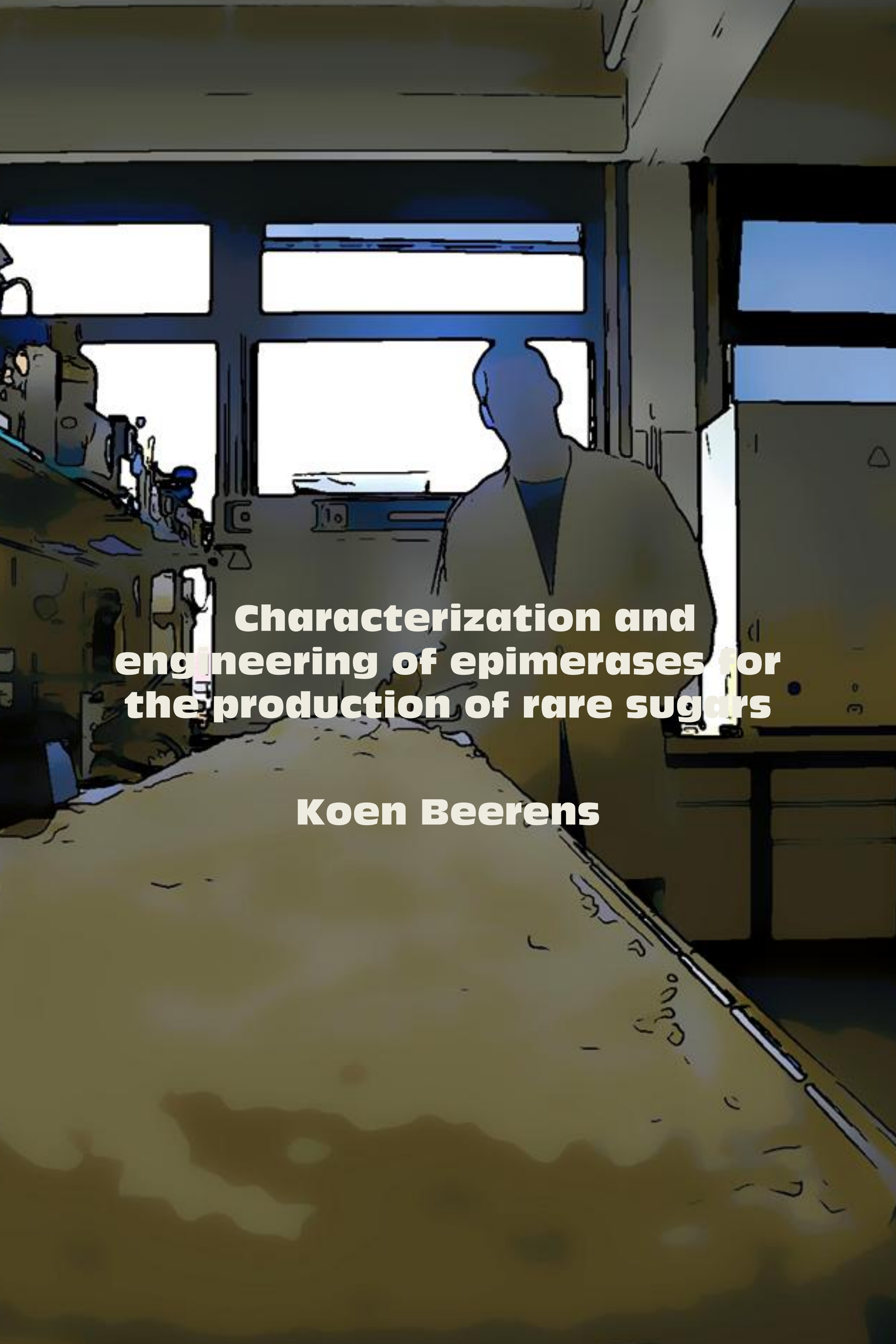


Characterization and engineering of epimerases for the production of rare sugars

Koen Beerens

Members of the jury: Prof. dr. ir. Wim Soetaert (promotor)
Prof. dr. Tom Desmet (promotor)
Prof. dr. Els Van Damme (secretary)
Prof. dr. ir. Chris Stevens (chairman)
Em. Prof. dr. ir. Erick Vandamme
Em. Prof. dr. Matti Leisola
dr. Jo Maertens

Promotors: Prof. dr. ir. Wim Soetaert
Prof. dr. Tom Desmet
Centre of Expertise - Industrial Biotechnology and Biocatalysis (InBio.be)
Department of Biochemical and Microbial Technology
Ghent University, Belgium

Dean: Prof. dr. ir. Guido Van Huylenbroeck

Rector: Prof. dr. Paul Van Cauwenberge

Koen Beerens was supported by a SB-fellowship (SB81309 and SB83309) of the Agency for Innovation by Science and Technology in Flanders (IWT-Vlaanderen). The research was conducted at the Centre of Expertise - Industrial Biotechnology and Biocatalysis, Department of Biochemical and Microbial Technology, Ghent University (Ghent, Belgium).

Koen Beerens

CHARACTERIZATION AND ENGINEERING OF EPIMERASES FOR THE PRODUCTION OF RARE SUGARS

Thesis submitted in fulfilment of the requirements for the degree of
Doctor (PhD) in Applied Biological Sciences (Cell and gene biotechnology)

Dutch translation of the title:
Karakterisering en engineering van epimerasen voor de productie van zeldzame koolhydraten

To refer to this thesis:

Beerens, K. (2013) . Characterization and engineering of epimerases for the production of rare sugars. PhD thesis, Faculty of Bioscience Engineering, Ghent University, Ghent, 198 p

Printed by University Press, Zelzate

ISBN-number: 978-90-5989-588-1

Copyright © 04/02/2013 by Koen Beerens. All rights reserved.

The author and the promotors give the authorization to consult and to copy parts of this work for personal use only. Every other use is subject to the copyright laws. Permission to reproduce any material contained in this work should be obtained from the author.

Dankwoord

De laatste weken heb ik ettelijke keren volgende of een gelijkaardige zin zien staan in de verschillende vacatures die ik bekeken heb: “Je kan zelfstandig werken, maar je bent ook een echte teamplayer”. Deze zin, hoe alledaags en triviaal die ook mag klinken, is misschien wel de kortste samenvatting die je van een doctoraatsonderzoek kan maken. Een doctoraat is immers een persoonlijk project maar wel eentje dat je niet kan realiseren zonder de steun en hulp van vrienden, familie en collega’s. Bij deze wil ik dan ook iedereen bedanken die me de voorbije vier jaar heeft gesteund en geholpen.

Eerst en vooral wil ik mijn promotoren Prof. dr. ir. Wim Soetaert en Prof. dr. Tom Desmet bedanken om vijf jaar geleden in een pasafgestudeerde biochemicus te geloven en zo mijn eerste werkervaring te geven. Maar ook voor de kans om nadien mijn eigen doctoraatsonderzoek te starten en mij hier het nodige vertrouwen en geduld voor te geven. Wim, jou wil ik nog bedanken om me destijds een andere kant van onderzoek te laten zien, namelijk het toegepaste i.p.v. het fundamentele onderzoek alsook om me de wondere wereld van suikers en de kristallisatie hiervan te laten ontdekken. Tom, voor jou een woordje van dank voor alle hulp bij mijn beursaanvraag, het schrijven van mijn artikels, de bemoedigende woorden in ontmoedigende tijden, me de kans te geven om in Cambridge over mijn onderzoek te gaan spreken, maar uiteraard ook voor minder wetenschappelijke momenten op congressen.

De volledige “Glycodirect”-groep wil ik bedanken voor de leuke sfeer op het bureau en het labo, maar ook voor de wetenschappelijke discussies. Laten we maar beginnen bij de intussen ex-collega’s. Manu, bedankt om me in te leiden in de *enzyme engineering* alsook in de automatisatie hiervan. Ook voor Eefje een dankuwel voor de hulp en uitleg bij InBios robotica. An, veel succes bij Ablynx en dat je er op de volgende InBio-(after)party wel mag geraken. Giang, let’s just thank you for being a €5-communist, your giggly smile and the delicious Vietnamese snacks. Chen Chao, Chen or Chao, I will never forget the amazing “Yes, but …” moment. Jef, zonder jou is de humor een stuk minder droog op de Glycodirect-bureau. Dirk, merci voor de vijf leuke jaren. Barbara, een superdikke merci voor al jouw hulp met de ontwikkeling van mijn screeningsassay en het andere werk op de epimerasen. En dat jullie nieuwe spruit je niet te veel slapeloze nachten bezorgt zodat je die cellulasen kan bedwingen. Karel, met jou in de buurt is er altijd veel plezier te beleven (en schunnige praat te horen). Tom, merci om op de goede (en soms minder gepaste) momenten de boel op te vrolijken met “Statler en Waldorf”-gewijze opmerkingen en de grapjes tot ver buiten de Glycodirect-bureau. De nieuwkomers Magali en Renfei zou ik vooral veel succes, geduld en doorzettingsvermogen willen toewensen. Daarnaast wil ik ook mijn twee thesisstudenten, Simon en Koen, bedanken voor de hulp bij het praktische werk van mijn onderzoek, ook al heb ik door jullie mijn eerste grijze haren gekregen. Iedereen veel succes met jullie onderzoek en/of verdere carrière en dat we elkaar zeker nog ergens mogen tegenkomen!

Verder wil ik natuurlijk ook alle andere (ex-)collega’s bedanken voor de leuke jaren samen op Limab/InBio. Emeritus Prof. dr. ir. Erick Vandamme, bedankt om voor mij jouw toga en wetenschappelijke kennis nog eens boven te halen. De mensen van het secretariaat voor de hulp met al het papierwerk. De geweldige gazet Liesbet, succes met je studies. Barbara, bedankt voor het last-minute nalezen van mijn doctoraat. Alle laboranten, doctoraatsstudenten en post-docs zowel voor hun hulp bij onderzoeksgerelateerde zaken als voor de geweldige sfeer die er altijd is geweest op InBio (en hopelijk nog lang mag blijven). Wouter aka dj Wally om steeds weer voor de ambiance te zorgen. Eric, merci voor de oppeppende gesprekken en de (zondag)avonden in ’t Spijker en de Hot Club de Gand. Kwok, Maarten, Catherine, Inge en Sophie voor de verticale ontspanning (lees: klimmuur en bijpassende Gulden Draak). Lien, wat een afscheidsfeestje om nadien gewoon weer terug te keren. Pieter, Gert, Frederik en Robin, dat er meer ups dan downs mogen zijn in jullie onderzoek de komende jaren. Sophie, veel succes met de laatste loodjes van jouw doctoraat! Marjan, supervrouw hier en in Amerika. Joeri, iemand die zowat alles weet staan en kent op InBio, merci. Jo, bedankt voor de leuke discussies en om je door mijn doctoraat te wurmen. Veel succes met jullie spin-off. Sofie, voor de hulp bij moleculair werk en de grappige momenten in het moleculair labo. En alle andere (ex-)InBio’tjes, bedankt voor alles! En natuurlijk de spelers en supporters van de InBio- en FBW-voetbalploegen niet vergeten bedanken voor zowel de sportieve als vloeibare ontspanning en uiteraard voor de kampioenschapstitel.

Als laatste wil ik ook mijn ouders, familie en vrienden bedanken voor alles wat jullie voor mij betekenen en gedaan hebben. Moeke en vake, een dikke knuffel voor jullie onvoorwaardelijke steun. Dries, meestal ver weg maar er valt altijd op je te rekenen. Jan en Marhaba, merci voor de steun in moeilijke tijden. Jelle, Sten, Flip, Ruben, Huub alsook de andere mannen van “den OLK” en “De Mart” om op tijd en stond de batterijen helemaal te ledigen en weer op te laden. Charlotte, bedankt om in de laatste weken van mijn doctoraat voor ontspanning en steun te zorgen.

Table of contents

I. INTRODUCTION 1

II. LITERATURE REVIEW 7

1 BIOCATALYTIC PRODUCTION ROUTES FOR RARE SUGARS 9

1.1. Rare sugars and their applications 9

1.2. Enzymes for rare sugar production 12

1.2.1. Oxidoreductases 17

1.2.2. Isomerases 18

1.2.2.1. L-arabinose isomerase for tagatose production 19

1.2.2.2. Isomerases for L-ribose production 20

1.2.2.3. Isomerases for deoxygenated and other modified sugars 20

1.2.3. Carbohydrate epimerases 20

1.2.3.1. Ketohexose 3-epimerase 21

1.2.3.2. UDP-Galactose 4-epimerase 21

1.2.3.3. Cellobiose 2-epimerase 22

1.2.3.4. L-ribulose-5-phosphate 4-epimerase and related aldolases 22

1.3. Rare sugars: conclusions and outlook 23

2 THE RARE SUGAR TARGET: D-TAGATOSE 24

2.1. Introduction 24

2.2. Properties 24

2.3. Production methods 26

2.3.1. Chemical 26

2.3.2. Biochemical 27

2.4. Applications 29

3 (SUGAR) EPIMERASES 31

3.1. General 31

3.2. L-ribulose-5-phosphate 4-epimerase 32

3.2.1. General 32

3.2.2. Mechanism 33

3.2.3. Structure 36

3.2.3.1. Tertiary and quaternary structure 36

3.2.3.2. Cation (zinc) binding site 37

3.2.3.3. Substrate binding site and potential catalytic acid/base residues 38

3.3. UDP-Galactose 4-epimerase 40

3.3.1. General 40

3.3.2. Structure 41

3.3.3. Mechanism 43

3.3.4. Substrate promiscuity 45

3.3.4.1. UDP-Glc/UDP-Gal and/or UDP-GlcNAc/UDP-GalNAc (N-acetylation of UDP-sugars) 46

3.3.4.2. UDP-pentose and UDP-uronic acid 4-epimerases 49

4 ENZYME ENGINEERING 51

4.1. Introduction 51

4.2. *Enzyme engineering and its types* ... *51*
4.2.1. Rational design ... 53
4.2.2. Directed evolution ... 53
4.2.3. Semi-rational design ... 54
4.2.4. *De novo* protein design ... 54
4.2.5. Selection vs. Screening ... 55
4.3. *Chemical modification of enzymes* ... *55*
5 CONCLUSION ... 56

III. CLONING AND EXPRESSION OF L-RIBULOSE-5-PHOSPHATE 4-EPIMERASE ... 57
1 INTRODUCTION ... 59
2 MATERIAL & METHODS ... 60
2.1. *Bacterial strains, growth conditions, plasmids and chemicals* ... *60*
2.2. *Construction of the expression vectors* ... *61*
2.2.1. Construction of pIXPtrc ... 61
2.2.2. Cloning the genes into pIXPtrc and pCXhPxx (Figure III.1) ... 61
2.3. *(Optimization of) Recombinant enzyme expression* ... *63*
2.4. *L-Ribulose-5-phosphate production and purification* ... *64*
2.5. *HPLC analysis of L-Ru-5-P* ... *65*
2.6. *Enzyme activity assays: in vivo and in vitro* ... *66*
3 RESULTS AND DISCUSSION ... 67
3.1. *Construction of the expression vectors* ... *67*
3.1.1. Preparation of the pIXhPtrc vector ... 67
3.1.2. pIXPtrc-AraD and pCXhPxx-AraD ... 67
3.2. *Recombinant L-ribulose-5-phosphate 4-epimerase expression* ... *68*
3.3. *Ribulokinase production, L-ribulose-5-phosphate production and purification70*
3.4. *Enzyme activity assays: in vivo and in vitro* ... *71*
4 CONCLUSION ... 73

IV. DEVELOPMENT OF A SELECTION SYSTEM FOR TAGATOSE 4-EPIMERASE ACTIVITY ... 75
1 INTRODUCTION ... 77
2 MATERIAL & METHODS ... 78
2.1. *Bacterial strains, plasmids, growth conditions and chemicals* ... *78*
2.2. *Construction of the SelTag-strains by gene disruption* ... *79*
2.2.1. Gene disruption ... 80
2.2.2. Verification by colony PCR ... 80
2.2.3. Elimination of the antibiotic resistance gene ... 82
2.2.4. Speeding up the gene disruption protocol ... 82
2.3. *Growth tests on different minimal media* ... *83*
2.3.1. Medium composition ... 83
2.3.2. Wash and growth conditions ... 83
3 RESULTS AND DISCUSSION ... 83
3.1. *Construction of the SelTag-strains by gene disruption* ... *83*
3.2. *Speeding up the gene disruption protocol* ... *84*
3.3. *Growth tests on different minimal media* ... *85*
3.3.1. SelTag1 = *E. coli* CGSC#10993 ΔsgbE ΔulaF ... 85

3.3.2. SelTag2 (3) = *E. coli* K-12 MG1655 ΔaraD ΔsgbE Δula (ulaF::kan) 86
4 CONCLUSION ... 87
V. MUTAGENESIS OF L-RU-5-P 4-EPIMERASE AND SELECTION TOWARDS IMPROVED MUTANTS ... 89
1 INTRODUCTION ... 91
2 MATERIAL & METHODS ... 92
2.1. Bacterial strains, plasmids, growth conditions and chemicals ... 92
2.2. Homology model of the L-Ru-5-P 4-epimerase ... 92
2.3. Random mutagenesis: error prone PCR ... 93
2.4. (Semi-)Rational mutagenesis: Site saturation mutagenesis ... 93
2.5. Selection of mutant libraries ... 95
2.5.1. Liquid minimal medium ... 95
2.5.2. Solid minimal medium ... 95
3 RESULTS AND DISCUSSION ... 95
3.1. Sequence analysis and homology model ... 95
3.2. Determining residues for SSM (Figure V.2) ... 97
3.3. Random mutagenesis: library quality and selection ... 98
3.4. Site saturation mutagenesis: library quality and selection ... 99
3.5. Solid minimal medium ... 102
3.6. Fructokinase problem ... 102
4 CONCLUSION ... 104
VI. DEVELOPMENT AND APPLICATION OF A SCREENING ASSAY ... 107
1 INTRODUCTION ... 109
2 MATERIAL & METHODS ... 109
2.1. Bacterial strains, plasmids, growth conditions ... 109
2.2. Chemicals and enzymes ... 109
2.3. Development of the screening assay for tagatose 4-epimerase activity 110
2.3.1. Soluble glucose isomerase ... 110
2.3.2. Glucose oxidase – Peroxidase (GOD–POD) ... 110
2.3.3. Coupling SGI and GOD-POD ... 110
2.4. Screening the libraries ... 111
3 RESULTS AND DISCUSSION ... 111
3.1. Development of the screening assay for tagatose 4-epimerase activity 111
3.2. Overview of the screening assay (Figure VI.2) ... 113
3.3. Standard curve and CV ... 114
3.4. Screening for tagatose 4-epimerase activity ... 115
4 CONCLUSION ... 117
VII. CHARACTERIZATION OF UDP-HEXOSE 4-EPIMERASE ... 119
1 INTRODUCTION ... 121
2 MATERIAL & METHODS ... 122
2.1. Bacterial strains and growth conditions, plasmids and chemicals. 122
2.2. Sequence analysis and homology models ... 122
2.3. Construction of the expression vector (pIXhPtrc-mGalE) ... 124
2.4. Recombinant enzyme expression and His-tag purification ... 124

2.5. *Enzyme activity assays* *125*
2.5.1. Activity on UDP-Gal 125
2.5.2. Activity on UDP-Glc 126
2.5.3. Conversion of N-acetylated UDP-sugars 126
3 Results and Discussion 127
3.1. *Sequence analysis and homology models* *127*
3.2. *Construction of the expression vectors* *129*
3.3. *Determination of optimal conditions* *130*
3.3.1. Temperature optimum and range 130
3.3.2. pH optimum and range 131
3.3.3. Thermostability at 45 °C and 60 °C: half-life (t_{50}) 131
3.4. *Substrate preference and kinetic parameters* *133*
4 Conclusion 134
VIII. MUTATIONAL ANALYSIS OF UDP-HEXOSE 4-EPIMERASE 135
1 Introduction 137
2 Material & Methods 137
2.1. *Bacterial strains, plasmids, growth conditions and chemicals* *137*
2.2. *Rational mutagenesis: Site directed mutagenesis (SDM)* *138*
2.3. *Expression, purification and activity assays* *139*
2.4. *mGalE activity on free monosaccharides and α-Glc-1-P* *139*
3 Results and Discussion 141
3.1. *Site directed mutagenesis (SDM)* *141*
3.2. *Mutational analysis of mGalE* *141*
3.2.1. Catalytic triad Tx_nYx_3K (S116A, T117A and T117S) 141
3.2.2. Two consecutive glycine residues (G118-G119) 142
3.2.3. The gatekeeper residue (S279Y) 143
3.2.4. Loop exchange mutants 144
3.3. *Free monosaccharides detection assays* *145*
3.4. *GalE activity on free monosaccharides and α-Glc-1-P* *146*
4 Conclusion 147
IX. GENERAL DISCUSSION & FUTURE PERSPECTIVES 149
1 Introduction 151
2 General Discussion & Future perspectives 153
2.1. *Rare sugars and potential of epimerases: 'Epimering'* *153*
2.2. *Evaluation of the present project and used epimerases* *155*
2.2.1. L-Ru-5-P 4-epimerase: a bad choice as starting point or not? 155
2.2.2. UDP-hexose 4-epimerase: epimerase for free monosaccharides? 159
2.3. *Where to find or how to create new epimerases?* *160*
REFERENCES 167
SUMMARY & SAMENVATTING 181
CURRICULUM VITAE 187
APPENDICES 193

Abbreviations

For terminology of nucleotides and amino acids see appendices I-IV.

Å	Ångström = 10^{-10} m = 0.1 nm
AraD	L-ribulose-5-phosphate 4-epimerase (enzyme)
araD	L-ribulose-5-phosphate 4-epimerase gene
ATP	adenosine-5’-triphosphate
BCA	bicinchoninic acid
BCCM	Belgian Co-ordinated Collections of Micro-organisms
BSA	bovine serum albumin
CASTing	combinatorial active site saturation test
CFU	colony forming units
CV	coefficient of variation
DH	dehydrogenase
DHAP	dihydroxyacetone phosphate
(g)DNA	(genomic) deoxyribonucleic acid
D-TE	D-tagatose 3-epimerase
D-Xu-5-P	D-xylulose-5-phosphate
EC	enzyme Commission number
EDTA	ethylenediaminetetraacetic acid
epPCR	error-prone PCR
FDA	US Food and Drug Administration
Fru	fructose
Gal	galactose
GalE	UDP-hexose 4-epimerase (enzyme)
galE	UDP-hexose 4-epimerase gene
Glc	glucose
GRAS	generally recognized as safe
HFCS	high-fructose corn syrup
IPTG	isopropyl β-D-thiogalactopyranoside
ISM	iterative saturation mutagenesis
ISRS	International Society of Rare Sugars
kb	1000 base pairs

k_{cat}	catalytic rate constant
kDa	kilo Dalton
K_m	Michaelis-Menten constant
L-AI	L-arabinose isomerase
L-Ara	L-arabinose
LB	Luria-Bertani broth
LPS	lipopolysaccharides
L-Fuc-1-P	L-fuculose-1-phosphate
L-Ru-5-P	L-ribulose-5-phosphate
Man	mannose
MTP	microtiter plate
NAD	nicotinamide adenine dinucleotide
Ni-NTA	Nickel-Nitrilotriacetic acid
OD	optical density
PCR	polymerase chain reaction
PDB	Protein Data Bank
PMSF	phenylmethylsulfonyl fluoride
r.m.s.d.	root-mean-square deviation
SDR	short-chain dehydrogenase/reductase and related enzymes
SDM	site directed mutagenesis
SSM	site saturation mutagenesis
Tag	tagatose
t_{50}	half-life time or the time required for a quantity to fall to half its value
TMM	tagatose minimal medium
Tris	tris(hydroxymethyl)aminomethane
TSM	tagatose synthetic medium
U	enzyme unit
UDP	uridine diphosphate
UDP-Gal	uridine diphosphate galactose
UDP-Glc	uridine diphosphate glucose
UDP-GalNAc	uridine diphosphate N-acetylgalactosamine
UDP-GlcNAc	uridine diphosphate N-acetylglucosamine
UMP	uridine monophosphate
α-Glc-1-P	α-glucose 1-phosphate

I. INTRODUCTION

Introduction

In light of the growing obesity pandemic, carbohydrates – also known as sugars – have obtained somewhat of a negative reputation. However, carbohydrates are much more than just a source of energy, and can also mediate a variety of recognition processes that are central to human health. As such, saccharides can be applied in the food and pharmaceutical industries to stimulate our immune system (e.g. prebiotics), to control diabetes (e.g. low-calorie sweeteners), or as building blocks for anti-cancer and anti-viral drugs (e.g. L-nucleosides) (see chapter II section 1 and Table II.1) [20]. It can thus be stated that the synthesis and availability of rare sugars holds high commercial value due to this wide range of applications. Unfortunately, only a small number of all possible monosaccharides is found in nature in sufficient amounts to allow their commercial exploitation. Consequently, so-called rare sugars have to be produced by (bio)chemical processes starting from cheap and widely available substrates.

In the Izumoring [91], which is a schematic representation of the current biochemical production routes for rare sugars, only one epimerase is mentioned: the D-tagatose 3-epimerase (D-TE). However, an epimerization reaction can in theory take place at each of the different chiral centers of the sugars and this means C2-, C4- and C5-epimerases and, since D-TE is only active on ketoses, C3-epimerases active on aldoses could still extend the production tools (biocatalysts) for rare sugars. Recently, some more epimerases have been found to be able to convert free monosaccharides or to show potential as new biocatalysts (see chapter II section 1 and Figure II.2) [20]. Nevertheless, the speed – or better the slowness (or almost inertia) – at which these enzymes can convert monosaccharides and the concentrations of substrate needed, make that these biocatalysts are not yet economically interesting [197].

Due to the immense progress of sciences like biochemistry, biotechnology, engineering and informatics, scientists nowadays have a variety of tools which they can use to change and modify naturally existing enzymes towards biocatalysts harboring the desired characteristics. Some examples of this biocatalyst optimization are improved stability of sucrose phosphorylase via enzyme engineering [44], enhanced or changed activity on natural and/or non-natural substrates [167, 197], but also to obtain new substrate specificities that are not found in natural enzymes [59].

The combination of high economical value of rare sugars, the potential of epimerases in rare sugar synthesis and the availability of techniques for enzyme engineering triggered the idea to apply enzyme engineering of epimerase for rare sugar production. The rare sugar D-tagatose[1] was taken as prime rare sugar target, and since its C4-epimer fructose is cheap and widely available due to efficient production routes starting from starch, it was chosen to develop a C4-epimerase that would be able to convert fructose into tagatose. Enzymes that naturally catalyze a C4-epimerization, however on different but similar substrates, were taken as starting point (Figure I.1 and Figure I.2). The two available epimerases are L-ribulose-5-phosphate 4-epimerase and UDP-galactose 4-epimerase. The L-ribulose-5-phosphate 4-epimerase would require the adaptation of the substrate binding domain around the phosphate group of the substrate, to enable acceptance of a non-phosphorylated hexose rather than a phosphorylated pentose (Figure I.1). On the other hand, the UDP-hexose 4-epimerase requires an adaptation to accept much smaller substrates, namely free monosaccharides instead of nucleotide activate sugars (Figure I.2). Next to fructose/tagatose interconversion, C4-epimerization of glucose into galactose is a second target since glucose is cheap and abundant as well and this new source of galactose can be applied in tagatose production via current (bio)chemical production routes.

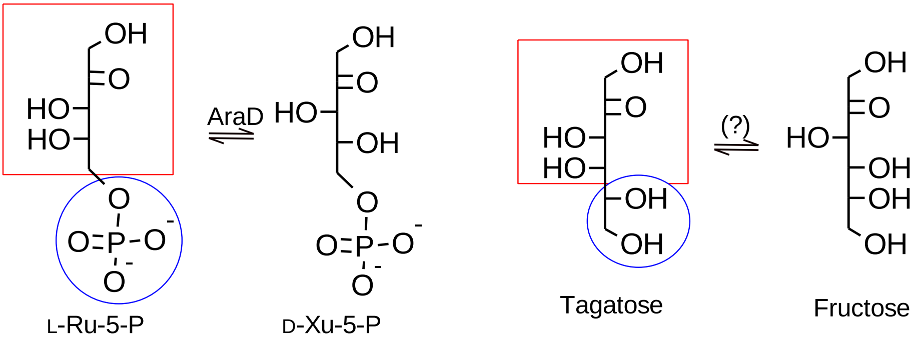


Figure I.1 Differences (in blue circle) and similarities (in red square) between the natural and desired substrates for the L-ribulose-5-phosphate 4-epimerase (AraD)

---

[1] Monosaccharides can occur in 2 different forms or enantiomers, namely the D- and L-form, which are each other's mirror images. Since D-enantiomers are naturally the most occurring forms of monosaccharides, by convention the 'D-' prefix can be omitted, but for L-sugars the prefix cannot be deleted. Example: D-tagatose = tagatose ≠ L-tagatose

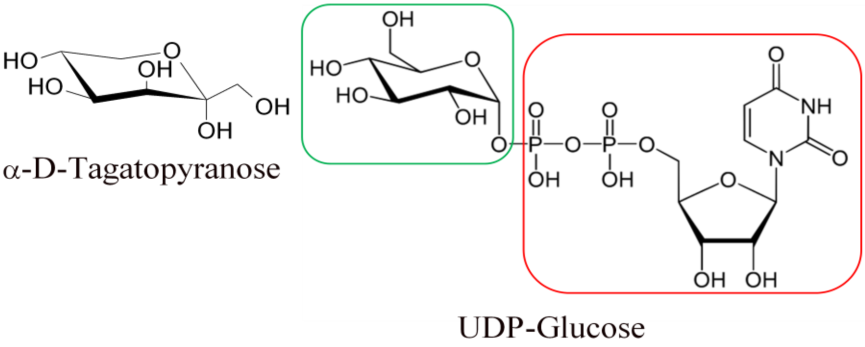


Figure I.2 The difference between the natural and desired substrates for the UDP-hexose 4-epimerase is even bigger than that needed for the L-ribulose-5-phosphate 4-epimerase. Loss of the UDP-moiety (red) is needed for activity on glucose (green), for tagatose additional differences are present between the sugar group

In summary, the goal of this thesis is to gain more knowledge in the determinants for substrate specificity of C4-epimerases and to use this information to create C4-epimerases that can be used for rare sugar production, with the focus on D-tagatose. At first, a thorough literature review on rare sugars, with a focus on the target tagatose, and their biochemical production routes as well as on epimerases and enzyme engineering is presented to provide the reader with sufficient background information (**chapter II**). Subsequently, the research results from this project are discussed, starting with the cloning and expression of the L-ribulose-5-phosphate 4-epimerase (**chapter III**). Then, the development of a selection system to identify improved enzyme variants and its application on mutant libraries are described (**chapter IV** and **chapter V**, respectively). This is followed by the development of a screening assay for the detection of mutants harboring tagatose 4-epimerase activity and the use thereof (**chapter VI**). The next two chapters deal with the cloning and characterization of the UDP-hexose 4-epimerase (**chapter VII**) and the mutational analysis of this epimerase (**chapter VIII**). Finally, an overall discussion as well as future perspectives are provided (**chapter IX**).

II. LITERATURE REVIEW

A part of this Literature review has been published as:

Beerens, K., Desmet, T., Soetaert, W. (2012). Enzymes for the biocatalytic production of rare sugars. *Journal of Industrial Microbiology and Biotechnology*, 39 (6), 823-834, DOI: 10.1007/s10295-012-1089-x

1 Biocatalytic production routes for rare sugars

1.1. Rare sugars and their applications

The International Society of Rare Sugars (ISRS) has classified monosaccharides and derivatives according to their abundance in Nature [91]. Of all possible hexoses and pentoses, only seven (Glc, Gal, Man, Fru, Xyl, Rib and L-Ara) were considered to be present in significant amounts, whereas twenty hexoses and nine pentoses were described as rare sugars. Another large group of rare sugars consists of deoxygenated monosaccharides, which often play a crucial role as recognition elements in bioactive molecules [87, 160, 264]. Furthermore, secondary modifications like amination or methylation can also occur. Rare sugars cannot be extracted from natural sources and thus have to be produced by (bio)chemical reactions. Nevertheless, several of these are now commercially available as bulk products, such as D-tagatose and D-sorbose. Others, in contrast, are specialty compounds that are used in high-value applications, which is the case for most L-sugars. It can be expected that more efficient production routes will increase the availability of rare sugars for research purposes, resulting in the discovery of new applications and/or as yet unidentified characteristics [91].

Despite their low natural abundance, rare sugars hold enormous potential for practical applications (Table II.1). In the pharmaceutical industry, for example, L-ribose can be used as a building block for drugs against cancers and viral infections. Its most important application is in antiviral therapy, where it is incorporated in L-nucleosides analogues [94]. The advantages of the L-enantiomer are increased antiviral activity, better metabolic stability and more favorable toxicological profiles. Since the discovery of Lamivudine (2',3'-dideoxy-3'-thiacytidine, mostly referred to as 3TC), more and more L-nucleoside analogues are undergoing clinical trials and/or preclinical studies. Furthermore, several other L-sugars can be used to produce L-nucleosides, such as L-gulose, L-xylose and L-galactose [94, 168, 276]. L-sugars can also be used as active compounds on their own, for instance as glycosidase inhibitors [145] or as insecticides [2].

