Verwendung von *Pichia pastoris* als geeigneter Expressionswirt für die Herstellung dirigierender Proteine. Eine enzymatische Deglykosylierung des Proteins induzierte strukturelle Veränderungen, welche zu einem Verlust der Funktion und anschließender Proteinaggregation führten. Dadurch konnte die essentielle Bedeutung der Glykanstrukturen von *At*DIR6 für die Proteinstruktur sowie die Löslichkeit und Funktion des Proteins demonstriert werden.

In der zweiten Fallstudie wurde eine das Substratspektrum einer Ribokinase aus *Saccharomyces cerevisiae* durch rationales Proteindesign verändert. Das Enzym konnte zum ersten Mal in hoher Ausbeute mit *Escherichia coli* als heterologer Expressionswirt hergestellt und biochemisch charakterisiert werden. Die optimalen Reaktionsbedingungen bezüglich der Abhangigkeit des Enzyms von mono-, di- und pentavalenten Ionen als auch dem pH Wert wurden ermittelt. Dabei zeigte das Wildtyp Enzym die höchste katalytische Aktivität in 100 mM Kaliumphosphatpuffer bei pH 6.0 mit 2,5 mM Magnesiumchlorid. Basierend auf einem Sequenzvergleich von 200 Ribokinase Sequenzen aus der Familie der Ribokinasen und rationalem Proteindesign mit einem Homologiemodell des Proteins konnte eine kleine, fokussierte Mutantenbibliothek erstellt werden. Insgesamt wurden 72 verschiedene Varianten mit Kombinationen aus einfach, zweifach, dreifach und vierfach Mutationen an vier unterschiedlichen Aminosäureresten in der Substratbindetasche erstellt. Die fokussierte Bibliothek wurde in Bezug auf Veränderungen im Substratspektrum evaluiert. Dabei lag der Fokus auf D-Xylose, welches ein Epimer zu dem natürlichen Substrat D-Ribose ist. Starke Veränderungen der Substratbindetasche durch dreifach und vierfach Mutationen führten zu einem nahezu vollständigen Verlust der katalytischen Aktivität gegenüber D-Ribose und D-Xylose. Mehrere Varianten zeigten ein erweitertes Substratspektrum mit geringen Aktivitäten gegenüber Tetrosen, Pentosen und Hexosen. Eine einzelne Variante wies eine etwa verdoppelte Aktivität im Vergleich zu dem Wildtyp Enzym gegenüber D-Xylose auf (0.77 s^{-1} gegenüber 0.38 s^{-1}), was als Basis für weitere Studien verwendet werden kann. Für weitere Studien wurde auch versucht, die Struktur des Proteins zu lösen. Dies könnte mehr Informationen über die Bindung von D-Xylose in der aktiven Tasche geben. Obwohl Diffraktionsmuster mehrerer Kristalle erhalten wurden, konnte die Struktur durch die „Molecular replacement“ Methode bisher nicht gelöst werden.

In der dritten Fallstudie wurde eine Cutinase aus dem Schlauchpilz *Humicola insolens* als neuartiger Biokatalysator für die Synthese von funktionalisierten Acrylestern eingesetzt. Dabei wurde als Modellreaktion die Transesterifizierung von Methylacrylat mit 6-Mercapto-1-hexanol in einem lösungsmittelfreien System untersucht. Das Enzym wurde auf mikroporösem Trägermaterial aus Polystyren immobilisiert um die Verwendbarkeit und Stabilität in einer nicht-wässrigen Umgebung sowie das einfache Wiederverwenden des Biokatalysators zu ermöglichen. Als Hauptprodukt wurde ausschließlich 6-Mercaptohexylacrylatester mit der freien Thiogruppe als funktionelle Endgruppe identifiziert. Die Bildung zweier Nebenprodukte, welche durch Michael-Additionsreaktionen entstanden waren, konnte durch die Optimierung der Reaktionsbedingungen verhindert werden. Die optimierte Transesterifizierung erreichte den Umsatz von 95,4 ± 0,3 % zu dem Hauptprodukt 6-Mercaptoacrylatester in 6 h bei 40°C in der Anwesenheit von 0,025 % (w/w) Wasser in einem lösungsmittelfreien System. Die Verwendung von Methylmethacrylat als Acyl-Akzeptor in einer Transesterifizierung mit 6-Mercapto-1-hexanol lieferte einen deutlich geringeren Umsatz zum entsprechenden Hauptprodukt (43,6 ± 0,1%).

VII. Abbreviations

(v/v)	volume per volume
(w/v)	weight per volume
(w/w)	weight per weight
°C	degrees Celsius
µL	microlitre
µm	micrometre
µM	micromolar
ADP	adenosine diphosphate
AMP-PNP	adenylyl imidodiphosphate
APS	ammonium persulfate
*At*DIR6	dirigent protein from *Arabidopsis thaliana*
ATP	adenosine triphosphate
bp	base pair
cdw	cell dry weight
CV	column volume
cww	cell wet weight
DAD	diode array detector
DIR	dirigent proteins
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
FPLC	fast protein liquid chromatography

g	gram
GC	gas chromatography
GC/MS	gas chromatography coupled to mass spectrometry
h	hour
HiC	cutinase from *Humicola insolens*
HPLC	high performance liquid chromatography
IPTG	isopropyl-β-D-thiogalactopyranoside
kb	kilo-base pair
k_{cat}	catalytic turnover number
k_{cat}/K_m	catalytic efficiency
KCl	potassium chloride
kDa	kilodalton
K_{iu}	inhibition constant for substrate (uncompetitve) inhibition
K_m	Michaelis-Menten constant
L	litre
LB	lysogenic broth
LC/MS	liquid chromatography coupled to mass spectrometry
M	molar
mg	milligram
$MgCl_2$	magnesium chloride
min	minute
mL	millilitre
mM	millimolar
mm	millimetre
NaCl	sodium chloride

NADH	nicotinamide adenine dinucleotide
nm	nanometre
NMR	nuclear magnetic resonance
OD_{600}	optical density measured at 600 nm
P. pastoris	*Pichia pastoris*
PCR	polymerase chain reaction
PEP	phosphoenolpyruvate
P_i	Inorganic phosphate
pmol	picomol
PMSF	phenylmethanesulfonylfluoride
PNGase F	peptide-*N*-glycosidase from *Flavobacterium meningosepticum*
*p*NP	*para*-nitrophenol
RBK1	ribokinase from *Saccharomyces cerevisiae*
RID	refractive index detector
rpm	rounds per minute
S. cerevisiae	*Saccharomyces cerevisiae*
S. peruvianum	*Solanum peruvianum*
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
TB	terrific broth
TEMED	N,N,N´,N´-tetramethylethylenediamine
Tris	tris(hydroxymethyl)aminomethane
U	units

1. Introduction

1.1 Challenges in biocatalysis

1.1.1 Industrial biocatalysis

In the field of biocatalysis enzymes or whole cells are applied as biocatalysts in synthetic chemistry for the production of bulk or fine chemicals [1]. Enzymes, comparable to chemical catalysts, lower the activation energy of (bio)chemical reactions with a high catalytic efficiency yielding catalytic improvements up to 10^{19}-fold compared to the corresponding uncatalyzed reactions [3]. Rapid developments and technological advances in the last two decades established biocatalysis as an economically successful and environmentally friendly alternative to the traditional metallo- and organocatalysis in chemical synthesis. While biocatalysts show many favorable advantages such as providing high selectivity with regard to regio- and stereochemistry, high catalytic turnover, and being environmentally benign offering safe processes at mild conditions, ultimately the economy dictates whether a biocatalytic process will be realized or if a traditional chemical approach is chosen [4].

Although fermentation processes such as the brewing of beer and baking of bread as early examples for “biotechnology” are known for centuries, the intentional use of biocatalysts started in the beginnings of the 20th century. The synthesis of *R*-(+)-mandelonitrile from benzaldehyde and hydrogen cyanide using a plant extract preparation, which included hydroxynitrile lyase, is one of the first descriptions of an asymmetric biocatalysis [5]. More examples for successful biocatalytic processes include the production of high-fructose corn syrup, i.e., isomerization of D-glucose to D-fructose by the action of xylose isomerase, which was established in the mid-1970s runs on a multi-million ton scale underlining the economic significance [6,7]. Furthermore, prominent examples are the preparation of semisynthetic penicillins by penicillin G acylase [8] and the use of proteases in laundry detergents [9]. In all cases, limited stability of the biocatalysts could be overcome by immobilization of the enzyme, which additionally facilitated the recycling of the biocatalyst [1].

The advent of protein engineering technologies in the late 1980s to 1990s allowed the optimization of the biocatalyst itself regarding its activity, process stability, and selectivity. In this light, extending the substrate spectrum enabled the synthesis of unusual, synthetic intermediates especially for pharmaceutical compounds and fine chemicals [1]. Examples include the lipase catalyzed resolution of alcohols and amines yielding enantiopure compounds [10] and the hydration of acrylonitrile to acrylamide for polymers catalyzed by a nitrile-hydratase [11,12].

Biocatalysis today is promoted by key developments like advanced protein engineering strategies, gene synthesis, sequence analysis, bioinformatics tools, and computer modeling [1,13,14]. The development of molecular tools for a fast modification of biocatalysts by iterative cycles of mutagenesis and screening for improved versions, which was termed "directed evolution", set the basics for many new processes featuring these improved biocatalysts [15]. Examples for engineered enzymes in industrial biocatalysis include an engineered amine transaminase (27 out of 330 amino acids were exchanged) in the biosynthesis of optically pure sitagliptin (pharmaceutical compound for treating of type 2 diabetes), which reduced total waste, eliminated all transition metals, and increased overall yield and productivity of the process compared to the chemical hydrogenation process [16]. Furthermore, several processes featuring engineered ketoreductases in (semi)purified form replacing whole cell reductions and metal-ligand based chemocatalysis were established recently [17]. The use of isolated enzymes compared to whole cell biocatalysis offers easier downstream-processing, removes diffusion limitations caused by cell membranes, tolerates harsher conditions, and simplifies logistical issues. Seven enzymatic approaches differing in the choice of engineered enzyme, starting material, and yield and purity of the product, for the synthesis of atorvastatin, the active ingredient in Lipitor (cholesterol-lowering drug with global sales of US$ 11,900,000,000 in 2010 [1]), have been developed, again pointing out the potential of biocatalysis.

Besides engineering of individual enzymes for specific reaction steps, biocatalysis also involves metabolic engineering of whole biochemical pathways. One of the still few success stories is the process set up by Genencor and DuPont for production of 1,3 propanediol in an metabolically engineered *Escherichia coli* (*E. coli*) strain, yet many new concepts especially for the conversion of biomass into second generation biofuels have been proposed [18,19]. Furthermore, heterologous pathways, which enable *Saccharomyces cerevisiae* (*S. cerevisiae*) to metabolize the pentoses D-xylose and D-arabinose from pretreated lignocellulosic feedstocks for second generation biofuels production have been introduced into this organism [20]. The rationale for developing pentose-utilizing *S. cerevisiae* strains is that this yeast has been used for the production of ethanol for millennia and that new processes can easily be integrated into existing plants already using this yeast for first generation biofuels from crops. While *S. cerevisiae*, cannot utilize D-xylose, it can ferment its isomer D-xylulose. Therefore, pathways converting D-xylose to D-xylulose have been introduced. Successful attempts included an introduction of xylose isomerases from the thermotolerant bacterium *Thermus thermophilum* [21] and the obligate anaerobe fungus *Pyromyces* spec [22] for direct conversion of D-xylose to D-xylulose. Another pathway involved the action of a xylose reductase and a xylitol dehydrogenase from the yeast *Pichia stipitis* yielding D-xylulose via xylitol [23,24]. However, the fermentation rates are still hampered by metabolic repression in the presence of D-glucose and the ethanol yield is still too low for industrial processes [25].

1.1.2 The biocatalysis cycle

Despite the advances in biocatalysis, many challenges regarding the production, characterization, engineering, and application of biocatalysts for establishing competitive biocatalytic processes remain. In many cases, traditional chemical processes are still economically more competitive. In establishing a biocatalytic process, several key aspects have to be considered [4]. This regards the availability, purity and costs of educts and medium components, whether a reduction of reaction steps compared to traditional processes can be achieved, space-time-yield or productivity and overall yield, the selectivity of the biocatalyst and potential by-products, and the final product quality. Furthermore, the next critical issue is how the biocatalyst is applied in the process. The biocatalyst can be used in free or immobilized form. While immobilization provides certain advantages like recycling and sometimes higher process stability, it also has to earn back the concomitant costs by allowing a more efficient process. Application of whole cells or isolated enzyme (crude preparation or purified) is often based on co-factor requirements. More decisions concern the type of reaction medium, whether the process will be performed in conventional aqueous environment or nonconventional media such as organic solvents. While water can be seen as a “green” solvent with low environmental impact, organic synthesis suffers from severe restrictions in aqueous media. Many apolar organic compounds are poorly soluble in water, side reactions due to hydrolysis, polymerization or racemization can take place in water, organic compounds can be degraded, the downstream processing is often inefficient and cost extensive since water cannot easily be evaporated, and extraction of products might not be ecological due to low product concentrations [26]. Additionally, chemical equilibria can be shifted in organic media allowing new reactions such as transesterifications instead of hydrolysis. Since it was shown that enzymes can work in neat organic solvents without being denatured, this was exploited for many asymmetric synthetic transformations [27]. Prominent examples are lipases like *Candida Antarctica* lipase B (CalB) used e.g. for the transacylation of industrially relevant compounds like acrylates [28]. Finally, on the basis of biocatalyst kinetics and decisions made for the application of the biocatalyst, the reactor configuration and mode of operation of the bioreactor have to be established [4].

A new biocatalytic process can be based on the availability of a biocatalyst with a novel function or on a desired product for which a new biocatalyst needs to be selected [2]. Therefore, the first key issue is producing the biocatalyst or providing an organism capable of performing the desired reaction (see Figure 1-1). For a robust process, the biocatalyst needs to be characterized regarding its activity, stability, kinetics, and other parameters, which might influence the process such as substrate and product inhibition and the optimum reaction conditions. Whenever possible, gaining structural information from X-ray crystallography or protein nuclear magnetic resonance (NMR) spectroscopy can be a huge asset especially for the next step, the engineering of the biocatalyst. Biocatalysis requires stable, selective and productive biocatalysts operating under the desired process conditions to be able to minimize costs and compete with chemical processes [1]. Since most biocatalysts often lack in one of these traits, engineering strategies to improve the biocatalyst were developed. Today, rational design based on structural information and directed evolution by mutagenesis and selection/screening of improved variants or a combination thereof allow the engineering of many desired traits, which can contribute to an improved biocatalytic process [17]. When a suitable biocatalyst is found or engineered, it needs to be prepared for application. This often includes steps such as lyophilization, immobilization on various resins, or cross-linking. These techniques can help increase the stability of the biocatalyst under process conditions, but also facilitate a recycling for repeated (fed)-batch processes or application in a continuous fashion. A critical issue in this regard is the choice of the solvent system. Since many enzymes show good activities in neat organic solvents, this allows many further advantageous applications compared to the traditional aqueous solvent [27]. In the last step of the process, the product is recovered, often in a second batch process after the separation of biomass or immobilisate in processes with whole cells or immobilized enzymes, respectively [2]. However, *in situ* product-recovery processes like crystallization, distillation, solid-phase-, and liquid-phase recovery, which are based on physical and chemical parameters of the reactants are used more often [29].

In the upcoming chapters, selected challenges in the distinct steps of establishing biocatalytic processes are introduced.

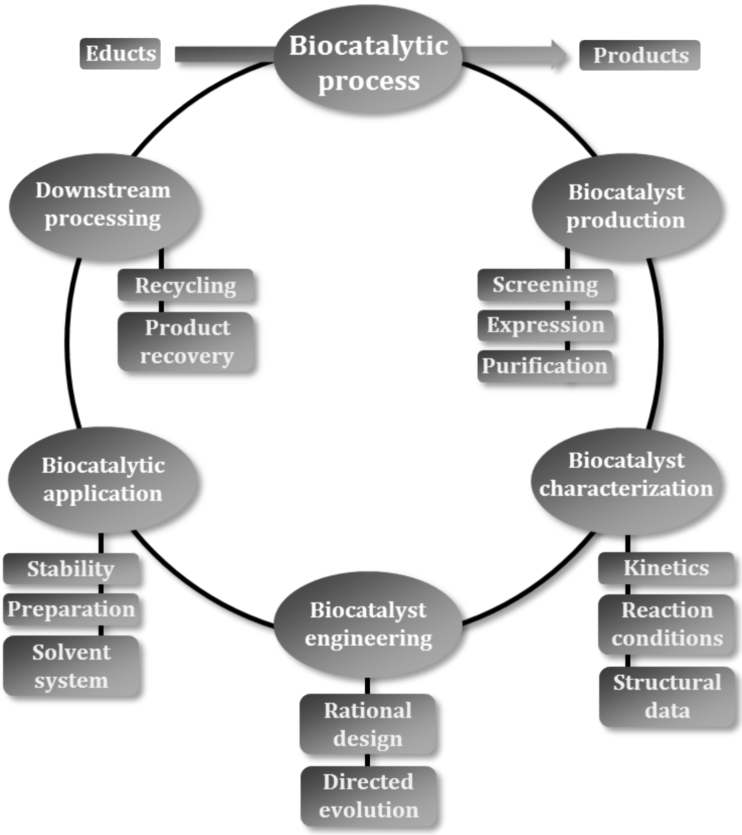
Educts
Biocatalytic process
Products
Biocatalyst production
Screening
Expression
Purification
Biocatalyst characterization
Kinetics
Reaction conditions
Structural data
Biocatalyst engineering
Rational design
Directed evolution
Biocatalytic application
Stability
Preparation
Solvent system
Downstream processing
Recycling
Product recovery

Figure 1-1 – The biocatalysis cycle (modified from [2]**).** The figure illustrates key steps and selected challenges in the creation of a biocatalytic process including the production, characterization, engineering and application of the biocatalyst itself as well as the downstream processing.

1.1.3 Biocatalyst production

Both, isolated enzymes and whole cell biocatalysts are used in industrial processes and are intensively researched. While whole cell biocatalysts are often used for synthetic reactions requiring cofactors, which need to be regenerated, isolated enzymes are typically used for hydrolytic or isomerization reactions [2]. Recently, focus has shifted towards the application of isolated enzymes as crude extracts or in a purified form yielding more economic and practical processes. Advantages include an easier removal of the biocatalyst (since less is used due to higher specific activities), preclusion of diffusion limitations by the cell membrane, toleration of harsh conditions, which might be toxic for whole cells, and easier logistics for global companies [1]. In this respect, e.g. ketoreductase based processes replaced whole cell reductions and metal-ligand based chemocatalysis, which used to be the industrial standard for years [17]. The development of recombinant DNA technology in the early 1970s enabled the production of recombinant proteins in pro- and eukaryotic hosts. This led to the production of recombinant biopharmaceuticals and facilitated a strong increase in the industrial enzyme market. It was then possible to produce enzymes in much larger quantities, lower costs, and to develop easier downstream-processing methods compared to the laborious extraction from plants and animals [30]. The main difference between industrial enzymes and biopharmaceuticals are the production costs. While industrial enzymes need to be produced at 7-14 € per kilogram with low margins on sales, biopharmaceuticals can be sold at prices 100 to 1,000 times higher. For the production of many biopharmaceuticals, correct post-translational modifications, especially glycosylation, is required and can only be achieved when using mammalian cell systems [30]. Here, Chinese hamster ovary cells are by far the most commonly used cell line for the production of glycoproteins. Industrial enzymes are mainly produced in filamentous fungi. These fungi species naturally secrete large amounts of industrially relevant enzymes, can grow on cheap substrates and are GRAS organisms (Generally Recognized As Safe). The most important species, which secrete enzymes like amylases, cellulases, lipases, and proteases, are *Aspergillus*, *Trichoderma*, and *Penicillum*.

For recombinant protein production (RPP), *E. coli* remains the most commonly used expression host [31]. In 2006, about 40% of all biopharmaceuticals were produced in *E. coli* [32]. *E. coli* offers several advantages, including well characterized genetics, broad spectrum of available molecular tools, growth in cheap media, rapid biomass accumulation, amenability to high cell-density fermentations, and a simple process scale-up [30,31]. Recently, other bacterial hosts than *E. coli* are used more often for RPP. The gram-positive *Bacillus* strains share many advantages with *E. coli*, but excel with their ability to secrete the recombinant protein into the medium [32,33]. Furthermore, unlike *E. coli*, they do not contain lipopolysaccharides (endotoxins), which are pyrogenic to humans and need to be specifically removed in additional purification steps for biopharmaceuticals. About 60% of the commercially used bulk enzymes such as alkaline proteases as washing agents are produced in *Bacillus* species. However, the lack of suitable expression vectors, plasmid instability, complicated transformation procedures, difficulty of high cell density cultivations and the presence of secreted proteases complicate the use in RPP.

Expression of recombinant proteins in bacteria such as *E. coli* often results in insoluble, nonfunctional protein aggregates (inclusion bodies) due to improper folding. This is particularly true for eukaryotic proteins requiring post-transcriptional modifications, such as glycosylation and disulfide bond formation, which can be important for stability and/or bioactivity [30]. In contrast to bacteria, yeasts possess an effective glycosylation machinery and they share the advantages of *E. coli* as an expression host, e.g. simple cultivation and well-characterized genetics and metabolism. In this regard, the yeast *Pichia pastoris* (*P. pastoris*) is of particular interest as this organism offers protein expression levels, which are often 10- to 100-fold higher compared to *S. cerevisiae* [34], and has already been used for the production of more than 1,000 proteins [35]. The typical glycosylation pattern of secretory *P. pastoris* proteins is $Man_8GlcNAc_2$ or $Man_9GlcNAc_2$ lacking α1,3-terminal mannose units of the oligosaccharide chains [36]. This differs from the paucimannosidic and complex types of glycosylation of plant and mammalian cells. Since many studies showed that a more stable and active protein can be achieved if the glycosylation pattern of recombinantly expressed proteins mimics the native one very well, various attempts to engineer the glycosylation pattern of *Pichia* for more human-like glycans have been reported [30]. In this regard, an attempt to introduce a functional glycosylation pathway from *Campylobacter jejuni* into *E. coli* has been made as well [37].

Besides bacterial, yeast, and mammalian expression hosts, insect and plant cell cultures are also used for RPP [38,39]. While the genetic modifications are generally more complicated compared to bacterial expression systems like *E. coli*, both systems offer the ability for post-translational modifications, low risk of contamination with pathogens, low to medium production costs, and the possibility for scale-up.

In addition to an optimum quality of the produced proteins, maximizing the volumetric or specific productivity is the second main goal in RPP. Many strategies can be used and have been implemented to improve the expression hosts such as *E. coli* and the downstream-processing to improve the yield of the target protein. While chromosomal integration of genes is necessary for higher eukaryotic hosts, plasmid-based RPP with self-replicating, extra-chromosomal vectors is common for bacterial hosts [30]. This allows a high gene dosage, easy genetic manipulation, and inducible expression of the target protein. A large series of *E. coli* strains is commercially available with key features like being deficient in distinct proteases (*E. coli* BL21 and derivatives), facilitating cytoplasmic disulfide bond formation (*E. coli* Origami), or adjusting the need for rare tRNAs to help with the codon bias of heterologous gene sequences (*E. coli* Rosetta) [32]. Various promoter systems allow the inducible expression of the target protein, e.g. the T7 RNA polymerase system and the P_{BAD} promoter of the arabinose operon.

Although the choice of the expression host and plasmid system can help overcome many problems with recombinant expression, overexpressed proteins still often accumulate as inclusion bodies [40]. There are two main strategies to solve this issue, either adapting the expression parameters to yield highly soluble protein or refolding the protein from inclusion bodies. For the first strategy, the simplest way to improve protein solubility is to lower the cultivation temperature, which leads to a reduced protein translation rate, reduced stress on the folding machinery of the cell, and lower proteolytic degradation of the target protein. Another method to solve the problem with rare codons is to use synthetic genes, which are codon optimized for the use in the specific expression host, i.e., a systematic study with 30 codon optimized human genes showed improved expression and solubility in *E. coli* [41]. Translational fusion to highly soluble proteins such as thioredoxin, maltose-binding protein, or bacterioferritin can drastically enhance the solubility of the target protein [42–44]. Furthermore, the co-expression of molecular chaperones aiding in the protein quality control can facilitate proper folding of overexpressed proteins and reduce aggregation [45,46]. Various combinations of the DnaK/DnaJ/GrpE and GroES/GroEL chaperone systems [47] and the *E. coli* trigger factor (TF) [48] are commercially available for co-expression from co-transformed plasmids. Although many proteins were successfully expressed in *E. coli* with this strategy, undesired side effects related to their activities such as promoting proteolysis of the target protein were also reported [49].

The second main strategy comprises the *in vitro* refolding of proteins from inclusion bodies [50,51]. While several more steps in downstream-processing are necessary, it avoids the need to optimize the soluble expression in RPP. Inclusion bodies contain the overexpressed protein in a paracrystalline form as insoluble aggregates. They have a high density allowing easy separation by centrifugation after cell homogenization. Further advantages are the resistance to proteolysis and often high purity (over 90%) after few washing steps. After solubilization with chaotropic buffers, the denatured protein is refolded at low protein concentrations (under 0.1 mg mL^{-1}). The correct folding competes with misfolding and aggregation and refolding yields decrease with higher initial protein concentration, since this favors aggregation of unfolded or partially folded proteins, which still contain accessible hydrophobic patches. Due to the low protein concentrations, large volumes of refolding buffer, huge reactors, and additional concentration steps are necessary. In industry, batch dilution is the standard method, which needs to be carefully optimized in order to reduce costs and waste disposal.

Although several studies suggest schemes or workflows, which should be followed in order to minimize costs and time and maximize the return of investment in RPP, the main problem with all the above mentioned strategies is that it almost always has to be empirically determined, how to operate for each individual protein [52–54]. Also, as it was shown that protein quality does not necessarily correlate with solubility, the latter should eventually be reconsidered as the main optimization goal and universal indicator for a successful process [55].

1.1.4 Biocatalyst characterization and engineering

A useful biocatalyst for industrial applications is characterized by high catalytic turnover, high selectivity (chemo-, regio-, and stereoselectivity), and high process stability under the required conditions [14,56]. Basic knowledge about the biocatalysts biochemistry like activity and kinetics, especially in regard to substrate and product inhibition, is indispensable. Further structural information about the native oligomerization state, overall secondary structure or post-translational modifications gained by techniques such as analytical gel filtration, circular dichroism spectroscopy, and MALDI-TOF or LC/MS-MS analysis can help in understanding the basic reaction mechanism and issues related to stability or solubility of the biocatalyst [57].

As most biocatalysts lack in some vital trait like activity, stability, or selectivity for the application in biocatalytic processes, many new strategies for improvement have been developed and are continuously advanced [13]. These protein engineering strategies differ in the degree of change introduced into the protein and the amount of information required about the protein. In the mid-1990s, Stemmer and Arnold set the foundation for the optimization of enzyme features by directed evolution [58,59]. Generation of a gene library by mutagenesis is followed by the expression of the genes and the screening of improved enzymatic traits like solvent tolerance or enantioselectivity [59,60]. Early methods like DNA-shuffling and error-prone PCR (epPCR) and advanced methods like sequence saturation mutagenesis (SeSaM) [61] are used to establish large gene libraries, which need efficient screening techniques like fluorescence-activated cell sorting, phage display, or growth assays [14]. In order to reduce the screening effort, techniques to reduce the library size with a concomitant increase in library quality have been made. In directed evolution, one way to reduce the library size is by eliminating duplicate codons by using a degenerate codon yielding only 12 instead of 20 possible amino acids [62]. Rational design offers another solution to generate smaller and smarter libraries [13]. Guided by crystal structures or homology models, certain amino acids or hotspots are identified, which are usually in the vicinity of the active site of the protein. Rational substitution of amino acids then leads to small focused libraries, which can easily be screened for the desired alteration. Preferentially, the structure of the protein should be solved with the substrate of interest, product, or an inhibitor, which resembles the substrate transition state [17]. This can help in identifying the residues (in)directly involved in catalysis. Combined approaches of rational design and directed evolution such as iterative saturation mutagenesis (ISM) or iterative CASTing (combinatorial active site testing) developed by Reetz *et al.* combine the advantages of both approaches [63–65].

1.1.5 Biocatalyst application and downstream-processing

Biocatalytic applications are frequently limited by the need to operate under conditions unsuited for biocatalysts [29]. While chemical catalysts are prepared for specific conversions at defined conditions, enzyme kinetics and protein stability under technical conditions are key features, which strongly differentiate biocatalysts from chemical catalysts [2]. Ideally, the biocatalyst operates under a broad range of physical parameters like different pH and temperatures. Also, additional factors like ion requirements or cofactors need to be taken in account.

In industrial biocatalysis, enzymes and also whole cells are mostly used in immobilized form. Various immobilization techniques, e.g. immobilization on hydrophobic, porous resins, cross-linked enzyme aggregates, or encapsulation [66], have been shown to be able to improve not only the stability, but also activity, specificity, and reduction of inhibition of biocatalysts [67]. Immobilized or lyophilized enzymes can be used in neat organic solvents [27]. While most proteins are denatured in organic/aqueous mixtures, they were shown to be very rigid and stable in neat (hydrophobic) organic solvents. This circumvents many restrictions of aqueous media as already discussed above (see 1.1.1). Prime examples here are many lipase-catalyzed asymmetric (trans)esterifications to produce optically pure intermediates and products [10].

Many enzymes show far lower catalytic activity in organic solvents compared to water [68]. In this regard, a low water activity can “activate” the rigid enzyme in the organic solvent by providing hydrogen bonds [69]. Interestingly, usage of various organic solvents can alter enzyme specificities like substrate-, enantio-, regio-, or chemoselectivity [70]. This “solvent engineering” can be synergistically used with protein engineering. In aqueous media, enzymes are strongly dependent on a distinct pH value or range. While the concept of pH has no meaning in organic solvents, it has been shown that the catalytic activity of an enzyme reflects the pH of its last aqueous solution [71]. Consequently, this pH memory effect should be exploited, i.e. enzymes should be lyophilized from or immobilized in aqueous solutions of the pH optimal for catalysis.

One of the most relevant advantages of an immobilization of biocatalysts is still the possibility to easily recover the biocatalyst for further process cycles [2]. This is usually done by filtration or centrifugation and precedes the product recovery. *In situ* product recovery methods, in which the product is removed during the biotransformation, are based on physical and chemical parameters of the reactants [29]. This includes ultrafiltration, crystallization, distillation, adsorption, solid-phase-, and liquid-phase recovery based on properties like molecular weight, volatility, charge, hydrophobicity, or solubility. Advantages of the *in situ* recovery are e.g. avoiding substrate or product inhibition and product decay in aqueous solvents. One example of *in situ* product recovery in an industrial process by simple precipitation is the acrylamide production from acrylonitrile [2].

1.2 Dirigent proteins

1.2.1 Secondary plant metabolites

Secondary plant metabolites are organic compounds, which are, unlike the primary metabolites, not directly involved in anabolic and catabolic reactions or the energy production in cells [72]. Since secondary plant metabolites often share enzymatic pathways with primary metabolites in their biosynthesis, a differentiation by the chemical structure is not possible. The assignment is therefore based on a physiological role in the organism, which is often an important role in plant defense against herbivors and pathogens, but also against abiotic stress factors like coldness, drought, high salinity or ultraviolet radiation resulting in an increased selective advantage [73,74]. Today, more than 10^5 different secondary metabolites in plants are known [75,76]. Despite this large number, all secondary metabolites can be traced back to a few basic chemical structures like terpenoids, alkaloids, glycosides, polyketides and phenolic compounds. Further chemical diversity is generated by oligomerization as well as glycosylation, hydroxylation, methylation or other chemical modifications.

Besides the physiological role in plants, many secondary metabolites exhibit remarkable pharmacological activities or can be used as flavoring agents or dyes [77,78]. In this regard, the chirality of the compounds is very important, since different enantiomers have been shown to exhibit strongly different bioactivities. A prime example for this is the use of the two enantiomers of thalidomide, which was introduced in the late 1950s as an anti-nausea and sedative drug especially for pregnant women leading to the “contergan” scandal [79]. While the (*R*)-enantiomer exhibits the desired sedative effect, the (*S*)-enantiomer was found to lead to a teratogenic effect on fetal development. For an organic synthesis of natural compounds, the regio- and stereoselectivity needed to achieve optically pure compounds can be a great challenge mostly involving complicated protecting and deprotecting steps [2,78].

1.2.2 Lignans

Lignans represent a structurally diverse class of secondary plant metabolites and belong together with flavonoids, stilbenes, and coumarines to the phenolic compounds [78,80]. The molecular backbone is formed by two C-C-coupled phenylpropanoids (C_6-C_3 units) [81].

While various coupling modes are possible, only 8-8´coupled compounds are referred to as lignans (see Figure 1-2a). Compounds coupled via other carbon atoms or via oxygen are referred to as neolignans and oxyneolignans, respectively, according to IUPAC nomenclature (see Figure 1-2a) [82]. Importantly, the coupling of two achiral phenylpropanoids leads to the formation of chiral centers. Depending on the number of chiral carbon atoms, several stereoisomers of a basic lignan structure do exist. Modifications of the basic structure lead to a wide variety of compounds. For example, a common modification is the formation of glycosides [83], which allows the storage of physiologically active compounds in the plant vacuole by increasing the water solubility of the compound.

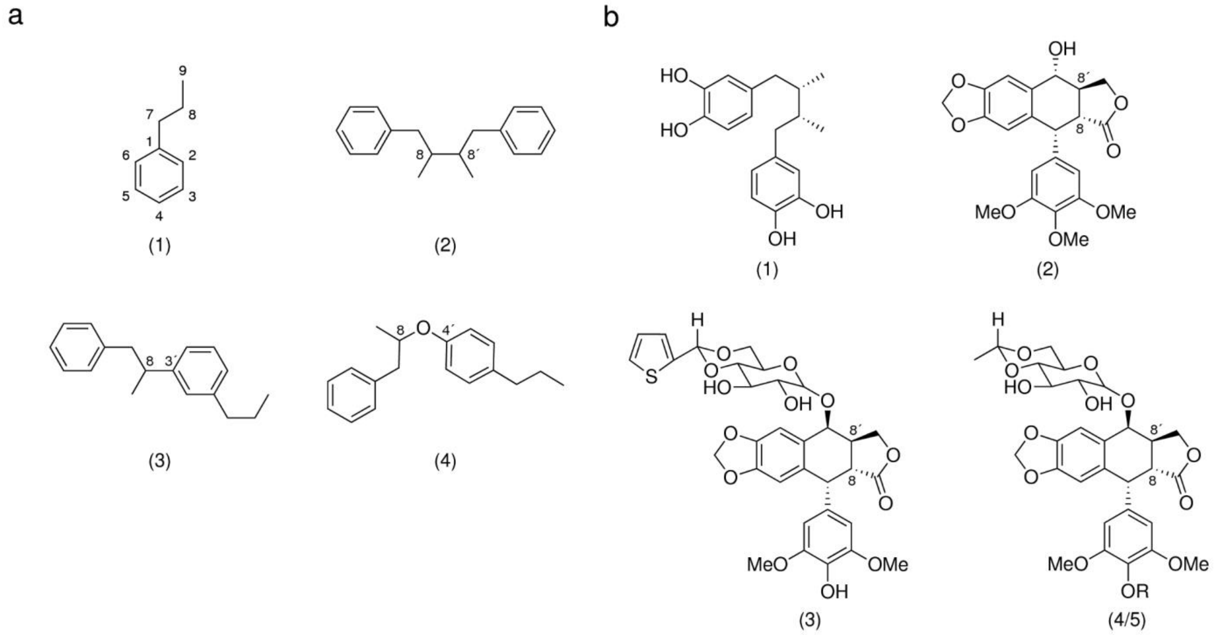

Figure 1-2 – Nomenclature of lignans (a) and selected examples (b). (a) Coupling of two phenylpropanes (1) and its hydroxylated derivatives lead to lignans (2), neolignans (3), or oxyneolignans (4). **(b)** Structures of lignans with pharmacologically interesting activities. NDGA (1), podophyllotoxin (2), teniposide (3), etoposide (4, R=H), and etopophos (5, R=$PO(OH)_2$)

Lignans have been found in liverworts [84], hornworts [85], ferns [86], and higher plants [80]. While the localization of lignans in plants can be diverse including roots, rhizomes, stems, leaves, seeds, and fruits [73], very high concentrations of lignans were found for example in the heart wood of *Thuja plicata* [87].

Lignans exhibit a variety of important physiological activities. While some lignans, e.g. nordihydroguiaretic acid (NDGA) (see Figure 1-2b), act as antioxidants in reducing cell damage by free oxygen species, many lignans exhibit antimicrobial, fungicidal and insecticidal properties for the defense against pathogens *in planta* [73]. Additionally, some lignans also show allelopathic effects. For example, NDGA effectively inhibits the root growth of many grasses [88].

Moreover, lignans have been demonstrated to possess significant pharmacological activities, including anti-inflammatory, anti-tumor, immunosuppression, cardiovascular, antioxidant and antiviral properties [77]. A well-known example is the antiviral podophyllotoxin from *Podophyllum peltatum* (may apple) used in the treatment of warts [89] and its semi-synthetic derivatives Etoposide™, Etopophos™ and Teniposide™, representing important anti-cancer drugs [90] (see Figure 1-2b). While the organic synthesis of derivatives of natural products is prevalent [78], first examples for a biotechnological production were presented [91,92].

The phenylpropanoid biosynthesis pathway generates various building blocks for the biosynthesis of coumarins, flavonoids, isoflavonoids, and stilbenes as well as the monolignols *p*-coumaryl-, coniferyl-, and sinapyl alcohol, which are used in the biosynthesis of lignin and lignans (see Figure 1-3) [93,94]. The starting molecules phenylalanine (1) and tyrosin (3) are derived from the shikimate pathway. From these, *p*-coumaric acid (4) is formed either by the action of phenylalanine ammonia lyase (PAL) and cinnamate 4-hydroxylase (C4H) or tyrosine ammonia lysase (TAL), respectively [94]. In a further enzymatic step, *p*-coumaric acid (4) is activated by the action of a hydroxycinnamoyl CoA ligase (4CL) yielding the central metabolite *p*-coumaryl CoA (5), which provides the basis for all subsequent branches and resulting metabolites. From here, further reduction steps lead to the monolignol *p*-coumaryl alcohol (7). The Cytochrome P450 enzyme C4H allows the entry into the broad majority of the phenylpropanoid pathway derived metabolites. Further hydroxylation steps performed by the P450 enzymes *p*-coumarate-3-hydroxylase (*p*C3H) and ferulate 5-hydroxylase (F5H) coupled with subsequent methoxylation and reduction steps lead to the monolignols coniferyl alcohol (11) and sinapylalcohol (14).

From the three monolignols, mainly coniferyl alcohol is used for the biosynthesis of lignans in higher plants [80]. The first step in the biosynthesis of lignans involves an oxidation of coniferylalcohol to the corresponding radical, which is usually catalyzed by laccases or peroxidases [95,96]. The lignan specific 8,8´-coupling of two radicals with the addition of two water molecules leads to the furofuran lignan pinoresinol. Other coupling modes include an

8,5´- and 8-*O*-4´-coupling yielding the neolignan (±)-dehydroconiferylalcohol and the oxylignans *erythro*/*threo*-(±)-guaiacyl-glycerol coniferyl ethers. Further reduction and oxidation steps catalyzed by pinoresinol-lariciresinol reductases (PLR) and secoisolariciresinol dehydrogenases (SDH) result in lariciresinol, secoisolariciresinol and finally matairesinol formation, which is an educt for the subsequent synthesis of various lignan structures by modification of the structure backbone or the aryl residues [97,98] (see Figure 1-4b). While all these enzymes exhibit a direct, stereoselective control leading to optically pure compounds [99–102], even the random, chemical coupling of two coniferyl alcohol radicals can be affected in a stereoselective fashion by so called dirigent proteins (DIR).

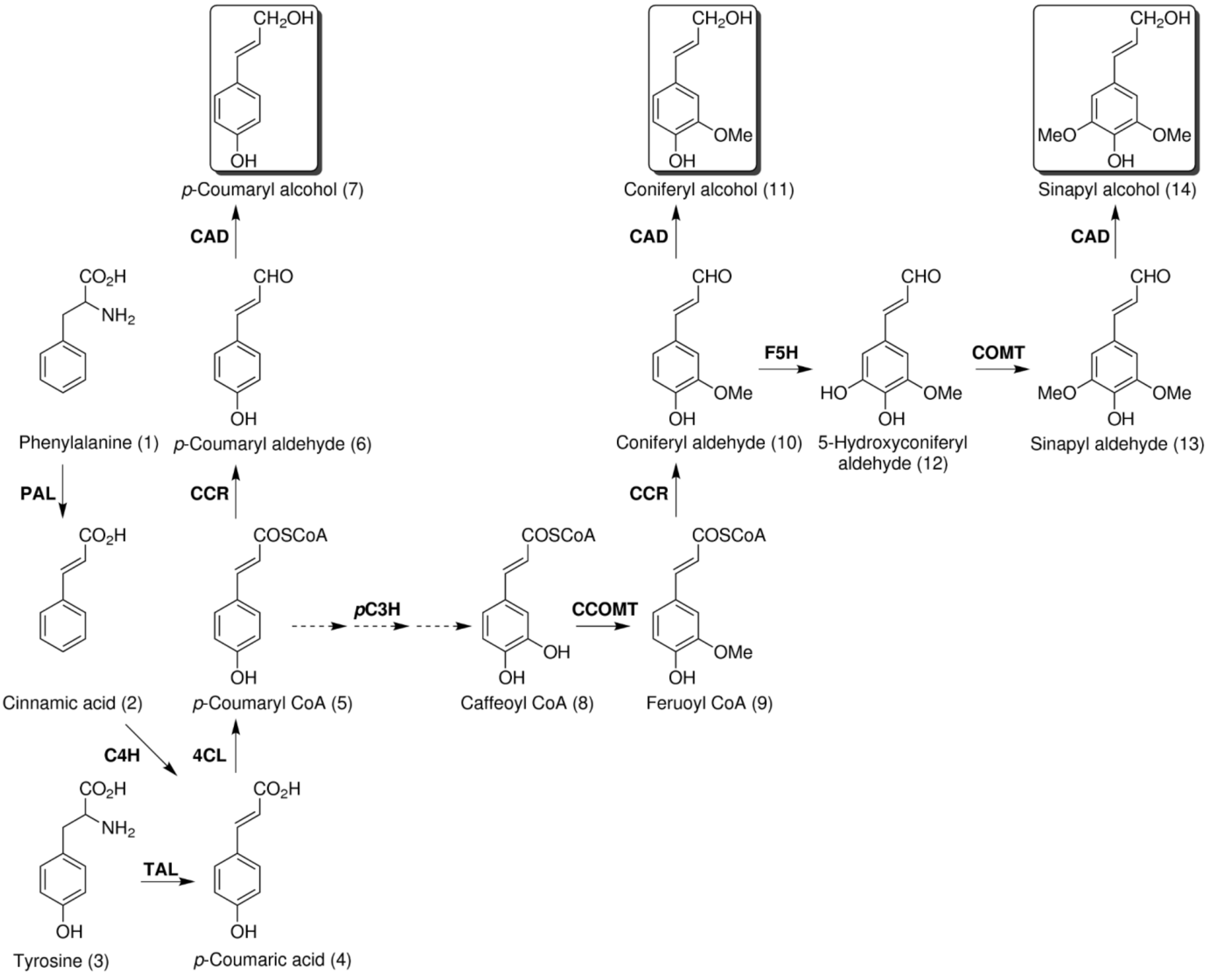
p-Coumaryl alcohol (7)
Coniferyl alcohol (11)
Sinapyl alcohol (14)
CAD
CAD
CAD
Phenylalanine (1)
p-Coumaryl aldehyde (6)
Coniferyl aldehyde (10)
F5H
5-Hydroxyconiferyl aldehyde (12)
COMT
Sinapyl aldehyde (13)
PAL
CCR
CCR
Cinnamic acid (2)
p-Coumaryl CoA (5)
pC3H
Caffeoyl CoA (8)
CCOMT
Feruoyl CoA (9)
C4H
4CL
TAL
Tyrosine (3)
p-Coumaric acid (4)

Figure 1-3 – Phenylpropanoid biosynthesis pathway leading to the monolignols *p*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol (modified from [94]**).** PAL: phenylalanine ammonia lyase; C4H: cinnamate 4-hydroxylase; TAL: tyrosine ammonia lyase; 4CL: hydroxycinnamoyl CoA ligases; *p*C3H: *p*-coumarate-3-hydroxylase; CCOMT: hydroxycinnammoyl CoA *O*-methyltransferases; CCR: cinnamoyl-CoA oxidoreductases; F5H: ferulate 5-hydroxylase; COMT: caffeic acid *O*-methyltransferases; CAD: cinnamyl alcohol dehydrogenases. Dotted arrows: Further enzymatic steps are involved in the formation of (8) from (5) (see [94]).

1.2.3 Stereoselective control of lignan biosynthesis by dirigent proteins

The first step in the biosynthesis of lignans is performed by the one-electron oxidation of monolignols followed by a phenoxy radical coupling reaction [94]. The resonance stabilized phenoxy radicals can couple in various ways leading to a mixture of different chemical compounds and racemates *in vitro* [73,103]. However, these reactions can also occur in a regio- and enantioselective fashion *in planta* [104]. The identification of the DIR *Fi*DIR1 from *Forsythia intermedia* by Davin and coworkers in 1997 offered first insights into the proteinaceous basis for these control mechanisms [95]. In this study, *Fi*DIR1 was found to guide the formation of (+)-pinoresinol from two coniferyl alcohohl radicals (see Figure 1-4a). Although other enzymes also exhibit selectivity in lignan biosynthesis (see 1.2.2), the existence of DIRs might explain the presence of optically pure 8,8´-linked lignans and led to a new paradigm for the stereospecificty of phenoxy radical coupling in biological systems [104,105].

Recent data indicates that the stereoselective control of phenoxy radical coupling reactions might not be limited to the phenylpropanoid metabolism, but rather plays a more general role in various biosynthetic pathways *in vivo*. In this respect, DIR genes have been identified in many plants and a large gene family including 35 members in spruce (*Picea* spp.), 25 in thale cress (*Arabidopsis thaliana*) and 54 in rice (*Oryza sativa*) was established [106]. This implied a more general role of DIRs in the stereoselective control of bimolecular phenol coupling. Further, the atropselective coupling of hemigossypol to (+)-gossypol has recently been shown to involve the activity of a not yet characterized DIR in cotton (*Gossypium hirsutum*) [107].

Up to now, eleven DIRs have been expressed and characterized including *Fi*DIR1 from *Forysthia intermedia* [95,105], *Tp*DIR1 to *Tp*DIR9 from *Thuja plicata* [108] and *At*DIR6 from *Arabidopsis thaliana* [109,110]. While *Fi*DIR1 and *Tp*DIR1-9 favor the formation of (+)-pinoresinol, *At*DIR6 mediates the enantiocomplementary formation of (-)-pinoresinol (see Figure 1-4a).

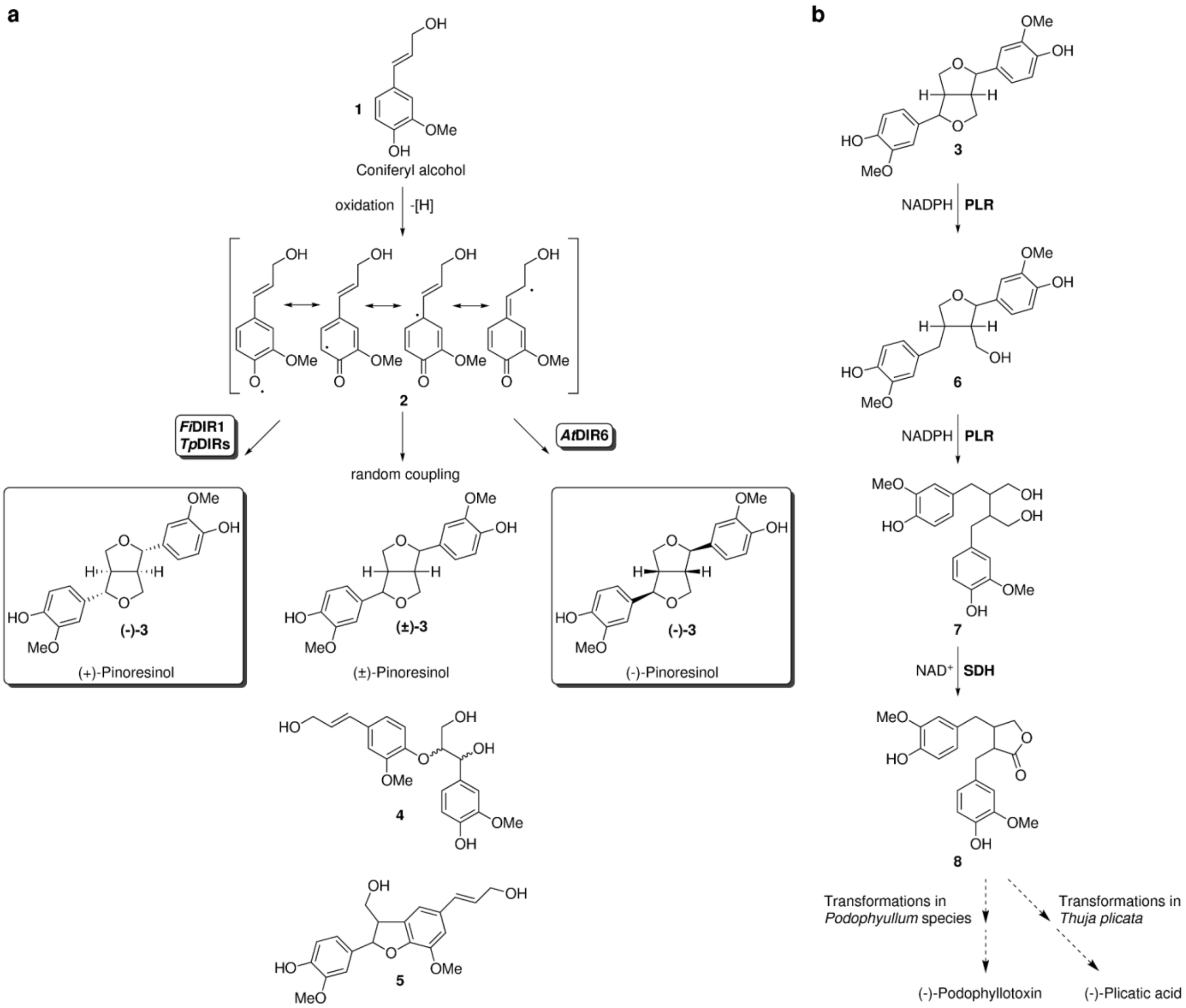
a
1
Coniferyl alcohol
oxidation
-[H]
2
FiDIR1
TpDIRs
random coupling
AtDIR6
(-)-3
(+)-Pinoresinol
(±)-3
(±)-Pinoresinol
(-)-3
(-)-Pinoresinol
4
5
b
3
NADPH
PLR
6
NADPH
PLR
7
NAD+
SDH
8
Transformations in Podophyullum species
Transformations in Thuja plicata
(-)-Podophyllotoxin
(-)-Plicatic acid

Figure 1-4 – Phenoxy radical coupling of coniferyl alcohol (a) and further biochemical pathway (b). (a) One electron oxidation of coniferyl alcohol (1) initiates the unspecific, oxidative coupling of resonance-stabilized radicals (2) to (±)-pinoresinol ((±)-3), *erythro*/*threo*-(±)-guaiacyl-glycerol coniferyl ethers (4), and (±)-dehydroconiferylalcohol (5). In the presence of *Fi*DIR1 and *Tp*DIRs, the product formation is strongly shifted to (+)-pinoresinol, while the enantiocomplementary *At*DIR6 shifts the product formation to (-)-pinoresinol. **(b)** The 8,8´-coupled lignan pinoresinol is subsequently converted into lariciresinol, secoisolariciresinol, and matairesinol by the action of pinoresinol-lariciresinol reductase (PLR) and secoisolariciresinol dehydrogenase (SDH).

1.2.4 Reaction mechanism of dirigent proteins

Although the precise reaction mechanism of DIRs is still elusive, a model was proposed based on the detailed biochemical and kinetic analysis of *Fi*DIR1 [95,104,111,112]. Since DIRs do not have any oxidative capacity for the initiation of the reaction on their own, free coniferyl alcohol radicals need to be generated by enzymes like laccases and peroxidases or by chemical oxidants such as ammonium peroxydisulfate in this model.

In a subsequent step, the two subunits of the functional DIR homodimer are thought to bind one coniferyl alcohol radical each. These radicals are then specifically oriented so that 8,8´-coupling yielding a quinone-methide intermediate and a subsequent intramolecular cyclization reaction is favored [112,113]. The model describes hence the formation of optically pure (+)- or (-)-pinoresinol by *Fi*DIR1 and *At*DIR6, respectively, instead of racemic mixtures of 8,8´-coupled (±)-pinoresinol, 8,5´-coupled (±)-dehydroconiferylalcohol and 8-O-4´-coupled *erythro*/*threo*-(±)-guaiacyl-glycerol coniferyl ethers obtained by random coupling [104,108,112].

1.2.5 Biochemical characterization of dirigent proteins

DIRs are generally small proteins of about 180 amino acids and a molecular mass of about 18 to 19 kDa [105,108,109]. So far, all characterized DIRs were either extracted from plant material or expressed in systems based on higher organisms such as plant and insect cell cultures, yielding only up to 200 µg per L of culture. Characterization of the quarternary structure of *Fi*DIR1 and *At*DIR6 implied a homodimeric assembly as the native form, which is in accordance with the proposed reaction mechanism (see 1.2.4) [110,111]. DIRs have been shown to be post-translationally modified by glycosylation *in planta* as well as in heterologous expression systems [105,108,110] at two to four *N*-glycosylation sites leading to a processed molecular mass of 21 to 26 kDa. In plants, the complex (2190 Da) and paucimannosidic (1170 Da) type of glycosylation patterns are the most prevalent [114]. While for most DIRs only the approximate size of the glycosylation, but not the exact pattern is known, a more detailed characterization of *At*DIR6 expressed in tomato cell cultures revealed a paucimannosidic type of glycosylation (Mannose$_3$–Xylose–*N*-acetylglucosamine$_2$–Fucose) with a core fucose residue at two *N*-glycosylation sites [110]. Further post-translational modification included a disulfide bond connecting the N- and C-terminus of *At*DIR6. Together with the glycosylation, both post-translational modifications were predicted to contribute to the stability and/or solubility of the protein.

Immunolocalization studies with *Fi*DIR1 revealed that the protein is located in the apoplast *in planta* [115]. This is in accordance with the observed N-terminal signalpeptides for all characterized DIRs [105,109]. The signal peptides are cleaved during secretion to form the mature protein. In this regard, it was found that the cleavage site in animal and plant expression systems can vary leading to a non-native amino terminus in the heterologous expression system [105]. Yet, this did not afflict the dirigent activity of the protein.

Recently, conceptual similarities between the proposed DIR mode of action and that of allene oxide cyclases (AOC) were noted [110,116]. AOCs also mediate a stereochemical control on a cyclization reaction, where the substrates are unstable allylic epoxides, which rapidly hydrolyze to ketols in aqueous media or can undergo spontaneous cyclization to yield racemic cyclopentenones [117,118]. However, in the presence of AOC, the hydrolysis is suppressed and the stereospecific cyclization of the substrate is favored. DIRs seem to share the ability of AOCs to bind hydrophobic ligands and impose a steric restriction on the substrate via the protein environment to preclude all but one possible reaction outcome.

Besides similarities in function, a homology model of *At*DIR6 based on the crystal structure of the allene oxide cyclase from *Arabidopsis thaliana* (*At*AOC2) could be established (see Figure 1-5) [110,119].

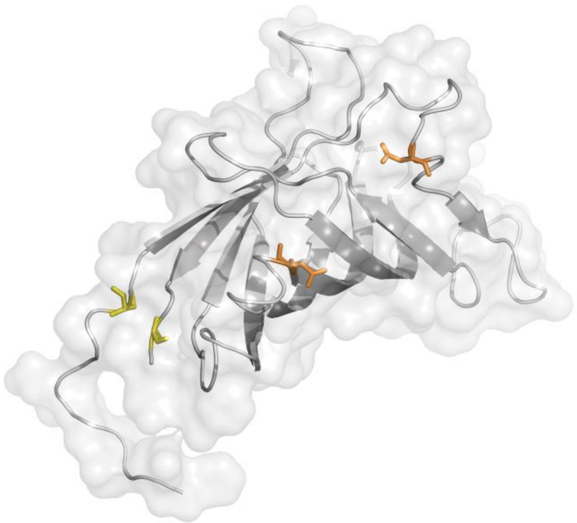

Figure 1-5 – Homology model of *At*DIR6 ([110]**).** The model of *At*DIR6 is shown in cartoon style with an 80% transparent surface. The two N- and C-terminal cysteine residues Cys40 and Cys186 forming a disulfide bond are shown as yellow sticks, while the two *N*-glycosylation sites N59 and N123 are shown as orange sticks. Numbering is according to Pickel *et al.* (2012) for the plant derived *At*DIR6 with the N-terminal secretion signal peptide. The model was visualized using Pymol 1.1. The model data is available in the PMDB database under the accession number PM0078038 (http://mi.caspur.it/PMDB/).

Although the sequence identity is rather low (11%) between *At*DIR6 and *At*AOC2, based on CD spectroscopic data the proteins (including *Fi*DIR1) all shared the fold of an AOC-like, eight-stranded antiparallel β-barrel with a central hydrophobic cavity for potential substrate binding. Based on this and the notion that DIRs like AOCs can be primarily seen as binding proteins for unstable lipophilic molecules, it was suggested to classify DIRs as homologs of AOCs belonging to the calycin superfamily of proteins. A distant relationship to lipocalins known for their capacity to bind hydrophobic ligands was proposed as well [110,120].

Recent mutational studies including alanine scanning mutagenesis and domain-swapping experiments between the (+)- and (-)-pinoresinol-forming DIRs *Sc*DIR from *Schizandra chinensis* and *At*DIR6 revealed further insights into the binding of the hydrophobic substrates to the DIR proteins [121]. Exchange of the phenylalanine residues F90, F113 and F163 in *Sc*DIR resulted in the loss of dirigent function supporting the presumed interaction in binding of the substrate. Based on the established homology model of *At*DIR6, residue F82 (F90 in *Sc*DIR), among other residues, was already suggested to partake in substrate binding [110]. In addition, domain swapping experiments uncovered that the region comprising residues K90 to L138 in *At*DIR6 is important for its stereoselectivity [121].

1.2.6 Physiological function

So far, the physiological function of DIRs is not completely clear. The molecular function in the stereoselective control of phenoxy radical coupling seems to be important for the first step of the lignan biosynthesis *in planta*. Based on the physiological function of lignans for the plants (see 1.2.2), a preventive function in the protection against pathogens by accumulation of lignans like plicatic acid in heartwood tissue seems reasonable [108]. This is supported by the induced transcription and expression of DIRs and DIR-like genes in e.g. *Picea spp*., *Gossypium hirsutum* and *Pisum sativum* upon infection by pathogens [106,122–124].