Table II.1 Overview of rare, unmodified monosaccharides and their (potential) applications (update from [2])

Sugar	Application(s)	Reference(s)
D-allose	Treatment of cancer, in particular chronic myeloid leukemia	[10, 102, 175, 182, 280]
	Suppression of thrombus formation and reperfusion injury	[13, 103, 184]
	Cryoprotectant for mammalian cells and organs	[245]
	Immunosuppressant	[104]
D-altrose	Synthesis of cyclic carbamates of derived glycosylamines (polymer chemistry)	[138, 143]
D-arabinose	Synthesis of antitumor compounds, such as dehydroamino acid derivatives	[86, 178, 290]
	Production of D-erythroascorbic acid and oxalic acid	[163]
D-gulose	Drug-formulation agent and food additive	[33]
D-idose	Synthesis of cyclic carbamates of derived glycosylamines (polymer chemistry)	[138, 143]
D-lyxose	Synthesis of antitumor and immunostimulatory agents	[180, 247]
D-psicose	Non-calorie sweetener, treatment of diabetes	[16, 97]
	Potential anthelmintic	[229]
	Precursor of xylosylpsicoses (used as prebiotics, cosmetics and therapeutics)	[210]
D-ribulose	Starting material for branched pentoses (useful in pharmaceutical chemistry)	[105]
D-sorbose	Building block for industrial and bio-active products	[107, 108]
	Insect control agent	[108]
D-tagatose	Low-calorie sweetener, treatment of diabetes	[80, 154, 164, 191]
	Improvement of human health (e.g. antiplaque, prebiotic)	[154, 164, 191]
	Additive in detergents, cosmetics, and pharmaceutical formulations	[154, 191]
D-talose	Anti-tumor and anti-microbial activities, including. marker of O-antigens	[190, 278]
D-xylulose	Starting material for branched pentoses (useful in pharmaceutical chemistry)	[105]
L-allose	Therapeutic agent for diseases involving vasculogenesis	[263]
L-altrose	Component of biologically important oligo- and polysaccharides	[106]
L-fructose	Potential inhibitor of various glucosidases	[145]
	Mixture of L- and D-fructose kills ants and house flies	[2]
L-galactose	Potential in synthesis of L-nucleoside-based antiviral medications	[276]
	Component of saponins, with applications in food, cosmetics and pharmaceuticals.	[195, 276]
L-glucose	Starting material for glycoconjugate vaccines against diseases caused by *Shigella sonnei*	[145]
	Cytostatic and cytotoxic properties with regards to neoplastic cells (cancer therapy)	[2]

Table II.1 Overview of rare, unmodified monosaccharides and their (potential) applications (update from [2]) (continuation)

Sugar	Application(s)	Reference(s)
L-gulose	Building block of bleomycin A_2, a glycopeptide antibiotic (potential anticancer agent)	[76, 276]
	Synthesis of nucleosides that exhibit very potent activity against HBV and HIV	[76]
	Starting material for the production of L-nucleoside-based antiviral medications	[276]
L-idose	Derivatives are required in the synthesis of sensitive substrates for α-L-iduronidase	[76]
	Derivatives are used as glycosyl donors in the synthesis of heparin oligosaccharides	[252]
L-lyxose	Component of the antibiotic avilamycin A	[100]
	Potential L-fucosidase inhibitor	[51]
L-mannose	Component of steroid glycosides	[106]
L-psicose	Starting material for the production of L-fructose	[113]
L-ribose	Building block for antiviral and anticancer L-nucleosides	[94, 171, 291]
	Building block for glycoconjugates, oligonucleotides and L-aptamers	[194]
	Starting material for the production of L-allose and L-altrose	[14]
	Potential against HBV and Epstein-Barr virus	[262]
L-ribulose	Starting material for L-ribose production	[66, 68, 236]
L-sorbose	Starting material for the production of L-tagatose	[113]
	Precursor for the synthesis of the potent glycosidase inhibitor 1-deoxygalactonojirimycin	[2]
	Starting material for the production of L-ascorbic acid, also known as vitamin C	[113]
	Starting material for the synthesis of L-talitol	[228]
L-tagatose	Potential as a functional sweetener	[210]
	Potential in chemotherapy	[210]
	Precursor of complex materials, such as 1,2,3,4-diisopropylidene tagatofuranose	[210]
	Starting materials for the synthesis of L-deoxygalactonojirimycin	[108]
L-talose	Precursor of L-talose nucleosides, inhibitors of the *in vitro* growth of leukemia L1210 cells	[151]
L-xylose	Starting material for the synthesis of the nucleosides against HBV	[168]
	Synthesis of L-ribofuranose derivatives	[47]
L-xylulose	Potential inhibitor of various glucosidases	[145, 155]
	Synthesis of L-xylose and L-lyxose	[90]
	Indicator of hepatitis or liver cirrhosis	[248]

Other rare sugars, such as D-tagatose, can serve as low-calorie sweeteners, replacing classical table sugar in the food industry. A major advantage is that tagatose has a low glycemic index, making it suitable for diabetic patients [164]. In that respect, it is interesting to note that tagatose has entered phase III clinical trials to investigate whether it can be used as diabetic medication [80]. Similar effects have been attributed to D-psicose, which shows potential as non-calorie sweetener as well as diabetic and obesity control agent [16]. In contrast, D-allose, an isomer of D-psicose, displays rather different properties. Besides its inhibitory effect on both carcinogenesis and cancer proliferation, it is also useful in surgery and transplantation as an anti-inflammatory agent, immunosuppressant and cryoprotectant [159].

1.2. Enzymes for rare sugar production

Basically, three different types of enzymes can be used for the interconversion of monosaccharides (Figure II.1). Two of these are classified within the class of isomerases, *i.e.* keto-aldol isomerases (EC 5.3.1) and carbohydrate epimerases (EC 5.1.3). The former (often referred to as aldose isomerases or simply as isomerases) catalyze an intramolecular redox reaction, exchanging the carbonyl functionality between the C1- and C2-positions [116]. The latter, in contrast, catalyze the re-orientation of a hydroxyl group, converting the substrate into one of its epimers [226]. The third group of enzymes consists of oxidoreductases (EC 1.1) that convert carbohydrates into their corresponding polyols, and vice versa [267]. Oxidoreductases acting on ketoses are typically designated as polyol dehydrogenases, whereas those that act on aldoses are known as aldose reductases [91].

All three of these enzyme classes have already been applied for the production of rare sugars (see overview in Figure II.2 and Figure II.3), but each have specific advantages and disadvantages (Table II.2). Isomerases, for example, are promiscuous biocatalysts that are active on a range of simple substrates, *i.e.* unsubstituted monosaccharides. However, promiscuity is not always an advantage because this can result in the formation of side products and complicate the downstream processing. This is nicely illustrated with the glucose-6-phosphate isomerase from *Pyrococcus furiosus*, which converts L-tagatose not only to L-talose but also to L-galactose (Figure II.4) [286]. Although both products are valuable, it would be more efficient to produce them separately with two different isomerases, each specific for one of the aldoses.

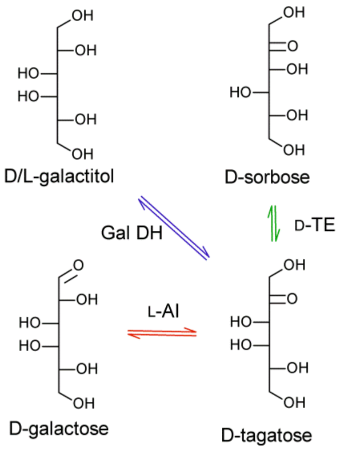

D/L-galactitol
D-sorbose
Gal DH
D-TE
L-AI
D-galactose
D-tagatose

Figure II.1 The three different enzyme classes that can be used for rare sugar production. Oxidoreductases are exemplified by galactitol dehydrogenase (Gal DH), aldose isomerases by L-arabinose isomerase (L-AI) and epimerases by D-tagatose 3-epimerase (D-TE)

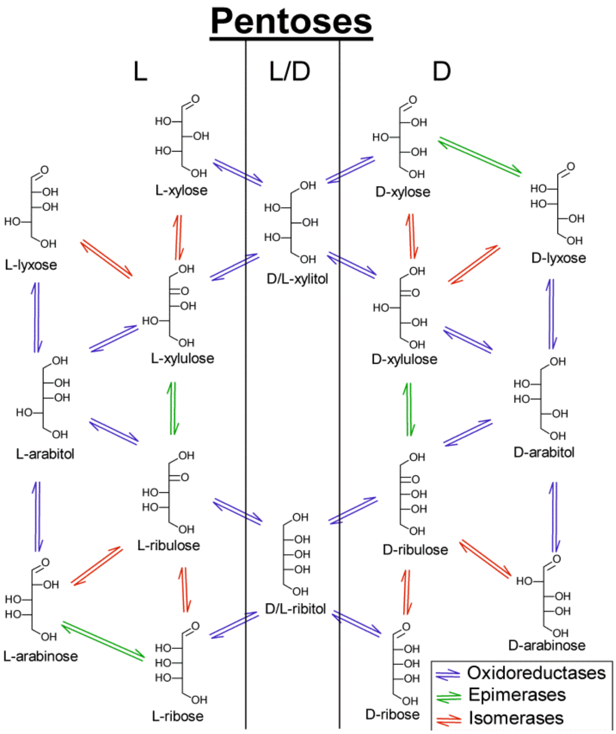

Pentoses
L
L/D
D
L-xylose
D/L-xylitol
D-xylose
L-lyxose
D-lyxose
L-xylulose
D-xylulose
L-arabitol
D-arabitol
L-ribulose
D/L-ribitol
D-ribulose
L-arabinose
D-arabinose
L-ribose
D-ribose
Oxidoreductases
Epimerases
Isomerases

Figure II.2 Overview of the enzymatic interconversions of unmodified monosaccharides that have been reported to date: pentoses. Update of the Izumoring process described in [91]

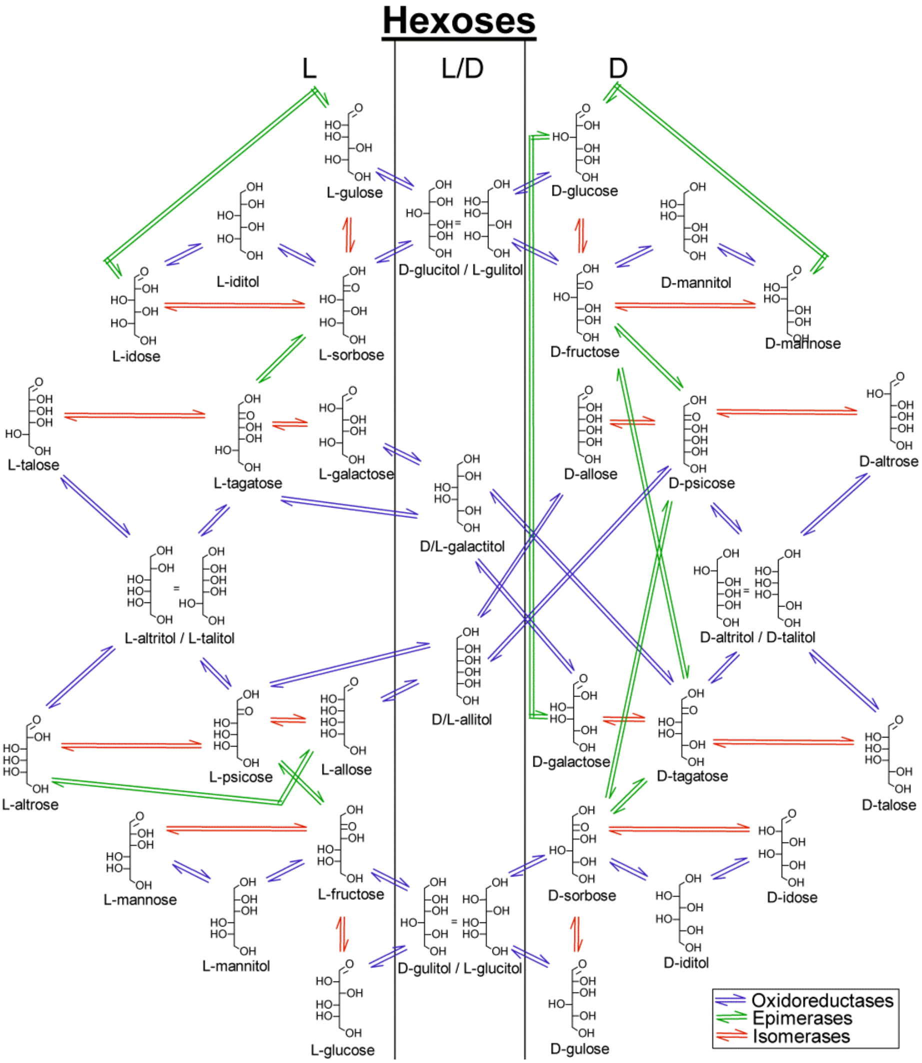

Hexoses
L
L/D
D
L-gulose
L-iditol
L-idose
L-sorbose
D-glucitol / L-gulitol
D-glucose
D-mannitol
D-fructose
D-mannose
L-talose
L-tagatose
L-galactose
D/L-galactitol
D-allose
D-psicose
D-altrose
L-altritol / L-talitol
D-altritol / D-talitol
D/L-allitol
L-psicose
L-allose
L-altrose
D-galactose
D-tagatose
D-talose
L-mannose
L-fructose
L-mannitol
D-sorbose
D-idose
D-iditol
D-gulitol / L-glucitol
L-glucose
D-gulose
Oxidoreductases
Epimerases
Isomerases

Figure II.3 Overview of the enzymatic interconversions of unmodified monosaccharides that have been reported to date: hexoses. Update of the Izumoring process described in [91]

Table II.2 Comparison of three enzyme classes in rare sugar production

Enzyme	Advantages	Disadvantages
Isomerase	Substrates often unsubstituted Broad substrate specificity	Product mixtures are sometimes formed
Epimerase	Shortcut in synthetic route Potential bridge between D- and L-sugars	Substrates often substituted Substrate specificity is rather strict
Oxidoreductase	Bridge between D- and L-sugars Substrates often unsubstituted	Need for cofactor regeneration

Epimerases are potentially the most useful biocatalysts for the production of rare sugars as they can give access to a wide range of structures, in contrast to the two other enzyme classes that are limited to modifications of the C1- and C2-positions. Unfortunately, most epimerases are only active on sugars that are substituted with a phosphate or nucleotide group, which drastically increases production costs. The recently discovered D-tagatose 3-epimerase is a noticeable exception that has allowed the production of D-psicose from the cheap substrate D-fructose [250]. The identification of new epimerases can create interesting shortcuts in current synthetic routes. A 2-epimerase, for example, can replace the double keto-aldol isomerization previously required for the conversion of D-xylose into D-lyxose (Figure II.2). Furthermore, 5-epimerases could form a new bridge between D- and L-hexoses, and the same is true for 4-epimerases acting on pentoses (Figure II.5). For instance, glucose could then serve as cheap substrate for L-idose production in a single reaction instead of three steps, which is currently the shortest route (Figure II.3).

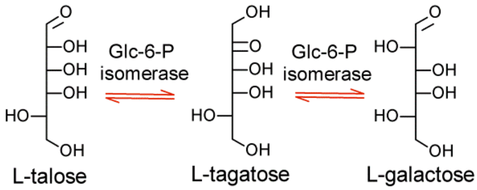


Figure II.4 Isomerization of L-tagatose with glucose-6-phosphate (Glc-6-P) isomerase. Because of the enzyme's low specificity, both L-talose and L-galactose are formed

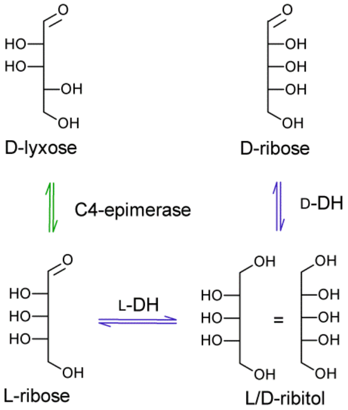


Figure II.5 Interconversion of D- and L-sugars. D-Ribose can only be converted to L-ribose through ribitol as intermediate, using oxidoreductases as biocatalysts. However, a putative C4-epimerase would be able to convert D-lyxose directly into L-ribose. DH Dehydrogenase

For the moment, D-sugars can only be converted into their L-isomers using the corresponding polyols as intermediate products. To that end, oxidoreductases are applied in a two-step process, *i.e.* sugar reduction followed by polyol oxidation (Figure II.5). However, polyols like xylitol and sorbitol are also valuable compounds in their own right and are used as sweeteners with a cooling sensation [88]. Despite their industrial relevance, oxidoreductases exhibit one great disadvantage, namely the need for NAD(P)H as expensive cofactor. As a result, reactions with oxidoreductases are often performed inside microbial cells so that the cellular metabolism can provide the reductive power. Alternatively, specific cofactor regeneration systems can be employed when isolated enzymes are to be used (Figure II.6) [72, 108].

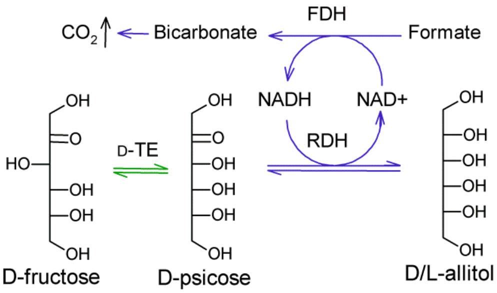


Figure II.6 The production of allitol from fructose. Allitol is produced from fructose by a coupling reaction using D-tagatose 3-epimerase (D-TE) and ribitol dehydrogenase (RDH). Cofactor regeneration is achieved with the help of the irreversible formate dehydrogenase (FDH) reaction

1.2.1. Oxidoreductases

Since the development of the chemical hydrogenation method, xylitol has been used on a large scale as alternative sweetener [88]. However, research has also been done in order to develop microbial production methods for xylitol. To that end, oxidoreductases are often employed in metabolically engineered organisms, such as *Saccharomyces cerevisiae* and *Candida* strains [88, 89]. The main challenge with *S. cerevisiae* is to increase the uptake of xylose as substrate, as well as the regeneration of NADPH through the pentose phosphate pathway. Although *Candida* yeasts are better at taking up xylose and maintaining the intracellular redox balance, their application in the food industry is hampered due to the opportunistic pathogenic nature of some species. In parallel, the use of xylitol 4-dehydrogenase (XDH) as isolated biocatalyst for the production of xylitol has also been optimized. The immobilization of the enzyme from *Rhizobium etli*, for example, has resulted in a 10-fold increased thermostability, a broader operational pH range, and excellent reusability [292]. In turn, xylitol can serve as substrate for the production of other rare sugars like L-xylulose and L-xylose. In a recent study, the XDH from *Bacillus pallidus* has been overexpressed in *Escherichia coli* for L-xylulose production. Although the conversion rates were lower than when the *B. pallidus* strain was used, the advantage of *E. coli* is that the formation of side products is drastically reduced [205, 248]. At higher temperatures, however, L-xylose started to accumulate instead of L-xylulose, perhaps due to the activity of endogenous D-arabinose isomerase in *E. coli*.

The combination of regio- and stereoselectivity for the C2 position allows the mannitol 1-dehydrogenase (MDH) from *Apium graveolens* to catalyze several interesting reactions, including the conversion of ribitol to L-ribose, D-sorbitol to L-gulose, and galactitol to L-galactose [243]. A recombinant *E. coli* harboring this MDH represents a significantly improved method for the large-scale production of L-ribose compared to previously used methods [277]. A threefold higher productivity was obtained compared to the double isomerization of L-arabinose with xylose isomerase [121], and a much higher conversion rate was obtained compared to the chemical process, while the environmental hazards could be avoided [124].

Another dehydrogenase that shows high potential as a biocatalyst in rare sugar production is the recently discovered D-arabitol 2-dehydrogenase (D-ADH) from *Thermotoga maritime* [125]. It is the first hyperthermophilic D-ribulose forming D-ADH and thus exhibits several industrial advantages. First of all, the enzyme can convert the inexpensive substrate D-arabitol with very high selectivity. In addition, the risk of contamination in food production is dramatically lower as it can be used at increased temperatures. And finally, it can be efficiently purified from recombinant *E. coli* by heat precipitation, providing a reliable and cost-effective supply of biocatalyst.

Oxidoreductases, and their producing strains, can also be used for the interconversion of deoxy-sugars. L-rhamnose is the only cheaply available deoxy-hexose but can be transformed into 1- and 6-deoxygenated D/L-psicose, D/L-fructose and L-tagatose by *Enterobacter aerogenes* IK7 [92, 93]. On the other hand, *Enterobacter agglomerans* 221e can be applied in the synthesis of 1- and 6-deoxy-D-tagatose from both enantiomers of fucose [288]. In these processes, the deoxy-sugar is first chemically hydrogenated towards the corresponding polyol, which then serves as substrate for the enzymatic oxidation. Furthermore, microbial oxidation can also be performed on methylated polyols, for instance for production of both enantiomers of 4-C-methyl-ribulose by *Gluconobacter thailandicus* NBRC 3254 [122, 211].

1.2.2. Isomerases

Due to the broad substrate specificity of isomerases, these enzymes can be applied in the synthesis of various aldoses and ketoses starting from their cheaper counterpart. A well known example to illustrate their promiscuity is the fact that xylose isomerase is better known as glucose isomerase because it is mainly used for the production of high fructose corn syrup (HFCS) rather than for its wild-type conversion (D-xylose to D-xylulose). The importance of substrate resemblance for side activities can be further illustrated with the following examples. The D-lyxose isomerase from *Providencia stuartii* was found to be active not only on D-lyxose/D-xylulose but also on nine other ketose-aldose couples. The highest specific activity was measured on aldose substrates with the C2 and C3 hydroxyl groups in the left-hand configuration, just as in the wild-type substrate. In contrast, very little activity could be detected on the mirror image substrates L-lyxose, L-mannose [142]. Many more isomerases have been shown to be active on sugars similar to their wild-type substrates [3, 52, 146, 172, 199, 200, 206, 215, 282].

The promiscuity of isomerases has not only been evaluated with resembling monosaccharides but also with respect to the presence of a phosphate group. Indeed, several isomerases that are naturally active on phosphorylated sugars have been shown to convert free carbohydrates as well. In particular, ribose-5-phosphate isomerase has been shown to be active on D/L-ribose, D/L-lyxose, D/L-talose, D/L-mannose, D/L-allose, L-lyxose and L-tagatose [199, 283, 285]. Similarly, mannose-6-phosphate isomerase and galactose-6-phosphate isomerase have been found to convert free monosaccharide, such as L-ribose [284] and D-allose [200], respectively.

A second feature to take into account is whether the isomerase is specific towards one of the two aldoses. If the enzyme is not specific enough, mixtures will be obtained containing three sugars, namely the substrate, the desired product and an undesired side product (Figure II.4). Consequently, the choice of the proper enzyme will result in both higher yields and easier purification. The L-rhamnose isomerase from *Bacillus pallidus* has been shown to convert D-psicose/D-allose, L-fructose/L-mannose, D-ribose/D-ribulose and L-talose/L-tagatose without the formation of by-products [206]. The opposite is true for the glucose-6-phosphate isomerase from *Pyrococcus furiosus*, which converts L-talose to both L-tagatose and L-galactose, and D-ribulose to both D-ribose and D-arabinose [286]. Xylose isomerase is another enzyme that has been shown to also catalyze C2-epimerization of different substrates in addition to the expected isomerization, leading to side product formation [201, 269].

To further outline the importance of isomerases in rare sugar synthesis, an overview is given from some isomerases that are currently in use or which show great potential in rare sugar conversion.

1.2.2.1. L-arabinose isomerase for tagatose production

L-arabinose isomerase (L-AI) has been thoroughly studied for the conversion of D-galactose into the low-calorie sweetener D-tagatose [49, 50, 55, 207, 215]. Tagatose is almost as sweet as sucrose but is metabolized in a different fashion, resulting in a caloric content that is about three times lower [80, 154]. Recently, the enzyme from *Bacillus stearothermophilus* was successfully mutated towards an improved industrial biocatalyst. Through rational design, three mutants were created based on previously reported data and sequence alignments. Acidotolerance and stability was improved by the mutation Q268K and a broader temperature range was achieved by the mutation N175H, with the combined double mutant displaying both characteristics [213]. In another study, ten positions identified through random mutagenesis of *Geobacillus thermodenitrificans* L-AI were examined in detail. Five of the corresponding variants were shown to display a two- to threefold increase in specific activity. Double mutants were then created, with C450S/N475K generating a 20% higher tagatose conversion and a fourfold increase in specific activity compared to the wild-type enzyme [193]. In addition, production processes have also been improved. The addition of boric acid was found to result in increased yields

[158], whereas immobilization of the L-AI was performed to obtain a stable and economic method for the industrial production of tagatose [192].

1.2.2.2. Isomerases for L-ribose production

Several isomerases have been shown to display a minor side activity on L-ribose, which has been the focus of enzyme engineering efforts. The activity of the mannose-6-phosphate isomerase from *Thermus thermophilus* for L-ribose production has, for example, been improved by the mutation R142N, resulting in a 1.4- and 1.6-fold increase in specific activity and catalytic efficiency (k_{cat}/K_M), respectively. This mutant was found through alanine scanning of several active-site residues, followed by partial randomization of the most important position. The catalytic efficiency of the resulting R142N mutant was 3.8-fold higher than that of *Geobacillus thermodenitrificans* mannose-6-phosphate isomerase, which exhibited the highest catalytic efficiency reported to that date. The purified R142N mutant had a volumetric productivity of 107 g liter^{-1} h^{-1} in a L-ribose production process [284]. However, a specific L-ribose isomerase (L-RI) has also been isolated from an *Acinetobacter* strain, with a K_M of 44 mM for L-ribose and a specific activity of 24.2 µmol mg^{-1} min^{-1} for L-ribulose formation [236]. Strangely, the enzyme's specific activity was about ten-fold higher when it was recombinantly expressed in *E. coli* [176]. The L-RI could be a promising biocatalyst for the production of L-ribose, but has not yet been evaluated in such a process.

1.2.2.3. Isomerases for deoxygenated and other modified sugars

Two isomerases that prefer deoxy-sugars as their substrates are L-fucose isomerase [82, 123] and L-rhamnose isomerase [146, 147]. These enzymes also tolerate other modifications at centers higher than C3 and can use epimeric or functionalized sugars as substrates. Examples hereof are terminus-modified fuculose-analogues for L-fucose isomerase [82]. Other related enzymes, including xylose (glucose) isomerase, were also shown to accept deoxygenated and/or substituted sugars [81, 96, 209]. It can thus be expected that several isomerases will be applied for the production of modified sugars in the near future.

1.2.3. Carbohydrate epimerases

All biochemical reactions that can be used for the production of rare sugars and polyols from readily available raw materials like starch, wood and lactose have been summarized in the Izumoring [91]. At that time, only one epimerase was available for such conversions, namely the D-tagatose 3-epimerase. Since then, however, two other epimerases have been shown to be naturally active on unsubstituted sugars (*i.e.* cellobiose 2-epimerase and UDP-galactose 4-

epimerase), whereas the L-ribulose-5-phosphate 4-epimerase has also been found to show promise for the production of rare sugars.

1.2.3.1. Ketohexose 3-epimerase

D-tagatose 3-epimerase (D-TE) was found to catalyze the epimerization of various ketoses at the C3 position, making it a very useful enzyme for rare sugar production. D-TE has already been used for the synthesis of various carbohydrates, both in single and multiple enzyme reactions. At the lab scale, this enzyme has even been used for the production of all possible ketohexoses [113, 115]. Furthermore, D-TE is a highly promiscuous enzyme that can accept a large range of unnatural substrates, like C-4-methylated pentoses [211], C-5-methylated hexoses [122], 5-deoxy-ketohexoses [209], as well as several 1- and 6-deoxy-ketohexoses [93]. On the other hand, immobilized D-TE has been applied in the mass production of D-psicose from D-fructose [250]. Addition of borate to the reaction mixture results in removal of the product as a psicose-borate complex and will thus result in improved yields [133]. Two examples of multiple enzyme reactions where D-TE is applied are allitol and D-arabinose production starting from fructose and xylose, respectively. To that end, the epimerase is combined with dehydrogenases or isomerases (Figure II.6) [246, 249].

A very similar enzyme has been described in *Agrobacterium tumefaciens*, namely D-psicose 3-epimerase (D-PE). Unfortunately, due to its short half-life (63 min at 50 °C), it is inefficient for the industrial production of psicose. Therefore, enzyme variants were created by error-prone PCR and tested for improved thermostability [53]. Two single mutants (I33L and S213C) displayed an increased optimal temperature and kinetic stability. Combining both mutations in a single enzyme further improved these parameters, resulting in a 30-fold increase in half-life at 50°C. In a continuous production process with the immobilized double mutant, no decrease in activity could be observed after 30 days. This suggests that the I33L/S213C variant may be useful as an industrial producer of D-psicose.

1.2.3.2. UDP-Galactose 4-epimerase

The second enzyme that shows epimerase activity on free monosaccharides is the *E. coli* UDP-galactose 4-epimerase, which normally uses nucleotide-activated galactose and glucose (UDP-Gal/UDP-Glc) as substrates. However, the purified enzyme was found to also catalyze the 4-epimerization of free galactose, glucose, fructose, tagatose, psicose, and sorbose, albeit with very low specific activities (0.3-9.7 nmol mg^{-1} min^{-1}). Three residues were submitted to site-saturation mutagenesis, resulting in the identification of a N179S mutant with a twofold improved activity on fructose and tagatose. The enzyme was also tested for 4-epimerization activity on allose, gulose, altrose, idose, mannose, and talose but no activity was found on these aldohexoses [130].

1.2.3.3. Cellobiose 2-epimerase

A third enzyme that was found to catalyze (rare) carbohydrate epimerizations is the cellobiose 2-epimerase (CE) from *Caldicellulosiruptor saccharolyticus*. This thermophilic epimerase not only catalyzes the epimerization of cellobiose, but also shows low activity on aldoses harboring hydroxyl group oriented in the left-hand configuration at the C3 position. The enzyme exhibited the highest side activity for mannose (to glucose), although this was 20 times lower than the activity on cellobiose. Low activities were also detected on D-xylose, L-altrose, L-idose, and L-arabinose [198]. Very recently, a new mannan catabolic pathway has been described in *Bacteroides fragilis*, including a CE that functions as mannobiose 2-epimerase (MBE) *in vivo*. These findings are supported by a lower K_m and higher catalytic efficiency for mannobiose than for cellobiose [233]. Further proof of its physiological role can also be found in its preference for mannose over glucose as monosaccharide substrate, which was reported for the *C. saccharolyticus* CE [198]. Surprisingly, the latter enzyme also shows slight isomerase activity on various monosaccharides when long reaction times and high amounts of enzyme were used [198]. This isomerase side activity will lower epimerization yield and complicate the purification process.

1.2.3.4. L-ribulose-5-phosphate 4-epimerase and related aldolases

Another epimerase that is promising for rare sugar conversions is L-ribulose-5-phosphate 4-epimerase. This enzyme is structurally and mechanistically related to the dihydroxyacetone phosphate dependent L-fuculose-1-phosphate aldolase and L-rhamnulose-1-phosphate aldolase [120, 166]. Despite the fact that the wild-type substrates of all these enzymes are phosphorylated (deoxy-) sugars, the epimerase shows potential for the production of rare sugars that lack a phosphate group. Indeed, by applying error-prone PCR to L-rhamnulose-1-phosphate aldolase, variants could be created that display activity towards unsubstituted L-rhamnulose [244]. Since the two enzymes are very similar, it is not unreasonable that the same or equivalent mutations would have an identical effect on the substrate specificity of the epimerase, thus making it active on free ketoses. Furthermore, aldolases can be used in the direct synthesis of rare sugars by aldol condensation of smaller carbohydrates. The previously mentioned L-fuculose-1-phosphate aldolase has been used in the production of D-psicose, D-sorbose, L-tagatose, and L-fructose starting from DL-glycerol 3-phosphate and D- or L-glyceraldehyde [156].

1.3. Rare sugars: conclusions and outlook

During the past decades, major progress has been made in research on rare carbohydrates. The discovery of new enzymes and the engineering of existing biocatalysts have generated new opportunities for their application in various industrial sectors. The discovery of a D-tagatose 3-epimerase (D-TE), for example, has not only allowed the synthesis of all possible ketohexoses, but also of methylated and deoxygenated sugars. For other biocatalysts, mutagenesis efforts have resulted in improved activity and stability, as illustrated by the R142N variant of mannose-6-phosphate isomerase and by the I33L/S213C variant of D-psicose 3-epimerase, respectively.

Epimerases, in particular, are exciting biocatalysts as they can facilitate rare sugar production through the introduction of major shortcuts in the current production routes. One major challenge, however, is the need for epimerases that are active on free monosaccharides instead of nucleotide-activated or phosphorylated sugars. Obtaining such epimerases by redesigning the active site of known enzymes is not a trivial task. Indeed, the substituents are often essential for strong binding of the substrates and can even be crucial for the enzyme's activity through an induced fit mechanism. Alternatively, suitable epimerases could be identified by further screening in natural environments, which still is a powerful approach and regularly redefines our knowledge of microbial physiology.

A very important step in any enzyme engineering project is the choice of the most suitable template. Recently, L-arabinose isomerase (L-AI) and tagatose-6-phosphate isomerase (T6PI) have been compared as starting points for increased isomerization activity on galactose [131]. Although this would require the loss of a phosphate group in the case of T6PI, this enzyme was concluded to be the best template for directed evolution because mutations are less likely to diminish its activity. Indeed, L-AI makes use of metal cofactor whose binding can be easily disrupted by random mutations. Nevertheless, L-AI would probably be the right choice for rational design studies because its crystal structure is available. Future work will reveal which enzyme and which strategy will generate powerful new biocatalysts for the production of rare sugars.

2 The rare sugar target: D-tagatose

2.1. Introduction

Tagatose, a rare natural ketohexose, is an isomer of D-galactose and an epimer (stereoisomer) of D-fructose (Figure II.7). It is found in small quantities in various foods such as sterilized and powdered cow's milk, a variety of cheeses, yoghurts, and other dairy products. In addition, tagatose also naturally occurs in the gum of the West-African tree *Sterculia setigera* [80, 191]. D-Tagatose has attracted a great deal of attention since it was first described as a sugar substitute [155]. Since then, several studies have been performed, investigating the different aspects of tagatose in food and pharmaceutical formulations, like stability in food and drinks [75, 165], clinical trials for drug application [80] as well as consumer evaluation [9, 253].

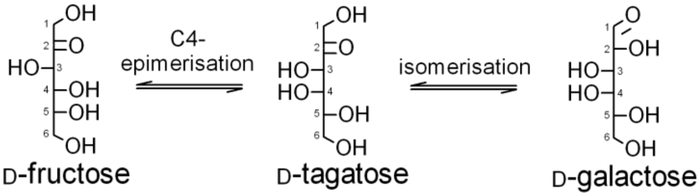


Figure II.7 Fisher projection of D-fructose, D-tagatose and D-galactose, including the (possible) link between the sugars. D-Fructose to D-tagatose conversion needs a C4-epimerization whereas D-galactose to D-tagatose conversion requires an isomerization

2.2. Properties

As all hexoses, the molecular formula of tagatose is $C_6H_{12}O_6$ and it has a molecular weight of 180.16 g/mol. It is an optically active compound containing 3 chiral carbon atoms. As it is a reducing sugar, it is involved in browning during heat treatment. The melting temperature of tagatose is 134 °C, which is lower than that of sucrose, resulting in a more readily decomposition at higher temperatures. It is stable in a pH 2-7 range and has a high solubility in water, 58% (*w*/*w*) at 21 °C. Its physical form is a white anhydrous odorless crystalline solid. It is less hygroscopic than fructose and has similar humectants properties as sorbitol [154].

As tagatose is poorly absorbed in the small intestine and the majority of ingested tagatose (75 %) will reach the large intestine, it is called a malabsorbing sugar [191]. In the large intestine, it is fermented by the intestinal microflora and is converted into short chain fatty acids, which are absorbed almost completely. Absorbance and metabolisation of these short chain fatty acids result in a low energy recovery. On the other hand, energy is lost due to increased biomass excretion of microflora [28]. Whereas reported caloric values range from -0.12 to 1.4 kcal/mol, the FDA approved caloric content is 1.5 kcal/g.