The participation of DIRs in the lignification process in plants is discussed controversially. Lignin, being the second most abundant organic plant substance in the terrestrial environment, is mainly composed of the three monolignols *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol [125]. It functions to impart strength to the cell walls, facilitate water transport, impede degradation of wall polysaccharides, and as a major defense against pathogens, insects and herbivores in vascular plants [126]. Historically, the lignin biosynthesis was described by a random (combinatorial) coupling model lacking stereoselective control [103]. With the discovery of DIRs and the differential targeting of the lignin monomers to distinct lignin initiation sites, evidence was found to support the theory that lignin primary structure is actually controlled at the proteinaceous level [105,115,125,127]. While several inconsistencies regarding the random coupling model such as implausible cross-linking frequencies in the lignin polymer and results based on unclear analytical results are discussed [125], the dirigent protein model also lacks proof for numerous problems. For example, there are no hints on the assembly of the lignin polymer by monolignol dimers, no evidence for optically active building blocks in the lignin polymer, the lack of DIRs facilitating 8-O-4´-coupling and the overall huge number of dirigent proteins needed in order to fulfill all the observed coupling specificities [126,128].

1.3 Ribokinases

1.3.1 D-ribose transport and metabolism

D-ribose is among the most abundant and important sugars in biological systems being a component of nucleic acids and many other biomolecules [129]. While there is only little knowledge of the ribose transport in eukaryotes, it has been intensively studied in *E. coli*. Here, the *rbs* operon encoding six proteins RbsDACBKR controls the D-ribose transport and is induced by the presence of ribose [130,131]. Like all monosaccharides, D-ribose exists in multiple configurations in solution, which are the β-pyranose (58.9%), α-pyranose (20.3%), β-furanose (13.3%), α-furanose (7.4%), and the acyclic aldehyde or hydrate forms (0.13%) [132] (see Figure 1-6a).

After D-ribose enters the periplasm of the cell by diffusing through pores in the outer membrane, the ribose binding protein RbsB recognizes and binds to β-D-pyranoribose, which is the major form in solution [133]. In an active transport step mediated by the permease RbsC and the ATP binding cassette component RbsA, β-D-pyranoribose is transported across the inner membrane into the cytoplasm [134].

Before entering the biosynthetic or metabolic pathways, D-ribose needs to be phosphorylated to D-ribose-5-phosphate by the action of the ribokinase RbsK [130]. This step also functions to trap the D-ribose in the cell after its uptake, since the negatively charged phosphate prevents leakage over the cell membrane [129]. Interestingly, while the most abundant β-D-pyranose form of D-ribose is transported, RbsK binds and phosphorylates the least populated cyclic α-furanose form of D-ribose, which is the only configuration that is used in biomolecules [135–137].

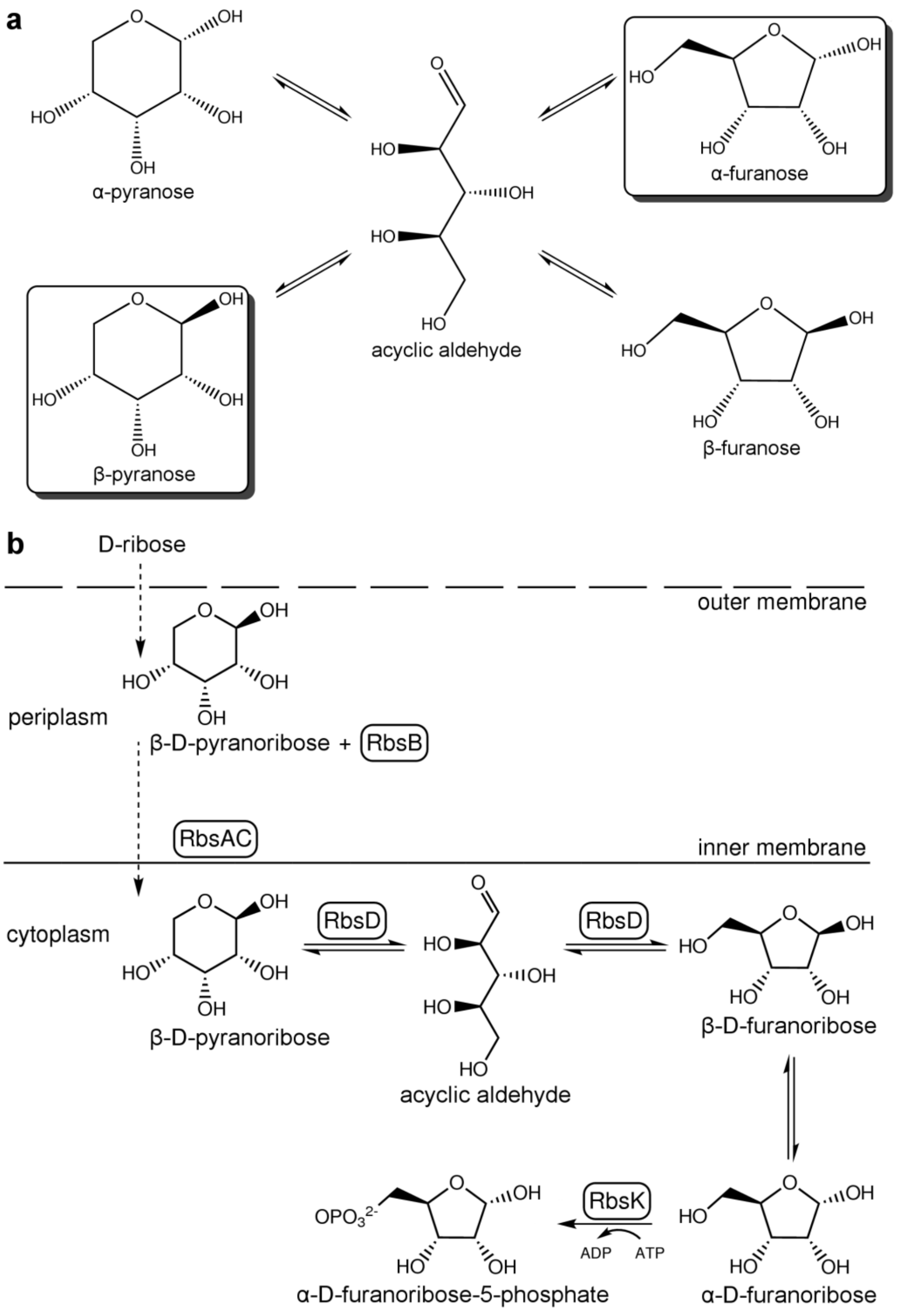
a
α-pyranose
acyclic aldehyde
α-furanose
β-pyranose
β-furanose
b
D-ribose
outer membrane
periplasm
β-D-pyranoribose + RbsB
RbsAC
inner membrane
cytoplasm
RbsD
RbsD
β-D-pyranoribose
acyclic aldehyde
β-D-furanoribose
RbsK
ADP
ATP
OPO_3^{2-}
α-D-furanoribose-5-phosphate
α-D-furanoribose

Figure 1-6 – D-ribose configurations in solution (a) and transport (b). (a) Monosaccharides like D-ribose exist in various configurations in solution. The cyclic forms of the sugar interconvert via the acyclic aldehyde form. The β-pyranose and the α-furanose configurations of D-ribose are highlighted since they represent the forms, which are transported into the cell and phosphorylated, respectively. **(b)** D-ribose diffuses into the periplasm, where the β-D-pyranose form is bound by RbsB, actively transported into the cytoplasm and converted into the β-D-furanose form by RbsD. The β-anomer spontaneously changes to the α-anomer, which is phosphorylated by RbsK to α-D-furanoribose-5-phosphate.

The cyclic forms of sugars can interconvert spontaneously via the acyclic aldehyde form. While it was shown for D-glucose that the α-β anomeric change (mutarotation) is a very slow process, which can be accelerated by a galactose mutarotase [138], the conversion between α- and β-furanoses is a fast and spontaneous process due to a lower energy barrier for the necessary ring opening for furanoses compared to pyranoses [137]. Although mutational studies showed that RbsD is not essential [139], it was characterized as a pyranase, which increases the exchange rate between the β-anomers β-ribopyranose and β-ribofuranose. This allows an efficient supply of the α-furanose form of D-ribose for the phosphorylation by RbsK and eliminates the discrepancy in the substrate specificities for uptake and phosphorylation. Figure 1-6b summarizes the transport steps, changes in configuration and phosphorylation of D-ribose leading to the formation of D-ribose-5-phosphate.

In further biosynthetic and metabolic steps, D-ribose-5-phosphate is used as a carbon source for the synthesis of nucleotides, cofactors like NADH and FAD, histidine, and tryptophane as well as for energy production after conversion to other sugar compounds in the non-oxidative phase of the pentose phosphate pathway [129,130]. Besides the formation of D-ribose-5-phosphate by phosphorylation of free D-ribose, two main pathways are involved in pentose phosphate biosynthesis: the oxidative branch of the pentose phosphate pathway and the phosphorylase mediated pathway [131]. For the *de novo* as well as the “salvage” synthesis of purine and pyrimidine nucleotides, the essential compound phosphoribosyl-5-pyrophosphate (PRPP) is generated from D-ribose-5-phosphate.

1.3.2 The ribokinase family of proteins

The first step in the metabolism of D-ribose is catalyzed by ribokinases, which are classified as phosphotransferases with an alcohol group as acceptor acting on the pentose sugar D-ribose (E.C. 2.7.1.15). Ribokinase genes have been found in both prokaryotes and eukaryotes, e.g. in *E. coli*, yeast and mammals [140–142]. Sequence comparisons have shown that they belong to the ribokinase (RK, also called PfkB) family of sugar kinases, which also includes other members like adenosinekinase, fructokinase, 2-dehydro-3-deoxygluconokinase, 1-phosphofructokinase, tagatose-6-phosphate kinase, inosine kinase, and 6-phosphofructo-2-kinase (PfkB or PFK2) [129]. All members of the RK family can be identified by the presence of two highly conserved sequence motifs. The first motif is a glycine-rich area in the N-terminal region with two consecutive glycine residues forming part of the hinge between lid and α/β domain (see 1.3.3). The second motif (GXGD) found in the C-terminal region contains a conserved aspartic acid serving as the catalytic base and is furthermore involved in ATP binding and formation of an oxyanion hole. While the overall sequence identity of members of the RK family is less than 30%, the structures are all very similar. The main structural differences can be found in the sugar substrate-binding site and the peripheral regions leading to broad substrate specificities within a distinct structural framework.

Besides the RK family, the hexokinase family and the galactokinase family form two other non-homologous branches of sugar kinases [141] classified according to the most abundant sugars. No significant sequence similarities or shared motifs between the different families can be found. Furthermore, the enzyme structures greatly differ as well. Interestingly, despite large differences in the primary and tertiary/quarternary structures, some enzymes e.g. fructokinases from various organisms show similar substrate specificities. Here, convergent evolution led to distinct three-dimensional enzyme folds, which can be classified into different enzyme families, but catalyze a chemically equivalent reaction on similar or identical substrates. Other examples for convergent evolution include e.g. serine proteases, tRNA synthases or superoxide dismutases.

In a comparison of sequence similarity and identity of genes annotated as ribokinases from eukaryotic and bacterial organisms (including *Homo sapiens*, *Drosophila melanogaster*, *Arabidopsis thaliana*, *S. cerevisiae*, *E. coli* and further organisms) it was noted that human ribokinase shared a higher sequence identity with protist and bacterial species (35 to 40% identity) as compared to the homologs from plants and fungi [143]. These results were unexpected since animals are more closely related to fungi and plants than to bacteria. The biochemical characterization of the human ribokinase protein established it as a true ribokinase, however, since the annotation of proteins from plants and fungi as ribokinases is mostly based on observed sequence similarities, a characterization of the proteins regarding their substrate specificity is also necessary to verify these as ribokinases and not some other kind of sugar kinases.

1.3.3 Structure and reaction mechanism of the *Escherichia coli* ribokinase

The first crystal structure of a protein of the ribokinase family, the ribokinase from *E. coli,* was solved and refined to a resolution of 1.8 Å in 1998 [135] (see Figure 1-7). The *E. coli* ribokinase is a protein with a molecular mass of about 33 kDa consisting of 309 amino acids. Earlier experiments suggested that it exists as a homodimer in solution [144], although only one monomer was found in the asymmetric unit of the original structure. Further structure solutions of the apo structure and various substrate containing structures contained dimers in the asymmetric unit and also lead to a more detailed understanding of the conformational changes upon substrate binding [136]. Each monomer consists of two domains. The large α/β domain is made up by a nine-stranded β-sheet, which is flanked by 10 α-helices, and the smaller domain is formed by four β-strands [135]. While the large domain provides for the specific binding of the sugar substrate D-ribose and the nucleotide ATP and therefore also contains the catalytically active residues, the smaller domain acts as a lid over the active site. Additionally, the outer face acts as the interface for the dimer interaction forming a flattened β-barrel with the β-sheet of the smaller domain of the second monomer.

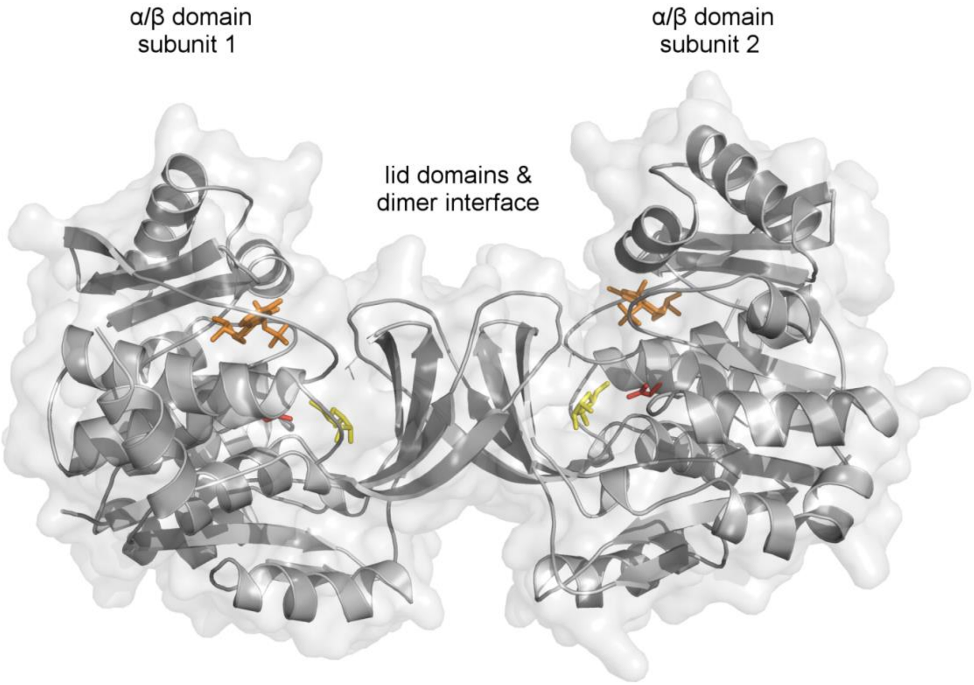

Figure 1-7 – Structure of the *E. coli* ribokinase. The biological assembly of the ribokinase dimer is shown in a cartoon style with an 80% transparent surface based on the PDB file 1RKD [135]. The data represents the closed form of the enzyme with an ADP analogue (orange) and D-ribose (yellow) bound. The conserved aspartate residues at position 255 of each monomer serving as the catalytic bases are shown as sticks (red).

The native homodimer harbors two active sites and each subunit contributes in forming the active site of the dimer partner by the interaction of the lid domains. D-ribose, in its α-furanose form, binds in the cleft of the α/β domain, where it is stabilized via various direct and indirect hydrogen bonds from amino acid side chains and main chain nitrogens. Comparison of the apo structure to the substrate-bound forms revealed drastic conformational changes [135,136], which can be seen as an opening and closing of the lid domain. The apo structure resembles the open conformation, while the substrate bound structure takes the closed conformation with the D-ribose buried almost completely under the small lid domain. This leads to further interactions of the lid domain with D-ribose, however these are mainly of hydrophobic nature involving four isoleucine and one alanine residue. ATP binds in a Rossmann fold-like domain known for NAD binding with fewer direct hydrogen bondings compared to the interactions with D-ribose. Yet, the closing of the lid upon D-ribose binding is supposed to enhance the affinity for the nucleotide. It was suggested that kinases in general use an ordered binding of substrates as a means of preventing premature ATP hydrolysis in an induced fit mechanism. In this light, the accepted reaction mechanism is an ordered Bi-Bi mechanism similar to those of other kinases [129,145]. After binding the co-substrate ATP, small conformational changes orient the γ-phosphate group of ATP in the vicinity of the O5´-hydroxyl group of D-ribose [136]. The conserved aspartate at position 255, which is found in the GXGD motif (see 1.3.2), serves as the catalytic base abstracting a proton from the O5´-hydroxyl group [135]. The negatively charged oxygen atom subsequently makes a nucleophilic attack on the γ-phosphate group of ATP in an in-line S_N2 reaction leading to a pentacovalent transition state, which is stabilized by an anion hole. Further movements of the lid domain lead to the breakdown of the transition state and the release of the products, ADP and D-ribose-5-phosphate.

1.3.4 Biochemical characterization of ribokinases

First characterization of a ribokinase was performed with the protein identified and purified from calf liver more than 50 years ago [140]. The biochemical characterization of eukaryotic ribokinases from *Homo sapiens* and *Leishmania major* and bacterial ribokinases from *E. coli, Salmonella typhimurium*, and *Staphylococcus aureus* has been reported up to now as well, with the best studied ribokinase coming from *E. coli* [143,146–149]. So far, although ribokinase genes have been also found in plants and fungi [141,150], no proteins from these kingdoms were characterized.

All characterized ribokinases show a strong dependence on mono-, di- and pentavalent ions for their activity [129]. The divalent cation Mg^{2+} is a prerequisite in the binding of ATP and the transfer of the phosphate group. The actual co-substrate for ribokinases is the metal chelate $(ATP^{4-}Mg^{2+})^{2-}$ and divalent cations are supposed to neutralize the negative charges on the phosphate groups of the nucleotide. No activity is found without divalent cations, however others such as Mn^{2+} can substitute for Mg^{2+}.

Ribokinases, like all other known carbohydrate kinases, show an additional requirement for monovalent cations. Here, two general types of activation can be distinguished. A direct (catalytic) activation with the monovalent cation binding in the active site can be found for enzymes such as the pyruvate kinase or the molecular chaperone GroEL [151]. In the case of ribokinases, an indirect (conformational) type of activation by the physiological cation K^+ has been suggested [152]. The structural basis for this activation was recently elucidated by comparison of the crystal structure of the *E. coli* ribokinase in presence of Cs^+ as monovalent cation and the structure of the *Staphylococcus aureus* ribokinase in the absence of a monovalent cation [148]. A structural rearrangement of the large ATP loop forming the nucleotide binding pocket upon binding of the monovalent cation was proposed to enable ATP binding and enhance catalytic efficiency.

A detailed study showed the importance of an NXXE motif in adenosine kinase for activation by the pentavalent cation phosphate and its effect on Mg^{2+} binding [153]. The human form of adenosine kinase shows a remarkable structural homology to ribokinase from *E. coli*. In the crystal structure of the *E. coli* ribokinase, a phosphate bound in the active site with close contacts to Asn187 and Glu190 forming an NXXE motif was also found [135]. The activation by phosphate was shown to increase the catalytic turnover and the substrate affinity for D-ribose [154]. Importantly, ribokinase and other enzymes depending on a activation by phosphate such as adenosine kinase, 6-phosphofructo-1-kinase and PRPP synthetase, which are all important for the synthesis of essential metabolites, can be fully regulated by physiological levels of phosphate [129]. While other pentavalent ions like vanadate or arsenate can substitute for phosphate, there are also a large number of phosphorylated compounds known, which serve as activators or inhibitors of adenosine kinase and ribokinase [143,155,156].

Ribokinases show high substrate specificity towards the phosphate-accepting substrate. While the phosphorylation of D-ribose is exceptionally efficient, many other comparable sugars are only very weak substrates or not accepted at all. The ribokinase from *E. coli* accepts 2-deoxy- D-ribose rather well, while D-arabinose and D-xylose, which differ at the C2 or C3 functional group, as well as D-fructose, which varies in the chain length (C6), are already very poor substrates. Further sugars such as D-lyxose, L-ribose, L-arabinose, and D-glucose were not phosphorylated [146]. Ribokinases from *Homo sapiens* and *Salmonella typhimurium* showed comparable substrate specificity with the *E. coli* enzyme, while the ribokinases from calf liver and *Leishmania major* did not show activity towards any other sugar [140,143,147,149]. In this respect, ribokinase from *E. coli* showed the highest catalytic efficiency of all characterized enzymes, while the relative activities towards other accepted sugars was only about 1%, which might explain why no activity towards these sugars was measured with the some of the other enzymes.

1.3.5 Biotechnological and medicinal applications

The metabolism of D-ribose has received much attention recently since the beneficial effects of D-ribose supplementation on the cardiovascular system have been studied [129]. By the action of ribokinase and PRPP synthetase (PRPPS), PRPP is formed from D-ribose, which is an essential precursor for the production of energy-rich purine and pyrimidine nucleotides. As the availability of PRPP under pathologic cardiac conditions is limited, nucleotide synthesis is hampered. As the repletion of nucleotides following ischemic insults is important for the recovery of the heart function, a steady supply with D-ribose, which is metabolized into ribose-5-phosphate and then PRPP is advantageous for the cells´ ability to regenerate ATP and restore proper function.

A more direct application for ribokinase was suggested in the enzymatic synthesis of nucleosides and analogues (see Figure 1-8) [146,157]. Chemical preparations of α-D-pentofuranose-1-phosphates (α-PFP-1) as precursors for the enzymatic synthesis of purine nucleosides by purine nucleoside phosphorylase (PNP) are rather complicated. An enzymatic approach is the transformation of pentofuranose-5-phosphates (PFP-5) via the action of phosphopentomutase (PPM) to α-PFP-1 [146]. PFP-5 can be synthesized chemically; however an enzymatic step employing ribokinase for the formation of PFP-5 from D-ribose was suggested. The high specificity found for many ribokinases could limit this approach, yet an alteration or extension of the substrate spectrum of the employed ribokinase by enzyme engineering, specifically rational design based on the available structural and biochemical information, might solve this limitation. Instead of the conversion of PFP-5 into α-PFP-1, it can also be used for the formation of PRPP by PRPPS, which can then be used for the synthesis of nucleosides [131]. Gross *et al.* compared the chemical synthesis of PRPP with an enzymatic transformation including ribokinase and PRPPS [157]. They found the enzymatic method to be the most attractive procedure in principle, however it lacked specifically in the high yield production of ribokinase, which otherwise showed beneficial traits regarding activity, stability and immobilization yield.

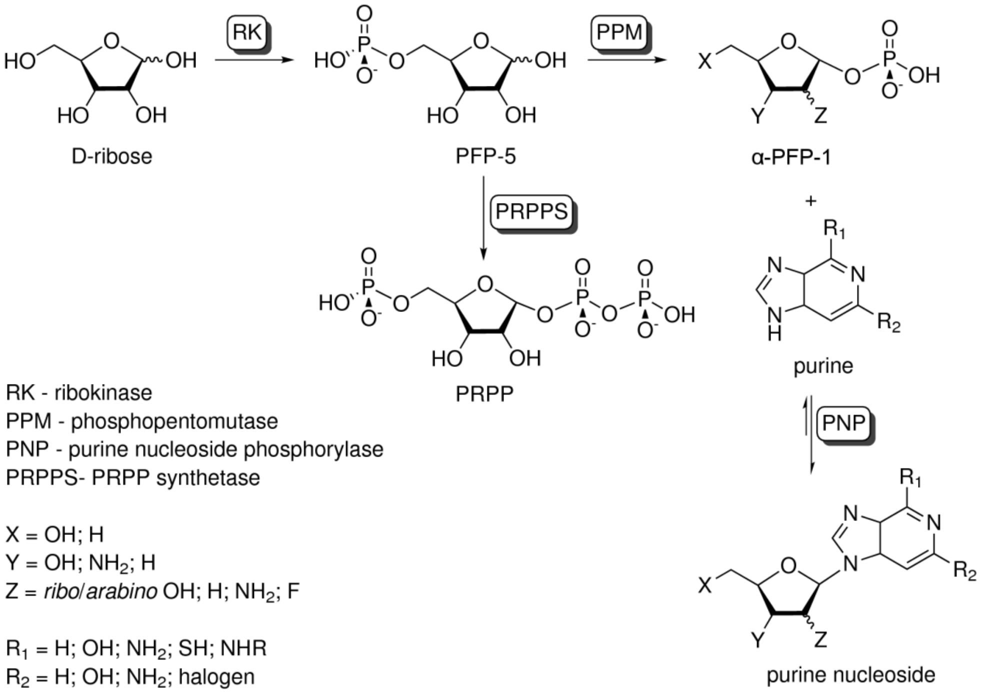
RK
PPM
PRPPS
PNP
D-ribose
PFP-5
α-PFP-1
PRPP
purine
purine nucleoside
RK - ribokinase
PPM - phosphopentomutase
PNP - purine nucleoside phosphorylase
PRPPS- PRPP synthetase
X = OH; H
Y = OH; NH2; H
Z = ribo/arabino OH; H; NH2; F
R1 = H; OH; NH2; SH; NHR
R2 = H; OH; NH2; halogen

Figure 1-8 – Enzymatic synthesis of purine nucleosides from pentoses (modified from [146] **and** [157]**).**

1.4 Cutinases

1.4.1 Structure and function

Cutinases (E.C. 3.1.1.74) belong to the enzyme class of hydrolases and are named for their ability to degrade cutin, the cuticular polymer matrix of higher plants [158]. Cutin is a polymer of hydroxy and epoxy fatty acids with common chain lengths of n-C_{16} and n-C_{18}. It is essential for plants in the protection from microbial infection and dehydration. Many phytopathogenic fungi produce extracellular cutinases for the enzymatic hydrolysis of cutin to support the initial infection of plants. Cutinases were found in a variety of different organism including fungi such as *Alternaria brassicicola*, *Aspergiulls oryzae*, *Fusarium solani f. pisi*, *Fusarium oxysporum* or *Humicola insolens* [159].

Based on the work with the cutinase from *Fusarium solani f. pisi* (FsC), cutinases were shown to belong to the class of serine esterases and the α/β hydrolase superfamily [158]. For FsC, the crystal structure has been solved several times up to a resolution of 1.0 Å [160–163]. With a molecular mass of 22 kDa, cutinases are the smallest members of this group, which have been described so far. The catalytic triad is comprised of an active serine, an essential histidine and a carboxyl residue (usually aspartate) and the reaction mechanism is based on the formation of an acyl-enzyme intermediate (see Figure 1-9) [164].

In FsC, the catalytical serine is situated in an α/β hairpin and the consensus sequence Gly-Tyr-Ser-Gln-Gly containing the active serine matches the consensus sequence in most lipases (Gly-His/Tyr-Ser-X-Gly) [165]. Interestingly, unlike in lipases, the catalytic site is solvent-accesible, i.e., no lid completely covers the catalytic site. With lipases, the presence of a lid leads to a so-called interfacial activation. For activity, most lipases need to interact with a hydrophobic interface (e.g. substrate emulsion in water), which leads to a conformational change of the lid rendering the catalytic site accessible to substrate and thereby increasing the rate of hydrolysis. Cutinase binding to interfaces does not lead to a main-chain reaarangement, but only the reorientation of a few hydrophobic residues [158,165]. Therefore, cutinases are considered to be intermediates between lipases and esterases [166].

In lipases, the lid movement also allows the formation of the oxyanion hole, which stabilizes the negative charge of the tetrahedral intermediate during the nucleophilic attack on the substrate [167]. In cutinases, the oxyanion hole seems to be pre-formed instead of being induced by ligand binding [162]. However, 3D NMR studies [168] and comparison of various crystal structures [160] showed some divergence indicating that the region comprising the oxyanion hole might also feature small conformational changes upon substrate binding.

Cutinases show a broad substrate spectrum including soluble synthetic esters and emulsified triglycerides under and above the critical micellar concentration [158]. They prefer short-chain length substrates such as *para*-nitrophenylbutyrate and tributyrin, which is related to the narrow and interrupted substrate binding site [169]. Like lipases, cutinases are able to catalyze esterification, transesterification and interesterification in organic media or at low water activities. When acting on triacylglycerol substrates, cutinases show high stereo- and regioselectivity towards the *sn*-3 position. Since cutinases and lipases both belong to the class of serine hydrolases, the basic mechanism of hydrolysis and esterification reactions is comparable [164,170] (see Figure 1-9).

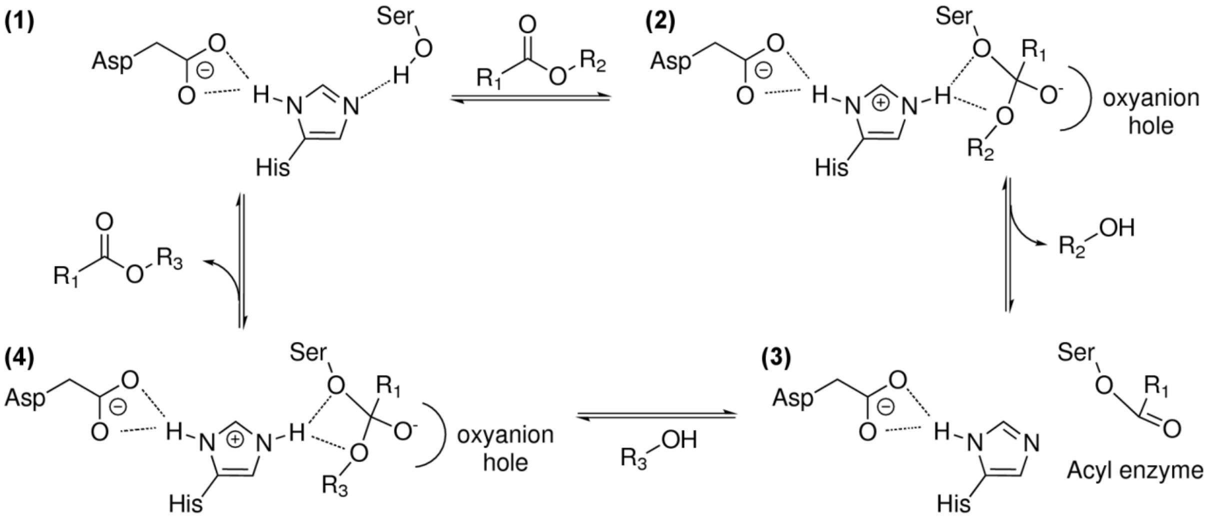

Figure 1-9 – Reaction mechanism of cutinases (modified from [171]**).** In a transesterification reaction (alcoholysis) with cutinases, first the ester binds to the enzyme (1). The catalytic serine, activated by the histidine and aspartate, starts the nucleophilic attack on the carbonyl forming the first tetrahedral intermediate (TI), which is stabilized by the oxyanion hole (2). The TI collapses and releases the first product leaving the enzyme acylated (3). During hydrolysis, a water molecule attacks the acyl enzyme, in a transesterification (alcoholysis) reaction an alcohol attacks forming a second TI (4). Dissociation releases an acid (hydrolysis) or a new ester (transesterification). The enzyme is free and ready for the next cycle (1).

1.4.2 Biocatalytic applications

Cutinases offer interesting biotechnological and industrial applications, which have been comprehensively reviewed [158,172]. Potential applications would be in the dairy industry, as industrial cleaning or household detergents, in the oleochemical-, polymer-, textile-, or food industry as well as for the synthesis of ingredients for personal-care products, agrochemicals and pharmaceuticals. For biocatalytic applications, cutinases are mostly used in immobilized form e.g. immobilized onto solid supports, covalently coupled, absorbed or microencapsulated in reversed micelles. Microencapsulation of cutinases in reversed micelles of surfactants such as AOT, SDS or Triton X-100 was successfully used in hydrolysis, esterification and transesterification. However, it was found that cutinases display relatively low stability in anionic surfactants like AOT [173]. Studies were performed to increase the stability in the presence of AOT [174,175]. Examples for the application of cutinases are the synthesis of aromatic esters by employing FsC in reversed micelles of AOT/isooctane [172], degradation and detoxification of the organophosphate malathion by the *Fusarium oxysporum* cutinase [176], degradation of the synthetic polyester polycaprolactone with FsC [177], and the enzymatic modification of polyesters like PET with FsC [178]. Furthermore, *Humicola insolens* cutinase (HiC) was used in polyester synthesis and lactone ring-opening polymerizations [179,180]. Recently, Baker *et al.* compared five cutinases from *Alternaria brassicicola*, *Aspergillus fumigatus*, *Aspergillus oryzae*, HiC and the well characterized FsC [181]. Interestingly, all newly characterized cutinases showed higher activity for the hydrolysis of small soluble esters as compared to FsC. In this regard, HiC exhibited the highest stability and activity versus small soluble esters and in the degradation of synthetic polyester (poly(ε-caprolactone)) compared to the other cutinases.

1.4.3 Functional monomers and polymers

Functional polymers with thiol and acrylate functional groups based on standard acrylic monomers have a great potential for industrial application. Traditionally, they are prepared using thiol-ene chemistry and are already in use as lacquers for improved scratch-resistance, adhesive formulations for high tack and instant bond strength or sizing of paper and textiles for improved printability and reduced abrasiveness [182]. However, the preparation of functionalized acrylic esters by chemical synthesis employs harsh reaction conditions often leading to premature polymerization and is further hampered by undesired side-reactions like the Michael addition of alcohols to acrylic esters and the thiol-Michael-addition of mercaptoalcohols to acrylic esters [183]. Polymerization inhibitors and acid catalysts have been used to avoid these problems and allow reactions at 80°C to 100°C [184]. Enzymatic processes employing biocatalysts such as hydrolases could be beneficial in the functionalization of acrylic ester monomers since the high chemoselectivity allows the introduction of functional groups into acrylic esters by transesterification.

The transesterification of acrylates with sterically hindered diols has been described for the first time with a lipase from *Chromobacterium viscosum* [185] and the lipase CalB from *Candida antarctica* was found to be the superior catalyst compared to other enzymes in the transesterification of methyl acrylate with 1-undecanol [186]. N-(2-hydroxyethyl)-acrylamide (HEA), which can be used as polymer for the application as matrix in DNA sequencing, was synthesized in a lipase catalyzed transesterification (aminolysis) reaction of ethyl acrylate with 2-amino-1-ethanol [187]. Highest yields of HEA were obtained in diisopropyl ether as solvent with addition of 3% (w/w acrylate ester) butylated hydroxytoluol (BHT). BHT is a radical inhibitor, which helped to reduce formation of Michael-addition by-products. It was shown for the acrylation of octanol in a solvent-free system with ethyl acrylate as donor of the acrylic group that the water activity is an important factor, which needs to be controlled [188]. Furthermore, the choice of the solvent is critical for high reaction rates of enzymatic acrylation [189]. Recently, the enzymatic synthesis of thiol functionalized acrylic monomers by CalB (Novozyme 435®) was coupled with a subsequent one pot reaction involving free radical polymerization, removal of thiol protection groups and thiol-ene ´click´reactions to yield thiol functionalized poly(meth)acrylates [190]. After removal of the enzyme by filtration, no further purification steps of monomers or intermediates were necessary and no side product formation influenced the reaction course. Many patents in this field by chemical companies like BASF SE emphasize the industrial interest for enzymatic synthesis of acrylates [191,192], e.g. the monoacrylation of polyols with CalB [193].

1.5 Objectives of the thesis

In the scope of the present thesis, selected steps are highlighted, which remain challenging in biocatalysis. The focus was set on case studies in the expression and characterization of proteins as well as in the enzyme engineering and optimization of reaction parameters in a biocatalytic application. These are key steps in establishing biocatalytic processes (see Figure 1-1).

1.5.1 Expression of the *Arabidopsis thaliana* dirigent protein *At*DIR6

Dirigent proteins are involved in the stereoselective control of the first step of lignan biosynthesis, the phenoxyl radical coupling of coniferyl alcohol radicals to yield optically pure (+)- or (+)-pinoresinol. Extraction of dirigent proteins from plant material or expression in systems based on higher organisms such as plant and insect cell cultures only resulted in low yields. For the solution of a structure, which might help in the elucidation of the exact reaction mechanism, and possible biotechnological applications, larger amounts of protein are needed.

The expression of *At*DIR6 was chosen to serve as a model for challenges in the production and characterization of biocatalysts. The goals of this case study were to:

- establish a heterologous expression platform, which allows the cost effective-, fast-, and robust expression of *At*DIR6 in high yields
- characterize *At*DIR6 in regard to function and structure compared to *At*DIR6 derived from other expression systems
- analyze the glycosylation pattern of *At*DIR6 derived from *Pichia pastoris* and evaluate its impact on function, structure, and solubility of the protein

1.5.2 Enzyme engineering of the *Saccharomyces cerevisiae* ribokinase RBK1

Ribokinase catalyzes the first step in the metabolism of D-ribose, which is the phosphorylation using ATP as co-substrate. The enzyme is highly conserved and shows high selectivity towards D-ribose since it catalyzes an essential step in the central metabolism. For a possible biocatalytic application, a broad substrate spectrum would be interesting. In this regard, an increase in activity of the ribokinase RBK1 from *S. cerevisiae* towards D-xylose was of interest. This could allow establishing a novel D-xylose metabolizing pathway in

S. cerevisiae for the production of second generation biofuels from lignocellulosic feedstocks.

The alteration of the substrate spectrum of RBK1 was chosen to serve as a model for challenges in the protein engineering and characterization of biocatalysts. The goals of this case study were to:

- establish a heterologous expression system for RBK1
- characterize the optimum reaction conditions and the substrate spectrum of RBK1
- establish a small focused library of RBK1 variants based on rational design and screen the library for alterations in the substrate spectrum with a focus on D-xylose
- solve a crystal structure of RBK1 with co-crystallized substrates to allow more precise bioinformatic predictions to support the rational design approach

1.5.3 Biocatalytic application of the *Humicola insolens* cutinase HiC

Cutinases are hydrolytic, cutin-degrading enzymes with a broad substrate spectrum. They offer potential applications in polymer chemistry and could be used, along with lipases, in the synthesis of functional polymers by transesterification of acrylates. For a biocatalytic application, the optimum process parameters need to be established.

The use of immobilized HiC for the synthesis of thiol-functionalized acrylic esters by transesterification was chosen to serve as a model for challenges in the application and characterization of biocatalysts. The goals of this case study were to:

- characterize the kinetics of HiC with various *p*NP esters
- prepare thiol-functionalized acrylic esters by employing immobilized HiC and optimize the reaction parameters
- purify and characterize the main product

2. Materials & Methods

2.1 Materials

2.1.1 Chemicals and enzymes

All chemicals and media were of analytical grade or higher and purchased from Fluka (Buchs, Switzerland), Sigma (Taufkirchen, Germany) or Bio-Rad (Munich, Germany) unless otherwise stated. All enzymes and enzyme buffers were from Thermo Scientific (St. Leon-Rot, Germany) or New England Biolabs (Ipswich, MA, USA) unless otherwise stated. Diaion HP20 was a gift from BASF SE (Ludwigshafen, Germany). Primary antibody towards *At*DIR6 was a kind gift of Prof. Dr. Andreas Schaller (University of Hohenheim, Stuttgart). Purified *At*DIR6 expressed in tomato cell cultures for comparison with our *Pichia pastoris* (*P. pastoris*) derived *At*DIR6 in SDS PAGE analysis and dirigent activity assays was also kindly provided by Prof. Dr. Andreas Schaller. The expression vector p4782.1 (pJOE) was kindly provided by Dr. Josef Altenbuchner (University of Stuttgart, Germany). D-xylose-5-phosphate (>95% purity) was synthesized and purchased from Synovo GmbH (Tübingen, Germany).

2.1.2 Media

BMGY (buffered glycerol complex medium for yeast)

20 g L^{-1}	peptone
10 g L^{-1}	yeast extract
10 %	1 M potassium phosphate buffer at pH 6.0
1 %	glycerol

BMMY (buffered minimal medium methanol yeast)

20 g L^{-1}	peptone
10 g L^{-1}	yeast extract
10 %	1 M potassium phosphate buffer at pH 6.0
0.5 %	methanol

YPD medium (yeast peptone dextrose)

10 g L^{-1}	yeast extract
20 g L^{-1}	peptone
16 g L^{-1}	agar (solid media only)
20 g L^{-1}	dextrose

YPDS medium (yeast peptone dextrose sorbitol)

10 g L^{-1}	yeast extract
20 g L^{-1}	peptone
182.2 g L^{-1}	sorbitol
16 g L^{-1}	agar (solid media only)
20 g L^{-1}	dextrose

LB (lysogenic broth) [194]

10 g L^{-1}	tryptone
5 g L^{-1}	sodium chloride
5 g L^{-1}	yeast extract
16 g L^{-1}	agar (solid media only)

TB (terrific broth)

12 g L^{-1}	tryptone
24 g L^{-1}	yeast extract
8 mL L^{-1}	glycerol (50%)
10 %	1 M potassium phosphate buffer at pH 7.0

All media were sterilized before use in an autoclave and supplemented with ampicillin (100 µg mL^{-1}), kanamycin (30 µg mL^{-1}) or zeocin™ (25 µg mL^{-1} for *Escherichia coli* (*E. coli*) and 100 µg mL^{-1} for *P. pastoris*) when necessary.

2.1.3 Plasmids

Table 2-1 – Plasmids

PLASMID	INSERT (USED RESTRICTION SITES)	SOURCE
pPICZαA		Invitrogen, Carlsbad, CA, USA
pET16b		Merck, Darmstadt, Germany
pET22b		Merck, Darmstadt, Germany
pET28a		Merck, Darmstadt, Germany
pET32a		Merck, Darmstadt, Germany
p4782.1 (pJOE)	Maltose binding protein "*mal*E" (NdeI, BamHI) and enhanced green fluorescent protein "*´eGFP*" (BamHI, HindIII)	Dr. Joseph Altenbuchner, University of Stuttgart
pRN754	*Saccharomyces cerevisiae* ribokinase *Rbk1*	DSM, Netherlands
pTf16		Takara, Otsu, Japan
pG-Tf2		Takara, Otsu, Japan
pG-KJE8		Takara, Otsu, Japan
pPICZαA_HIC	*Humicola insolens* cutinase *hic* (EcoRI, KpnI)	this work (pITB118)
pPICZαA_*At*DIR6	*Arabidopsis thaliana* dirigent protein *Atdir6* (EcoRI, XbaI)	this work (pITB762)
pET16b_*At*DIR6	*Arabidopsis thaliana* dirigent protein *Atdir6* (NdeI, BamHI)	this work (pITB763)
pET22b_*At*DIR6	*Arabidopsis thaliana* dirigent protein *Atdir6* (NcoI, HindIII)	this work (pITB764)
pET28a_*At*DIR6	*Arabidopsis thaliana* dirigent protein *Atdir6* (NdeI, BamHI)	this work (pITB765)
pET32a_*At*DIR6	*Arabidopsis thaliana* dirigent protein *Atdir6* (NcoI, HindIII)	this work (pITB766)
pJOE_*At*DIR6	*Arabidopsis thaliana* dirigent protein *Atdir6* (NdeI, HindIII)	this work (pITB767)
pET28a_RBK1	*Saccharomyces cerevisiae* ribokinase *Rbk1* (NdeI, HindIII)	this work (pITB768)
pJOE_RBK1	*Saccharomyces cerevisiae* ribokinase *Rbk1* (NdeI, HindIII)	this work (pITB769)

Vector maps of the pET series vectors, the Takara chaperone plasmids, pJOE4782.1 (pJOE) and pRN754 can be found in the supplementary figures (see 9.1).

2.1.4 Strains

Table 2-2 – Strains

NAME	**GENOTYPE**	**SOURCE**
Escherichia coli DH5α	F$^-$ *φ*80/*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(r_k^- m_k^-) *pho*A *sup*E44 λ$^-$ *thi*-1 *gyr*A96 *rel*A1	Invitrogen, Carlsbad, CA, USA
Escherichia coli XL-1 blue	*end*A1 *gyr*A96(nal^R) *thi*-1 *rec*A1 *rel*A1 *lac* *gln*V44 F'[::Tn*10* *pro*AB$^+$ *lac*I^q Δ(*lac*Z)M15] *hsd*R17(r_K^- m_K^+)	Agilent, USA
Escherichia coli BL21(DE3)	F$^-$ *ompT* *hsdS*$_B$ (r_B^- m_B^-) *gal dcm* (DE3)	Invitrogen, Carlsbad, CA, USA
Escherichia coli Origami(DE3)	F$^-$ *ompT* *hsdS*$_B$ (r_B^- m_B^-) *gal dcm lac*Y1 *ahpC gor522*::Tn*10* (Tc^R) *trxB*::kan (DE3)	Invitrogen, Carlsbad, CA, USA
Escherichia coli Rosetta(DE3)	F$^-$ *ompT* *hsdS*$_B$ (r_B^- m_B^-) *gal dcm lac*Y1 (DE3) pRARE^6 (Cm^R)	Invitrogen, Carlsbad, CA, USA
Escherichia coli C43 (DE3)	F$^-$ *ompT* *hsdS*$_B$ (r_B^- m_B^-) *gal dcm* (DE3) (and further uncharacterized mutations) [195]	University of Hohenheim, Stuttgart
Pichia pastoris X-33	wild-type	Invitrogen, Carlsbad, CA, USA

2.1.5 PCR and sequencing primers

Table 2-3 – PCR and sequencing primers

PRIMER	**FUNCTION**	**SEQUENCE**
CKA_I101S_for	QuikChange® primer RBK1 focused library (I101S mutation)	5´- GGGTACCGCTACC*AGC*CT AATTGAAGAG -3´
CKA_I101S_rev	QuikChange® primer RBK1 focused library (I101S mutation)	5´- CTCTTCAATTAG*GC*GGTAGCGG TACCC -3´
CKA_I101V_for	QuikChange® primer RBK1 focused library (I101V mutation)	5´- GGGTACCGCTACC*GTT*CTAATT GAAGAG -3´
CKA_I101V_rev	QuikChange® primer RBK1 focused library (I101V mutation)	5´- CTCTTCAATTAG*AAC*GGTAGCG GTACCC -3´
CKA_I101T_for	QuikChange® primer RBK1 focused library (I101T mutation)	5´- GGGTACCGCTACC*ACC*CTAATT GAAGAG -3´
CKA_I101T_rev	QuikChange® primer RBK1 focused library (I101T mutation)	5´- CTCTTCAATTAG*GGT*GGTAGCG GTACCC -3´
CKA_I113V_for	QuikChange® primer RBK1 focused library (I113V mutation)	5´- GGTGGCCAAAATAGG*GTT*TG ATTGTAGAAGG -3´
CKA_I113V _rev	QuikChange® primer RBK1 focused library (I113V mutation)	5´- CCTTCTACAATCAA*AAC*CCTATT TTGGCCACC -3´
CKA_I113T_for	QuikChange® primer RBK1 focused library (I113T mutation)	5´- GGTGGCCAAAATAGG*ACC*TTG ATTGTAGAAGG -3´
CKA_ I113T_rev	QuikChange® primer RBK1 focused library (I113T mutation)	5´- CCTTCTACAATCAA*GGT*CCTATT TTGGCCACC -3´

PRIMER	FUNCTION	SEQUENCE
CKA_ T279S_for	QuikChange® primer RBK1 focused library (T279S mutation)	5´- GGTTGTTGATACT*AGC*GGAGCTGGAGATAC -3´
CKA_ T279S_rev	QuikChange® primer RBK1 focused library (T279S mutation)	5´- GTATCTCCAGCTCC*GCT*AGTATCAACAACC -3´
CKA_ A319G_for	QuikChange® primer RBK1 focused library (A319G mutation)	5´- CATTCAAAGAAAAGGT*GGT*GCTGAAAGCATGC -3´
CKA_ A319G _rev	QuikChange® primer RBK1 focused library (A319G mutation)	5´- GCATGCTTTCAGC*ACC*ACCTTTTCTTTGAATG -3´
CKA_ A319S_for	QuikChange® primer RBK1 focused library (A319S mutation)	5´- CATTCAAAGAAAAGGT*AGC*GCTGAAAGCATGC -3´
CKA_ A319S _rev	QuikChange® primer RBK1 focused library (A319S mutation)	5´- GCATGCTTTCAGC*GCT*ACCTTTTCTTTGAATG -3´
CKA_*At*DIR6_for	PCR amplification of the *Atdir6* gene cloning into pET22b and pET32a	5´- TTTCGTAAAACCATTGATCAGAAAAAACCG -3´
CKA_*At*DIR6_HindIII_rev	PCR amplification of the *Atdir6* gene cloning into pET22b and pET32a	5´- CCC**AAGCTT**ATAGCATTCATACAGTTTAATATCC -3´
CKA_RBK1_NdeI_for	PCR amplification of gene for RBK1 cloning into pET16b and pJOE	5´- GGAATTC**CATATG**GGTATTACAGTAATAGGTTC -3´
CKA_RBK1_BamHI_rev	PCR amplification of gene for RBK1 cloning into pET16b	5´- CG**GGATCC**TATGCATCTTTCTG - 3´
CKA_RBK1_HindIII_rev	PCR amplification of gene for RBK1 cloning into pJOE	5´- CCC**AAGCTT**CTATGCATCTTTCTGAACATC -3´
CKA_pPICZ_Seq	Sequencing of pPICZαA constructs	5´- CTGTTCTAACCCCTAC-3´
CKA_*At*DIR6_EcoRI_for	Control of integration of the *Atdir6* gene in pPICZαA-*At*DIR6	5´- G**GAATTC**TTCAGAAAGACTATCGACCAG -3´
CKA_*At*DIR6_XbaI_rev	Control of integration of the *Atdir6* gene in pPICZαA-*At*DIR6	5´- GC**TCTAGA**GCGTAACACTCGTACAAC -3´
CKA_HiC_EcoRI_for	Control of integration of *Hic* gene in pPICZαA-HiC	5´- CG**GAATTC**GGCGCAATCGAAAAC -3´
CKA_HiC_KpnI_rev	Control of integration of *Hic* gene in pPICZαA-HiC	5´- GG**GGTACC**TATGCGCGGATG - 3´

Mutation sites in QuikChange® primers are shown in italic letters. Restriction sites are shown in bold letters.

2.2 Molecular- and microbiological methods

2.2.1 Agarose gel electrophoresis

Agarose gel electrophoresis was performed to separate DNA by size. Gels with an agarose concentration of 1% were run at 120 V for 35 min with 1x TAE buffer (40 mM Tris-acetate, 1 mM ethylenediaminetetraacetate (EDTA)) as running buffer. For loading the DNA samples, 5x loading dye (40% sucrose, 2 g L^{-1} Orange G) was used. To determine the size of the DNA bands, 2.5 µL of a 1 kB DNA ladder (Generuler™ 1 kb DNA ladder, Fermentas, Germany) was loaded as a standard. GelRed™ (Biotium, Hayward, CA, USA) was used in a 1x concentration directly in the agarose gel to detect DNA under UV light.

2.2.2 Plasmid DNA preparation

The Zippy™ Plasmid Miniprep kit (Zymo Research, CA, USA) was used according to the manufacturer´s manual to purify recombinant plasmid DNA from *E. coli*. 5 mL of an *E. coli* overnight culture in LB supplemented with the according antibiotics were used for plasmid preparations. Plasmids were eluted with 30 µL of deionized water and analyzed by agarose gel electrophoresis.

2.2.3 DNA restriction digests

All restriction enzymes and buffers were acquired from and used under recommendations of Thermo Scientific (St. Leon-Rot, Germany) and New England Biolabs (Ipswich, MA, USA). Agarose gel electrophoresis was used to determine the result of the restriction digests.

2.2.4 DNA gel extraction and PCR product purification

The QIAquick® gel extraction kit from Qiagen (Hilden, Germany) was used according to the manufacturer´s manual to extract plasmid DNA and DNA fragments from an agarose gel. DNA was eluted from the columns using deionized water.

The QIAquick® PCR purificiation kit from Qiagen (Hilden, Germany) was used according to the manufacturer´s manual to purify PCR products after restriction digests. DNA was eluted from the columns using deionized water.

2.2.5 DNA ligation

T4 DNA ligase obtained from Thermo Scientific (St. Leon-Rot, Germany) was used according to the manufacturer´s manual to ligate insert fragments into vectors cut with the corresponding restriction enzymes. The concentrations of the digested gene fragments and plasmids were measured using a NanoDrop 1000 (Thermo Fisher, MA, USA). For the ligation reactions, the insert to plasmid ratio was adjusted to a molar ratio of about three to one. The ligation reactions were carried out for about 1 h at room temperature.

2.2.6 Polymerase chain reaction

Polymerase chain reaction was used to either amplify the genes for *At*DIR6 (*Atdir6*) and RBK1 (*Rbk1*) from the original vectors with new restriction sites, to control the integration of the pPICZαA_*At*DIR6 construct into the *P. pastoris* genomic DNA or for QuikChange® mutagenesis to create the focused RBK1 library. The primers used are listed in Table 2-3.

Amplification of *Atdir6* from the original Geneart® (Regensburg, Germany) construct with primers CKA_*At*DIR6_for and CKA_*At*DIR6_HindIII_rev (see Table 2-3) for cloning into pET22b and pET32a vectors was performed according to the setup shown in Table 2-4. The PCR program used is shown in Table 2-5. The same setup was used for the amplification of *Rbk1* from pRN754 (DSM, Netherlands) with primers CKA_RBK1_NdeI_for and CKA_RBK1_BamHI_rev or CKA_RBK1_NdeI_for and CKA_RBK1_HindIII_rev for cloning into pET16b or pJOE, respectively (see Table 2-3).

Table 2-4 – PCR setup for the amplification of the *Atdir6* and *Rbk1* genes

VOLUME	COMPONENT	FINAL CONCENTRATION
5 µL	10x *Pfu* buffer with $MgSO_4$	1x
1 µL	dNTP mixture (10 mM)	0.2 mM
1 µL	Template DNA	0.1 ng μL^{-1}
10 µL	Forward primer (1 pmol μL^{-1})	0.1 pmol μL^{-1}
10 µL	Reverse primer (1 pmol μL^{-1})	0.1 pmol μL^{-1}
1 µL	*Pfu* DNA polymerase (2.5 U μL^{-1})	0.05 U μL^{-1}
up to 50 µL	ddH_2O	-

Table 2-5 – PCR program for the amplification of the *Atdir6* and *Rbk1* genes

CYCLES	TEMPERATURE	TIME	FUNCTION
1	95°C	2 min	Initial denaturation
35	95°C 55°C (*Atdir6*) or 64°C (*Rbk1*) 72°C	0.5 min 0.5 min 1.5 min (*Atdir6*) or 2.5 min (*Rbk1*)	Denaturation Annealing Extension
1	72°C	5 min	Final extension
	4°C	hold	

Integration of the *Atdir6* gene via the respective pPICZαA constructs into the genome of *P. pastoris* was verified by PCR using the following setup.

Table 2-6 – PCR setup for the integration control of the *Atdir*6 gene

VOLUME	COMPONENT	FINAL CONCENTRATION
5 µL	10× Taq Reaction Buffer without $MgCl_2$	1x
1 µL	dNTP mixture (10 mM)	0.2 mM
1 µL	Template DNA	-
10 µL	Forward primer (1 pmol $µL^{-1}$)	0.1 pmol $µL^{-1}$
10 µL	Reverse primer (1 pmol $µL^{-1}$)	0.1 pmol $µL^{-1}$
1 µL	*Taq* DNA polymerase (5 U $µL^{-1}$)	0.1 U $µL^{-1}$
3 µL	$MgCl_2$ (25 mM)	1.5 mM
up to 50 µL	ddH_2O	-

The PCR program used is shown in the following table.

Table 2-7 – PCR program for the integration control of the *Atdir*6 gene

CYCLES	TEMPERATURE	TIME	FUNCTION
1	95°C	5 min	Initial denaturation
30	95°C	1 min	Denaturation
	55°C	1 min	Annealing
	70°C	1 min	Extension
1	70°C	2 min	Final extension
	4°C	hold	

The PCR products were analyzed on an agarose gel.

The focused library of RBK1 with single, double, triple, and quadruple amino acid exchanges was created by employing the QuikChange® method (Stratagene, CA, USA). For single mutants, the entire wild-type plasmid pJOE_RBK1 was synthesized via PCR with primers containing the chosen mutation (see Table 2-3). Double, triple, and quadruple mutants were subsequently created using the according single, double, and triple mutants as template. The mutation is located in the middle of the primers, thus allowing annealing despite the mismatch at the mutated positions. Forward and reverse primers, each complementary to opposite strands of the vector, were designed as reverse complements to each other. Subsequent to the synthesis, the product was treated with DpnI, an endonuclease specific for methylated and hemimethylated DNA (target sequence 5´-Gm^{6}ATC-3´). DNA isolated from most *E. coli* strains including DH5α and XL-1 blue is dam methylated and therefore susceptible to *Dpn*I digestion. Hence, the wild-type plasmid is digested, while the PCR product remains stable.

Primers (see Table 2-3) were designed with regard to optimal melting temperature and *E. coli* codon usage. The QuikChange® PCR reaction for site-directed mutagenesis was performed according to the following setup.

Table 2-8 – QuikChange® PCR reaction setup

VOLUME	COMPONENT	FINAL CONCENTRATION
5 µL	10× *PfuUltra*™ II Reaction Buffer	1x
1 µL	dNTP mixture (10 mM)	0.2 mM
1 µL	Template DNA (50-100 ng)	1-2 ng µL^{-1}
5 µL	5′ forward primer (1 pmol µL^{-1})	0.1 pmol µL^{-1}
5 µL	3′ reverse primer (1 pmol µL^{-1})	0.1 pmol µL^{-1}
1 µL	*PfuUltra*™ II Fusion HS DNA Polymerase	-
up to 50 µL	ddH$_2$O	-

The PCR program used is shown in the following table.

Table 2-9 – QuikChange® PCR program

CYCLES	TEMPERATURE	TIME	FUNCTION
1	95°C	0.5 min	Hot start
18	95°C	0.5 min	Denaturation
	50°C	1 min	Annealing
	72°C	4 min	Extension
1	72°C	7 min	Final extension
	4°C	hold	

The PCR program was chosen after recommendations in the QuikChange® manual. Since *PfuUltra*™ II Fusion HS DNA Polymerase with an extension rate of about 2 kb min^{-1} was used, the extension step for the approximately 7.4 kb pJOE_RBK1 was set to 4 min.

After the QuikChange® PCR, 1 µL DpnI (5 U µL^{-1}) was added to each reaction, mixed, spun down briefly, and incubated for 2 h at 37°C. Subsequently, *E. coli* XL-1 blue cells (100 µL) were transformed with 10 µL of each reaction.

The complete QuikChange® protocol, transformation of *E. coli* cells, subsequent plasmid preparations, and sequence verifications were performed by Dr. Sandra Facey, Melanie Allgaier and Silke Bastian (ITB, University of Stuttgart).