The taste of tagatose is very comparable to that of sucrose, with a sweetness of 92 % compared to sucrose at 10 % solutions. In contrast to polyols, no cooling effect is observed with tagatose, nor bitter or off-flavor aftertastes as found with artificial sweetener aspartame [183]. It is similar to polyols in having a low caloric value and is also tooth-friendly [191], yet without the laxative effect. The glycemic index of tagatose is very low [154], namely 2-3. An overview of the physical, chemical, and biological properties of tagatose is given in Table II.3.

Table II.3 D-tagatose and its physical, chemical, and biological properties (Adapted from [80, 154, 191])

Property	Value and/or comment
Common name	D-tagatose, Tagatose
IUPAC name	(3S,4S,5R)-1,3,4,5,6-Pentahydroxy-hexan-2-one
Chemical family	Carbohydrate; monosaccharide; ketohexose
Molecular formula	$C_6H_{12}O_6$
Molecular weight	180.16 g/mol
Structure	3 chiral carbons (C3, C4 & C5); C4-epimer of fructose; isomer of galactose
Distribution of cyclic forms	α-D-tagato-2,6-pyranose (79%), β-D-tagato-2,6-pyranose (14%), α-D-tagato-2,5-furanose (2%), and β-D-tagato-2,5-furanose (5%)
Physical form	White anhydrous crystalline solid
Odor	None
Melting point	134 °C
Decomposition temperature	120 °C
Optical rotation	$\alpha_D^{20} = -5°$ (c = 1 in water)
Solubility in water	High solubility: 58% wt/wt at 21°C
pH stability range	2-7
Relative sweetness	92% of sucrose (compared in 10% solutions)
Sweetness profile	Mimics sucrose, but with a faster onset than fructose
Cooling effect	None
Caloric value	1.5 kcal/g approved (reported range: -0.12 – 1.4 kcal/g)
Glycemic index	Very low: 2-3
Cariogenicity	None
Health promotion	Low calorie, prebiotic, low glycemic, no elevation of blood glucose, suitable for diabetics, healthy foods, dietary supplements, beneficial drugs or drug adjuvants, antioxidant, cytoprotective
Bulk	Similar to sucrose
Humectant/Hygroscopicity	Similar to sorbitol / Less than fructose
Metabolism	Malabsorbing sugar: 25% absorbed, unabsorbed fraction (75%) is fermented by the intestinal microflora towards short chain fatty acids which results in a low energy recovery. Energy loss due to increased biomass excretion of microflora
Maillard reaction	Yes, browning like sucrose
Applicable in	Chocolate candy, soft confectioneries, hard confectioneries, diet soft drinks, ready to eat cereals, frosting, ice cream, frozen yogurt, diet chewing gum
Regulatory status	GRAS status by the FAO/WHO since 2001 for food and beverages

2.3. Production methods

In both the chemical and biochemical manufacturing process, tagatose is made by isomerization of galactose. This galactose is obtained by hydrolysis of lactose which is present in milk whey, a byproduct of the cheese industry. The released galactose and glucose in the hydrolysate are then separated by chromatography to yield pure galactose as substrate for the isomerization reaction (Figure II.8).

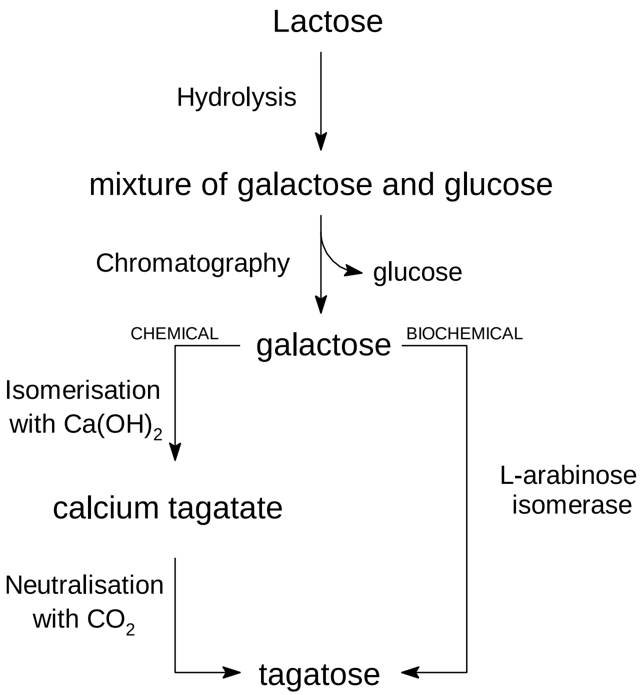

Lactose
Hydrolysis
mixture of galactose and glucose
Chromatography
glucose
CHEMICAL
galactose
BIOCHEMICAL
Isomerisation with Ca(OH)2
calcium tagatate
Neutralisation with CO2
L-arabinose isomerase
tagatose

Figure II.8 Tagatose manufacturing starting from lactose. Left: chemical isomerization under alkaline conditions ($Ca(OH)_2$) and neutralization using CO_2. Right: biochemical isomerization using L-arabinose isomerase

2.3.1. Chemical

The current production process is based on the chemical isomerization of galactose which was developed by Spherix Incorporated [19]. Chemical isomerization of galactose to tagatose is carried out under alkaline conditions using a hydroxide. Calcium hydroxide is the preferred hydroxide because this complexing agent will shift the isomerization equilibrium between galactose and tagatose towards tagatose. This is a result of insoluble complex formation with tagatose at elevated pH. The calcium tagatate suspension is then treated with carbon dioxide, the preferred acid, to neutralize the mixture and hereby liberate tagatose as well as precipitate calcium as calcium carbonate (Figure II.8). Finally, tagatose is further purified, crystallized and dried [154, 164].

2.3.2. Biochemical

Several biocatalysts have been studied for biological manufacturing of tagatose (Figure II.9). Conversion of galactitol into tagatose has been reported for *Arthrobacter globiformis* [117], *Gluconobacter oxydans* [170, 220], *Mycobacterium smegmatis* [118], *Enterobacter agglomerans*, and *Klebsiella pneumoniae* [191]. In these micro-organisms, a sorbitol dehydrogenase is the responsible enzyme for the biotransformation [220]. A second biocatalytic route that has been proposed is the conversion of sorbose into tagatose with the help of D-psicose 3-epimerase from *Agrobacterium tumefaciens* [129] or D-tagatose 3-epimerase from *Pseudomonas cichorii* [109, 114, 287]. However, both sorbose and galactitol are expensive substrates and thus have limited potential in the commercial production of tagatose.

More economical production routes start from cheap substrates such as galactose or fructose. The current production route that starts from fructose is a two step method in which fructose is first converted to psicose which is then applied as substrate in the biotransformation towards tagatose with the help of *Mucoraceae* fungi. Various *Mucoraceae* strains have been reported to be able to carry out this reaction. Neither the availability nor the price of psicose are limiting factors here, as psicose mass production became industrially feasible after the discovery of ketohexose 3-epimerases [129, 250]. The two-step tagatose production from fructose over psicose is a good alternative method; however, it still requires further intensive investigation [191].

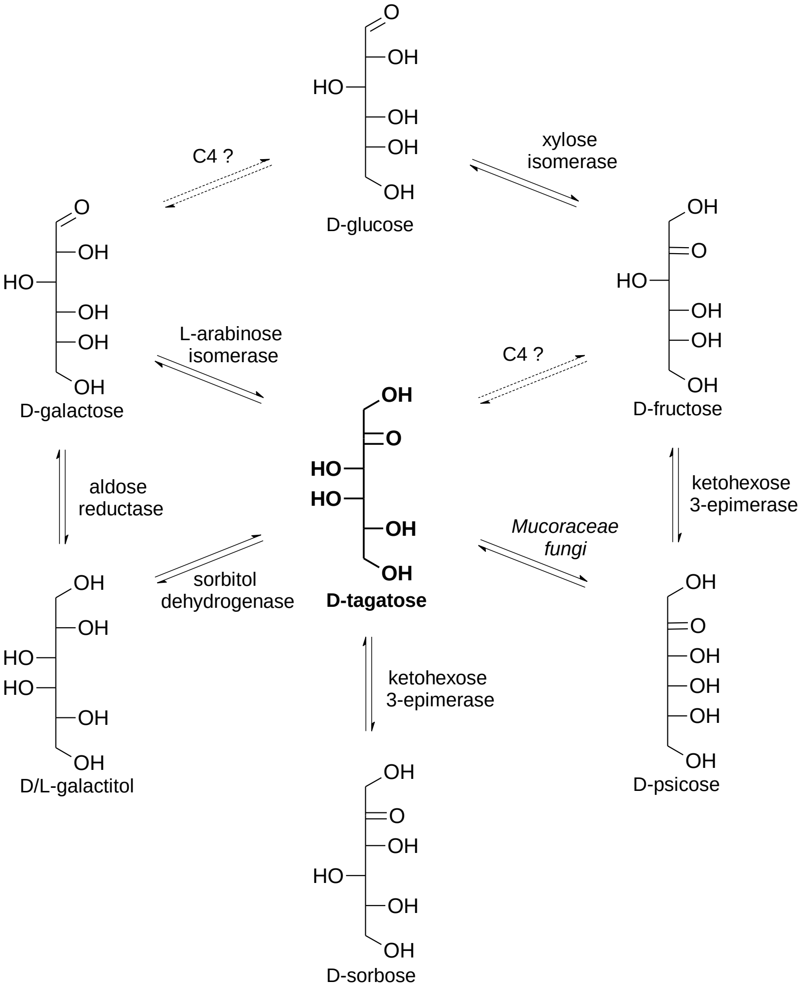

D-glucose
C4 ?
xylose isomerase
D-galactose
L-arabinose isomerase
C4 ?
D-fructose
D-tagatose
aldose reductase
ketohexose 3-epimerase
Mucoraceae fungi
sorbitol dehydrogenase
D/L-galactitol
ketohexose 3-epimerase
D-psicose
D-sorbose

Figure II.9 Different biological manufacturing routes for tagatose. C4? indicates possibilities for C4-epimerization reactions in manufacturing of tagatose

A biocatalyst that is currently receiving much interest, and probably the most studied one for biochemical isomerization of galactose, is L-arabinose isomerase. The isomerization can be both achieved biocatalytically, with purified L-arabinose isomerase, or in a biotransformation setting in which the isomerase is still inside living/resting cells. As discussed earlier (see 1.2.2.1), site directed and site saturation as well as random mutagenesis were applied to improve L-arabinose isomerase as an industrial biocatalyst. Characteristics that were improved are acidotolerance and stability [213], a broader temperature range [213], tagatose conversion and specific activity [193]. On the other hand, tagatose can also be produced using L-arabinose isomerase inside living cells. The advantage of working inside cells is that an equilibrium shift is achieved towards tagatose. This is the result of the differential selectivity of the cell membrane, leading to higher uptake and lower release rates for galactose than for tagatose and *vice versa* [132]. Alginate immobilized *Lactobacillus* strains are an attractive substitute for recombinant *Escherichia coli*, as they are approved as GRAS (generally recognized as safe) organisms [279]. Another example of *in vivo* production is that of tagatose production in milk fermentation by *Lactobacillus* and *Streptococcus* harboring recombinant L-arabinose isomerase [214].

2.4. Applications

Tagatose has numerous health benefits, which form the basis of its applications in dietary food and beverages, as well as its drugs applications. The main health benefits are summarized here. Due to its low caloric content, its consumption results in promotion of weight loss at medically desirable rates [41, 154]. Consequently it can also be used for the treatment of obesity [41, 177]. As it has no glycemic effect, it is safe for diabetic patients [77], for whom it can also be used to reduce symptoms associated with diabetes type 2, as well as symptoms of hyperglycemia, anemia and hemophilia [153]. Thus, the low-caloric sweetener is mainly applied in dietary foods and beverages such as low-carbohydrate diets, cereals, yoghurt, milk-based drinks, soft drinks, health bars, candy, chocolate and other confectionery [191].

Besides being of low-caloric value, tagatose has other benefits in health food products, like prebiotic properties, as well as anti-biofilm, anti-plaque, and non-cariogenic [29, 56]. These latter features make tagatose very useful in tooth paste and mouth wash. Furthermore, it can be used for improvement of pregnancy and fetal development [153], flavor enhancements [191] and for organ transplants where it acts as a powerful antioxidant and cytoprotective [202]. Other applications for tagatose are in the synthesis of other optically active compounds and as an additive in detergents, cosmetics, and pharmaceutical formulations [191].

More recently, preliminary studies in humans pointed out that tagatose has a low postprandial blood glucose and insulin response. Currently, tagatose is being studied in clinical phase III trials for use as medication for the treatment of diabetes type 2 as well as prediabetes and obesity [80]. Different mechanisms of action have been proposed and combinations of these effects are also possible. Tagatose might compete with or partially inhibit glucose transporters in the small intestine leading to delayed absorption. However, this has not been demonstrated in animal models [237]. Another proposed mechanism is the inhibition of sucrase and maltase which would decrease the digestion of sucrose and starches in the small intestine, respectively. *In vitro* and animal studies support these mechanisms. Another putative mechanism states that tagatose may act through its liver metabolism. Tagatose-1-phosphate formation might induce glucokinase and inhibit glycogen phosphorylase, both resulting in increased glycogen accumulation and decrease of hepatic glucose output. In a phase II clinical trial study, a 1% decrease in glycated hemoglobin (HbA_{1c}) was found in type 2 diabetic patients. This overall significant decrease in HbA_{1c} indicates a highly improved blood sugar control by the body. Furthermore, HDL cholesterol had increased whereas no changes were observed in blood pressure, non-HDL cholesterol or other parameters. The ongoing clinical phase III trials will provide better insights in tagatose, its mechanism and potential in diabetes and obesity treatment [80].

As indicated for other rare sugars (see 1.1 and Table II.1), it can be stated that tagatose has plenty of applications, mainly as low-caloric sweetener in food and flavor industry. Furthermore, it still holds great potential in a variety of other industries, such as medication in diabetes type 2, prediabetes and obesity. As such, tagatose has a big economical value.

3 (Sugar) Epimerases

3.1. General

Carbohydrates and their derivatives are essential to different life forms, in which they exhibit a wide variety of functions. Among these vital tasks in biology are source of energy, structural elements, molecular recognition markers, and they are used as precursors for the biosynthesis of other building blocks/molecules such as aromatic amino acids. Furthermore, as already stated above in the section on rare sugar production, they have many applications in industrial processes (Table II.1). In nature, several kinds of enzymes are used in the metabolism and conversion of carbohydrates, such as dehydrogenation, oxidation, reduction, acetylation, isomerization, and epimerization.

Epimerization is defined as the inversion of the configuration of asymmetrically substituted carbon in linear or cyclic molecules like carbohydrates. Oversimplified, it can be seen as the removal of a hydrogen atom on one face of the substrate followed by readdition of the hydrogen at the opposite site of that carbon atom (Figure II.10). However, despite the simplicity of representation of this reaction, the chemistry behind these transformations is much more complex; epimerization does not occur spontaneously as carbohydrates are extremely stable. On the other hand, sugar epimerases have been found in all branches of life and, in theory, they can change stereochemistry at all different stereocenters of sugars. Some enzymes even epimerize two centers simultaneously during turnover, like dTDP-6-deoxy-D-xylo-4-hexulose 3,5-epimerase.

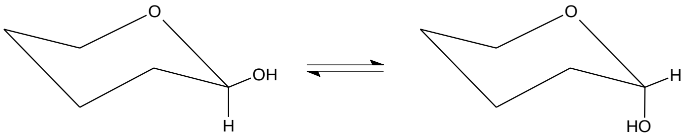


Figure II.10 Epimerization; inversion of chirality at a asymmetrically substituted carbon

To date, over 20 types of sugar epimerases have been reported active on C1, C2, C3, C4, C5, and C6 centers of mono- and oligosaccharides, using several completely different reaction mechanisms to invert chirality. Among these epimerase mechanisms are mutarotation or ring opening (e.g. aldose 1-epimerase), proton abstraction/readdition (e.g. D-ribulose-5-phosphate 3-epimerase), nucleotide elimination (e.g. UDP-N-acetylglucosamine 2-epimerase), carbon-carbon bond cleavage (e.g. L-ribulose-5-phosphate 4-epimerase), and transient keto intermediate (e.g. UDP-galactose 4-epimerase) (Table II.4). These latter two mechanisms will be further outlined in the literature part on L-ribulose-5-phosphate 4-epimerase (see 3.2) and UDP-galactose 4-epimerase (see 3.3), respectively. These two C4-epimerases will be discussed in further detail as they are applied in this research project. More details on the mechanisms for epimerization can be found in the following reviews [5, 112, 226, 251].

Table II.4 Mechanistic types of epimerization (adapted from [5])

Epimerization mechanism	Example enzyme	Epimerization site
Mutarotation or ring opening	aldose 1-epimerase	C1
Proton abstraction/readdition	D-ribulose-5-phosphate 3-epimerase	C3
	tagatose/psicose 3-epimerase	C3
	GDP-4-keto-6-deoxy-D-mannose epimerase/reductase	C3, C5
	dTDP-6-deoxy-D-xylo-4-hexulose 3,5-epimerase	C3, C5
Nucleotide elimination	UDP-N-acetylglucosamine 2-epimerase	C2
	N-acyl-D-glucosamine 2-epimerase	C2
Carbon-carbon bond cleavage	**L-ribulose-5-phosphate 4-epimerase**	**C4**
Transient keto intermediate	**UDP-galactose 4-epimerase**	**C4**
	ADP-L-glycero-D-mannoheptose 6-epimerase	C6
	CDP-tyvelose 2-epimerase	C2

As can be seen in Table II.4, two epimerases active on the C4-stereocenter have been found in nature until now. As a C4-epimerization is the desired reaction, these two enzymes will serve as a template for enzyme engineering. Therefore, both the L-ribulose-5-phosphate 4-epimerase (3.2) and UDP-galactose 4-epimerase (3.3) will be discussed in further detail now.

3.2. L-ribulose-5-phosphate 4-epimerase

3.2.1. General

The bacterial L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4, AraD) is one of the enzymes that enables many bacteria to use L-arabinose as a source of energy by coupling the L-arabinose pathway to the pentose phosphate pathway [5]. The gene is located in the *araBAD* operon and, together with L-arabinose isomerase (AraA) and ribulokinase (AraB), it metabolizes L-arabinose into D-xylulose-5-phosphate which is subsequently taken up in the pentose phosphate cycle (Figure II.11). Other genes involved in this pathway are *araC*, *araE* and *araFG* which are necessary for gene regulation, low-affinity and high-affinity transport, respectively [157].

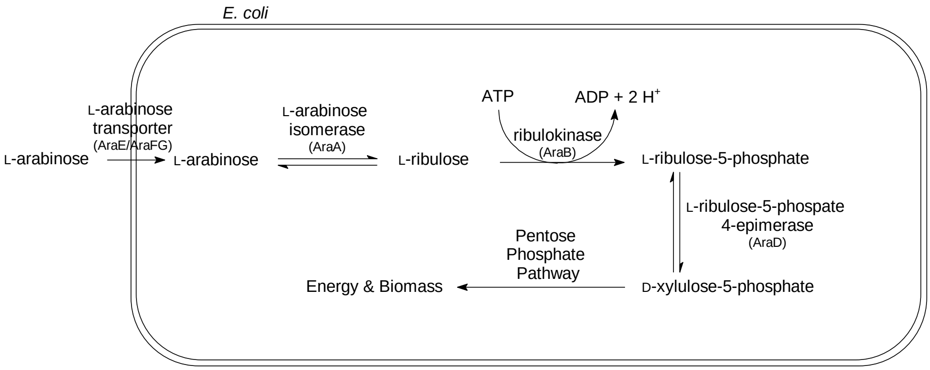

E. coli
L-arabinose
L-arabinose transporter
(AraE/AraFG)
L-arabinose
L-arabinose isomerase
(AraA)
L-ribulose
ATP
ADP + 2 H+
ribulokinase
(AraB)
L-ribulose-5-phosphate
L-ribulose-5-phospate 4-epimerase
(AraD)
Pentose Phosphate Pathway
Energy & Biomass
D-xylulose-5-phosphate

Figure II.11 L-Arabinose pathway in *E. coli*

In *E. coli*, two other homologous L-ribulose-5-phosphate 4-epimerases are expressed besides AraD, namely the gene products from *sgbE* and *ulaF*. These two enzymes also catalyze the same reaction but are differently induced, making them active in other pathways. Whereas the gene product for *araD* is expressed in L-arabinose catabolism, SgbE and UlaF are active in the degradation of L-lyxose and L-ascorbic acid, respectively. All three epimerases reversibly interconvert L-ribulose-5-phosphate and D-xylulose-5-phosphate by inverting the configuration of the hydroxyl group at the C4-stereocenter via the same mechanism with the use of a divalent metal ion (Figure II.12).

3.2.2. Mechanism

As the L-ribulose-5-phosphate 4-epimerase epimerizes a stereocenter that does not bear an acidic proton, it cannot apply a simple deprotonation/reprotonation mechanism. Furthermore, it was proven that no NAD^+ cofactor is involved in the reaction and the oxidation/reduction mechanism is thereby excluded. Initially, two mechanisms were proposed for this enzyme, namely an aldolase/retroaldolase mechanism applying a carbon-carbon bond cleavage and re-addition or a dehydratation/rehydratation mechanism in which an enone intermediate is formed [74]. Extensive research provided clear evidence for the retroaldolase/aldolase mechanism which is describe in more detail below.

Despite the low sequence identity between L-ribulose-5-phosphate (L-Ru-5-P) 4-epimerase and L-fuculose-1-phosphate aldolase (FucA), the structural resemblance is striking. The L-fuculose-1-phosphate (L-Fuc-1-P) aldolase is the best characterized example of the class II aldolases and catalyzes the condensation of L-lactaldehyde and dihydroxyaceton phosphate into

L-fuculose-1-phosphate or the reverse cleavage reaction [120]. For the aldolase, conserved residues include three histidines (His92, His94, and His155) and a glutamate (Glu73) which are necessary for binding zinc into the metallo-enzyme. These four residues are structured in more or less tetrahedral position around the metal ion. Upon binding of the substrate, the bond between the glutamate and zinc ion is broken and L-Fuc-1-P is positioned towards the zinc ion with its carbonyl function as well as the hydroxyl group at the C3-stereocenter. The now liberated acidic moiety of the glutamate residue, subsequently deprotonates the hydroxyl group at the C4-stereocenter and hereby initiates the cleavage of the carbon-carbon bond between C3 and C4. In this manner, L-lactaldehyde and the metal bound enolate of dihydroxyaceton phosphate (DHAP) are formed. After cleavage, the acidic form of Glu73 will protonate the enolate intermediate in order to form dihydroxyaceton phosphate (Figure II.12a).

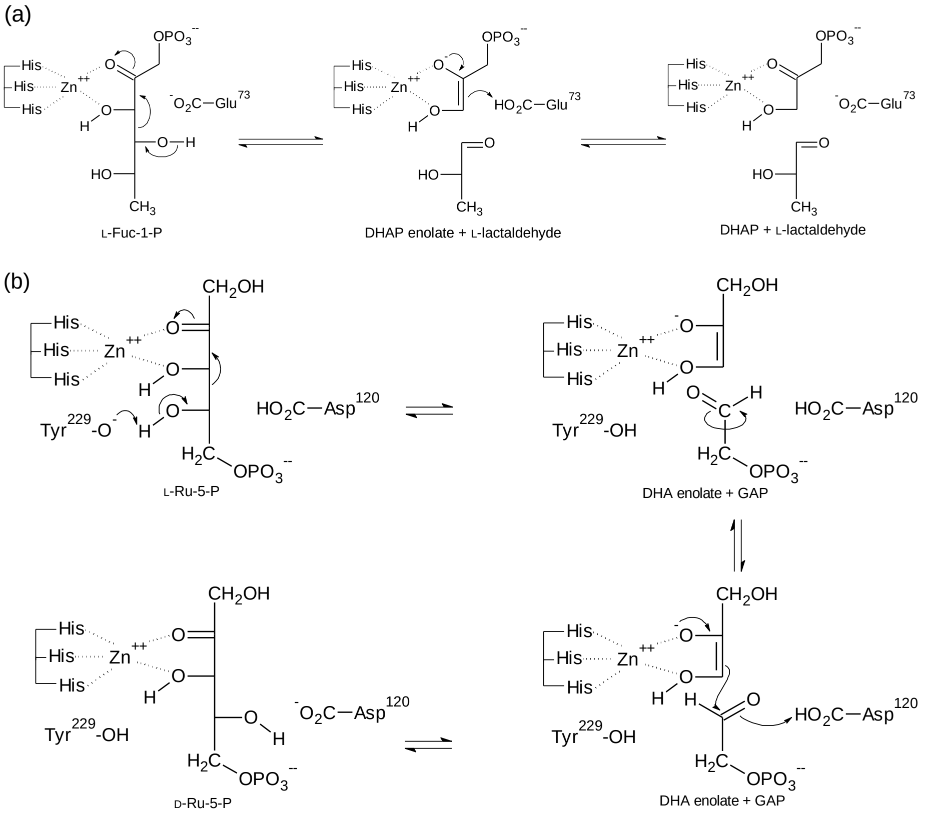

(a)
L-Fuc-1-P
DHAP enolate + L-lactaldehyde
DHAP + L-lactaldehyde
(b)
L-Ru-5-P
DHA enolate + GAP
D-Ru-5-P
DHA enolate + GAP
His
Zn
Glu73
Asp120
Tyr229

Figure II.12 Comparison between the mechanisms of (a) L-Fuc-1-P aldolase and (b) L-Ru-5-P 4-epimerase

The strong structural similarity suggested that L-Ru-5-P 4-epimerase and L-Fuc-1-P aldolase belong to the same superfamily or have evolved from a common ancestor [251]. Furthermore, it provided a first provisional evidence that the epimerase applies the retroaldolase/aldolase mechanism. Evidence that supports this relatedness was provided by mutagenesis of some of the conserved residues in L-Ru-5-P 4-epimerases, namely His95, His97 and Asp76. Residues were mutated towards asparagines and the resulting enzyme variants exhibited a reduced epimerase activities, in the range of 3-200-fold reduction. More importantly, enzyme variants also showed reduced affinity for divalent zinc ions, and thus require exogenous Zn^{2+} for full activity. This demonstrates their importance in catalysis, namely via their role in binding of the divalent metal ion. Further sequence alignments of the epimerase and aldolase suggests His171 to be the fourth zinc binding residue. However, this hypothesis remains to be tested by mutagenic studies [120].

For the epimerase, the initial retroaldolase cleavage of the L-Ru-5-P results in the enolate of dihydroxyaceton (DHA) and glycolaldehyde phosphate (GAP) as enzyme bound intermediates. In the epimerase, the protonation of the enolate intermediate is not promoted by acid/base catalytic residues. In this case, the formed aldehyde turns around in the catalytic cleft facing its opposite side towards the enolate, which is then followed by the reverse reaction, namely the aldol condensation. As a result of the condensation at the other side of the aldehyde, D-xylulose-5-phosphate is formed resulting in an overall epimerization at C4 (Figure II.12b). Acid/base catalysts responsible for this difference were found to be an aspartic acid (Asp120) and a tyrosine residue (Tyr229) rather than the Asp76 which corresponds to the Glu73 in the aldolase [148, 225].

The overall structural similarity and common metal-binding motif are insufficient to prove that both enzymes use the same or a very similar reaction mechanism. This was evidenced by the fact that both the wild-type epimerase and its D76N variant catalyzed the aldol condensation of dihydroxyaceton and glycolaldehyde phosphate to form a mixture of both L-Ru-5-P and D-xylulose-5-phosphate [120]. In addition, ^{13}C and deuterium isotopic effect measurements both ruled out the dehydration mechanism and conformed epimerization via the retroaldolase/aldolase mechanism [149]. Finally, the solving of the epimerases structure by means of X-ray crystallography showed more resemblances between both enzymes and gave more insights in acid/base catalysts for the epimerase. Overall structural features and important residues and motifs, like metal binding motif, acid/base catalysts, phosphate binding site, are described in the next paragraphs.

3.2.3. Structure

3.2.3.1. Tertiary and quaternary structure

The crystal structure of the *Escherichia coli* L-ribulose-5-phosphate 4-epimerase was solved by X-ray crystallography (Figure II.13a). Like the L-fuculose-1-phosphate aldolase, this enzyme consist of four identical subunits and displays a C4 symmetry. Each subunit consists of 231 amino acids resulting in a molecular weight of 25.5 kDa per subunit, and thus a total mass of 102 kDa for the tetramer. Each subunit contains one single domain with a typical α/β fold. In this fold, a central β-sheet formed by nine β-strands is sandwiched by two layers of α-helices. The smaller layer consist of three helices (a2, a5 and a6), whereas the other five helices shape the larger layer (a1, a3, a4, a7 and a8). The β-sheet is predominantly antiparallel except between b7 and b8. At all four subunit interfaces, a zinc ion is present which is necessary for catalytic activity. Residues from both adjacent subunits contribute to the active site clefts [166].

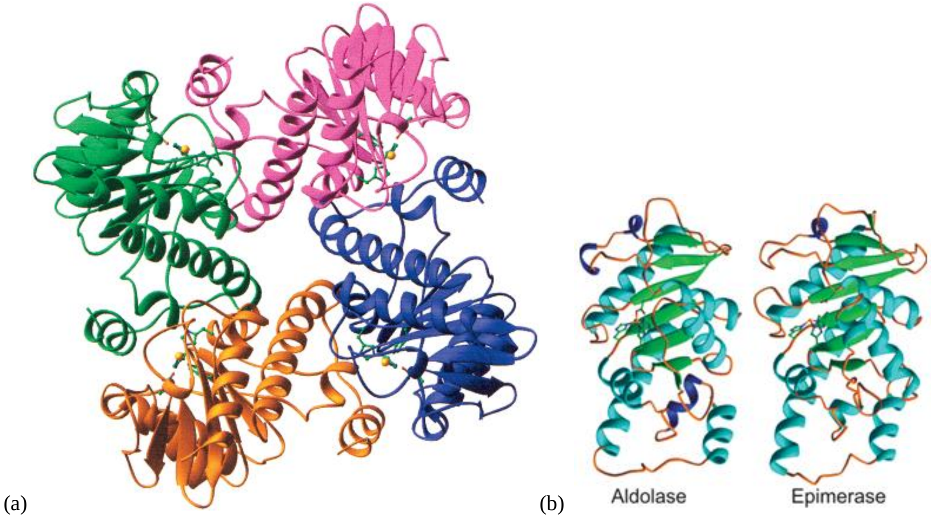


Figure II.13 (a) Overall structure of the L-ribulose-5-phosphate 4-epimerase tetramer (Adapted from [166]) (b) Comparison of the monomer of the epimerase (left) and the L-fuculose-1-phosphate aldolase (Adopted from [251])

Even though the sequence identity between both enzymes is only 26 %, both enzymes display strikingly similar tertiary and quaternary structures. This is very clear when the enzymes subunit structures and tetrameric structures are superimposed (Figure II.13b). For subunits, 93 % of the α-carbon atoms overlap with standard deviation of only 1.5 Å, whereas the overlay of the tetrameric structures results in an α-carbon atoms overlap with 1.9 Å standard deviation.

Furthermore, all secondary structure elements from the smaller aldolase are returning in the larger epimerase. The main difference is an additional α-helix in the smaller α-helix layer of the epimerase. Smaller differences are a longer α-helix a6 and a 3_{10}-helix instead of the α-helix a4 in the epimerase. Despite the subtle difference between a 3_{10}-helix and an α-helix, this difference is likely to affect its contribution to catalysis as this helix participates in active site formation [166]. In the crystal structure, the first 223 residues of each subunit are ordered, whereas the latter eight residues (224-231) are undefined indicating a highly flexible loop. As postulated for the aldolase, the epimerases loop is most likely also able to close the active site [166]. The importance of this loop is supported by mutagenic studies of the Tyr229' residue which is part of this loop. Mutagenesis of the tyrosine towards a phenylalanine resulted in a dramatically decreased activity, more specifically 1700-fold slower [148, 251].

3.2.3.2. Cation (zinc) binding site

As was earlier stated, three conserved histidine residues (His95, His97 and His 171) are ligands for binding of the zinc ion. In the aldolase the Glu73 serves as fourth ligand, however in the epimerase this residue is replaced by a shorter but topological variant, namely Asp76. Due to the shorter side chain of the aspartate, the Asp76 cannot act as the fourth Zn^{2+} binding ligand. A water molecule positioned in front of the Asp76 will serve as the Zn^{2+} binding ligand (Figure II.14) [166]. However, the amount and exact positioning of water molecules remains unclear as three waters (octahedral: 3 times His, 3 times water) were observed in EPR spectrum of Mn^{2+} substituted epimerase [148]. On one hand, changing metal ion could influence ligand coordination and on the other hand both Zn^{2+} and Mn^{2+} are able to accommodate both 4 (tetrahedral) and 6 (octahedral) ligands in their coordination sphere [166].

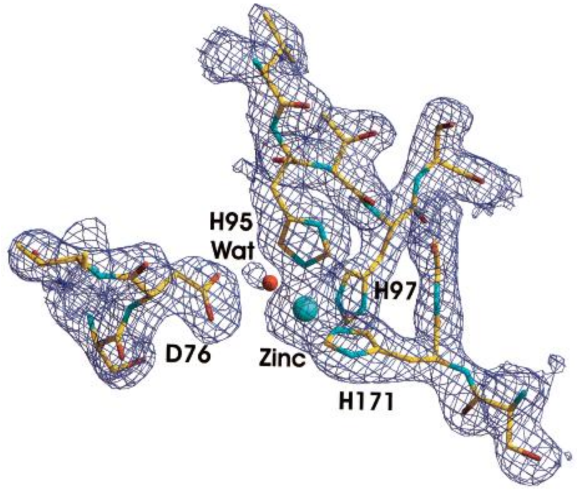


Figure II.14 Electron density map of the cation binding site of L-Ru-5-P 4-epimerase shows that the three conserved histidines interact with the Zn^{2+}, whereas a water molecule is needed to link the aspartate to the zinc ion (Adopted from [166])

3.2.3.3. Substrate binding site and potential catalytic acid/base residues

As mentioned earlier, the L-Ru-5-P 4-epimerase and L-Fuc-1-P aldolase are evolutionary related and share some common features, such as structure, divalent cation as cofactor and are able to deprotonate a hydroxyl group of a phosphoketose at the C4 position. Despite these striking similarities, these enzymes catalyze different reactions and act on distinctly phosphorylated substrates, at C1 and C5 for the aldolase and epimerase, respectively. Therefore, significant differences between residues involved in substrate binding and/or activity are observed when comparing both enzymes [225]. Based on sequence alignments between both enzymes, residues were chosen that could play a possible role in substrate binding and/or might serve as catalytic acid/base residues [166]. Mutagenic studies were performed to determine the influence of amino acid substitutions on affinity and activity of the epimerase activity [225]. By these studies, the residues that play an important role in substrate binding and reaction mechanism (acid/base catalysts) could be indicated with high probability.

Substrate/Phosphate binding site

Among the conserved residues are some amino acids that are known to be of importance or show potential for phosphate binding, among which are Asn28, Ser44, Gly45, Ser 74 and Ser 75 (Figure II.15) [166]. Besides these five residues that are conserved among both enzymes, a conserved lysine is found among epimerases, namely Lys42. Site directed mutagenesis of Asn28 and Lys42 revealed their importance in substrate binding, as the N28A and K42M mutations drastically reduced affinity of the epimerase for L-Ru-5-P. The asparagine is expected to form a hydrogen bond with the phosphate moiety as was observed for its topological variant in the aldolase. The lysine in its turn is located just below the substrates phosphate and impairment with the substrate is achieved through electrostatic attraction between the positively charged lysine and negatively charged phosphate moiety. Its loss thus resulted in a much higher K_M value [225].

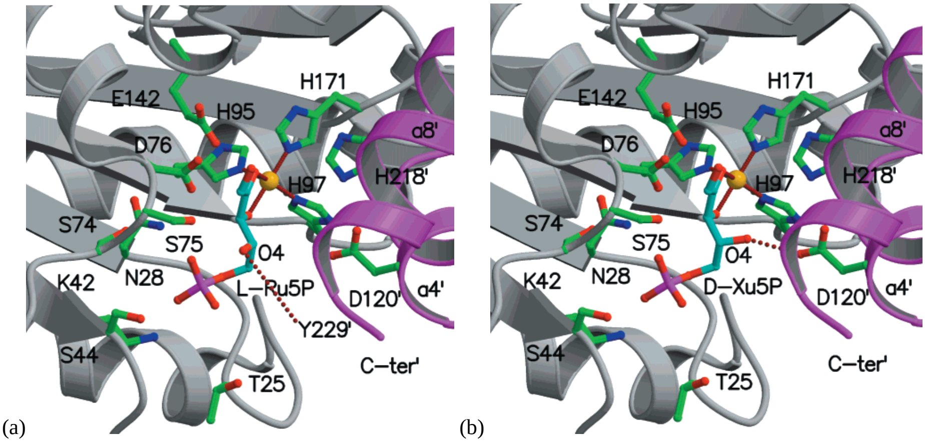


Figure II.15 Substrate binding site of L-Ru-5-P 4-epimerase. The catalytic site is formed within the cleft between 2 adjacent subunits (grey and purple, respectively). The bound substrates are (a) L-Ru-5-P and (b) D-Xu-5-P and the red dotted lines connect the C4 hydroxyl group with the assumed deprotonation residue (Adopted from [225])

As the substrates of both enzymes are differently phosphorylated (C5 for the epimerase vs. C1 for the aldolase), a common phosphate-binding pocket means that the substrates are bound in a reversed or ‘flipped’ orientation. This further implicates that acid/base catalytic residues of both enzymes are different.

Catalytic acid/base residues

Based on a structural search, four conserved residues were pointed out to be potential acid/base catalytic residues. These residues are Asp76 and Glu142 from the subunit containing the zinc cofactor and Asp120’ and His218’ from the adjacent subunit [166]. The Asp76 is most likely to affect epimerization by providing a good bonding with the metal cofactor, whereas loss of the carboxylic group of the Asp120 resulted in a 3000-fold decrease in k_{cat} and is thus likely to serve as a acid/base residue in epimerization. On the other hand, mutagenesis of the other two residues (E142Q and H218N) only reduced activity by 9- and 24-fold, respectively. This rules them out as catalytic residues, nevertheless, they are important for activity and stability. The histidine is expected to assist Asp120’ in catalysis and in addition stabilize the active site through hydrogen bonding with the conserved Tyr141. Similarly, the glutamate interacts with Arg221’ from the adjacent subunit via a salt bridge and is thus likely to stabilize the active site interface [225]. Mutagenesis of a tyrosine (Tyr229’) in the flexible C-terminal end of the epimerase resulted in a 1700-fold decreased k_{cat} value and was therefore pointed out as the second acid/base residue [148]. Docking of the substrates into the enzymes structure predicts that Tyr229’ is responsible for L-Ru-5-P deprotonation and Asp120’ for deprotonation of D-Xu-5-P [225]. It is important to notice that residues from both subunits are needed for epimerase activity, either for substrate or cofactor binding or for catalysis.

3.3. UDP-Galactose 4-epimerase

3.3.1. General

Uridine diphosphate galactose 4-epimerase, or UDP-Gal 4-epimerase or shortly GalE (EC 5.1.3.2), is one of the enzymes in the Leloir pathway. This pathway is responsible for the conversion of galactose into glucose-1-phosphate with the help of four enzymes. At first, galactose mutarotase converts β-galactose into its α-anomer. Next, this α-galactose is phosphorylated by a galactokinase to produce galactose-1-phosphate. In the third step, galactose-1-phosphate uridylyltransferase exchanges the uridylylgroup of UDP-glucose (UDP-Glc) onto the galactose-1-phosphate, hereby forming UDP-galactose (UDP-Gal) and glucose-1-phosphate. Glucose-1-phosphate will be metabolized further in glycolysis. Finally, UDP-Gal will be epimerized by UDP-Gal 4-epimerase for regeneration of UDP-Glc, which can then be applied in new cycles with the transferase (Figure II.16) [223]. In case of detrimental mutations in one of these enzymes, carriers will suffer from a disease called galactosemia. Vomiting, jaundice, and lethargy are early symptoms of this rare but deadly disease. Delayed complications include mental retardation, liver cirrhosis, renal failure, and cataracts [101]. Nevertheless, the disease can be easily prevented or treated by a galactose and lactose free diet.

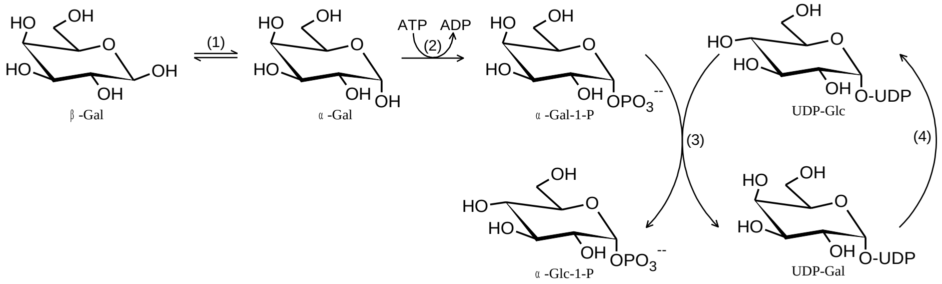


Figure II.16 The Leloir pathway. (1) galactose mutarotase, (2) galactokinase, (3) galactose-1-phosphate uridylyltransferase and (4) UDP-galactose 4-epimerase

On the other hand, UDP-sugars are also important as precursors in lipopolysaccharides (LPS). LPS is involved in several aspects of cell-cell interactions, such as host-pathogen interactions [110] and biofilm formation [187]. Apart from non-modified sugars, also N-acetylated and carboxylated sugars are found in LPS [32, 83]. Consequently, the corresponding UDP-sugars have to be made by the cells and GalE-like UDP-sugars 4-epimerases are used for this purpose. The UDP-sugars 4-epimerase family consists of epimerase active on UDP-Glc/UDP-Gal, N-acetylated forms (UDP-GlcNAc/UDP-GalNAc) [32], uronic acid forms (UDP-GlcUA/UDP-GalUA) [83] as well as UDP-pentoses (UDP-L-Ara/UDP-Xyl) [137] or on combinations thereof [71, 110, 187]. Attempts have been made to elucidate the determinants for substrate specificity of UDP-sugars 4-epimerases as this can render information to use these epimerases as targets for new antibiotics and drugs [58]. Substrate specificity will be further discussed in part 3.3.4.

The UDP-Gal 4-epimerases (and the other GalE-like UDP-sugar 4-epimerases) belong to the short-chain dehydrogenase/reductase (SDR) superfamily of proteins. These enzymes show great functional diversity and despite their lower sequence identities (typically only 15-30%) specific sequence motifs are detectable, reflecting their common folding patterns. SDRs are widely spread in nature and involved in different physiological processes such as normal and metastatic growth, hypertension and fertility [101].

3.3.2. Structure

UDP-Gal 4-epimerases are formed as homodimers, and the *E. coli* monomer, for example, contains 338 amino acids resulting in a molecular weight of 37.3 kDa [5]. The human epimerase is slightly longer (348 amino acids) and its dimer weighs 76.6 kDa (2 times 38.3 kDa). Given that UDP-Gal 4-epimerase is part of the big SDR superfamily, it displays some typical characteristics of this superfamily. As such, the Sx_nYx_3K motif as well as the GxxGxxG motif are two of the encountered signature sequences [95, 261]. The Sx_nYx_3K contains three conserved residues which play a key role in catalysis. The repetitive glycine motif is located in the Rossmann fold, which is typical for nucleotide binding enzymes, and the first 2 glycines participate in NAD^+ binding whereas the third glycine facilitates close packing of the helix to the beta-strand [152].

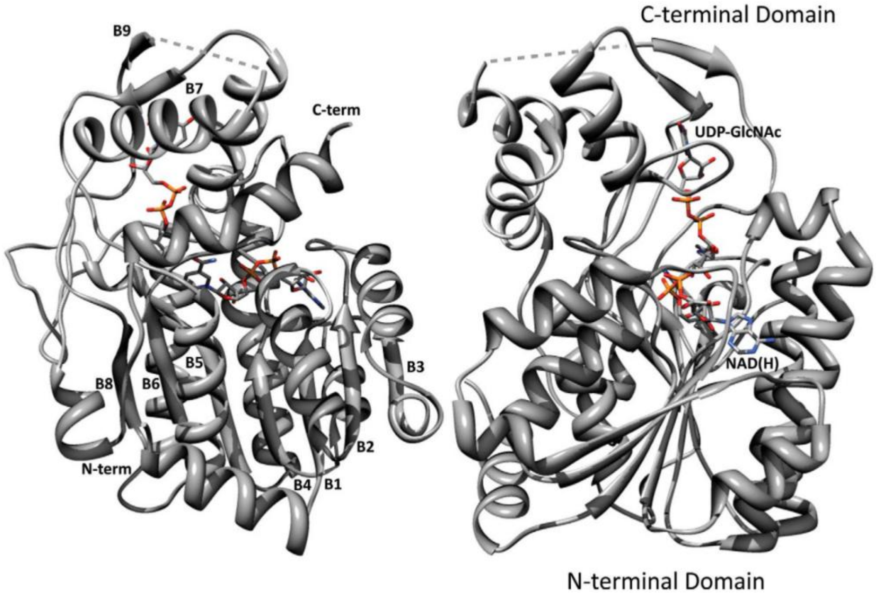


Figure II.17 Overall structure of the UDP-GalNac 4-epimerase from *Pleisomonas shigelloides*. The N-terminal domain is a modified Rossmann fold binding the cofactor NAD(H), whereas the C-terminal domain binds the substrate UDP-GlcNAc. Strand numbering from N- to C-terminus, dashed lines represent a non-modeled loop (Adopted from [32])

Crystal structures of UDP-Gal 4-epimerases (GalE) from several sources have been determined using X-ray crystallography (Figure II.17). The crystal structure from the *E. coli* enzyme is further explained here. Two different domains are distinguished, namely a N-terminal nucleotide binding domain and a smaller C-terminal domain which is responsible for the correct positioning of its substrate, a UDP-sugar. The N-terminal domain comprises seven stranded parallel β-sheet which are flanked on both sides by α-helices and shape the Rossmann fold [257, 259]. Two paired Rossmann folds tightly bind one NAD^+ cofactor per subunit. In *E. coli* GalE, the NAD^+ interacts more extensively with the protein than was observed with other SDR enzymes. A total of 35 protein-NAD^+ contacts was observed with distances up to 3.2 Å, of which seven were contributed by ordered water molecules, while other SDR enzymes showed 22-27 contacts. This results in irreversible denaturation of the enzyme after removal of the cofactor [254]. However, the NAD^+ cofactor could be removed from human GalE without denaturation. Here, fewer protein-NAD^+ contacts were observed in the crystal structure, which explains the reversible character of cofactor binding [260]. The C-terminal domain is built from five β-strands and four α-helices. As the domains are necessary for the binding of the cofactor and the substrate, respectively, the active site is located between these two domains [257, 259]. Determination of the structure of human UDP-Gal 4-epimerase, revealed an active site which was 15 % larger than that of the *E. coli* enzyme. A possible explanation was found in the secondary role of the human enzyme, namely epimerization of UDP-N-acetylgalactosamine (UDP-GalNAc). Activity on the larger acetylated substrates would require a larger active site [5, 261].

3.3.3. Mechanism

The GalE mechanism was already studied and broadly outlined in the 1970s [272, 273]. The first step after substrate binding is the abstraction of the hydroxylic proton at C4 by an enzymatic base and the transfer of a hydride from the C4 position of the sugar to NAD^+ to form NADH and a transient keto sugar. Three important residues are located in the conserved Sx_nYx_3K motif. At first, the tyrosine was postulated to be the enzymatic base responsible for deprotonation, as occurs in its phenolic form. The nearby lysine residue has a stabilizing effect on this phenolic tyrosine [162]. However, except for human GalE, the distance between this tyrosine and the sugar moiety of the substrate is too large (4.3 Å) to directly abstract the proton [256]. To overcome this distance, a proton shuttle is needed. In GalE, it is provided by the conserved serine from the Sx_nYx_3K motif [255]. The enzyme retains the transient keto sugar by anchoring of the substrate by its UDP-group. Finally, the reverse of the first step occurs, namely the transfer of the hydride back to the C4 of the sugar. This is now transferred on the opposite face, resulting in the inversion of the configuration at C4 of the sugar. The proton that was extracted by the serine-tyrosine proton shuttle is also transferred back to the sugar. A schematic overview of the mechanism is given in Figure II.18.

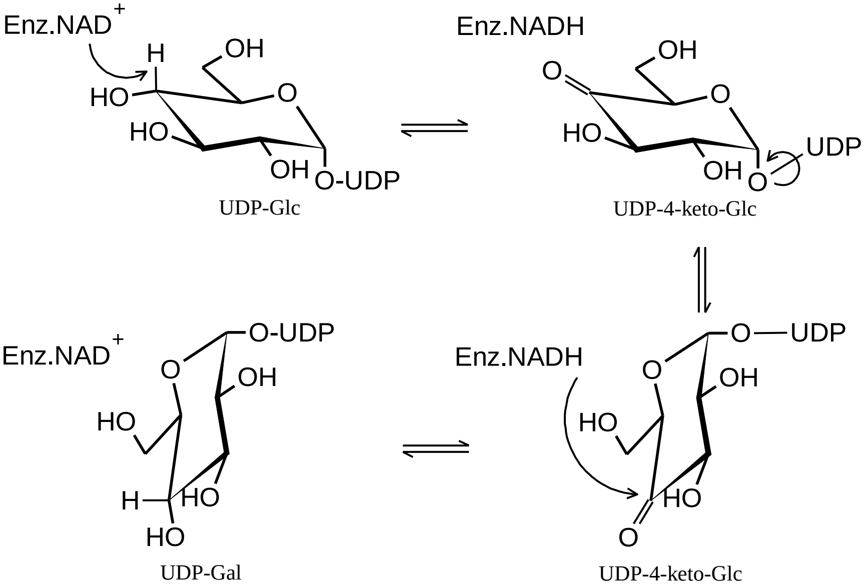


Figure II.18 Mechanism of UDP-galactose 4-epimerase

A first important feature is the capability of the enzyme to bind the substrates sugar moiety in different positions. The substrate is mainly retained by the enzyme by interactions with the UDP-group, whereas only weak binding with the sugar moiety is observed. This led to the proposition of a rotation around the bond connecting the glycosyl oxygen atom and the β-phosphorus atom in the pyrophosphoryl linkage. X-ray crystallography of inactivated enzyme (S124A/Y229F) that was crystallized with each substrate separately gave evidence of the rotation of the sugar moiety. Significant changes were also observed in the two dihedral angles that define the substrate configuration; a 130° rotation was observed for the angle delineated by the α,β-bridging oxygen, the β-phosphorus, the glycosyl oxygen, and the hexose C-1, whereas that delineated by the β-phosphorus, the glycosyl oxygen and the hexose C1 and C2 differ 30°. This mechanism is called 'revolving door mechanism' and is unusual in biology [5, 259].

A second important feature is the fact that the enzyme undergoes a conformational change upon binding of the UDP-sugar, which is in fact a result of the binding of the UMP-moiety of the substrate. The presence of the UDP-group in the substrate is thus not only necessary to increase the affinity of the epimerase for the glucose/galactose, but it also increases the reactivity of the enzyme-bound NAD^+ cofactor by inducing a different protein conformation [272, 275]. Several studies support this proposal; the enzyme oxidizes free sugars in the presence of UMP or UDP and occasionally releases the ketose intermediate into solution, leaving behind an inactivated enzyme containing NADH (E·NADH) [30, 34], and enzyme can be chemically reduced by treating it with $NaBH_4$, under anaerobic conditions [161]. However, the resulting E·NADH is rapidly auto-oxidized in the presence of oxygen which indicates the importance of the uridine nucleotide in stabilizing the reduced enzyme. In addition, it has been shown that UMP and UDP lower the activation energy by 5.7 and 4.1 kcal mol^{-1}, respectively [62, 161]. NMR studies with uridine nucleotide-dependent perturbation of the ^{31}P, ^{13}C and ^{15}N provide further evidence for the postulated protein conformational change [42, 135]. The conserved lysine from the Sx_nYx_3K motif plays an important role in the activation of the cofactor, as due to the conformational change, the 6-ammonium group is hydrogen-bonded to both the 2'- and 3'-hydroxylgroups of the nicotinamide riboside of NAD^+. This results in a charge displacement from nicotinamide-N1 to nicotinamide-C4 which activates the NAD^+ to attract a hydride (Figure II.19) [256].

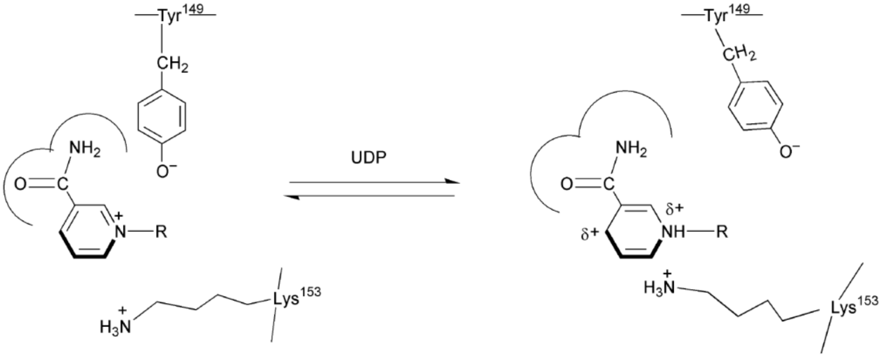


Figure II.19 Schematic representation of the activation of the GalE·NAD$^+$ complex due to the conformational change induced by binding of UDP (Adopted from [226])

3.3.4. Substrate promiscuity

As mentioned before, a variety of UDP-sugars 4-epimerases have been found with different substrate specificities and promiscuities (see 3.3.1). Among these were epimerases active on UDP-Glc/UDP-Gal [261], UDP-GlcNAc/UDP-GalNAc [32], UDP-GlcUA/UDP-GalUA (uronic acid form) [83], UDP-L-Ara/UDP-Xyl (pentoses) [137], or multiple of these substrates [71, 110, 187]. As they are involved in different pathways and functions, such as protein glycosylation and production and secretion of virulence factors, these enzymes are possible targets for new therapeutics. Therefore, it is of great importance to understand the molecular basis of substrate promiscuity and/or selectivity [71].

3.3.4.1. UDP-Glc/UDP-Gal and/or UDP-GlcNAc/UDP-GalNAc (N-acetylation of UDP-sugars)

Research conducted to substrate promiscuity of UDP-sugars 4-epimerases has mainly focused on the determinants that make the epimerases discriminate between non-acetylated and N-acetylated UDP-Glc/UDP-Gal. The discovery of the genuine UDP-GlcNAc 4-epimerase (WbpP) from *Pseudomonas aeruginosa* has led to a classification of UDP-hexose 4-epimerases into three groups with distinct substrate promiscuity [58]. Group 1 contains the 4-epimerases that exhibit a strong preference for non-acetylated substrates, such as *E. coli* GalE (eGalE), group 2 members can epimerize both non-acetylated and N-acetylated substrates equally well, such as the human epimerase (hGalE), and group 3 epimerases are strongly specific for N-acetylated substrates, like the WbpP from *P. aeruginosa* [110].

With the structural characterization of WbpP in the presence of both substrates, structures from all distinct groups were available for research to understand the structural determinants for substrate specificity [110]. Interactions between the enzyme and the substrate in the binding pocket were scrutinized, which has led to a structural model showing that only a limited number of residues predominantly contribute to the molecular basis for substrate specificity. In this model, the substrate-binding pocket is represented as a hexagonal-shaped box (Figure II.20), with the bottom formed by the nicotinamide ring of the cofactor and an open top to accommodate the ring-flipping movement during catalysis. Three of the six walls of the hexagonal box are formed by highly conserved residues: Ser142, Tyr166 and Asn195 in WbpP (Figure II.20 - yellow, blue and orange walls, respectively). The serine and tyrosine are part of the Sx_nYx_3K catalytic triad. The other three walls (Gly102, Ala209 and Ser306 for WbpP) have been proposed to be key determinants for substrate specificity.

The green wall is occupied by a bulky residue Tyr299 in eGalE, which is unable to catalyse the epimerization of acetylated substrates, whereas enzymes with a smaller residue are able to convert acetylated substrates. A cysteine is found in hGalE [261] and *Yersinia enterocolitica* Gne (yGne) [26], leucine in *Campylobacter jejuni* Gne (cGne) [27] and Ser306 in WbpP [23] (Figure II.20, green wall). Therefore, it was suggested that a wider binding pocket would allow both substrates to enter for catalysis, whereas a narrow one would limit catalysis to non-acetylated substrates only and that the enzymes substrate spectrum could be predicted based on the sequence only [110]. This theory is supported by mutagenic studies of hGalE and yGne, as narrowing binding pockets size, by a C307Y and a C297Y mutant, respectively, resulted in significant loss of activity on acetylated substrates without affecting conversion of non-acetylated substrates [26, 232]. However, reverse mutations challenge this theory, a Y299C mutation in eGalE resulted in significant loss of catalysis of non-acetylated substrates [258] and a S306Y mutation of WbpP totally abolished the activity of the enzyme [110]. It was suggested that broadening the binding pocket, might result in unproductive conformations of the non-acetylated sugar moiety or that the tyrosine might have a role in catalysis itself. As such, the serine residue (S306) in WbpP could achieve a tighter binding pocket packing due to hydrogen bond formation with active-site water molecules to other residues in the wall of the hexagonal box, such as S142 and N195 [71, 110].

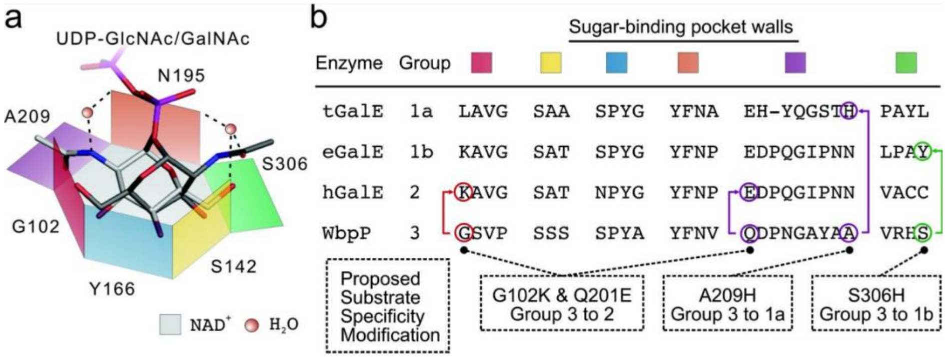

a
UDP-GlcNAc/GalNAc
N195
A209
S306
G102
S142
Y166
NAD⁺
H₂O
b
Sugar-binding pocket walls
Enzyme
Group
tGalE 1a LAVG SAA SPYG YFNA EH-YQGSTH PAYL
eGalE 1b KAVG SAT SPYG YFNP EDPQGIPNN LPAY
hGalE 2 KAVG SAT NPYG YFNP EDPQGIPNN VACC
WbpP 3 GSVP SSS SPYA YFNV QDPNGAYAA VRHS
Proposed Substrate Specificity Modification
G102K & Q201E Group 3 to 2
A209H Group 3 to 1a
S306H Group 3 to 1b

Figure II.20 (a) A hexagonal box model was suggested as a representation of substrate specificity of UDP-hexose 4-epimerases (b) The corresponding residues found in other types of UDP-hexose 4-epimerases are indicated, including neighboring residues (adopted from [71])

Of the other hexagon walls, experimental data are available dealing with the fifth purple wall, namely the A209 in WbpP whereas more bulkier residues like asparagine or histidine are found at the equivalent positions in the other epimerases. This will partially or totally close the hexagonal box and hereby limit the ability to accommodate acetylated substrates [71]. An A209H mutant was made to check the effect of a re-introducing of a bulkier residue on activity on acetylated substrates. The mutation resulted in limited ability to epimerize acetylated residues and thus partially validates the predictions on this residues function [110]. This position was also mutated towards an asparagine, resulting in the A209N mutant. It was expected to change specificity from group 3 (preference for acetylated substrates) to group 2 (accept both acetylated and non-acetylated equally well); however, it was found to enhance specificity for acetylated substrates, and this was accompanied by a lower catalytic efficiency. The effect is most likely arising from a hydrogen bond between the introduced asparagine and Gln201, which retains the asparagine in the open conformation. In group 2 members, the asparagine is known to swing back and forth, depending on whether or not the substrate is acetylated [71].

The effect of residues in the third non-conserved wall was checked by two single mutants, G102K and Q201E, as well as the combined double mutant, G102K/Q201E, in the genuine UDP-GlcNAc 4-epimerase WbpP. Both single mutants showed slightly reduced activity on acetylated substrates and (almost) abolished activity on non-acetylated substrates. As a result of the introduction of both mutations at the same time, a salt bridge could be formed which resulted in a rescue of the activity for acetylated substrates, probably due to restoration of the slight distortion that was observed in both single mutants. The side chains of these residues were found not to be directly involved in controlling access of large substrates to the binding site, but seemed to be involved in establishing other interactions that keep the binding pocket more or less flexible. Flexibility is needed here to allow a proper positioning of the substrates into the cavity, and in this way avoid unproductive configurations [71].

Other residues that were targeted by site directed mutagenesis were the two serine residues that are located next to the catalytic serine, namely S143 and S144. The S143A mutation abolished activity on non-acetylated substrates, probably due to loss of the hydrogen bonding, whereas the mutant remained active on UDP-GlcNAc/UDP-GalNAc, as additional stabilizing interactions with the N-acetyl moiety are present. The second serine was mutated towards a lysine and this mutant (S144K) showed no activity at all, nor was dehydratase activity observed. This serine thus seems essential for activity; however, the exact function is yet unknown [71].

The availability of a second group 3 epimerase, the WbgU from *Pleisomonas shigelloides*, and its structure provided more information on the substrate specificity of epimerases in this group [32]. Despite the relatively low sequence identity among all three groups, the similarity of the enzymes' tertiary structures is striking with an overall r.m.s.d. of the multiple structure alignment being 1.08 Å and variation is most pronounced at the C-terminal end. The model of the '297-308 belt' was proposed to determine substrate specificity in group 3 members. The belts conformation supports (i) the formation of a hydrophobic cluster which interact with the methyl group of the N-acetyl moiety, (ii) a correct positioning of the Asn195, and (iii) orients the substrate so the GlcNAc moiety will form hydrogen bonds with Ser143 and Ser144. Due to this belt and the resulting hydrogen bond network, the group 3 members have a distinct conformation at this region whereas the conformation of group 1 and group 2 enzymes is very similar. As a result, the S306Y mutation, which allows a switch from group 2 to group 1, forms steric clashes between the group 3 epimerases and their substrates which results in the observed loss of activity [32]. The importance and flexibility of the hydrogen bond network was evidenced by multiple single mutants as well as their combined double, triple and quadruple mutants [31].

3.3.4.2. UDP-pentose and UDP-uronic acid 4-epimerases

As not only glucose and galactose, acetylated or not, are found in cell wall polysaccharides, there are more UDP-sugars and likewise more specificities for UDP-sugar 4-epimerases. As such, epimerases harboring activity on UDP-xylose (UDP-Xyl) or UDP-galacturonic acid (UDP-GalUA) have been found in plants and bacteria [83, 137]. One UDP-glucose 4-epimerase from *Pisum sativum* (PsGalE) and two from *Arabidopsis thaliana* (AtGalE) have been found to possess activity on both UDP-Glc/UDP-Gal as well as the corresponding UDP-pentoses, UDP-Xyl and UDP-L-arabinose (UDP-L-Ara). Despite the significant similarities with human GalE, the PsGalE failed to act on N-acetylated UDP-sugars, most likely due to the presence of a large 'gatekeeper' residue (Val309) (*cfr.* Tyr in group 1 vs. Ser in group 2). However, the same 'gatekeeper' allows a very low UDP-GlcNAc 4-epimerase activity in barley GalE [137]. Similarly, the UDP-galacturonic acid 4-epimerase from *Klebsiella pneumonia* shows a broad *in vitro* substrate specificity. It is capable of interconverting UDP-Glc/UDP-Gal and the acetylated forms thereof, although with a much lower activity than was observed for its wild-type substrate, UDP-GlcUA/UDP-GalUA [83]. However, no intense research has been conducted on the determinants for specificity and promiscuity of the UDP-substrates.

In summary, the first proposed model, that of the hexagonal box, provides a good representation for determinants and interchangeability of the substrate specificity of group 1 and group 2 epimerases. A bulky residue (Tyr) in the green wall most likely indicates activity only on non-acetylated substrates, whereas epimerases with a small residue (Ser, Cys) are likely to possess activity on acetylated substrates. However, as also group 3 members contain this smaller residue here, a biochemical analysis is still necessary for confirmation to which group the enzyme belongs. A reason for this is also that interchangeability between group 3 and the other two groups is much smaller due to the presence of the '297-308 belt' and the importance of the hydrogen bond network that are of great importance for activity and substrate specificity. Further experiments are still needed for a proper understanding of the residues that determine substrate specificity in order to validate and strengthen the current models, not only in specificity on substrate with or without N-acetylation, but also on UDP-pentoses and uronic acid forms of UDP-sugars.

4 Enzyme Engineering

4.1. Introduction

Micro-organisms and their enzymes have been applied throughout the ages. The oldest and best known examples are beer brewing and production of cheese and wine. As enzymes are able to catalyze all kinds of chemical reactions, they become more and more available for industrial and household applications [197]. Nowadays, enzymes are available in different industries such as agriculture, bioenergy, biopharma, food and beverages, household care and many others [188].

Enzymes apply the same basic principles as conventional chemical catalysts. However, due to their more intimate interaction with the substrate, they have an exquisite control of enantio- and regioselectivity as well as stereochemistry [59, 197]. Other advantages of the use of enzymes in chemistry are the high activity and specificity rates (up to 10^{17}-fold accelerations) [59], the sustainability of biocatalysts [197] and they often require less energy and raw material consumption as well as generate less waste and toxic side-products [4]. Biocatalysis can thus be considered as a form of green chemistry [197]. However, not many enzymes meet the needs of industrial or academic chemists as they are more interested in converting unnatural substrates, and this in many cases under conditions that are not favored by the enzymes, like high temperatures and in the presence of solvents [38]. A good industrial biocatalyst should combine high activity and specificity with great enantioselectivity and, on top of that, should be able to deal with the harsh conditions of industrial processes [38]. The major drawbacks of many natural enzymes are narrow substrate specificity, small pH and temperature optima, and low pH and temperature stability [197]. Another problem can be that no enzyme is available in nature to catalyze the desired reaction. Here, enzyme engineering can be applied to overcome these issues. Furthermore, chemical modification of enzymes can also improve their stability, for example via immobilization and/or cross-linking [43, 45, 46, 69, 234, 235].

4.2. Enzyme engineering and its types

Enzyme engineering is the technique of modifying an enzyme's structure, and thus its function and characteristics, and/or its catalytic activity towards natural and/or new substrates. Successful enzyme engineering will result in an improved biocatalyst harboring the desired characteristics. Several characteristics of the biocatalyst can be improved, such as catalytic efficiency (or reaction speed), reaction conditions (improved pH or temperature optimum and stability), as well as the enzyme’s selectivity [39]. The latter can be subdivided in chemoselectivity, regioselectivity and enantioselectivity referring to substrate promiscuity and acceptance, and discrimination of stereoisomers which is of great importance in pharmaceuticals (*cfr*. softenon), respectively.