2.2.7 Preparation of competent *Escherichia coli* cells

Chemically competent *E. coli* cells were prepared using standard methods [196]. Briefly, 50 mL of LB medium was inoculated with 0.5 mL of an *E. coli* overnight culture (5 mL of LB medium inoculated with the respective *E. coli* strain) and incubated at 37°C and 180 rpm. After the OD$_{600}$ reached approximately 0.5 the culture was chilled on ice for 15 min. Afterwards the cells were centrifuged for 10 min at 3,220 *g*. The pellet was resuspended in 20 mL TfBI buffer (30 mM potassium acetate, 100 mM rubidium chloride, 10 mM calcium chloride, 50 mM manganese chloride, 15% glycerol, adjusted to pH 5.8 with acetic acid and sterile filtered with 0.2 µm filters) and again centrifuged for 10 min at 3,220 *g*. The pellet was now resuspended in 2 mL TfBII buffer (10 mM 3-(N-morpholino)propanesulfonic acid (MOPS), 75 mM calcium chloride, 10 mM rubidium chloride, 15% glycerol adjusted to pH 6.5 with potassium hydroxide and sterile filtered with 0.2 µm filters) and incubated on ice for 15 min. Aliquots of 200 µL were immediately frozen in liquid nitrogen and stored at -80°C until use.

2.2.8 Heat-shock transformation of competent *Escherichia coli* cells

Transformation of competent *E. coli* cells was performed using standard methods [196]. Briefly, competent *E. coli* cells were thawed on ice. Plasmid (1 µL) for retransformation or ligated vector (10 µL) was added to 100 µL of competent cells. The mixtures were gently mixed and incubated for 30 min on ice. After a heat-shock for 45 s at 42°C, the cells were chilled for 1 min on ice. 500 µL of LB was added and the cells were incubated for 45 min at 37°C and 180 rpm. The cells were then centrifuged at 20,800 *g* for 1 min and plated onto LB plates with the appropriate antibiotics. The plates were incubated overnight at 37°C.

2.2.9 Transformation of competent *Pichia pastoris* cells by electroporation

Transformation of *P. pastoris* cells was performed according to the pPICZαA,B,C manual (Invitrogen, Carlsbad, CA USA). *P. pastoris* X-33 was grown as an overnight culture in 5 mL of YPD in a 50 mL conical tube at 30°C. About 0.1-0.5 mL of the overnight culture was used to inoculate 200 mL of YPD medium in a 1 L baffled flask. This culture was again grown overnight to an OD_{600} of 1.3-1.5. The cells were then centrifuged at 1500 *g* for 5 min at 4°C. The pellet was resuspended in 200 mL of ice-cold (0°C), sterile water. Centrifugation and resuspension were repeated three times. The cells were then resuspended once in 100 mL of ice-cold (0°C), once in sterile water, once in 8 mL of ice-cold (0°C) 1 M sorbitol and eventually in 0.4 mL of ice-cold (0°C) 1 M sorbitol. Since the cells are not exceedingly stable and cannot be stored at -80°C, they were kept on ice and used for transformation immediately.

80 µL of competent *P. pastoris* cells were mixed with linearized pPICZαA construct DNA (5-10 µg) and transferred into an ice-cold (0°C) electroporation cuvette (Biorad, Hercules, CA, USA). The cuvette was then incubated on ice for 5 min. Subsequently, the cells were pulsed with 1.5 kV, 200 ohm and 25 µF. Immediately 1 mL of ice-cold 1 M sorbitol was added to the cuvette to stabilize the cells. The suspension was transferred to 15 mL Falcon tubes and incubated at 30°C without shaking for 2 h. The cells were centrifuged and plated on YPDS

plates containing 100 µg mL^{-1} Zeocin™. The plates were incubated for up to 5 d at 30°C. Colonies were picked and streaked out for single colonies on fresh YPDS plates containing 100 µg mL^{-1} of Zeocin™.

2.2.10 Isolation of genomic DNA of *Pichia pastoris*

To confirm the transformation of *P. pastoris* X-33 with pPICZαA constructs, genomic DNA of *P. pastoris* was isolated. 10 mL of an overnight culture in YPD were centrifuged at 1500 *g* and 4°C for 3 min. The pellet was resuspended in 500 µL of PBS (140 mM sodium chloride, 10 mM potassium chloride, 6.4 mM disodium hydrogen phosphate, 2 mM potassium dihydrogen phosphate) containing 1 mg mL^{-1} proteinase K. After 20 min incubation at 65°C, 500 µL of glass beads (mean average of 0.75-1.00 mm) were added. The tubes were then shaken for 30 min at 4 °C in a swing mill (MM 2000, Retsch, Germany). Subsequently, the glass beads and cell debris were centrifuged for 5 min at 20,800 *g* and 4°C. The supernatant was transferred into a new reaction tube, resuspended in 150 µL of TE (100 mM Tris at pH 7.5, 10 mM EDTA) with 2% (w/v) SDS and incubated for 5 min at 65°C. Afterwards, 150 µL of 5 M potassium acetate at pH 6.2 was added and the solution was incubated for 40 min on ice. The DNA was precipitated twice with isopropanol. Then, 200 µL of 3 M sodium acetate at pH 5.2 and 1 mL of isopropanol were added, incubated for 10 min at -80°C and centrifuged for 10 min at 20,800 *g* and 4°C. The pellet was dried and resuspended in 90 µL of deionized water. For the second precipitation, 9 µL 3 M sodium acetate and 90 µL isopropanol were added. The DNA was again precipitated at –80°C for 10 min and centrifuged as described above. The pellet was washed in 1 mL of 70% (v/v) ethanol. After another round of centrifugation, the pellet was dried in a Concentrator 5301 (Eppendorf, Hamburg, Germany) for 15 min at 45°C under vacuum. The pellet was finally resuspended in 100 µL of TE with 400 µg mL^{-1} RNase A and incubated at 37°C for 30 min to degrade any contaminating RNA.

2.2.11 Sequencing of plasmid constructs

The sequences of all plasmid constructs were verified by DNA sequencing performed by GATC Biotech AG (Köln, Germany). Received sequences were checked with Multalin [197].

2.2.12 Storage of *Escherichia coli* and *Pichia pastoris* strains

For long-term storage, *E. coli* and *P. pastoris* strains were stored as glycerol stocks. For *E. coli*, a single colony of each strain was cultured overnight in 5 mL of LB and 750 µL of this culture was mixed with 750 µL of 60% sterile glycerol. For *P. pastoris*, a single colony of each strain was cultured overnight in 10 mL of YPD, harvested at 20,800 *g*, again resuspended in 750 µL of YPD and mixed with 750 µL of sterile 60% glycerol. Cells were stored at -80°C until use.

2.2.13 Cloning

2.2.13.1 *Humicola insolens* cutinase

The *P. pastoris* strain transformed with the pPICZαA_HiC construct containing the *Humicola insolens* cutinase gene was already available from a previous work [198]. The *hic* wild-type gene was cloned from pEF-Tu_HiC (kindly provided by Dr. Steffen Maurer, BASF) and subcloned into pPICZαA using restriction enzymes EcoRI and KpnI. *P. pastoris* X-33 was transformed as described for *At*DIR6 (see 2.2.9).

2.2.13.2 *Arabidopsis thaliana* dirigent protein

The gene for *At*DIR6 from *Arabidopsis thaliana* was obtained as a synthetic gene from GeneArt® (Regensburg, Germany) without the N-terminal secretion signal sequence and codon-optimized for *E. coli* and *P. pastoris*.

The *Atdir*6 gene optimized for expression in *P. pastoris* was subcloned into pPICZαA using restriction enzymes EcoRI and XbaI generating the expression plasmid pPICZαA_*At*DIR6. The open reading frame (ORF) was cloned without a stop codon allowing the fusion of a vector-encoded C-terminal His_6-tag sequence for protein purification. *P. pastoris* X-33 cells were transformed with SacI linearized pPICZαA_*At*DIR6 (see 2.2.9) and genomic integration was verified (see 2.2.10). Cells were plated on YPDS agar plates containing Zeocin™ (100 µg mL^{-1}) and incubated at 30°C for up to 5 d or until colonies were observed.

The *Atdir*6 gene optimized for expression in *E. coli* was subcloned into pET16b and pET28a using the restriction enzymes NdeI and BamHI generating the N-terminal His_{10}- (pET16b_*At*DIR6) or His_6-tag fusions (pET28a_*At*DIR6), respectively. For generation of the pJOE_*At*DIR6 construct, restriction enzymes NdeI and HindIII were used. For cloning into pET22b and pET32a, the *Atdir*6 gene was amplified by PCR from the original GeneArt® (Regensburg, Germany) vector (see 2.2.6 and Table 2-3). To generate N-terminal translation fusions with a *pel*B leader (pET22b) and thioredoxin (pET32a), respectively, NcoI was used as a restriction site. However, the *Atdir*6 gene already contained an NcoI restriction site. Therefore, a 5´- blunt end ligation was performed. The vectors were cut with NcoI and HindIII, while the insert from PCR amplification was only cut with HindIII. The 5´overhang in the cut vectors was filled in using the Klenow fragment according to the manufacturer´s recommendations (Fermentas, Germany). After purification of vector and insert, the genes were ligated into pET22b resulting in a C-terminal His_6-tag translational fusion with an N-terminal *pel*B leader sequence (pET22b_*At*DIR6) and an N-terminal translational fusion with thioredoxin-His_6 (pET32a_*At*DIR6). All sequences were verified by DNA sequencing (GATC Biotech AG, Köln, Germany) (see 2.2.11). Chemically competent *E. coli* strains were transformed using standard methods (see 2.2.8). Chaperone expression plasmids pG-Tf2, pTf16 and pG-KJE8 (Takara, Japan) were co-transformed with pET16b_*At*DIR6.

2.2.13.3 *Saccharomyces cerevisiae* ribokinase

The gene for RBK1 from *Saccharomyces cerevisiae* was obtained as the construct pRN754 (DSM, Netherlands). The gene was amplified from this plasmid using primers harboring NdeI and HindIII restriction sites. After restriction digest and purification of vectors and insert, the insert was ligated into pET16b and pJOE, respectively. RBK1 variants of the focused library were prepared with the QuikChange® protocol as described in 2.2.6. All sequences were verified by DNA sequencing (GATC Biotech AG, Köln, Germany) (see 2.2.11). Chemically competent *E. coli* strains were transformed using standard methods (see 2.2.8).

2.3 Protein expression and purification

2.3.1 Expression of *At*DIR6 with *Escherichia coli*

Expression of *At*DIR6 in *E. coli* was performed in 50 mL of TB medium containing the appropriate antibiotics (e.g. 100 µg mL^{-1} ampicillin or 30 µg mL^{-1} kanamycin) in 500 mL Erlenmeyer flasks without baffles. Cultures were inoculated with overnight cultures of *E. coli* transformants to an OD_{600} of 0.05 and grown at temperatures ranging from 15°C to 37°C using a rotary shaker at 180 rpm. After reaching an OD_{600} of 0.4-0.6, the cultures were induced with 0.1 mM to 1 mM isopropyl-β-D-thiogalactopyranosid (IPTG) for the pET series vector constructs or 0.5% (w/v) L-rhamnose for pJOE_*At*DIR6, respectively. Expression of chaperones was induced for 4-24 h by the addition of the appropriate inducer (1 mg mL^{-1} L-arabinose and/or 2 ng mL^{-1} tetracyclin) directly after inoculation of the main culture. For maintaining chaperone plasmids, 20 µg mL^{-1} chloramphenicol was added to the medium. For purification by fast protein liquid chromatography (FPLC), the *E. coli* cultures were centrifuged at 4°C with 11,000 *g* for 30 min and the pellet was resuspended in 100 mM potassium phosphate buffer at pH 6.0 containing 150 mM KCl. After addition of 0.2 mM of the protease inhibitor phenylmethanesulfonylfluoride (PMSF) the cells were disrupted by ultrasonification. The cell debris was centrifuged at 4°C with 34,400 *g* for 1 h and the lysate was filtered with 0.2 µm sterile filters before the purification procedure.

2.3.2 Expression of HiC and *At*DIR6 with *Pichia pastoris*

For the expression of HiC and *At*DIR6 in shaking flasks, selected transformants were inoculated in 10 mL of BMGY medium in 50 mL falcon tubes and cultivated at 30°C and 180 rpm overnight. The precultures were used to inoculate 200 mL of BMMY medium to an OD_{600} of 0.5. Subsequently, the cultures expressing HiC and AtDIR6 were grown at 30°C and 140 rpm in 1 L Erlenmeyer flasks with baffles for 3 d and 6 d, respectively. Methanol (100%) was added daily to a final concentration of 0.5% (v/v) to maintain induction. Cells were harvested by centrifugation at 4°C with 1,800 *g* for 15 min. The supernatants were concentrated and buffer exchanged to 100 mM potassium phosphate buffer at pH 6 containing 150 mM potassium chloride at 4°C with a PALL Minimate™ tangential flow filtration capsule (PALL, Port Washington, NY, USA) with a 5 kDa nominal molecular mass limit. Further concentration was achieved by ultrafiltration using an Amicon stirred cell model 8050 (Merck Millipore, Schwalbach, Germany) with a 5 kDa nominal molecular mass limit membrane.

2.3.3 Fed-batch fermentation of *Pichia pastoris* expressing *At*DIR6

For fed-batch fermentation, a 50 mL falcon tube with 10 mL of BMGY medium was inoculated with a single colony of strain *P. pastoris* pPICZαA_AtDIR6 and incubated at 30°C at 180 rpm. After 24 h of growth, this preculture was used to inoculate 200 mL of BMGY medium in a 2 L Erlenmeyer flask with baffles and the cells were grown at 30°C and 180 rpm for 48 h. Following incubation, the cells were harvested by centrifugation (22°C, 1200 *g*, 5 min), washed with 200 mL of 100 mM potassium phosphate buffer at pH 6.0, centrifuged again, and finally resuspended in 20 mL of buffer. Fermentation was performed in a 7.5 L reactor (Infors, Bottmingen, Switzerland) containing 4 L basal salt medium (9.1 g L^{-1} K_2SO_4, 7.5 L^{-1} $MgSO_4{\cdot}7\ H_2O$, 4.2 g L^{-1} KOH, 0.47 g L^{-1} $CaSO_4{\cdot}2\ H_2O$, 8 mL L^{-1} H_3PO_4 (85%), 50 g L^{-1} glycerol (87%), adjusted to pH 5 with NH_4OH (28%)). Additionally, 0.87 mg L^{-1} filter sterilized D-biotin, 4.35 mg L^{-1} PTM_1 trace salt solution (6 g L^{-1} $CuSO_4{\cdot}5\ H_2O$, 0.08 g L^{-1} NaI, 3 g L^{-1} $MnSO_4{\cdot}H_2O$, 0.5 g L^{-1} $CoCl_2$, 20 g L^{-1} $ZnCl_2$, 0.02 g L^{-1} H_3BO_3, 0.2 g L^{-1} $Na_2MoO_4{\cdot}2\ H_2O$, 65 g L^{-1} $FeSO_4{\cdot}7\ H_2O$, 0.2 g L^{-1} D-biotin, 30 mL L^{-1} H_2SO_4 (6 N)) were added to the medium. The bioreactor was inoculated with the washed cells to an OD_{600} of 0.5 and the growth temperature was kept constant at 30°C for 24 h and subsequently shifted to 22°C. The pH was maintained at 5.0 using NH_4OH (28%) and H_3PO_4 (10%). The airflow was maintained at 8 L min^{-1} and the stirrer speed set to 800 rpm. After depletion of the initial 5% (v/v) glycerol, which was indicated by an abrupt increase in dissolved oxygen (DO), 20 mL of 0.6% (v/v) PTM_1 in methanol was automatically added. During the fermentation process, a threshold of 30% DO was used for automatic addition of methanol with a 2 h hold-off period before the next addition. Methanol consumption was monitored using a balance that was interfaced with the IRIS process control system (Infors, Bottmingen, Switzerland). The addition of Antifoam 204 (Sigma, Germany) was regulated by an antifoam electrode. Samples for OD_{600}, cell wet weight (cww) and cell dry weight (cdw) measurements were taken daily. For OD_{600}, samples were diluted appropriately in 0.9% (w/v) NaCl solution. For CWW, 10 mL samples were centrifuged (4°C, 1800 *g*, 5 min) and the pellet was washed once with 10 mL of 0.9% (w/v) NaCl solution. The sample was centrifuged again and the pellet was weighed. For CDW, 10 mL samples were centrifuged (4°C, 1800 *g*, 5 min) and the pellet was washed once with 10 mL of 0.9% (w/v) NaCl solution. The sample was centrifuged again and the pellet was incubated at 100°C till mass constancy was achieved and subsequently weighed. The fermentation with cells expressing *At*DIR6 was performed twice and both fermentations proceeded comparably and were ended after 13 d. The culture broth was harvested and centrifuged (4°C, 1800 *g*, 15 min). The dirigent protein containing supernatant was filtrated by cross-flow filtration (Merck Millipore, Schwalbach, Germany) using a set-up according to the manufacturer´s instructions with a silicone membrane with a 10 kDa cut-off value (PALL, Port Washington, NY, USA). During this step, the buffer was exchanged to a 100 mM potassium phosphate buffer at pH 6 containing 300 mM KCl.

2.3.4 Expression of RBK1 with *Escherichia coli*

Expression of RBK1 in *E. coli* was performed in 200 mL of TB medium containing 30 µg mL^{-1} kanamycin in 1 L Erlenmeyer flasks without baffles. For maintaining the pRARE plasmid in *E. coli* Rosetta(DE3), 20 µg mL^{-1} chloramphenicol was added to the medium. Cultures were inoculated with overnight cultures of *E. coli* transformants to an OD_{600} of 0.05 and grown at 20°C, 30°C or 37°C using a rotary shaker at 180 rpm. After reaching an OD_{600} of 0.4-0.6, the cultures were induced with 0.1 mM to 0.5 mM IPTG (pET16b_RBK1) or 0.5% (w/v) L-rhamnose (pJOE_RBK1), respectively. All RBK1 variants were expressed using pJOE vector constructs in *E. coli* C43(DE3) for 24 h at 30°C. For purification by FPLC, the *E. coli* cultures were centrifuged at 4°C with 11,000 *g* for 30 min and the pellet was resuspended in 100 mM potassium phosphate buffer at pH 7.5 containing 100 mM KCl. After addition of 0.2 mM of the protease inhibitor PMSF the cells were disrupted by using a homogenizer (Emulsiflex C5, Avestin, Mannheim, Germany). The cell debris was centrifuged at 4°C with 34,400 *g* for 1 h and the lysate was filtered via a 0.2 µm sterile filter before the purification procedure.

2.3.5 Protein purification of *At*DIR6

The dirigent protein *At*DIR6 was purified from concentrated fermentation supernatants and *E. coli* lysates using immobilized metal affinity chromatography (IMAC) on an Äkta Explorer FPLC system (GE Healthcare, Little Chalfont, United Kingdom). A Hitrap Chelating HP 5 mL column (GE Healthcare, Little Chalfont, United Kingdom) charged with Ni^{2+} was equilibrated with five column volumes (CV) of 100 mM potassium phosphate buffer at pH 6.0 containing 300 mM KCl (buffer A). Samples of 30 mL were loaded onto the column at a linear flow rate of 3 mL min^{-1}. Afterwards, the column was washed with four CV of buffer A containing 20 mM imidazole. *At*DIR6 was eluted from the column with a step gradient using four CV of buffer A containing 300 mM imidazole at a flow rate of 5 mL min^{-1}. To remove residual protein, the column was washed with two CV of buffer A containing 500 mM imidazole. Protein-containing fractions from the elution step were detected by absorption measurements at 280 nm, pooled and desalted with PD10 columns (GE Healthcare, Little Chalfont, United Kingdom) using buffer A according to manufacturer´s instructions. Subsequently, samples were concentrated using Amicon Ultra 4 centrifugal filter units with a nominal molecular mass limit of 10 kDa (Merck Millipore, Schwalbach, Germany) to protein concentrations ranging from 5-8 mg mL^{-1}.

2.3.6 Protein purification of RBK1

The ribokinase RBK1 and all variants from the focused library were purified from *E. coli* lysates using IMAC on an Äkta Explorer FPLC system. A Hitrap Chelating HP 5 mL column charged with Ni^{2+} was equilibrated with three CV 100 mM potassium phosphate buffer at pH 7.5 containing 100 mM KCl (buffer B). Samples of 10 mL were loaded onto the column at a linear flow rate of 3 mL min^{-1}. Afterwards, the column was washed with three CV of buffer B containing 15 mM imidazole. RBK1 and its variants were eluted from the column with a step gradient using four CV of buffer B containing 200 mM imidazole at a flow rate of 5 mL min^{-1}. To remove residual protein, the column was washed with two CV of buffer B containing 500 mM imidazole. Protein-containing fractions from the elution step were detected by absorption measurements at 280 nm, pooled and dialyzed against buffer B at 4°C overnight using SpectraPor dialysis tubings with a nominal molecular mass limit of 6-8 kDa (Spectrum Labs, Rancho Dominguez, CA, USA).

2.3.7 Protein quantification

Protein concentration was determined using the Bradford [199] or BCA assay [200] in a micro titer plate format. The absorption at 595 nm or 562 nm, respectively, was measured using a Spectramax 340PC micro titer plate reader (Molecular Devices, Sunnyvale, CA, USA). Each sample was measured in quadruplicates from individual dilutions and the standard deviation was always below 2%. Bovine serum albumin was used to establish standard curves.

2.3.8 SDS-PAGE analysis

Samples for SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) with cell extracts or purified proteins were mixed with 6x sample buffer (200 mM Tris pH 6.8, 600 mM dithiothreitol, 30% glycerol, 10% sodium dodecyl sulfate, 0.04% bromphenol blue) in a 1 to 5 ratio and boiled at 95°C for 5-10 min. For SDS-PAGE analysis of proteins from growing cultures aliquots with 5x 10^8 cells (OD_{600} of 1 corresponding to 1 x 10^9 cells per mL) were centrifuged (14,000 *g*, 1 min). Cell pellets were resuspended in 50 µl of 2x sample buffer and boiled at 95°C for 5-10 min. 10 µl of these samples were applied to SDS-PAGE. Gels were run in a Hoefer SE 250 electrophoresis chamber (Hoefer, San Francisco, CA, USA) with SDS-PAGE running buffer (20 mM Tris pH 8.3, 190 mM glycine, 0.1% SDS) and stained with Coomassie Brilliant Blue R250 (2.5 g L^{-1} Coomassie Brilliant Blue R250, 30% methanol, 10% acetic acid).

Subsequently, gels were destained (30% methanol, 10% acetic acid) and documented on a Quantum ST 4 system (Vilber Lourmat, Eberhardzell, Germany).

Densitometric analysis of SDS gels was performed using the ImageJ 1.45 software [201]. Background was substracted with a rolling ball filter (50 pixels). Relative amounts of purified protein and impurities as well as glycosylated and deglycosylated *At*DIR6 species were calculated from band intensities obtained by peak integration.

2.3.9 Western blot analysis

For Western blot analysis a Trans-blot SD semi-dry transfer cell (Bio-Rad, Munich, Germany) was used for protein transfer onto nitrocellulose membranes (Pall, Port Washington, NY, USA). Rabbit polyclonal IgG directed toward *At*DIR6 (Eurogentec, Seraing, Belgium) and mouse monoclonal IgG directed toward His_6-tag (GE Healthcare, Little Chalfont, United Kingdom) were used as primary antibodies. Goat anti-mouse IgG-horseradish peroxidase (HRP) conjugate (Sigma, Germany) and goat anti-rabbit IgG-HRP conjugate (Sigma, Germany) were used as secondary antibodies for detection. All antibodies were used as 1:10,000 dilutions in 10 mL of Tris-buffered saline with Tween® 20 (TBST, 25 mM Tris·Cl at pH 7.4, 150 mM NaCl, 0.3 mM KCl, 0.1% Tween® 20) containing 5% non-fat milk powder. For visualization, blots were developed using the Amersham ECL Western blot analysis kit (GE Healthcare (Amersham Biosciences), Little Chalfont, United Kingdom) and analyzed using a Fusion SL image acquisition system (Vilber Lourmat, Eberhardzell, Germany).

2.4 Biochemical characterization of *At*DIR6

2.4.1 Deglycosylation of *At*DIR6

Recombinant *At*DIR6 from *P. pastoris* fermentations was enzymatically deglycosylated with PNGase F (New England Biolabs, Ipswich, MA, USA) according to the manufacturer´s protocol. To demonstrate that *At*DIR6 is post-translationally glycosylated, 20 µg denatured *At*DIR6 was incubated with 500 U of PNGase F at 37°C for 24 h and analyzed via SDS-PAGE. For further deglycosylation, additional 1,000 U of PNGase F were added to the samples, incubated for another 48 h, and analyzed via SDS-PAGE. For evaluation of dirigent activity after deglycosylation, 1 mg native *At*DIR6 was incubated with 2,500 U PNGase F at 30°C for up to 24 h. Samples were taken after 0 h, 0.5 h, 1 h, 2 h, 4 h and 24 h of incubation and subjected to SDS-PAGE analysis (5 µg) and dirigent activity assays (51 µg, see below). For solubility tests, 1 mg native *At*DIR6 was incubated with 2,500 U PNGase F at 30°C for up to 96 h. Samples were taken after 24 h, 48 h, 72 h and 96 h and centrifuged for 15 min at 20,800 *g* and 4°C. Soluble and insoluble fractions were resuspended in appropriate amounts of sample buffer and subjected to SDS-PAGE analysis.

2.4.2 Circular dichroism

Circular dichroism (CD) was measured with the help of Dr. Uwe Gerken (University of Hohenheim, Stuttgart) with a Jasco J-715 spectropolarimeter (Jasco, Tokyo, Japan) using quartz cuvettes (Hellma AG, Müllheim, Germany). Protein samples of 1 mg mL^{-1} *At*DIR6 and deglycosylated *At*DIR6 were dialyzed against 5 mM potassium phosphate buffer (pH 6.0) containing 5 mM potassium chloride at 4°C. Samples were diluted to concentrations of 0.07, 0.04, and 0.02 mg mL^{-1} for measuring CD spectra or titration with guanidinium hydrochloride (GndCl), respectively.

CD spectra were recorded in the range of 190-260 nm with a scanning speed of 20 nm min^{-1} and a step width of 1 nm using a quartz cuvette of 1 mm path length. A response time of 4 s and an excitation bandwidth of 2 nm were chosen to improve signal-to-noise ratio. Each spectrum was measured in triplicate and background corrected by subtracting a buffer blank spectrum. When necessary, the spectra were smoothed with the Savitzky-Golay filter. In order to compare spectra from different samples, all spectra were normalized to mean residue weight ellipticity (θ_{MRW}) [deg cm^2 $dmol^{-1}$] with

$$\theta_{MRW}(\lambda) = \theta_{mdeg}(\lambda)\cdot MW\cdot(n\cdot c\cdot d\cdot 10)^{-1}$$

where $\theta_{mdeg}(\lambda)$ is the recorded ellipticity in millidegrees for a given wavelength, MW the molecular weight of the protein in Da, d the path length of the cuvette in centimeters, n the number of amino acid residues and c the sample concentration in mg mL^{-1}. Calculations of the secondary structure were performed with the DICHROWEB server (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) [202,203] using a combination of the SP175short dataset [204] with the CDSSTR algorithm [205].

For chemical denaturation, *At*DIR6 and deglycosylated *At*DIR6 was titrated by adding small aliquots of 8 M GndCl and denaturation was followed at constant wavelength of 220 nm using a quartz cuvette of 5 mm path length (Hellma AG, Müllheim, Germany). Samples were incubated with GndCl aliquots under constant stirring for 5 min at 20°C prior to measuring the ellipticity. The response time and excitation bandwith of the spectropolarimeter were set to the same parameters as mentioned above.

All measurements were done in triplicate and the ellipticity values were corrected for dilution.

2.4.3 Mass spectrometric analysis and glycan structure

Glycosylated and deglycosylated *At*DIR6 were subjected to SDS-PAGE analysis and the corresponding bands were cut out of the gel. Further steps were performed by Dr. Jens Pfannstiel (Proteomics Core Facility, Life Science Center, University of Hohenheim, Stuttgart) and are briefly summarized.

Proteins were in-gel-digested using either trypsin or chymotrypsin (Roche, Penzberg, Germany) according to Shevchenko et al. [206]. For chymotryptic digests a modified incubation buffer was used (100 mM Tris-HCl, pH 7.6, 10 mM $CaCl_2$). After digestion, the gel pieces were extracted with a mixture of 50% of acetonitrile and 50% of 0.1% formic acid (v/v) for 15 min. The supernatant was collected and the gel pieces were covered with 5% formic acid for 15 min before the same volume of acetonitrile was added. After incubation for 10 min the supernatant was collected. The pooled supernatants were then dried in a vacuum centrifuge and stored at -20°C. Dried samples were dissolved in 0.1% formic acid.

Characterization of intact *At*DIR6 and identification of *At*DIR6 glycosylation sites by mass spectrometry was performed as described by Cedzich *et al.* [207].

2.4.4 Dirigent activity assay

For assay of dirigent activity, recombinant *At*DIR6 was used in the oxidative coupling of coniferyl alcohol catalyzed by *Trametes versicolor* laccase (Sigma, Germany) with molecular oxygen. The standard reaction condition consisted of 4 mM coniferyl alcohol, 5.7 µM laccase, and 4.2 µM *At*DIR6 in 100 mM potassium phosphate buffer at pH 6 containing 300 mM KCl. The reactions were incubated at 30°C with constant shaking at 500 rpm for 15 min. Subsequently, substrate and products were extracted three times with 250 µl of ethyl acetate. The solvent was evaporated at 40°C for 1 h using a GeneVac EZ 2 plus evaporator (GeneVac, Gardiner, NY, USA). The residue was resuspended in 50 µl of 50% acetonitrile containing 0.01% formic acid and analyzed by high-performance liquid chromatography (HPLC) (see below). In further experimental setups, concentration of coniferyl alcohol, *At*DIR6 or laccase were varied in a range from 1-7 mM, 0-53 µM, and 0.1-23 µM, respectively. Dirigent activity assays for each sample were done in triplicates.

For comparison with the plant derived *At*DIR6, molar concentrations of *At*DIR6 from *P. pastoris* fermentations were calculated based on the molecular mass of the active homodimer, which slightly differs between plant- and *Pichia*-derived *At*DIR6 due to the C-terminal His-tag. The molecular mass of the *P. pastoris* derived, His-tagged *At*DIR6 and the untagged, native *At*DIR6 homodimer as derived from *S. peruvianum* is 41.5 kDa and 36.1 kDa, respectively. Although Pickel (2010) used the molecular mass of the glycosylated *At*DIR6 homodimer expressed in *S. peruvianum* (40.8 kDa) for calculation of molar concentrations in the dirigent activity assay, here the molecular mass of the unglycoslated forms was used for the calculations to be able to compare *At*DIR6 derived from both systems since the degree of glycosylation and therefore contribution to the molecular mass differs.

2.4.5 HPLC analytics

Reaction products from dirigent assays were first subjected to RP-HPLC in order to fractionate the pinoresinol peak. The analysis was carried out on an Agilent 1200 series HPLC system (Agilent, Santa Clara, CA, USA) equipped with an Agilent HPLC Zorbax Eclipse XDB-C18 column (150x4.6 mm, 5 µm particle size, Agilent, USA). A gradient of 10-40% acetonitrile containing 0.1% formic acid with a flow rate of 1 ml min^{-1} was used for elution (10-27% from 0-20 min, 40% from 20-26 min, 10% from 26-35 min). A diode array detector (DAD; 1260 infinity diode array detector; Agilent, Santa Clara, CA, USA) was used for detection at 280 nm and 200-320 nm. The pinoresinol peak was fractionated and the solvent evaporated at 40°C for 1.5 h in a GeneVac EZ 2 plus evaporator (GenVac, Ipswich, United Kingdom). The residue was resuspended in 50 µl of chiral solvent (ethanol and hexane in a 1:1 ratio, 0.1% formic acid). The enantiomeric composition of pinoresinol was analyzed by chiral HPLC using a Daicel Chiralpak IB column (150x4.6 mm, 5 µm particle size, Daicel Chiral Technologies, Illkirch, France) run isocratically in chiral solvent at 0.7 ml min^{-1} on an Agilent 1200 series HPLC system with DAD. The enantiomeric excess of (-)-pinoresinol was calculated by integration of the peak areas of both enantiomers. Pure (+)-pinoresinol (ArboNova, Turku, Finland) was used as standard for RP-HPLC and chiral HPLC.

2.4.6 Analytical gel filtration

Native molecular mass (MM) of *At*DIR6 was determined by analytical gel filtration on a XK-16/100 Sephacryl S-200 HR column (GE Healthcare, Little Chalfont, United Kingdom) with an Äkta Explorer FPLC system (GE Healthcare, Little Chalfont, United Kingdom). The column (total bed volume V_t of 183 mL) was packed according to the manufacturer´s recommendations (maximum pressure 0.28 MPa), washed with 3 CV of deionized water and equilibrated with 3 CV of buffer C (100 mM potassium phosphate buffer pH 6, 100 mM NaCl) at a flow rate of 1.0 mL min^{-1}. An *At*DIR6 sample was buffer exchanged to buffer C using Amicon Ultra 4 centrifugal filter units with a nominal molecular mass limit of 10 kDa (Merck Millipore, Schwalbach, Germany). Three samples with a volume of purified *At*DIR6 of 0.5 mL each and a concentration of about 10 mg mL^{-1} were applied in analytical gel filtration. The column void volume V_0 was determined using blue dextran as the excluded solute. A gel filtration standard (Bio-Rad gel filtration standard kit, Bio-Rad, Munich, Germany) was run three times through the column and a calibration curve was generated by semi-logarithmic plotting of molecular mass (MM) versus the gel phase distribution coefficient (K_{av}). The K_{av} value of *At*DIR6 was used to calculate its molecular mass using the established calibration curve. The K_{av} value was calculated according to the following formula, where V_e is the elution volume of the protein, V_0 the column void volume and V_t the total bed volume [208]:

$$K_{av} = \frac{V_e - V_0}{V_t - V_0}$$

2.5 Biochemical characterization and structure elucidation of RBK1

2.5.1 Kinase activity assay

Kinase activity was determined using a coupled NADH assay monitoring the oxidation of NADH to NAD$^+$ at 340 nm (ε = 6,220 M^{-1} cm^{-1}) (modified from [142]). In this assay, the ATP consumed in the kinase reaction is regenerated by the phoshphorylation of ADP to ATP catalyzed by pyruvate kinase using phosphoenolpyruvate. The generated pyruvate is then reduced to lactate by lactate dehydrogenase under oxidation of NADH (see Figure 2-1).

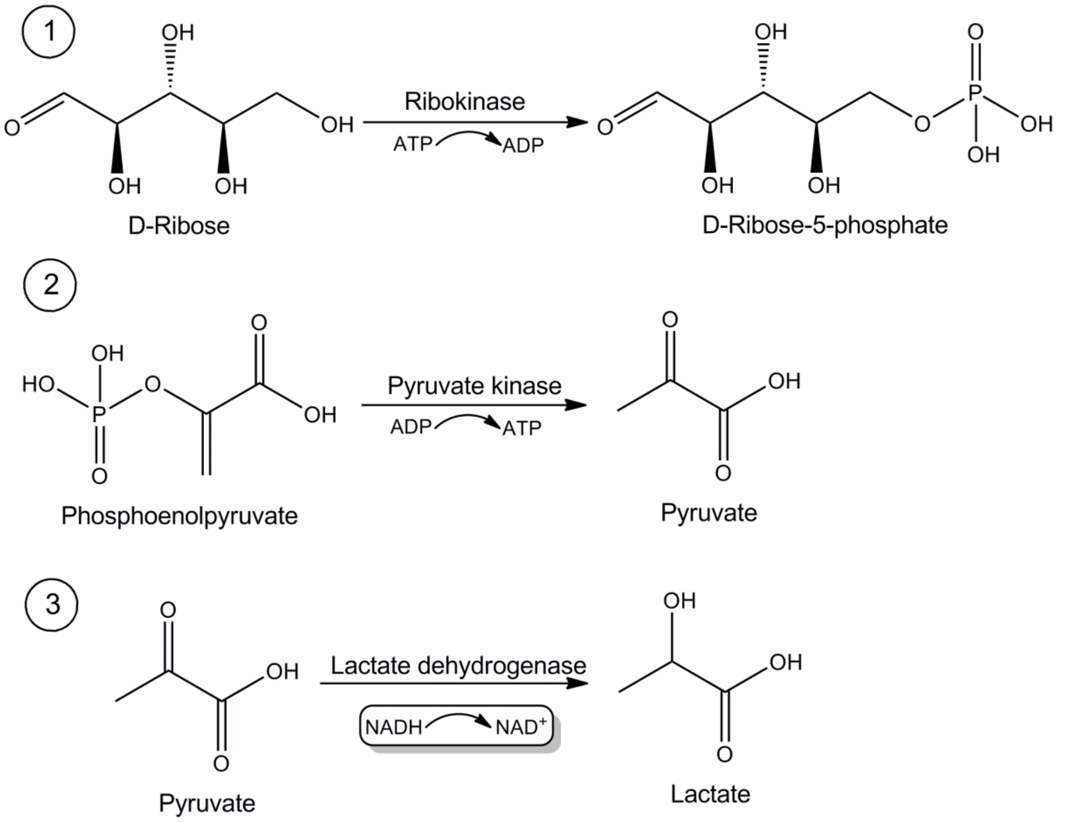

Figure 2-1 – Kinase activity assay. In the first step, ribokinase phosphorylates D-ribose to D-ribose-5-phosphate consuming the co-substrate ATP. In the coupled enzyme reactions, pyruvate kinase phosphorylates the generated ADP from the first step consuming phosphoenolpyruvate and yielding pyruvate. The pyruvate is then reduced to lactate in the last step of the coupled NADH assay by lactate dehydrogenase. The NADH oxidation in the last step can be measured spectrophotometrically at 340 nm.

One unit of enzymatic activity was defined as the amount of free, soluble ribokinase that phosphorylated 1 µmol sugar substrate in 1 min at room temperature. Since the coupling to the NADH consumption is in a stoichiometric ratio, the reaction rate of the kinase corresponds to the reaction rate of the NADH oxidation in the coupled NADH assay. The basic setup of the assay included 10 mM sugar substrate, 1 mM ATP, 0.6 mM phosphoenolpyruvate, 0.5 mM NADH, 10 µl pyruvate kinase/lactate dehydrogenase mixture (Sigma Germany, 600-1,000 U and 9,000-14,000 U, respectively) and 30 nM to 5 µM RBK1 in a 1 mL reaction scale unless otherwise stated. The reaction was monitored spectrophotometrically in cuvettes using an Ultrospec 3100 pro spectrophotometer (GE Healthcare (Amersham Biosciences), Little Chalfont, United Kingdom).

The reaction was optimized for the wild-type RBK1 in respect to magnesium, potassium and phosphate concentrations as well as pH value using different buffer compositions.

A temperature profile of the RBK1 activity at increased temperatures was recorded. Therefore, the enzyme was incubated at the respective temperature for 5 min in the reaction mixture prior to starting the reaction by substrate addition. All measurements at varying temperatures were performed in 100 mM potassium phosphate buffer at pH 6.0 including 2.5 mM magnesium chloride.

For analysis of enzyme stability, aliquots of wild-type RBK1 (5.5 mg mL^{-1}) were incubated at 22°C and 30°C for up to 24 h and at -20°C and 4°C for up to 8 weeks. Aliquots were tested at several time points and the enzyme was appropriately diluted prior to determination of activity versus D-ribose. All measurements for stability analysis were performed in 75 mM triethanolamine buffer at pH 7.5 including 5 mM potassium phosphate and 5 mM magnesium chloride.

RBK1 variants from the focused library were initially screened for increased specific activity with 10 mM D-ribose and 50 mM D-xylose as substrates in 75 mM triethanolamine buffer at pH 7.5 including 5 mM potassium phosphate and 5 mM magnesium chloride. Selected variants were analyzed for their substrate specificity towards D-glucose, D-fructose, D-arabinose, L-arabinose, D-lyxose, D-erythrose, D-threose (50 mM each) and 2-deoxy-D-ribose (10 mM) in 100 mM potassium phosphate buffer at pH 6 including 2.5 mM magnesium chloride.

For kinetic data, activity of wild-type RBK1 and variants towards D-ribose were measured at concentrations from 0.1 µM to 20 mM D-ribose. Furthermore, kinetic data with varying concentrations of the co-substrate ATP (0.05 µM to 5 mM), D-xylose (1 mM to 500 mM), D-glucose (0.01 mM to 100 mM) and 2-deoxy-D-ribose (1 mM to 500 mM) were collected. For the K_m determination of ATP, the concentration of D-ribose was held constant at 10 mM D-ribose. For the K_m determination of D-ribose, 2-deoxy-D-ribose, D-xylose and D-glucose, the concentration of ATP was held constant at 1 mM. All measurements for kinetic data were performed in 100 mM potassium phosphate buffer at pH 6.0 including 2.5 mM magnesium chloride. Apparent kinetic constants K_m and k_{cat} were calculated using Origin Pro 8.5 from the non-linear regression of the Michaelis-Menten plots.

For 2-deoxy-D-ribose, D-xylose and D-glucose, the data was fitted to the equation:

$$v_0 = \frac{v_{max}[S]}{(K_m + [S])}$$

The Michaelis-Menten model basically describes an enzyme-catalyzed reaction with the conversion of a single substrate into a product. However, it can also describe the rate of a multi-substrate reaction as long as the dependency of only one substrate is studied and the other substrate is kept constant at a saturating concentration [209]. Here, v_0 is the initial velocity, v_{max} the maximal velocity for infinite substrate concentrations, K_m the Michaelis-Menten constant and [S] the substrate concentration.

For ATP and D-ribose, the data was fitted to the following equation:

$$v_0 = \frac{v_{max}\ [S]}{K_m + \left(1 + \frac{[S]}{K_{iu}}\right)[S]}$$

This equation describes the rate of a reaction with substrate inhibition at higher substrate concentrations due to a second substrate molecule acting as an uncompetitive inhibitor binding to the enzyme-substrate complex. Here, K_{iu} is the inhibitory constant for the uncompetitive competition.

2.5.2 Biotransformations with RBK1

RBK1 wild-type and variants were employed in biotransformations with D-ribose, D-xylose and D-glucose. The basic setup was in a 0.5 mL scale including 5 mM sugar substrate, 2 mM ATP and 30 nM to 3 µM enzyme in 100 mM potassium phosphate buffer at pH 6.0 with 2.5 mM magnesium chloride. The samples were incubated at 30°C for 10 min to 60 min at 750 rpm in a Thermomixer comfort (Eppendorf, Hamburg, Germany). Aliquots of 100 µL were taken after addition of the enzyme and after 10 min to 60 min. The reactions were stopped by addition of 6 µL sulfuric acid (1 M) and spun down briefly at 20,800 *g*. The samples (100 µL) were transferred into glass vials and analyzed by HPLC.

The dependency of RBK1 activity on mono- and divalent cations was analyzed in biotransformations with D-ribose. For monovalent cation dependency, biotransformations were performed in 100 mM Tris buffer at pH 6.0 with 2.5 mM magnesium chloride. Monovalent cations tested were potassium, sodium, ammonium, cesium, lithium, rubidium (each as the respective chloride salt) at a concentration of 100 mM. Additionally, the reaction was performed without monovalent cations. Here, 1 µM RBK1 wild-type, which was purified by size exclusion chromatography (see 2.5.4) and buffer exchanged to 10 mM Tris at pH 7.5 was used in these reactions. Samples were taken after 0 min and 45 min and processed as described above.

For divalent cation dependency, biotransformations with ribose were performed in 100 mM potassium phosphate buffer at pH 6.0. Divalent cations tested were magnesium, manganese and calcium (each as the respective chloride salt) at a concentration of 2.5 mM. Additionally, the reaction was performed without divalent cations. In all cases, 30 nM RBK1 wild-type from IMAC purification was used. Samples were taken after 0 min and 10 min and processed as described above.

Further biotransformations with D-ribose were performed with varying concentrations of the pentavalent anion phosphate in 100 mM Tris pH 6.0 with 2.5 mM magnesium chloride. Here, potassium phosphate was used and the potassium concentration was always 100 mM by addition of potassium chloride. Samples were taken after 0 min and 10 min and processed as described above.

2.5.3 HPLC and LC/MS analytics

The analysis of biotransformations with RBK1 was carried out on an Agilent 1200 series HPLC system (Agilent, Santa Clara, CA, USA) equipped with an Aminex HPX-87H column (300 x 7.8 mm, 9 µm particle size, Bio-Rad, Munich, Germany). Eluation was achieved with 5 mM sulfuric acid as solvent run in an isocratic mode at 0.5 or 0.7 mL min^{-1} flow rate for 25 or 15 min, respectively, at 60°C. The injection volume was 10 µL. A refractive index detector (RID; 1260 infinity refractive index detector; Agilent, Santa Clara, CA, USA) was used for detection of substrates and products. Additionally, a DAD (1260 infinity diode array detector; Agilent, Santa Clara, CA, USA) was used for detection of ATP and ADP at 254 nm. All substrates and products were validated by their retention time in comparison to standard substances. The conversion of substrates and generation of products was calculated by integration of the peak areas from detection with RID and DAD in comparison to standard curves. Standards were measured and standard curves were generated for D-ribose, D-xylose and D-glucose (2.6 mM to 5.4 mM, refractive index detection) as well as ATP and ADP (0.4 mM to 2.4 mM, diode array detection). All standard curves showed excellent linearity with r^2 values of minimum 0.99.

The reaction products D-ribose-5-phosphate, D-xylose-5-phosphate, and D-glucose-6-phosphate were qualitatively validated by analysis of the biotransformations by mass spectrometry using an LC/MS system (Agilent 1260 Infinity series with a 6130 quadrupole MS system). Biotransformations for LC/MS analysis were performed in the MS compatible buffer ammonium acetate at pH 5.8 with 2.5 mM magnesium chloride and 20 mM potassium chloride. Fractionation was achieved using an Aminex HPX-87H column with 0.1% formic acid as solvent at a flow rate of 0.7 mL min^{-1} for 15 min at 60°C. The injection volume was 10 µL. Besides detection of ATP and ADP at 254 nm by DAD, a mass spectrometer (MSD) was used as detector of reaction products. The following ESI (electrospray ionization) parameters were used: The drying gas temperature was set to 350°C, the drying gas flow to 10.0 L min^{-1}, the nebulizer pressure to 35 psig, the capillary voltage to +/- 3000 V and the fragmentor voltage to 70 eV. Mass spectrometric data was acquired in positive and negative scan mode from *m*/*z* 100 to 300. Additionally, negative SIM mode was used to aquire data of the reaction products D-ribose-5-phosphate (229 *m*/*z*), D-xylose-5-phosphate (229 *m*/*z*) and D-glucose-6-phosphate (259 *m*/*z*). The retention times and mass spectra of the substrates and reaction products were compared to standard substances.

2.5.4 Size exclusion chromatography

Further purification of IMAC purified RBK1 for crystallization and native molecular mass determination was performed by size exclusion chromatography (SEC) on a XK-16/100 Sephacryl S-200 HR column (GE Healthcare, Little Chalfont, United Kingdom) with an Äkta Explorer FPLC system (GE Healthcare, Little Chalfont, United Kingdom). The column (total bed volume V_t of 183 mL) was packed according to the manufacturer´s recommendations (maximum pressure 0.28 MPa) and validated using an aceton standard. Therefore, the column was washed with 3 CV of deionized water. An acetone sample (200 µL, 2%) was injected and the column was run at a flow rate of 1.0 mL/min. For a measure of column efficiency, the number of theoretical plates (N) was calculated with the following formula, where V_e is the peak elution volume (in mL), $W_{1/2}$ the peakwidth at half peak height (in mL) and L the total bed height (in mm) [208]:

$$N = 5.54 \left(\frac{V_e}{W_{1/2}}\right)^2 \frac{1{,}000}{L}$$

Additionally, the peak symmetry A_s was calculated where a is the first half peak width at 10% peak height and b the second half peak width at 10% peak height:

$$A_s = \frac{b}{a}$$

For the Sephacryl S-200 HR resin applied for this column these values should be N > 9,000 and A_s = 0.80 - 1.50.

For purification, the column was equilibrated with 3 CV of buffer D (20 mM Tris pH 7.5, 150 mM sodium chloride) at a flow rate of 1.0 mL min^{-1}. An RBK1 sample was buffer exchanged to buffer D using Amicon Ultra 4 centrifugal filter units with a nominal molecular mass limit of 10 kDa (Merck Millipore, Schwalbach, Germany). Three samples with a volume of purified RBK1 of 2 mL each and a concentration of about 8.5 mg mL^{-1} were applied in gel filtration chromatography. Protein containing fractions detected with UV at 280 nm were collected.

Furthermore, the native molecular mass of RBK1 was determined by analytical gel filtration using the same column. The K_{av} value of RBK1 was compared to the values of standard proteins as described for *At*DIR6 in 2.4.6.

2.5.5 Crystallization of RBK1

For crystallization screening, RBK1 samples from SEC purification were buffer exchanged to 10mM Tris buffer at pH 7.5 and concentrated to 10 mg mL^{-1} and 20 mg mL^{-1} using Amicon Ultra 4 centrifugal filter units with a nominal molecular mass limit of 10 kDa (Merck Millipore, Schwalbach, Germany). Samples were then filtrated using centrifugal filter units with a pore size of 0.22 µm (Merck Millipore, Schwalbach, Germany) and stored for shipment at 4°C. An initial crystallization screening was performed by Dr. Jürgen Kopp at the Protein crystallization platform of the "Biochemie-Zentrum der Universität Heidelberg" (University of Heidelberg, Germany). Prior to crystallization, RBK1 was mixed with 10 mM D-ribose and 10 mM AMP-PNP as ligands. A liquid handling system was used to dispense sitting drops of 384 non redundant conditions and plates were incubated at 18°C. Several conditions yielded protein crystals suitable for diffraction studies. Additionally, the "Optimizer and Additive" screen was performed to obtain crystals of higher quality. Based on the results from the initial screening, a home-made fine screen was performed at the University of Hohenheim, Stuttgart, with PEG3350 concentrations varying between 18 to 23% containing 200 mM or 400 mM potassium citrate, potassium fluoride, or potassium sulfate, respectively. RBK1 was incubated using the sitting drop method at a final concentration of 3.3 mg mL^{-1} with 5 mM AMP-PNP and 25 mM D-ribose or D-xylose at 18°C. Additionally, the protein was incubated in the apo form without substrates.

For data collection, crystals were mounted on cryoloops and flash-frozen in liquid nitrogen. Native datasets were recorded at 100 K using a rotation angle of 0.15° per image on a Pilatus 6M detector at beamline ID29 of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France (Dr. Daniele de Sanctis, ESRF, and Dr. Domenico Lupo, University of Hohenheim).

2.6 Biochemical characterization and biotechnological use of *Humicola insolens* cutinase

2.6.1 Cutinase activity assay

Cutinase activity was determined using *para*-nitrophenyl (*p*NP) esters of butyrate, octanoate and palmitate. Activity was measured by absorbance increase at 410 nm (ε = 13,200 M^{-1} cm^{-1}) due to free *para*-nitrophenol using an Ultrospec 3100 pro spectrophotometer (GE Healthcare (Amersham Biosciences), Little Chalfont, United Kingdom). The assay consisted of 800 µL Tris buffer (100 mM Tris pH 7.5 with 0.5% Triton X-100 and 0.1% gum arabic), 100 µL substrate (2.5 mM *p*NP ester in 100% DMSO) and 100 µL cutinase solution. One unit of enzymatic activity was defined as the amount of free, soluble cutinase that releases 1 µmol of *para*-nitrophenol in 1 min at room temperature. For kinetic data, activity of HiC towards *p*NP esters was measured at concentrations of 10 µM to 10 mM for *p*NPP and of 10 µM to 2 mM for *p*NPB and *p*NPO, respectively. Apparent kinetic constants K_M and k_{cat} were calculated using Origin 8.0 from the non-linear regression of the Michaelis-Menten plots. Standard deviations were calculated from at least three individual measurements and were below 6.2%.

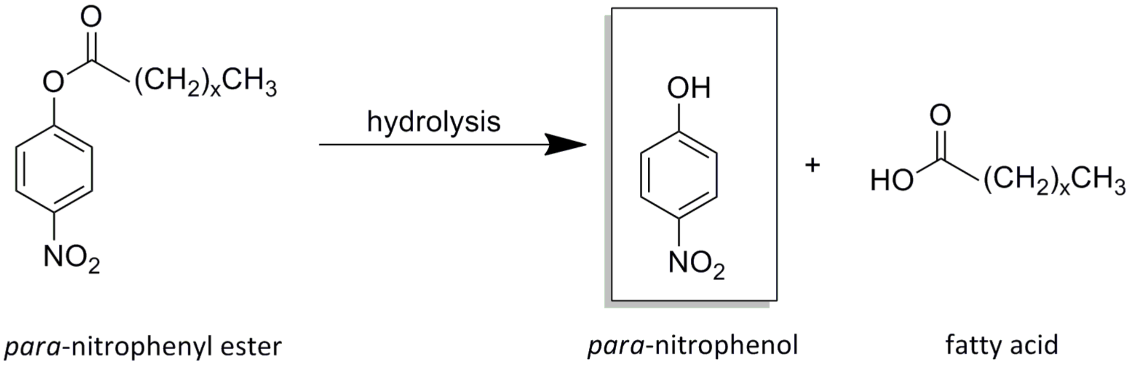

Figure 2-2 – Cutinase activity measurement with the *para*-nitrophenol assay. Hydrolysis of *para*-nitrophenyl ester yields fatty acid and free *para*-nitrophenol leading to an absorbance increase at 410 nm.

2.6.2 Immobilization of HiC

HiC was immobilized onto microporous resin Diaion HP20 (average particle size 440 µm, 30 nm pore radius) with an enzyme loading of 1% (w/w) or 4.5 µmol HiC. The resin was degreased using ethanol (100%) and dried under vacuum for 24 h at room temperature prior to use. Three mg of HiC were diluted in immobilization buffer (50 mM potassium phosphate buffer pH 7.8 containing 25% ethanol) to a final concentration of 0.25 mg mL^{-1}. Afterwards, 300 mg Diaion HP20 was added and the solution was incubated at 4°C for 24 h under constant agitation. Subsequently, the immobilisate was recovered by filtration, washed twice with 15 mL of potassium phosphate buffer (50 mM, pH 7.8) and dried in a desiccator over silica gel for 72 h at room temperature to achieve a low water activity a_w towards 0. The immobilisate was stored at 4°C.

2.6.3 Transesterification reactions with HiC

In the standard reaction setup methyl acrylate (300 µL, 10.3 M) as donor of the acrylic group and sole solvent, 6-mercapto-1-hexanol (20 µL, 275 mM) and 10 mg immobilisate (1% (w/w) enzyme loading; 4.5 µmol HiC) were applied. Reactions were carried out in screw-capped glass vials under constant shaking (700 rpm) at a temperature of 40°C and a water concentration of 0.025% (w/w) for up to 24 h. The influence of the radical inhibitor butylated hydroxytoluol (BHT) was tested by adding BHT (3-15% (w/w) acrylic ester, 0.14-0.71 M) to the standard reaction. For the substrate specificity in transesterifications, methyl propionate was additionally used as acyl donor.

2.6.4 GC/MS analysis

Samples (50 µL) from transesterifications were centrifuged for 1 min at 20,800 *g*, diluted in 1 mL diethyl ether dried over $MgSO_4$ (anhydrous) and analyzed by GC/MS measurements. GC/MS analysis was carried out on a Shimadzu GCMS-QP2010 (equipped with a FS-Supreme-5 column length 30 m, internal diameter 0.25 mm) using helium as carrier gas. The following GC program was used: 70°C for 1 min, 15°C min^{-1} to 280°C, 30°C min^{-1} to 310°C. Standard curves for substrates were generated and conversion of mercaptoalcohols was calculated from the peak area ratios of mercaptoalcohols and products. Transesterification products were identified by their characteristic MS fragmentation patterns.

2.6.5 Product purification and identification by NMR and FT-IR

For the identification of the main product (6-mercaptohexyl acrylic ester) by nuclear magnetic resonance (NMR) and FT-IR (Fourier transform infrared) spectroscopy an upscaled reaction was performed. Methylacrylate (3 mL, 10.3 M) was incubated with 200 µL of 6-mercapto-1-hexanol (275 mM), 50 mg HiC immobilisate and 0.025% (w/w) water at 40°C for 24 h using a reflux condenser. 6-mercaptohexyl acrylic ester was purified using silica column chromatography with a mobile phase of hexane and diethyl ether in a 10 to 4 ratio. To determine the product containing fractions, TLC was performed with the same mobile phase and potassium permanganate solution was used to visualize the compounds. The solvent from product containing fractions was evaporated using a Rotavapor R-200 evaporator (Büchi, Essen, Germany) and the substance was dissolved in deuterated chloroform containing 1% (v/v) tetramethylsilane (TMS) for NMR measurements. The structure was confirmed by ^1H (500 mHz) and ^{13}C (125 MHz) NMR measurements and FT-IR The NMR measurements were performed by the service unit of the Institute of Organic Chemistry (University of Stuttgart) and the spectra were interpreted with the help of Dr. Holger Beuttler (Institute of Technical Biochemistry, University of Stuttgart).

3. Results

3.1 Optimized expression of the dirigent protein *At*DIR6 in *Pichia pastoris* and impact of glycosylation on protein structure and function

Recently, the expression and initial characterization of the dirigent protein (DIR) *At*DIR6 from *Arabidopsis thaliana* was reported [109]. In the presence of laccase and molecular oxygen, this protein was shown to exhibit a novel dirigent activity, mediating the formation of (-)-pinoresinol from coniferyl alcohol. Although a hypothesis based on detailed biochemical studies of *Fi*DIR1 was proposed, the exact reaction mechanism is still elusive [95,104,111,112]. Up to now, all characterized DIRs were either extracted from plant material or expressed in systems based on higher organisms such as plant and insect cell cultures, yielding only up to 200 µg per L of culture [95,105,108,109]. Initial trials with *E. coli* as expression host resulted only in insoluble expression and the formation of inclusion bodies [210].

The aim of this case study was the heterologous expression of the dirigent protein *At*DIR6 in biotechnologically established expression hosts like *Escherichia coli* (*E. coli*) or *Pichia pastoris* (*P. pastoris*) to obtain high yields of the protein for further characterization or possible biotechnological use. Furthermore, the biochemical characterization especially with respect to the dirigent activity of *At*DIR6 and the impact of glycosylation on protein structure, solubility and function was investigated.

3.1.1 Heterologous expression of *At*DIR6 with *Escherichia coli*

In a recent study, it has been shown that *At*DIR6 could not be expressed in soluble form using a pET21a construct under the control of the IPTG inducible T7 promoter in *E. coli* BL21(DE3) or Rosetta-gami B(DE3) [210]. Neither lowering expression temperature from 30°C to 4°C nor refolding of purified, insoluble *At*DIR6 yielded soluble protein. It was hypothesized that the glycosylation might influence the soluble expression and possibly also the aggregation propensity of the protein. A more systematic evaluation regarding *E. coli* as a heterologous expression system was necessary to exclude *E. coli* as a viable alternative for the expression of dirigent proteins. Therefore, several strategies for a soluble expression in *E. coli* were tested.

Since the synthetic *Atdir6* gene was optimized for *E. coli* in terms of codon usage, *E. coli* strains like Rosetta (Merck, Darmstadt, Germany), which can substitute for rare tRNAs [211], were not required. Initial expressions were performed using the standard *E. coli* BL21(DE3) expression strain as well as the BL21(DE3) derivatives C41(DE3) and C43(DE3) [195] at a temperature of 37°C. In these initial experiments, the formation of a band at about 19 kDa was observed during SDS-PAGE analysis (see Figure 3-1) and identified as *At*DIR6 by Western blot analysis using the primary *At*DIR6 or His_6-tag antibodies (data not shown). *At*DIR6 was therefore readily expressed in *E. coli* BL21(DE3) (see Figure 3-1a) as well as C41(DE3) and C43(DE3) (data not shown) strains either as N-terminal His_{10} or C-terminal His_6 fusion protein using the pET16b or pET28a constructs, respectively. However, in all attempts, *At*DIR6 could only be found in the insoluble fraction after cell lysis.

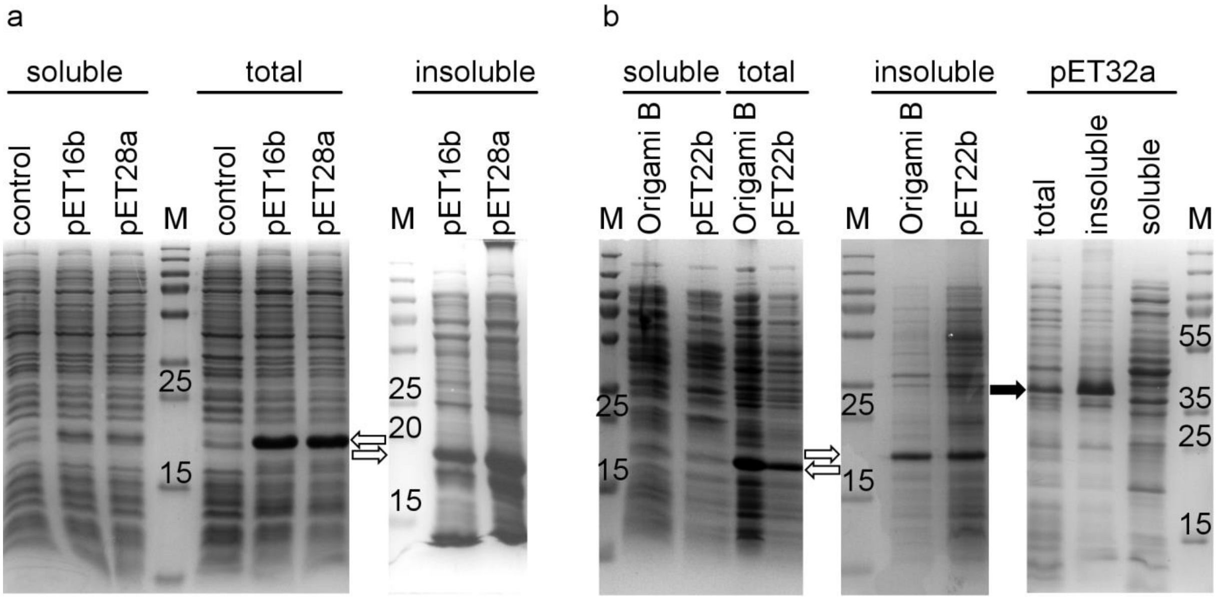
a
soluble
total
insoluble
control
pET16b
pET28a
M
control
pET16b
pET28a
M
pET16b
pET28a
25
15
25
20
15
b
soluble
total
insoluble
pET32a
M
Origami B
pET22b
Origami B
pET22b
M
Origami B
pET22b
total
insoluble
soluble
M
25
15
25
15
55
35
25
15

Figure 3-1 – Comparison of the expression of *At*DIR6 in *E. coli* BL21(DE3) and Origami B(DE3) hosts by SDS-PAGE analysis. (a) Expression of *At*DIR6 with *E. coli* BL21(DE3) at 37°C. Control: pET16b vector control without *Atdir*6 gene; pET16b: expression with pET16b_*At*DIR6 construct; pET28a: expression with pET28a_*At*DIR6 construct. The *white arrows* indicate the protein band at about 19 kDa identified as *At*DIR6 by Western blot analysis (data not shown). **(b)** Expression of *At*DIR6 with *E. coli* BL21(DE3) and Origami B(DE3). Origami B: expression with *E. coli* Origami B(DE3) pET16b_*At*DIR6; pET22b: expression with *E. coli* BL21(DE3) pET22b_*At*DIR6; pET32a: expression with *E. coli* BL21(DE3) pET32a_*At*DIR6. The *white arrows* indicate the protein identified as *At*DIR6 by Western blot analysis (data not shown). The *black arrow* indicates the thioredoxin-*At*DIR6 fusion protein at about 38 kDa. **Lane abbreviations:** Soluble: soluble protein fraction after cell lysis; total: total protein fraction from expression culture; insoluble: insoluble (pellet) fraction after cell lysis. 15 µg of total protein was loaded in each lane. *Lanes M* represent molecular size markers. Molecular size is indicated in kDa in *lanes M*.

Variation of expression parameters such as inducer concentration (0.01 mM to 1 mM IPTG) or temperature (15°C to 37°C) did not result in increased solubility of *At*DIR6. To rule out problems with the strong T7 promoter used in the pET series vectors, *At*DIR6 was expressed with *E. coli* BL21(DE3) using a rhamnose-inducible pJOE construct under the control of the rhamnose promoter [212–214]. However, no detectable change in expression strength or solubility was found (data not shown).

As already described in the introduction, a homology model of *At*DIR6 based on the crystal structure of an allene oxide cyclase was recently established indicating a possible disulfide bond connecting the N- and C terminus, which was then verified by mass-spectrometric data [110]. To facilitate the formation of this disulfide bond during expression with *E. coli*, *At*DIR6 was expressed as pET22b construct in BL21(DE3) or as pET16b construct in Origami B(DE3). While the pET22b construct allowed targeting of the protein into the periplasm in a Sec-dependent fashion via the PelB leader to provide an oxidative environment for disulfide bond formation, Origami B(DE3) already provides an oxidative cytoplasm due to various mutations. Again, *At*DIR6 was strongly expressed, but not found in the soluble fraction (see Figure 3-1b).

Another possibility to enhance soluble protein production is to fuse the target protein to other, highly soluble proteins like maltose binding protein or thioredoxin (Trx) [42–44]. Using a pET32a construct, *At*DIR6 was expressed as C-terminal fusion to Trx-His$_6$ in *E. coli* BL21(DE3) and Origami B(DE3). A band at about 38 kDa corresponding to the expected size of the fusion protein was observed. Despite the high solubility of Trx alone, the Trx-*At*DIR6 fusion was found in the insoluble fraction (see Figure 3-1b). Western blot analysis verified that not even small amounts of the fusion protein were obtained as soluble protein (data not shown).

To explore the effect of increased chaperone activity in *E. coli* on the expression and correct folding of *At*DIR6, some of the commercially available Takara chaperone plasmids (Clontech, USA) were used (see Figure S 6, Figure S 7 and Figure S 8). Directly after inoculation, *E. coli* Origami B(DE3) cultures were induced with the appropriate inducer for co-expression of chaperones DnaK-DnaJ-GrpE and/or GroES-GroEL, GroES-GroEL-TF or TF alone from pG-KJE8, pG-Tf2 or pTf16, respectively. Chaperone expression, especially for GroEL (about 60 kDa) and TF (about 48 kDa), was visible from SDS-PAGE analysis. Depending on the expression parameters, *At*DIR6 expression was weak and often only visible in Western blot analysis (data not shown). Figure 3-2 shows the results of the SDS-PAGE and Western blot analysis from the co-expression of DnaK-DnaJ-GrpE from pG-KJE8 and *At*DIR6 from the pET16b construct. Here, the co-expression of chaperones did result in soluble *At*DIR6 for the first time as verified by Western blot analysis, albeit only in very low amounts not clearly visible in the SDS-PAGE analysis (see Figure 3-2a). In order to gain substantial amounts of *At*DIR6 expressed in *E. coli* for further biochemical characterization, a larger cultivation, expression, and purification procedure was performed. *E. coli* Origami B(DE3) cultures were inoculated to an OD_{600} of 0.05, induced with 0.5 mM IPTG at an OD_{600} of about 0.5 after 2 d, and harvested at an OD_{600} of about 3 after 6 d. Cells were lysed and directly purified with Ni-NTA chromatography. Lysate from 1 L cell culture yielded approximately 40 µg *At*DIR6 after Ni-NTA chromatography. In addition to these low amounts of soluble form, the protein was still only partially pure as seen in Figure 3-2b.

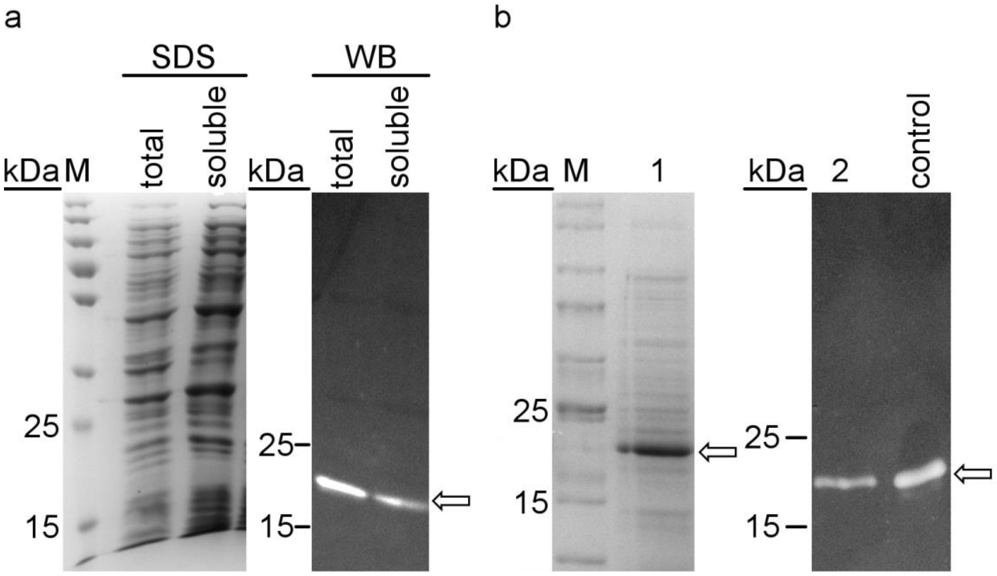

Figure 3-2 – Expression of *At*DIR6 in *E. coli* Origami B(DE3) with the co-expression of chaperones DnaK, DnaJ and GrpE and purification of *At*DIR6. (a) Expression of *At*DIR6 was analyzed by SDS-PAGE and Western blotting. *Lanes total* and *soluble* represent the total protein fraction from expression cultures and the soluble protein fraction after cell lysis, respectively. Fifteen micrograms of total protein was loaded in each lane. **(b)** Purification of *At*DIR6 by Ni-NTA chromatography. For SDS-PAGE analysis (*lane 1*), 5 µg of purified *At*DIR6 was used. For Western blot analysis (*lane 2*), 0.2 µg of purified *At*DIR6 was used. For the c*ontrol,* 5 µg of total protein with insoluble *At*DIR6 from a previous expression was used. **Lane abbreviations:** *Lanes M* represent a size marker. Protein bands are marked with their respective molecular weight in kilodaltons (kDa). The *white arrows* indicate the protein band at about 19 kDa identified as *At*DIR6 by Western blot analysis.

3.1.2 High yield fed-batch fermentation of *At*DIR6 with *Pichia pastoris*

Due to the problems with the expression in *E. coli*, *P. pastoris* was investigated as an alternative expression host. Dirigent proteins were found to be *N*-glycosylated *in planta* [105] as well as in insect [108] and plant cell expression hosts [109,110] at two *N*-glycosylation sites. Thus, the most obvious advantage with *P. pastoris* is the ability for glycosylation of recombinant proteins, although this glycosylation might vary strongly from the original pattern in the natural organism [36].

Initial experiments were performed in Erlenmeyer flasks at 30°C and 180 rpm up to 6 d. The gene of *At*DIR6 was fused to the alpha-mating factor pre-pro leader sequence (α-MF) from *Saccharomyces cerevisiae* in the pPICZαA construct. This allows the secretion of the dirigent protein into the medium. After concentration of the supernatant, several bands could be identified by SDS-PAGE analysis (Figure 3-3a). Surprisingly, none of these bands was detected at the size of unglycosylated *At*DIR6 of about 21 kDa when expressed from the pPICZαA_AtDIR6 construct in *P. pastoris*. However, the band at about 33 kDa was identified as *At*DIR6 by Western blot analysis using anti-*At*DIR6 primary antibodies (Figure 3-3a). Additionally, an indistinct signal at about 15 kDa was visible.

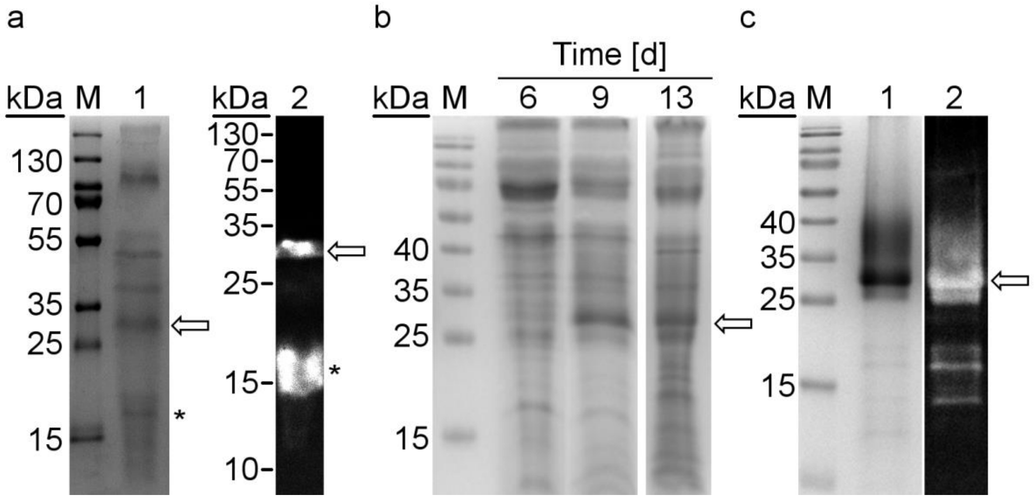

Figure 3-3 – SDS-PAGE and Western blot analysis of the *At*DIR6 expression, fermentation and purification (modified from [215]**). (a)** SDS-PAGE (1) and Western blot analysis (2) of purified *At*DIR6. For SDS-PAGE analysis (*lane 1*), 15 µg of concentrated supernatant was used. For Western blot analysis (*lane 2*), 1 µg of concentrated supernatant was used. The asterisks indicate the indistinct signal in the lower molecular weight range. **(b)** SDS-PAGE of fermentation supernatants from *P. pastoris* cells expressing *At*DIR6 after 6, 9 and 13 days. 15 µg of total protein was loaded in each lane. **(c)** SDS-PAGE (1) and Western blot analysis (2) of purified *At*DIR6. For SDS-PAGE analysis (*lane 1*), 10 µg of purified *At*DIR6 was used. For Western blot analysis (*lane 2*), 0.5 µg of purified *At*DIR6 was used. **Lane abbreviations:** *Lanes M* represent a size marker. Protein bands are marked with their respective molecular weight in kilodaltons (kDa). The *white arrows* indicate the protein band identified as *At*DIR6 by Western blot analysis.

For higher yields of *At*DIR6, high cell density fed-batch fermentation of the *At*DIR6 expressing *P. pastoris* strain was performed as duplicates. To accumulate cell mass before induction of protein expression, the cells grew at 30°C on the initial 5% (v/v) glycerol for the first 28 h to an OD_{600} over 100 during the batch phase. Then the process was switched to fed-batch with the addition of methanol, which was performed at 22°C to increase the level of dissolved oxygen as this is the main restraint of the growth of *P. pastoris* at high cell densities. Both fermentations proceeded comparably and the process is shown for the second fermentation in Figure 3-4. Before harvest, the cells reached a final OD_{600} of approximately 500 and cell wet weight as well as cell dry weight increased up to 414 g L^{-1} and 115 g L^{-1}, respectively, and were therefore in good correlation with the OD_{600} [215].

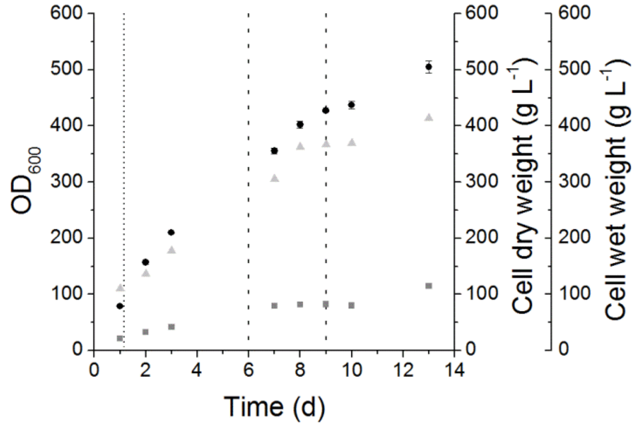

Figure 3-4 – Fed-batch fermentation process of *At*DIR6 with *Pichia pastoris* (adapted from [215]**).** Samples were taken at various time points and analyzed for the OD_{600} (*black circles*), cell dry weight (*grey squares*), and cell wet weight (*light grey triangles*). The *dotted line* represents the shift from the batch to the fed-batch process. The *dashed lines* represent sampling of 0.5 L fermentation broth for protein expression analysis and addition of new media. Standard deviations of OD_{600} were calculated from three individual samples.

For evaluation of the increase in *At*DIR6 yield during the fermentation process, 500 mL samples were taken after 6 d and 9 d. The fermenter was replenished with 500 mL fresh medium at these points. After 13 d, the fed-batch fermentation processes were finished with a final volume of 5.1 L and 4.9 L. After centrifugation, the volume of the supernatants amounted to 3.9 and 3.4 L, respectively. Since *At*DIR6 was secreted into the supernatant, a concentration via a cross-flow filtration unit with a 10kDa molecular weight limit membrane was performed as the first downstream processing step. During this, the buffer was exchanged to a 100 mM potassium phosphate buffer at pH 6 containing 300 mM KCl. Subsequently, the protein content of the retentate was analyzed by SDS-PAGE. Comparably to the shaking flask experiment, a distinct band was observed around 33 kDa (white arrow; Figure 3-3b) after 9 d and 13 d of fermentation. The protein was purified by immobilized metal affinity chromatography (IMAC) and verified as *At*DIR6 by Western blot analysis with primary anti-*At*DIR6 antibodies (Figure 3-3c).

Besides the major band around 33 kDa, the fermentation resulted also in impurities in the lower molecular range as seen before in the expression in shaking flasks. Additionally, a smear between 35 and 40 kDa was visible on both the SDS gel and the Western blot. Based on the quantification of overall band intensities by densitometric analysis of the SDS gels of purified *At*DIR6 samples, a purity of 90% and yield of 47 mg L^{-1} were calculated for the 33 kDa protein from the fermentation process [215].

3.1.3 Characterization of the glycan structure and enzymatic deglycosylation

Since the apparent molecular mass estimated from SDS-PAGE analysis (about 33 kDa) was higher than the calculated size for the *P. pastoris* derived *At*DIR6 (21 kDa), it was investigated whether this is due to glycosylation of *At*DIR6 in *P. pastoris*. In a first attempt, a sample of *At*DIR6 was analyzed with MALDI-TOF-MS analysis (Dr. Jens Pfannstiel, Proteomics Core Facility, Life Science Center, University of Hohenheim) and the molecular mass of *At*DIR6 was determined as 24.6 kDa, which is significantly lower than the 33 kDa estimated from the SDS-PAGE analysis, but still higher than the calculated size (see Figure 3-5) [215].

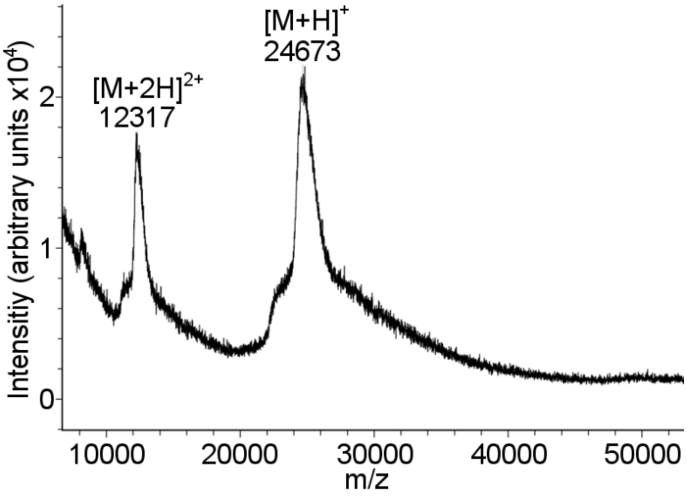

Figure 3-5 – MALDI-TOF mass spectrum of *At*DIR6 isolated from *Pichia pastoris* (modified from [215]**, supplementary information).** Approximately 10 pmol of the protein were applied to a stainless steel target after C_4-ZipTip purification. The mass spectrum was recorded in the linear detection mode with external calibration. Singly and doubly charged molecule ions are indicated. The average mass observed for the singly charged molecule ion (24.6 kDa) is considerably larger than the calculated average mass for the singly charged non-glycosylated *At*DIR6 protein.

In this regard, the possibility of glycosylation of *At*DIR6 was further investigated by using a deglycosylation approach. For this, *At*DIR6 was incubated with PNGase F under denaturing conditions to enzymatically remove possible *N*-glycans. A shift in the molecular mass of the protein indicated the presence of an *N*-linked glycan structure (Figure 3-6). Two distinct bands at about 25 kDa and 21 kDa were formed after 24 h of incubation with sizes comparable to glycosylated *At*DIR6 derived from cell cultures of *S. peruvianum*. During further incubation with additional PNGaseF for another 48 h (72 h total incubation) no significant changes in the protein pattern was observed, indicating the formation of two discrete protein species, which likely differed in the glycosylation pattern [215].

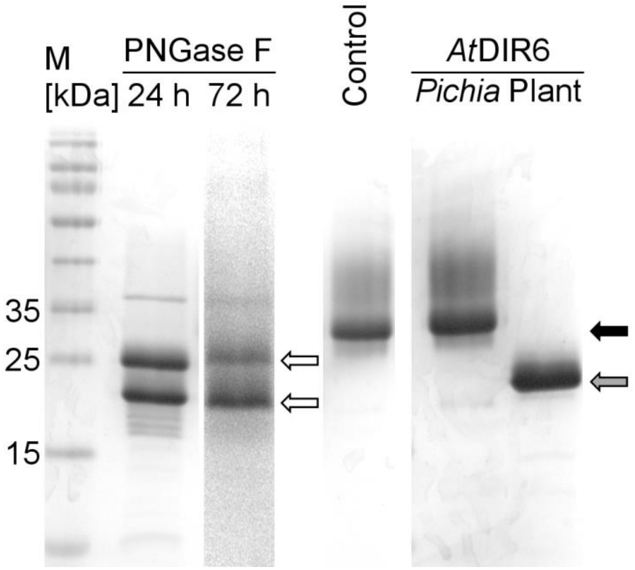

Figure 3-6 – SDS-PAGE analysis of *At*DIR6 samples after deglycosylation (adapted from [215]**).** PNGase F 24 h: 24 h deglycosylation with 500 U PNGase F; PNGase F 72 h: 72 h deglycosylation with additional 1000 U PNGase F added after 24 h of incubation. The bands at 36 kDa represent PNGase F; Control: Negative control under the same buffer conditions but lacking PNGase F; *At*DIR6 *Pichia*: *At*DIR6 produced by *P. pastoris* cells; *At*DIR6 Plant: *At*DIR6 produced by cell cultures of *Solanum periuvanum*. The *black arrow* indicates native *At*DIR6 derived from *P. pastoris* fermentations (33 kDa), the *grey arrow* indicates glycosylated *At*DIR6 derived from cell cultures of *Solanum peruvianum* (22 kDa) and the *white arrows* indicate deglycosylated *At*DIR6 species (25 kDa and 21 kDa). Lane M represents a size marker.

In order to investigate possible differences in the glycosylation pattern of the two observed protein species, samples from untreated *At*DIR6 and the two distinct deglycosylated *At*DIR6 species were prepared for analysis via MALDI-TOF and LC-ESI-MS (Dr. Jens Pfannstiel, Proteomics Core Facility, Life Science Center, University of Hohenheim). In this analysis, two *N*-glycosylation sites at position N30 and N94 could be identified. These sites correspond to the *N*-glycosylation sites N59 and N123 in the plant derived *At*DIR6 with the N-terminal secretion signal peptide [110]. Both sites exhibited a high mannose type glycan chain with a mass of about 2 kDa for each structure (Hexose$_9$GlcNAc$_2$), which is typical for *P. pastoris* (Figure 3-7, Figure S 11, Figure S 12 and Figure S 13). Different isoforms with nine to thirteen hexose residues at both sites were found by ESI-MS/MS (data not shown). In the peptide mass fingerprint (PMF) of the 25 kDa band a deglycosylated peptide including N30 was identified, whereas the position at N94 remained glycosylated (Figure 3-7b). In contrast, in the PMF of the 21 kDa band deglycosylated peptides for both sites, N30 and N94, were identified (Figure 3-7c) [215].

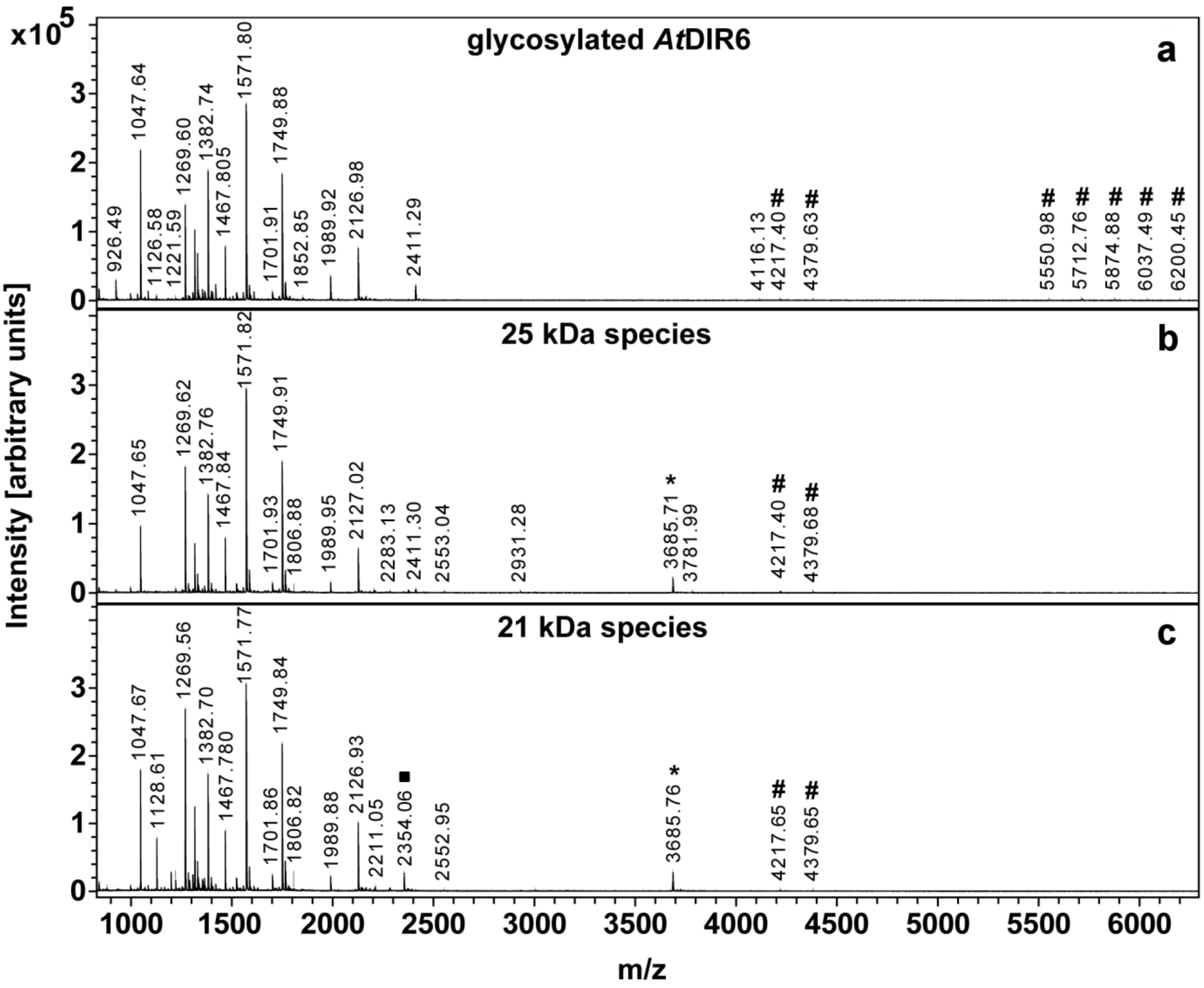

Figure 3-7 – MALDI-TOF peptide mass fingerprints (PMF) of tryptic digests from glycosylated and PNGase F treated *At*DIR6 isolated from *Pichia pastoris* (adapted from [215]**).** PMF spectra of glycosylated *At*DIR6 (**a**), the 25 kDa species (**b**) and the 21 kDa species (**c**) were recorded in the reflector mode with external calibration. *At*DIR6 glycopeptides including the N30 and N94 glycosylation sites are marked by hashes, the corresponding deglycosylated peptides are marked by an asterisk and a filled square, respectively (b, c). Zoomed views of the PMF's showing the glycosylated and deglycosylated peptides are given in supplementary figures Figure S 11 and Figure S 12a-c. The identity of the glycopeptides and the corresponding deglycosylated peptides was verified by MS/MS analysis (see Figure S 13a-d). The figure was prepared by Dr. Jens Pfannstiel (Proteomics Core Facility, Life Science Center, University of Hohenheim).

In correspondence with the proposed reaction mechanism for dirigent proteins, *At*DIR6 was shown to be a native homodimer when derived from a plant cell expression system [109,110]. To estimate the native molecular size, analytical gel filtration was performed. Here, *At*DIR6 exhibited a K_{av} value of 0.182, which corresponded to a molecular mass of 97.7 ± 0.1 kDa (Figure 3-8). However, this indicated a homotetrameric structure for native *At*DIR6 produced in *P. pastoris* based on the molecular mass of 24.6 kDa obtained by MALDI-TOF-MS analysis [215].

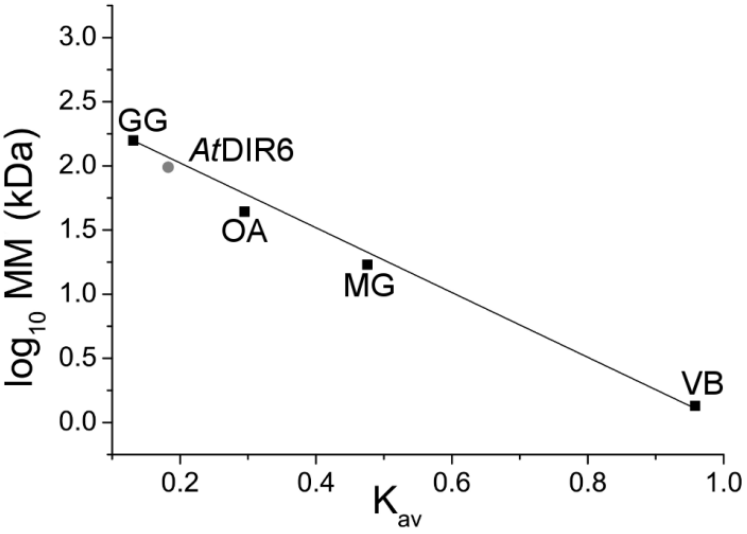

Figure 3-8 – Analytical gel filtration of *At*DIR6 isolated from *Pichia pastoris* (modified from [215]**).** Analytical gel filtration was performed on a XK16/100 column (GE Healthcare, England) packed with Sephacryl S-200. The column was calibrated with γ-globulin (GG, 158 kDa), ovalbumin (OA, 44 kDa), myoglobin (MG, 17 kDa) and vitamin B12 (VB, 1.35 kDa). The K_{av} was determined three times for 500 µg purified *At*DIR6 and the $\log_{10}$(MM) was derived by comparison with the standard proteins.

3.1.4 Dirigent activity of *At*DIR6

It has already been shown that *At*DIR6, like other dirigent proteins, does not possess catalytic activity for the initiation of the phenoxy radical coupling, but can increase overall pinoresinol formation and lead to an enantiomeric excess of (-)-pinoresinol in biotransformations of coniferyl alcohol with an oxidizing agent such as laccases [95,109]. To validate these results and the functionality of the *P. pastoris* produced *At*DIR6, the purified protein was employed in laccase-catalyzed bioconversions of coniferyl alcohol (1), which were analyzed with RP-HPLC (Figure 3-9a). The collected pinoresinol fraction was subsequently separated into (+)- and (-)-pinoresinol using chiral HPLC (Figure 3-9b). As expected no products were formed in the presence of *At*DIR6 without laccase (Figure 3-9a, *solid line*). In contrast, RP-HPLC analysis of biotransformations with *Trametes versicolor* laccase showed the formation of erythro/threo-(±)-guaiacylglycerol 8-O-4´-coniferyl alcohol ether (4a and 4b), (±)-dehydrodiconiferyl alcohol (5) and (±)-pinoresinol (3) (Figure 3-9a, *dotted line*). In the presence of dirigent proteins, quantitative formation of products (4) and (5) was slightly decreased, while formation of (3) was favoured (Figure 3-9a, *dashed line*).

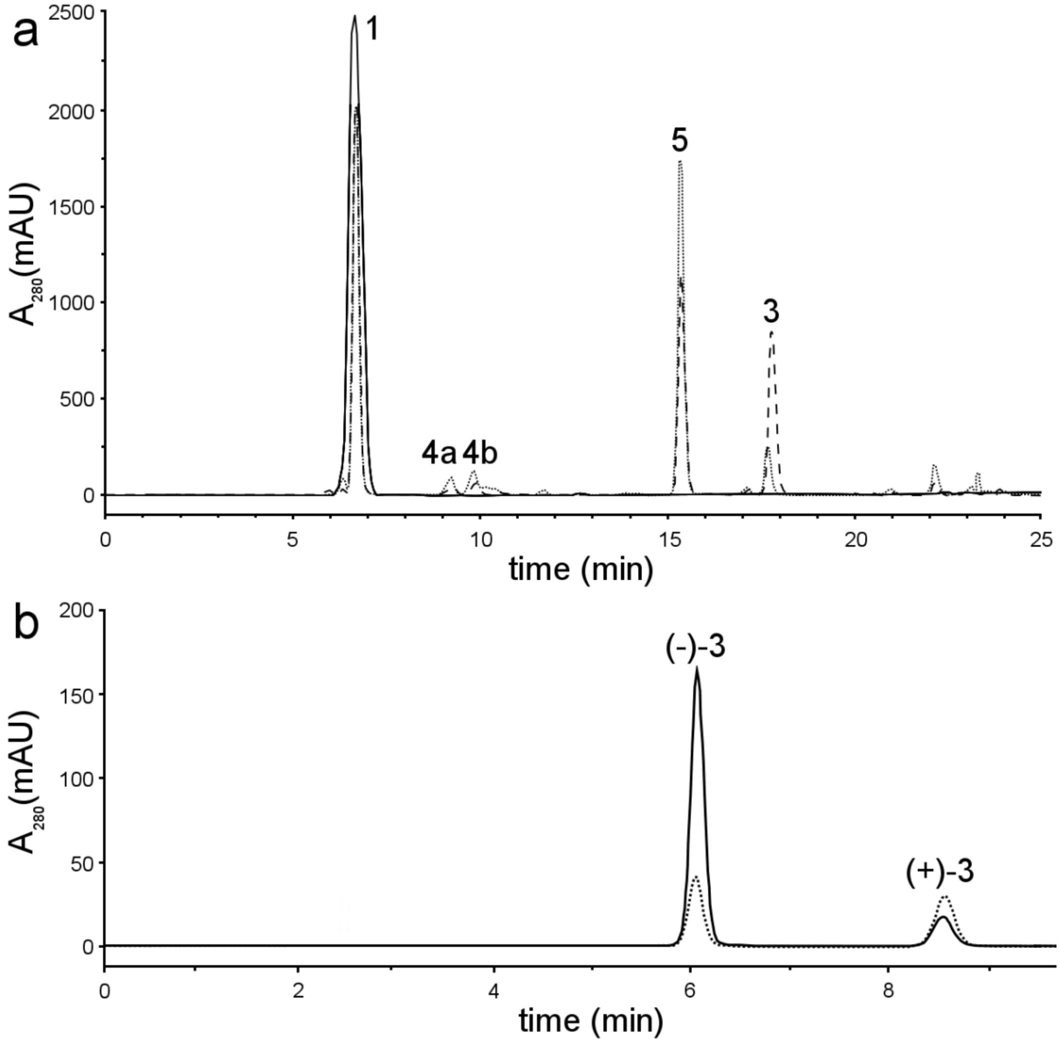

Figure 3-9 – Reversed phase and chiral HPLC analysis of biotransformations with *At*DIR6 isolated from *Pichia pastoris*. **(a)** Reversed phase HPLC analysis of biotransformations of coniferyl alcohol with *At*DIR6 (*solid line*), laccase (*dotted line*) or *At*DIR6 and laccase (*dashed line*). No products are formed in biotransformations with only *At*DIR6, however the product composition of biotransformations with laccase is shifted to pinoresinol formation in the presence of *At*DIR6. Concentrations of coniferyl alcohol, *At*DIR6 and laccase were 4 mM, 8.5 µM and 5.7 µM, respectively. 1: coniferyl alcohol; 3: (±)-pinoresinol 4a/4b: *erythro*/*threo*-(±)-guaiacylglycerol 8-O-4´-coniferyl alcohol ether; 5: (±)-dehydrodiconiferyl alcohol. **(b)** Chiral HPLC analysis of the pinoresinol fraction from RP-HPLC analysis without *At*DIR6 (*dotted line*) and with 25 µM *At*DIR6 (*solid line*). (-)-3: (-)-pinoresinol; (+)-3: (+)-pinoresinol. The figure was created from the original chromatograms using GIMP2.0 software.

Chiral HPLC analysis of the pinoresinol fraction from RP-HPLC showed an increased enantiomeric excess of (-)-pinoresinol validating the functionality of *At*DIR6 from *P. pastoris* (Figure 3-9b).

The enantiomeric excess of (-)-pinoresinol was further evaluated as a function of *At*DIR6, coniferyl alcohol, and laccase concentrations and the results were compared to the dirigent activity of *At*DIR6 derived from cell cultures of *S. peruvianum*. It was found that the enantiomeric excess of (-)-pinoresinol increased with higher concentrations of the dirigent protein (Figure 3-10a). At 83 µM, which corresponds to two mole percent of *At*DIR6 per mole of coniferyl alcohol, an enantiomeric excess of 89.0 ± 0.5% was reached. Although *At*DIR6 from *P. pastoris* fermentations exhibited the typical high mannose type glycan structure and varied therefore strongly from the plant specific glycosylation pattern, *At*DIR6 from *S. peruvianum* and *P. pastoris* showed comparable activities, of 59.4 ± 1.5% and 53.6 ± 1.6% at 8.5 µM, respectively [215].

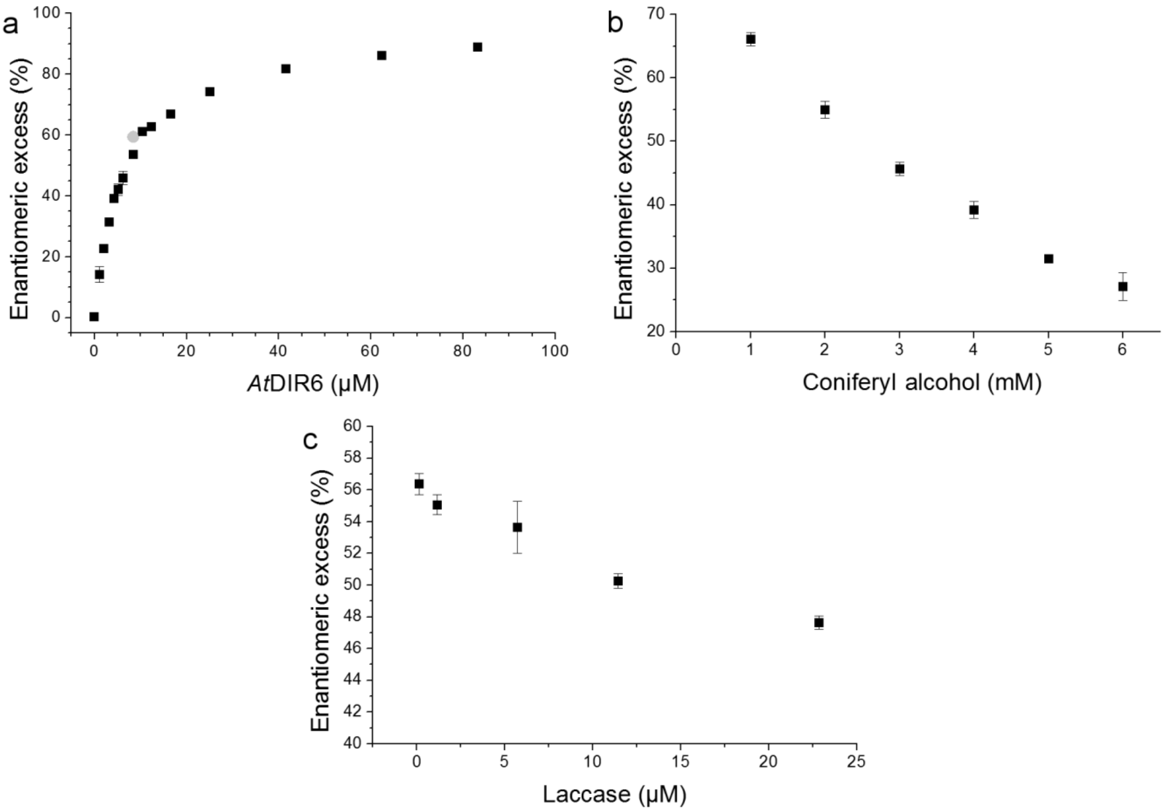

Figure 3-10 – Enantiomeric excess of (-)-pinoresinol as a function of the (a) *At*DIR6, (b) coniferyl alcohol, and (c) laccase concentration (modified from [215]**).** Black squares: *At*DIR6 derived from *P. pastoris*; grey circle: *At*DIR6 derived from *S. peruvianum.* **(a)** Concentrations of coniferyl alcohol and laccase were kept constant at 4 mM and 5.7 µM, respectively. **(b)** Concentrations of *At*DIR6 and laccase were kept constant at 4.2 µM and 5.7 µM, respectively. **(c)** Concentrations of *At*DIR6 and coniferyl alcohol were kept constant at 8.5 µM and 4 mM, respectively. The standard deviations did not exceed 2.6% enantiomeric excess in (**a**), 2.2% in (**b**) and 1.6% in (**c**), respectively.

In contrast, increasing the coniferyl alcohol or laccase concentrations led to a decrease in enantiomeric excess of (-)-pinoresinol. An increase in coniferyl alcohol concentration from 1 to 6 mM, resulted in a marked reduction in enantiomeric excess from 66.1 ± 1.0% to 27.1 ± 2.2% (Figure 3-10b). Altering the laccase concentration from 0.11 to 22.9 µM resulted in a similar, but less prominent effect on the enantiomeric excess, which decreased from 56.4 ± 0.7% to 47.6 ± 0.4%, respectively (Figure 3-10c) [215].

3.1.5 Dirigent activity of deglycosylated *At*DIR6

Since glycosylation can have not only an influence on folding or stability of proteins, but also on the bioactivity [30], the relevance of glycosylation for dirigent activity was evaluated. Therefore, *At*DIR6 was enzymatically deglycosylated with PNGase F and samples from different time intervals were subjected to SDS-PAGE analysis and dirigent activity assays (Figure 3-11). From these experiments it could be shown that the dirigent activity was stable with 81.0 ± 0.8% of the initial activity remaining after 24 h when *At*DIR6 was incubated in the absence of PNGase F (Figure 3-11b).

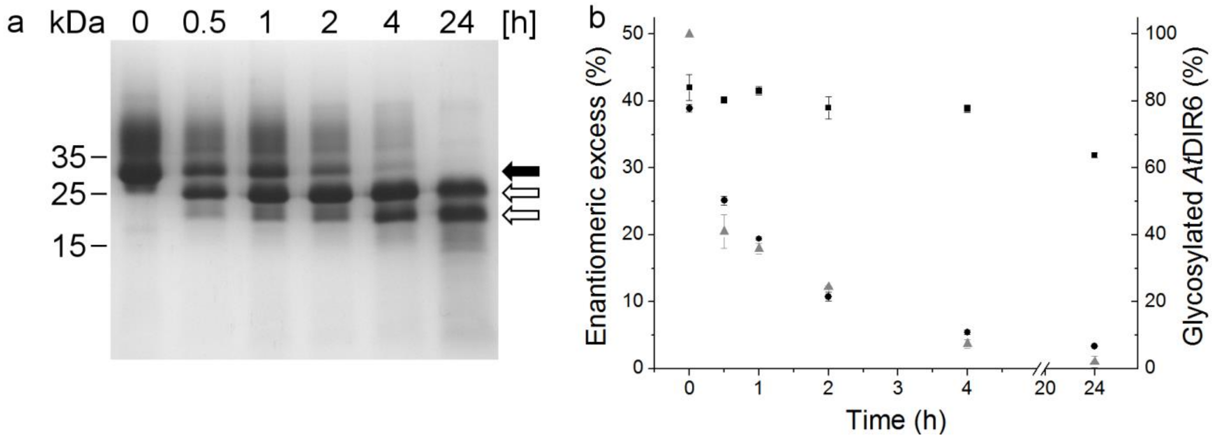

Figure 3-11 – SDS-PAGE and dirigent activity analysis of deglycosylated *At*DIR6 derived from *P. pastoris* (adapted from [215]**). (a)** SDS-PAGE analysis of enzymatic deglycosylation of *At*DIR6 over time. The *black arrow* indicates glycosylated *At*DIR6 from *P. pastoris* (33 kDa) and the *white arrows* indicate deglycosylated *At*DIR6 species (25 kDa and 21 kDa). **(b)** Enantiomeric excess of (-)-pinoresinol as a function of *At*DIR6 from incubations with (*black circles*) or without (*black squares*) PNGase F at different time intervals. The percentage of fully glycosylated *At*DIR6 at a given timepoint (*grey triangles*) from PNGase F incubations were calculated by densitometry. Standard deviations are shown for the densitometric analysis of three independent experiments.

In contrast, the enantiomeric excess considerably decreased over time when *At*DIR6 was incubated in the presence of PNGaseF, with 3.4% remaining after 24 h corresponding to a

90% loss of activity. The loss of activity over time corresponded well with the disappearance of the fully glycosylated *At*DIR6 band at 33 kDa, which showed a 97.9 ± 1.7% reduction after 24 h quantified by densitometric analysis (Figure 3-11a) [215].

To further characterize, whether loss of function of deglycosylated AtDIR6 was due to reduced solubility of the protein, stability and solubility of *At*DIR6 during incubations for up to 96 h at 30°C were investigated. SDS-PAGE analysis of the soluble and insoluble protein fractions verified that fully glycosylated *At*DIR6 was stable during 96 h of incubation (see Figure 3-12). In contrast, a gradual loss of soluble protein was identified for deglycosylated *At*DIR6 [215].

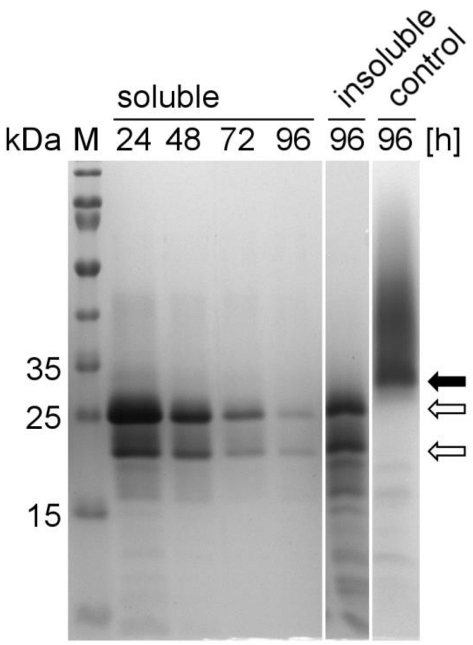

Figure 3-12 – Solubility of deglycosylated *At*DIR6 isolated from *Pichia pastoris* during long-term incubations (adapted from [215]**).** Soluble: soluble *At*DIR6 fractions; insoluble: insoluble *At*DIR6 fraction after 96 h incubation at 30°C; control: soluble fraction of the fully glycosylated *At*DIR6 after 96 h incubation at 30°C. The *black arrow* indicates glycosylated *At*DIR6 derived from *P. pastoris* (33 kDa) and the *white arrows* indicate deglycosylated *At*DIR6 species (25 kDa and 21 kDa). *Lane M* represents a size marker.

3.1.6 Changes in protein structure induced by deglycosylation of *At*DIR6

As loss of function is often accompanied by an alteration in the structural composition of proteins, samples of *At*DIR6 were analyzed by circular dichroism (CD) spectroscopy in the far-UV range (190-260 nm). The spectrum of fully glycosylated *At*DIR6 revealed a single minimum at 220 nm, indicating a high content of β structures, and a characteristic shoulder around 200 nm (Figure 3-13a). By deconvolution of the spectrum, a secondary structure composition of 9% α-helices, 34% β-sheets, 25% turns and 31% unordered elements was found. After deglycosylation of *At*DIR6, a significant change in the CD spectrum was observed. The 220 nm minimum and the wavelength characterized by a mean residue weight ellipticity of 0 deg cm^2 dmole^{-1} were shifted by about 4 nm to 216 nm. Furthermore, the intensity of the positive band decreased by almost two thirds from 27500 to 11400 deg cm^2 dmole^{-1}. However, deconvolution of the spectrum also revealed that the fractions of structural elements were not altered significantly with 10% α-helices, 32% β-sheets, 24% turns and 32% unordered elements for deglycosylated *At*DIR6 [215].

For a further evaluation of the effect of deglycosylation on protein structure, glycosylated and deglycosylated *At*DIR6 was exposed to chemical denaturation by guanidinium hydrochloride (GndCl) and related changes in ellipticity at 220 nm were analyzed (Figure 3-13b). At low concentrations of GndCl, the fully glycosylated *At*DIR6 showed a slight decrease in ellipticity at 220 nm followed by a strong increase up to concentrations of about 1 M GndCl. In contrast, deglycosylated *At*DIR6 showed a slow and gradual increase in ellipticity at 220 nm over the entire concentration range. Above 1.33 M GndCl, both, glycosylated and deglycosylated *At*DIR6, showed identical behavior with a slow increase in ellipticity [215].

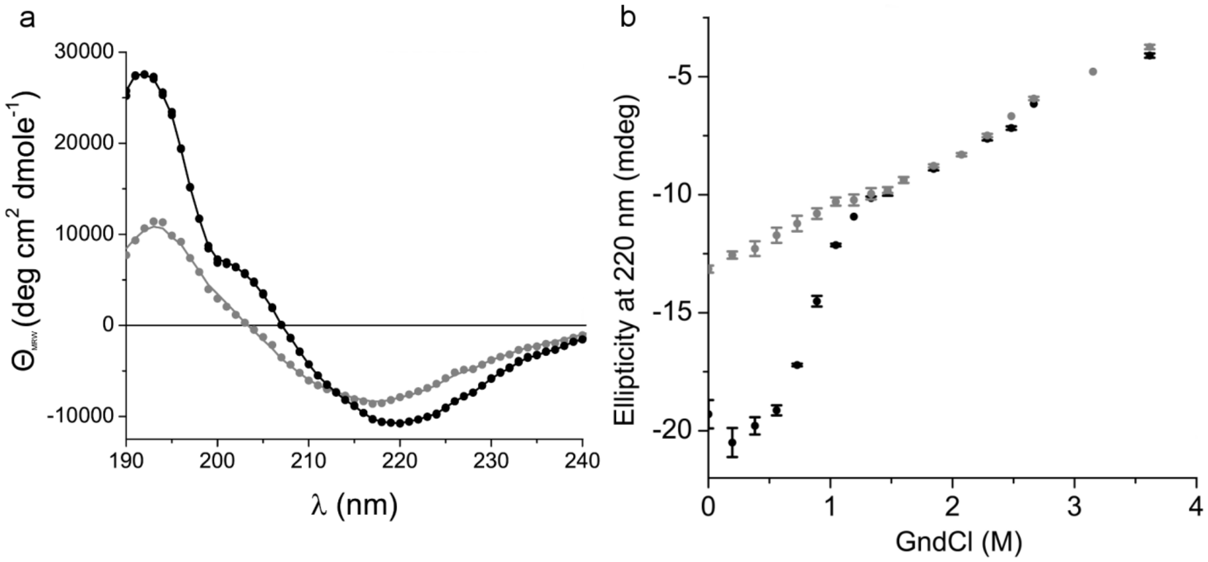

Figure 3-13 – CD spectroscopy of glycosylated and deglycosylated *At*DIR6 derived from *P. pastoris* (modified from [215]**). (a)** Far-UV spectrum of fully glycosylated (*black*) or deglycosylated *At*DIR6 (*grey*). Identical concentrations of glycosylated and deglycosylated *At*DIR6 were used (0.07 mg mL^{-1}). **(b)** Chemical denaturation followed by CD spectroscopy at 220 nm of fully glycosylated glycosylated (*black*) or deglycosylated *At*DIR6 (*grey*). Identical concentrations of glycosylated and deglycosylated *At*DIR6 were used (0.04 mg mL^{-1}).

3.2 Enzyme engineering of *Saccharomyces cerevisiae* ribokinase

3.2.1 A novel ᴅ-xylose metabolizing pathway

Ribokinases catalyze the phosphorylation of ᴅ-ribose to ᴅ-ribose-5-phosphate with ATP as co-substrate [130]. They are ubiquitous among prokaryotes and eukaryotes, highly conserved due to their central role in pentose-phosphate metabolism, and show high substrate specificity towards ᴅ-ribose [129,131]. Ribokinase from *E. coli* was shown to be able to weakly accept ᴅ-xylose as substrate yielding ᴅ-xylose-5-phosphate [146].

While *Saccharomyces cerevisiae* (*S. cerevisiae*) is able to ferment hexoses like ᴅ-glucose, it cannot utilize pentoses like ᴅ-xylose and ᴅ-arabinose, which are the main constituents from hemicellulose in lignocellulosic feedstocks [25]. However, *S. cerevisiae* can ferment the ᴅ-xylose isomer ᴅ-xylulose by phosphorylation to ᴅ-xylulose-5-phosphate and funneling of this intermediate into the pentose phosphate pathway. Metabolic engineering of *S. cerevisiae* strains by implementing pathways for the isomerization of ᴅ-xylose to ᴅ-xylulose have been reported (see Figure 3-14), however they lack the fermentation rate and ethanol yield necessary for an industrial process.

In this case study, a novel ᴅ-xylose metabolizing pathway was proposed. This would include the action of a ribokinase engineered to phosphorylate ᴅ-xylose to ᴅ-xylose-5-phosphate with a specific activity comparable to the natural activity towards ᴅ-ribose (see Figure 3-14). In a second step, the novel intermediate ᴅ-xylose-5-phosphate could be converted to ᴅ-xylulose-5-phosphate, which is an intermediate of the pentose phosphate pathway, e.g. by the action of an engineered ribose-5-phosphate-isomerase from *S. cerevisiae* [25]. Since the novel pathway should be implemented by metabolic engineering into *S. cerevisiae* in further studies, the homologous ribokinase candidate RBK1 from *S. cerevisiae* itself was chosen for the enzyme engineering approach. For the enzyme engineering of RBK1 in regard to the alteration of the substrate specificity, a rational design approach was chosen.

The rationale for the novel pathway was devised in an industrial cooperation with DSM (Heerlen, The Netherlands) for developing novel ᴅ-xylose metabolizing *S. cerevisiae* strains for the production of second generation biofuels from lignocellulosic feedstocks.

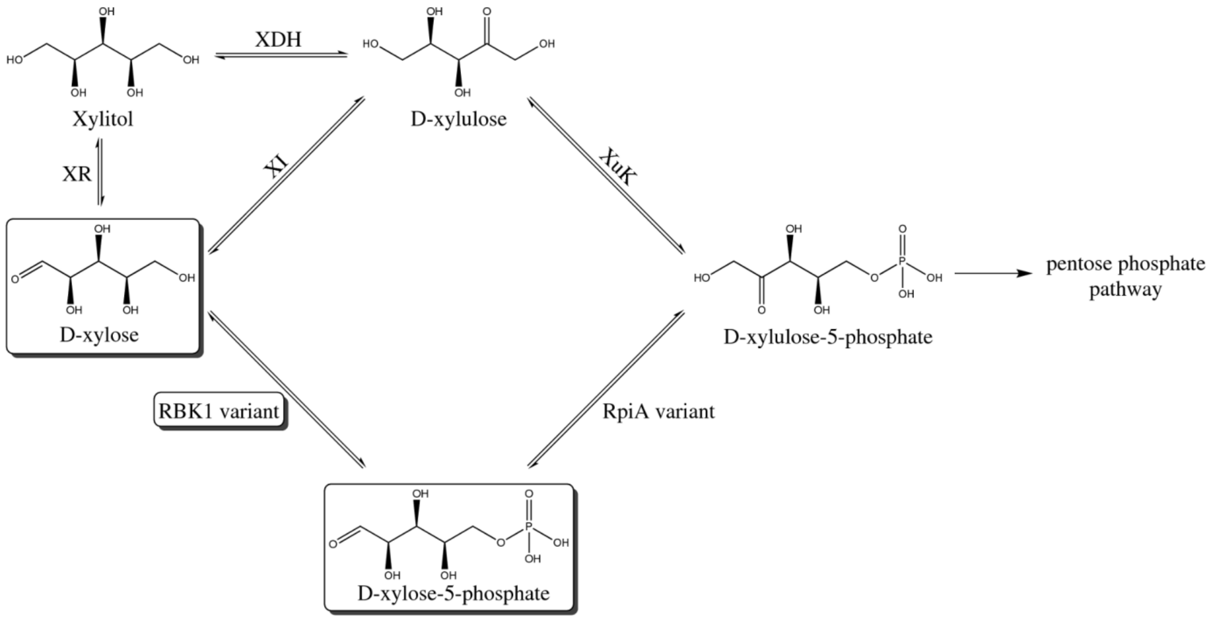

Figure 3-14 – Metabolically engineered pathways for D-xylose metabolism and novel pathway. In the implemented pathways, D-xylose is isomerized to D-xylulose by xylose isomerase (XI) or converted to D-xylulose by the action of xylose reductase (XR) and xylitol dehydrogenase (XDH) *via* xylitol. *S. cerevisiae* can convert D-xylulose to D-xylulose-5-phosphate by the action of xylulokinase (XuK). In the proposed, novel pathway, an engineered RBK1 variant phosphorylates D-xylose to yield D-xylose-5-phosphate. The novel intermediate could then be isomerized to D-xylulose-5-phosphate by an engineered variant of ribose-5-phosphate-isomerase (RpiA). In all cases, D-xylulose-5-phosphate is funneled into the pentose phosphate pathway.