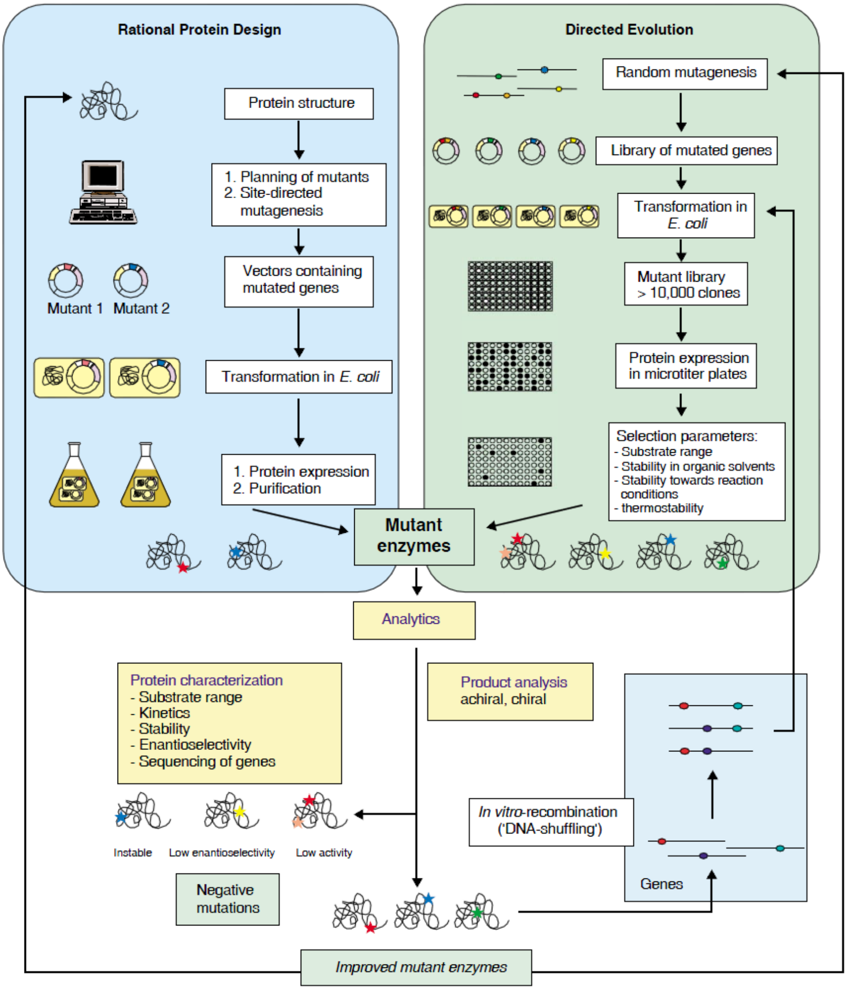

Rational Protein Design
Protein structure
1. Planning of mutants
2. Site-directed mutagenesis
Vectors containing mutated genes
Mutant 1
Mutant 2
Transformation in *E. coli*
1. Protein expression
2. Purification
Directed Evolution
Random mutagenesis
Library of mutated genes
Transformation in *E. coli*
Mutant library
> 10,000 clones
Protein expression in microtiter plates
Selection parameters:
- Substrate range
- Stability in organic solvents
- Stability towards reaction conditions
- thermostability
Mutant enzymes
Analytics
Protein characterization
- Substrate range
- Kinetics
- Stability
- Enantioselectivity
- Sequencing of genes
Product analysis
achiral, chiral
Instable
Low enantioselectivity
Low activity
Negative mutations
In vitro-recombination
('DNA-shuffling')
Genes
Improved mutant enzymes

Figure II.21 Comparison of rational protein design and directed evolution. In rational design, mutants are created based on available information, such as protein structure. Each variant is expressed, purified and analyzed for desired properties. On the other hand, mutagenesis in directed evolution is achieved at random. Protein libraries are then usually screened in microtiter plates using previously set selection parameters or in a selection system (not given in this figure). Newly gained information based on desired and negative mutants can be used in new round of mutagenesis and found mutants can be used as new templates for further improvements (Adopted from [40])

All types of enzyme engineering follow the same pattern, namely a mutagenesis step followed by an identification step [39]. In the first step, changes are made at DNA-levels in order to generate genetic diversity. Secondly, the library is tested to identify whether improved variants were created. These two distinct steps are typically applied in an iterative fashion in order to fully optimize the biocatalyst to perform the desired function according to the specified criteria [12]. In many ways, the identification step is the most challenging, as without an efficient identification strategy, the ability to identify improved variants is fairly low.

Several techniques have been developed to introduce mutations in the corresponding gene. In general, the type of enzyme engineering is determined by the type of mutagenesis. When mutations are made at random, the engineering is called Directed evolution, whereas Rational design deals with directed mutagenesis mainly based on structural and functional insights (Figure II.21). These different enzyme engineering techniques will be discussed here, including some possible mutagenesis techniques. Afterwards, the difference between selection and screening will be clarified.

4.2.1. Rational design

When rational design is applied, only specific mutations are made through site directed mutagenesis (SDM), or the replacement of one residue by another. It is mostly applied to identify important residues, for example for catalysis or specificity, or to conform a proposed hypothesis, like a proposed mechanism. In order to use rational design to improve a biocatalyst via enzyme engineering, a very good understanding of the enzyme is needed at different aspects, such as structure-function relationship of residues, preferably a 3D-structure, the reaction mechanism and/or determinants for substrate specificity.

4.2.2. Directed evolution

In contrast to rational design, directed evolution does not require any knowledge on mechanism or structure as mutations are incorporated at random. It can in fact be seen as a mimic of natural or Darwinian evolution in a lab scale timeframe [197]. Generation of mutant libraries can be achieved in two manners, namely asexual or sexual. In asexual evolution, mutagenesis is applied on preferentially one parental gene encoding the template protein. The most applied system is error-prone PCR (epPCR), which exploits the fact that DNA polymerases occasionally make errors during replication [208]. To increase error rate, unbalanced dNTP concentrations can be used or $MnCl_2$ can be added to the PCR mixtures [39]. However, despite the ease of these methods, they resulted in a mutational bias favoring replacements of As and Ts. Today, commercial kits are available containing a mix of two mutator polymerases in order to overcome the bias (e.g. GeneMorph II Random Mutagenesis Kits (Stratagene)). On the other hand, sexual evolution starts with a pool of homologue parental genes, which are then ‘shuffled’ to obtain chimers of two or more parent genes, which are then tested for improved variants [185, 241, 242].

A major disadvantage of directed evolution is that very large libraries are created, which results in a lot of screening effort, and only shows a limited success rate. Therefore, nowadays researchers try to minimize mutant libraries on the one hand and try to increase the quality of libraries on the other hand [167]. It thus combines rational design and directed evolution and is called semi-rational design.

4.2.3. Semi-rational design

In semi-rational design, structural information is combined with random elements to achieve high quality libraries [167]. Residues that were selected based on structural insights or that were found by random approaches can be saturated by site saturation mutagenesis (SSM). In contrast to SDM, in SSM the targeted residue is replaced by all other possible residues using a degenerated primer. Saturations can be applied in an iterative fashion to accumulate beneficial effects of single mutants (iterative saturation mutagenesis, ISM) or multiple residues can be targeted simultaneously to obtain synergetic effects (CASTing, combinatorial active site saturation test) [38]. Other optimizations of mutant libraries are achieved by using different degenerates codons, like NDT (12 codons, 12 amino acids) instead of NNS or NNK (32 codons, 20 amino acids) [212] and the use of computational approaches [11, 48, 127].

4.2.4. *De novo* protein design

Besides the engineering of existing proteins, research is also performed to create totally novel proteins by *de novo* protein design. Here, biocatalysts are not engineered starting from a natural protein but are created with the aid of the computational protein design methodology. The RosettaMath algorithm was developed by David Baker’s group and is used to create idealized active sites, for the introduction of additional functional groups, and to search for appropriate protein backbones [216]. The algorithm has been applied in the creation of an enzyme catalyzing the Kemp elimination reaction, for which no natural enzyme exists, [221] and the *de novo* design of a retro-aldolase [119]. Despite the enormous value and potential of *de novo* design, the *de* *novo* created enzymes are not ideal biocatalysts and further optimization is needed both for the created catalysts as for the computational method. Catalyst optimization is mainly done by general enzyme engineering techniques such as directed evolution and (semi-)rational design and analysis of these improved variants will generate additional knowledge to improve the design methodology [196].

4.2.5. Selection vs. Screening

After the first step of creating mutant libraries, there is the need to identify improved enzyme variants, which can be achieved in two different ways. These methods should be specific and sensitive enough to identify positive mutants. The first method is called selection and applies the principle of Darwinian evolution or ‘survival of the fittest’ [196]. In case enzyme function can provide an advantage to the host, a library containing different variants of the enzyme is transformed into the selection host, which is then tested for growth on a minimal medium. If the mutant expresses an active enzyme variant, the strain will able to convert the provided carbon source into useful building blocks and energy to grow. As a result hereof, cells expressing good enzyme variants will overgrow the rest of the culture (inactive and slow mutants). Repetitive inoculation in fresh medium will further enrich the culture with the best mutant/variant which can then be isolated and tested for further confirmation.

For the second method, called screening, each variant is tested individually for the desired reaction. Therefore, the library is transformed into an expression host and single colonies are transferred from solid medium to liquid medium. A major disadvantage of screening is that each mutant has to be tested separately, even those that are inactive, which is typically 50-80 % of a library [196]. A simple visual screening uses a colorimetric assay, via chemical or enzymatic reactions, to identify product formation. High throughput screening (HTS) refers to automated screening in miniaturized vessels in order to process large amounts of mutans [196].

4.3. Chemical modification of enzymes

In addition to enzyme engineering, enzyme characteristics can also be modified by chemically modifications. As such, immobilization and/or cross-linking can enhance solvent and thermostability [45, 46], while imprinted cross-linked enzyme aggregates or iCLEAs have been shown to improve activity on natural and non-natural substrates [43, 69].

5 Conclusion

It is clear that the economical value of tagatose as well as other rare sugars is big. However, it still has a bigger potential when one compares the current tagatose market with its competitors. In 2006, about 4.6 and 11.9 million metric tons of refined sugars and corn-derived sweeteners were used in foods and beverages in the US market, whereas only 205,000 metric tons of other caloric sweeteners, like sugar alcohols, were used [17]. Still, global sales of non-sugar sweeteners in 2010 was worth $9.2 billion [18]. Nevertheless, seen the big difference between the current use of traditional and high caloric sweeteners and the low-caloric alternatives as well as the increasing number of obese and diabetic persons (163 and 16 million Americans in 2003, respectively [189]), the potential market and with it the economical value of low-caloric sweeteners (e.g. tagatose) is huge. Despite this big potential, the current limiting factor to increase tagatose production worldwide is the lack of enough substrate (galactose) for the current production routes. A way to circumvent this problem is by tapping another substrate pool to make tagatose from cheap raw materials. Cheap and widely available alternatives would be fructose or glucose; however, no (bio)catalysts are available to directly convert these inexpensive and abundant sugars into tagatose. Multiple step approaches are available but due to the large number of steps, the overall conversion is low and several expensive purification steps are needed. Ideally, a C4-epimerase would be needed to convert fructose into tagatose or glucose into galactose, which can then be used as substrate in the current production routes. No such enzymes occur naturally, but enzymes that catalyze C4-epimerazations on similar substrates are available. These enzyme can be subjected to enzyme engineering in order to change their substrate specificity and make them active on fructose/tagatose or glucose/galactose. Here, the engineering of two C4-epimerases with respect to their potential in fructose to tagatose conversion will be described. These two enzymes are L-ribulose-5-phosphate 4-epimerase and UDP-hexose 4-epimerase and they will need adaptation to accept free monosaccharides instead of phosphorylated or nucleotide-activated sugars, respectively. This thesis comprises the cloning and expression of the L-ribulose-5-phosphate 4-epimerase (chapter III), the development of a selection system for tagatose 4-epimerase activity and the application thereof (chapter IV and chapter V, respectively) as well as the development and use of a screening assay for the same purpose (chapter VI). Finally, it deals with the cloning, expression and characterization of the UDP-hexose 4-epimerase (chapter VII) and its mutational analysis (chapter VIII).

III. CLONING AND EXPRESSION OF L-RIBULOSE-5-PHOSPHATE 4-EPIMERASE

1 Introduction

To date, biochemical production routes are available to produce all simple rare sugars [20, 91], nevertheless, a lot of these biochemical steps are not very efficient or they have to start from substrates that are either too expensive or too scarce or in the worst case both of the above. Therefore, it would be interesting to create new biocatalysts that are able to convert widely available and cheap substrates into rare, but valuable products. Within the area of sugars, only seven monosaccharides (Glc, Gal, Man, Fru, Xyl, Rib and L-Ara) were considered to be present in significant amounts, whereas the other twenty hexoses and nine pentoses were described as rare sugars. Despite the fact that many isomerases and oxidoreductases are available for rare sugar production, only one epimerase (ketohexose 3-epimerase) is available that can be used therefore with economical viability [91], whereas some other epimerases that were found more recently, show some potential in rare sugar synthesis (see literature review: chapter II sections 1.2.3 and 1.3) [20].

We chose to engineer an existing 4-epimerase towards a tagatose 4-epimerase, therefore, the gene of the template epimerase – our starting point, L-ribulose-5-phosphate 4-epimerase from *Geobacillus thermodenitrificans* – had to be cloned into an expression vector. In this chapter, an inducible expression vector was made that is fully complementary with the 4 constitutive expression vectors (pCXhPxx) available at the Laboratory for Industrial Biotechnology and Biocatalysis (InBio) at Ghent University [1]. Subsequently, the cloning of this L-ribulose-5-phosphate 4-epimerase gene into these pIXPtrc and pCXhPxx expression vectors is described. Then, recombinant expression of the gene is compared between all 5 expression vectors in two *E. coli* strains, namely XL10 Gold and BL21 (DE3).

Furthermore, the L-ribulokinase gene (AraB) from *E. coli* K-12 MG1655 was cloned in this pIXPtrc vector and heterologously expressed in *E. coli* BL21 (DE3) cells. Next, the recombinant L-ribulokinase was purified and used in the production of L-ribulose-5-phosphate. After purification, the L-ribulose-5-phosphate is then used in the wild-type activity test of the recombinant *G. thermodenitrificans* epimerase.

2 Material & Methods

2.1. Bacterial strains, growth conditions, plasmids and chemicals

Geobacillus thermodenitrificans LMG 17532T was obtained from the Belgian Co-ordinated Collections of Micro-organisms (BCCM) and was grown at 55 °C on Tryptone Soya medium containing 15 g/L pancreatic digest of casein, 5 g/L papaic digest of soybean meal, 5 g/L NaCl, pH 7.3 and supplemented with 15 g/L of agar for solid media. *Escherichia coli* K-12 MG1655 and the other *E. coli* strains were cultured at 37 °C and 200 rpm in Luria broth (LB) containing 5 g/L yeast extract, 10 g/L tryptone, 10 g/L NaCl at pH 7.0 or LB agar plates (LB plus 10 g/L agar). For cultures of cells containing a plasmid, media was supplemented with 0.1 g/L ampicillin.

E. coli XL10 Gold (Stratagene) and *E. coli* BL21 (DE3) (Stratagene) were used as a cloning strain and for recombinant expression, respectively, of both the L-ribulose-5-phosphate 4-epimerase gene (*araD*) from *G. thermodenitrificans* and the L-ribulokinase (AraB) from *E. coli* K-12 MG1655. *E. coli* cells were made chemically competent using the Inoue method for heat shock transformation [224].

The pGEM-T vector system and pTrc99A expression vector were obtained from Promega and the Netherlands Culture Collection of Bacteria (NCCB), respectively, while the four pCXhPxx expression plasmids were developed earlier at the Laboratory for Industrial Biotechnology and Biocatalysis (InBio) at Ghent University [1, 65].

Clone Manager Professional 8 (Sci Ed Software, USA) was used for designing primers and sequence alignments. Standard protocols were used for routine recombinant DNA methodology and nucleotide sequencing was performed by LGC Genomics (former Agowa Genomics, Germany). Primers were synthesized by Sigma, the High-Fidelity PCR Master mix used for gene amplification was obtained from Roche, while restriction enzymes and T4 DNA ligase were purchased from New England Biolabs. Kits for gel extractions, PCR purifications and plasmid isolation were obtained from Qiagen. All chemicals were obtained from Sigma Aldrich unless otherwise stated.

2.2. Construction of the expression vectors

2.2.1. Construction of pIXPtrc

The sequence encoding for the His-tag and linker in the pCXhP22 [1] was amplified by high fidelity PCR using a 50 µL PCR mixture containing 50 % (v/v) of High Fidelity PCR Master mix (Roche), 40 pmol of forward and reverse primer (Table III.1), 80 ng pCXhP22 as template and a standard PCR protocol as given in Table III.2. After amplification, the PCR fragment was treated with 10 U of NcoI and SacI each to prepare the fragment for cloning into a likewise treated empty pTrc99A vector (Pharmacia Biotech Inc.). DpnI (10 U) restriction enzyme was added to the restriction mixture as well in order to remove the PCR template. Ligation of the cut and purified fragments (QiaQuick PCR purification kit, Qiagen) was achieved by using a 3/1 insert/vector ratio and 3 Weiss units of T4 DNA ligase at 22 °C for 1h. The ligation mixture was then transformed into *E. coli* XL10 Gold cells, the plasmid was purified and checked for correctness by sequencing.

Table III.1 Primers for the amplification of the His-tag from pCXhP22, *araD* and *araB* from *G. thermodenitrificans* and *E. coli* K-12 MG1655, respectively. Restriction sites are underlined in the primer sequence

Gene product	Primer	Sequence (5'→3')	Restriction site
His-tag	Fwd	CCATGGGGGTTCTCATCATCATC	NcoI
	Rev	GAGCTCTCCCATATGGTCGAC	SacI
araD_Gt	Fwd	CTTAAGATGCTTGAGGAGCTGAAACGG	AflII
	Rev	ACTAGTTTATTGTCCATAGTAAGCGTTTAC	SpeI
araB_Ec	Fwd	GGATCCCATGGCGATTGCAATTGGCCTC	BamHI
	Rev	CTGCAGTTATAGAGTCGCAACGGCCTGG	PstI

2.2.2. Cloning the genes into pIXPtrc and pCXhPxx (Figure III.1)

At first, genomic DNA (gDNA) was extracted from an overnight grown culture using the 'GenElute Bacterial Genomic DNA kit' from Sigma. The protocol for Gram-positive bacteria was used for the gDNA from *G. thermodenitrificans,* while the Gram-negative protocol was applied to obtain the *E. coli* K-12 MG1655 gDNA. The *araD* and *araB* genes from *G. thermodenitrificans* and *E. coli* K-12 MG1655, respectively, were amplified from gDNA by high fidelity PCR using the primer pair listed in Table III.1. The PCR mixtures contained 25 µL of High Fidelity PCR Master mix (Roche), 40 pmol of forward and reverse primer and 8 % (v/v) DMSO in a total volume of 50 µl. The PCR cycling conditions are shown in Table III.2.

Table III.2 Standard PCR cycling conditions

Cycles	Time	Temperature
1 x	4 min	94 °C
35 x	45 s	94 °C
	45 s	55 °C
	1 min/kb + 30 s	72 °C
1 x	7 min	72 °C

After amplification, the resulting fragment was purified and cloned into the pGEM-T vector according to supplier's protocol. The ligation mixture was then transformed into competent *E. coli* XL10 Gold cells by a heat shock [224] and plated out on Luria Broth (LB) agar. Single colonies were then grown in liquid LB for plasmid extraction to both check the gene sequence and to propagate the gene fragments. Both pGEM-T plasmids were then cleaved using 10 U of the appropriate restriction enzymes (AflII/SpeI and BamHI/PstI for *araD* and *araB*, respectively) at 37 °C for 1 h. Subsequently, both genes were ligated in a similarly treated pIXhPtrc vector again using a 3/1 insert/vector ratio and 3 Weiss units of T4 DNA ligase at 22 °C for 1h. The resulting plasmids are called pIXhPtrc-araD and pIXhPtrc-araB, respectively, and were checked by sequencing.

The *araD* gene was also cloned into the set of constitutive expression vectors that was created earlier at the Laboratory for Industrial Biotechnology and Biocatalysis (InBio) at Ghent University [1]. Therefore, the *araD* gene fragment was cleaved from the pIXhPtrc-araD vector using 10 U of both AflII and SpeI restriction enzymes, the 4 empty constitutive expression vectors were treated in the same way, and the fragments were connected using the same ligation condition as mentioned above. The constitutive expression vectors obtained after plasmid extraction were named pCXhPxx-araD.

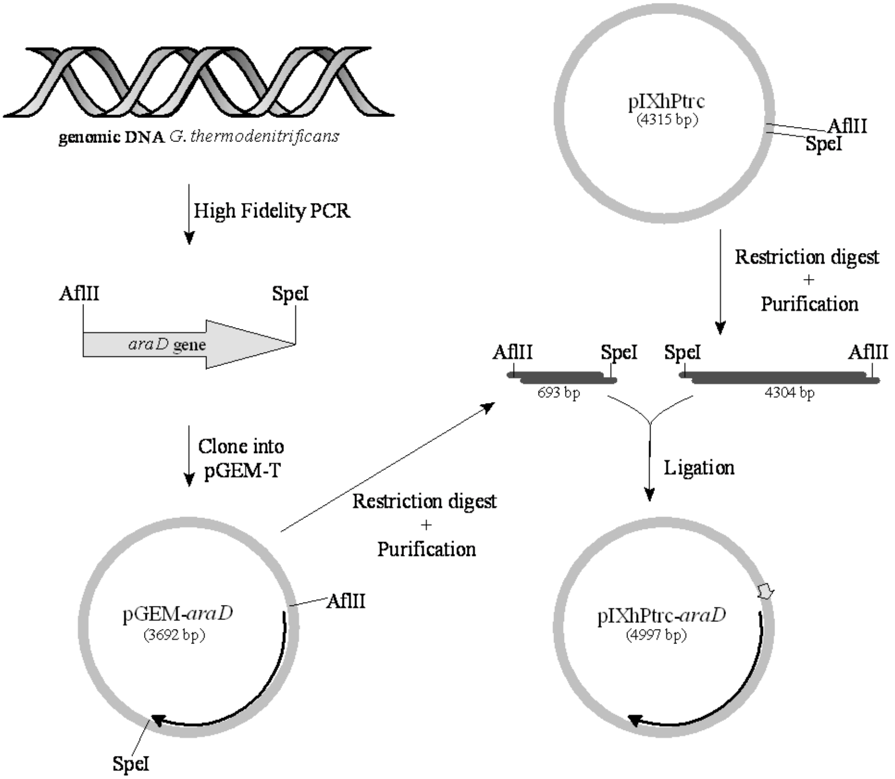

genomic DNA *G. thermodenitrificans*
High Fidelity PCR
AflII
SpeI
araD gene
Clone into pGEM-T
pGEM-*araD*
(3692 bp)
AflII
SpeI
Restriction digest + Purification
pIXhPtrc
(4315 bp)
AflII
SpeI
Restriction digest + Purification
AflII
SpeI
693 bp
SpeI
AflII
4304 bp
Ligation
pIXhPtrc-*araD*
(4997 bp)

Figure III.1: Cloning strategy for the construction of the inducible expression vector pIXhPtrc-araD

2.3. (Optimization of) Recombinant enzyme expression

The constructed expression plasmids were transformed into *E. coli* XL10 Gold and BL21 (DE3) cells. A single colony was picked from each transformation and inoculated into 5 mL LB medium supplemented with ampicillin. After overnight growth, the culture was used as inoculum (2 % v/v) for an enzyme expression in 50 ml LB medium with ampicillin. The cultures were grown until the beginning of the exponential phase ($OD_{600} \approx 0.6$) and subsequently 0.1 mM IPTG was added to the culture with the inducible expression plasmid to induce heterologous expression. Subsequently, all cultures were grown for another 6 h to express heterologous protein.

For the inducible expression system, the ideal amount of IPTG was checked by inducing recombinant protein with different amounts of IPTG. Fresh expression culture was grown until $OD_{600} \approx 0.6$ and then 0.1 mM, 0.5 mM or 1 mM IPTG was added to the culture and growth was continued for another 6 h. Afterwards, 1 ml samples of the cultures were centrifuged for 5 min at 14000 rpm and the obtained pellets were frozen at -20 °C.

Crude enzyme extracts were obtained by chemo-enzymatic lysis of the frozen cell pellets containing the heterologously expressed enzyme. Therefore, frozen pellets were thawed on ice, dissolved in 200 µL of lysis solution (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 4 mM MgCl2, 50 mM NaCl, 0.1 mM PMSF and 1 mg/ml lysozyme) and lysis was achieved by 5 min incubation at room temperature. Insoluble fraction and cell debris were pelleted by centrifugation in a table top centrifuge at full speed and supernatant was transferred to a new tube. Both soluble and insoluble fractions were checked on SDS-PAGE and Western blotting with anti-His_6 antibodies.

The expression of the L-Ru-5-P 4-epimerase (AraD) was optimized by comparing expression from the inducible and constitutive vectors. Furthermore, expression was analyzed in *E. coli* BL21 and XL10 Gold, and finally IPTG concentration for induced expression was sampled with 3 different IPTG concentrations, namely 0.1 – 0.5 – 1.0 mM IPTG. For L-ribulokinase (AraB) only expression in BL21 on the inducible expression vector was checked with 3 different IPTG concentrations (0.1 – 0.5 – 1.0 mM IPTG).

2.4. L-Ribulose-5-phosphate production and purification

L-ribulose-5-phosphate (L-Ru-5-P) was needed for the wild-type activity tests of the L-ribulose-5-phosphate epimerase, therefore, the gene coding for the ribulokinase from *E. coli* K-12 MG1655 was cloned into the pIXhPtrc expression vector and brought to expression in *E. coli* BL21 (DE3) as described above. Due to the presence of an N-terminal His-tag, the ribulokinase could easily be purified from the obtained cell extract by Ni-NTA chromatography according to the supplier’s protocol (ThermoScientific), with wash and elution buffers containing 20 mM and 500 mM imidazole, respectively. After elution, the buffer from the purified ribulokinase was exchanged to 50 mM glycyl glycine buffer (gly-gly) using a Centricon YM-30 (Millipore). The protein concentration of the ribulokinase mixture was analyzed with the Pierce® BCA Protein assay kit (ThermoScientific).

The purified ribulokinase was then applied in the phosphorylation reaction of L-ribulose, similarly to the production described earlier by Anderson [7]. The L-ribulose had been made earlier by dehydrogenation of ribitol by De Muynck [66] at the Laboratory for Industrial Biotechnology and Biocatalysis (InBio) at Ghent University. For this reaction, equimolar amounts of L-ribulose, $MgCl_2$ and ATP (30 mM each) were dissolved together with 50 mM KF, 5 mM reduced L-glutathion and 12.5 μM EDTA in 40 ml of 50 mM gly-gly buffer pH 7.5. The phosphorylation reaction was initiated by addition of large amounts His-tag purified ribulokinase and performed at room temperature. The pH of the reaction was followed and maintained at pH 7.5 by titration of NaOH as the phosphorylation of L-ribulose results in the acidification of the solution.

Purification of the produced L-Ru-5-P was accomplished as described earlier by Anderson [7] and the potassium salt was obtained after barium precipitation with K_2SO_4 [238]. Summarized, after the reaction had stopped, acetic acid is added to a final concentration of 0.2 M. Subsequently, the precipitate is filtered and the pH of the supernatant solutions set to pH 6.7 with NaOH. The L-Ru-5-P is purified and separated from the nucleotides by addition of barium acetate to a concentration of 0.1 M, after which the precipitate is filtered off. The supernatant is acidified to pH 2.0 with HCl, and absorbance is checked at $OD_{260\ nm}$. Contaminants were removed by repetitive treatment with activated charcoal until $OD_{260\ nm}$ had reached a minimum. The activated charcoal was removed by filtration and in each round smaller amounts of activated charcoal were used. After absorbance had reached its minimum, the solution was adjusted to pH 6.7 with NaOH and 4 volumes of ethanol were added. This 80 % (v/v) ethanol mixture was then chilled overnight in ice water to precipitate the L-Ru-5-P. The precipitate is collected by centrifugation and washed twice with fresh and chilled 80 % (v/v) ethanol. The solution was poured off and the precipitate dried using a vacuum to remove traces of water and ethanol (SpeedVac, Savant). The resulting barium salt of L-Ru-5-P is quickly washed with water twice, before addition of equimolar amounts of K_2SO_4 to re-dissolve the L-Ru-5-P as potassium salt and precipitate the barium as sulfate. The $BaSO_4$ precipitate was removed by filtration and a powder of L-Ru-5-P potassium salt was obtained by vacuum drying using a SpeedVac.

2.5. HPLC analysis of L-Ru-5-P

Two High Performance Liquid Chromatography (HPLC) methods were used to analyze the purity of the produced L-Ru-5-P. The first method applies an isocratic flow of 5 mM H_2S0_4 over an Aminex HPX-87H column (Bio-Rad Laboratories) at 30 °C at a flow rate of 0.6 ml/min. The HPLC system (Varian Prostar) was composed of a Varian Prostar 410 auto sampler, a Varian Prostar 230 pump, a Varian Prostar 320 UV/Vis detector and a Varian Prostar 350 RI differential refractive index. The second method uses a similar HPLC system attached to a ELSD 2000ES detector (Alltech). The Hypercarb column (Thermoscientific) was incubated at 30 °C and the gradient used is given in Table III.3. Solvents A, B and C are milliQ water, 100 % acetonitril and 15 % formic acid in milliQ water, respectively.

Table III.3 Elution gradient profile of the Hypercarb sugar method

Time (min)	A (%)	B (%)	C (%)
0	96	4	0
5	92	8	0
7	75	25	0
10	75	25	0
20	0	25	75
22	50	50	0
27	50	50	0
30	96	4	0
50	96	4	0

2.6. Enzyme activity assays: *in vivo* and *in vitro*

Activity of the recombinant L-Ru-5-P 4-epimerase is checked both *in vivo* and *in vitro*. For *in vivo* activity, the selection strain SelTag2 or SelTag3 (see chapter IV) was transformed with the pIXhPtrc-araD plasmid and its growth in L-arabinose minimal medium (L-AMM) was compared with that of the selection strain without the expression plasmid in the same medium. More details about the composition of the minimal medium and growth conditions are found in the Material & Methods section of chapter IV.

For *in vitro* activity assays, the epimerization reaction was tested from L-Ru-5-P towards D-Xu-5-P as was earlier described [63, 225]. The reaction mixture contained 25 mM gly-gly buffer pH 7.6, 5 mM D-ribose-5-P, 0.1 mM thiamine pyrophosphate, 0.3 mM of $MgCl_2$, 0.15 mM NADH, 0.25 U/ml transketolase, 50 U/ml triosephosphate isomerase, 5 U/ml α-glycerol phosphate dehydrogenase, 0.25 mM L-Ru-5-P (potassium salt). The reaction mixture was incubated at 37 °C for 10 min to remove impurities in the L-Ru-5-P by reaction with the coupled enzymes. The reaction was started by addition of 2 µg of purified L-Ru-5-P 4-epimerase to the reaction mixture (finally 200 µl) and followed by measuring the drop in absorbance at 340 nm.

Affinity for L-Ru-5-P (K_m) was determined using 12 different L-Ru-5-P concentrations from 0.078 mM to 10 mM and results were plotted in a Lineweaver-Burk linearization for calculation. Thermostability of the enzyme was determined by measuring the half-life (t_{50}) at 37 °C. Effect of Zn^{2+}, Co^{2+} and Mg^{2+} concentration were determined by addition of different concentrations of these metal ions (0-0.5 mM) to EDTA treated enzyme. EDTA treatment performed as described earlier [74] and afterwards EDTA was removed by a wash step with 25 mM gly-gly buffer pH 7.6 over a Amicon® Ultra 30K Centricons (Millipore).

3 Results and Discussion

3.1. Construction of the expression vectors

3.1.1. Preparation of the pIXhPtrc vector

High fidelity amplification of the sequence containing the coding sequence (CDS) of the His-tag and the multiple closing site of the pCXhP22 plasmid was successful, resulting in the amplification of a 156 bp fragment, as expected. The PCR fragment was then cloned into the pTrc99A vector as described in the Material & Methods section and the pIXhPtrc was obtained by plasmid extraction. Plasmid sequencing revealed that the cloning was successful and that no errors were introduced. Thus, an inducible expression vector – pIXhPtrc – complementing the 4 constitutive expression vectors available at the Laboratory for Industrial Biotechnology and Biocatalysis (InBio) at Ghent University [1] – pCXhPxx – has been created successfully.

3.1.2. pIXPtrc-AraD and pCXhPxx-AraD

The high fidelity PCRs on the genomic DNA of *G. thermodenitrificans* and *E. coli* K-12 MG1655 resulted in fragments of the expected length, namely 699 bp for *araD* and 1714 bp *araB*. Both amplified fragments were successfully cloned in the pGEM-T vector for propagation of the gene fragment. After plasmid extraction, both the *araD* gene fragment and ribulokinase gene were cloned in the inducible pIXhPtrc expression vector. Plasmid sequencing revealed that the ribulokinase gene had been correctly inserted into the inducible expression vector; however, the plasmid containing the L-Ru-5-P 4-epimerase gene (*araD*) was found to possess a 5 nucleotide insert between the His-tag and the gene. As this 'cttaa' insert results in a frame shift in the gene, it had to be removed first to be able to express the L-Ru-5-P 4-epimerase correctly.

For the deletion of the 'cttaa' insert in the pIXhPtrc-AraD plasmid, a high fidelity and whole plasmid PCR was performed using the QuickChange® XL Site-Directed Mutagenesis Kit (Stratagene) according to supplier's protocol and the deletion primers and cycling conditions mentioned in Table III.4 and Table III.5, respectively. After DpnI treatment and transformation, extracted plasmid was sequenced and this revealed that the insert had been removed and thus now the pIXhPtrc-AraD is correct for further work.

Table III.4 Primers for the deletion of the 'cttaa' insert in pIXhPtrc-AraD (* indicates the position of the deletion)

Primer	Sequence (5'→3')
Deletion_Fwd	GACGATAAGGATCCAACC*CTTAAGATGCTTGAGGAG
Deletion_Rev	CTCCTCAAGCATCTTAAG*GGTTGGATCCTTATCGTC

Table III.5 Whole plasmid PCR cycling conditions for deletion of the 'cttaa' insert

Cycles	Time	Temperature
1 x	1 min	95 °C
18 x	50 s	95 °C
	50 s	60 °C
	1 min/kb + 30 s	68 °C
1 x	7 min	68 °C

Subsequently, the AraD gene fragment was cloned into the 4 constitutive pCXhPxx expression vectors as mentioned in the Material & Methods section above. All 4 constitutive expression vectors were found to contain no errors and thus cloning was successful.

3.2. Recombinant L-ribulose-5-phosphate 4-epimerase expression

For high-throughput screening purpose, L-Ru-5-P 4-epimerase expression was checked under different conditions in order to determine ideal conditions for high enzyme concentration and activity. Since no substrate (L-Ru-5-P) was yet available to test the epimerization reaction, it was chosen to compare enzyme expression on SDS-PAGE. Both inducible and constitutive expression resulted in an L-Ru-5-P 4-epimerase containing a N-terminal His-tag and linker with an overall calculated molecular mass of 29.4 kDa/subunit.

At first, expression in *E. coli* BL21 and XL10 Gold cells using the pIXhPtrc-AraD vector was compared. Under the same conditions, higher recombinant protein concentrations were observed when expressing the enzyme in *E. coli* BL21. This observation is expected since *E. coli* BL21 is lacking the protease genes *lon* and *ompT*, therefore resulting in higher protein concentrations.