3.2.2 Homology modelling and smart library design

RBK1 from *S. cerevisiae* shows 30.7% identity and 48.9% similarity in the amino acid sequence to the ribokinase from *E. coli* [143]. In addition, members of the ribokinase family were found to be structurally highly similar [129]. Based on the crystal structure of the *E. coli* ribokinase (PDB code: 1RKD) [135], a homology model was designed by the bioinformatics group of the Institute of Technical Biochemistry of the University of Stuttgart (Jannik Vollmer) (see Figure 3-15a). The crystal structure of the *E. coli* ribokinase includes D-ribose bound in the active site. Based on the position of D-ribose in its cyclic α-furanose form, residues were identified, which are in direct contact and are likely to contribute to substrate binding. These residues included isoleucine at position 101 (I101) and 113 (I113), threonine at position 279 (T279), and alanine at position 319 (A319) in the yeast RBK1. D-ribose is bound in a manner that the more hydrophobic side of the sugar ring faces towards the hydrophobic isoleucines, which are situated in the lid domain [135]. While the two residues aspartate 12 (D12) and asparagine 42 (N42) are also in direct proximity and were shown to interact with the hydroxyl groups on the other side of the sugar via hydrogen bonds, both are strictly conserved. While D12 is involved in interactions with the protein backbone, the catalytic base aspartate 283 is supported by N42.

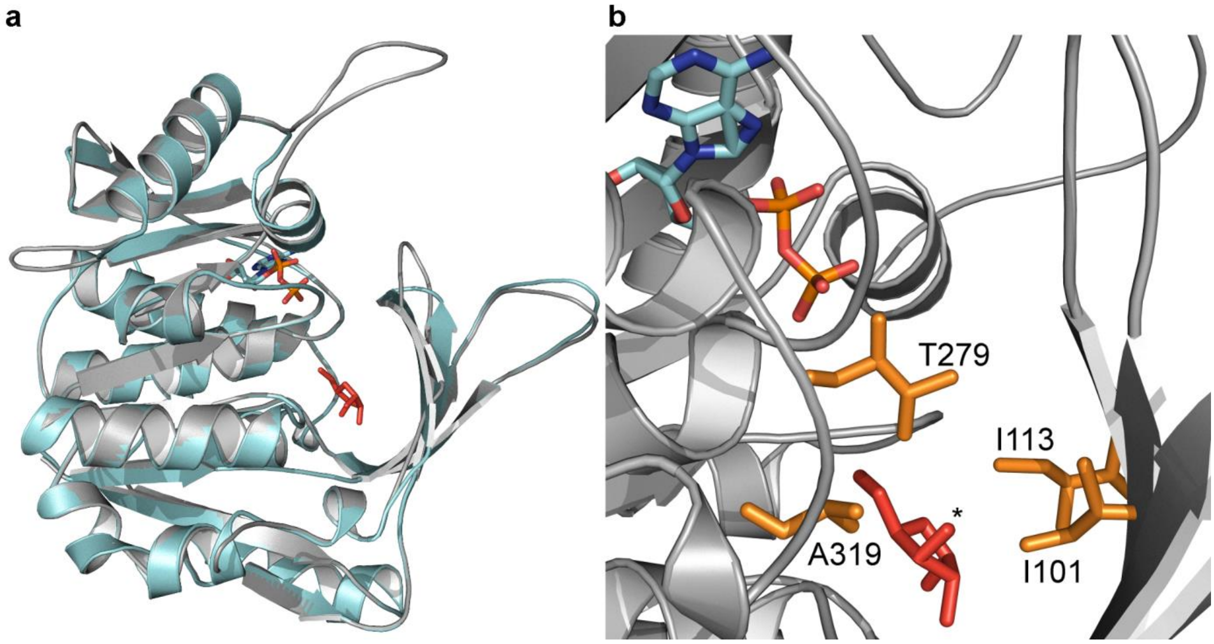

Figure 3-15 – Homology model of the yeast RBK1 protein. (a) Overlay of the homology model of RBK1 (grey) and the *E. coli* ribokinase protein structure (cyan) used as template for homology modeling (PDB: 1RKD). The ADP analogue and α-D-furanoribose bound in the *E. coli* structure are shown in the standard atomic colors and red, respectively. **(b)** A zoomed view of the substrate binding site of the homology model. The sites I101, I113, T279, and A319 chosen for mutagenesis are represented as orange sticks. The ADP analogue is again shown colored according to the standard atomic colors. The bound α-D-furanoribose is exchanged by an α-D-furanoxylose (red) with the altered C3 hydroxyl group indicated by an asterisk.

In D-xylose, the hydroxyl group at carbon atom three (C3-OH) faces in the other direction. Based on the assumption that D-xylose binds in the same manner as D-ribose, the C3-OH group would face mainly towards I101 and I113, while T279 and A319 are in close proximity (see Figure 3-15b). Examination of a family alignment of 200 ribokinase sequences with over 25% sequence identity showed that the four identified residues are able to tolerate mutations, as some minor amino acid variants were observed at these positions (René de Jong, DSM, The Netherlands) (data not shown).

Based on these results, a small focused library of RBK1 variants was designed with the aim to generate several active enzyme variants, while at the same time creating diversity in hydrogen bonding patterns for alternative sugars to bind in the active site. For all amino acid exchanges, the family alignment showed that these variations can be tolerated in other ribokinase sequences.

Table 3-1 – Variants of the focused library.

Single	Double	Triple	Quadruple
I101S	I101S I113T	I101S I113T T279S	I101S I113T T279S A319G
I101V	I101S I113V	I101S I113T A319G	I101S I113T T279S A319S
I101T	I101S T279S	I101S I113T A319S	I101S I113V T279S A319G
I113T	I101S A319G	I101S I113V T279S	I101S I113V T279S A319S
I113V	I101S A319S	I101S I113V A319G	I101T I113T T279S A319G
T279S	I101T I113T	I101S I113V A319S	I101T I113T T279S A319S
A319G	I101T I113V	I101S T279S A319G	I101T I113V T279S A319G
A319S	I101T T279S	I101S T279S A319S	I101T I113V T279S A319S
	I101T A319G	I101T I113T T279S	I101V I113T T279S A319G
	I101T A319S	I101T I113T A319G	I101V I113T T279S A319S
	I101V I113T	I101T I113T A319S	I101V I113V T279S A319G
	I101V I113V	I101T I113V T279S	I101V I113V T279S A319S
	I101V T279S	I101T I113V A319G	
	I101V A319G	I101T I113V A319S	
	I101V A319S	I101T T279S A319G	
	I113T T279S	I101T T279S A319S	
	I113T A319G	I101V I113T T279S	
	I113T A319S	I101V I113T A319G	
	I113V T279S	I101V I113T A319S	
	I113V A319G	I101V I113V T279S	
	I113V A319S	I101V I113V A319G	
	T279S A319G	I101V I113V A319S	
	T279S A319S	I101V T279S A319G	
		I101V T279S A319S	
		I113T T279S A319G	
		I113T T279S A319S	
		I113V T279S A319G	
		I113V T279S A319S	
8 variants	23 variants	28 variants	12 variants

The focused library is based on a combinatorial approach of single, double, triple, and quadruple variants. For this, the hydrophobic I101 and I113 were changed to the similar valine. The rationale for this was to accommodate the altered position of the C3-OH group potentially pointing in this direction for D-xylose as well as allowing larger sugar substrates more space in the active site. Additionally, I101 was changed to two small, polar residues, serine and threonine, and I113 was changed to threonine. These changes might facilitate hydrogen bonds with hydroxyl groups facing towards these residues. In the case of T279, only minor variations were suggested by the family alignment. This residue was changed to the smaller serine, again allowing more space to accommodate larger substrates or stereoisomeric changes in the substrates. For A319, the same rationale was applied for a change to glycine, while more operational freedom suggested by the family alignment also allowed a change to serine, which might again facilitate hydrogen bonding. By combination of the original amino acids at all positions with these new variations, a small focused library of 72 combinatorial variants (including wild-type RBK1) with single, double, triple, and quadruple mutations was established (see Table 3-1).

3.2.3 Expression and purification of RBK1

Since the *E. coli* ribokinase protein was successfully expressed using a pET vector construct in the *E. coli* BL21(DE3) derivative ER2566 (New England Biolabs) [146], expression of the *S. cerevisiae* RBK1 protein was initially investigated with a pET16b vector construct expressing RBK1 as N-terminal His_{10} fusions. Strains used were the common BL21(DE3) expression strain, the derivative C43(DE3) [195] and Rosetta(DE3) since the *rbk1* gene was not codon optimized for *E. coli*. Expression in BL21(DE3) and Rosetta(DE3) at 37°C did not result in an easily distinguishable band at about 38 kDa in SDS-PAGE analysis (see Figure 3-16a). Faint bands at the expected size of RBK1 were visible in Western blot analysis with anti-His antibodies supporting weak expression strength, however no signals were observed in the soluble fraction (*lanes* S, Figure 3-16a). Interestingly, a signal was already observed for the sample before inducing with 0.5 mM IPTG (*lanes* 0, Figure 3-16a). This indicated that the repression of the T7 promoter was not completely tight as verified by Western blot analysis. In contrast, expression with C43(DE3) resulted in strong expression of RBK1 as analyzed by SDS-PAGE and verified by Western blot analysis (see Figure 3-16b). Soluble protein expression as indicated by Western blot analysis was achieved at a low temperature of 20°C. At this temperature, the *E. coli* cells grew slowly and produced only low protein yields of

RBK1. Expression at a higher temperature (37°C) did not lead to soluble protein and the Western blot analysis also showed signals in the lower molecular range indicating protein degradation. Variation of inducer concentration (0.01 mM to 1 mM IPTG), expression temperature (30°C) or expression time (2 h to 24 h) could not improve the soluble expression of RBK1 using the pET16b_RBK1 vector construct harboring the *rbk1* gene (data not shown).

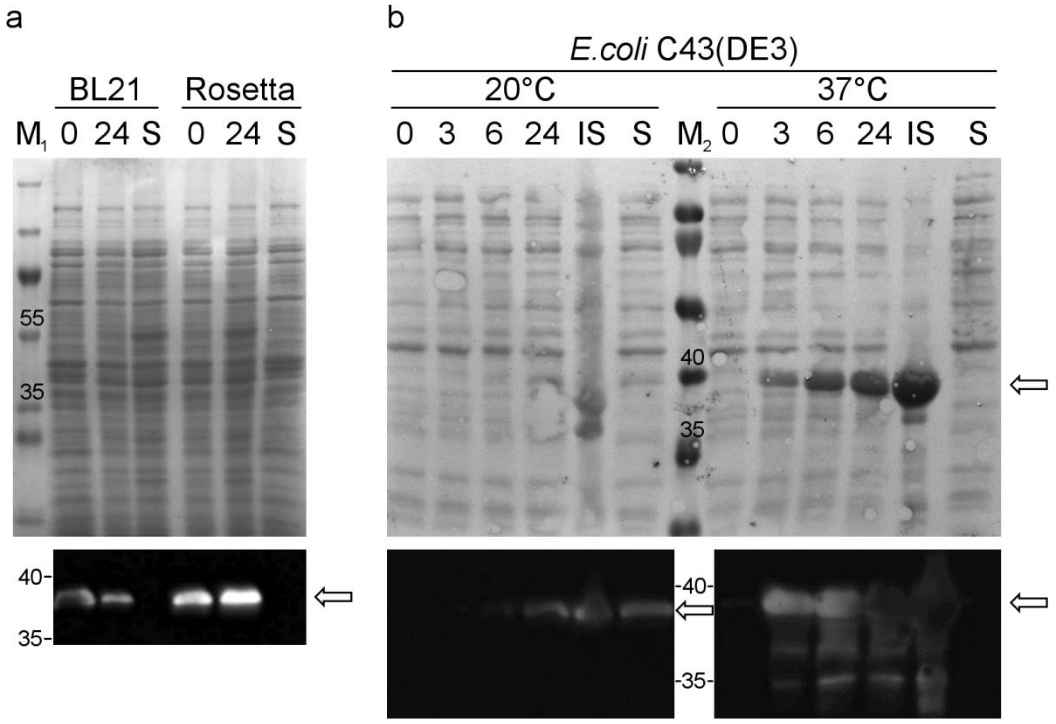

Figure 3-16 – Comparison of the expression of RBK1 in *E. coli* BL21(DE3), Rosetta(DE3) and C43(DE3) hosts by SDS-PAGE and Western blot analysis. (a) Expression of RBK1 in *E. coli* BL21(DE3) and Rosetta(DE3) at 37°C. (**b)** Expression of RBK1 in *E. coli* C43(DE3) at 20°C and 37°C. **Lane abbreviations**: *Lanes* 0 or 3, 6 and 24 show total cell protein fractions before or after IPTG induction, respectively. *Lanes S* and *IS* show the soluble protein fraction after cell lysis and the insoluble (pellet) fraction after cell lysis, respectively. *Lanes* M_1 and M_2 represent molecular size markers. Protein bands are marked with their respective molecular weight in kilodaltons (kDa). 15 µg of total protein was loaded in each lane. The *white arrows* indicate the protein bands identified as RBK1 by SDS-PAGE and Western blot analysis.

Expression in C43(DE3) using a pJOE vector construct (Dr. Joseph Altenbuchner, Institute for Industrial Genetics, University of Stuttgart) [212–214] with the *rbk1* gene under the control of the rhamnose promoter resulted in high yields of soluble protein when expressed for 24 h at 30°C with 0.5% rhamnose as inducer. Still, a large percentage of recombinantly expressed RBK1 was found in the insoluble fraction (see Figure 3-17a).

For further characterization of RBK1, the C-terminally His_6-tagged protein was purified by immobilized metal affinity chromatography (IMAC) and again verified by Western blot analysis with anti-His antibodies (data not shown). After elution with 200 mM imidazole, the RBK1 containing fraction was dialyzed against 100 mM potassium phosphate buffer at pH 7.5 containing 100 mM KCl. After determination of the protein concentration (approximately 8-10 mg mL^{-1}), the sample was subjected to SDS-PAGE analysis (see Figure 3-17b). The expression with the pJOE vector construct harboring the *rbk1* gene in *E. coli* C43(DE3) resulted in an almost pure protein (> 95% purity as assessed by densitometric analysis of SDS gels, data not shown) after one purification step (IMAC) and gave excellent yields of 150-200 mg purified protein per litre cell culture. Only minor protein impurities were faintly visible in the 70-85 kDa and 15-20 kDa range.

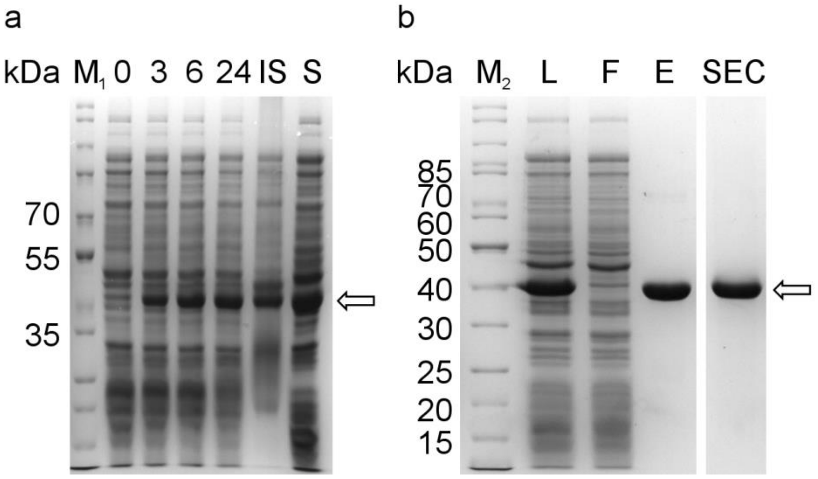

Figure 3-17 – Expression and purification of RBK1 from *E. coli* C43(D3) cells. (a) Expression of RBK1 in *E. coli* C43(DE3) at 30°C. **(b)** Purification of RBK1 by IMAC. **Lane abbreviations:** *Lanes M_1* and *M_2* represent molecular size markers. Protein bands are marked with their respective molecular weight in kilodaltons (kDa). The *white arrows* indicate the protein bands identified as RBK1 by SDS-PAGE. For unpurified samples, 15 µg of total protein was loaded in each lane. For purified RBK1, 5 µg protein was loaded in each lane. *Lanes* 0 or 3, 6 and 24 show total cell protein fractions before or after rhamnose induction, respectively. *Lanes S* and *IS* show the soluble protein fraction after cell lysis and the insoluble (pellet) fraction after cell lysis, respectively. *Lanes L, F, E* and *SEC* show the soluble protein fraction used for the IMAC purification, the flow-through of the IMAC purification, the RBK1 containing eluate fraction after the IMAC purification and the RBK1 containing eluate fraction after size exclusion chromatography, respectively.

A subsequent purification step was included using size exclusion chromatography (SEC) to prepare the protein for crystallization experiments (see 3.2.7). Therefore, a XK-16/100 column was packed with Sephacryl S-200 HR resin (GE Healthcare, Little Chalfont, United Kingdom). The column efficiency was validated by an acetone standard. The minimum specifications for a self-packed column with this kind of resin are theoretical plates N > 9,000 and an asymmetry factor A_s = 0.8 - 1.5. The column (183 mL CV) showed excellent values with N = 14,600 and A_s = 1.1 supporting an effective polishing step. The pooled RBK1 protein fractions from IMAC purification were buffer exchanged to 20 mM Tris at pH 7.0 with 150 mM NaCl, applied to the SEC column in three 2 mL samples (8.5 mg mL^{-1}) and run with the same buffer at 1.0 mL min^{-1} flow rate. The major protein containing fractions were pooled (see Figure 3-18b), buffer exchanged to 10 mM Tris pH 7.5 and concentrated to 10 mg mL^{-1} and 20 mg mL^{-1}, respectively, using Amicon Ultra 4 centrifugal filter units with a nominal molecular mass limit of 10 kDa (Merck Millipore, Schwalbach, Germany). Although the RBK1 samples were almost pure after IMAC (> 95% purity), it was possible to remove the minor impurities in the 70-85 kDa and 15-20 kDa range leaving an apparently pure RBK1 after SEC (see Figure 3-17b).

Analysis of the elution time of RBK1 samples during size exclusion chromatography/analytical gel filtration in comparison to standard proteins resulted in a K_{av} value of 0.223 for RBK1 corresponding to a molecular mass of 75.7 kDa (see Figure 3-18a). This would indicate a dimeric structure for native RBK1 from *S. cerevisiae* produced in *E. coli* based on the molecular mass of 37.7 kDa calculated from the amino acid sequence.

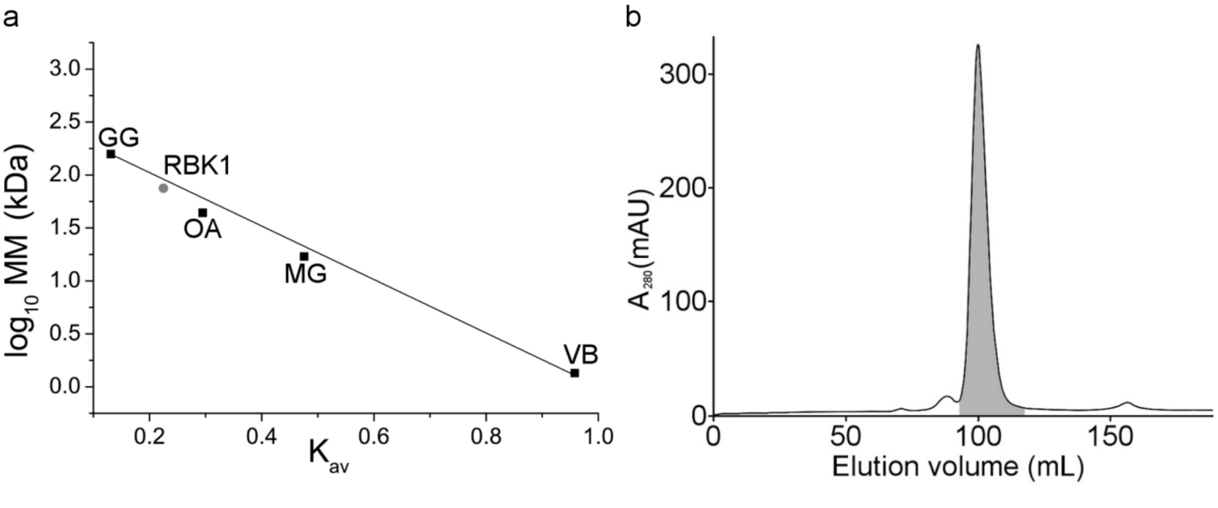

Figure 3-18 – Gel filtration purification of RBK1. (a) Analytical gel filtration was performed on a XK16/100 column (GE Healthcare, England) packed with Sephacryl S-200. The column was calibrated with γ-globulin (GG, 158 kDa), ovalbumin (OA, 44 kDa), myoglobin (MG, 17 kDa) and vitamin B12 (VB, 1.35 kDa). The K_{av} was determined three times for 1 mg purified RBK1 and the $\log_{10}$(MM) was derived by comparison with the standard proteins. **(b)** Chromatogram of a size exclusion purification of RBK1. The shaded fractions were pooled. The figure was created from the original chromatogram using the GIMP2.0 software.

After the expression system for the RBK1 wild-type was established, all sequence verified Quikchange™ mutants of the focused RBK1 library were expressed with the pJOE vector constructs in *E. coli* C43(DE3) induced with 0.5% rhamnose at an OD_{600} of 0.5 for 24 h at 30°C in a 200 mL scale in 1 L unbaffled shaking flasks. All variants showed strong expression comparable to the RBK1 wild-type in SDS-PAGE analysis (data not shown). The C-terminally His_6-tagged RBK1 variants were purified with IMAC using the setup established for the wild-type. Typical yields of 30-40 mg from 200 mL culture in a 4 mL fraction (7-10 mg mL^{-1}) of purified protein were achieved. The purity of all variants was comparable to the RBK1 wild-type (> 95% purity as assessed by densitometric analysis of SDS gels, data not shown) (see Figure 3-19).

In respect to the oligomerization state assessed by analytical gel filtration, RBK1 variants I101V and I113V, which were also purified for crystallization studies, showed K_{av} values similar to the wild-type and therefore indicated a dimeric native structure as well (data not shown).

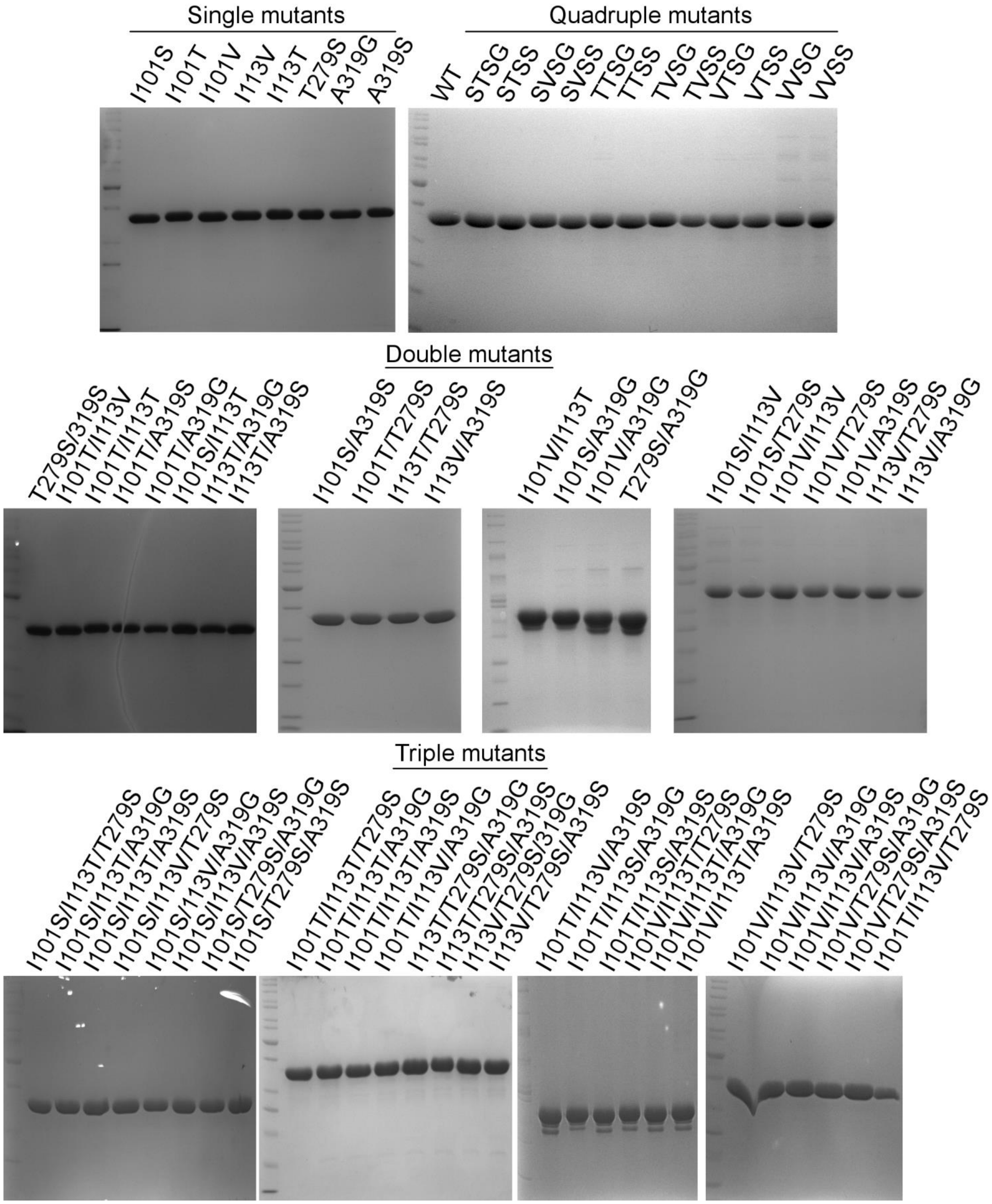
Single mutants
I101S
I101T
I101V
I113V
I113T
T279S
A319G
A319S
Quadruple mutants
WT
STSG
STSS
SVSG
SVSS
TTSG
TTSS
TVSG
TVSS
VTSG
VTSS
VVSG
VVSS
Double mutants
T279S/319S
I101T/I113V
I101T/I113T
I101T/A319S
I101T/A319G
I101S/I113T
I113T/A319G
I113T/A319S
I101S/A319S
I101T/T279S
I113T/T279S
I113V/A319S
I101V/I113T
I101S/A319G
I101V/A319G
T279S/A319G
I101S/I113V
I101S/T279S
I101V/I113V
I101V/T279S
I101V/A319S
I113V/T279S
I113V/A319G
Triple mutants
I101S/I113T/T279S
I101S/I113T/A319G
I101S/I113T/A319S
I101S/I113V/T279S
I101S/113V/A319G
I101S/I113V/A319S
I101S/T279S/A319G
I101S/T279S/A319S
I101T/I113T/T279S
I101T/I113T/A319G
I101T/I113T/A319S
I101T/I113V/A319G
I113T/T279S/A319G
I113T/T279S/A319S
I113V/T279S/319G
I113V/T279S/A319S
I101T/I113V/A319S
I101T/I113S/A319G
I101T/I113S/A319S
I101V/I113T/T279S
I101V/I113T/A319G
I101V/I113T/A319S
I101V/I113V/T279S
I101V/I113V/A319G
I101V/I113V/A319S
I101V/T279S/A319G
I101V/T279S/A319S
I101T/I113V/T279S

Figure 3-19 – SDS-PAGE analysis of the purified RBK1 variants. 15 µg purified RBK1 variants were loaded in each lane. Abbreviations for quadruple mutants stand for the amino acid exchanges (in one letter code) in the order of positions I101(S/T/V), I113(T/V), T279(S) and A319(G/S). WT: RBK1 wild-type.

3.2.4 Biochemical characterization of RBK1

For a valid comparison of the RBK1 wild-type and mutants from the focused library, an intense biochemical characterization of RBK1 was performed. Therefore, the activity was assessed by a coupled NADH assay. Additionaly, substrate conversion and product formation during biotransformations were analyzed by HPLC. Furthermore, the stability as well as the ion and pH requirements of the enzyme were investigated and kinetic data of the bi-substrate reaction with the natural substrates and additional substrates were collected.

3.2.4.1 Activity and stability

Purified RBK1 from *S. cerevisiae* was strongly dependent on ion concentrations as already described for ribokinase from *E. coli* [146,152,154] and other organisms [143]. With a turnover number k_{cat} of 82.3 ± 0.3 s^{-1} the highest activity versus D-ribose at 22°C was measured in 100 mM potassium phosphate (KP_i) buffer at pH 6.0 with 2.5 mM magnesium chloride ($MgCl_2$). When measured at 37°C, the turnover number reached 165.0 ± 4.3 s^{-1}, which is comparable to the highest measured rate of ribokinase from *E. coli* at 37°C (130 s^{-1}) [146]. At higher temperatures, activity increased up to 401.7 ± 16.2 s^{-1} at 65°C, however when incubated for 10 min at 65°C prior to the measurement, already about 50% of the activity was lost.

The stability of RBK1 was evaluated using the coupled NADH assay to assure a conclusive comparison in regard to activity between the wild-type and variants from the focused library. Therefore, a short-term stability under conditions used in the coupled NADH assay and biotransformations and a long-term stability under storage conditions was assessed. For incubation at 30°C and 22°C, the enzyme lost only about 3% and 1%, respectively, of its activity in 2 h. After 24 h, the residual activities were still 78.9 ± 2.0% and 99.3 ± 2.4%, respectively. Especially for the biotransformations with RBK1 at 30°C, the high stability of the enzyme should guarantee comparable results. For long-term stability at 4°C and -20°C, residual activities were 74.7 ± 1.6% and 97.2 ± 3.8%, respectively, after eight weeks of storage. Again, the high stability of the enzyme proved unproblematic for the production and evaluation of the focused library variants of RBK1.

3.2.4.2 Optimization of ion concentrations and pH

Since the ribokinase from *E. coli* showed strong dependence on adequate mono, di- and pentavalent ion concentrations [146,152,154], optimized parameters regarding the concentration of potassium, magnesium and phosphate ions as well as pH were examined for the purified RBK1 from *S. cerevisiae*. Changes in pH and ion concentrations could influence the activity of the two coupling enzymes, pyruvate kinase and lactate dehydrogrenase, which are used in the coupled NADH assay (see 2.5.1). Therefore, in all experiments it was verified that the coupling reactions catalyzed by pyruvate kinase and lactate dehydrogenase did not represent the rate limiting steps. Initial activities were measured in 75 mM triethanolamine (TEA) buffer at pH 7.5 containing 5 mM KP_i and 5 mM $MgCl_2$ and resulted in a k_{cat} of only 27.8 ± 0.9 s^{-1} (versus 82.3 ± 0.3 s^{-1} in 100 mM KP_i buffer at pH 6.0 with 2.5 mM $MgCl_2$). By optimization of the ion concentrations and pH to 100 mM KP_i buffer at pH 6.0 with 2.5 mM $MgCl_2$, it was possible to improve the turnover number by a factor of three.

Bacterial and mammalian ribokinases were shown to be affected by pentavalent ions increasing both the velocity and affinity of the enzyme for D-ribose [143,154]. For RBK1, an increase in activity from 7.6 ± 0.1 s^{-1} to 53.4 ± 0.9 s^{-1} at 0 mM up to 85 mM inorganic phosphate (P_i), respectively, was found (see Figure 3-20a). Higher concentrations of P_i resulted in a decreased activity, however also higher concentrations of potassium might influence the activity. Measurements were performed in KP_i buffer at pH 7.5 with 5 mM $MgCl_2$ and potassium chloride (KCl) up to a final concentration of 100 mM potassium. These non-optimized parameters reflect the difference to the maximum activity with 100 mM KP_i buffer at pH 6.0 with 2.5 mM $MgCl_2$ (82.3 ± 0.3 s^{-1}). Relative activities under 2% were reported for mammalian and bacterial ribokinases without inorganic phosphate in the reaction buffer [143,154]. The relative activity of about 14% without inorganic phosphate found in the coupled NADH assay for RBK1 is likely due to the activating effect of phosphoenolpyruvate (PEP) present in the coupled NADH assay reaction buffer, which was already described by Maj and Gupta [154]. Doubling the standard concentration of PEP at low phosphate concentrations had a measurable effect. However, at high P_i concentrations the enzyme was already fully activated by P_i and no further activation via PEP was measurable (data not shown).

The best activity with different magnesium concentrations was found at 2.5 mM $MgCl_2$ (see Figure 3-20a), which is in the same range as for the concentrations used with bacterial (5 mM) and mammalian (2 mM) ribokinases [146,154].

In contrast to the *E. coli* ribokinase whose highest activities were at a pH of 8 to 9 [146], the evaluation of the optimum pH value resulted in the highest activity of RBK1 at pH 6.0 while the enzyme showed high activity in a rather broad pH spectrum (see Figure 3-20b). Below pH 6.0, the activity decreased rather fast with less than 60% residual activity at pH 4.5, while at pH 9.0 still over 75% residual activity was measurable.

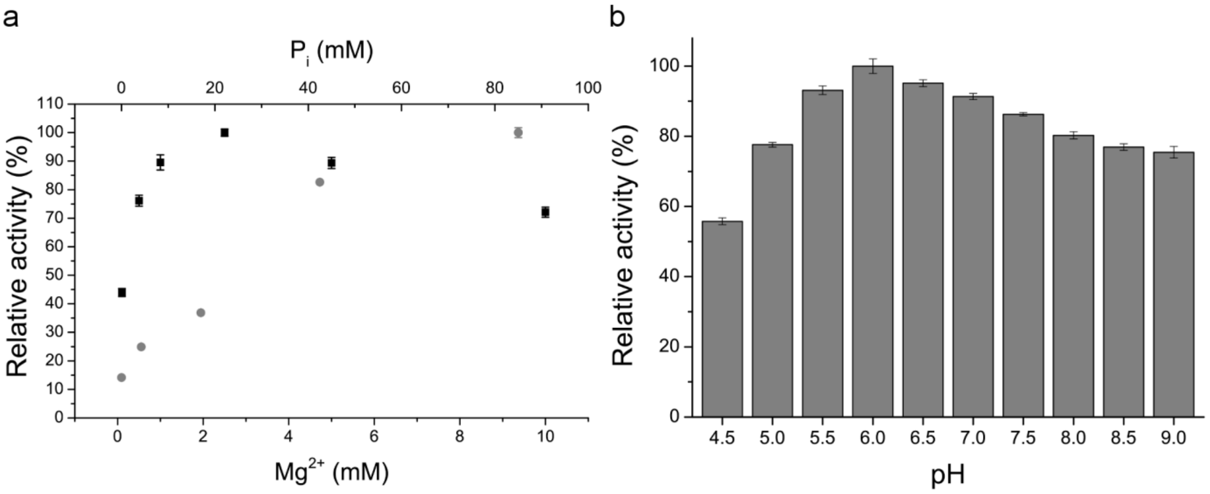

Figure 3-20 – Optimization of magnesium and phosphate concentrations and pH value for RBK1 activity. (a) Grey circles: relative activity as a function of the phosphate (P_i) concentration; black squares: relative activity as a function of magnesium (Mg^{2+}) concentration. Standard deviations from three independent samples are shown and were always below 3%. **(b)** Relative activity as a function of the pH value. Standard deviations from three independent samples are shown and were always below 2%.

3.2.4.3 Biotransformations and analytics via HPLC

In addition to the indirect measurement of the RBK1 activity by the coupled NADH assay, biotransformations of sugar substrates with RBK1 and subsequent analysis by HPLC and LC/MS were performed. This setup had the advantage of allowing further characterization of the ion requirements of RBK1 since no other enzymatic steps and cofactors are involved. Furthermore, this allowed the direct detection of the phosphorylated reaction products e.g. D-ribose-5-phosphate.

Since the sugar substrates and phosphorylated sugar products are not detectable with a diode array detector (DAD; 1260 infinity diode array detector; Agilent, Santa Clara, CA, USA), a refractive index detector (RID; 1260 infinity refractive index detector; Agilent, Santa Clara, CA, USA) was used with an Agilent 1200 series HPLC system. This however limited the used method to an isocratic flow, because a gradient would lead to a severe baseline drift in the refraction index. For a complete separation of all reaction substrates and products, namely ribose, D-ribose-5-phosphate, ATP and ADP, several methods and columns for polar substances were tested. While there was no satisfying separation of substrates and products achieved with an Polaris NH_2 column (4.6x150 mm, 5 µm particle size, Agilent, Santa Clara, CA, USA) (data not shown), ribose and D-ribose-5-phosphate could be fully separated on a TSKgel Amide-80 column (4.6x250 mm, 3 µm particle size, Tosoh, Tokyo, Japan) with 65% acetonitrile in 0.1% formic acid as solvent at 1.0 mL min^{-1} flow rate and 30°C. However, the ATP and ADP peaks showed strong peak fronting and overlapped completely precluding any quantification via DAD at 254 nm. Furthermore, the peak produced from 50 mM sulfuric acid from quenching of the biotransformations overlapped with ribose and D-ribose-5-phosphate preventing routine quantification of these compounds with the RID as well (see Figure 3-21a). A separation of all compounds was achieved using an Aminex HPX-87H (7.8x300 mm, 9 µm particle size, Bio-Rad, Munich, Germany) with 5 mM sulfuric acid as solvent at 0.5 mL min^{-1} flow rate and 60°C. This column was further used for standard operation.

For the calculation of conversion, standard curves for ribose, D-ribose-5-phosphate, ATP and ADP were generated. In the basic setup with 5 mM ribose, 2 mM ATP and 30 nM RBK1 in 100 mM KP_i buffer at pH 6 with 2.5 mM $MgCl_2$, 1.51 ± 0.01 mM ribose were converted in 10 min at 30°C and 750 rpm (Thermomixer comfort, Eppendorf, Germany) (see Figure 3-21b). A high correlation with the conversion of ATP (1.55 ± 0.01 mM) and the formation of ADP (1.52 ± 0.03 mM) was found.

Since the D-ribose-5-phosphate peak detected with the RID overlapped with the ATP peak and the sulfate peak from quenching of the biotransformations, it was not used for routine quantification. However, biotransformations, which were not stopped with sulfuric acid, showed that a complete conversion of 1.97 ± 0.08 mM D-ribose and 1.99 ± 0.10 mM ATP correlated with the formation of 1.93 ± 0.05 mM D-ribose-5-phosphate. Due to the high correlation of conversion of substrates and formation of products, routinely only the conversion of ribose and ATP were quantified.

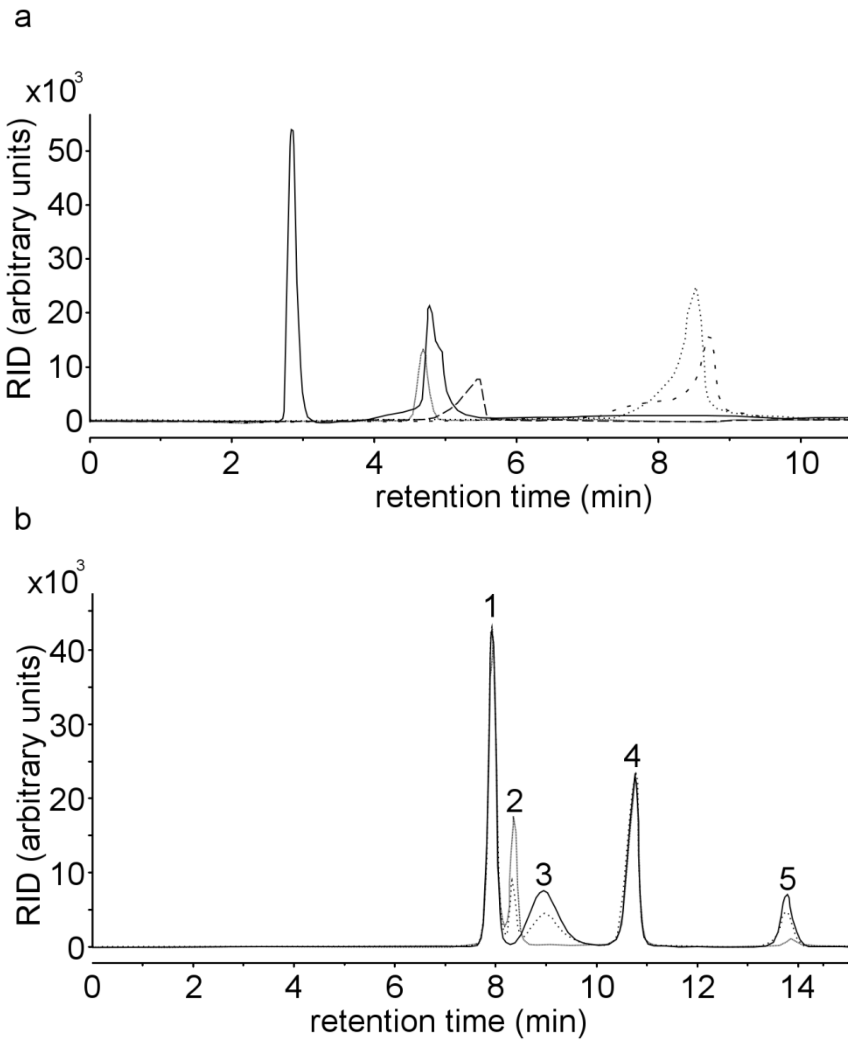

Figure 3-21 – HPLC analysis of biotransformations. (a) Chromatograms of various standards using an Amide-80 column (Tosoh, Tokyo, Japan) column. *Black line*: 50 mM potassium phosphate buffer at pH 6.0 with 50 mM sulfuric acid; *grey line*: 10 mM D-ribose; *dashed line (long dashes)*: 10 mM D-ribose-5-phosphate; *dotted line*: 10 mM ATP; *dashed line (short dashes)*: 10 mM ADP. **(b)** Chromatograms of biotransformations of D-ribose and ATP with RBK1 analyzed with an Aminex HPX-87H (column). 10 mM D-ribose (*peak* 5, retention time of 13.8 min) and 9 mM ATP (*peak* 3, retention time of 8.9 min) were incubated for 10 min at 30°C. *Peak 1*: 50 mM sulfuric acid (retention time of 7.9 min); *peak 2*: D-ribose-5-phosphate (retention time of 8.4 min; *peak 4*: 50 mM potassium phosphate at pH 6.0 (retention time of 10.8 min); *black line*: biotransformation after 0 min; *dotted line*: biotransformation after 5 min; *grey line*: biotransformation after 10 min. Formation of ADP (Retention time of 22.5 min) is not shown.

For comparison of the enzyme activity of RBK1 in the coupled NADH assay and biotransformations, the conversion in the basic setup was evaluated as a function of time (see Figure 3-22a). In the first 4 min an initial, maximum conversion rate of about 200 µM min^{-1} for 30 nM of RBK1 was reached. After this, the conversion rate decreased until the limiting substrate ATP was depleted. This reduce in conversion rate was likely due to a combination of product inhibition and substrate shortage. To avoid a rate limiting effect like this, a conversion of 0.8 mM was not exceeded in the basic setup for further comparative evaluations. The initial conversion rate corresponds well with the activity measured in the coupled NADH assay at 30°C. A k_{cat} of 110.3 ± 2.5 s^{-1} matches a conversion of 199 ± 5 µM min^{-1} for 30 nM enzyme indicating a normal behavior of RBK1 in the coupled NADH assay (see also Table 3-3).

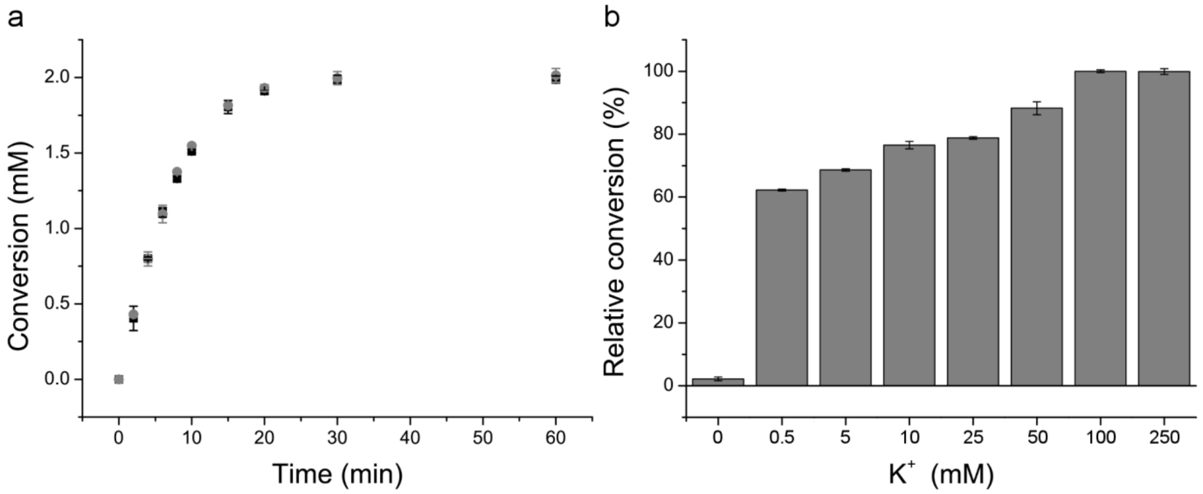

Figure 3-22 – Conversion rate as a function of time and potassium cation concentration. (a) Conversion rate as a function of time. D-ribose (5 mM) and ATP (2 mM) were converted with 30 nM RBK1 at 30°C and 750 rpm. Standard deviations are shown for 3 individual samples. **(b)** Relative conversion of D-ribose at different potassium ion concentrations. Ribose (5 mM) and ATP (2 mM) were converted with 1 µM RBK1 in 45 min at 30°C and 750 rpm. Standard deviations are shown for three individual samples.

The evaluation of the potassium concentration needed for full activation of RBK1 was performed with biotransformations and subsequent HPLC analysis. The coupled NADH assay was not used for this analysis due to several reasons. Since no potassium phosphate could be used, overall activities were rather low as the activation of RBK1 by phosphate was missing. The use of sodium phosphate was avoided as well, since at this point it was not clear whether sodium would activate RBK1 on its own. The resulting low conversions were easier to measure with biotransformations over a longer time-span. Furthermore, the coupling enzyme pyruvate kinase in the coupled NADH assay is dependent on potassium, which would preclude a control without any potassium. The biotransformations with RBK1 were performed in 100 mM Tris buffer at pH 6.0 with 2.5 mM $MgCl_2$. Further, 1 µM RBK1, which was buffer exchanged to 10 mM Tris at pH 7.5, was used in these reactions. This had two reasons, one being the need to circumvent contamination of the buffer with potassium from the enzyme preparation after IMAC purification. Secondly, the high amount of enzyme was necessary since no P_i was present in the reaction. As has been shown, the conversion of ribose dropped to about 3% without P_i. Although 2mM ATP available as disodium salt contaminated the buffer with 4 mM sodium, the negative control showed almost no conversion (0.02 ± 0.01 mM). Comparable to the results for the ribokinase from *E. coli* [146,152], a jump-like increase in relative conversion from 2.2 ± 0.6% to 62.2 ± 0.3% was found with an increase of the potassium concentration from 0 mM to 0.5 mM. The highest conversion was measured at 100 mM potassium ion concentration and no further influence on conversion rate was found up to 250 mM (see Figure 3-22b).

As seen in the phosphate screening with the coupled NADH assay, PEP can also function as activator comparable to P_i. RBK1 showed only 3.4 ± 0.8% conversion in biotransformations without P_i relative to a maximum conversion when the enzyme was fully activated by 100 mM P_i in 100 mM Tris buffer at pH 6.0 with 100 mM KCl and 2.5 mM $MgCl_2$. This is comparable to the results for bacterial and mammalian ribokinases [143,154].

It was reported that other mono- and divalent cations can facilitate the activation of bacterial ribokinases. Therefore, in addition to potassium, the influence of the monovalent cations ammonium, cesium, lithium, rubidium and sodium was tested. As full activation of RBK1 was observed with 100 mM potassium, all other monovalent cations were used as the respective chloride salts at a concentration of 100 mM in 100 mM Tris buffer at pH 6.0 with 2.5 mM $MgCl_2$. While all monovalent cations activated RBK1 to some degree, only rubidium almost reached the level of potassium (see Figure 3-23a).

For the influence of divalent ions, magnesium, manganese and calcium were tested. Since the optimum concentration of magnesium was found to be 2.5 mM, manganese and calcium were tested with 2.5 mM as well. While conversion with calcium was nearly negligible, manganese was comparable to magnesium (see Figure 3-23b).

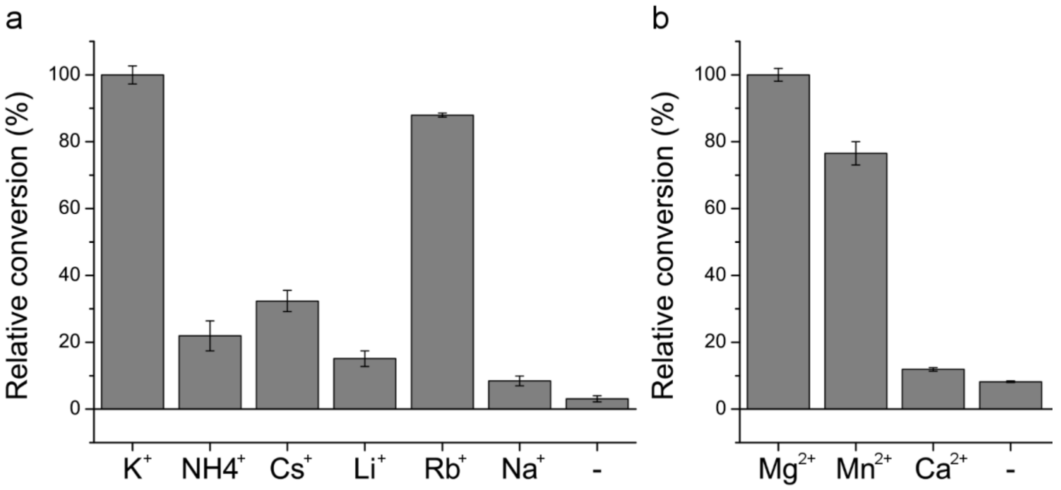

Figure 3-23 – Activation of RBK1 by different mono- and divalent cations. (a) Relative conversion for the activation of RBK1 by different monovalent cations. -: negative control without monovalent ions. Standard deviations are shown for three individual samples. **(b)** Relative conversion for the activation by different divalent cations. -: negative control without divalent ions.

3.2.5 Evaluation of the focused library

The small focused library of the RBK1 variants, which was established by evaluation of the tolerated variations based on a family alignment (René de Jong, DSM) and rational design based on the homology model of the *S. cerevisiae* ribokinase RBK1 (Jannik Vollmer, Institute of Technical Biochemistry, University of Stuttgart), was screened for an altered substrate spectrum compared to the wild-type enzyme. Therefore, a variety of sugars with four to six carbon atoms (C4 to C6) and stereoisomeric differences in the C2 or C3 hydroxyl group position were chosen. Of special interest was the change in activity towards the D-ribose epimer D-xylose, which differs in the position of the hydroxyl group at the C3 atom, since the initial aim of the project was the phosphorylation of this sugar for a novel D-xylose metabolizing pathway in yeast.

3.2.5.1 Initial screening

All variants from the focused library were initially screened under the same conditions (75 mM triethanolamine buffer at pH 7.5 including 5 mM potassium phosphate and 5 mM $MgCl_2$) using the coupled NADH assay at 22°C. Hereby, the activity of RBK1 was evaluated either with 10 mM D-ribose or 50 mM D-xylose. It was found that all triple and quadruple mutant variants did not show any activity or less than 1% relative to wild-type activity towards both substrates. Also, the mutations I101T and I113T in the double variants resulted in a complete loss of activity. For the I101S mutation, only the double variants I101S/I113V and I101S/T279S still showed activity towards D-xylose (see Figure 3-24b). However, the activity was reduced to 72.6 ± 1.6% and 56.9 ± 3.9%, respectively, compared to the wild-type activity. While all the double variants with a I101V and I113V mutation (except with a I101S, I101T or I113T mutation in combination) still showed activity towards D-ribose and D-xylose, only the double variant I101V/I113V showed higher activity than the wild-type towards D-ribose (116.3 ± 1.8%). From the eight single mutation variants, all showed at least some activity towards D-ribose and D-xylose. Especially the more conservative mutations I101V, I113V and A319G, but also A319S showed higher activity than the wildtype towards D-ribose (see Figure 3-24a). For the activitiy versus D-ribose, the best performing single variant was I113V. It showed an relative activity of 173.3 ± 2.4%.

From all the variants, only the single variant I101V showed marginally higher activity towards D-xylose than the wild-type (113.2 ± 3.7%). Therefore, the targeted increase in activity towards D-xylose in the order of two magnitudes could not be achieved with the focused library. Furthermore, the aim to generate as many active variants as possible was also not achieved, since 72% of the variants exihibited completely abolished or strongly diminished activity towards D-ribose and D-xylose.

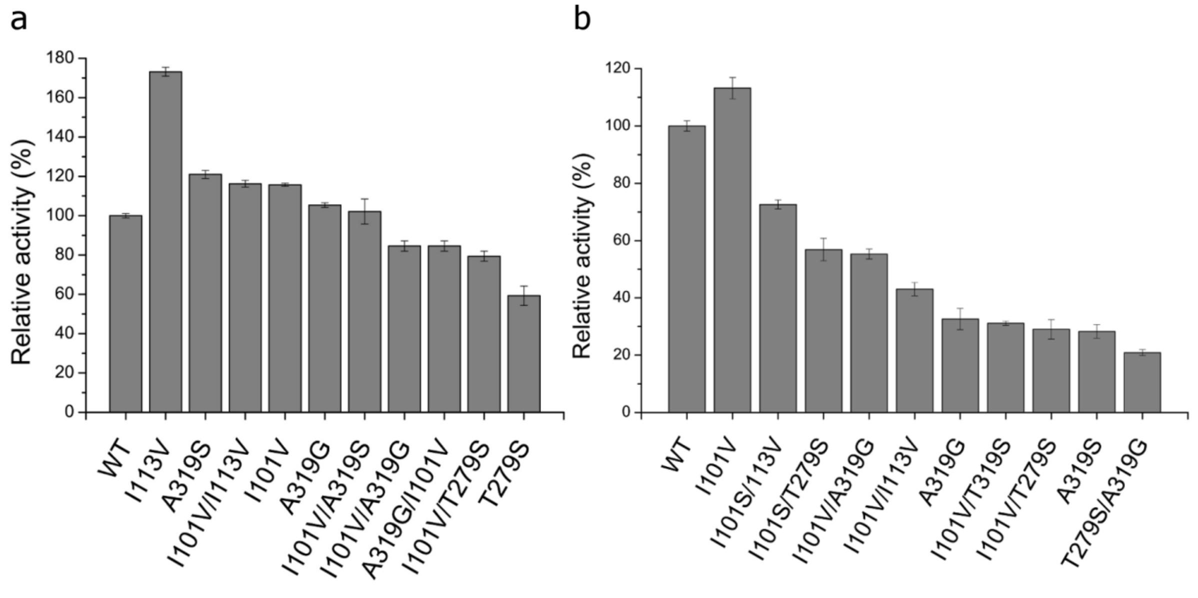

Figure 3-24 – Initial activity screening of the RBK1 focused library towards (a) 10 mM D-ribose and (b) 50 mM D-xylose. The ten best variants are shown in comparison to the wild-type (WT). Standard deviations are shown for at least three individual samples and were calculated from standard deviations in the activity assay and concentration measurements.

3.2.5.2 Phosphate dependency

The best performing mutants from the initial screen were further characterized. Interestingly, under conditions optimized for the wild-type (100 mM $\mathrm{KP_i}$ at pH 6.0 with 2.5 mM $\mathrm{MgCl_2}$) the increase in activity towards 10 mM D-ribose, especially of the I113V variant, could not be reproduced. Instead, an activity comparable to the wild-type activity was found (80.8 ± 1.1 s^{-1} versus 82.3 ±0.3 s^{-1} for the wild-type). A more detailed screen at various P_i concentrations showed that the variant I113V reached its maximum screened activity already at 50 mM P_i (84.9 ± 0.4 s^{-1}) and was almost fully activated at 5 mM and 25 mM P_i (78.0% and 97.9%, respectively, of the maximum activity) (see Figure 3-25a). This behavior was also found for the variants I101V, A319G and A319S, although to a lesser degree starting at 25 mM P_i (70.6%, 92.0% and 92.4%, respectively). In contrast, the wild-type needed 100 mM P_i for the maximum screened activity and only reached 46.4% relative activity at 5 mM P_i. Also, while variants I113V, A319G and A319S reached the maximum screened activity at 50 mM P_i (84.9 ± 0.4 s^{-1}, 78.5 ± 2.2 s^{-1} and 88.2 ± 3.0 s^{-1}, respectively), the wild-type and variant 101V reached it at 100 mM P_i (82.3 ±1.2 s^{-1} and 85.6 ± 3.8 s^{-1}, respectively). As the enzymes where diluted 100 to 300 times in the corresponding buffer, buffer effects from the enzyme preparations were unlikely. Still, to preclude any effects from this, different enzyme preparations from each variant were tested.

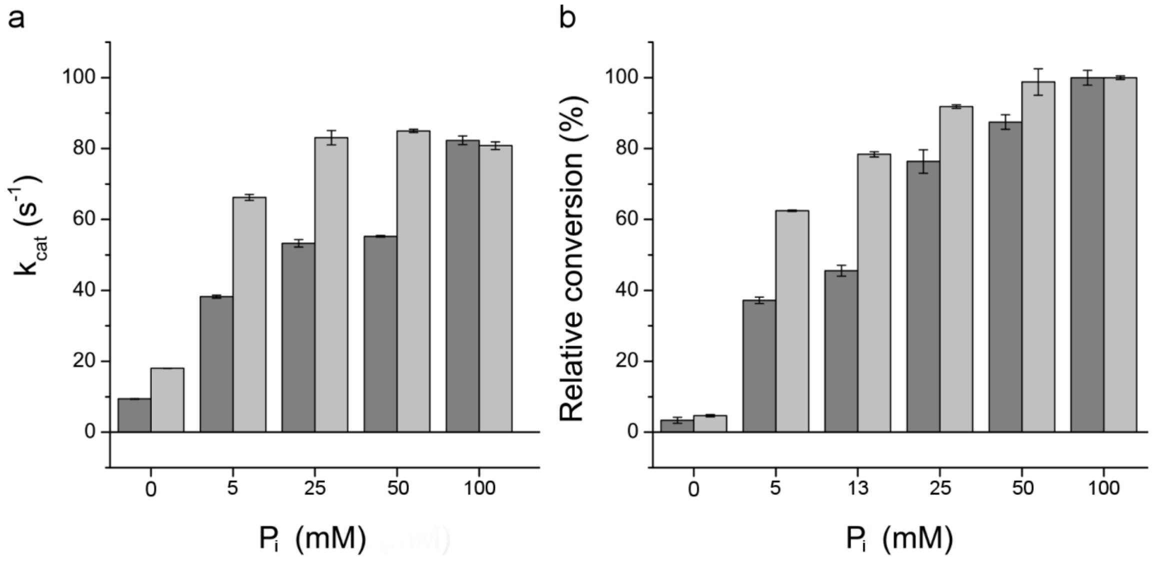

Figure 3-25 – Phosphate dependency of RBK1 wild-type and I113V variant. (a) Turnover number k_{cat} as a function of P_i concentration. Standard deviations are shown for three individual samples. *Dark grey columns*: RBK1 wild-type; *light grey columns*: RBK1 I113V variant. **(b)** Relative conversion as a function of P_i concentration. Standard deviations are shown for three individual samples. *Dark grey columns*: RBK1 wild-type; *light grey columns*: RBK1 I113V variant.

Since PEP present in the coupled NADH assay can facilitate activation of RBK1 in a similar manner as P_i, the wild-type and I113V mutant were also evaluated in biotransformations with different P_i concentrations. A similar behavior was found verifying the results from the coupled NADH assay (see Figure 3-25b). Lacking the activating effect of PEP in the biotransformations, a lower relative conversion was found at 0 mM P_i when compared to the coupled NADH assay.

3.2.5.3 Substrate spectrum of the focused library variants

While human ribokinase and ribokinase from *E. coli* are both highly specific for the phosphorylation of D-ribose, they accept a number of other sugar substrates with differences in the C2 or C3 functional group (e.g. replacement of the hydroxyl group by a proton, altering the conformation from *ribo* to *arabino* or *xylo*) or chain length (tetroses or hexoses) [143,146]. These include for example 2-deoxy-D-ribose, D-arabinose, D-xylose and D-fructose. However, in all cases higher K_m values indicated lower substrate affinities and the total turnover numbers k_{cat} were about two orders of magnitude lower compared to D-ribose (about one order of magnitude for the *E. coli* ribokinase towards 2-deoxy-D-ribose). The focused library of RBK1 was established to aim for a shift in the substrate specificity especially towards the C3 epimer D-xylose. While the initial screening of the focused library indicated only a minor increase in activity towards D-xylose, a verification of these results with biotransformations and HPLC as well as LC/MS analytics was performed.

While 2-deoxy-D-ribose was tested at 10 mM concentration, all further sugar substrates were tested at a concentration of 50 mM in 100 mM KP_i buffer including 2.5 mM $MgCl_2$. The wild-type showed high substrate specificity for D-ribose and only minor activities towards other sugar substrates (see Table 3-2). Although RBK1 from *S. cerevisiae* accepted similar sugar substrates including D-xylose, D-arabinose, D-erythrose and 2-deoxy-D-ribose, the activity towards these substrates was still one to two orders of magnitude lower compared to the *E. coli* ribokinase [146]. Interestingly, from the selected variants tested with these sugar substrates, several showed an increased activity. The I101V variant showed an increased activity towards 2-deoxy-D-ribose (0.57 ± 0.00 s^{-1}). Furthermore, while the wild-type showed no detectable activity with D-glucose, D-fructose, D-arabinose, D-lyxose and D-threose, the variants I101S/I113V and I101S/T279S exhibited low activity with these substrates with the highest increase in activity towards D-glucose (0.33 ± 0.00 s^{-1} and 0.16 ± 0.00 s^{-1}, respectively). For D-xylose, the results from the initial screening regarding the I101V variant could be verified with the conditions optimized for the RBK1 wild-type. The I101V variant showed an activity of 0.24 ± 0.00 s^{-1} at 50 mM D-xylose, which was approximately twice the activity of the wild-type (0.11 ± 0.00 s^{-1}) (see Table 3-2). Still, the highest turnover numbers were not reached for both until a concentration of 300 mM D-xylose (0.76 ± 0.00 s^{-1} and 0.38± 0.00 s^{-1}, respectively) (see also 3.2.6).

Table 3-2 – Substrate spectrum of RBK1 wild-type and variants. D-ribose and 2-deoxy-D-ribose were used at a concentration of 10 mM, all other sugars at 50 mM. Standard deviations are shown for at least three individual samples. -: no activity detected.

Substrate	**k_{cat} [s^{-1}]**			
	Wild-type	**I101V**	**I101S/I113V**	**I101S/T279S**
D-ribose	82.6 ± 1.2	75.8 ± 2.3	1.45 ± 0.04	1.29 ± 0.03
2-deoxy-D-ribose	0.29 ± 0.00	0.57 ± 0.00	-	0.14 ± 0.00
D-xylose	0.11 ± 0.00	0.24 ± 0.00	0.10 ± 0.00	0.08 ± 0.00
D-lyxose	-	-	0.06 ± 0.00	0.04 ± 0.00
D-arabinose	0.02 ± 0.00	0.07 ± 0.00	0.05 ± 0.00	0.05 ± 0.00
L-arabinose	-	-	0.07 ± 0.00	0.06 ± 0.00
D-glucose	-	-	0.33 ± 0.00	0.16 ± 0.00
D-fructose	-	-	0.07 ± 0.00	0.06 ± 0.00
D-erythrose	0.04 ± 0.00	-	0.10 ± 0.00	0.09 ± 0.00
D-threose	-	-	0.07 ± 0.00	0.07 ± 0.00

For validation of the results from the coupled NADH assay, biotransformations of D-xylose and D-glucose with RBK1 were performed. Here, D-xylose and D-glucose showed different retention times (11.9 min and 11.2 min, respectively) in the standard HPLC analytic than D-ribose (13.8 min) and were also identified with standards.

With 500 mM D-xylose an initial conversion rate of about 165 µM min^{-1} and 80 µM min^{-1} for 3 µM of the I101V variant and wild-type, respectively, was found at 30°C. This corresponds to a turnover number k_{cat} of 0.93 and 0.45, respectively, validating the rates from the coupled NADH assay (0.98 ± 0.05 and 0.44 ± 0.02, respectively) at 30°C (see Table 3-3). The variants I101S/I113V and I101S/T279S were tested at a D-glucose concentration of 10 mM since preliminary tests indicated saturation with substrate already at this concentration. Initial conversion rates of 65 µM min^{-1} and 30 µM min^{-1} were found corresponding to the rates calculated with the coupled NADH assay (0.38 ± 0.06 and 0.20 ± 0.03, respectively.

Table 3-3 – Comparison of the reaction rates from the coupled NADH assay and initial conversion rates from biotransformations with selected RBK1 variants and various substrates. [a] Activity and initial conversion rates were measured at 30°C. [b] Initial conversion rates of D-ribose are shown for 30 nM enzyme. [c] Conversion rates towards D-xylose and D-glucose are shown for 1 µM enzyme.

Substrate	RBK1 variant	rate [s^{-1}][a]	conversion [µM min^{-1}][a]
D-ribose	**Wild-type**	110.3 ± 2.5	200 [b]
	I113V	106.2 ± 4.5	195 [b]
D-xylose	**Wild-type**	0.44 ± 0.02	80 [c]
	I101V	0.98 ± 0.05	165 [c]
D-glucose	**I101S/I113V**	0.38 ± 0.06	65 [c]
	I101S/T279S	0.20 ± 0.03	30 [c]

A qualitative verification of the phosphorylated sugar products was performed with LC/MS analysis. Therefore, the compounds were determined on an Agilent 1260 Infinity series LC system equipped with an Aminex HPX-87H column (Bio-Rad, Munich, Germany) and detected with a DAD and MSD. All retention times and mass spectra were verified with standard substrates. The sugar substrates D-ribose, D-xylose and D-glucose were detected as $[M+NH_4^{+}]^{+}$ and $[M+K^{+}]^{+}$ ions in positive scan mode and $[M+Cl^{-}]^{-}$, $[M-H^{+}+HCO_2H]^{-}$ and $[M-H^{+}]^{-}$ ions in negative scan mode (see Figure 3-26 and Table 3-4). Here, the dominant ions in positive scan mode were the ammonia adducts $[M+NH_4^{+}]^{+}$ with relative abundancies of over 95% for all substrates. The dominant ions in negative scan mode were the formic acid adducts, also with relative abundancies over 95%. The co-substrate ATP and the phosphorylated products were only detected as $[M-H^{+}]^{-}$ ions in negative scan mode (see Table 3-4 and Figure 3-26) with relative abundancies over 95%. For all biotransformations with the sugar substrates D-ribose, D-xylose and D-glucose, negative controls without enzyme did show the respective substrate ions, but did not show the respective product ions (data not shown). A qualitative detection and increase of the integrated total ion chromatogram

(TIC) signals of product ions and decrease of the TIC signals of substrate ions were found in samples from biotransformations with enzyme (data not shown).

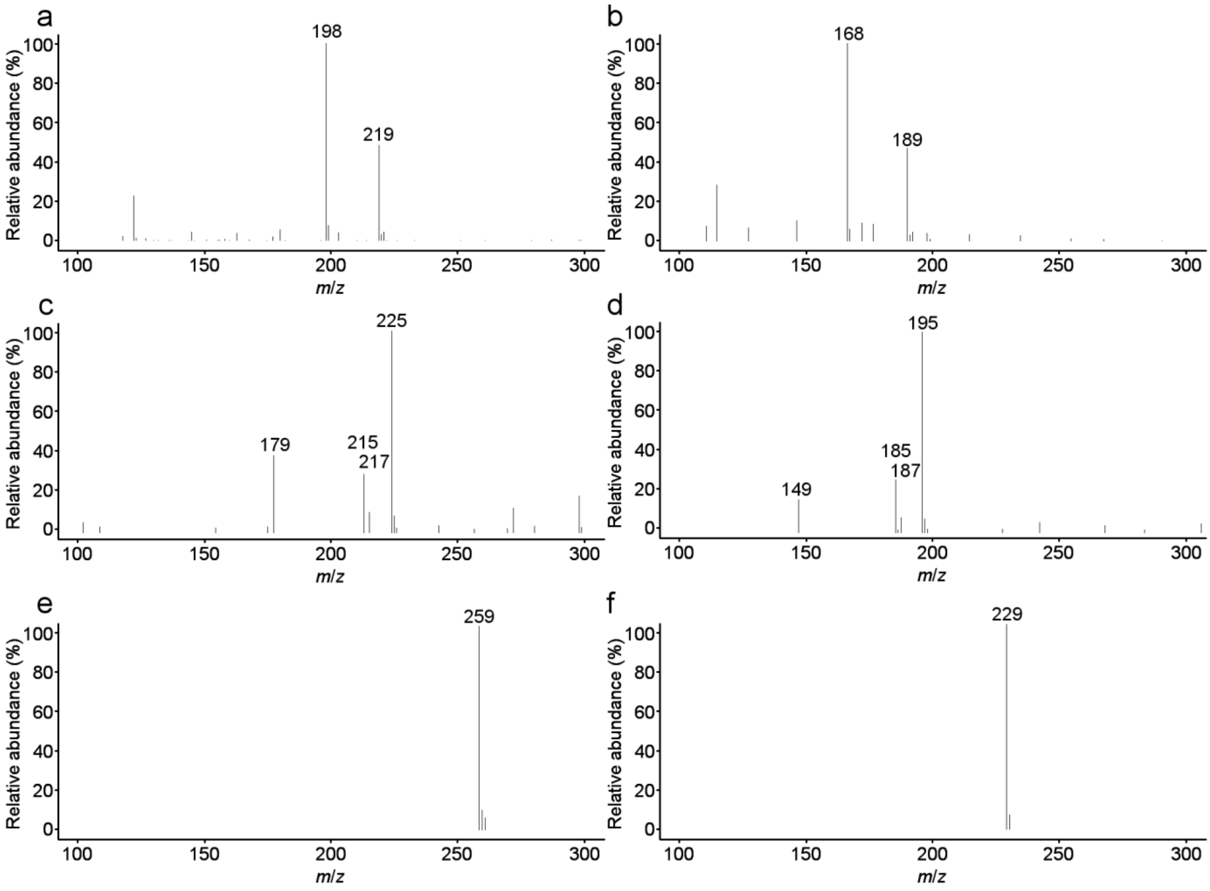

Figure 3-26 – Mass spectra from LC/MS analysis. (a) Mass spectrum of D-glucose in positive scan mode. **(b)** Mass spectra of D-xylose in positive scan mode. **(c)** Mass spectra of D-glucose in negative scan mode. **(d)** Mass spectra of D-xylose in negative scan mode. **(e)** Mass spectra of D-glucose-6-phosphate in negative scan mode. **(f)** Mass spectra of D-xylose-5-phosphate in negative scan mode. The figures were created from the original mass spectra using GIMP2.0 software.

Table 3-4 – Adducts and ions of reaction substrates and products formed during LC/MS analysis.

Substrates/ Products	Retention time (min)	Adducts/ions formed (m/z)				
		$[M+NH_4^+]^+$	$[M+K^+]^+$	$[M+Cl^-]^-$	$[M-H^++HCO_2H]^-$	$[M-H^+]^-$
ATP	5.3	-	-	-	-	506
D-ribose	8.9	168	189	185/187	195	149
D-xylose	7.9	168	189	185/187	195	149
D-glucose	7.4	198	219	215/217	225	179
D-ribose-5-P	5.0	-	-	-	-	229
D-xylose-5-P	4.9	-	-	-	-	229
D-glucose-6-P	4.3	-	-	-	-	259

3.2.6 Kinetic data of RBK1

Since the rational design approach with the focused library of RBK1 variants concentrated on the sugar substrate binding site, not only an alteration in the maximum velocity of the enzyme, but also in the substrate affinity was to be expected. Therefore, kinetic data of the wild-type and selected variants for the natural substrates D-ribose and ATP as well as for 2-deoxy-D-ribose, D-xylose and D-glucose were collected. When possible, the wild-type was compared to the best performing mutant versus the various substrates.

While the maximum activity of the RBK1 wild-type and the I113V variant at high phosphate concentrations (100 mM KP_i at pH 6.0 including 2.5 mM $MgCl_2$) were comparable (82.3 ± 1.2 s^{-1} and 80.8 ± 1.1 s^{-1}, respectively), the K_m for D-ribose increased significantly from 0.89 ± 0.02 mM for the wild-type to 2.47 ± 0.05 mM for the I113V variant at a constant ATP concentration of 1.0 mM (see Table 3-5 and Figure 3-27a). This resulted in a decrease of the catalytic efficiency for the variant by a factor of about three. Concomitantly, the increase in the K_m value was also reflected in the higher inhibition constant K_{iu}. However, the lower substrate inhibition did not lead to significantly higher activity at high D-ribose concentrations over 15 mM, but resulted in a stable activity around 80.0 ± 1.9 s^{-1}. For a detailed assessment of the kinetic data regarding this bi-substrate, enzymatic reaction, the influence of the ATP concentration at a constant D-ribose concentration (10 mM) was investigated as well. Here, no significant change in the K_m value was found (see Table 3-5 and Figure 3-27b).

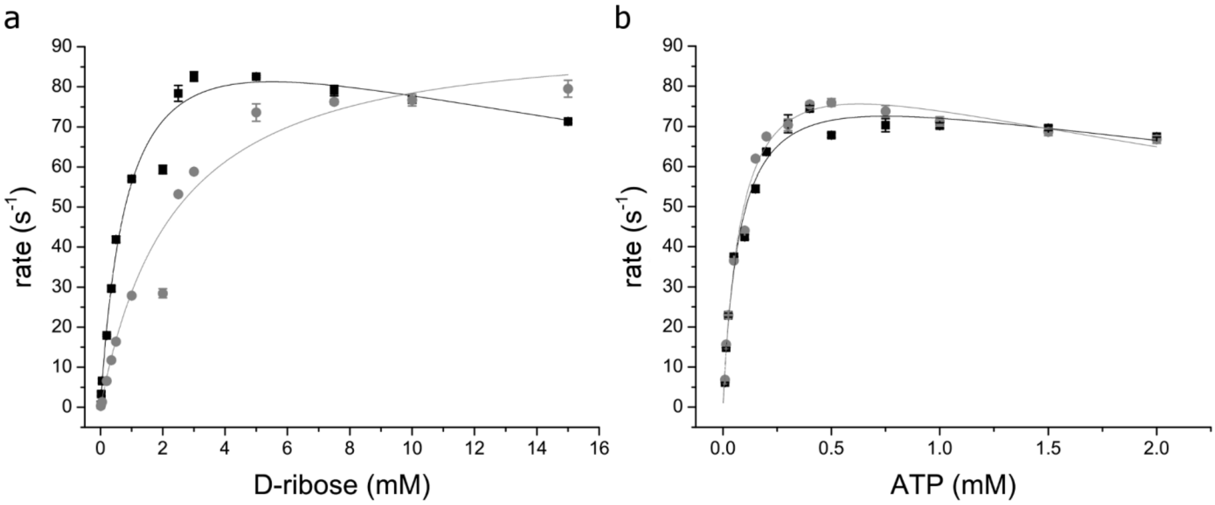

Figure 3-27 – Michaelis-Menten kinetics of RBK1 wild-type and I113V variant with D-ribose (a) and ATP (b). *Black squares*: experimental values for the RBK1 wild-type; *grey circles*: experimental values for the RBK1 I113V variant; *black line*: non-linear regression calculation of the Michaelis-Menten kinetics for the RBK1 wild-type; *grey line*: non-linear regression calculation of the Michaelis-Menten kinetics for the RBK1 I113V variant. Standard deviations are shown and calculated from at least three samples.

Since the I101V variant performed slightly better than the wild-type towards D-xylose in the initial screening, the kinetic data for both were collected. All measurements were performed in the optimized buffer (100 mM KP_i at pH 6.0 including 2.5 mM $MgCl_2$) and showed no significant change in the K_m values (170 ± 2 mM versus 148 ± 2 mM for the wild-type) (see Table 3-5). Still, the I101V variant performed better than the wild-type. The k_{cat} value at 500 mM D-xylose was 0.77 ± 0.00 s^{-1} and 0.38 ± 0.01 s^{-1} for the I101V variant and the wild-type, respectively.

Comparable to D-xylose, a slight increase in activity towards 2-deoxy-D-ribose was seen for the I101V variant in the screening at 50 mM substrate concentration (5.2 ± 0.2 s^{-1} versus 4.1 ± 0.3 s^{-1}). However, in contrast to D-xylose, kinetic data showed that the I101V variant actually has a better substrate affinity with a K_m value of 163 ± 17 mM compared to the wild-type (275 ± 58 mM) (see Table 3-5). This is about two orders of magnitude higher compared to the K_m value of the wild-type towards D-ribose (0.89 ± 0.02 mM).

In the evaluation of the substrate spectrum of the focused library, variants I101S/I113V and I101S/T279S showed activity towards substrates that were not phosphorylated by the wild-type, e.g. D-glucose and D-fructose. The variant I101S/I113V was further investigated with D-glucose as substrate. Here, a K_m of 0.72± 0.01 mM was found, which is in the range of the affinity for the natural substrate D-ribose (see Table 3-5). Since the wild-type enzyme did not show a measurable activity towards D-glucose, no kinetic data could be collected.

Table 3-5 – Kinetic data with RBK1 wild-type and selected variants with various substrates. The k_{cat} values shown are the highest measured rates in the standard setup. The K_m and K_{iu} values are theoretical values calculated from the non-linear regression of the kinetic data collected. The r^2 values of the fit of the non-linear regression to the experimental values were always at least 0.97. N/A: no values available.

Substrate	RBK1 variant	k_{cat} [s^{-1}]	K_m [mM]	K_{iu} [mM]	k_{cat}/K_m [s^{-1} mM^{-1}]
ATP	**Wild-type**	74.5 ± 0.7	0.079 ± 0.000	7.06 ± 0.39	945.7 ±4.3
	I113V	76.0 ± 1.0	0.088 ± 0.001	4.52 ±0.07	862.0 ± 0.2
D-ribose	**Wild-type**	82.6 ± 1.2	0.89 ± 1.2	33.9 ± 3.1	92.3 ±1.1
	I113V	79.5 ± 2.1	2.47 ± 0.05	416.5 ±1.5	32.2 ± 0.3
D-xylose	**Wild-type**	0.38 ± 0.0	148 ± 2	N/A	0.002
	I101V	0.77 ± 0.0	162 ± 2	N/A	0.005
2-deoxy-D-ribose	**Wild-type**	4.1 ± 0.3	275 ± 58	N/A	0.015
	I101V	5.2 ± 0.2	163 ± 17	N/A	0.032
D-glucose	**Wild-type**	0.00 ± 0.00	N/A	N/A	N/A
	I101S/I113V	0.35 ± 0.0	0.72 ± 0.01	N/A	0.49 ± 0.00

3.2.7 Crystallization and x-ray diffraction of RBK1

In order to gain structural information of the yeast ribokinase RBK1, the protein was expressed and purified for crystallization trials (see 3.2.3). The apparently pure protein was concentrated, filtrated and mixed with the substrates D-ribose, D-xylose, or AMP-PNP when necessary. The protein was crystallized in an initial screening at the Biochemiezentrum, University Heidelberg (Dr. Jürgen Kopp). Further optimization of crystals was performed using the “Optimizer and Additive” screen. Based on the initial screening, a home-made fine screen was performed at the University of Hohenheim, Stuttgart. Crystals of apparently good quality appeared already after a few hours in the different screens. Various crystals from the initial and the fine screening were analyzed at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. Data of well diffracting crystals were collected and processed (Dr. Daniele de Sanctis, ESRF, Grenoble, and Dr. Domenico Lupo, University of Hohenheim, Stuttgart). The best data were obtained from crystals grown from the initial screen (0.18 M ammonium citrate and 20% (w/v) PEG 3350 at pH 7) and the fine screen (0.2 M potassium citrate and 20% (w/v) PEG 3350) with AMP-PNP and D-xylose or D-ribose as ligands, respectively. The data are summarized in Table 3-6. However, it was not possible to solve the structure using molecular replacement based on the available data so far [216].

Table 3-6 – Data collection during x-ray diffraction for RBK1.

Data collection	RBK1 (fine screen)	RBK1 (initial screen)
Space group	C222	I121
Cell parameters Axis a, b and c (in Å) Angles α, β and γ (in °)	 78.435; 168.5; 58.949 90.00; 90.00; 90.00	 123.26; 79.806; 160.75 90.00; 90.40; 90.00
Beamline	ESRF ID 29	ESRF ID 29
Wavelenght (Å)	1.06124	1.06124
Resolution (Å)*	48-2.8 (2.9-2.8)	49-2.3 (2.4-2.3)
Completeness (%)*	99.5 (99.1)	98.5 (97.3)
R_{sym} (%)*	4.5 (79.0)	7.5 (52.0)
I/σ*	18.0 (1.6)	6.0 (1.5)
Reflections (Total/Unique)	45,334/10,068	214,572/68,932

3.3 Thiol-functionalization of acrylic ester monomers catalyzed by immobilized *Humicola insolens* cutinase

Functionalized acrylic monomers are important precursors for refined polymers used in various industrial applications. Enzymatic processes with biocatalysts like lipases have been for the production of polymers with thiol and acrylate functional groups have already been established [217]. Cutinases have been evaluated for and applied in various reactions like the polyester synthesis and lactone ring-opening polymerizations [179,180].

In this case study, the cutinase from the ascomycete *Humicola insolens* (HiC) was evaluated as a novel biocatalyst for the synthesis of functionalized acrylic esters by transesterification. In the first stage of the project [198], the expression and immobilization of the enzyme on a microporous resin was optimized. Initial transesterification reactions with the model substrates 6-mercapto-1-hexanol and methylacrylate were performed (see Figure 3-28a) and the influences of water concentration and by-product inhibition as well as the temperature optimum were evaluated. For a comprehensive analysis, the substrate spectrum (see Figure 3-28b) and the influence of a radical inhibitor on by-product formation were investigated in the present study. Furthermore, the scope of this work involved the purification and identification of the main product as well as kinetics of HiC in hydrolysis reactions.

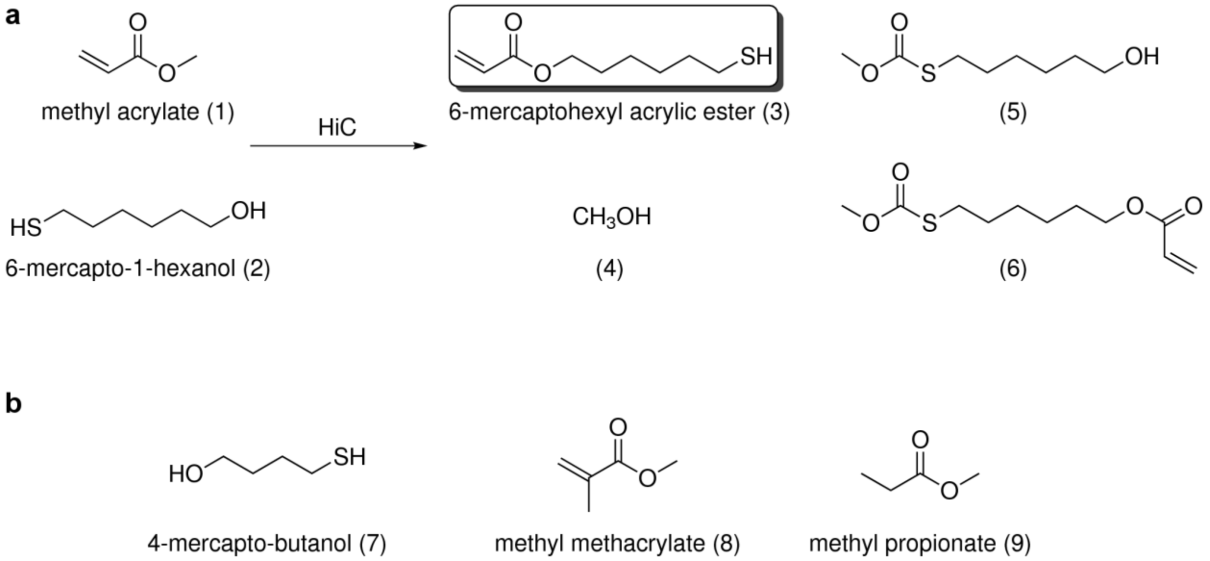

Figure 3-28 – Transesterification of methyl acrylate and 6-mercapto-1-hexanol by HiC (a) and further substrates tested (b). (4) methanol, (5) methyl-3-(6-hydroxyhexylthio)propanoate, and (6) 6-(3-methoxy-3-oxopropylthio)hexyl acrylate.

3.3.1 Preparation of immobilized *Humicola insolens* cutinase

Expression of HiC was performed with a *P. pastoris* strain, which had earlier been transformed with the pPICZαA_HiC construct [198]. The expression in shaking flasks resulted in volumetric yields of secreted HiC of up to 200 mg L^{-1} with a purity of over 90% without any purification steps (see Figure 3-29).

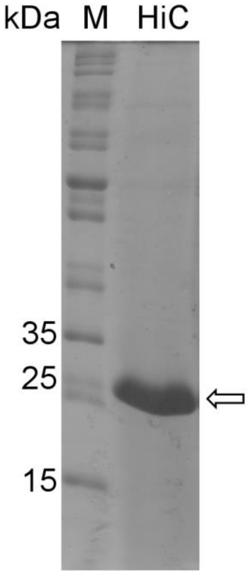

Figure 3-29 – Expression of HiC from *P. pastoris* and analysis with SDS-PAGE. The *white arrow* indicates the molecular size of the HiC protein band at about 22 kDa. *Lane M* represents a molecular size marker. Ten µg of total protein were loaded.

For the use of HiC in transesterification reactions in organic media or substrate, the enzyme was immobilized on the polystyrenic, microporous resin Diaion HP20 [218]. After incubation of the resin with the enzyme, filtration and drying for 72 h at room temperature under vacuum to achieve a water activity towards zero, the immobilisate was stored at 4°C until further use.

3.3.2 Kinetic data of *Humicola insolens* cutinase

Before HiC was used in transesterification reactions for the thiol-functionalization of acrylic esters a comprehensive biochemical characterization of HiC was performed also collecting kinetic data with various *para*-nitrophenyl (*p*NP) esters. Here, HiC showed a preference for shorter chain substrates (C4 and C8) with higher substrate affinity and turnover numbers (see Table 3-7) [219]. This was comparable to literature data on other cutinases from e.g. *Fusarium solani pisi* and *Alternaria brassicicola* [220–222]

Table 3-7 – Kinetic parameters of HiC.

Substrates	k_{cat} [s^{-1}]	K_m [µM]	k_{cat}/K_m [s^{-1} mM^{-1}]
***p*NP-butyrate (C_4)**	191 ± 6	454 ± 29	0.42
***p*NP-octanoate (C_8)**	187 ± 24	207 ± 24	0.90
***p*NP-palmitate ($C_{18:1}$)**	18 ± 1	2246 ± 347	0.0082

3.3.3 Transesterification reactions and substrate specificity

For comparison with earlier results [198,219] and to validate the functionality of the newly prepared batch of immobilized HiC, 10 mg of the immobilisate were used in a transesterification reaction under previously optimized conditions (methyl acrylate (300 µL, 10.3 M), 6-mercapto-1-hexanol (20 µL, 275 mM), 0.025% (w/w) deionized water, 40°C, 700 rpm, 6 h). The formation of 83.9 ± 0.9% 6-mercaptohexyl acrylic ester corresponded well with earlier results (84.1 ± 1.7%) validating the functional expression, reproducible immobilization of HiC on Diaion HP20, and activity in transesterification reactions [198,219].

Based on the results from earlier experiments regarding the substrate specificity of HiC, it was found that the sterically less demanding methyl acrylate (1) was favored over methyl methacrylate (8) as acyl donor in transesterifications with 6-mercapto-1-hexanol (2) and 4-mercapto-1-butanol (7) [198,219] (see Figure 3-28 and Table 3-8). To further evaluate the influence of the predominantly flat substrate geometry of the acrylate, transesterifications were additionally performed with methyl propionate. While transesterifications of methyl acrylate (1) with 6-mercapto-1-hexanol (2) yielded 84.1 ± 1.7% main product after 6 h at 40°C, the transesterification with methyl propionate resulted in 93.2 ± 0.2% product already after 2 h. No Michael-addition products were found in this reaction, which is likely since methyl propionate lacks the acrylic double bond.

Table 3-8 – Substrate specificity of HiC in transesterification with various acyl donors and acceptors (adapted from [219]**.** Reactions were performed with 10 mg immobilisate and 0.025% (w/w) deionized water at 40°C for 6 h for transesterifications of methyl acrylate (1), 24 h for methyl methacrylate (8) and 2 h for methyl propionate (9). -: not detected.

	Ratio substrate to products [%]				
	1+2[a]	**1+7** [b]	**8+2** [c]	**8+7** [d]	**9+2** [e]
Mercaptoalcohol	15.2 ± 1.6	21.3 ± 0.8	55.5 ± 0.3	50.0 ± 1.5	6.8 ± 0.2
Acrylic ester	84.1 ± 1.7	76.3 ± 0.9	43.6 ± 0.1	48.0 ± 1.1	93.2 ± 0.2
Michael-addition #1 [f]	-	-	0.9 ± 0.2	2.0 ± 0.3	-
Michael addition #2 [f]	0.6 ± 0.1	2.4 ± 0.0	-	-	-

[a] Transesterifications of methyl acrylate (1) with 6-mercapto-1-hexanol (2)
[b] Transesterifications of methyl acrylate (1) with 4-mercapto-1-butanol (7)
[c] Transesterifications of methyl methacrylate (8) with 6-mercapto-1-hexanol (2)
[d] Transesterifications of methyl methacrylate (8) with 4-mercapto-1-butanol (7)
[e] Transesterifications of methyl propionate (9) with 6-mercapto-1-hexanol (2)
[f] Michael addition #1 and #2 refers to product (5) and (6) and corresponding Michael-addition products formed in transesterification of (1) + (7), (8) + (2), and (8) + (7), respectively.

3.3.4 Influence of a radical inhibitor on transesterifications

For a further optimization of the transesterification of methyl acrylate with 6-mercapto-1-hexanol, the influence of the radical inhibitor butylated hydroxytoluol (BHT) on the formation of by-products and overall conversion was studied. The addition of 3% (w/w) BHT enhanced the overall conversion in diisopropyl ether as solvent at equimolar concentrations of substrates (150 µM) from 5.9 ± 1.1% to 38.9 ± 2.1% 6-mercaptohexyl acrylic ester. When used in the standard setup in methyl acrylate as the sole solvent, BHT reduced the formation of the Michael-addition product (6) at concentrations of 0.14 to 0.71 M (3 to 15% (w/w acrylic ester)), but also increasingly lowered the conversion of the main product (75.5 ± 0.4% 6-mercaptohexyl acrylic ester with 0.71 M BHT.

3.3.5 Product purification and identification by NMR and FT-IR

For the identification of the main product via nuclear magnetic resonance (NMR) spectroscopy and Fourier transform infrared spectroscopy (FT-IR), a large scale transesterification reaction was performed to gain the necessary amount of product. Analysis by GC/MS showed that in the large scale batch, no Michael-addition products were detected, which might contaminate the main product. The product was then purified with silica column chromatography and the structure of the compound was confirmed by ^{1}H and

^{13}C NMR (see Figure 3-30). Thereby, the acrylic ester was confirmed as the exclusive product leaving the thiol group as the reactive, functional end group.

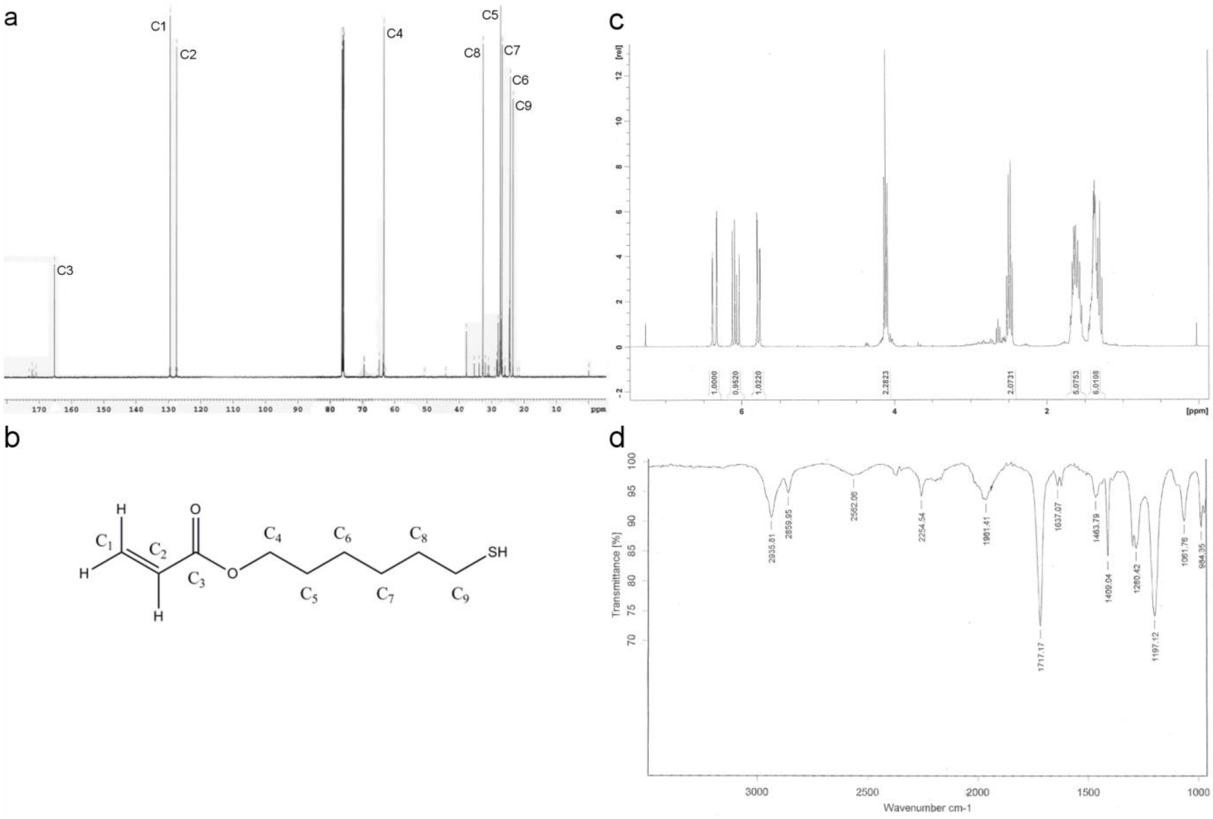

Figure 3-30 – Structure elucidation of 6-mercaptohexyl acrylic ester by NMR and FT-IR. (a) 13C NMR spectrum of 6-mercaptohexyl acrylic ester. C-atoms are marked in the spectrum and in the (b) corresponding structure. (c) 1H NMR spectrum 6-mercaptohexyl acrylic ester. Chemical shifts are marked in the structure on the right. (d) FT-IR spectrum of 6-mercaptohexyl acrylic ester. Chemical shifts (NMR) and wavenumbers (FT-IR): 13C NMR (125 MHz, CDCl3): δ = 23 (C-9), 24 (C-6), 27 (C-7), 28 (C-5), 33 (C-8), 63 (C-4), 128 (C-1), 130 (C-2), 165 (C-3) ppm. 1H NMR (500 MHz, CDCl3): δ = 1.2-1-6 (m, 9 H), 2.4 (dt, J = 7.5 Hz, J = 10.5 Hz, 2 H), 4.1 (t, J = 6.7, 2 H), 5.8 (dd, J =1.3 Hz, J = 10.5 Hz, 1 H), 6.0 (dd, J = 10.5 Hz, J = 17.2 Hz, 1 H), 6.3 (dd, J = 1.3, J = 17.2, 1 H) ppm. FT-IR: 2936 (w), 2860 (w), 2562 (w, S-H), 2255 (w), 1961 (w), 1717 (m), 1637 (w), 1464 (w), 1409 (w), 1280 (w), 1197 (m), 1061 (w), 984 (w).

4. Discussion

4.1 Optimized expression of the dirigent protein *At*DIR6 in *Pichia pastoris* and impact of glycosylation on protein structure and function

In this case study, the goal was to improve the heterologous expression of the dirigent protein *At*DIR6 from *Arabidopsis thaliana* (*A. thaliana*) by using expression hosts such as *Escherichia coli* (*E. coli*) or *Pichia pastoris* (*P. pastoris*) to serve as an example for selected, critical challenges in the production of biocatalysts. The dirigent protein *At*DIR6 derived from a high yield fed-batch fermentation process with *P. pastoris* was studied in regard to the impact of its glycosylation on protein structure, solubility, and function. This exemplifies some challenges in the second step of establishing a biocatalytic process, the characterization of the protein or biocatalyst.

4.1.1 Recombinant protein production of *At*DIR6

The biotechnologically interesting dirigent proteins were so far only produced with insect and cell culture expression systems such as *Spodoptera frugiperda* Sf9 cells (with the baculovirus expression system using *Autographa californica* multiple nuclear polyhedrosis virus) or *Solanum peruvianum* cells [95,105,108,109]. These heterologous expression systems, which are also applied in biotechnology and offer some advantages compared to bacterial expression systems, generally result in low overall yields of the target protein [38,39].

For a high yield production of *At*DIR6, *E. coli* was chosen as the primary expression host. Critical challenges in the expression of this eukaryotic protein in a bacterial expression system were particularly the correct formation of post-translational modifications, which might be necessary for the soluble expression, structural integrity, and function of the protein. First indications for a certain difficulty to express *At*DIR6 in soluble form were made in a previous study using a standard pET expression vector and *E. coli* expression strain [210]. In the same study, refolding of the insoluble *At*DIR6 failed.

For a comprehensive evaluation, several strategies, which addressed different problems in the use of *E. coli* as a heterologous expression system, were tested in the scope of this work. The focus here was on allowing a soluble expression, and not refolding of the protein from inclusion bodies, since this was already shown to be ineffective [210]. This included strategies to allow a correct formation of the putative disulfide bond in *At*DIR6, which was verified only recently [110]. While attempts to introduce a functional glycosylation pathways into *E. coli* have been made [37], this technology is not yet ready for common use and the organism is naturally not able to add glycan structures to proteins. Glycosylation has been shown to be important for the soluble expression, stability and/or bioactivity of many, especially eukaryotic, proteins [30]. Therefore, several basic and more advanced strategies, which have been shown to be advantageous in facilitating a soluble expression of a target protein, have been pursued.

The *Atdir6* gene was synthesized and codon-optimized to enhance expression in *E. coli* [223]. Although solving the problem with rare codons by using synthetic, codon-optimized genes was shown to improve soluble expression of eukaryotic proteins [41], initial attempts to express *At*DIR6 in a soluble form failed. Neither lowering the cultivation temperature nor the variation of inducer concentration, expression time or promoter system, which possibly could all reduce the stress on the folding machinery of the cell, resulted in a soluble expression of *At*DIR6. In some cases, purification tags like the common His$_6$-tag result in a decreased solubility of the target protein and variation of the position (N- or C-terminal) or deletion can have an advantageous effect [53,224]. However, expression of *At*DIR6 without His$_6$-tag from the rhamnose-inducible, very tight expression vector pJOE did not lead to soluble protein. Since the homology model of *At*DIR6 [110] indicated that the N- and C-terminus are not buried in the core of the enzyme, a disruption of the structure by the His$_6$-tag is unlikely. Therefore, the presence or absence of the His$_6$-tag did not seem to have an effect on the soluble expression of *At*DIR6.

The *E. coli* BL21 derivatives C41 and C43 contain a mutation in the T7 polymerase promoter region and additional, uncharacterized mutations shown to be advantageous for toxicity issues and plasmid stability in the expression of recombinant proteins [195,225]. Several cytosolic and membrane proteins from bacterial and eukaryotic origin only expressed in low yields or as inclusion bodies in BL21 could be expressed successfully using these strains. *At*DIR6 was still found in inclusion bodies when expressed in C41 or C43. These results, taken together with the results from the previously mentioned, basic strategies to improve soluble expression, indicated an important role of the post-translational modifications.

Therefore, two strategies to allow the formation of correct disulfide bonds were tested. *At*DIR6 includes two cysteines near the N- and C-terminus and it was recently shown that these form a disulfide bond in the native protein, which was suggested to be structurally relevant [110]. While the expression of soluble proteins featuring correct disulfide bonds in the cytoplasm of *E. coli* can be hampered by the oxidative milieu, this problem can be solved by localizing the target protein into the periplasm. The latter features enzymes involved in the protein quality control like oxidoreductases and isomerases, which allow the correct formation of disulfide bonds [46]. Furthermore, special *E. coli* strains like Origami B have mutations in the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes, which strongly enhance disulfide bond formation in the *E. coli* cytoplasm. These strategies did not solve the problem of *At*DIR6 forming inclusion bodies. While an important effect of the disulfide bond on the structural stability of *At*DIR6 cannot be excluded, its formation was not enough to allow a soluble expression.

Proteins, which are prone to aggregation during heterologous overexpression can be often produced soluble by a translational fusion to highly soluble proteins such as thioredoxin or maltose-binding protein [42,43]. Fusion of *At*DIR6 to the former resulted in an insoluble protein species indicating a strong aggregation propensity of *At*DIR6 expressed in *E. coli*. Also, the combination of this strategy with the expression in *E. coli* Origami B(DE3) did not result in soluble protein.

The majority of proteins in the *E. coli* cytoplasm, especially large multi-domain and overexpressed proteins, do not fold correctly on their own, but require the assistance of folding modulators [31]. Folding modulators include molecular chaperones, which favor productive folding by shielding interactive surfaces from each other and from the solvent as well as folding catalysts accelerating rate-limiting steps, such as the isomerization of peptidyl-prolyl bonds and reshuffling of disulfide bonds. The correct folding and soluble expression of many recombinant proteins was successfully improved by co-expression of the well-studied DnaK/DnaJ/GrpE or GroES/GroEL chaperone systems as well as the *E. coli* trigger factor TF [45]. Indeed, *At*DIR6 was found in the soluble form during a co-expression of DnaK/DnaJ/GrpE in *E. coli* Origami B(DE3) at a low temperature. The protein could be purified by Ni-NTA chromatography, however only low yields (about 40 µg L^{-1}) and purity were achieved. This was most likely connected to the disadvantageous expression conditions at 15°C and co-expression of chaperones leading to a strongly reduced growth rate. However, other expression conditions during co-expression of chaperones, e.g. higher temperatures did not yield soluble protein. Importantly, the partially purified *At*DIR6 exhibited aggregation during storage. Therefore, this aggregation-prone behaviour indicated a further problem with *At*DIR6, which was potentially correctly folded during the expression by the aid of the molecular chaperones.

Based on these data and compared to results from the expression in insect and cell cultures, in which *N*-glycosylation of proteins is possible [95,105,108,109], it was hypothesized that this post-translational modification might be the critical issue in failing to express *At*DIR6 in a soluble form in *E. coli*. This conclusion led to the evaluation of *P. pastoris* as potential expression host.

The rationale for using *P. pastoris* cells in recombinant protein production is mainly based on the simple production and secretion into the supernatant via gene-fusions to the alpha-mating factor pre-pro leader sequence (α-MF) from *Saccharomyces cerevisiae* and the ability of yeast cells to efficiently glycosylate recombinant proteins, although the glycosylation clearly varies from the pattern found in the originating organism [36]. Furthermore, this host is widely used for ultra-high cell density and high yield fermentations [35].

Based on initial results in shaking flasks, a fed-batch fermentation process was established yielding 47 mg L^{-1} *At*DIR6. This corresponds to an overall yield of about 160 mg *At*DIR6 from only one fermentation procedure over 13 days. Compared to plant-based expression systems using transfected *S. peruvianum* cell lines, where only about 200 µg L^{-1} of purified protein could be obtained, this represents a more than 250-fold increase in recombinant protein production. In addition to these high yields, the ease of molecular manipulation and cheap fermentation procedure distinguish *P. pastoris* as the superior expression host for the production of dirigent proteins compared to all previously reported expression systems [95,105,109].

4.1.2 Characterization of *At*DIR6

The positive results with *P. pastoris* as expression host for *At*DIR6 compared to the failure of producing stable protein in *E. coli* again strongly indicated an important role of the glycosylation in this regard. Therefore, a biochemical characterization of *At*DIR6 was performed and the impact of its glycosylation on protein structure and function was evaluated.

First of all, to verify the basic suitability of *P. pastoris* as an expression host for the production of dirigent proteins, the dirigent activity of *At*DIR6 was confirmed by activity assays. For this, a combination of reversed-phase and chiral HPLC methods were used, which were modified from Pickel *et al.* (2010) [109]. While it was shown that the overall pinoresinol formation was increased in the presence of *At*DIR6 derived from *P. pastoris*, only the increase in the enantiomeric excess (*ee%*) of (-)-pinoresinol was routinely quantified. This increase in *ee%* is defined as dirigent activity. *At*DIR6 from *P. pastoris* exhibited 90% of the dirigent activity when compared to the protein derived from *S. peruvianum*. This slight difference in activity may be attributed to a somewhat lower purity and possible degradation products assessed by densitometric analysis of the SDS-PAGE gels as compared to the *S. peruvianum* preparation.

The obtained bioconversion data validate previous work on dirigent proteins, in which higher *ee*-values were observed with increasing amounts of *At*DIR6 and decreasing coniferyl alcohol concentrations [109]. The *ee%* increased from 14.1 ± 2.6% to 89.0 ± 0.5% in the presence of 4 mM coniferyl alcohol and 1 µM or 89 µM *At*DIR6, respectively. When comparing to an industrial process, 89 µM *At*DIR6 would correspond to two mole percent catalyst per mole substrate indicating a high efficiency of the dirigent protein in formation of the product enantiomer. These results are in agreement with the mode of action of dirigent proteins as proposed by Lewis and coworkers, according to which the guided coupling of protein-bound coniferyl alcohol radicals (CA*) competes with unselective coupling of free radicals in solution (as shown in Figure 1-4a) [104,112]. Furthermore, the effective concentration of CA* depends not only on the concentration of dirigent protein and coniferyl alcohol, but also on the concentration of the oxidizing agent. Therefore, as expected, experiments with lower laccase concentrations lead to higher enantiomeric access, which demonstrates that selective coupling is favored at low levels of CA*.

While *At*DIR6 produced by *P. pastoris* and *S. peruvianum* appeared to be functionally equivalent, there are subtle differences, which can most likely be attributed to differences in the glycosylation in the two expression hosts. Molecular mass analysis by MALDI-TOF-MS revealed a mass of 24.6 kDa for the fully glycosylated *At*DIR6 produced by *P. pastoris*. This varied significantly from the mass of 33 kDa estimated by SDS-PAGE analysis. The discrepancy can be explained by the fact that varying conformations of the sugar moieties can alter the interactions with the electrophoresis gel or extent of SDS binding, which finally results in an electrophoretic behavior which does not reflect the correct size of the corresponding protein [226]. Results from analytical gel filtration indicated a homo-tetrameric structure of the protein derived from *P. pastoris* fermentations with a size of about 98 kDa. In contrast, a homo-dimeric structure was shown for *At*DIR6 expressed in *S. peruvianum* and for *Fi*DIR1 purified from *Forsythia intermedia* stem tissue [109,111]. Interestingly, in the first report of dirigent proteins, Davin *et al.* (1997) [95] proposed a homo-trimeric structure for *Fi*DIR1 with a molecular weight estimation of 78 kDa from analytical gel filtration for the native protein based on an observed band at 26 kDa in SDS-PAGE analysis. More recent MALDI-TOF analysis of *Fi*DIR1 verified the actual size of a monomer to be 21 to 23 kDa, which again would hint to a homo-tetrameric structure instead of the earlier proposed structure of a homotrimer (Davin and Lewis 2000). Although there is no final clarification found in the corresponding literature to this discrepancy of the earlier studies and the later established homo-dimeric structure and reaction mode, this indicates a possible dependence of the oligmerization state on other factors. As the oligomerization state may depend on the purification conditions used, the different glycosylation pattern of the expression host, and/or on the handling of the protein during the analysis, the observation of a tetrameric structure for *At*DIR6 derived from *P. pastoris* is not necessarily in conflict with the proposed homodimer as the functional unit of a dirigent protein. Most importantly, even if the oligomerization states of the protein preparations from *P. pastoris* and *S. peruvianum* are different, the function of the protein is not impaired as shown by the bioconversions with *At*DIR6 from both sources.

Analysis of the *N*-glycosylation revealed a high mannose type glycan structure ($Hex_9GlcNAc_2$; about 2 kDa) typical for *P. pastoris* at the two *N*-glycosylation sites N30 and N94. These sites correspond to N59 and N123 reported as glycosylation sites in the plant-derived *At*DIR6 (numbering including the signal peptide) [110]. ESI-MS/MS measurements showed that different isoforms with 9-13 hexose residues exist. This heterologous glycosylation can explain the broad peaks in the MALDI-TOF chromatogram as well as the smear above the major band at 33 kDa in the SDS-PAGE analysis. In contrast to proteins produced by *P. pastoris*, plant-derived *At*DIR6 has paucimannosidic type glycans, which hinders efficient deglycosylation with most of the commercially available enzymes due to the presence of fucose residues at the core of the glycan structure [114]. Consequently, the impact of *At*DIR6 glycosylation on structure, solubility, and function of the protein remained untested so far. With the option to enzymatically deglycosylate *At*DIR6 from *P. pastoris*, it was possible to evaluate the influence of the glycosylation on the native structure and function of the protein.

Interestingly, treatment of *At*DIR6 from *P. pastoris* with the enzyme peptide-*N*-glycosidase F (PNGaseF) resulted not only in the formation of one lower molecular weight band as expected, but of two major bands at 25 kDa and 21 kDa. MALDI-TOF and ESI-MS/MS analysis revealed that the higher band was quantitatively deglycosylated at only one *N*-glycosylation site (N30), while the lower band was deglycosylated at both sites (N30 and N94). The molecular mass of the partially deglycosylated protein (about 23 kDa) still did not correspond with the molecular mass assessed by SDS-PAGE analysis (25 kDa). This could again be addressed to the remaining glycan influencing the electrophoretic behavior as discussed earlier. In contrast, the calculated mass of 21 kDa for the fully deglycosylated protein species corresponded well with its mobility during SDS-PAGE analysis.

Despite the fact that PNGase treatment of native *At*DIR6 did result in a heterogeneous population of partially and fully deglycosylated proteins, it was still possible to address the relevance of glycosylation for dirigent activity. Samples of *At*DIR6 taken during the enzymatic deglycosylation were directly employed in laccase-catalyzed biotransformations of coniferyl alcohol. A clear correlation was found between the loss of dirigent activity and the disappearance of fully glycosylated protein. After 24 h of deglycosylation, 90% of the initial dirigent activity was lost, which corresponded well with the disappearance of 97.9 ± 1.7% of the fully glycosylated *At*DIR6. Importantly, a negative control of *At*DIR6 incubated under the same conditions without PNGase F still showed 81.0 ± 0.8% of the initial dirigent activity after 24 h validating the important effect of the glycosylation on the function of the protein. Interestingly, purified and soluble *At*DIR6 from the *E. coli* expression with chaperones did not show any dirigent activity (0.1 ± 0.2% enantiomeric excess) in bioconversions (data not shown). Since the unglycosylated *At*DIR6 from *E. coli* showed aggregation over time (see 3.1.1), it was used in activity assays directly after a purification procedure before complete aggregation. Taken together, this species behaved comparably to the deglycosylated *At*DIR6 species derived from *P. pastoris*.

A proper glycosylation of *At*DIR6 may thus be necessary for the binding and orientation of CA* to maximize the probability for the specific 8,8´-coupling event yielding (-)-pinoresinol. Alternatively, glycosylation may affect the tertiary structure with an indirect effect on dirigent activity. To test this assumption, CD spectroscopy was performed with glycosylated and deglycosylated *At*DIR6. CD spectroscopy is a powerful and easy to use methodology to assess protein secondary structure information without the need for a crystal structure of the protein [203]. Based on reference databases, the obtained CD spectra can be deconvoluted, which allows an estimation of the secondary structure composition [204]. This analysis revealed that the composition in the glycosylated and degylcosylated forms of *At*DIR6 from *P. pastoris* is almost identical. It can therefore be assumed that both proteins adopt the same overall β-barrel structure as proposed for *At*DIR6 from *S. peruvianum* [110]. However, despite the similarity in secondary structure composition, differences in the shape of the CD spectra hinted at conformational differences between the deglycosylated and glycosylated forms of the protein. This was further supported by the diverse behavior in chemical denaturation experiments. While deglycosylated *At*DIR6 exhibited a gradual increase in ellipticity value at 220 nm, glycosylated *At*DIR6 showed a two-phase behavior with a strong increase in ellipticity during low concentrations of guanidinium hydrochloride indicating a more pronounced change in tertiary structure. Concluding, the results obtained by CD spectroscopy indicated a change in tertiary structure concomitant with a loss in stability upon deglycosylation of *At*DIR6.

Based on the data indicating structural changes upon deglycosylation, it was hypothesized that these changes in tertiary structure could also explain the partial deglycosylation observed under native conditions. Intramolecular conformation changes upon deglycosylation at position N30 of *At*DIR6 could be made accountable for a steric hindrance of PNGase F on the glycan at position N94. However, longer incubation of *At*DIR6 with freshly added PNGase F under denaturated conditions did not support this hypothesis. Thus, the reason for incomplete deglycosylation is still unknown and needs to be further investigated.

In addition to changes in tertiary structure, the deglycosylated protein mixture also exhibited a significant reduction in long-term solubility compared to the fully glycosylated protein, indicating increased protein aggregation over time. This phenomenon again correlates with the results obtained from the expression of *At*DIR6 in *E. coli*, where the obtained, unglycosylated protein is not functional and also shows aggregation over time.

From these results, it can be concluded that only fully glycosylated *At*DIR6 is functional and already the removal of the glycan structure at site N30 induces conformational changes in the monomer structure, which are sufficient for the complete loss in dirigent activity. Whether loss of function is solely due to changes in the protein structure or coupled to higher oligomerization (n>4) preceding the aggregation or a combination thereof remains to be determined.

4.1.3 Outlook

A high yield fed-batch fermentation process for *At*DIR6 was established and the importance of the protein glycosylation for the structure, solubility, and function of the protein was confirmed. With the availability of recombinant protein, a crystal structure could be solved aiming at a better understanding of the reaction mechanism and the molecular basis for enantioselectivity in dirigent proteins. In this regard, a co-crystallization of a stable, radical substrate analogue would be interesting to identify residues involved in substrate binding and conferring enantioselectivity. Crystallization studies based on protein obtained from *S. peruvianum* and *P. pastoris* will be conducted by the group of Prof. Schaller (University of Hohenheim, Germany). Mutation of the N-glycosylation sites or alteration of the glycan pattern using special yeast strains manufactured for glycoengineering could further help in understanding whether the glycans have a direct influence on the activity of the dirigent protein, are imperative for correct folding during expression, effect the solubility of the protein, or a combination thereof. Furthermore, high protein yields also allow a more extensive biochemical characterization of *At*DIR6. This could include an evaluation of the optimum of important parameters such as temperature or the pH value. Also, a more detailed characterization of the kinetics involved in the radical coupling process directed by dirigent proteins might be useful for high space-time yields and low by-product formation in possible biotechnological applications. Rational design based on the newly established homology model [110] could already be useful for studies aiming at an alteration of the substrate spectrum of dirigent proteins for possible biocatalytic applications such as enzymatic steps in the enantioselective synthesis of natural products. In this regard, the evaluation and optimization of protein stability by enzyme engineering or immobilization techniques would be important for the biotechnological use of dirigent proteins.

4.2 Enzyme engineering of *Saccharomyces cerevisiae* ribokinase

In this case study, the aim was to produce the ribokinase RBK1 from *Saccharomyces cerevisiae* (*S. cerevisiae*) and alter its substrate spectrum to serve as an example for selected, critical challenges in the production, characterization, enzyme engineering, and application of biocatalysts. A small focused, library of RBK1 variants was established by rational design based on a homology model of the protein with the focus to alter the substrate spectrum to D-xylose, which might allow a novel pathway for the production of second generation biofuels. RBK1 wild-type and its variants were derived from a high yield, heterologous expression in *E. coli* and characterized for conditions optimized for activity and the substrate spectrum. For a better understanding of the substrate binding, RBK1 was prepared for crystallization and the solution of the structure was attempted.

4.2.1 Recombinant protein production of RBK1

While the first biochemical characterization of a ribokinase was already performed in 1956 [140], the first recombinant overexpression of a ribokinase was performed for the *E. coli* ribokinase 30 years later [142]. In 1998, the crystal structure of this protein was solved up to a resolution of 1.8 Å [135]. For this, the protein was overexpressed in an *E. coli* strain with a non-functional ribose operon and purified to near homogeneity [144]. Typical yields of about 8 mg purified protein per litre cell culture after three purification steps including ammonium sulfate precipitation and ion-exchange chromatography were achieved. Park *et al*. reported the first recombinant expression of a eukaryotic (human) ribokinase using a pET vector construct in *E. coli* BL21(DE3) as a heterologous host [143]. While ribokinase genes were also found in plants and fungi [141,150], the production of RBK1 from *S. cerevisiae* in this study is the first report of a functional, recombinant expression of a fungal ribokinase.

Based on the positive results for the human ribokinase, overexpression of RBK1 was initially tested with *E. coli* BL21(DE3) and a pET16b construct. However, the expression was very weak and the low amounts of produced protein were not soluble. The use of the Rosetta(DE3) strain featuring rare tRNAs, which might help with the codon bias for the *rbk1* gene, which was not codon optimized for the use in *E. coli*, did not alter this result. Since overexpression of an enzyme involved in the central metabolism like RBK1 might be problematic for the expression host, the *E. coli* strain C43(DE3) [195], which is often used for the expression of problematic and toxic proteins, was evaluated as well. Here, strong, yet insoluble expression at 37°C and weak, soluble expression at 20°C was observed. The failure to express RBK1 soluble in high yields in a system comparable to the one used for human ribokinase might be related to the fact that human ribokinase actually shares higher sequence identity with homologs from bacterial and protist species, and not with fungal and plant species [143].