Secondly, expression using the inducible and the constitutive vectors was compared against each other. Two times of induction were tested for the inducible expression vector, namely induction at the beginning of the exponential phase ($OD_{600} \approx 0.6$) and induction from the start on. SDS-PAGE analysis and Western blotting revealed that recombinant enzyme concentration was correlated with the promoter strength of the constitutive vector used. The lowest L-Ru-5-P 4-epimerase concentration was observed with the promoter 78, intermediate concentrations were seen with both P34 and P22, while the use of promoter 14 yielded the highest enzyme concentrations for the constitutive expression vectors. When comparing the inducible Trc promoter with the 4 constitutive promoters, induced enzyme expression is at least as high for the highest constitutive expression (P14). Furthermore, no big difference is observed for the time of induction of the pIXhPtrc-AraD vector. IPTG can be added both at the start of growth as well as in the beginning of the exponential phase, these conditions will be used for screening and enzyme production, respectively.

Finally, the concentration of IPTG to induce the *LacIq – Trc* operator and promoter system of pIXhPtrc-AraD was sampled. Therefore, the influence of low (0.1 mM), intermediate (0.5 mM) and high (1.0 mM) IPTG concentrations were determined. Since no big improvements were observed using higher IPTG concentrations, low IPTG concentration (0.1 mM) will be used. Under all of the above conditions, the L-Ru-5-P 4-epimerase was present in the soluble enzyme fraction, which is also a good indication since increased portions in the insoluble fraction would have led to loss of active enzyme in the form of aggregates or inclusion bodies.

To summarize, recombinant expression will be achieved in *E. coli* BL21 cells using the inducible pIXhPtrc-AraD vector and induction by addition of 0.1 mM IPTG. For larger enzyme productions, cells will first be grown to $OD_{600} \approx 0.6$, before inducing them with IPTG, whereas for (high throughput) screening, IPTG will be added from the start on to minimize pipetting steps and workload.

3.3. Ribulokinase production, L-ribulose-5-phosphate production and purification

To easily obtain pure recombinant ribulokinase, it was expressed containing an N-terminal His-tag, resulting in a ribulokinase with a calculated molecular mass of 64.7 kDa/subunit (Figure III.2). In order to obtain high levels of recombinant ribulokinase, the production thereof was optimized in *E. coli* BL21 cells using the pIXhPtrc-AraB vector. The cells were grown until exponential phase was reached and then induced with 3 different IPTG concentrations (0.1 mM, 0.5 mM and 1.0 mM IPTG). Levels of recombinant protein production were analyzed on SDS-PAGE. Comparison of soluble and insoluble fraction clearly indicate that the recombinant ribulokinase is only present in the soluble phase. Since no great increase in recombinant protein content was observed by increasing IPTG concentrations, a larger scale ribulokinase production was achieved using 0.1 mM IPTG.

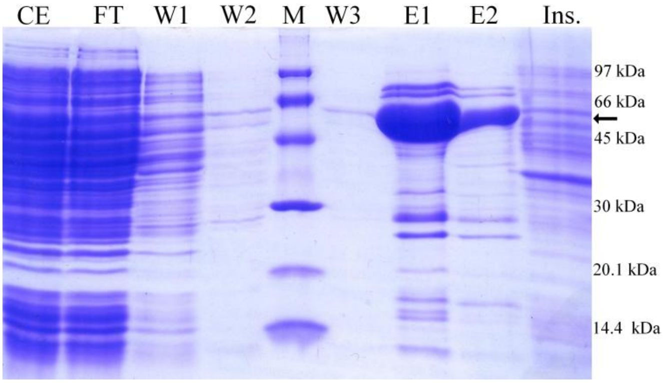


Figure III.2 SDS-PAGE of the recombinant expressed ribulokinase and its His-tag purification. CE, cell extract; FT, flow through; W1-W3, Wash fractions 1-3; E1-2, Eluent fractions 1-2; Ins., Insoluble; M, Marker.

Ribulokinase activity was first checked by performing the reaction as mentioned in the Material & Methods section on a small scale (10 ml). Since pH of the reaction mixture after addition of the purified enzyme gradually decreased, it indicates that the purified ribulokinase is active and could be used in a large scale production of L-Ru-5-P.

Large scale production of L-Ru-5-P was then achieved according to the protocol described earlier. After the acidification of the solution had stopped, extra ribulokinase was added to the reaction mixture to ensure maximal conversion. No further acidification was observed, meaning that the enzyme had not lost its activity but that the reaction had completed. The produced L-Ru-5-P was subsequently purified as described earlier. However, the barium salt of L-Ru-5-P is obtained by this purification and was found to contain remainders of L-ribulose by HPLC analysis. Since the barium salt of L-Ru-5-P has low solubility in water, the L-ribulose could easily be washed away with water. After 2 wash steps, the barium was removed from the L-Ru-5-P by barium precipitation with K_2SO_4, leaving behind the potassium salt of L-Ru-5-P in solution. Barium sulfate could easily be removed by filtration and the L-Ru-5-P was obtained from the filtrate by vacuum drying. HPLC analysis showed that the produced L-Ru-5-P was now pure (Figure III.3). After purification and vacuum drying around 238 mg of L-Ru-5-P was obtained, which means that an overall production and recovery of 54 % (238 mg/438 mg) was achieved.

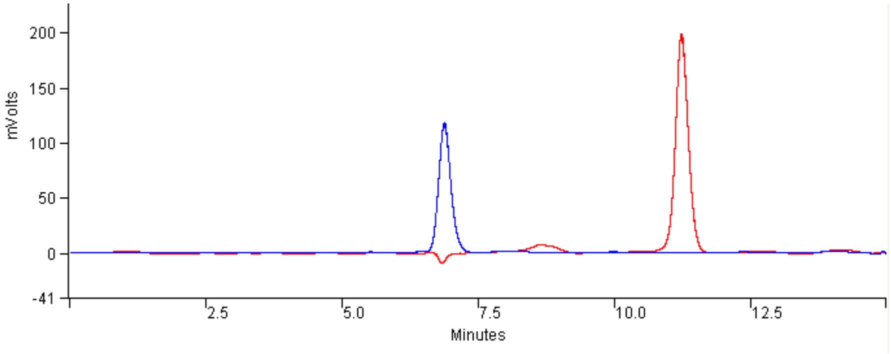


Figure III.3 HPLC analysis (Aminex) of L-Ru-5-P K-salt (blue) and L-ribulose (red)

3.4. Enzyme activity assays: *in vivo* and *in vitro*

In vivo activity test is described in the next chapter at section 3.3. In vitro activity assay determined that the His-tagged purified epimerase had a specific activity of 1 U/mg. With a Lineweaver-Burk linearization, it was calculated to have a K_m of 1.4 mM for L-Ru-5-P (Figure III.4). This is almost three times as high as the K_m of the *E. coli* L-Ru-5-P 4-epimerase under similar conditions (22°C, pH 7.5, activated by 0.1 mM $MgCl_2$) [148].

The enzyme was found to have a half-life of slightly more than 1 day (24,3 h) at 37 °C (Figure III.5). Effect of metal ions revealed that the highest activity was found with 0.3 mM of Mg^{2+}, while equivalent amounts of Co^{2+} activated the enzyme almost equally. Effect of the desired cofactor, Zn^{2+}, could not be determined since zinc has an inhibitory effect on the coupled enzyme assay. Nevertheless, it is likely to have a similar activating effect as was found for the *Aerobacter* epimerase [74].

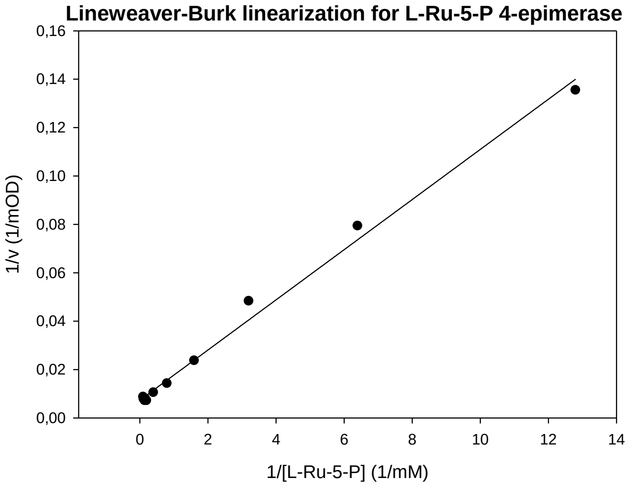


Figure III.4 Lineweaver-Burk linearization for L-Ru-5-P 4-epimerase

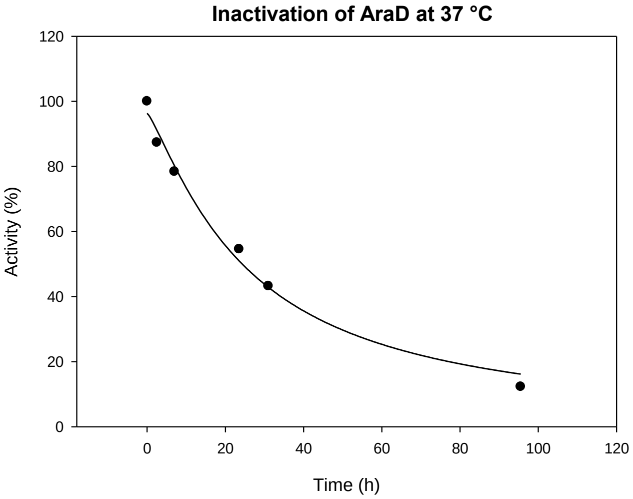


Figure III.5 Inactivation of the L-Ru-5-P 4-epimerase at 37 °C

4 Conclusion

The His-tag, linker and multiple cloning site of the pCXhP22 plasmid for constitutive expression [1] was cloned successfully into the pTrc99A vector, resulting in the inducible expression vector pIXhPtrc. This pIXhPtrc vector is now fully complementary with the 4 constitutive expression vectors that were already available at the Laboratory for Industrial Biotechnology and Biocatalysis (InBio) at Ghent University. Now, a set of 5 plasmids is available in which cloning can be achieved using the same restriction enzymes. Depending on the restriction enzymes used, one can add a short His-tag (NheI) or a longer His-tag with linker and enterokinase cleavage site (AflII) to the recombinant protein or express it as native protein (NcoI). Both His-tags are N-terminal and facilitate purification of the expressed enzyme by means of Ni-NTA chromatography.

The L-Ru-5-P 4-epimerase gene from *G. thermodenitrificans* was successfully cloned into all 5 expression vectors. Heterologous expression was found to give highest recombinant enzyme concentrations in *E. coli* BL21 cells. Since expression using the inducible pIXhPtrc-AraD resulted in the highest L-Ru-5-P 4-epimerase concentrations, this vector will be used from here on and induction will be achieved by addition of 0.1 mM IPTG. Since no big difference was found between immediate induction and induction at $OD_{600} \approx 0.6$, induction time will depend on the type of experiment. In case of high throughput screening (chapter VI), IPTG will be added from the beginning to minimize pipetting steps and chance on contamination, while for larger recombinant enzyme production cells will first be grown to the start of the exponential phase before induction.

The ribulokinase gene from *E. coli* K-12 MG1655 was also successfully cloned into the inducible expression vector, resulting in the pIXhPtrc-AraB plasmid. Subsequently, the plasmid was transformed into *E. coli* BL21 cells for recombinant expression, resulting in a 64.7 kDa His-tagged ribulokinase. This was then purified and applied in the production of L-Ru-5-P, starting from L-ribulose and ATP. After purification, around 238 mg of L-Ru-5-P (as potassium salt) could be obtained with an overall production and recovery rate of 54 %. The produced L-Ru-5-P was then used for the characterization of the epimerase. The enzyme has a K_m value of 1.4 mM, a half-life at 37 °C of slightly more than 1 day and, as expected, it can be activated by metal ions.

IV. DEVELOPMENT OF A SELECTION SYSTEM FOR TAGATOSE 4-EPIMERASE ACTIVITY

1 Introduction

After correctly cloning the *araD* gene into the expression vectors, the next hurdle could be taken, namely the development of a selection strain which can be used to look for improved variants after mutagenesis has been applied on the *G. thermodenitrificans* epimerase gene. In a selection system, Darwinian evolution is applied to identify the best variant in a mutant library. As such, if an active (or improved) enzyme is expressed in the selection strain, this mutant will be able to consume the given carbon source (better) and have an advantage over all other mutants in the selection culture. Through this advantage the fittest in the mixture will finally overgrow the others, a classic example of Darwin's theory of 'survival of the fittest' [60, 150].

In order to obtain a selection strain suitable for Darwinian evolution experiments, an *E. coli* strain had to be able to grow on the product of the tagatose 4-epimerase reaction and unable to grow on the substrate, nevertheless, it should be able to absorb the substrate. Since *E. coli* strains are natural fructose consumers and unable to use tagatose as carbon source, it was chosen to select for the reverse reaction. Uptake of tagatose in *E. coli* K-12 MG1655 is primarily regulated via the methylgalactoside transport system, mglABC [132]. As such, tagatose will be given to the cells as a sole carbon source and if the mutated epimerase is able to convert it into fructose, the cell will be able to grow and enrich itself in the culture medium. By means of inoculating enriched cultures into fresh medium, the best mutant will finally overgrow all inactive and less active variants and the plasmid can be obtained to identify this mutation.

To prevent potential loss of the *G. thermodenitrificans* L-Ru-5-P 4-epimerase gene by recombination with homologue genes present in the *E. coli* genome, these homologous genes were removed by the method of gene disruption [61]. Three homologous genes are present in the E. coli genome, namely *araD, ulaF* and *sgbE*, all encoding a L-Ru-5-P 4-epimerase in different pathways. *AraD, ulaF* and *sgbE* are located in the L-arabinose operon for L-arabinose degradation*, ulaA*p operon for anaerobic L-ascorbate utilization and *yiaK*p operon for metabolism of L-lyxose and L-xylulose, respectively [126, 128]. It was chosen to create the selection strains using different *E. coli* strains as starting points. *E. coli* CGSC#10993 was chosen as this strain already contains 2 of the 3 knock-outs (KO) needed, namely full KO of the *araBAD* operon, thus loss of the *araD* gene, and the replacement of the *ulaF* gene by the kanamycin resistance marker. Furthermore this strain contained a lacZ deletion, lacks the operon for rhamnose degradation and some other characteristics [15]. On the other hand, *E. coli* K-12 MG1655, of which the full genome is known, is seen as a standard *E. coli* [36]. In the CGSC#10993, only the sgbE gene has to be removed, whereas all three L-Ru-5-P 4-epimerase genes (*araD*, *ulaF* and *sgbE*) have to be removed sequentially. Selection strains are named SelTag for **Sel**ection strain for **Tag**atose 4-epimerase activity.

2 Material & Methods

2.1. Bacterial strains, plasmids, growth conditions and chemicals

Two strains were used as starting point for the creation of a selection strain for detection of tagatose 4-epimerase activity, namely *E. coli* CGSC#10993, which was obtained from the Coli Genetic Stock Center (CGSC) of the Yale University (USA), and *E. coli* K-12 MG1655 [36]. They were routinely grown at 37 °C and 200 rpm on LB medium (5 g/L yeast extract, 10 g/L tryptone, 10 g/L NaCl, pH 7.0) or LB agar plates (LB plus 10 g/L agar), supplemented with 0.1 g/L ampicillin or carbenicillin. The inducible pIXhPtrc-AraD and constitutive pCXhPxx-AraD expression plasmids encoding the L-Ru-5-P 4-epimerase from *G. thermodenitrificans* were created earlier (chapter III).

L-Arabinose, ampicillin, carbenicillin, chloramphenicol and kanamycin were purchased from Sigma-Aldrich. Taq polymerase and restriction enzyme DpnI were purchased from New England Biolabs and High-Fidelity DNA polymerase was obtained from Roche Applied Science. PCR primers were designed with Clone Manager Professional8 (Sci Ed Software, USA) and synthesized by Sigma. The Clone Manager Professional 8 software was also used for sequence alignments.

The plasmids used for the creation of the selection strain by knock-out strategy, pKD3, pKD4, pKD46 and pCP20, were a kind gift from prof. R. Cunin (Laboratory for Microbiology and Genetics, Free University of Brussels, Belgium). pKD3 (Cm^R) and pKD4 (Km^R) contain an FRT-flanked antibiotic resistance gene for chloramphenicol [67] and kanamycin (kan) resistance, respectively. The two helper plasmids, pKD46 (Amp^R) and pCP20 (Amp^R), carry the genes for the λ Red recombinase and the FLP recombinase, respectively. Both plasmids are low copy number plasmids and are easily curable through their temperature sensitive replicons. Therefore, strains containing the pKD46 or pCP20 helper plasmids were grown at 30 °C, unless the strains had to be cured from these plasmids, in which case they were grown at 42 °C.

2.2. Construction of the SelTag-strains by gene disruption

Creation of gene knock-outs is accomplished by the method of gene disruption described earlier by Datsenko & Wanner [61]. In summary, a first helper plasmid containing the λ Red recombinase (pKD46) is brought in the cells and the recombinase is induced by growth on L-arabinose. Subsequently, an antibiotic resistance gene flanked by two FRT sites (flippase recognition target) and DNA fragments homologous to the flanking of the target gene is brought in the cells. The recombinase will exchange the gene with the antibiotic marker (2.2.1) and correct integration can be tested by a simple PCR (2.2.2). After correct integration has been achieved, antibiotic marker can be removed by the flippase brought to expression by the second helper plasmid (pCP20) (2.2.3). If more genes have to be knocked out, the protocol can be repeated from the start.

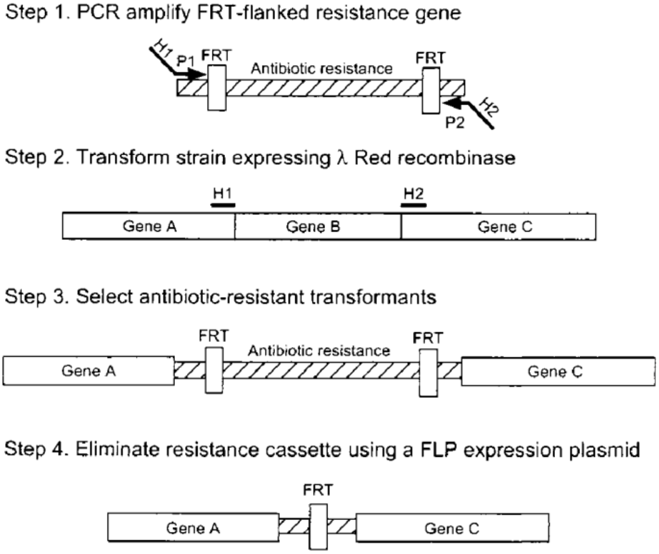

Step 1. PCR amplify FRT-flanked resistance gene
H1
P1
FRT
Antibiotic resistance
FRT
H2
P2
Step 2. Transform strain expressing λ Red recombinase
H1
H2
Gene A
Gene B
Gene C
Step 3. Select antibiotic-resistant transformants
FRT
Antibiotic resistance
FRT
Gene A
Gene C
Step 4. Eliminate resistance cassette using a FLP expression plasmid
FRT
Gene A
Gene C

Figure IV.1 Schematic overview of the gene disruption strategy. H1 and H2 refer to the homology extensions, P1 and P2 refer to priming sites. Taken from [61]

2.2.1. Gene disruption

At first, pKD46 was transformed into both *E. coli* CGSC#10993 and *E. coli* K-12 MG1655. The pKD46 plasmid carries the L-arabinose induced araC-P_{BAD} promoter as well as the λ Red recombinase genes: *γ*, *β* and *exo*, whose products are called Gam, Bet and Exo, respectively. Gam inhibits the host RecBCD exonuclease V, in order to allow Bet and Exo to access DNA ends to promote recombination with the genomic DNA (gDNA). Transformants were then grown in 5 mL LB cultures supplemented with 0.1 g/l carbenicillin and 1 mM L-arabinose at 30 °C until $OD_{600} \approx 0.6\text{-}1.0$. Subsequently, these cells were harvested by centrifugation and made electrocompetent by washing them 3 times with ice-cold distilled water and finally concentrating them up to 200 times.

Secondly, PCR products that had been created by amplifying pKD3 and pKD4 with appropriate primers for recombination (see Table IV.1) and the PCR mixture and cycling conditions mentioned in Table IV.2 and Table IV.3, were digested with DpnI, gel purified and finally transformed into the pKD46 containing cells. Electroporation was achieved by using a GenePulser II (Bio-Rad), 2 mm cuvettes (Eurogentec), 50 µl of freshly made electrocompetent cells and 3 µl of PCR product. After electroporation, 1 ml of LB medium was added to the cells and incubated for 1h at 37 °C, before spreading them on LB-agar containing the appropriate antibiotics. This was done to select for cells with built-in Cm^R or Km^R resistance gene in their gDNA. The resulting transformants were then tested with colony-PCR to ensure that the resistance gene had been introduced at the correct position in the gDNA. Meanwhile, cells were plated on fresh LB-agar and grown at 42 °C to cure them from the pKD46 plasmid.

2.2.2. Verification by colony PCR

Verification whether the resistance gene had been introduced at the correct position was performed by inoculating a fresh colony into a PCR mix containing 2.5 nmol of each control primer (Table IV.1), 0.2 mM dNTP mix, appropriate buffer and 0.5 U of Taq polymerase, with milliQ water up to 25 µl. The PCR program consisted of the following cycles: initial denaturation at 95 ° C for 4 min, 30 cycles of 95 °C (30 s), 58 °C (30 s) and 72 °C (1 min/kb) for amplification and final elongation for 7 min at 72 °C. PCR products were verified by electrophoresis using 1 % agarose gels.

Table IV.1 Primers used for gene disruption and control thereof. Primer sequences in capital letters are homologous (H1 and H2) to the flanking regions of the target genes, to create a place for recombination. The sequences in small letters are the parts homologous to the pKD3 and pKD4 helper plasmids, P1 and P2.

Primer	Sequence (5’→3’)
Fw-o-sgbE-P1	GGAGATTATTCAGGCGCGGCGTTGGATTGAAGCGCGTATGCAGGAGGCTGGATTTATgtgtaggctggagctgcttc
Rv-o-sgbE-P2	GCAAGGAACATATCAATTCGTAGTGCCGGGGCGATGAAGCCCCGGCGTGAGGGAcatatgaatatcctccttag
Fw-sgbE-out	GCCGGTTCTTTCCTGATTGAGATG
Rv-sgbE-out	TGGAAGCGGCGTTACAG
Fw-o-AraD-P1	GCTTGAGTATAGCCTGGTTTCGTTTGATTGGCTGTGGTTTTATACAGTCAgtgtaggctggagctgcttc
Rv-o-AraD-P2	CGATTTTGTAGGCCGGATAAGCAAAGCGCATCCGGCACGAAGGAGTCAACcatatgaatatcctccttag
Fw-AraD-out *	GCCAGAAGGAGACTTCTGTCCCTTG
Rv-AraD-out *	AAAGCGCATCCGGCATTCAACGCCTG
Fw-AraD-out2	TAACCTGCAACGGCCCGTTGTC
Rv-AraD-out3	GTATTACGGGTTTCGTCGCTAAG
Fw-o-UlaF-P1	ATTGGGTGAAAGCGCGCATGGCGAAAGCGGGCATGGTGGAGGCGGCATAAgtgtaggctggagctgcttc
Rv-o-UlaF-P2	GGAATTAGACCAGTTATCTCCCGAGGAAGGAAATTTCCGCAGCGCGTGTTcatatgaatatcctccttag
Fw-UlaF-out	TTGAGATGTGGAGCGAAACG
Rv-UlaF-out	TCGCTAGCACCAGGTATAAC

* Since these primers were found to bind on multiple sites, a new pair of control primers was used (Fw-AraD-out2 and Rv-AraD-out3)

Table IV.2 PCR mixture for the creation of linear DNA for gene disruption

Component	Amount
Pfu Ultra DNA polymerase	2.5 U
10x Pfu Ultra HF AD buffer	1 x
dNTPs	0.2 mM
Primers	0.4 µM each
Template: pKD3 or pDK4	50-100 ng
Total	50 µl

Table IV.3 PCR cycling conditions for linear DNA creation for gene disruption

Cycles	Time	Temperature
1 x	1 min	95 °C
30 x	30 s	95 °C
	30 s	60 °C
	1 min/kb + 30 s	72 °C
1 x	7 min	72 °C

2.2.3. Elimination of the antibiotic resistance gene

The pCP20 plasmid shows a thermal induction of the FLP synthesis and therefore the created Cm^R and Km^R mutants were made electrocompetent and transformed with pCP20, and plated at LB-agar with ampicillin and placed at 30 °C. After growth, transformants were tested for loss of the FRT-flanked resistance gene with colony PCR as mentioned above. Finally cells were cured from the helper plasmid pCP20 by growth at 42 °C and then tested again with colony PCR and for carbenicillin sensitivity to confirm loss of pCP20. At last, verification fragments were amplified using a high fidelity polymerase, purified and sequenced as to confirm that the correct gene had been disrupted. Sequencing was performed by VIB Genetic Service Facility (Belgium) or Agowa (Germany). After the cells are cured from the pCP20 plasmid, cells are made competent again and transformed with pKD46 to prepare the disruption of another gene.

2.2.4. Speeding up the gene disruption protocol

Since three genes had to be removed from the *E. coli* K-12 MG1655 strain, we tried to speed up the gene disruption protocol. Optimization of the protocol was achieved by eliminating some of the labor intensive and time consuming steps. Instead of curing the strain from the pKD46 plasmid (growth at 42 °C) after a first gene was replaced with one of the antibiotic markers, the cells were grown at 30 °C and made competent to transform with a second linear DNA made with the second antibiotic marker. After good disruption of the second gene, the cells were cured from pKD46 and transformed with pCP20 to simultaneously lose both antibiotic markers. As such, the work and time of two curing steps as well as two rounds of competent making and transformation could be saved.

2.3. Growth tests on different minimal media

2.3.1. Medium composition

The minimal medium (MM) used in the growth test of the selection strains is an M9 based medium. Apart from 1X M9 salts (6 g/l Na_2HPO_4, 3 g/l $KHPO_4$, 1 g/l NH_4Cl, 0.5 g/l NaCl), the medium contains 20 mg/l proline, 1 mM thiamine-HCl, 0.1 mM $CaCl_2$, 1 mM $MgS0_4$, 18 µM $FeCl_2$ and 6.7 µM $ZnCl_2$. To this, different sugars were added as a sole carbon source at a concentration of 20 g/l. As such, the created selection strains (SelTagN°) were tested for growth on L-Arabinose MM (L-AMM), Fructose MM (FMM) and Tagatose MM (TMM). IPTG and the appropriate antibiotics were added at concentration mentioned earlier for induction and prevention of plasmid loss and contamination, respectively.

2.3.2. Wash and growth conditions

For the growth tests of the SelTag strains, the strains with or without pIXhPtrc-AraD, were first grown in LB medium at 37 °C and 200 rpm. Cells were then harvested by centrifugation and washed twice with phosphate buffered saline (PBS; 8 g/l NaCl, 0.2 g/l KCl, 1.44 g/l Na_2HPO_4, 0.27 g/l KH_2PO_4, pH 7.4) and then inoculated (2 % inoculum) into 3 ml of fresh minimal medium supplemented with IPTG and the necessary antibiotics. Cells were grown in a 24-deepwell plate at 37 °C and 220 rpm. Cell growth was followed by measuring cell density ($OD_{600\ nm}$).

3 Results and Discussion

3.1. Construction of the SelTag-strains by gene disruption

Starting from 2 different *E. coli* strains, 3 selection strains were created that can be used for the selection of mutants bearing tagatose 4-epimerase activity. Starting from *E. coli* CGSC#10993, the SelTag1 strain was created by knocking out the *sgbE* gene and removal of the kanamycin resistance marker that was inserted at the position of the *ulaF* gene. Initially, this had happened by accident, but it triggered the idea of a possible speeding up of the gene disruption of the selection strain based on *E. coli* K-12 MG1655 (see paragraph 2.2.4 and 3.2). Using the same method of gene disruption, all three genes encoding an L-Ru-5-P 4-epimerase in *E. coli* K-12 MG1655 (*araD*, *ulaF* and *sgbE*) had successfully been knocked-out (Figure IV.2). This mutant was named SelTag2. Due to the presence of repetitive parts in between the *araA* and *araD* genes in the *E. coli* K-12 MG1655 genome, a slightly bigger part of DNA (153 bp) had been knocked-out. Nonetheless, the removal of this extra DNA part should not be a problem for the selection strain as the removed part does not code for a gene and also the flanking *araA* gene stays intact.

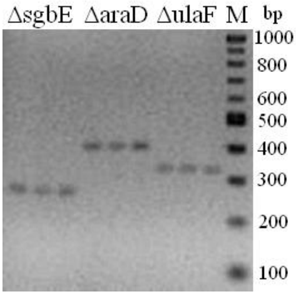


Figure IV.2 DNA fragments obtained after colony PCR on SelTag2 confirm successful gene disruption. Fragments obtained are around 264 bp, 414 bp and 326 bp, corresponding to the theoretical lenght of successful gene knock-outs.

Due to problems of contamination in the selection cultures (see chapter V), it was later chosen to use the last ‘intermediate step’ of the creation of the SelTag2 strain as this strain still contained the kanamycin resistance gene at the position of the *ulaF* gene (*ulaF*::kan in SelTag3 instead of Δ*ulaF* in SelTag2). Due to the presence of an additional antibiotic, chances of contamination will decrease. By analogy with the 2 other strains, this one was called SelTag3.

3.2. Speeding up the gene disruption protocol

Since, both antibiotic resistance genes were deleted by the flippase at the same time, the optimizing of the gene disruption protocol was successful. In this manner, the work load and time of two curing steps (first pKD46, then pCP20) as well as two rounds of competent making and transformation (first pCP20, then pKD46) could be saved. This had resulted in speeding up the protocol for 2 gene disruptions by 2-3 days. Since this simultaneous removal of both antibiotic markers was successful, it would not be unlikely that more than 2 antibiotic markers could be removed at the same time as well. However, it would be best to have extra antibiotic markers to be able to select for disruption of a third, fourth, … gene, for instance a gentamycin resistance gene. With 3 antibiotic resistance genes, disruption of 3 genes following this optimized protocol would save twice the 2-3 days of work, or a full week. Since there is a small risk that an earlier introduced marker is replaced by a new one, it is important to plate the cells on medium containing all antibiotics. On the other hand, disruption of the correct gene is checked by colony PCR using gene specific primers as well, nevertheless, it can be good to check previous genes by colony PCR too.

3.3. Growth tests on different minimal media

3.3.1. SelTag1 = *E. coli* CGSC#10993 ΔsgbE ΔulaF

As expected, the SelTag1 strain was found to be unable to grow on tagatose (TMM) and on L-arabinose (L-AMM) since *E. coli* are naturally no tagatose consumers and, on the other hand, the full araBAD operon for L-arabinose degradation has been deleted, respectively. Surprisingly, the SelTag1 strain was also found to be unable to grow on fructose (FMM). The mother strain of this SelTag1, namely *E. coli* CGSC#10993, was also found to be unable to grow on fructose, which explains why the SelTag1 does not grow on fructose (Table IV.4). However, considering the mutations mentioned on the CGSC website for this strain, it would be expected that the *E. coli* CGSC#10993 strain would be able to grow on fructose. Nonetheless, the strain is most likely to harbor extra mutations than only those mentioned on the website, and seen the inability to grow on fructose, these mutations are likely to be found in fructose degradation pathways. Since SelTag1 is unable to grow on fructose, it cannot be used as a selection strain for the detection of tagatose 4-epimerase activity. Therefore, the creation of a new selection strain was started from the standard *E. coli* K-12 MG1655, which is known to grow on fructose.

Table IV.4 Growth of the different strains on Fructose minimal medium (MM), Tagatose MM and L-Arabinose MM. Ability and inability to grow are indicated with + and – , respectively. The expectation of growth is given between brackets

Strain	L-AMM	FMM	TMM
E. coli CGSC#10993	– (–)	– (+)	– (–)
SelTag1	– (–)	– (+)	– (–)
E. coli K-12 MG1655	+ (+)	+ (+)	– (–)
SelTag2	– (–)	+ (+)	– (–)
SelTag2+pIXhPtrc-AraD	+ (+)	+ (+)	– (–)
SelTag2+pCXPhxx-AraD	+ (+)	+ (+)	– (–)
SelTag3	– (–)	+ (+)	– (–)
SelTag3+pIXhPtrc-AraD	+ (+)	+ (+)	– (–)
SelTag3+pCXPhxx-AraD	+ (+)	+ (+)	– (–)

3.3.2. SelTag2 (3) = *E. coli* K-12 MG1655 ΔaraD ΔsgbE Δula (ulaF::kan)

The growth tests of the SelTag2 strain revealed that growth on L-arabinose had been made impossible, which is normal since one of the genes in the L-arabinose degradation pathway has been knocked out (Table IV.4). On the other hand, supplementation of this L-Ru-5-P 4-epimerase through the recombinant expression of the *G. thermodenitrificans* L-Ru-5-P 4-epimerase of the inducible or constitutive expression plasmids could undo the loss of the native L-Ru-5-P 4-epimerase. On one hand, this confirms that the removal of the extra 153 bp DNA part is not deleterious for the nearby L-arabinose isomerase (*araA*) gene (cfr. 3.1). More importantly, this also confirms that the *G. thermodenitrificans* L-Ru-5-P 4-epimerase is expressed in an active form, which is a prime requirement for enzyme engineering. Furthermore, SelTag2 was able to grow on fructose and unable to grow on tagatose, making it useful as selection strain for the detection of enzymes harboring tagatose 4-epimerase activity.

The same results concerning growth are expected for the SelTag3 strain, since it only differs from the SelTag2 strain at the position of the *ulaF* gene. It still contains the kanamycin marker here, instead of the deletion of this gene (*ulaF*::kan instead of Δ*ulaF*). As such, the same growth curves were observed for the SelTag3 strain (*E. coli* K-12 MG1655 ΔaraD ΔsgbE ulaF::kan) compared to SelTag2 (*E. coli* K-12 MG1655 ΔaraD ΔsgbE ΔulaF). Like SelTag2, this SelTag3 strain can thus also be used for selection towards a tagatose 4-epimerase. The advantage of SelTag3 over SelTag2 lies in the presence of a kanamycin resistance marker which can reduce the risk of contamination due to the presence of extra antibiotic in the medium.

4 Conclusion

In this chapter, 3 strains have been created starting from the *E. coli* CGSC#10993 or from the standard *E. coli* K-12 MG1655 by knocking out all genes encoding homologues L-Ru-5-P 4-epimerases. Removal of these genes will reduce the risk of loss of the *G. thermodenitrificans* L-Ru-5-P 4-epimerase gene present on the pIXhPtrc-AraD plasmid as a result of recombination. The first selection strain, SelTag1, was based on *E. coli* CGSC#10993 and is not useful since it is unable to grow on fructose as a sole carbon source.