For the characterization and crystallization of RBK1, a robust expression system with high yields was necessary. Therefore, further options were evaluated. Expression of RBK1 was performed using a pJOE vector construct with the *rbk1* gene under the control of the tight rhamnose promoter in C43(DE3) [212–214]. This resulted in a strong expression with excellent yields of 150-200 mg purified protein per litre cell culture and high purity (>95%) after Ni-NTA purification. Compared to the overexpression of the *E. coli* ribokinase in various studies [143,144,146,152], this represents a 15 to 25 times higher yield with comparable purity after only one chromatographic step. All RBK1 variants performed comparably during expression and purification with adequate yield and purity indicating no massive structural changes, e.g. leading to insolubility due to the mutations.

4.2.2 Characterization of RBK1

Purified RBK1 from high yield, heterologous expression in *E. coli* showed a turnover number k_{cat} of 82.3 ± 0.3 s^{-1} versus D-ribose at 22°C in 100 mM potassium phosphate (KP_i) buffer at pH 6.0 with 2.5 mM magnesium chloride ($MgCl_2$). The turnover number of 165.0 ± 4.3 s^{-1} at 37°C was comparable with the highest measured rate of the *E. coli* ribokinase of 130 s^{-1}. Many annotations of genes from plants and fungi as ribokinases are only based on sequence similiarities without available biochemical characterization [143]. Based on the sequence similarity to other characterized ribokinases and the high activity versus D-ribose compared to other sugars (see Table 3-2), it can be assumed that RBK1 from *S. cerevisiae* is a true ribokinase.

For the measurement of activity and kinetic data, a NADH coupled assay was used [142] (see Figure 2-1). The assay was reported to be complicated to use, probably due to the coupled reaction in three steps and the influence of the composition of sample and buffer [144,154]. Other assays were reported such as a radioactive assay [154] and an assay based on a drop in pH [152]. However, these assays have various disadvantages. The radioactive assay is expensive, necessitates an isotope laboratory and is not usable for medium or high throughput screening. The pH assay was shown to have reaction rates, which were lower by a factor of 40 compared to the rates observed with the NADH coupled assay, and reproducibility was questionable. In this light, the NADH coupled assay was evaluated in this study and it was found that it was highly reproducible and amenable to medium or high-throughput screening when performed conscientiously. Since it was already shown that ribokinases are dependent on mono-, di-, and pentavalent ion concentrations for their activity and can also be activitated by the assay component phosphoenol pyruvate (PEP) [152,154], it was especially important to assure a constant assay composition. Various controls were performed to assure that the other enzymatic steps catalyzed by pyruvate kinase and lactate dehydrogenase were not the rate limiting steps under the screened conditions.

In addition to the NADH coupled assay, biotransformations with RBK1 and subsequent analysis by HPLC were performed. A HPLC method using a Bio-Rad Aminex HPX-87H column and a refractive index detector was established allowing the detection of the sugar substrates and phosphorylated products. The initial conversion rates achieved in biotransformations matched well with the activities measured in the NADH coupled assay verifying the performance of the latter.

The RBK1 wild-type was characterized for the optimum concentrations of mono-, di-, and pentavalent ions as well as the pH value. The optimum activity of RBK1 was found at a pH of 6.0. At a pH of 5 and up to pH 9, the relative activity was about 80%. A stronger drop in activity was observed below pH 5. At low pH values, aspartate is protonated and not able to function as the catalytic base anymore. While *E. coli* ribokinase showed a pH optimum around 7 to 8, a lower optimum for yeast RBK1 was expected, since yeast grows well at low pH values and the enzymes generally are adapted to these values. For the optimum ion concentrations, the best activity was found with a concentration of 100 mM potassium phosphate and 2.5 mM magnesium at a pH of 6.0. The optimum concentration of magnesium is in the same range as found for bacterial and mammalian ribokinases [146,154] and activity decreases at higher concentrations. Since the chelate of ATP and magnesium is the real co-substrate for ribokinases [129], the optimum concentration of magnesium is likely to depend on the ATP concentration. However, this was not systematically evaluated. The characterized ribokinases from *E. coli* and *Homo sapiens* show an activation profile for potassium, which is comparable to RBK1. Low concentration of potassium up to 5 mM strongly activates the enzymes, while full activation is achieved at around 100 mM [143,152]. Interestingly, in the case of phosphate activation, five to ten times higher concentrations are needed for RBK1 compared to human and bacterial ribokinase [143,154]. When comparing the results from the NADH coupled assay and biotransformations at low phosphate concentrations, it was clear that PEP can substitute inorganic phosphate as activator to some extent as already found for bacterial ribokinase [154]. In a comparable manner, other mono- and divalent cations can facilitate the activation of ribokinases [146]. For divalent ions, yeast RBK1 behaved as was shown for the *E. coli* ribokinase with manganese facilitating activation, while calcium only showed marginal activation. For monovalent ions, caesium activated slightly more compared to ammonium, which was reversed for the *E. coli* ribokinase. Additionally, rubidium was tested, which activated the yeast enzyme comparably to potassium. It was shown for the *E. coli* ribokinase that the monovalent ion site favors binding of larger ions such as potassium, rubidium, or ammonium [152]. The yeast RBK1 exhibits a high structural similarity compared to the bacterial ribokinase and its binding site seems to favor larger ions as well.

In order to enable a comparative evaluation of the RBK1 wild-type and all variants of the focused library, the stability of the enzyme was assessed. This was important for reproducible results in biotransformations of variants, which were produced at different times. Long- and short-term stability tests over 8 weeks and 24 h, respectively, showed excellent stability at -20°C and 30°C, respectively. The high stability of the enzymes allowed an unproblematic, comprehensive comparison of the focused library variants of RBK1.

4.2.3 Enzyme engineering and evaluation of the focused library

The small focused library was established by a combination of evaluation of a family alignment, which indicated variations that can be tolerated, and rational design based on a homology model, which identified interesting residues involved in substrate binding. The library was screened for an altered substrate spectrum compared to the wild-type enzyme. The wild-type exhibits high activity versus D-ribose (82.3 ± 0.3 s^{-1}) and very low activities towards other pentoses (D-xylose, D-arabinose, and 2-deoxy-D-ribose) and the tetrose D-erythrose. Other tetroses, pentoses, and hexoses are not accepted. This underlines the high substrate specificity of RBK1. Human and bacterial ribokinase exhibit a substrate specificity comparable to RBK1, however *E. coli* ribokinase is able to phosphorylate 2-deoxy-D-ribose quite well (26.7 s^{-1}) [143,146]. The high specificity reflects the importance of ribokinases for the central metabolism.

Initial screening of the complete RBK1 focused library at low phosphate concentrations (5 mM) revealed a few variants with higher activity towards the natural substrate D-ribose including, I101V, I113V, A319G, A319S, and I101V/I113V, while only I101V showed slightly higher activity towards D-xylose. Therefore, the goal to achieve a variant exhibiting an increase in activity towards D-xylose in the range of one to two orders of magnitude to support a possible D-xylose metabolizing pathway could not be achieved with the focused library. Also, the goal to generate as many active variants as possible was not realized, since all triple and quadruple variants exhibited no detectable activity at standard substrate concentrations of 10 to 50 mM. The quadruple mutants were also tested at very high D-ribose concentrations up to 1 M (data not shown). Here, variant I101S/I113V/T279S/A319G exhibited about 50% activity of the wild-type (at 10 mM D-ribose) indicating that multiple mutations negatively affected the substrate binding. All variants showing increased activity are rather conservative, while stronger mutations completely abolished measurable activity. In this regard, additionally to the focused library single and double mutants at positions I101 and I113 were established (data not shown). The isoleucines were exchanged against asparagine residues, which were shown to be frequently involved in sugar substrate binding sites [227]. Here, they can offer bidentate or cooperative hydrogen bonds important for effective binding [228,229]. However, the more drastic exchange to arginine completely abolished activity towards D-xylose as well as D-ribose.

Further evaluation of the best performing variants from the initial screening revealed an altered phosphate dependency, which was strongest for the I113V variant. Compared to the wild-type, the variant was already strongly activated by phosphate at low concentrations of

25 to 50 mM, while the wild-type reaches full activation not until 100 mM. However, at the buffer conditions optimized for the wild-type (100 mM potassium phosphate at pH 6.0, 2.5 mM magnesium chloride), all variants initially screened for a higher activity exhibited only about wild-type activity. The activation by phosphate was already shown to increase the catalytic turnover and the substrate affinity for D-ribose [154]. Therefore, it sounds interesting that the activation profile of enzymes by important ions can be altered in this way. However, for the evaluation of the focused library, this behavior was problematic, since not all variants could be screened at various buffer compositions.

Although the initial goal in dramatically increasing the activity towards D-xylose could not be met with the focused library, variant I101V showed about twice the activity (0.76 ± 0.00 s^{-1}) at a concentration of 300 mM D-xylose compared to the wild-type. The chemical difference between the two epimers D-ribose and D-xylose seems marginal. They only differ at the position of the hydroxyl group at carbon atom three (C3-OH). For the generation of the variants, it was assumed that D-xylose binds in the same orientation and conformation as is known for D-ribose from the *E. coli* ribokinase crystal structure [135]. However, this assumption could be problematic. One way to elucidate the binding orientation is to establish a crystal structure of the yeast RBK1 with D-xylose bound as substrate, which will be discussed later. A problem could be basic thermodynamics of D-xylose in solution. While about 7% of D-ribose in solution is found in the α-furanose form, only under 1% of D-xylose is found in this form [132]. Assuming that D-xylose is also needed in the α-furanose form for efficient phosphorylation, this decreases the available concentration of D-xylose by about an order of a magnitude and thereby increases the apparent K_M value.

For the alteration of the substrate spectrum to other sugars, two interesting variants were identified, I101S/I113V and I101S/T279S. While the wild-type did not accept a variety of tetroses, pentoses, and hexoses, these two variants showed low activities for most of the tested substrates. The more open active site due to the smaller residues serine and valine and new hydrogen bonding possibilities with serine might help better facilitate binding of these non-natural substrates. For more information on the mechanisms, a bioinformatics approach featuring substrate docking studies, at best with a crystal structure of the yeast RBK1 itself, could be insightful.

To support the observed phosphorylation of the non-natural substrates, LC/MS analysis was performed exemplarily with D-ribose, D-xylose, and D-glucose. Samples from biotransformations were analyzed with LC/MS and masses corresponding to the substrates as well as phosphorylated products were identified validating the successful phosphorylation of the above-mentioned substrates.

The hotspot residues identified during rational design of the focused library are found in the active site and are supposed to contribute to binding of the sugar substrate. Therefore, not only changes in the catalytic turnover, but also changes in the substrate affinity were expected. Kinetic data were collected for the wild-type and variants, which showed the best activity towards D-ribose, D-xylose, and D-glucose, namely I113V, I101V, and I101S/I113V, respectively. The substrate affinity for the co-substrate ATP was tested with the wild-type and variant I113V, which showed high activity towards D-ribose. As expected, no changes were found in the affinity for ATP since the mutations were not in the ATP binding site. In contrast, the affinity for D-ribose decreased about three-fold from 0.89 ± 1.2 mM to 2.47 ± 0.05 mM for the I113V variation compared to the wild-type. At a high substrate concentration of 15 mM, the I113V variant showed a higher activity than the wild-type. The reason for this is that RBK1 shows substrate inhibition at higher concentrations of ATP and D-ribose. Therefore, the kinetic data was fitted to a Michaelis-Menten equation describing a reaction with substrate inhibition (see 2.5.1). Substrate inhibition is a form of uncompetitive inhibition leading to a decline in the maximum velocity at high substrate concentrations. The inhibition constant K_{iu} for the I113V variant (416.5 ± 1.5 mM) was approximately an order of magnitude higher compared to the wild-type (33.9 ± 3.1 mM), indicating dramatically decreased substrate inhibition, which could be useful for possible applications at high substrate concentrations. Interestingly, Chuvikowsky *et al.* [146] found a about three-fold lower K_m for D-ribose (0.279 mM) for the ribokinase from *E. coli*, while the K_m for ATP (0.213) was about three-fold higher. It was shown that the activation by mono- and pentavalent ions not only increases catalytic turnover, but also increases the substrate affinity for ATP and D-ribose, respectively [148,152,154]. This complicates collection of kinetic data. It was imperative to use the same setup regarding buffer conditions in all experiments. Since the results for the *E. coli* RBK1 were achieved with other ion concentrations, the results are not necessarily comparable and might also explain differences found for the substrate affinity as well as catalytic turnover.

The kinetic data for D-xylose showed high K_M values of 148 ± 2 mM and 162 ± 2 mM for the wild-type and the I101V variant, respectively. This could be related to the problem, that D-xylose exists only in less than 1% in the α-furanose form, which is necessary for phosphorylation, as discussed earlier. A basic problem with D-xylose is likely that the altered position of the C3-OH group prevents the formation of cooperative and bidentate hydrogen bonds with functional groups of adjacent residues and the backbone such as hydroxyl, carboxy, or amino groups. Again, a bioinformatics approach featuring docking studies and/or elucidation of a crystal structure with D-xylose bound in the active site could help in understanding these issues. While the mutation did not affect the binding affinity for D-xylose compared to the wild-type, it increased the activity by about two-fold. However, this is still about two orders of magnitude lower compared to the natural activity versus D-ribose.

4.2.4 Crystallization of RBK1

Presently, crystal structures are only available for the bacterial ribokinases from *E. coli* and *Staphylococcus aureus* [135,148]. Enzyme engineering by rational design with the aim to alter the substrate spectrum of a given enzyme greatly benefits from structural data including the natural substrate, the targeted substrate, or a transition state analogue bound in the active site [13]. Since no crystal structure for the *S. cerevisiae* ribokinase RBK1 was available, homology modeling based on the *E. coli* structure was performed in this project. However, this led to some limitations like an unclear positioning of the targeted substrate D-xylose. Also, it is not clear whether D-xylose binds in its α-furanose form comparable to D-ribose. In parallel to this approach, the elucidation of the RBK1 crystal structure was promoted. For this, the protein was purified in two chromatographic steps including Ni-NTA and size exclusion chromatography leading to an apparently pure form of the protein. In this regard, analytical gel filtration showed that the protein is a native homodimer as was expected based on biochemical information for other ribokinases. Activity measurements verified the activity after the enzyme was prepared for crystallization trials to verify the structural integrity. Crystals, which were achieved from various screenings, were analyzed by x-ray diffraction and several datasets were collected (Dr. Daniele de Sanctis, ESRF, Grenoble, France). Based on the homologous *E. coli* structure (PDB: 1RKD) [135], molecular replacement was attempted to solve the structure and build a model (Dr. Domenico Lupo, University Hohenheim, Stuttgart). However, this attempt was unsuccessful. This is most likely due to a non-crystallographic symmetry (pseudotranslational) effect leading to ghost peaks in the diffraction pattern [230], which cannot be correctly interpreted by the molecular replacement software. Since the crystals of RBK1 grew very fast (in a few hours), the reason for this could be that several crystal seeds grew together during the crystallization leading to a larger crystal.

4.2.5 Outlook

The established focused library of RBK1 variants was evaluated for alterations in the substrate spectrum and could show potential for further studies and applications. The performed biochemical characterization of the protein regarding optimum pH and ion concentrations as well the high, intrinsic stability of the protein serve as a strong basis for interesting applications. The focused library could be further evaluated for the phosphorylation of modified sugar substrates, which could be used for the enzymatic synthesis of nucleosides and their analogues (see 1.3.5) [146]. In this regard, the established high yield production of ribokinase could serve to eliminate the limitation of recombinant ribokinase availability for the chemo-enzymatic synthesis of phosphoribosyl-5-pyrophosphate (PRPP) [157]. For a further pursuit of the work regarding the novel D-xylose metabolizing pathway, a crystal structure with D-xylose bound as ligand would be extremely helpful. While initial attempts to solve a structure by molecular replacement were unsuccessful, further options are still in process. This includes a heavy metal soaking of high-quality crystals to allow another solution of the phase problem not involving molecular replacement [230]. An alternative approach would be the production of a seleno-methionine labeled RBK1 protein, which could allow solving of the structure by multiple anomalous dispersion (MAD) phasing. Although the characterized variant I101V, which showed twice the activity compared to the wild-type, could serve as a basis for further mutations at other hotspots, evaluating the basic binding mode and conformation of D-xylose should be attempted first. Since the low percentage of the α-furanose form in solution might be a severe limitation, approaches featuring mutarotases catalyzing the α-β anomeric change could solve this issue. Other approaches such as saturation transfer difference nuclear magnetic resonance (STD-NMR) studies might allow more insight whether low substrate affinity towards D-xylose or unproductive binding leading to low catalytic efficiencies is the main problem in the phosphorylation of this sugar epimer of D-ribose. Also, studying whether D-xylose fully induces the conformational changes in ribokinases necessary for efficient phosphorylation could be interesting.

4.3 Thiol-functionalization of acrylic ester monomers catalyzed by immobilized *Humicola insolens* cutinase

In this case study, immobilized cutinase from *Humicola insolens* (*H. insolens*) was evaluated as a novel biocatalyst in the synthesis of functionalized acrylic esters by transesterification to serve as an example for critical challenges in the application and characterization of biocatalysts. The transesterification reaction was optimized in regard to the catalytic turnover, which is strongly influenced by parameters like the water concentration in the reaction and the optimum temperature. Furthermore, the formation of by-products, which can also be influenced by reaction parameters like the temperature as well as by the presence of radical inhibitors, was minimized. Furthermore, the biocatalyst was characterized regarding its substrate spectrum in transesterifications of acrylates. The evaluation of the kinetics for standard substrates in hydrolysis reactions allows a better comparison with other cutinases.

4.3.1 Characterization of *Humicola insolens* cutinase

Cutinases usually show a broad substrate spectrum and a strong dependence on the size of the substrate with a preference for shorter-chain substrates like the soluble ester *para*-nitrophenylbutyrate (*p*NPB). The activity and kinetics of various cutinases has been tested with soluble *p*NP esters in the literature. For a comparative evaluation, the biochemical data were also collected for HiC for a range of *p*NP ester substrates in this study. Hic showed a preference for the shorter-chain substrates *p*NP-butyrate (C4) and *p*NP-octanoate (C8), while the substrate affinity K_M and enzymatic activity towards *p*NP-palmitate were one order of magnitude higher and lower, respectively. The data corresponds well with values published for cutinases from *Thermobifida fusca* (TfC) and *Fusarium solani pisi* (FsC) [231]. Recently, Baker *et al*. compared five cutinases from *Alternaria brassicicola* (AbC), *Aspergillus fumigatus* (AfC), *Aspergillus oryzae* (AoC), *F. solani pisi* (FsC), and *H. insolens* (HiC) and found that HiC demonstrated the second highest catalytic efficiency towards *p*NPB after AoC. However, the catalytic data from Baker *et al.* [181] is presented in an unusual way and the k_{cat} values are about two orders of magnitude lower compared to the results in this work and other studies. HiC was found to exhibit a high stability up to 50°C with over 90% residual activity after 48 h of incubation, yet activity strongly decreased when incubated at 60°C [198,219]. In this regard, HiC showed the highest stability among the five cutinases at elevated temperatures [181]. The highest activity towards soluble *p*NP esters and triglycerides was found at 50°C [198,219], which is slightly higher compared to the cutinases FsC and AbC (35 to 40°C) [220,221].

4.3.2 Transesterification of acrylic monomers

HiC was used for the first time in the transesterification of acrylic monomers for thiol-functionalized acrylic esters as a novel biocatalyst in this industrially relevant application. In the application of a biocatalyst for a biocatalytic process, many critical challenges have to be considered.

For the application in organic media in the transesterification reactions, HiC was immobilized on the polystyrenic, micropourous resin Diaion HP20 [218]. In the first stage of the study [198], the immobilization was improved in regard to the pH value of the aqueous solution used during immobilization. As it was shown, that immobilized or lyophilized enzymes retain the ionization state influenced by the pH from the last aqueous solution, the activity in organic solvents can be optimized by working at the pH optimum [27,71]. For this study, HiC was immobilized at a pH of 7.8 since the pH optimum was found to be between 7 to 8 [198]. Interestingly, HiC showed no activity for the transesterification of acrylic monomers when used in lyophilized form (data not shown) [219]. Immobilization allows the easy separation of the biocatalyst after a batch production and can also enhance the stability of the enzyme in organic media [67].

The choice of the organic solvent and concentrations of substrates can have an remarked influence on the overall conversion and other factors like chemo-, regio- and stereoselectivitiy [70]. Compared to a setup of equimolar amounts of methyl acrylate and 6-mercapto-1-hexanol (150 µM) in diisopropyl ether as solvent, the reaction in a solvent free system in methyl acrylate with a high molar ratio of methyl acrylate to 6-mercapto-1-hexanol of 22 to 1 was far superior in the achieved conversion [198,219].

While non-aqueous enzymology has emerged as a major area of biotechnology, it was soon discovered that enzymes are oftentimes up to 10^4 to 10^5 times less active in organic solvents than in water [68]. Reasons for this include diffusional limitations, structural changes or suboptimal pH conditions, which can occur during lyophilization or immobilization, unfavorable energetics of substrate dissolvation, transition state destabilization, or reduced conformational mobility. Most of these issues can be solved by selecting an optimum way of preparing the biocatalyst, a solvent, which favors substrate binding and stabilization of the transition state, or a water activity, which allows the conformational mobility of the enzymes. The latter was optimized for HiC since this is especially important for transesterification reactions, which is a reverse reaction to hydrolysis in aqueous conditions [198,219]. Although the specific hydration of an enzyme is important for its activity, the water activity has to be low in transesterification reactions, which are usually conducted in organic solvents, in order not to favor hydrolysis [69,232–234]. The effect of the water activity on hydrolysis and transesterification has been studied for immobilized FsC in acetonitrile for the reaction between vinyl butyrate and 2-phenyl-1-propanol [235]. An increase in the water activity a_w between 0.2 and 0.4 improved transesterification, while a further increase of a_w to 0.7 shifted the equilibrium to hydrolysis. For the transesterification of acrylic monomers in this study, HiC was prepared with a water activity towards zero. Addition of 0.025% (w/w immobilisate) almost doubled the conversion in 6 h at 40°C from 54.5 ± 1.0% to 84.1 ± 1.7% [198,219]. Formation of Michael-addition by-products did not increase, but was shifted to the Michael-addition product (6) (see Figure 3-28a). The increase in conversion was seen for the transesterification of ethyl acrylate and methyl methacrylate with 6-mercapto-1-hexanol as well.

Radical inhibitors can be applied to avoid formation of unfavored Michael-addition products in transesterification reaction with acrylates. Recently, the addition of 3% (w/w acrylic ester) of the radical inhibitor butylated hydroxytoluol was used to reduce formation of Michael-addition by-products in the lipase catalyzed synthesis of *N*-(2-hydroxyethyl)-acrylamide from ethyl acrylate and 2-amino-1-ethanol in diisopropyl ether [187]. While the formation of the Michael-addition products in the presence of BHT was further reduced, the overall conversion to the main product 6-mercaptohexyl acrylic ester was lowered as well. Although it was not further tested, this might be due to an inhibiting effect of BHT on the cutinase since the inhibition increased with increasing BHT concentration.

The stability of immobilized HiC was tested by preincubation in methyl acrylate for up to 24 h at 40°C in the presence or absence of 0.025% (w/w immobilisate) water [198,219]. Without water HiC displayed almost no residual activity, while with 0.025% (w/w) water, Hic showed about 68% conversion to the main product without preincubation in methyl acrylate. A loss of water from the enzyme or immobilisate during preincubation might explain the much stronger loss in activity in the absence of water. In polar organic solvents, unlike in nonpolar solvents, water is found fragmented in small clusters and loosely bound to the enzyme´s surface [236]. Leakage of water from the enzyme might have occurred during preincubation and addition of water prior to transesterification might have compensated for the loss, regaining the critical hydration state of the enzyme.

With the optimized reaction conditions (40°C, 0.025% (w/w immobilisate) water, reaction in methyl acrylate without further solvent) [198,219], a large-scale conversion (5 mL) was performed. Here, no Michael-addition products were detected during routine GC/MS analysis, which can be explained by higher reaction rates leading to higher conversions concomitantly reducing by-product formation. The reason for this is likely a better and faster homogenization of the substrates and the immobilisate in the reaction mixture by more efficient stirring in the larger setup. This could have prevented sedimentation of the immobilisate and avoided diffusional limitations in the large scale batch compared to the small scale setup. The main product was purified by silica column chromatography and the structure of the compound was subsequently elucidated by NMR and FT-IR. The compound 6-mercaptohexyl acrylic ester was verified as the main product, and importantly only the acrylic ester, not a possible thioester, was formed. This underlines the high chemoselectivity of the employed cutinase. The thiol group is thereby left as the reactive, functional end group, thus providing the possibility to synthesize functionalized polyacrylates from acrylic monomers.

For the evaluation of the substrate spectrum, 4-mercapto-1-butanol, methyl methacrylate, and methyl propionate were applied in transesterification reactions under optimized conditions (see Table 3-8). HiC strongly preferred methyl acrylate over methyl methacrylate. Since methyl acrylate is sterically less demanding than methyl methacrylate, it might fit better into the active site of HiC, resulting in higher conversion compared to methyl methacrylate. Sterical hindrance at the α-carbon and electronic effects of the acyl donors as well as the nature of the acyl acceptor was shown to influence transacylation of several acrylates and corresponding saturated esters catalyzed by CalB [237]. Transesterification of methyl propionate with 6-mercapto-1-hexanol was about three times faster compared to the reaction with methyl acrylate. Syrén and Hult [237] explored the problem of the low reaction rates with chemical and enzymatic transacylations of acrylates and compared apparent k_{cat} values of CalB, HiC and *Rhizomucor miehei* lipase with molecular dynamics simulations. They found that acrylates, which exist in two low-energy conformations s-*cis*/s-*trans* due to a stabilizing π-system, react with small deviations from their low-energy ground state. When comparing the reaction rates for the transacylation of methyl propionate and methyl acrylate with 1-propanol as acyl acceptor, respectively, a five to six times higher reaction rate for methyl propionate was found [238]. This was explained with the unnatural, predominantly flat substrate geometry of methyl acrylate. While there was only little difference observed for the transesterification with both mercaptoalcohols as acyl acceptors, the geometry of the acrylate seemed to have the strongest influence on conversion. Therefore, highest conversion can be achieved with sterically undemanding substrates like methyl acrylate.

4.3.3 Outlook

Cutinase from *H. insolens* was shown to be a promising candidate for the functionalization of polyacrylates for various industrial applications. The optimization of the reaction condition, especially controlling the critical water content and enzyme hydration state, will be most important for high space-time yields in industrial scale applications. HiC could now be further improved regarding its process stability, either by evaluation of more immobilization techniques or by improving the enzyme itself. By using rational protein design, directed evolution, or a combined approach, the stability as well as the substrate spectrum of HiC might be optimized to tailor this enzyme to a highly efficient transesterification biocatalyst.

4.4 Conclusion

The biocatalytic cycle [2] for establishing biocatalytic processes involves several key steps, the production, characterization, engineering, and application of a biocatalyst as well as down-stream processing. The present thesis features attempts and solutions to several critical challenges in these steps.

As a case study on challenges in the production and characterization of biocatalysts, a high yield fed-batch fermentation process of the dirigent protein *At*DIR6 in *Pichia pastoris* was established and the importance of the protein glycosylation for the structure, solubility, and function of the protein was evaluated. The most critical step involved the heterologous production of the eukaryotic protein, which was found to be highly problematic for bacterial expression hosts.

As a case study on challenges in the production, characterization, and engineering of biocatalysts, the ribokinase RBK1 from *Saccharomyces cerevisiae* was expressed in a recombinant host, characterized for optimum reaction conditions, and engineered towards altered substrate specificity. The rationale for this engineering approach was a strategy for developing a novel D-xylose metabolizing *Saccharomyces cerevisiae* strain for the production of second generation biofuels from lignocellulosic feedstocks. The ribokinase exhibits high substrate specificity for its natural substrate D-ribose. In this regard, the most critical step involved the alteration of the substrate spectrum of the ribokinase by rational design, which proved to be extremely problematic. From the established focused library only one variant showed improved activity towards D-xylose, which can now serve as a basis for further studies since the targeted increase in catalytic efficiency was not yet sufficient for the envisioned process.

As a case study on challenges in the characterization and application of biocatalysts, immobilized cutinase from *Humicola insolens* was characterized regarding its substrate spectrum and kinetics and employed in the thiol-functionalization of acrylic ester monomers by transesterification. The critical step involved the meticulous optimization of the process parameters, especially controlling the critical water content in the non-aqueous reactions.

In summary, in the scope of this thesis, numerous critical challenges involved in the key steps of establishing biocatalytic processes were highlighted. While few issues remained to be challenging, reasonable solutions were presented for several of these issues.

5. References

1 Bornscheuer, U. T. *et al.* (2012) Engineering the third wave of biocatalysis. *Nature*. 485, 185–194

2 Schmid, A. *et al.* (2001) Industrial biocatalysis today and tomorrow. *Nature*. 409, 258–268

3 Wolfenden, R. and Snider, M. J. (2001) The depth of chemical time and the power of enzymes as catalysts. *Accounts of Chemical Research*. 34, 938–45

4 Tramper, J. (1996) Chemical versus biochemical conversion: when and how to use biocatalysts. *Biotechnology and Bioengineering*. 52, 290–5

5 Rosenthaler, L. (1908) Durch Enzyme bewirkte asymmetrische Synthese. *Biochemische Zeitschrift*. 14, 238–253

6 Jensen, V. J. and Rugh, S. (1987) Industrial scale production and application of immobilized glucose isomerase. *Methods in Enzymology*. 136, 356–370

7 Liese, A. *et al.* (2006) *Industrial Biotransformations*, (2nd edn) Wiley-VCH Verlag.

8 Bruggink, A. *et al.* (1998) Penicillin acylase in the industrial production of β-lactam antibiotics. *Organic Process Research & Development*. 2, 128–133

9 Maurer, K.-H. (2004) Detergent proteases. *Current Opinion in Biotechnology*. 15, 330–4

10 Koeller, K. M. and Wong, C. H. (2001) Enzymes for chemical synthesis. *Nature*. 409, 232–40

11 Nagasawa, T. and Yamada, H. (1990) Application of nitrile converting enzymes for the production of useful compounds. *Pure and Applied Chemistry*. 62, 1441–1444

12 Nagasawa, T. *et al.* (1990) Production of acrylic acid and methacrylic acid using *Rhodococcus rhodochrous* J1 nitrilase. *Applied Microbiology and Biotechnology*. 34,

13 Kazlauskas, R. J. and Bornscheuer, U. T. (2009) Finding better protein engineering strategies. *Nature Chemical Biology*. 5, 526–9

14 Turner, N. J. (2009) Directed evolution drives the next generation of biocatalysts. *Nature chemical biology*. 5, 567–73

15 Arnold, F. H. and Volkov, A. A. (1999) Directed evolution of biocatalysts. *Current Opinion in Chemical Biology*. 3, 54–9

16 Desai, A. A. (2011) Sitagliptin manufacture: a compelling tale of green chemistry, process intensification, and industrial asymmetric catalysis. *Angewandte Chemie (International Ed. in English)*. 50, 1974–6

17 Strohmeier, G. *et al.* (2011) Application of designed enzymes in organic synthesis. *Chemical Reviews*. 111, 4141–64

18 Steen, E. J. *et al.* (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature*. 463, 559–62

19 Keasling, J. D. and Chou, H. (2008) Metabolic engineering delivers next-generation biofuels. *Nature Biotechnology*. 26, 298–299

20 Hahn-Hägerdal, B. *et al.* (2007) Metabolic Engineering for Pentose Utilization in *Saccharomyces cerevisiae*. *Advances in Biochemical Engineering and Biotechnology*. 108, 147–177

21 Dekker, K. *et al.* (1991) Xylose (glucose) isomerase gene from the thermophile *Thermus thermophilus*: cloning, sequencing, and comparison with other thermostable xylose isomerases. *Journal of Bacteriology*. 173, 3078–3083

22 Kuyper, M. *et al.* (2003) High-level functional expression of a fungal xylose isomerase: the key to efficient ethanolic fermentation of xylose by *Saccharomyces cerevisiae*? *FEMS Yeast Research*. 4, 69–78

23 Tantirungkij, M. *et al.* (1993) Construction of xylose-assimilating *Saccharomyces cerevisiae*. *Journal of Fermentation and Bioengineering*. 75, 83–88

24 Kötter, P. and Ciriacy, M. (1993) Xylose fermentation by *Saccharomyces cerevisiae*. *Applied Microbiology and Biotechnology*. 38, 776–783

25 Matsushika, A. *et al.* (2009) Ethanol production from xylose in engineered *Saccharomyces cerevisiae* strains: current state and perspectives. *Applied Microbiology and Biotechnology*. 84, 37–53

26 Wenda, S. *et al.* (2011) Industrial biotechnology - the future of green chemistry? *Green Chemistry*. 13, 3007–3047

27 Klibanov, A. M. (2001) Improving enzymes by using them in organic solvents. *Nature*. 409, 241–246

28 Nestl, B. M. *et al.* (2011) Recent progress in industrial biocatalysis. *Current Opinion in Chemical Biology*. 15, 187–193

29 Lye, G. and Woodley, J. M. (1999) Application of in situ product-removal techniques to biocatalytic processes. *Trends in Biotechnology*. 17, 395–402

30 Waegeman, H. and Soetaert, W. (2011) Increasing recombinant protein production in Escherichia coli through metabolic and genetic engineering. *Journal of Industrial Microbiology & Biotechnology*. 38, 1891–1910

31 Baneyx, F. and Mujacic, M. (2004) Recombinant protein folding and misfolding in *Escherichia coli*. *Nature Biotechnology*. 22, 1399–1408

32 Terpe, K. (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*. 72, 211–22

33 Westers, L. *et al.* (2004) *Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. *Biochimica et Biophysica Acta*. 1694, 299–310

34 Higgins, D. R. (2001) Overview of Protein Expression in *Pichia pastoris*. In *Current Protocols in Protein Science.* John Wiley & Sons, Inc.

35 Damasceno, L. M. *et al.* (2011) Protein secretion in *Pichia pastoris* and advances in protein production. *Applied Microbiology and Biotechnology*. 93, 31–39

36 Cereghino, J. L. and Cregg, J. M. (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiology Reviews*. 24, 45–66

37 Wacker, M. *et al.* (2002) *N*-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science*. 298, 1790–3

38 Hellwig, S. *et al.* (2004) Plant cell cultures for the production of recombinant proteins. *Nature Biotechnology*. 22, 1415–22

39 Ikonomou, L. *et al.* (2003) Insect cell culture for industrial production of recombinant proteins. *Applied Microbiology and Botechnology*. 62, 1–20

40 Sahdev, S. *et al.* (2008) Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Molecular and Cellular Biochemistry*. 307, 249–64

41 Burgess-Brown, N. A. *et al.* (2008) Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. *Protein Expression and Purification*. 59, 94–102

42 LaVallie, E. R. *et al.* (1993) A Thioredoxin Gene Fusion Expression System That Circumvents Inclusion Body Formation in the *E. coli* Cytoplasm. *Nature Biotechnology*. 11, 187–193

43 Sachdev, D. and Chirgwin, J. M. (1998) Solubility of proteins isolated from inclusion bodies is enhanced by fusion to maltose-binding protein or thioredoxin. *Protein Expression and Purification*. 12, 122–32

44 Davis, G. D. *et al.* (1999) New fusion protein systems designed to give soluble expression in *Escherichia coli*. *Biotechnology and Bioengineering*. 65, 382–388

45 Kolaj, O. *et al.* (2009) Use of folding modulators to improve heterologous protein production in *Escherichia coli*. *Microbial Cell Factories*. 8,

46 Leichert, L. I. (2011) Proteomic methods unravel the protein quality control in *Escherichia coli*. *Proteomics*. 11, 3023–35

47 Bukau, B. and Horwich, a L. (1998) The Hsp70 and Hsp60 chaperone machines. *Cell*. 92, 351–66

48 Hoffmann, A. *et al.* (2010) Structure and function of the molecular chaperone trigger factor. *Biochimica et Biophysica Acta*. 1803, 650–61

49 Martínez-Alonso, M. *et al.* (2010) Side effects of chaperone gene co-expression in recombinant protein production. *Microbial Cell Factories*. 9, 64

50 Eiberle, M. K. and Jungbauer, A. (2010) Technical refolding of proteins: Do we have freedom to operate? *Biotechnology Journal*. 5, 547–559

51 Vallejo, L. F. and Rinas, U. (2004) Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microbial Cell Factories*. 3, 11

52 Peti, W. and Page, R. (2007) Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. *Protein Expression and Purification*. 51, 1–10

53 Pacheco, B. *et al.* (2011) A screening strategy for heterologous protein expression in *Escherichia coli* with the highest return of investment. *Protein Expression and Purification*. 81, 33–41

54 Marco, A. De (2009) Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. *Microbial Cell Factories*. 8, 26

55 González-Montalbán, N. *et al.* (2007) Recombinant protein solubility - does more mean better? *Nature Biotechnology*. 25, 718–20

56 Fox, R. J. and Clay, M. D. (2009) Catalytic effectiveness, a measure of enzyme proficiency for industrial applications. *Trends in Biotechnology*. 27, 137–140

57 Berg, J. M. *et al.* (2011) *Biochemistry*, 7th (rev).Palgrave MacMillan.

58 Stemmer, W. P. (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature*. 370, 389–91

59 Moore, J. C. and Arnold, F. H. (1996) Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nature Biotechnology*. 14, 458–67

60 Reetz, M. T. *et al.* (1997) Creation of enantioselective biocatalysts for organic chemistry by *in vitro* evolution. *Angewandte Chemie (International Ed. in English)*. 36, 2830–2832

61 Wong, T. S. *et al.* (2004) Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. *Nucleic Acids Research*. 32, e26

62 Reetz, M. T. *et al.* (2008) Addressing the numbers problem in directed evolution. *ChemBioChem*. 9, 1797–804

63 Reetz, M. T. *et al.* (2006) Directed evolution of enantioselective enzymes: Iterative cycles of CASTing for probing protein-sequence space. *Angewandte Chemie*. 118, 1258–1263

64 Reetz, M. T. *et al.* (2010) Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods. *Journal of the American Chemical Society*. 132, 9144–52

65 Reetz, M. T. *et al.* (2006) Expanding the substrate scope of enzymes: Combining mutations obtained by CASTing. *Chemistry - A European Journal*. 12, 6031–6038

66 Zhou, Z. and Hartmann, M. (2012) Recent progress in biocatalysis with enzymes immobilized on mesoporous hosts. *Topics in Catalysis*. 55, 1081–1100

67 Mateo, C. *et al.* (2007) Improvement of enzyme activity, stability and selectivity via immobilization techniques. *Enzyme and Microbial Technology*. 40, 1451–1463

68 Klibanov, A. M. (1997) Why are enzymes less active in organic solvents than in water? *Trends in Biotechnology*. 15, 97–101

69 Zaks, A. and Klibanov, A. M. (1988) The effect of water on enzyme action in organic media. *Journal of Biological Chemistry*. 263, 8017–8021

70 Carrea, G. *et al.* (1995) Role of solvents in the control of enzyme selectivity in organic media. *Trends in Biotechnology*. 13, 63–70

71 Xu, K. and Klibanov, A. M. (1996) PH control of the catalytic activity of crosslinked enzyme crystals in organic solvents. *Journal of the American Chemical Society*. 118, 9815–9819

72 Hartmann, T. (1996) Diversity and variability of plant secondary metabolism: a mechanistic view. *Entomologia Experimentalis et Applicata*. 80, 177–188

73 Lewis, N. G. and Davin, L. B. (1999) Lignans: Biosynthesis and function. In *Comprehensive Natural Products Chemistry* 1 (Barton Sir, D. H. R. *et al.*, eds), pp. 639–712, Elsevier

74 Parr, A. J. and Bolwell, G. P. (2000) Phenols in the plant and in man. The potential for possible nutritional enhancement of the diet by modifying the phenols content or profile. *Journal of the Science of Food and Agriculture*. 80, 985–1012

75 Lewinsohn, E. and Gijzen, M. (2009) Phytochemical diversity: The sounds of silent metabolism. *Plant Science*. 176, 161–169

76 Gershenzon, J. and Dudareva, N. (2007) The function of terpene natural products in the natural world. *Nature Chemical Biology*. 3, 408–14

77 Saleem, M. *et al.* (2005) An update on bioactive plant lignans. *Natural Product Reports*. 22, 696–716

78 Pan, J.-Y. *et al.* (2009) An update on lignans: natural products and synthesis. *Natural Product Reports*. 26, 1251–92

79 Lenz, W. and Knapp, K. (1962) Die Thalidomid-Embryopathie. *Deutsche medizinische Wochenschrift*. 87, 1232–1242

80 Umezawa, T. (2003) Diversity in lignan biosynthesis. *Phytochemistry Reviews*. 2, 371–390

81 Haworth, R. D. (1936) Natural Resins. *Annual Reports on the Progress of Chemistry B*. 33, 266–279

82 Moss, G. P. (2000) Nomenclature of Lignans and Neolignans (IUPAC Recommendations 2000). *Pure and Applied Chemistry*. 72, 1493–1523

83 Su, D.-M. *et al.* (2007) Glucosides from the roots of *Capparis tenera*. *Chemistry & Biodiversity*. 4, 2852–62

84 Scher, J. M. *et al.* (2003) Lignan derivatives from the liverwort *Bazzania trilobata*. *Phytochemistry*. 62, 769–777

85 Takeda, R. *et al.* (1990) The first isolation of lignans, megacerotonic acid and anthocerotonic acid, from non-vascular plants, anthocerotae (hornworts). *Tetrahedron Letters*. 31, 4159–4162

86 Rui Chaio, L. *et al.* (1994) Phenolic constituents of *Selaginella doederleinii*. *Planta Medica*. 60, 168–170

87 Swan, E. P. *et al.* (1969) The lignans of *Thuja plicata* and the sapwood-heartwood transformation. *Phytochemistry*. 8, 345–351

88 Elakovich, S. D. and Stevens, K. L. (1985) Phytotoxic properties of nordihydroguaiaretic acid, a lignan from *Larrea tridentata* (Creosote bush). *Journal of Chemical Ecology*. 11, 27–33

89 Imbert, T. F. (1998) Discovery of podophyllotoxins. *Biochimie*. 80, 207–222

90 Gordaliza, M. *et al.* (2004) Podophyllotoxin: distribution, sources, applications and new cytotoxic derivatives. *Toxicon*. 44, 441–59

91 Seidel, V. *et al.* (2002) Biosynthesis of podophyllotoxin in *Linum album* cell cultures. *Planta*. 215, 1031–9

92 Kim, H. J. *et al.* (2009) Metabolic engineering of lignan biosynthesis in *Forsythia cell* culture. *Plant & Cell physiology*. 50, 2200–9

93 Vogt, T. (2010) Phenylpropanoid biosynthesis. *Molecular Plant*. 3, 2–20

94 Davin, L. B. *et al.* (2008) Dissection of lignin macromolecular configuration and assembly: comparison to related biochemical processes in allyl/propenyl phenol and lignan biosynthesis. *Natural Products Reports*. 25, 1015–1090

95 Davin, L. B. *et al.* (1997) Stereoselective bimolecular phenoxy radical coupling by an auxiliary (dirigent) protein without an active center. *Science*. 275, 362–366

96 Frías, I. *et al.* (1991) Purification of a new peroxidase catalysing the formation of lignan-type compounds. *The Biochemical Journal*. 273(Pt 1), 109–13

97 Suzuki, S. and Umezawa, T. (2007) Biosynthesis of lignans and norlignans. *Journal of Wood Science*. 53, 273–284

98 Croteau, R. *et al.* (2000) Natural Products (Secondary Metabolites). *Biochemistry & Molecular Biology of Plants, B*.

99 Fujita, M. (1999) Recombinant Pinoresinol-Lariciresinol Reductases from Western Red Cedar (*Thuja plicata*) Catalyze Opposite Enantiospecific Conversions. *Journal of Biological Chemistry*. 274, 618–627

100 Min, T. *et al.* (2003) Crystal structures of pinoresinol-lariciresinol and phenylcoumaran benzylic ether reductases and their relationship to isoflavone reductases. *The Journal of Biological Chemistry*. 278, 50714–23

101 Youn, B. *et al.* (2005) Crystal structures of apo-form and binary/ternary complexes of *Podophyllum secoisolariciresinol* dehydrogenase, an enzyme involved in formation of health-protecting and plant defense lignans. *The Journal of Biological Chemistry*. 280, 12917–26

102 Von Heimendahl, C. B. I. *et al.* (2005) Pinoresinol-lariciresinol reductases with different stereospecificity from *Linum album* and *Linum usitatissimum*. *Phytochemistry*. 66, 1254–63

103 Freudenberg, K. (1965) Lignin: Its constitution and formation from *p*-hydroxycinnamyl alcohols. *Science*. 148, 595–600

104 Davin, L. B. and Lewis, N. G. (2005) Dirigent phenoxy radical coupling: advances and challenges. *Current Opinion in Biotechnology*. 16, 398–406

105 Gang, D. R. *et al.* (1999) Regiochemical control of monolignol radical coupling: A new paradigm for lignin and lignan biosynthesis. *Chemistry & Biology*. 6, 143–151

106 Ralph, S. G. *et al.* (2007) Dirigent proteins in conifer defense II: Extended gene discovery, phylogeny, and constitutive and stress-induced gene expression in spruce (*Picea* spp.). *Phytochemistry*. 68, 1975–1991

107 Liu, J. *et al.* (2008) Stereoselective coupling of hemigossypol to form (+)-gossypol in moco cotton is mediated by a dirigent protein. *Phytochemistry*. 69, 3038–3042

108 Kim, M. K. *et al.* (2002) The western red cedar (*Thuja plicata*) 8-8’ DIRIGENT family displays diverse expression patterns and conserved monolignol coupling specificity. *Plant Mol Biol*. 49, 199–214

109 Pickel, B. *et al.* (2010) An Enantiocomplementary Dirigent Protein for the Enantioselective Laccase-Catalyzed Oxidative Coupling of Phenols. *Angewandte Chemie (International Ed. in English)*. 49, 202–204

110 Pickel, B. *et al.* (2012) A model of dirigent proteins derived from structural and functional similarities with allene oxide cyclase and lipocalins. *FEBS Journal*. 279, 1980–1993

111 Halls, S. C. and Lewis, N. G. (2002) Secondary and quaternary structures of the (+)-pinoresinol-forming dirigent protein. *Biochemistry*. 41, 9455–9461

112 Halls, S. C. *et al.* (2004) Kinetic study of coniferyl alcohol radical binding to the (+)-pinoresinol forming dirigent protein. *Biochemistry*. 43, 2587–2595

113 Hapiot, P. *et al.* (1994) Mechanism of oxidative coupling of coniferyl alcohol. *Phytochemistry*. 36, 1013–1020

114 Lerouge, P. *et al.* (1998) *N*-Glycoprotein biosynthesis in plants: recent developments and future trends. *Plant Molecular Biology*. 38, 31–48

115 Burlat, V. *et al.* (2001) Dirigent proteins and dirigent sites in lignifying tissues. *Phytochemistry*. 57, 883–897

116 Schaller, A. and Stintzi, A. (2009) Enzymes in jasmonate biosynthesis - Structure, function, regulation. *Phytochemistry*. 70, 1532–1538

117 Ziegler, J. *et al.* (2000) Molecular cloning of allene oxide cyclase. The enzyme establishing the stereochemistry of octadecanoids and jasmonates. *The Journal of Biological Chemistry*. 275, 19132–8

118 Hofmann, E. and Pollmann, S. (2008) Molecular mechanism of enzymatic allene oxide cyclization in plants. *Plant Physiology and Biochemistry*. 46, 302–8

119 Hofmann, E. *et al.* (2006) The crystal structure of *Arabidopsis thaliana* allene oxide cyclase: Insights into the oxylipin cyclization reaction. *The Plant Cell Online*. 18, 3201–3217

120 Flower, D. R. (1996) The lipocalin protein family: structure and function. *The Biochemical Journal*. 318 (Pt 1), 1–14

121 Kim, K.-W. *et al.* (2012) Opposite Stereoselectivities of Dirigent Proteins in *Arabidopsis and Schizandra* Species. *Journal of Biological Chemistry*. 287, 33957–72

122 Ralph, S. *et al.* (2006) Dirigent proteins in conifer defense: gene discovery, phylogeny, and differential wound- and insect-induced expression of a family of DIR and DIR-like genes in spruce (Picea spp.). *Plant Mol Biol*. 60, 21–40

123 Culley, D. E. *et al.* (1995) Molecular characterization of disease-resistance response gene DRR206-d from *Pisum sativum (L.)*. *Plant Physiology*. 107, 301–2

124 Zhu, L. *et al.* Isolation and characterization of two novel dirigent-like genes highly induced in cotton (*Gossypium barbadense* and *G. hirsutum*) after infection by *Verticillium dahliae*. , *Journal of Plant Pathology*, 89. 03-Jan-(2007) , 41–45

125 Davin, L. B. and Lewis, N. G. (2005) Lignin primary structures and dirigent sites. *Current Opinion in Biotechnology*. 16, 407–415

126 Hatfield, R. and Vermerris, W. (2001) Lignin formation in plants. The dilemma of linkage specificity. *Plant Physiology*. 126, 1351–1357

127 Davin, L. B. and Lewis, N. G. (2000) Dirigent proteins and dirigent sites explain the mystery of specificity of radical precursor coupling in lignan and lignin biosynthesis. *Plant Physiology*. 123, 453–462

128 Ralph, J. *et al.* (1999) Are Lignins Optically Active? *Journal of Agricultural and Food Chemistry*. 47, 2991–2996

129 Park, J. and Gupta, R. S. (2008) Adenosine kinase and ribokinase – the RK family of proteins. *Cellular and Molecular Life Sciences*. 65, 2875–2896

130 Anderson, A. and Cooper, R. A. (1969) The significance of ribokinase for ribose utilization by Escherichia coli. *Biochimica et Biophysica Acta (BBA) - General Subjects*. 177, 163–165

131 Tozzi, M. G. *et al.* (2006) Pentose phosphates in nucleoside interconversion and catabolism. *The FEBS journal*. 273, 1089–101

132 Drew, K. N. *et al.* (1998) 13C-labeled aldopentoses: detection and quantitation of cyclic and acyclic forms by heteronuclear 1D and 2D NMR spectroscopy. *Carbohydrate Research*. 307, 199–209

133 Mowbray, S. L. and Cole, L. B. (1992) 1.7 Å X-ray structure of the periplasmic ribose receptor from Escherichia coli. *Journal of Molecular Biology*. 225, 155–175

134 Kim, M.-S. *et al.* (2003) Crystal structures of RbsD leading to the identification of cytoplasmic sugar-binding proteins with a novel folding architecture. *The Journal of Biological Chemistry*. 278, 28173–80

135 Sigrell, J. A. *et al.* (1998) Structure of *Escherichia coli* ribokinase in complex with ribose and dinucleotide determined to 1.8 Ã resolution: insights into a new family of kinase structures. *Structure*. 6, 183–193

136 Sigrell, J. A. *et al.* (1999) Induced fit on sugar binding activates ribokinase. *Journal of Molecular Biology*. 290, 1009–1018

137 Ryu, K.-S. *et al.* (2004) NMR application probes a novel and ubiquitous family of enzymes that alter monosaccharide configuration. *The Journal of Biological Chemistry*. 279, 25544–8

138 Thoden, J. B. *et al.* (2003) The catalytic mechanism of galactose mutarotase. *Protein Science*. 12, 1051–9

139 Oh, H. (1999) A mutated PtsG, the glucose transporter, allows uptake of D-ribose. *Journal of Biological Chemistry*. 274, 14006–14011

140 Agranoff, W. *et al.* (1956) Purification and properties of calf liver ribokinase. *Journal of Biological Chemistry*. 219, 221–229

141 Bork, P. *et al.* (1993) Convergent evolution of similar enzymatic function on different protein folds: The hexokinase, ribokinase, and galactokinase families of sugar kinases. *Protein Science*. 2, 31–40

142 Hope, J. N. *et al.* (1986) Ribokinase from *Escherichia coli* K12. Nucleotide sequence and overexpression of the rbsK gene and purification of ribokinase. *Journal of Biological Chemistry*. 261, 7663–7668

143 Park, J. *et al.* (2007) Identification and characterization of human ribokinase and comparison of its properties with *E. coli* ribokinase and human adenosine kinase. *FEBS Letters*. 581, 3211–3216

144 Sigrell, J. A. *et al.* (1997) Purification, characterization, and crystallization of *Escherichia coli* ribokinase. *Protein Science*. 6, 2474–2476

145 Kawai, S. *et al.* (2005) Hypothesis: structures, evolution, and ancestor of glucose kinases in the hexokinase family. *Journal of Bioscience and Boengineering*. 99, 320–30

146 Chuvikovsky, D. V *et al.* (2006) Ribokinase from *E. coli*: Expression, purification, and substrate specificity. *Bioorganic & Medicinal Chemistry*. 14, 6327–6332

147 Ogbunude, P. O. J. *et al.* (2007) Molecular Cloning, Expression and Characterization of Ribokinase of *Leishmania major*. *Acta Biochimica et Biophysica Sinica*. 39, 462–466

148 Li, J. *et al.* (2012) Crystal structure of Sa239 reveals the structural basis for the activation of ribokinase by monovalent cations. *Journal of Structural Biology*. 177, 578–582

149 Schimmel, S. D. *et al.* (1974) Deoxyribokinase from *Salmonella typhimurium*. *Archives of Biochemistry and Biophysics*. 164, 560–570

150 Thierry, A. *et al.* (1990) The complete sequence of the 8.2kb segment left of MAT on chromosome III reveals five ORFs, including a gene for a yeast ribokinase. *Yeast*. 6, 521–534

151 Di Cera, E. (2006) A structural perspective on enzymes activated by monovalent cations. *The Journal of Biological Chemistry*. 281, 1305–8

152 Andersson, C. E. and Mowbray, S. L. (2002) Activation of ribokinase by monovalent cations. *Journal of Molecular Biology*. 315, 409–419

153 Maj, M. C. *et al.* (2002) Pentavalent ions dependency is a conserved property of adenosine kinase from diverse sources: Identification of a novel motif implicated in phosphate and magnesium ion binding and substrate inhibition. *Biochemistry*. 41, 4059–4069

154 Maj, M. C. and Gupta, R. S. (2001) The effect of inorganic phosphate on the activity of bacterial ribokinase. *Journal of Protein Chemistry*. 20, 139–144

155 Park, J. *et al.* (2004) Phosphorylated derivatives that activate or inhibit mammalian adenosine kinase provide insights into the role of pentavalent ions in AK catalysis. *The Protein Journal*. 23, 167–177

156 Park, J. *et al.* (2006) Inhibition of adenosine kinase by phosphonate and bisphosphonate derivatives. *Molecular and Cellular Biochemistry*. 283, 11–21

157 Gross, A. *et al.* (1983) Practical synthesis of 5-phospho-D-ribosyl alpha-1-pyrophosphate (PRPP): enzymatic routes from ribose 5-phosphate or ribose. *Journal of the American Chemical Society*. 105, 7428–7435

158 Carvalho, C. M. *et al.* (1999) Cutinase: from molecular level to bioprocess development. *Biotechnology and Bioengineering*. 66, 17–34

159 Dutta, K. *et al.* (2009) Production, characterization and applications of microbial cutinases. *Process Biochemistry*. 44, 127–134

160 Longhi, S. *et al.* (1997) Atomic resolution (1.0 A) crystal structure of Fusarium solani cutinase: stereochemical analysis. *Journal of Molecular Biology*. 268, 779–799

161 Longhi, S. *et al.* (1997) Crystal structure of cutinase covalently inhibited by a triglyceride analogue. *Protein Science*. 6, 275–286

162 Martinez, C. *et al.* (1994) Cutinase, a lipolytic enzyme with a preformed oxyanion hole. *Biochemistry*. 33, 83–89

163 Martinez, C. *et al.* (1993) Engineering cysteine mutants to obtain crystallographic phases with a cutinase from *Fusarium solani pisi*. *Protein Engineering Design and Selection*. 6, 157–165

164 Koller, W. and Kolattukudy, P. E. (1982) Mechanism of action of cutinase: chemical modification of the catalytic triad characteristic for serine hydrolases. *Biochemistry*. 21, 3083–3090

165 Longhi, S. and Cambillau, C. (1999) Structure-activity of cutinase, a small lipolytic enzyme. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*. 1441, 185–196

166 Egmond, M. R. and Van Bemmel, C. J. (1997) Impact of structural information on understanding lipolytic function. *Methods in Enzymology*. 284, 119–129

167 Van Tilbeurgh, H. *et al.* (1993) Interfacial activation of the lipase-procolipase complex by mixed micelles revealed by X-ray crystallography. *Nature*. 362, 814–820

168 Prompers, J. J. *et al.* (1997) 1H, 13C, and 15N resonance assignments of *Fusarium solani pisi* cutinase and preliminary features of the structure in solution. *Protein Science*. 6, 2375–2384

169 Liu, Z. *et al.* (2009) Structural and functional studies of *Aspergillus oryzae* cutinase: enhanced thermostability and hydrolytic activity of synthetic ester and polyester degradation. *Journal of the American Chemical Society*. 131, 15711–15716

170 Carter, P. and Wells, J. A. (1988) Dissecting the catalytic triad of a serine protease. *Nature*. 332, 564–568

171 Magnusson, A. (2005) Rational redesign of *Candida antarctica* lipase B. *Department of Biochemistry, Royal Institute of Technology, School of Biotechnology*.

172 Pio, T. F. and Macedo, G. A. (2009) Cutinases: properties and industrial applications. *Advances in Applied Microbiology*. 66, 77–95

173 Ternström, T. *et al.* (2005) Unfolding and inactivation of cutinases by AOT and guanidine hydrochloride. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. 1748, 74–83

174 Brissos, V. *et al.* (2008) Improving activity and stability of cutinase towards the anionic detergent AOT by complete saturation mutagenesis. *Protein Engineering Design and Selection*. 21, 387–393

175 Brissos, V. *et al.* (2008) Biochemical and structural characterisation of cutinase mutants in the presence of the anionic surfactant AOT. *Biochimica et biophysica acta*. 1784, 1326–1334

176 Kim, Y.-H. *et al.* (2005) Biodegradation and detoxification of organophosphate insecticide, malathion by *Fusarium oxysporum f. sp. pisi* cutinase. *Chemosphere*. 60, 1349–55

177 Murphy, C. A. *et al.* (1996) *Fusarium* polycaprolactone depolymerase is cutinase. *Applied and Environmental Microbiology*. 62, 456–460

178 Vertommen, M. A. *et al.* (2005) Enzymatic surface modification of poly(ethylene terephthalate). *Journal of Biotechnology*. 120, 376–386

179 Hunsen, M. *et al.* (2008) *Humicola insolens* cutinase-catalyzed lactone ring-opening polymerizations: kinetic and mechanistic studies. *Biomacromolecules*. 9, 518–522

180 Hunsen, M. *et al.* (2007) A cutinase with polyester synthesis activity. *Macromolecules*. 40, 148–150

181 Baker, P. J. *et al.* (2012) Identification and comparison of cutinases for synthetic polyester degradation. *Applied Microbiology and Biotechnology*. 93, 229–40

182 Hoyle, C. E. *et al.* (2004) Thiol-enes: Chemistry of the past with promise for the future. *Journal of Polymer Science Part A: Polymer Chemistry*. 42, 5301–5338

183 Li, G.-Z. *et al.* (2010) Investigation into thiol-(meth)acrylate Michael addition reactions using amine and phosphine catalysts. *Polymer Chemistry*. 1, 1196

184 Ohrui, T. *et al.* (1975) Process for continuously synthesizing acrylic acid esters., U.S. Patent 3875212

185 Hajjar, A. B. *et al.* (1990) Preparation of monomeric acrylic ester intermediates using lipase catalysed transesterifications in organic solvents. *Biotechnology Letters*. 12, 825–830

186 Warwel, S. *et al.* (1996) An efficient method for lipase-catalysed preparation of acrylic and methacrylic acid esters. *Biotechnology Techniques*. 10, 283–286

187 Rustoy, E. M. and Baldessari, A. (2006) Chemoselective enzymatic preparation of *N*-hydroxyalkylacrylamides, monomers for hydrophilic polymer matrices. *Journal of Molecular Catalysis B: Enzymatic*. 39, 50–54

188 Nordblad, M. and Adlercreutz, P. (2008) Efficient enzymatic acrylation through transesterification at controlled water activity. *Biotechnology and Bioengineering*. 99, 1518–1524

189 Nordblad, M. and Adlercreutz, P. (2008) Effects of acid concentration and solvent choice on enzymatic acrylation by Candida antarctica lipase B. *Journal of Biotechnology*. 133, 127–133

190 Hrsic, E. *et al.* (2012) Synthesis of thiol functionalized poly(meth)acrylates through enzymatic catalysis and a subsequent one pot reaction process. *European Polymer Journal*. 48, 761–768

191 Boeckh, D. *et al.* (2007) Enzymatic synthesis of sugar acrylates., U.S. Patent 1448785

192 Häring, D. *et al.* (2007) Enzymatic production of (meth)acrylic acid esters., U.S. Patent 1599594

193 Paulus, W. *et al.* (2006) Enzymatic Synthesis of Polyol Acrylates., U.S. Patent 2006030013 (A1)

194 Sambrook, J. *et al.* (1989) Commonly used techniques in molecular cloning. *Molecular cloning: a laboratory manual*.

195 Miroux, B. and Walker, J. E. (1996) Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology*. 260, 289–298

196 Mandel, M. and Higa, A. (1970) Calcium-dependent bacteriophage DNA infection. *Journal of Molecular Biology*. 53, 159–162

197 Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research*. 16, 10881–10890

198 Kazenwadel, C. (2010) Transesterification of acrylate ester monomers with immobilized *Humicola insolens* cutinase. *Diploma thesis, Institute of Technical Biochemistry, Universität Stuttgart*.

199 Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*. 72, 248–254

200 Smith, P. K. *et al.* (1985) Measurement of protein using bicinchoninic acid. *Analytical Biochemistry*. 150, 76–85

201 Schneider, C. A. *et al.* (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*. 9, 671–675

202 Whitmore, L. and Wallace, B. A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research*. 32, W668–W673

203 Whitmore, L. and Wallace, B. a (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers*. 89, 392–400

204 Lees, J. G. *et al.* (2006) A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics*. 22, 1955–62

205 Johnson, W. C. (1999) Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins: Structure, Function, and Bioinformatics*. 35, 307–312

206 Shevchenko, A. *et al.* (1996) Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels. *Analytical Chemistry*. 68, 850–858

207 Cedzich, A. *et al.* (2009) The protease-associated domain and c-terminal extension are required for zymogen processing, sorting within the secretory pathway, and activity of tomato subtilase 3 (SlSBT3). *Journal of Biological Chemistry*. 284, 14068–14078

208 Striegel, A. *et al.* (2009) *Modern size-exclusion liquid chromatography : practice of gel permeation and gel filtration chromatography*, (2nd edn) John Wiley & Sons, Inc.

209 Bisswanger, H. (2002) *Enzyme Kinetics - Principles and Methods*, (2nd edn) WILEY-VCH Verlag GmbH.

210 Pickel, B. (2011) Heterologous expression and molecular characterization of the dirigent protein *AtDIR6* from *Arabidopsis thaliana*. *PhD thesis, Institute for Physiology and Biotechnology of Plants, University of Hohenheim, Stuttgart*.

211 Sørensen, H. P. and Mortensen, K. K. (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *Journal of Biotechnology*. 115, 113–128

212 Brass, J. *et al.* (2008) Rhamnose Promoter Expression System., U.S. Patent 2008/0206817 A1

213 Stumpp, T. *et al.* (2000) Ein neues L-Rhamnose-induzierbares Expressionssystem für Escherichia coli. *Biospektrum*. 6, 33

214 Jeske, M. and Altenbuchner, J. (2010) The Escherichia coli rhamnose promoter rhaP(BAD) is in Pseudomonas putida KT2440 independent of Crp-cAMP activation. *Applied Microbiology and Biotechnology*. 85, 1923–33

215 Kazenwadel, C. *et al.* (2012) Optimized expression of the dirigent protein *At*DIR6 in *Pichia pastoris* and impact of glycosylation on protein structure and function. *Applied Microbiology and Biotechnology*. DOI: 10.1007/s00253-012-4579-x

216 Rossmann, M. G. and Blow, D. M. (1962) The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica*. 15, 24–31

217 Hedfors, C. *et al.* (2005) Thiol End-Functionalization of Poly(e-caprolactone), catalyzed by Candida antarctica Lipase B. *Macromolecules*. 38, 3–5

218 Adachi, T. and Isobe, E. (2004) Fundamental characteristics of synthetic adsorbents intended for industrial chromatographic separations. *Journal of Chromatography A*. 1036, 33–44

219 Kazenwadel, C. *et al.* (2012) Thiol-functionalization of acrylic ester monomers catalyzed by immobilized *Humicola insolens* cutinase. *Enzyme and Microbial Technology*. 51, 9–15

220 Kwon, M. A. *et al.* (2009) High-level expression and characterization of *Fusarium solani* cutinase in *Pichia pastoris*. *Protein Expression and Purification*. 68, 104–109

221 Koschorreck, K. *et al.* (2010) Heterologous expression, characterization and site-directed mutagenesis of cutinase CUTAB1 from *Alternaria brassicicola*. *Applied Microbiology and Biotechnology*. 87, 991–997

222 Melo, E. P. *et al.* (1995) Triglyceride hydrolysis and stability of a recombinant cutinase from *Fusarium solani* in AOT-iso-octane reversed micelles. *Applied Biochemstry and Biotechnology*. 50, 45–56

223 Welch, M. *et al.* (2009) You’re one in a googol: optimizing genes for protein expression. *Journal of the Royal Society - Interface*. 6 Suppl 4, S467–76

224 Zhu, S. *et al.* (2012) A simple and effective strategy for solving the problem of inclusion bodies in recombinant protein technology: His-tag deletions enhance soluble expression. *Applied Microbiology and Biotechnology*. 97, 837–45

225 Dumon-Seignovert, L. *et al.* (2004) The toxicity of recombinant proteins in Escherichia coli: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expression and Purification*. 37, 203–206

226 Durchschlag, H. *et al.* (1991) Comparative determination of the particle weight of glycoproteins by SDS-PAGE and analytical ultracentrifugation. In *Progress in Analytical Ultracentrifugation* 86 (Borchard, W., ed), pp. 41–56, Springer

227 Taroni, C. *et al.* (2000) Analysis and prediction of carbohydrate binding sites. *Protein Engineering*. 13, 89–98

228 Vyas, N. K. (1991) Atomic features of protein-carbohydrate interactions. *Current Opinion in Structural Biology*. 1, 732–740

229 Quiocho, F. A. (1989) Protein-carbohydrate interactions: basic molecular features. *Pure & Applied Chemistry*. 61, 1293–1306

230 Rupp, B. (2010) *Biomolecular crystallography*, Garland Science.

231 Chen, S. *et al.* (2010) Biochemical characterization of the cutinases from *Thermobifida fusca*. *Journal of Molecular Catalysis B: Enzymatic*. 63, 121–127

232 Micaelo, N. M. *et al.* (2005) Water dependent properties of cutinase in nonaqueous solvents: a computational study of enantioselectivity. *Biophysical Journal*. 89, 999–1008

233 Parker, M. C. *et al.* (1995) Measuring enzyme hydration in nonpolar organic solvents using NMR. *Biotechnology and Bioengineering*. 46, 452–458

234 Soares, C. M. *et al.* (2003) Protein structure and dynamics in nonaqueous solvents: insights from molecular dynamics simulation studies. *Biophysical Journal*. 84, 1628–1641

235 Vidinha, P. *et al.* (2004) Effect of immobilization support, water activity, and enzyme ionization state on cutinase activity and enantioselectivity in organic media. *Biotechnology and Bioengineering*. 85, 442–449

236 Micaêlo, N. M. and Soares, C. M. (2007) Modeling hydration mechanisms of enzymes in nonpolar and polar organic solvents. *FEBS Journal*. 274, 2424–2436

237 Syrén, P.-O. and Hult, K. (2010) Substrate conformations set the rate of enzymatic acrylation by lipases. *ChemBioChem*. 11, 802–10

238 Syrén, P.-O. *et al.* (2010) Increased activity of enzymatic transacylation of acrylates through rational design of lipases. *Journal of Molecular Catalysis B: Enzymatic*. 65, 3–10

6. List of figures

Figure 1-1 – The biocatalysis cycle 24

Figure 1-2 – Nomenclature of lignans (a) and selected examples (b) 32

Figure 1-3 – Phenylpropanoid biosynthesis pathway 34

Figure 1-4 – Phenoxy radical coupling of coniferyl alcohol (a) and further biochemical pathway (b) 36

Figure 1-5 – Homology model of *At*DIR6 38

Figure 1-6 – D-ribose configurations in solution (a) and transport (b) 41

Figure 1-7 – Structure of the *E. coli* ribokinase 44

Figure 1-8 – Enzymatic synthesis of purine nucleosides from pentoses 48

Figure 1-9 – Reaction mechanism of cutinases (modified from [171]) 50

Figure 2-1 – Kinase activity assay 77

Figure 2-2 – Cutinase activity measurement with the *para*-nitrophenol assay 84

Figure 3-1 – Comparison of the expression of *At*DIR6 in *E. coli* BL21(DE3) and Origami B(DE3) hosts by SDS-PAGE analysis 87

Figure 3-2 – Expression of *At*DIR6 in *E. coli* Origami B(DE3) with the co-expression of chaperones DnaK, DnaJ and GrpE and purification of *At*DIR6 89

Figure 3-3 – SDS-PAGE and Western blot analysis of the *At*DIR6 expression, fermentation and purification 90

Figure 3-4 – Fed-batch fermentation process of *At*DIR6 with *Pichia pastoris* 91

Figure 3-5 – MALDI-TOF mass spectrum of *At*DIR6 isolated from *Pichia pastoris* 92

Figure 3-6 – SDS-PAGE analysis of *At*DIR6 samples after deglycosylation 93

Figure 3-7 – MALDI-TOF peptide mass fingerprints (PMF) of tryptic digests from glycosylated and PNGase F treated *At*DIR6 94

Figure 3-8 – Analytical gel filtration of *At*DIR6 95

Figure 3-9 – Reversed phase and chiral HPLC analysis of biotransformations with *At*DIR6 isolated from *Pichia pastoris* 96

Figure 3-10 – Enantiomeric excess of (-)-pinoresinol as a function of the (a) *At*DIR6, (b) coniferyl alcohol, and (c) laccase concentration 97

Figure 3-11 – SDS-PAGE and dirigent activity analysis of deglycosylated *At*DIR6 derived from *P. pastoris* 98

Figure 3-12 – Solubility of deglycosylated *At*DIR6 isolated from *Pichia pastoris* during long-term incubations 99

Figure 3-13 – CD spectroscopy of glycosylated and deglycosylated *At*DIR6 derived from *P. pastoris* 100

Figure 3-14 – Metabolically engineered pathways for D-xylose metabolism and novel pathway 102

Figure 3-15 – Homology model of the yeast RBK1 protein 103

Figure 3-16 – Comparison of the expression of RBK1 in *E. coli* BL21(DE3), Rosetta(DE3) and C43(DE3) hosts by SDS-PAGE and Western blot analysis 106

Figure 3-17 – Expression and purification of RBK1 from *E. coli* C43(D3) cells 107

Figure 3-18 –Gel filtration purification of RBK1 108

Figure 3-19 – SDS-PAGE analysis of the purified RBK1 variants 110

Figure 3-20 – Optimization of magnesium and phosphate concentrations and pH value for RBK1 activity 113

Figure 3-21 – HPLC analysis of biotransformations 115

Figure 3-22 – Conversion rate as a function of time and potassium cation concentration .. 116

Figure 3-23 – Activation of RBK1 by different mono- and divalent cations 118

Figure 3-24 – Initial activity screening of the RBK1 focused library towards (a) 10 mM D-ribose and (b) 50 mM D-xylose 119

Figure 3-25 – Phosphate dependency of RBK1 wild-type and I113V variant 120

Figure 3-26 – Mass spectra from LC/MS analysis 124

Figure 3-27 – Michaelis-Menten kinetics of RBK1 wild-type and I113V variant with D-ribose (a) and ATP (b) 126

Figure 3-28 – Transesterification of methyl acrylate and 6-mercapto-1-hexanol by HiC (a) and further substrates tested (b) 129

Figure 3-29 – Expression of HiC from *P. pastoris* and analysis with SDS-PAGE 130

Figure 3-30 – Structure elucidation of 6-mercaptohexyl acrylic ester by NMR and FT-IR 133

7. List of supplementary figures

Figure S 1 – Vector map of pPICZαA,B,C (Invitrogen, USA) 180

Figure S 2 – Vector map of pET16b 180

Figure S 3 – Vector map of pET22b 181

Figure S 4 – Vector map of pET28a 181

Figure S 5 – Vector map of pET32a 182

Figure S 6 – Vector map of pTf16 182

Figure S 7 – Vector map of pG-Tf2 183

Figure S 8 – Vector map of pG-KJE8 183

Figure S 9 – Vector map of pJOE4782.1 184

Figure S 10 – Vector map of pRN754 184

Figure S 11 – Detailed view of the high molecular mass range of the PMF of glycosylated *At*DIR6 185

Figure S 12 – Detailed view of the PMFs in Figure 3-7 showing deglycosylated peptides after PNGaseF treatment 186

Figure S 13 – MALDI-TOF MS/MS-Spectra of *At*DIR6 glycopeptides and the corresponding deglycosylated peptides 187

8. List of tables

Table 2-1 – Plasmids 57

Table 2-2 – Strains 58

Table 2-3 – PCR and sequencing primers 58

Table 2-4 – PCR setup for the amplification of the *Atdir6* and *Rbk1* genes 61

Table 2-5 – PCR program for the amplification of the *Atdir6* and *Rbk1* genes 61

Table 2-6 – PCR setup for the integration control of the *Atdir*6 gene 62

Table 2-7 – PCR program for the integration control of the *Atdir*6 gene 62

Table 2-8 – QuikChange® PCR reaction setup 63

Table 2-9 – QuikChange® PCR program 63

Table 3-1 – Variants of the focused library 104

Table 3-2 – Substrate spectrum of RBK1 wild-type and variants 122

Table 3-3 – Comparison of the reaction rates from the coupled NADH assay and initial conversion rates from biotransformations with selected RBK1 variants and various substrates 123

Table 3-4 – Adducts and ions of reaction substrates and products formed during LC/MS analysis 125

Table 3-5 – Kinetic data with RBK1 wild-type and selected variants with various substrates127

Table 3-6 – Data collection during x-ray diffraction for RBK1 128

Table 3-7 – Kinetic parameters of HiC. 131

Table 3-8 – Substrate specificity of HiC in transesterification with various acyl donors and acceptors 132

9. Appendix

9.1 Supplementary figures

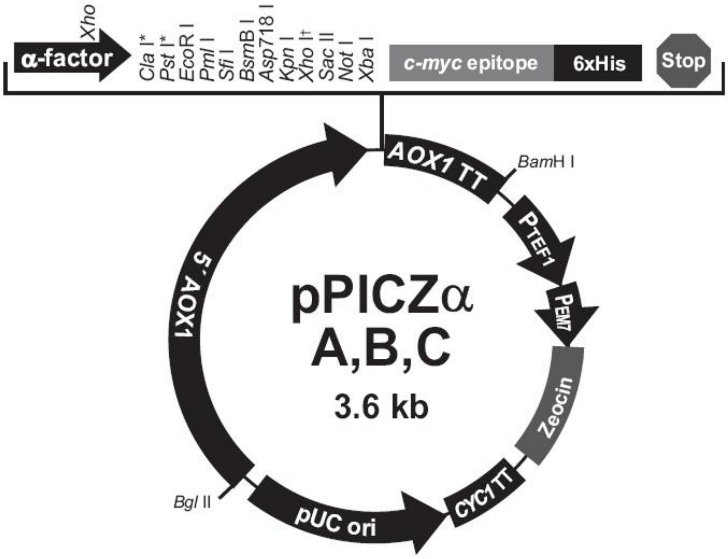
α-factor
Xho I†
Cla I*
Pst I*
EcoR I
Pml I
Sfi I
BsmB I
Asp718 I
Kpn I
Xho I†
Sac II
Not I
Xba I
c-myc epitope
6xHis
Stop
AOX1 TT
BamH I
PTEF1
PEM7
Zeocin
CYC1 TT
pUC ori
Bgl II
5´AOX1
pPICZα
A,B,C
3.6 kb

Figure S 1 – Vector map of pPICZαA,B,C (Invitrogen, USA). 5´AOX1: methanol-inducible AOX1 promoter, also targets integration to the AOX1 locus; α-factor: secretion signal peptide; c-myc epitope: permits detection of recombinant fusion protein with anti-myc antibody; 6xHis: Polyhistidine tag; AOX1 TT: AOX1 transcription terminator; PTEF1: TEF1 promoter for expression of Zeocin™ resistance gene in *P. pastoris*; PEM7: EM7 promoter for expression of Zeocin™ resistance gene in *E. coli*; Zeocin: Zeocin™ resistance gene; CYC1 TT: transcription terminator of the Zeocin™ resistance gene; pUC ori: origin of replication for replication and maintenance in *E. coli*.

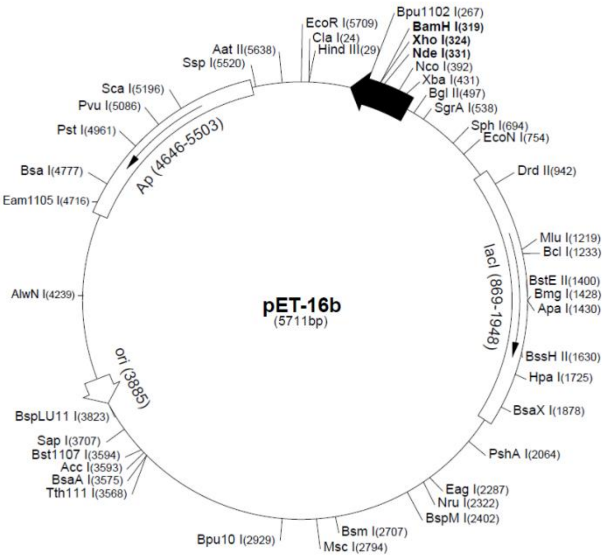
EcoR I(5709)
Cla I(24)
Hind III(29)
Bpu1102 I(267)
BamH I(319)
Xho I(324)
Nde I(331)
Nco I(392)
Xba I(431)
Bgl II(497)
SgrA I(538)
Sph I(694)
EcoN I(754)
Drd II(942)
Mlu I(1219)
Bcl I(1233)
BstE II(1400)
Bmg I(1428)
Apa I(1430)
BssH II(1630)
Hpa I(1725)
BsaX I(1878)
PshA I(2064)
Eag I(2287)
Nru I(2322)
BspM I(2402)
Bsm I(2707)
Msc I(2794)
Bpu10 I(2929)
Tth111 I(3568)
BsaA I(3575)
Acc I(3593)
Bst1107 I(3594)
Sap I(3707)
BspLU11 I(3823)
AlwN I(4239)
Eam1105 I(4716)
Bsa I(4777)
Pst I(4961)
Pvu I(5086)
Sca I(5196)
Ssp I(5520)
Aat II(5638)
Ap (4646-5503)
lacI (869-1948)
ori (3885)
pET-16b
(5711bp)

Figure S 2 – Vector map of pET16b (Merck, Darmstadt, Germany). Ap: ampicillin resistance gene; ori: pBR322 origin of replication; *lac*I: coding sequence for the *lac* repressor; black arrow: multiple cloning site. Furthermore, single-cutting restriction enzyme sites are shown. The numbers in brackets indicate the position of genes and restriction enzymes in base pairs. The sequence is numbered according to the pBR322 convention.

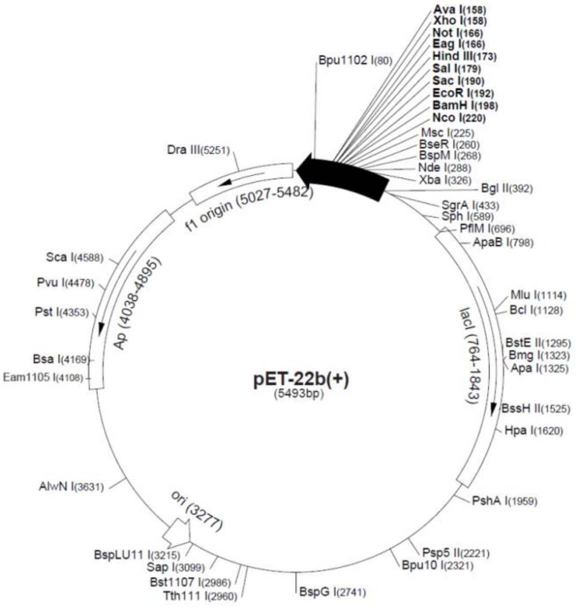
Ava I(158)
Xho I(158)
Not I(166)
Eag I(166)
Hind III(173)
Sal I(179)
Sac I(190)
EcoR I(192)
BamH I(198)
Nco I(220)
Msc I(225)
BseR I(260)
BspM I(268)
Nde I(288)
Xba I(326)
Bgl II(392)
SgrA I(433)
Sph I(589)
PflM I(696)
ApaB I(798)
Mlu I(1114)
Bcl I(1128)
BstE II(1295)
Bmg I(1323)
Apa I(1325)
BssH II(1525)
Hpa I(1620)
PshA I(1959)
Psp5 II(2221)
Bpu10 I(2321)
BspG I(2741)
Tth111 I(2960)
Bst1107 I(2986)
Sap I(3099)
BspLU11 I(3215)
AlwN I(3631)
Eam1105 I(4108)
Bsa I(4169)
Pst I(4353)
Pvu I(4478)
Sca I(4588)
Dra III(5251)
Bpu1102 I(80)
f1 origin (5027-5482)
Ap (4038-4895)
lacI (764-1843)
ori (3277)
pET-22b(+)
(5493bp)

Figure S 3 – Vector map of pET22b (Merck, Darmstadt, Germany). Ap: ampicillin resistance gene; ori: pBR322 origin of replication; *lac*I: coding sequence for the *lac* repressor; black arrow: multiple cloning site; f1 origin: F1 phage origin. Furthermore, single-cutting restriction enzyme sites are shown. The numbers in brackets indicate the position of genes and restriction enzymes in base pairs. The sequence is numbered according to the pBR322 convention.

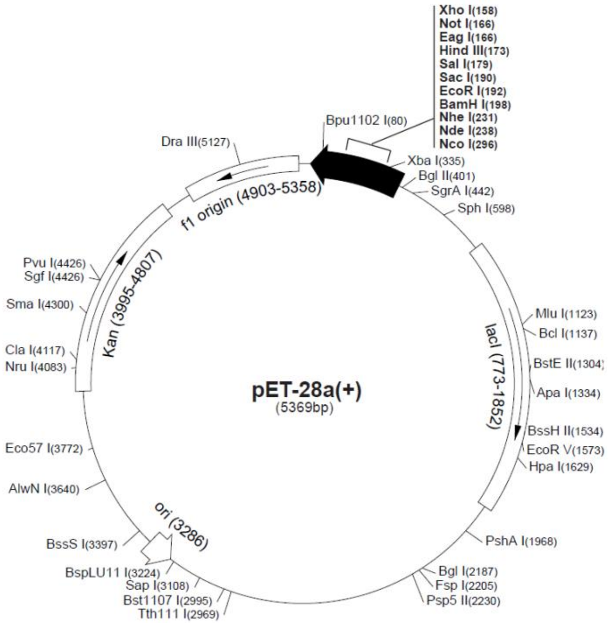
Xho I(158)
Not I(166)
Eag I(166)
Hind III(173)
Sal I(179)
Sac I(190)
EcoR I(192)
BamH I(198)
Nhe I(231)
Nde I(238)
Nco I(296)
Bpu1102 I(80)
Xba I(335)
Bgl II(401)
SgrA I(442)
Sph I(598)
Mlu I(1123)
Bcl I(1137)
BstE II(1304)
Apa I(1334)
BssH II(1534)
EcoR V(1573)
Hpa I(1629)
PshA I(1968)
Bgl I(2187)
Fsp I(2205)
Psp5 II(2230)
Tth111 I(2969)
Bst1107 I(2995)
Sap I(3108)
BspLU11 I(3224)
BssS I(3397)
AlwN I(3640)
Eco57 I(3772)
Nru I(4083)
Cla I(4117)
Sma I(4300)
Sgf I(4426)
Pvu I(4426)
Dra III(5127)
f1 origin (4903-5358)
Kan (3995-4807)
lacI (773-1852)
ori (3286)
pET-28a(+)
(5369bp)

Figure S 4 – Vector map of pET28a (Merck, Darmstadt, Germany). Kan: kanamycin resistance gene; ori: pBR322 origin of replication; *lac*I: coding sequence for the *lac* repressor; black arrow: multiple cloning site; f1 origin: F1 phage origin. Furthermore, single-cutting restriction enzyme sites are shown. The numbers in brackets indicate the position of genes and restriction enzymes in base pairs. The sequence is numbered according to the pBR322 convention.

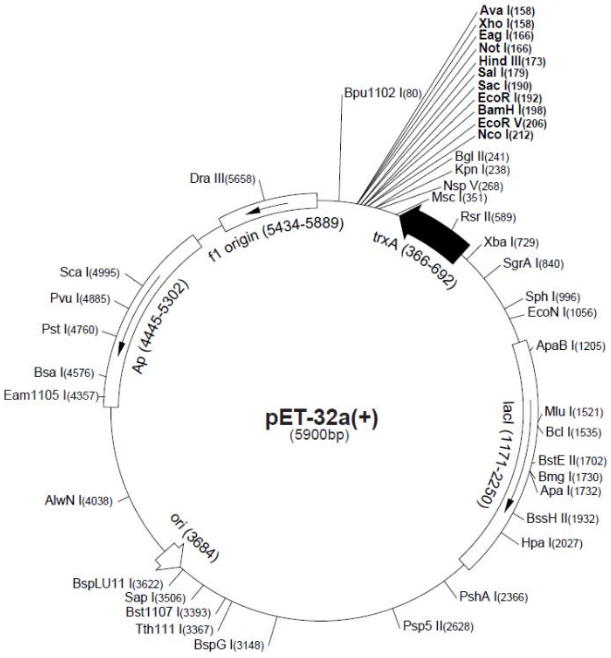
Ava I(158)
Xho I(158)
Eag I(166)
Not I(166)
Hind III(173)
Sal I(179)
Sac I(190)
EcoR I(192)
BamH I(198)
EcoR V(206)
Nco I(212)
Bpu1102 I(80)
Bgl II(241)
Kpn I(238)
Nsp V(268)
Msc I(351)
Dra III(5658)
Rsr II(589)
f1 origin (5434-5889)
trxA (366-692)
Xba I(729)
SgrA I(840)
Sca I(4995)
Pvu I(4885)
Pst I(4760)
Ap (4445-5302)
Sph I(996)
EcoN I(1056)
ApaB I(1205)
Bsa I(4576)
Eam1105 I(4357)
pET-32a(+)
(5900bp)
lacI (1171-2250)
Mlu I(1521)
Bcl I(1535)
BstE II(1702)
Bmg I(1730)
Apa I(1732)
AlwN I(4038)
ori (3684)
BssH II(1932)
Hpa I(2027)
BspLU11 I(3622)
Sap I(3506)
Bst1107 I(3393)
Tth111 I(3367)
BspG I(3148)
Psp5 II(2628)
PshA I(2366)

Figure S 5 – Vector map of pET32a (Merck, Darmstadt, Germany). Ap: ampicillin resistance gene; ori: pBR322 origin of replication; *lac*I: coding sequence for the *lac* repressor; *trx*A: thioredoxin gene included in the multiple cloning site (black arrow); f1 origin: F1 phage origin. Furthermore, single-cutting restriction enzyme sites are shown. The numbers in brackets indicate the position of genes and restriction enzymes in base pairs. The sequence is numbered according to the pBR322 convention.

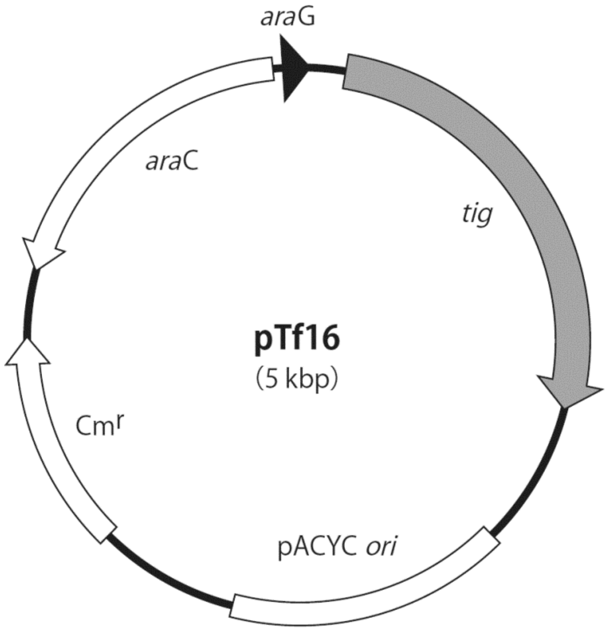
araG
araC
tig
pTf16
(5 kbp)
Cmr
pACYC ori

Figure S 6 – Vector map of pTf16 (Takara, Japan). *ara*G: *ara*G promoter induced by L-arabinose; *ara*C: gene for AraC repressor protein; Cm^r: chloramphenicol resistance gene; pACYC ori: pACYC origin of replication; *tig*: gene for *E. coli* trigger factor protein.

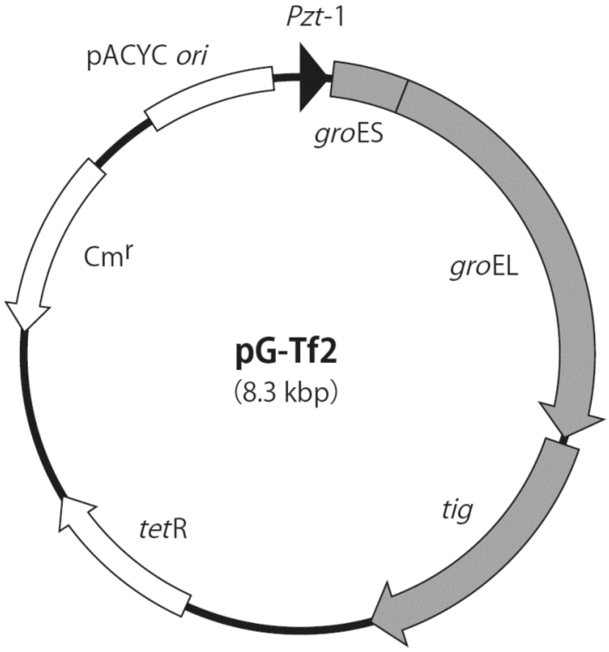
Pzt-1
pACYC ori
groES
groEL
Cmr
pG-Tf2
(8.3 kbp)
tig
tetR

Figure S 7 – Vector map of pG-Tf2 (Takara, Japan). Pzt-1: tetracycline-inducible promoter; pACYC ori: : pACYC origin of replication; Cm^r: chloramphenicol resistance gene; *tet*R: gene for Tet repressor protein; *tig*: gene for *E. coli* trigger factor protein; *gro*EL: gene for GroEL chaperone protein; *gro*ES: gene for GroES chaperone protein.

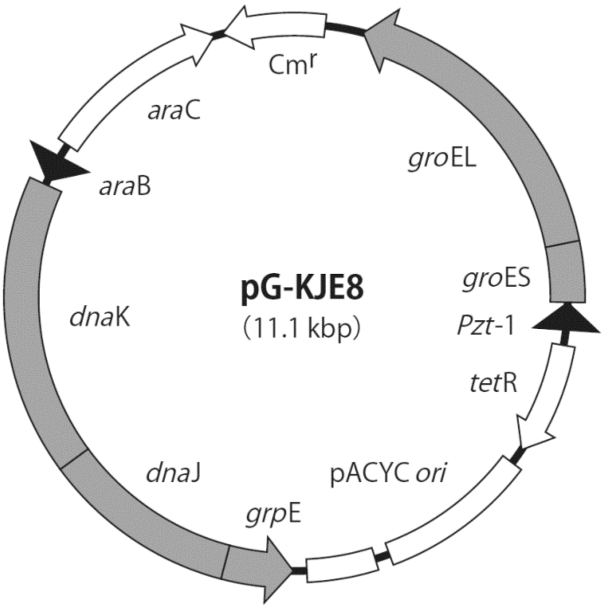
Cmr
araC
groEL
araB
pG-KJE8
(11.1 kbp)
groES
Pzt-1
dnaK
tetR
dnaJ
pACYC ori
grpE

Figure S 8 – Vector map of pG-KJE8 (Takara, Japan). Cm^r: chloramphenicol resistance gene; araC: gene for AraC repressor protein; *ara*B: *ara*B promoter induced by L-arabinose; *dna*K: gene for DnaK chaperone protein; *dna*J: gene for DnaJ chaperone protein; *grp*E: gene for GrpE chaperone protein; pACYC ori: pACYC origin of replication; *tet*R: gene for Tet repressor protein; Pzt-1: tetracycline-inducible promoter; *gro*ES: gene for GroES chaperone protein; *gro*EL: gene for GroEL chaperone protein.

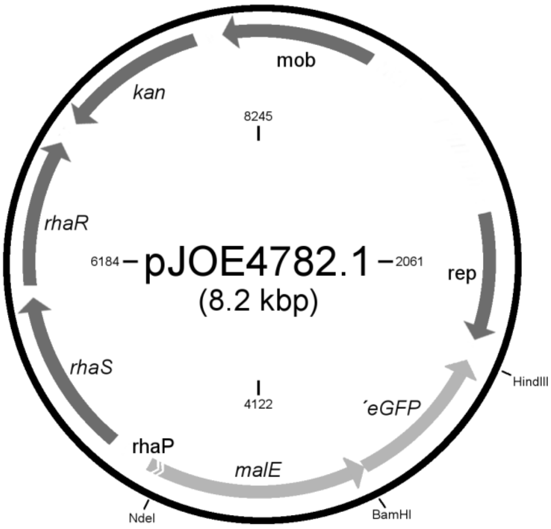
mob
kan
8245
rhaR
6184
pJOE4782.1
(8.2 kbp)
2061
rep
rhaS
4122
´eGFP
HindIII
rhaP
malE
NdeI
BamHI

Figure S 9 – Vector map of pJOE4782.1. Mob: mobilization site (origin of transfer); *kan*: kanamycin resistance gene; *rha*R: transcriptional activator of the rhamnose operon; *rha*S: transcriptional activator of the rhamnose operon; *rha*P: L-rhamnose promoter; *mal*E: gene for maltose-binding protein; *´eGFP*: gene for enhanced green fluorescent protein; rep: pBBR1 origin of replication. Restriction sites NdeI and HindIII were used for cloning thereby eliminating the originally existing malE and ´eGFP genes. The vector was kindly provided by J. Altenbuchner (University of Stuttgart) [212,214].

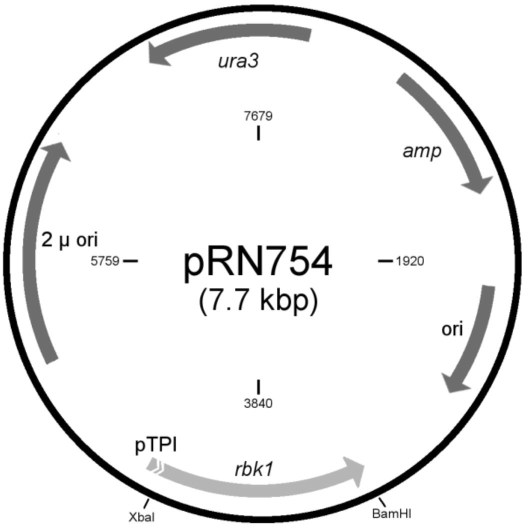
ura3
7679
amp
2 µ ori
5759
pRN754
(7.7 kbp)
1920
ori
3840
pTPI
rbk1
XbaI
BamHI

Figure S 10 – Vector map of pRN754. *Ura*3: gene for orotidine 5-phosphate decarboxylase (auxotrophy selection marker); 2 µ ori: yeast origin of replication; pTPI: triose-phosphate-isomerase promoter; *rbk1*: gene for *S. cerevisiae* ribokinase; ori: *E. coli* origin of replication; *amp*: ampicillin resistance gene. The *rbk1* gene was introduced into the vector using restriction sites XbaI and BamHI. The vector was provided by DSM, Netherlands.

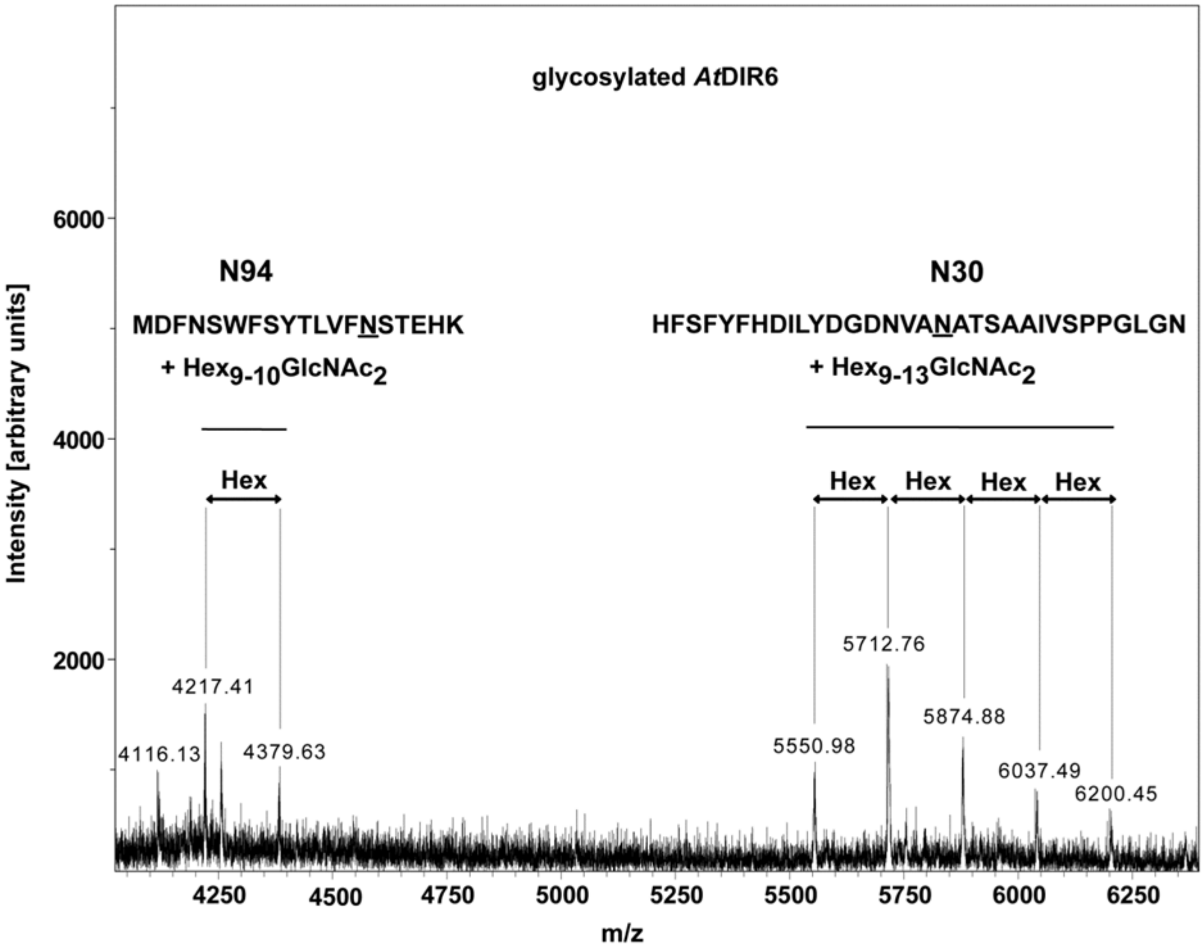

Figure S 11 – Detailed view of the high molecular mass range of the PMF of glycosylated *At*DIR6 (adapted from [215]**).** The figure was prepared by Dr. Jens Pfannstiel (Proteomics Core Facility, Life Science Center, University of Hohenheim). In the high molecular mass range of the PMF of glycosylated *At*DIR6 characteristic groups of glycopeptide ion signals spaced by 162 Da (Hexose) were observed. m/z values of these signals corresponded to tryptic peptides glycosylated at N30 and N94 with attached glycans structures of the high mannose type Hex_{9-13} $GlcNAc_2$. The peptide sequences and the predicted glycan structures are indicated. Peptide sequence and glycan structure of the $Hex_{10}GlcNAc_2$ glycoforms were verified by MS/MS analysis (supplementary Figure S 13a and b).

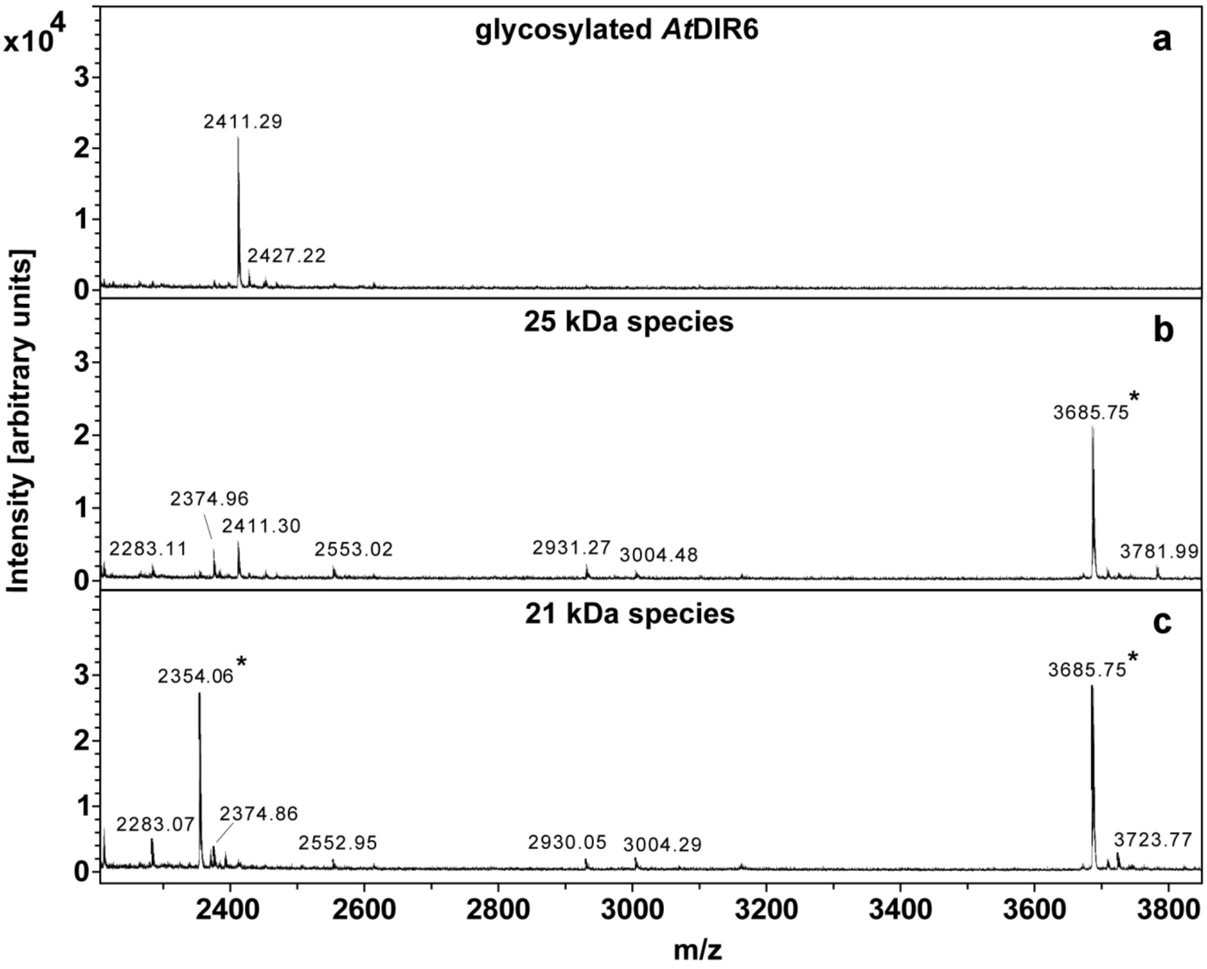

Figure S 12 – Detailed view of the PMFs in Figure 3-7 showing deglycosylated peptides after PNGaseF treatment (adapted from [215]**).** The figure was prepared by Dr. Jens Pfannstiel (Proteomics Core Facility, Life Science Center, University of Hohenheim). Deglycosylated peptide ions corresponding to the glycosylation sites at N30 and N94 are marked by an asterisk. In the PMF of glycosylated *At*DIR6 (**a**) none of the deglycosylated peptides could be identified. After PNGaseF treatment a 25 kDa and a 21 kDa band were detected in SDS-gels. In the PMF analysis of the 25 kDa band (**b**) only a peptide ion corresponding to the deglycosylated peptide including N30 was detected. In the PMF of the 21 kDa band (**c**) peptide ions corresponding to deglycosylated peptides including both sites N30 and N94 were detected. Peptide sequences were verified by MS/MS analysis (supplementary Figure S 13c and d).

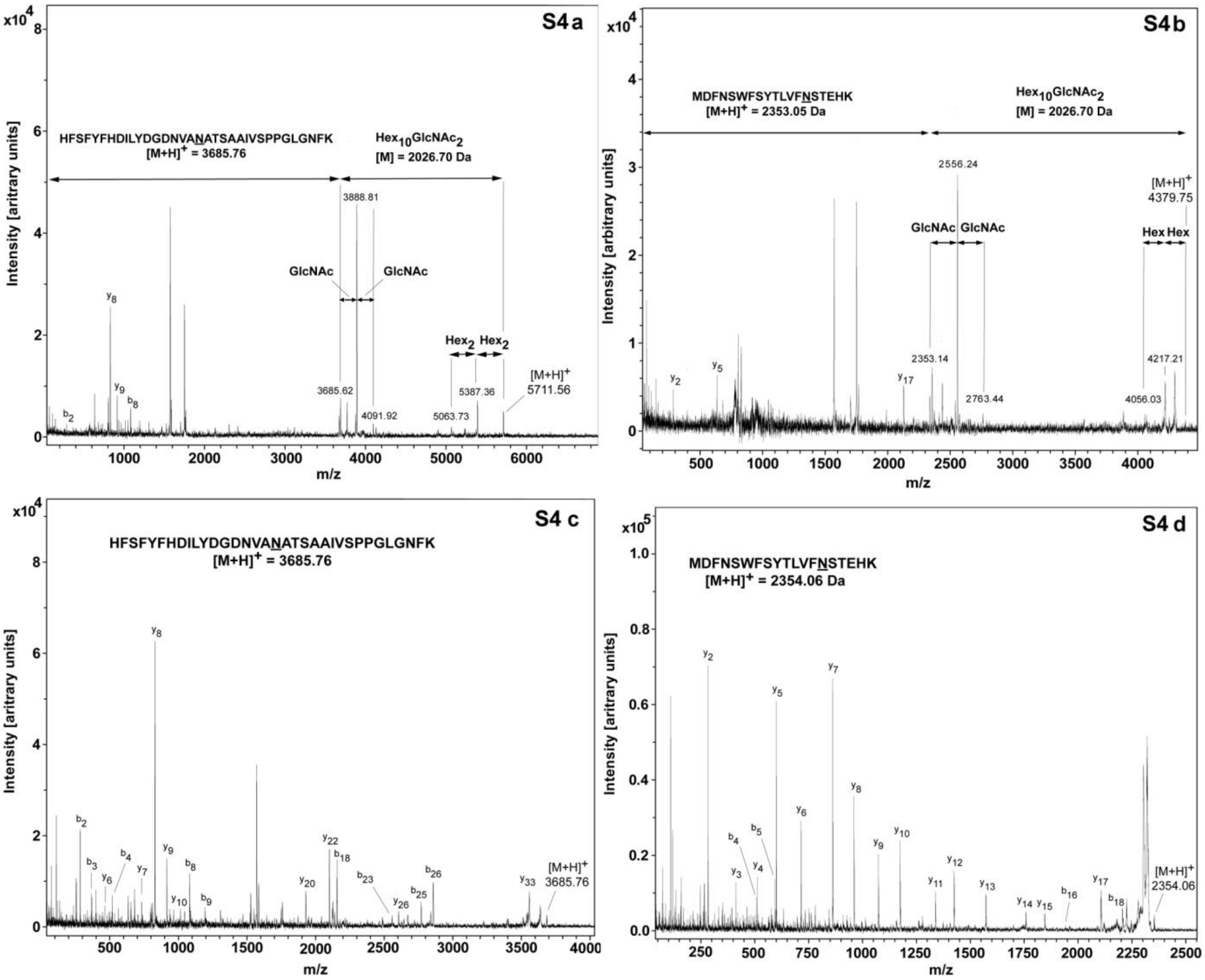

Figure S 13 – MALDI-TOF MS/MS-Spectra of *At*DIR6 glycopeptides and the corresponding deglycosylated peptides (modified from [215]**).** The figure was prepared by Dr. Jens Pfannstiel (Proteomics Core Facility, Life Science Center, University of Hohenheim). MALDI-TOF MS/MS analysis established the presence of glycans on N34 (**a**), and N90 (**b**). Glycan compositions and the fragment ions corresponding to oligosaccharide chain fragmentation are shown. The glycan composition depicted from the spectra is indicated. The amino acid sequence of the tryptic peptides is shown and the glycosylation site is underlined. The C- and N-terminal peptide fragment ions (y-ion and b-ion series, respectively) are indicated. The presence of the deyglycosylated peptides including N30 (**c**) and N94 (**d**) after PNGaseF treatment was verified by MS/MS analysis. The peptide sequence is shown above the spectrum and the deglycosylated Asparagines are underlined. The C- and N-terminal peptide fragment ions (y-ion and b-ion series, respectively) are indicated.

9.2 Nucleotide sequences

9.2.1 *At*DIR6 codon optimized for *Escherichia coli*

The open reading frame (ORF) for *At*DIR6 from *Arabidopsis thaliana* was obtained as synthetic gene from GeneArt (Germany) with NdeI and BamHI restriction sites flanking the gene, without the N-terminal secretion signal sequence and codon-optimized for *E. coli*. The sequence is shown below:

GGAATTCCATATGTTTCGTAAAACCATTGATCAGAAAAAACCGTGCAAACATTTTAGCTT

TTATTTTCATGATATTCTGTATGATGGCGATAATGTTGCAAATGCAACCAGCGCAGCAAT

TGTTAGTCCTCCGGGTCTGGGTAATTTCAAATTTGGCAAATTCGTGATTTTTGATGGTCC

GATTACCATGGATAAAAATTATCTGAGCAAACCGGTTGCACGTGCACAGGGTTTTTATTT

CTATGATATGAAAATGGATTTTAATAGCTGGTTTAGCTATACCCTGGTGTTTAATAGCAC

CGAACATAAAGGCACCCTGAATATTATGGGTGCAGATCTGATGATGGAACCGACCCGTGA

TCTGAGCGTTGTTGGTGGCACCGGTGATTTTTTTATGGCACGTGGTATTGCAACCTTTGT

TACCGACCTGTTTCAGGGTGCCAAATATTTTCGCGTGAAAATGGATATTAAACTGTATGA

ATGCTATGGATCCCG

9.2.2 *At*DIR6 codon optimized for *Pichia pastoris*

The ORF for *At*DIR6 from *Arabidopsis thaliana* was obtained as synthetic gene from GeneArt (Germany) with EcoRI and XbaI restrictions sites flanking the gene, without the N-terminal secretion signal sequence and codon-optimized for *P. pastoris*. The sequence is shown below:

GGAATTCTTCAGAAAGACTATCGACCAGAAGAAGCCTTGTAAGCACTTCTCCTTCTACTT

CCACGACATCTTGTACGACGGTGACAACGTTGCTAACGCTACTTCTGCTGCTATCGTTTC

TCCACCAGGTTTGGGTAACTTCAAGTTCGGTAAGTTCGTTATCTTCGACGGTCCAATCAC

TATGGACAAGAACTACTTGTCCAAGCCAGTTGCTAGAGCACAGGGTTTCTACTTCTACGA

TATGAAGATGGACTTCAACTCCTGGTTCTCCTACACTTTGGTTTTCAACTCCACTGAGCA

CAAGGGTACTTTGAACATCATGGGTGCTGACTTGATGATGGAACCTACTAGAGACTTGTC

CGTTGTTGGTGGTACTGGTGACTTCTTCATGGCTAGAGGTATCGCTACTTTCGTTACTGA

CTTGTTCCAGGGTGCTAAGTACTTCAGAGTTAAGATGGACATCAAGTTGTACGAGTGTTA

CGCTCTAGAGC

9.2.3 Ribokinase RBK1 from *Saccharomyces cerevisiae*

The gene for RBK1 from *Saccharomyces cerevisiae* was amplified with PCR from the construct pRN754 (DSM, Netherlands) using primers harboring NdeI and HindIII restriction sites. The sequence is shown below:

GGAATTCCATATGGGTATTACAGTAATAGGTTCTCTAAACTATGATTTGGACACATTTAC

GGATAGATTACCTAACGCTGGAGAAACTTTCAGGGCTAACCACTTCGAAACACATGCTGG

TGGTAAGGGATTGAACCAAGCTGCGGCCATTGGTAAATTAAAAAACCCCAGCAGCAGATA

TAGTGTTCGAATGATTGGTAATGTTGGAAATGATACATTTGGTAAACAATTGAAGGACAC

TTTATCCGATTGCGGAGTCGATATCACTCACGTCGGTACTTACGAAGGCATTAATACGGG

TACCGCTACCATACTAATTGAAGAGAAAGCTGGTGGCCAAAATAGGATATTGATTGTAGA

AGGTGCTAACAGCAAGACTATTTATGACCCGAAACAGTTGTGTGAAATTTTTCCAGAGGG

CAAGGAGGAAGAAGAGTATGTTGTTTTTCAACACGAAATTCCTGATCCTCTTTCCATTAT

TAAATGGATACATGCGAACAGGCCGAATTTTCAGATCGTATATAACCCCTCACCTTTCAA

GGCCATGCCTAAGAAAGATTGGGAGTTGGTAGACCTTTTGGTCGTTAATGAAATTGAGGG

TCTTCAAATCGTGGAAAGTGTATTTGATAATGAACTTGTTGAAGAAATAAGGGAGAAGAT

AAAGGACGACTTTTTAGGAGAATATCGTAAAATTTGTGAGCTTTTGTATGAAAAACTCAT

GAATCGAAAGAAAAGAGGAATTGTGGTTATGACTTTGGGTTCGAGAGGGGTGCTTTTCTG

TTCGCACGAAAGCCCTGAAGTACAATTCCTTCCGGCTATTCAAAATGTTTCGGTTGTTGA

TACTACAGGAGCTGGAGATACTTTCCTGGGCGGTTTGGTTACTCAATTGTATCAAGGAGA

GACCTTGTCTACGGCTATAAAGTTCTCTACATTAGCTAGTTCATTGACCATTCAAAGAAA

AGGTGCTGCTGAAAGCATGCCACTGTATAAAGATGTTCAGAAAGATGCATAGAAGCTTGG

G

9.2.4 Cutinase HiC from *Humicola insolens*

The gene for HiC from *Humicola insolens* was available from a previous study [198]. The sequence is shown below:

GGCGCAATCGAAAACGGCCTGGAATCCGGCTCCGCAAACGCATGCCCAGATGCAATCCTG

ATCTTCGCACGCGGCTCCACCGAACCAGGCAACATGGGCATCACCGTGGGCCCAGCACTG

GCAAACGGCCTGGAATCCCACATCCGCAACATCTGGATCCAGGGCGTGGGCGGCCCATAC

GATGCAGCACTGGCAACCAACTTCCTGCCACGCGGCACCTCCCAGGCAAACATCGATGAA

GGCAAGCGCCTGTTCGCACTGGCAAACCAGAAGTGCCCAAACACCCCAGTGGTGGCAGGC

GGCTACTCCCAGGGCGCAGCACTGATCGCAGCAGCAGTGTCCGAACTGTCCGGCGCAGTG

AAGGAACAGGTGAAGGGCGTGGCACTGTTCGGCTACACCCAGAACCTGCAGAACCGCGGC

GGCATCCCAAACTACCCACGCGAACGCACCAAGGTGTTCTGCAACGTGGGCGATGCAGTG

TGCACCGGCACCCTGATCATCACCCCAGCACACCTGTCCTACACCATCGAAGCACGCGGC

GAAGCAGCACGCTTCCTGCGCGATCGCATCCGCGCA

10. Curriculum vitae

Persönliche Daten

Name:	Christian Kazenwadel
Geburtsdatum/-Ort	25.10.1982 in Schorndorf, Deutschland
Nationalität:	deutsch
Email	christian.kazenwadel@itb.uni-stuttgart.de

Ausbildung

seit 02/2010	**Promotion bei Prof. Dr. Bernhard Hauer am Institut für Technische Biochemie der Universität Stuttgart** *Dissertation: „ Challenges in Biocatalysis – Case Studies in Protein Expression, Engineering, and Application“*
05/2009 – 01/2010	**Diplomarbeit bei Prof. Dr. Bernhard Hauer am Institut für Technische Biochemie der Universität Stuttgart** *Diplomthema: “Transesterification of acrylate ester monomers with immobilized* Humicola insolens *cutinase”*
03/2008 – 09/2008	**Studienarbeit bei Prof. Phillip Coffino am Department of Microbiology & Immunology der University of California, San Francisco, USA** *Thema: “An evaluation of ornithine decarboxylase interactions with the 26S proteasome”*
10/2003 – 01/2010	**Studium der Technischen Biologie an der Universität Stuttgart** *Abschluss: Diplom Biologe technisch orientiert (t.o.)*
09/1993 – 06/2002	**Max-Planck Gymnasium Schorndorf** *Abitur*
09/1989 – 08/1993	**Grundschule Adelberg**