Nevertheless, the selection strains based on the *E. coli* K-12 MG1655, namely SelTag2 and SelTag3, are found to be useful in a selection system towards a tagatose 4-epimerase. since they are able to grow on fructose, unable to do so on tagatose and express the *G. thermodenitrificans* L-Ru-5-P 4-epimerase in an active form. SelTag3 has the advantage of harboring an extra antibiotic resistance marker which reduces the chance of contamination in the case of slow growth of the selection strain.

While making these different selection strains, the protocol of gene disruption was optimized speaking of time frame and workload. By simultaneous removal of 2 antibiotic resistance markers, 2 to 3 days of work could be spared, hereby speeding up the protocol. This also opens the door for further speeding up the protocol for simultaneous removal of more than 2 genes. However, it is necessary to use new resistance markers to be able to select for gene disruption, a possible new resistance marker can be that for gentamicin.

V. MUTAGENESIS OF L-RU-5-P 4-EPIMERASE AND SELECTION TOWARDS IMPROVED MUTANTS

1 Introduction

After the development of two useful selection strains for tagatose to fructose conversion, the focus was shifted towards the enzyme, namely an L-Ru-5-P 4-epimerase. Amongst the different strategies available for enzyme engineering are random and (semi-)rational mutagenesis. Random mutagenesis can always be applied since no or little information about the target enzyme is needed [40]. Mutations are made at random via different strategies, such as error prone PCR and shuffling [208, 241]. In the research project, random mutagenesis was achieved by applying error-prone PCR on the full gene as well as parts of the gene.

In recent decades, the field of informatics and mathematics has become an important field in biochemical or biotechnological sciences, since several mathematical models have been translated in useful bio-informatica tools. Among these programs are visualizing programs such as the PyMOL Molecular Graphics System [70] but also molecular modeling programs like YASARA [139]. With the use of this latter program, a homology model can be made for new enzymes based on available crystal structures of homologous enzymes. Since multiple structures of other L-Ru-5-P 4-epimerases and related homologous enzymes are available [78, 140, 141, 166], a homology model for the *G. thermodenitrificans* L-Ru-5-P 4-epimerase can be made using the molecular modeling program YASARA [139] and the mentioned crystal structures. The creation of this homology model allows us to have a look at the enzyme's active site, including the residues important for substrate binding. Furthermore, previous (engineering) studies of other L-Ru-5-P 4-epimerases and the homologous aldolases provide us with a lot of information as well. This information will be used together with the created homology model to more rationally identify residues that might be of importance to change the substrate specificity of the epimerase from phosphorylated pentoses (L-Ru-5-P/D-Xu-5-P) towards free hexoses (tagatose/fructose) (Figure I.1). Target residues will then be mutated towards all possible amino acids. This approach is called semi-rational as residues are chosen based on rational information but mutagenesis is not fully rational since saturation is applied.

In this chapter, the creation of a homology model for the *G. thermodenitrificans* L-Ru-5-P 4-epimerase is described, along with the random and site saturation mutagenesis of the enzyme. After mutagenesis, mutant libraries were transformed into the previously created selection strains, SelTag2 or SelTag3 (chapter IV), and grown in tagatose minimal medium in order to identify improved enzyme variants.

2 Material & Methods

2.1. Bacterial strains, plasmids, growth conditions and chemicals

Both the inducible expression plasmid pIXhPtrc-AraD containing the *G. thermodenitrificans* L-Ru-5-P 4-epimerase gene and the selection strains SelTag2 and SelTag3 have been created earlier (chapter III and chapter IV, respectively). Standard growth conditions and media were used as mentioned in the previous chapters. Primers were ordered at Sigma-Aldrich, the same counts for chemical unless otherwise stated. Aligments of the L-Ru-5-P 4-epimerase with other enzymes and sequence analysis of this 4-epimerase were performed using the online available ClustalW2 alignement program [144] as well as the NCBI BLAST tool [6].

Minimal medium (TMM) composition was used as described in chapter IV section 2.3.1, however, with 250 mM or 500 mM of tagatose. Experiments were also performed in tagatose synthetic medium (TSM). This synthetic medium contained 2 g/l NH_4Cl, 5 g/l $(NH_4)_2SO_4$, 2.993 g/l KH_2PO_4, 7.315 g/l K_2HPO_4, 8.372 g/l MOPS, 0.5 g/l NaCl, 0.5 g/l $MgSO_4.7H_2O$, 250 mM or 500 mM of tagatose (Nutrilab), 1 ml/l vitamin solution, 100 µl/l molybdate solution, and 1 ml/l selenium solution. The medium was set to a pH of 7 with 1 M KOH. Vitamin solution consisted of 3.6 g/l $FeCl_2.4H_2O$, 5 g/l $CaCl_2.2H_2O$, 1.3 g/l $MnCl_2.2H_2O$, 0.38 g/l $CuCl_2.2H_2O$, 0.5 g/l $CoCl_2.6H_2O$, 0.94 g/l $ZnCl_2$, 0.0311 g/l H_3BO_4, 0.4 g/l $Na_2EDTA.2H_2O$ and 1.01 g/l thiamine-HCl. The molybdate solution contained 0.967 g/l $Na_2MoO_4.2H_2O$. The selenium solution contained 42 g/l SeO_2.

2.2. Homology model of the L-Ru-5-P 4-epimerase

To determine the residues and/or regions of the 4-epimerase that might be important to change the substrate specificity of the epimerase, alignment studies with other members of the *araD*-like aldolase/epimerase family were performed, important residues were identified in previously published literature, and finally important residues were chosen by looking at the substrate binding domain of the *G. thermodenitrificans* L-Ru-5-P 4-epimerase. Therefore, a homology model of the *G. thermodenitrificans* L-Ru-5-P 4-epimerase was made using the molecular modeling program YASARA [139] based on the crystal structures of L-Ru-5-P 4-

epimerase (PDB: 1JDI), L-rhamnulose-1-phosphate aldolase (PDB: 1GT7) and L-fuculose-1-phosphate aldolase (PDB: 1DZU), all from *E. coli*, and with default settings. The resulting homology model was saved as a pbd-file, which was then used to look at the substrate binding domain, and more specifically the phosphate binding site, using the YASARA program [139] and the PyMOL Molecular Graphics System [70].

The substrates L-Ru-5-P and D-Xu-5-P were docked into the created homology model using the docking tool of the molecular modeling program YASARA [139]. Furthermore, the substrate analogue phosphoglycolohydroxamic acid (PGH) that was found in the crystal structure of *E. coli* L-fuculose-1-phosphate aldolase (PDB: 4FUA) was brought in the homology model by superposition of both structures with the help of the MUSTANG alignment tool [134] in YASARA [139]

2.3. Random mutagenesis: error prone PCR

For random mutagenesis, the GeneMorph II EZClone Domain Mutagenesis kit (Stratagene) and primers mentioned in Table V.1 were used, applying a mutation frequency of 0-10 mutations/kb, which would result in 0-7 mutations on the *araD* gene.

Table V.1 Primers used for random mutagenesis of the *araD*_Gt gene

Primer	Sequence (5’→3’)
EP_AraD_Fwd	GGATCCAACCCTTAAGATG
EP_AraD_Fwd_internal	GACTTGGGCGACTGTTTGG
EP_AraD_Rev_internal	AACAACGGCATTGTGGACTG
EP_AraD_Rev	GAGCTCTCCCATATGGTCGAC

2.4. (Semi-)Rational mutagenesis: Site saturation mutagenesis

For site saturation mutagenesis (SSM), the Sanchis protocol was used [227]. Briefly summarized, reaction mixtures contained 50–100 ng of wild-type plasmid as template, 5 pmol of mutagenic primer and a non-mutagenic reverse primer or two mutagenic primers (primer sequence see Table V.2), 0.2 mM dNTP mix, 2.5 U of PfuUltra^{TM} High-Fidelity DNA polymerase and the supplied buffer (Stratagene) in a final volume of 50 μL. The PCR program consists of the following steps: initial denaturation 3 min at 95 °C, followed by 5 cycles of 30 s denaturation at

95 °C, annealing 45 s at 53 °C, extension 1 min/kb according to the megaprimer size at 72 °C. Second stage contained the whole plasmid amplification starting from the created megaprimer by 20-25 cycles of 30 s at 95 °C and extension at 2 min/kb of template at 68 °C, and program was ended by a final extension of 2 min/kb of template at 68 °C. PCR mixtures where then digested with DpnI to remove template DNA, purified using the QIAquick PCR purifcation kit (Qiagen) and transformed into electrocompetent BL21 cells as described above. Plasmids were then extracted and mutagenesis was checked by sequencing.

Table V.2 Mutagenic primers used for site saturation mutagenesis of the *araD*_Gt gene; target codon are underlined

Primer	Sequence (5’→3’)
N28X_Fw	TGTGACATTTACATGGGGANNSGTGAGCGGGATTGACCGGG
W26X-G27X-N28X_Fw	AGCTCCCACAATACCGCCTTGTGACATTTACANNKNNKNNKGTGAGCGGGATTGACCGGGAGCGCG
K42X_Fw	GGGAGCGCGGGTTGGTTGTCATTNNSCCAAGTGGGTTGG
K42X-P43X-S44X_Fw	GATTGACCGGGAGCGCGGGTTGGTTGTCATTNNKNNKNNKGGGTTGGCGTATGACAGACT
S73X_Rv	ATGGGTCGGAGTGTCCGAWNNCGGTTTCCACTCTCCTTC
S74X_Rv	GAGATGGGTCGGAGTGTCWNNTGACGGTTTCCACTCTCC
D75X_Rv	CCAGAGATGGGTCGGAGTWNNCGATGACGGTTTCCACTC
S73X-S74X _Rv	GCCAGAGATGGGTCGGAGTGTCMNNMNNCGGTTTCCACTCTCCTTCC
S73X-S74X-D75X_Rv	GTTTATACAGCCAGAGATGGGTCGGAGTWNNWNNWNNCGGTTTCCACTCTCCTTCCACCACC
T95X_Rv	AACAGTCGCCCAAGTCGAATGMNNATGCACAATTCCTCCGATTCC

2.5. Selection of mutant libraries

2.5.1. Liquid minimal medium

Mutant libraries are transformed into the selection strain by electroporation and a small portion (10%) of the transformation mix is plated on LB-agar to calculate the amount of colony forming units (CFU) in the selection culture. A second part (10-25 %) was grown overnight in liquid Luria broth to check the libraries' quality by sequencing. The rest of the library was grown in liquid Luria broth to start the selection pre-culture. This selection pre-culture was first grown for 2-6 h and then tagatose was added for overnight growth. After overnight growth, a portion of the cells was harvested by centrifugation, washed twice with phosphate buffered saline (PBS; 8 g/l NaCl, 0.2 g/l KCl, 1.44 g/l Na_2HPO_4, 0.27 g/l KH_2PO_4, pH 7.4) and subsequently inoculated (2 % inoculum) in fresh TMM (chapter IV 2.3.1) or TSM (2.1) supplemented with the appropriate antibiotics and IPTG. Cells were grown at 37 °C, 200 rpm and growth was followed by measuring cell density at $OD_{600\ nm}$.

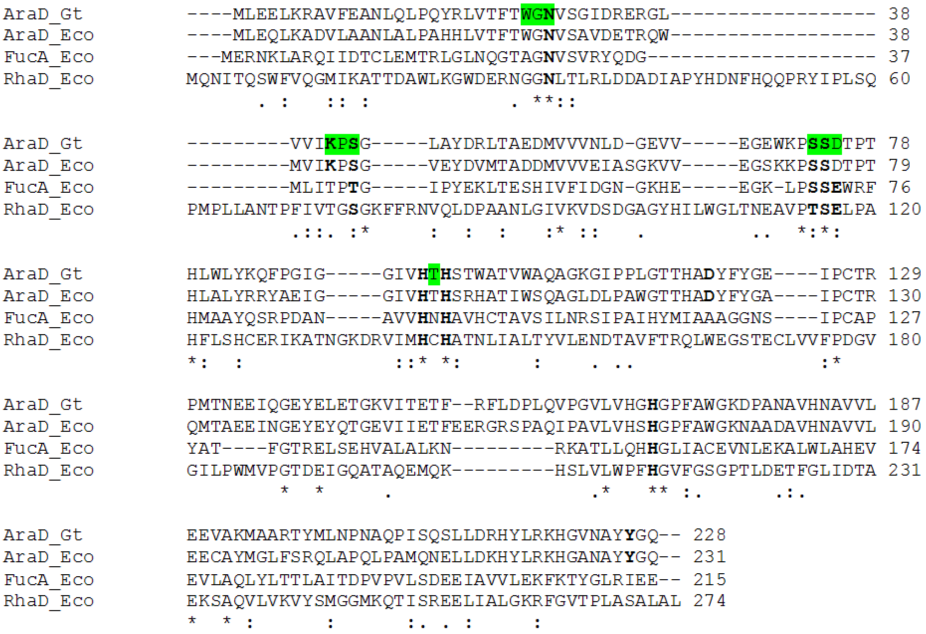
2.5.2. Solid minimal medium

Growth tests were also performed using solid minimal medium containing tagatose. General composition of the TMM-agar medium was the same as described above, supplemented with 15 g/l agar (Difco). Here, carbenicillin was added to allow longer incubation periods since it is found to be more stable in culture media. Triphenyl tetrazolium chloride (BD Difco™ TTC Solution) was added to a concentration of 0.1 g/l after the medium was autoclaved in order to obtain a red color in the colonies that are able to grow on the provided tagatose. The red color is obtained by reduction of the TTC when cells are able to grow on tagatose and hereby create reductive power [37].

3 Results and Discussion

3.1. Sequence analysis and homology model

A sequence alignment (Figure V.1) of the *Geobacillus* L-Ru-5-P 4-epimerase with the related aldolases reveals low amino acids identity with the L-rhamnulose-1-P aldolase and L-fuculose-1-P aldolase, and around 65 % identity with the *E. coli* epimerase. The conserved residues from the AraD-like aldolase/epimerase family are found in the *Geobacillus* epimerase as well.



```
AraD_Gt         ----MLEELKRAVFEANLQLPQYRLVTFTWGNVSGIDRERGL------------------ 38
AraD_Eco        ----MLEQLKADVLAANLALPAHHLVTFTWGNVSAVDETRQW------------------ 38
FucA_Eco        ---MERNKLARQIIDTCLEMTRLGLNQGTAGNVSVRYQDG-------------------- 37
RhaD_Eco        MQNITQSWFVQGMIKATTDAWLKGWDERNGGNLTLRLDDADIAPYHDNFHQQPRYIPLSQ 60
                      . :   :: :            . **::

AraD_Gt         ---------VVIKPSG-----LAYDRLTAEDMVVVNLD-GEVV-----EGEWKPSSDTPT 78
AraD_Eco        ---------MVIKPSG-----VEYDVMTADDMVVVEIASGKVV-----EGSKKPSSDTPT 79
FucA_Eco        ---------MLITPTG-----IPYEKLTESHIVFIDGN-GKHE-----EGK-LPSSEWRF 76
RhaD_Eco        PMPLLANTPFIVTGSGKFFRNVQLDPAANLGIVKVDSDGAGYHILWGLTNEAVPTSELPA 120
                         .::. :*     :  :  :   :* ::   .         ..  *:*:

AraD_Gt         HLWLYKQFPGIG-----GIVHTHSTWATVWAQAGKGIPPLGTTHADYFYGE----IPCTR 129
AraD_Eco        HLALYRRYAEIG-----GIVHTHSRHATIWSQAGLDLPAWGTTHADYFYGA----IPCTR 130
FucA_Eco        HMAAYQSRPDAN-----AVVHNHAVHCTAVSILNRSIPAIHYMIAAAGGNS----IPCAP 127
RhaD_Eco        HFLSHCERIKATNGKDRVIMHCHATNLIALTYVLENDTAVFTRQLWEGSTECLVVFPDGV 180
                *:  :             ::* *:      :    . ..                :*

AraD_Gt         PMTNEEIQGEYELETGKVITETF--RFLDPLQVPGVLVHGHGPFAWGKDPANAVHNAVVL 187
AraD_Eco        QMTAEEINGEYEYQTGEVIIETFEERGRSPAQIPAVLVHSHGPFAWGKNAADAVHNAVVL 190
FucA_Eco        YAT----FGTRELSEHVALALKN---------RKATLLQHHGLIACEVNLEKALWLAHEV 174
RhaD_Eco        GILPWMVPGTDEIGQATAQEMQK---------HSLVLWPFHGVFGSGPTLDETFGLIDTA 231
                        *  *     .                 .*   ** :.      .:.

AraD_Gt         EEVAKMAARTYMLNPNAQPISQSLLDRHYLRKHGVNAYYGQ-- 228
AraD_Eco        EECAYMGLFSRQLAPQLPAMQNELLDKHYLRKHGANAYYGQ-- 231
FucA_Eco        EVLAQLYLTTLAITDPVPVLSDEEIAVVLEKFKTYGLRIEE-- 215
RhaD_Eco        EKSAQVLVKVYSMGGMKQTISREELIALGKRFGVTPLASALAL 274
                *  * :      :      :. . :     :
```

Figure V.1 Sequence alignment of L-Ru-5-P 4-epimerase from Geobacillus (AraD_Gt) and E. coli (AraD_Eco) with L-rhamnulose-1-P aldolase and L-fuculose-1-P aldolase (FucA_Eco and RhaD_Eco, respectively) . Conserved residue are marked in bold and randomized residues are highlighted in green

The homology model of the *Geobacillus* L-Ru-5-P 4-epimerase was created based on structures of the AraD-like aldolase/epimerase family members. It seems to be of good quality since it aligns very well with the *E. coli* epimerase, with a rmsd of only 0.5 Å over the 217 aligned residues. Substrate analogue phosphoglycolohydroxamic acid (PGH) was introduced by superposition with the L-fuculose-1-P aldolase the crystal structure. This was possible since these enzymes' overall structure and conserved residues align very well.

A look at the residues close to the phosphate group of the phosphoglycolohydroxamic acid (PGH) reveals that most of the residues mentioned in the next paragraph (3.2) are indeed in close contact with the phosphate of the substrate analogue, making them good candidates for SSM in order to obtain a binding region for the hydroxylic groups of ketohexoses.

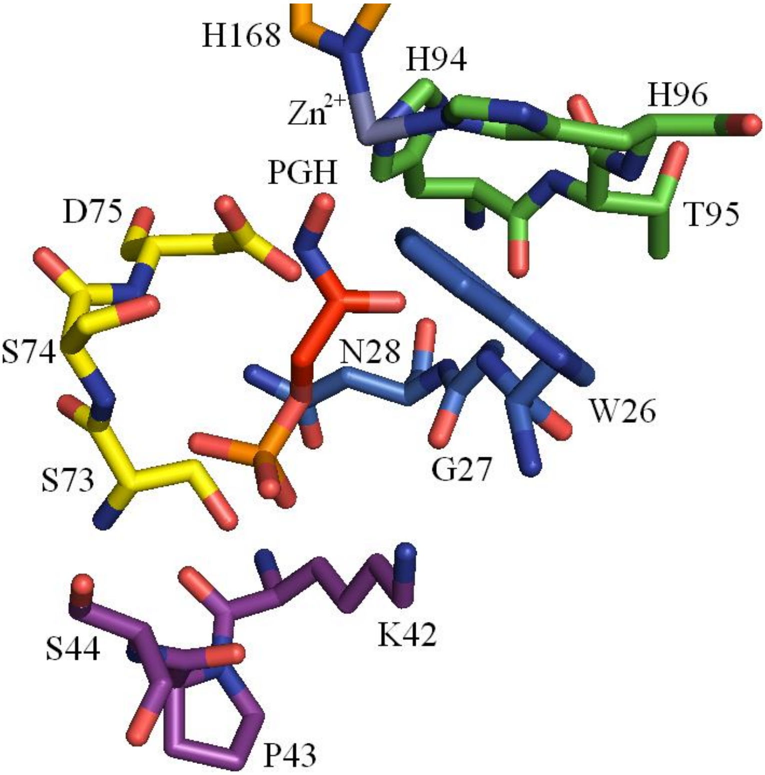


Figure V.2 Homology model of the *Geobacillus* L-Ru-5-P 4-epimerase with the substrate analogue phosphoglycolohydroxamic acid (PGH). The residues targeted with SSM and histidines for Zn2+ binding are highlighted

3.2. Determining residues for SSM (Figure V.2)

The residues targeted by SSM were chosen based on previous engineering results on L-Ru-5-P 4-epimerase [166] as well as the related L-rhamnulose-1-phosphate aldolase [141] and L-fuculose-1-phosphate aldolase [78]. Furthermore, the created homology model of the epimerase and available crystal structures from the above mentioned enzymes were used to have a look in the active site, and more specifically at the substrate and phosphate binding site, of the epimerase. With these tools, 5 residues were identified as important for SSM (N28, K42, S73, S74 and T95) and 5 neighboring residues were taken along with them (W26, G27, P43, S44 and D75). The following paragraphs will further explain why these particular residues were chosen.

Four residues have been proposed to form the phosphate binding pocket in *E. coli* L-Ru-5-P 4-epimerase based on alignment with the 2 related aldolases, namely S44, G45, S74, S75. These residues constitute the phosphate binding pocket in the L-fuculose-1-P aldolase and its corresponding residues in the *Geobacillus* epimerase are S44, G45, S73, S74 [149]. Later on, a conserved asparagine (N28) and lysine (K42) in the active site were added to this list [166]. Most of these conserved residues are suspected to form hydrogen bonds with the phosphate moiety via their β-hydroxyl groups (S), α- or γ-amide groups (G/N). On the other hand, the lysine was found to be positioned at the bottom of the putative phosphate binding pocket, where it is likely to form a salt bridge with the phosphate group [166]. These latter two residues had been subject of mutagenesis for removal of their specific side groups (N28A and K42M). Due to these mutations, affinity of the epimerase for L-Ru-5-P had drastically dropped. An increase in K_m from 0.047 mM to 1.2 mM and > 2 mM was observed, respectively, confirming their role in substrate binding [225]. The corresponding residues of the mentioned amino acids (N28, K42, S73, S74, S44) were targeted by SSM to introduce variety in the phosphate binding region and in this way engineer this pocket towards a pocket that will retain a poly-hydroxyl moiety instead. Extra interaction could be formed with the hydroxyl group at C5 and C6 as well with the carbon atoms themselves.

Since most of the residues are likely to be able to interact with C5, C6 and their hydroxyl groups, and an altered positioning of these residues might also improve binding of the non-phosphorylated ketohexose, some neighboring residues were targeted by SSM as well: W26, G27, P43 and D75. The 10th residue that was subjected to SSM is the threonine (T95) found between two of the histidines (H94, H96) that are known to bind the metal ion. In a similar engineering project, a double mutation was found to convert the L-rhamnulose-1-P aldolase from *E. coli* into an L-rhamnulose aldolase [244]. This double mutant contained the following two mutations: C142Y and T158S. The first (C142) corresponds with the T95 in the *Geobacillus* epimerase, while the T158 is located in a helix which is absent in the epimerase. With the help of the primer listed in Table V.2, several single to quadruple SSM libraries were created combining mutagenesis of the above mentioned residues (Table V.3).

3.3. Random mutagenesis: library quality and selection

It was chosen to work with different levels of mutation frequency to obtain divergent libraries. A low mutation frequency might not introduce enough mutations, whereas the chance of detrimental mutations is bigger with a high mutation frequency. Sequencing of 3 times 3 single colonies grown on LB medium showed that random mutagenesis was successful. They were found to have 0-7 nucleotide changes which resulted in 0-3 amino acid substitutions. The observed mutation frequency matched with the theoretical mutation frequency. More variation was introduced in the libraries by applying random mutagenesis on the entire gene (full enzyme) or only parts of the gene (N-terminal or C-terminal part of the enzyme).

The mutant libraries showed very slow growth and in many cases cultures only reached $OD_{600\ nm} \approx 0.4$ after around 10 days. Due to this slow growth, some of the selection cultures of the SelTag2 selection strain became contaminated. Therefore, it was chosen to work with an intermediate strain from the making of SelTag2. This intermediate strain still contains the kanamycin marker at the position of the *ulaF* gene (*ulaF*::kan) and so kanamycin can be added to the medium to reduce the chance of contamination (see also chapter IV 0).

Over 40 different random mutagenesis libraries, with distinct mutation frequencies, were created and transformed in the SelTag2 or SelTag3 selection strain. Around 3 million CFU were inoculated in TMM or TSM, which resulted in the finding of a potential hit, containing 2 mutations A8V-T23A and a silent one (T115T). Based on their position in the structure, near the enzyme’s surface, and the small difference between the original and new residue, it is unlikely that this mutant is a true hit. Nonetheless, the T23A mutation could have an effect on the alignment of the residues a bit more C-terminal, more specifically those around N28.

3.4. Site saturation mutagenesis: library quality and selection

A total a 10 different residues was subjected to site saturation mutagenesis using 10 different primers containing NNS or NNK degenerated codons. As such, the residue could be switched to all 20 amino acids using 32 codons (Appendix III). Of the 10 primers, 6 contained a single degenerated codon, one a two degenerated codons and the other three could result in triple SSM mutants. Some primers were used separately (with non-mutagenic primers), while different combinations of 2 mutagenic primers were made as well, resulting in SSM of 1-4 positions.

The library part plated out on solid media was used to determine the size of each mutant library (see below and Table V.3). Sequencing results confirm that SSM gave libraries of good to very good quality (Figure V.3), approximating the ideal distribution. An ideal library would consist of percentage A/C/G/T-distributions of 25/25/25/25 for ‘N’ and 0/50/50/0 for ‘S’. Nevertheless, in many cases the natural nucleotide (or codon) is in many cases overrepresented. This is logical since degenerated codons with less difference to the natural codon experience less repulsion during primer annealing. Nonetheless, due to oversampling by using bigger libraries, the chance of starting with a full library (all possible codons) is increased. For a single SSM library with NNS or NNK degenerated codon, the library should contain around 100 CFU to have 95 % coverage (Appendix III) [212].

	Position in codon			
Res.	1	2	3	nucl.
N28	49	44	0	A
	11	10	31	C
	20	26	47	G
	20	20	22	T
K42	30	35	0	A
	14	20	48	C
	28	18	52	G
	28	27	0	T
S73	31	22	12	A
	24	19	63	C
	20	39	25	G
	25	20	0	T
S74	26	26	4	A
	25	36	46	C
	17	20	50	G
	32	18	0	T
D75	30	43	4	A
	23	20	57	C
	31	20	37	G
	16	17	2	T

	Position in codon			
Res.	1	2	3	nucl.
N28	64	59	0	A
	7	8	21	C
	14	18	37	G
	15	15	42	T
K42	51	51	2	A
	12	15	34	C
	20	16	63	G
	17	18	1	T
S73	27	18	25	A
	20	50	47	C
	16	13	23	G
	37	19	5	T
S74	18	15	0	A
	19	61	39	C
	13	13	61	G
	50	11	0	T
D75	20	72	7	A
	16	12	74	C
	55	9	19	G
	9	7	0	T

Figure V.3 Example of the quality of a SSM library (NNS) at different positions. Percentage of each nucleotide at the different positions is given, original nucleotide is underlined. Ideal library would have a N: 25/25/25/25 and S: 0/50/50/0 A/C/G/T-distribution

An overview of the number of CFU that was transformed in the selection strains and selected for growth, and thus epimerase activity, in tagatose minimal medium (TMM) or tagatose synthetic medium (TSM) is given in Table V.3. For most of the libraries, the number of CFU needed to reach the screening effort for at least 95 % coverage [212] was reached by single libraries or combining different transformations in the same library or when multiple smaller libraries were combined.

Table V.3 Overview of the created libraries, their size, screening effort and coverage

Residue(s) targeted	# NNS	Screening effort needed (CFU)	Library size (CFU)	Coverage (%)
N28X	1	94	> 10000	100 %
K42X	1	94	> 10000	100 %
S73X	1	94	± 1100	100 %
S74X	1	94	± 1700	100 %
D75X	1	94	± 900	100 %
T95X	1	94	> 1500	100 %
N28X, S73X	2	3066	± 4400	100 %
N28X, S74X	2	3066	± 3800	100 %
N28X, D75X	2	3066	± 4400	100 %
N28X, T95X	2	3066	> 5000	100 %
K42X, S73X	2	3066	± 550	18 %
K42X, S74X	2	3066	± 2800	90 %
K42X, D75X	2	3066	± 2200	72 %
K42X, T95X	2	3066	± 4400	100 %
S73X-S74X	2	3066	± 20000	100 %
N28X, S73X-S74X	3	98163	± 10000	± 10 %
K42X, S73X-S74X	3	98163	± 10000	± 10 %
W26X-G27X-N28X	3	98163	± 5000	± 5 %
K42X-P43X-S44X	3	98163	± 2500	± 2.5 %
S73X-S74X-D75X	3	98163	> 130000	100 %
N28X, S73X-S74X-D75X	4	3141251	> 20000	± 1 %
K42X, S73X-S74X-D75X	4	3141251	> 25000	± 1 %

As was observed for the selection cultures of random mutagenesis, the SSM libraries also showed very slow growth to $OD_{600\ nm} \approx 0.4$ (again in ± 10 days). Nevertheless, some of the libraries were found to be fully enriched towards one mutant after several inoculations in fresh TMM. Among these mutations were K42S, K42A, K42V, N28R and S73R. Three smaller residues were found in libraries at which K42 was randomized: more specifically serine, alanine and valine. As the distance between the hydroxyl group of serine and the introduced hydroxymethyl group is over 5 Å, it is too far away for H-bond formation. Due to the three smaller residues, the active site might have been slightly prolonged. The introduction of bulkier and charged groups (N28R, S73R) in the phosphate binding region would have the opposite effect and reduce the size of the active site as well as occupy the space that is needed for the extra hydroxymethyl of the hexoses. Surprisingly, one of the libraries had fully enriched to a stop codon mutant, resulting in a truncated epimerase unlikely to be active. Perhaps this minimal growth observed is due to traces of Luria broth from the inoculums, nevertheless, it is still strange that this would lead to the enrichment of a single mutant. Inoculated libraries had been checked by sequencing of the inoculums (examples are given in Figure V.3), while five single colonies had been sequenced after plating out the selection culture, clearly pointing out towards enrichment.

3.5. Solid minimal medium

Due to very slow growth at the solid tagatose minimal medium, only very tiny colonies were observed after several days of incubation at 37 °C. Slight red coloring was observed on the colonies. However, as a result of the very slow growth, the plates and colonies had dried out before the colonies could be inoculated on fresh medium. This could mean either two things, at first, the selection on solid medium works but the initial activity in the strains is to low to achieve growth before the plates and colonies or dried out. On the other hand, this minimal growth could also be a result of growth on traces of Luria broth from the inoculum. This could have been checked by inoculation of the selection strain not expressing the epimerase.

3.6. Fructokinase problem

Due to the (extreme) slow growth of the libraries, the selection system was scrutinized in order to double check the strain and the assumptions that were made to use it as a selection strain. Problems that could occur are the lack of tagatose transport systems, too low concentration of tagatose in the cells and problems downstream of the epimerization reaction.

Uptake of tagatose should not be the problem, since *E. coli* K-12 MG1655 is able to absorb tagatose via the methylgalactoside transport system, *mglABC* [132]. Even though it has a higher flux transporting tagatose out than in, due to high concentration of tagatose in the medium, the *mlgABC* transport system should be able to transport it inside the cells [132]. Nonetheless, intracellular levels of tagatose might not reach high levels.

According to the growth tests that were performed, *E. coli* K-12 MG1655 and the selection strains based on it were able to grow well on fructose. However, there is a difference in fructose consumption between the growth test and in the selection strain, that was overlooked. In *E. coli*, three routes of fructose utilization are present, of which one is predominant [136]. Here, fructose is transported in the cell via the phosphotransferase system or PTS and is simultaneously phosphorylated to fructose-1-P, to which then a second phosphate group is coupled (Fru-1,6-bis-P) and broken down to yield energy and building blocks. The second route is similar but yields fructose-6-P, which is then converted to Fru-1,6-bis-P. In the third route, fructose diffuses into the cell via an isoform of the major glucose permease of the PTS (no PTS involved). It enters the cell as fructose and has to be phosphorylated by ATP and a manno(fructo)kinase (Mak$^+$). However, it was found that the manno(fructo)kinase from *E. coli* K-12 MG1655 works very slow, partially due to lower expression as a result of a rare start codon (personal communication Joeri Beauprez) [174, 240].

In our selection system, tagatose is absorbed and epimerization would result in non-phosphorylated fructose in the cell. As a result, two consecutive bottlenecks are present, namely the epimerase and the manno(fructo)kinase. This second bottleneck hampers the utilization of formed fructose. In a perfect selection strain, only one bottleneck should be present and that is the target enzyme, in this case the epimerase. The fructokinase problem could be solved by constitutive overexpression of a fructokinase on a plasmid or via introduction in the genome (personal communication Joeri Beauprez).

4 Conclusion

In this chapter, two types of mutagenesis are applied on the L-Ru-5-P 4-epimerase: random mutagenesis via error prone PCR and a semi-rational approach through site saturation mutagenesis of carefully selected residues. Both types of libraries are then transformed into the earlier developed selection strains SelTag2 and SelTag3, followed by inoculation in minimal medium with tagatose as a sole carbon source (TMM) for selection of improved variants.

Of over 40 different random mutagenesis libraries created with different levels of mutation frequency, a total of around 3 million CFU were transformed in the selection strains. Up to 300,000 mutants were created in 22 different saturation libraries, in which 1 to 4 codons were simultaneously degenerated using NNS or NNK codons. For smaller libraries (SSM of 1-2 residues), the amount of mutants/CFU for at least 95 % coverage was reached, whereas for bigger libraries (3-4 residues), these numbers were not achieved. Nonetheless, still several thousands of CFU were inoculated in selection cultures, covering 1-10 % of the CFU needed for 95 % coverage.

The majority of the targeted residues is located in the phosphate binding pocket of the epimerase, since the desired substrates are hexoses instead of phosphorylated pentoses. A disadvantage with the targeted or phosphate binding residues is that they are (semi-)conserved residues, which could also mean that they might be of importance for other enzyme features as well, such as folding or stability. On the other hand, since all AraD-like aldolase/epimerase family members are active on phosphorylated carbohydrates and the substrate specificity of the epimerase is desired to be extended or changed to non-phosphorylated sugars in this project, it is not unlikely that (semi-)conserved residues have to be targeted.

Via this selection approach, the following mutants had enriched after several rounds of selection in TMM: K42S, K42A, K42V, N28R and S73R in site saturation libraries and the double mutation A8V-T23A(-T115T) after random mutagenesis. Since one of the libraries had fully enriched to a mutant containing a stop codon, it is not unlikely that there are more false positives among the other mentioned hits. This stop codon would make that only a truncated, very likely inactive enzyme is expressed. Whether they are true hits rather than false positives will be examined later on using the screening assay for detection of tagatose 4-epimerase that was developed (see chapter VI).

Due to the very slow growth of the libraries in the SelTag strains ($OD_{600\ nm} \approx 0.4$ only reached in approximately 10 days), the selection system was scrutinized. This inspection revealed that there is likely to be a second bottleneck, namely the first step in the utilization of fructose. In the growth tests, fructose was imported as fructose-1-P that is easily further broken down, resulting in proper growth. In the selection system, on the other hand, fructose has to be phosphorylated first and the fructokinase, in the strain used, is not sufficient. Due to this second bottleneck, the growth and consequently also the selection is hampered. A perfect selection strain would only have one bottleneck and that is the target enzyme. Constitutive overexpression of an active fructokinase could overcome this second bottleneck and improve the selection strain. More information and possibilities on how to improve the selection strain are discussed in the General discussion & Future Perspectives (Chapter IX).

Nevertheless, engineering of the L-Ru-5-P 4-epimerase towards a 4-epimerase active on non-phosphorylated sugars still seems feasible since similar results were encountered through this approach on the related L-rhamnulose-1-phosphate aldolase [244]. In the next chapter, the development of a screening assay for detection of tagatose 4-epimerase is described. It will also be applied on several libraries as well as to confirm whether the found mutants are true positive hits.

VI. DEVELOPMENT AND APPLICATION OF A SCREENING ASSAY

1 Introduction

In chapter IV and chapter V the development and application of a selection system are described, respectively. However, next to the use of selection as a way to detect improved enzyme variants in a mutant library, one can also perform screening to do this. In a screening assay, each mutant will be tested separately for the desired function or characteristic. Nevertheless, one must have a preferably easy, yet robust and accurate reaction to test the created mutant libraries. In this chapter, the development of a screening assay for the detection of tagatose 4-epimerase activity is described.

Shortly summarized, the screening assay goes as follows (Figure VI.2). If an enzyme variant is able to convert tagatose, its C4-epimer fructose will be formed (epimerization step). In the first step of the screening assay, this formed fructose will be converted to glucose with the help of soluble glucose (xylose) isomerase (SGI). In the second step, the formed glucose will react with two coupling enzymes, namely glucose oxidase (GOD) and peroxidase (POD), and a chromogenic agent (2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) or ABTS) towards a green color [274]. This screening protocol is called the SGI-GOD-POD method according to the 3 enzymes that are applied in a coupled fashion.

2 Material & Methods

2.1. Bacterial strains, plasmids, growth conditions

E. coli BL21 (DE3) cells will be used for the expression of the recombinant L-Ru-5-P 4-epimerase. The epimerase will be expressed from the inducible pIXhPtrc-AraD expression vector created earlier (chapter III). Growth will be achieved on solid and in liquid LB medium supplied with ampicillin to select for the plasmid and IPTG for induction as described above. Mutagenesis was performed as described in chapter V.

2.2. Chemicals and enzymes

Soluble glucose isomerase (Gensweet SGI) was kindly provided by Genencor Belgium. This glucose isomerase is in fact a xylose isomerase from *Streptomyces rubiginosus* and has a molecular weight of 160 kDa. Since this SGI solution contains not only the isomerase but also many stabilizers, both PD-10 desalting columns (GE Healthcare) and Amicon® Ultra 30K Centricons (Millipore) were applied to clean up the enzyme solution. Chemicals were purchased from Sigma-Aldrich unless otherwise stated.

2.3. Development of the screening assay for tagatose 4-epimerase activity

2.3.1. Soluble glucose isomerase

Isomerization of fructose to glucose by the SGI solution was performed similar to the reaction described by Tukel and Alagoz [266]. The final SGI reaction mixture contained 20 mM $MgSO_4$, 1 mM $COCl_2$ in 250 mM Tris-HCl buffer pH 7.5. The mixture was incubated during 20 min at 60 °C and inactivated by a 10 min heat treatment at 100 °C.

2.3.2. Glucose oxidase – Peroxidase (GOD–POD)

In the next step, the glucose concentration was measured, using the coupled enzyme assay of glucose oxidase and peroxidase [274]. In the GOD-POD method, the present glucose will serve as substrate for the coupled enzymes and finally oxidase a chromogenic substrate, in this case 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) or shortly ABTS. The GOD-POD mixture contains 45,2594 mg GOD, 173 µL POD stock solution (40 mg/mL) and 50 mg ABTS in 100 mL 0.2 M acetate buffer pH 4.5. The GOD-POD method has successfully been implemented earlier in enzyme engineering screening assays [64].

2.3.3. Coupling SGI and GOD-POD

In order to remove any GOD-POD disturbing effects of chemicals (like stabilizers) present in the SGI solution, purification was accomplished by PD-10 desalting columns (GE Healthcare) as well as by buffer exchange using Amicon® Ultra 30K Centricons (Millipore). To identify ideal enzyme concentrations, the SGI solution has been tested undiluted and 10 to 1000 times diluted in SGI buffer (20 mM $MgSO_4$, 1 mM $COCl_2$ as cofactors in Tris-HCl buffer at different buffer strengths and pH).

Buffer strengths and pH of the different buffers were also checked to fully optimize the consecutive steps of the screening protocol and provide each step its ideal pH value. First, 125 mM to 1M of Tris-HCl buffer pH 8.5 was tested for its effect on the pH of both the SGI step and the GOD-POD step. Secondly, the ideal pH of the SGI buffer was tested at 250 mM Tris-HCl buffer in a range of pH 8.5 to pH 9.5.

Subsequently, the full SGI-GOD-POD reaction was tested for different fructose concentrations and high tagatose concentrations. Finally, a standard curve of fructose (0–2 mM) in the presence of excess of tagatose (0.5 M) was tested to determine the minimal detectable concentration.

2.4. Screening the libraries

Mutant libraries were picked for solid LB media by using the QPix2 colony picker (Genetix) and transferred to microtiter plates (MTP, Nunc) containing 175 µL of liquid LB in each well and grown overnight at 37 °C and vigorous shaking (220-250 rpm). This first plate is called master plate and after overnight growth a copy was made by inoculation of the grown cultures into fresh LB with 0.1 mM IPTG to trigger induction of heterologous expression (production plate) and grown again overnight for epimerase production. Then, cells were harvested by centrifugation of the production plates during 10 min at 3500 rpm (Rotixa 50 RS Hettich zentrifugen) and medium is discarded. The plates containing the cell pellets were frozen at -20 °C for at least 2 h. Subsequently, pellets are defrosted (30 min at 37 °C) and lysed by adding 100 µl of TrueLyse (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 4 mM $MgCl_2$, 50 mM NaCl, 1 mg/ml lysozyme, 0.1 mM PMSF), lysis is then achieved by 30 min incubation at 37 °C. Cell debris was removed by centrifugation of the production plates during 10 min at 3500 rpm and 50 µl of cell extract was transferred to a new plate to which the tagatose solution was added to start the epimerization reaction at 37 °C for 1 h.

3 Results and Discussion

3.1. Development of the screening assay for tagatose 4-epimerase activity

It was found that undiluted SGI solution gave rise to color formation equal to that of around 350 µM of glucose when coupling the SGI step and the GOD-POD method (Figure VI.1). Since this is around the upper limit of linear detection by the GOD-POD, this disturbing effect had to be solved. Two different routes were used to overcome this issue. Removal of disturbing compounds was achieved by buffer exchange using PD-10 desalting columns or Amicon® Ultra 30K Centricons. Since the SGI solution is a highly concentrated enzyme solution, this purification did not proceed as smoothly as desired. On the other hand, the SGI solution is highly concentrated and thus dilution of hereof might solve this problem to a large extent as well.

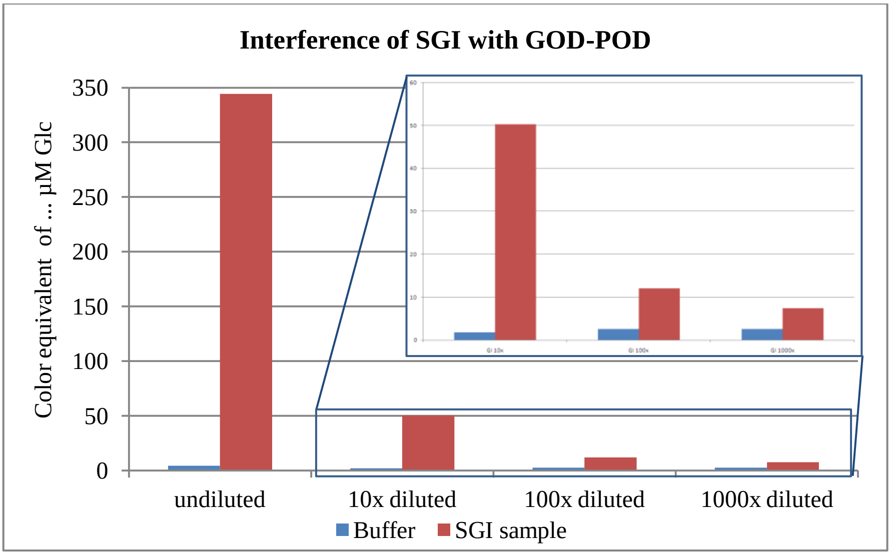


Figure VI.1 Interference of the SGI solution on the GOD-POD measurement

Indeed, it was found that a 10 times dilution of the SGI solution reduced the ‘contaminating factor’ of SGI towards a level around 50 µM glucose. The 100 and 1000 times diluted samples further reduced the ‘contaminating factor’ to levels of around 10 µM of glucose, rendering these dilutions very useful for further screening assay development (Figure VI.1). Besides a nice base-line or zero-point, the mixture should contain enough enzyme to convert the present fructose relatively fast into glucose in order to obtain an easy, fast and robust screening assay. In other words, do the 100x and 1000x dilution still contain sufficient catalytic power to convert the fructose in a desirable time frame or have they become too diluted?

Our results suggest that the 100x dilution still has enough enzyme units present to convert most of the fructose (1 mM) present in 20 min, whereas further dilution (1000x) reduces the activity too much. This suggests that a 100 times diluted SGI solution still contains both enough catalytic potential and reduces the GOD-POD coloring in the absence of glucose as a result of contaminants in the SGI solution. On the other hand, the 1000x dilutions reduced this latter coloring. However, it was too diluted concerning enzyme units. Another way to purify the glucose (xylose) isomerase (GI/XI) would be a cooling crystallization process, starting with the GI/XI in a 0.17 M $MgSO_4$ solution [270].

At last, the buffer strength and the pH of the SGI buffer was optimized so that all enzymes in each step could react at optimal conditions. An SGI buffer containing 250 mM of Tris-HCl was found to be strong enough to increase pH after the epimerization step to an optimal range for the SGI and this without affecting the pH of the GOD-POD step too hard. Under these conditions, an SGI buffer set at pH 9.5 gave more optimal conditions than at lower pH values.

3.2. Overview of the screening assay (Figure VI.2)

Step 1: Epimerization

The screening assay will thus consist of a first epimerization step, in which hopefully a mutant will convert tagatose into fructose. In the epimerization reaction, 50 µl of crude cell extract (50 mM Tris-HCl pH 7.5) will be added to 150 µl of 0.5 M tagatose in 33 mM glycylglycine buffer pH 7.6 containing 0.4 mM of $MgCl_2$ as cofactor. This reaction will then proceed during 1 h at 37 °C or 60 °C.

Step 2: SGI

Then, 40 µl of SGI solution (0.25 M Tris-HCl, 100 mM $MgSO_4$, 5mM $CoCl_2$, pH 9.5) will be added to 160 µl of the epimerization reaction mix from step 1. The resulting mix will then be incubated at 60 °C during 1.5 h to ensure that equilibrium between fructose and glucose will be reached.

Step 3: GOD-POD

After cooling the SGI mix, 50 µl of this solution will be mixed with 200 µl of GOD-POD solution (composition see above) and incubated at 37 °C during 30 min to allow color development. Afterwards, the absorbance at 405 nm is checked to look for increased fructose/glucose concentrations.

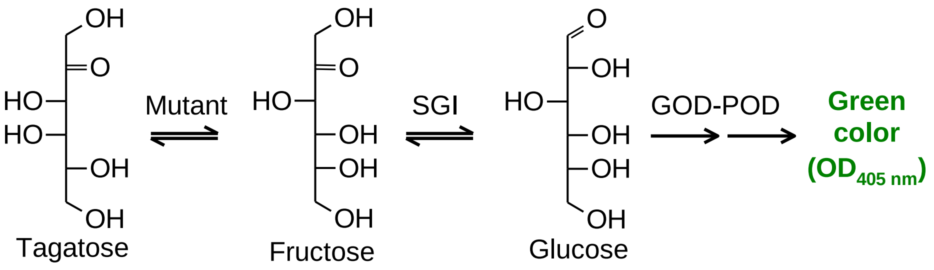


Figure VI.2 Overview of the screening protocol

3.3. Standard curve and CV

When applying this 3 step protocol on a standard curve containing 0 to 2 mM of fructose in an excess of 500 mM tagatose, a clear linear curve is obtained up to 800 µM of fructose, while the detection limit of this protocol lies around 50-100 µM of fructose even in the presence of excess amounts of tagatose (500 mM) (Figure VI.3). Repetition of the protocol for the standard curve was found to be high and thus an easy and robust screening protocol has been developed that can be used in the detection of tagatose 4-epimerase activity. Variation on two plates containing all wild-type epimerases showed that the assay had little variation, namely a coefficient of variation (CV) of around 10 % (Figure VI.4).

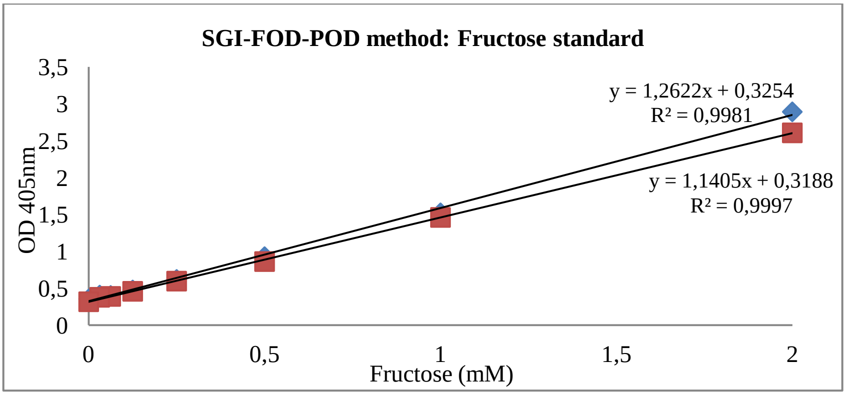


Figure VI.3 Standard curve of the screening protocol

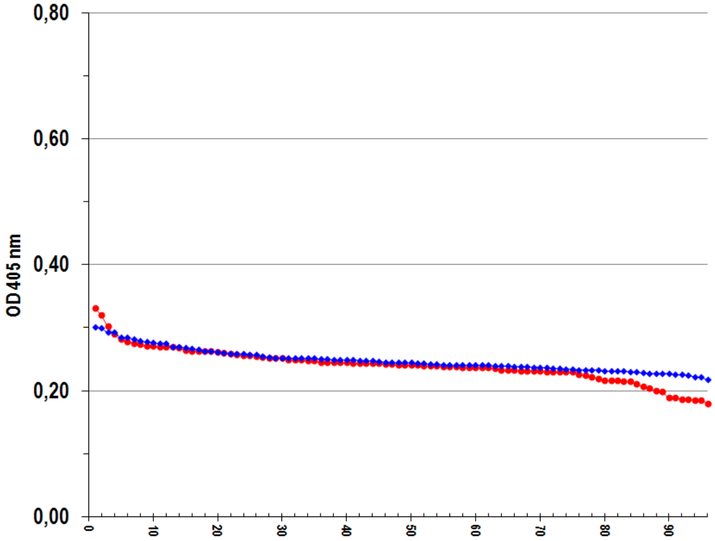


Figure VI.4 Screening of two plates containing the wild-type epimerase shows little variation for the developed screening assay, CV of 7.4 % (blue) and 11.3 % (red)

3.4. Screening for tagatose 4-epimerase activity

Several of the different libraries mentioned in chapter V section 3.4, were subjected to screening. From all 6 single SSM libraries, at least 2 MTPs (192 CFU) had been screened, covering the 94 CF needed for full coverage of the library (Appendix III) [212]. In none of these libraries, hits could be found that clearly and unambiguously showed (improved) tagatose 4-epimerase activity. This contradicts the results from selection in which enrichment towards one single variant was found, suggesting the hits from selection might not possess any tagatose 4-epimerase activity at all.

From both the triple mutant library (S73X-S74X-D75X) and quadruple mutant library (N28X combined with S73X-S74X-D75X), 8 MTPs had been screened, covering smaller portions of the libraries. From the other libraries mentioned in chapter V section 3.4, the colonies formed on LB-agar after plating out a part of the library to count the amount of CFU in the selection libraries (see chapter V) were picked and transferred to liquid media for growth and subsequent screening. As such, another 20 MTP or almost 2000 CFU had been grown and tested with the screening assay. The colonies that were found to potentially hold tagatose 4-epimerase activity (green color formation) were sequenced and tested in multiples, together with the potential hits that were found via selection.

In order to check whether the potential hits found via the selection strains or screening assay were true hits, all these mutants had been grown in eight-fold together with the wild-type and a negative control (no epimerase), both in eight-fold as well. For none of these variants, higher levels of fructose could be detected than observed for the wild-type or the negative control. Even with the use of longer incubation times for the epimerization step (up to 24 h) and/or epimerization at higher temperature (60 °C), no tagatose 4-epimerase activity could be detected for the potential hits. As a result hereof, these potential hits were classified as false positives.

The wild-type epimerase, the K42S and the N28R mutants had been tested by using high amounts of purified epimerase as well, but for none of these fructose formation could be proven by either HPLC analysis (Hypercarb sugar method in chapter III section 2.5) or by the coupled screening assay described in this chapter.

4 Conclusion

In this chapter, a coupled enzyme assay is described that can be applied in the search for C4-epimerase activity on tagatose. The assay is easy, accurate and has the big advantage of being able to detect as little as 50-100 µM of fructose in the presence of excess amount of tagatose (500 mM). This means that high substrate concentrations can be used, which is necessary since it is expected that the epimerase will only have (very) low affinity for the non-natural substrates (free monosaccharide). On the other hand, the low detection limit – even in excess amounts of tagatose – has the advantage that samples do not have to be diluted as the substrate does not interfere with the assay and, as such, even very low activity should be detectable.

Furthermore, a test with a full plate of wild-type epimerase revealed that the epimerase has no or only extremely low activity on tagatose and only low variation is found on this plate. This together with the developed screening assay, would mean that an improved variant would immediately pop out of the plate since it would be the only one with green color formation. Subsequently, the developed screening assay was applied on single SSM libraries of the targeted residues as well as on combinations of these residues. In total, over 8000 CFU had been tested for tagatose 4-epimerase activity, but despite the ability of the screening assay to detect low amounts of fructose in excess amounts of tagatose, none of the created mutants showed to possess proper amounts of tagatose 4-epimerase activity. Even using high concentrations of purified epimerase and longer incubation times, no tagatose to fructose conversion could be detected for the wild-type epimerase and two mutants (K42S, N28R) with HPLC analysis or the developed screening assay.

VII. CHARACTERIZATION OF UDP-HEXOSE 4-EPIMERASE

A part of this chapter has been published as:

Beerens, K., Desmet, T., Soetaert, W. (Epub ahaed of print, 2012). Characterization and mutational analysis of the UDP-Glc(NAc) 4-epimerase from *Marinithermus hydrothermalis*. *Applied Microbiology and Biotechnology*, DOI: 10.1007/s00253-012-4635-6

1 Introduction

As mentioned in the general introduction (chapter I), two naturally occurring enzymes are able to epimerize (modified) carbohydrates at the C4-OH position, namely L-Ru-5-P 4-epimerases and UDP-hexose 4-epimerases. In the previous 4 chapters (chapter III-VI), the different steps in the enzyme engineering of the *G. thermodenitrificans* L-Ru-5-P 4-epimerase are described. In the following 2 chapters (chapters VII-VIII), the other type of 4-epimerase is utilized as a template for enzyme engineering, that is a UDP-hexose 4-epimerase.

The putative UDP-hexose 4-epimerase (GalE) from the thermophilic bacterium *Marinithermus hydrothermalis* will be used as template. This enzyme was chosen since it had not been characterized previously and is expected to be more thermostable than the *E. coli* GalE, an enzyme that has been reported to show very slight epimerase activity on free monosaccharides [130].

Despite the similar reaction of both enzymes (both C4-epimerizations), there is a big difference between the chemistry of both enzymes. Whereas the L-Ru-5-P 4-epimerase is active on a phosphorylated pentose and applies an aldolase/retroaldolase mechanism (C-C bond cleavage), the UDP-hexose 4-epimerase uses nucleotide activated sugars as substrates and acts via a transient keto intermediate with NAD^+ as cofactor (more info see chapter II 3 (Sugar) Epimerases). To make the GalE epimerase active on free monosaccharides, a bigger adaptation of the substrate binding site is needed compared to the adaptation needed for the L-Ru-5-P 4-epimerase, namely the loss of the UDP-group (the biggest part of the substrate) (Figure I.2).

In this chapter, the cloning, expression and characterization of the putative UDP-hexose 4-epimerase from *M. hydrothermalis* (mGalE) is described. This characterization is focused on optimal reaction conditions (pH and temperature), its substrate specificity (non-acetylated vs. N-acetylated UDP-sugars and kinetics) as well as its thermostability.

2 Material & Methods

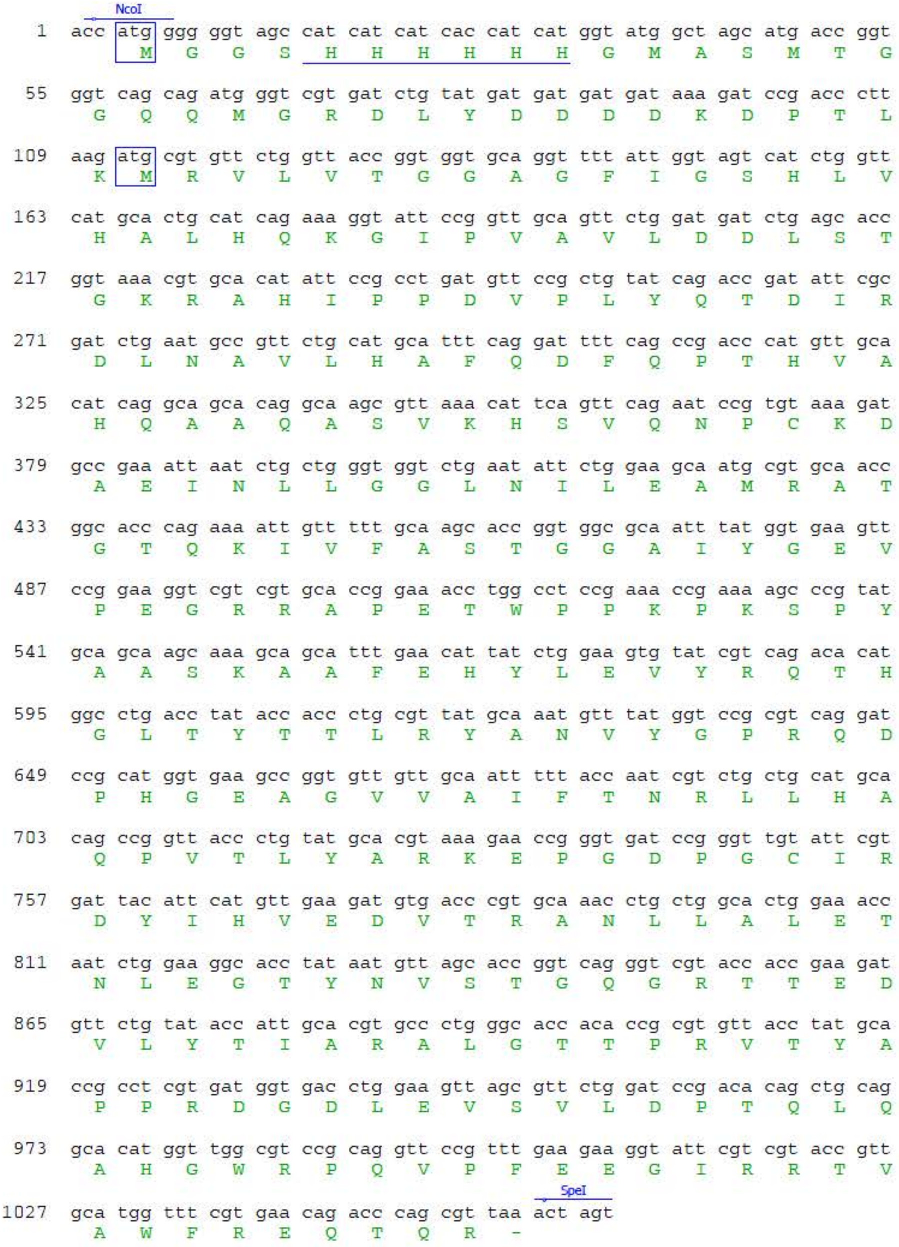
2.1. Bacterial strains and growth conditions, plasmids and chemicals

The codon optimized sequence encoding mGalE (NCBI YP_004368664.1) with a His_6-tag attached to its N-termimus (Figure VII.1) was ordered from GenScript (USA). The empty pIXhPtrc plasmid was created earlier (see chapter III 2.2.1). Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs. Kits for gel extractions, PCR purifications and plasmid isolation were obtained from Qiagen. UDP-sugars were obtained from Carbosynth, while all other chemicals were obtained from Sigma Aldrich, unless otherwise stated.

E. coli BL21 (DE3) was used both for cloning the *galE* gene corresponding to the *Marinithermus hydrothermalis* UDP-glucose/N-acetyl-glucosamine 4-epimerase and for the recombinant expression of the epimerase. *E. coli* cells were cultured at 37 °C in Luria broth (LB: 5 g/L yeast extract, 10 g/L bactotrypton, 10 g/L NaCl, pH 7.0) or LB agar plates (LB plus 10 g/L agar) containing 100 µg/ml ampicillin. For recombinant expression, cells were grown until OD_{600nm} reached 0.6, before induction with 150 µM of IPTG and subsequently grown for another 6h before being harvested by centrifugation.

2.2. Sequence analysis and homology models

Determination of the conserved residues and/or regions of the UDP-hexose 4-epimerases, and to a larger extent the short-chain dehydrogenases/reductases (SDR), was performed by alignment studies. Alignments with other GalEs and sequence analysis of the mGalE enzyme were performed using the online available ClustalW2 alignment program [144], as well as the NCBI BLAST tool [6]. Homology models from both the wild-type enzyme as well as the different mutants were made using the molecular modeling program YASARA [139] based on 5 crystal structures from *E. coli* GalE each containing a different substrates or substrate analogues (eGalE, pdb-codes: 1LRJ, 1UDA, 1UDB, 1UDC and 1XEL) and with default settings. Homology models were also created based on the available crystal structures from the other types of UDP-hexoses 4-epimerases, namely human GalE (hGalE, PDB entries 1EK5 and 1EK6) and WbpP (PDB entries 1SB8 and 1SB9) representing group 2 and group 3. Resulting homology models were saved as pbd-files and compared to each other as well as with other available crystal structures with the YASARA program [139], the YASARA/WHATIF twinset [268] and the PyMOL Molecular Graphics System [70].



```
    1  acc atg ggg ggt agc cat cat cat cac cat cat ggt atg gct agc atg acc ggt
            M   G   G   S   H   H   H   H   H   H   G   M   A   S   M   T   G

   55  ggt cag cag atg ggt cgt gat ctg tat gat gat gat gat aaa gat ccg acc ctt
        G   Q   Q   M   G   R   D   L   Y   D   D   D   D   K   D   P   T   L

  109  aag atg cgt gtt ctg gtt acc ggt ggt gca ggt ttt att ggt agt cat ctg gtt
        K   M   R   V   L   V   T   G   G   A   G   F   I   G   S   H   L   V

  163  cat gca ctg cat cag aaa ggt att ccg gtt gca gtt ctg gat gat ctg agc acc
        H   A   L   H   Q   K   G   I   P   V   A   V   L   D   D   L   S   T

  217  ggt aaa cgt gca cat att ccg cct gat gtt ccg ctg tat cag acc gat att cgc
        G   K   R   A   H   I   P   P   D   V   P   L   Y   Q   T   D   I   R

  271  gat ctg aat gcc gtt ctg cat gca ttt cag gat ttt cag ccg acc cat gtt gca
        D   L   N   A   V   L   H   A   F   Q   D   F   Q   P   T   H   V   A

  325  cat cag gca gca cag gca agc gtt aaa cat tca gtt cag aat ccg tgt aaa gat
        H   Q   A   A   Q   A   S   V   K   H   S   V   Q   N   P   C   K   D

  379  gcc gaa att aat ctg ctg ggt ggt ctg aat att ctg gaa gca atg cgt gca acc
        A   E   I   N   L   L   G   G   L   N   I   L   E   A   M   R   A   T

  433  ggc acc cag aaa att gtt ttt gca agc acc ggt ggc gca att tat ggt gaa gtt
        G   T   Q   K   I   V   F   A   S   T   G   G   A   I   Y   G   E   V

  487  ccg gaa ggt cgt cgt gca ccg gaa acc tgg cct ccg aaa ccg aaa agc ccg tat
        P   E   G   R   R   A   P   E   T   W   P   P   K   P   K   S   P   Y

  541  gca gca agc aaa gca gca ttt gaa cat tat ctg gaa gtg tat cgt cag aca cat
        A   A   S   K   A   A   F   E   H   Y   L   E   V   Y   R   Q   T   H

  595  ggc ctg acc tat acc acc ctg cgt tat gca aat gtt tat ggt ccg cgt cag gat
        G   L   T   Y   T   T   L   R   Y   A   N   V   Y   G   P   R   Q   D

  649  ccg cat ggt gaa gcc ggt gtt gtt gca att ttt acc aat cgt ctg ctg cat gca
        P   H   G   E   A   G   V   V   A   I   F   T   N   R   L   L   H   A

  703  cag ccg gtt acc ctg tat gca cgt aaa gaa ccg ggt gat ccg ggt tgt att cgt
        Q   P   V   T   L   Y   A   R   K   E   P   G   D   P   G   C   I   R

  757  gat tac att cat gtt gaa gat gtg acc cgt gca aac ctg ctg gca ctg gaa acc
        D   Y   I   H   V   E   D   V   T   R   A   N   L   L   A   L   E   T

  811  aat ctg gaa ggc acc tat aat gtt agc acc ggt cag ggt cgt acc acc gaa gat
        N   L   E   G   T   Y   N   V   S   T   G   Q   G   R   T   T   E   D

  865  gtt ctg tat acc att gca cgt gcc ctg ggc acc aca ccg cgt gtt acc tat gca
        V   L   Y   T   I   A   R   A   L   G   T   T   P   R   V   T   Y   A

  919  ccg cct cgt gat ggt gac ctg gaa gtt agc gtt ctg gat ccg aca cag ctg cag
        P   P   R   D   G   D   L   E   V   S   V   L   D   P   T   Q   L   Q

  973  gca cat ggt tgg cgt ccg cag gtt ccg ttt gaa gaa ggt att cgt cgt acc gtt
        A   H   G   W   R   P   Q   V   P   F   E   E   G   I   R   R   T   V

 1027  gca tgg ttt cgt gaa cag acc cag cgt taa act agt
        A   W   F   R   E   Q   T   Q   R   -
```

Figure VII.1 Codon optimized gene and protein sequence of mGalE, including a His-tag with linker and enterokinase cleavage site. The start codon of the His-tagged mGalE and native mGalE are given in the blue boxes, the His-tag is underlined and the NcoI and SpeI restriction sites are marked on the gene sequence

2.3. Construction of the expression vector (pIXhPtrc-mGalE)

A codon optimized sequence encoding the mGalE epimerase, including an N-terminal His-tag was ordered at GenScript (USA). The gene was cut from the supplied plasmid using 10 U of both NcoI and SpeI restriction enzymes (NEB) and subsequently cloned into a likewise treated empty pIXhPtrc plasmid using a 3/1 insert/vector ratio and 3 Weiss units of T4 DNA ligase (NEB) at 22 °C for 1h. After transformation in *E. coli* BL21 cells, plasmid was extracted using the QIAprep spin miniprep kit (Qiagen) after which correct insertion was checked by sequencing (Agowa, Germany). Transformation was performed using 1-2 µL of DNA, 40 µL of electro competent BL21 cells, and a Bio-Rad Gene Pulser (2.5 kV, 25 µF, 200 Ω) in an 0.2 cm electroporation cuvette.

2.4. Recombinant enzyme expression and His-tag purification

For recombinant enzyme production, 10 mL of an overnight culture grown on LB medium was incubated in 500 mL of LB medium containing 100 µg/mL ampicillin in a 2 L shake flask which was then incubated at 37 °C and at 200 rpm shaking speed. Cells were grown until the beginning of the exponential phase ($OD_{600nm} \approx 0.6$) had been reached, after which induction was achieved by addition of IPTG to a final concentration of 150 µM. Cells were kept at 37 °C, 200 rpm for another 6 h for expression and then biomass was harvested by centrifugation for 20 min at 5000 g at 4 °C. Finally, the cell pellets were stored at -20 °C.

For the purification of the expressed epimerase, frozen pellets were thawed on ice for 15 min and consequently dissolved in lysis buffer containing 300 mM NaCl, 50 mM NaH_2PO_4, 10 mM imidazole, 0.1 mM PMSF, 1 mg/mL lysozyme and 6 U/L Benzonase® nuclease (Merck). Lysis was achieved during a 30 min incubation on ice followed by sonication with a Branson sonifier 250 (3 times 2.5 min at level 3 and 50 % duty cycle). Cell debris was pelleted by centrifugation (30 min, 13000 g, 4 °C) and the His-tagged protein was purified from the lysate by Ni-NTA chromatography according to the suppliers protocol (ThermoScientific), wash and elution buffer contained 35 mM and 250 mM imidazole, respectively. After elution, a buffer exchange was achieved using Centricon YM-30 (Millipore) and 10 mM Tris-HCl buffer, pH 7.5 to remove imidazole. The protein concentration of the samples were analyzed with the Pierce® BCA Protein assay kit (ThermoScientific).

2.5. Enzyme activity assays

2.5.1. Activity on UDP-Gal

The activity (as initial velocity) of mGalE on UDP-Gal was spectrophotometerically determined using a colorimetric assay described earlier [179]. This assay determines the amount of glucose which is released after acid hydrolysis of the formed UDP-glucose. In general, epimerase reaction mixtures consisted of 5 mM Tris-HCl buffer pH 8.6 (for temperature studies) or pH 7.5 (for kinetic experiments) and 2 mM of UDP-galactose at 45 °C. Both substrate and enzyme (50 µg/mL final concentration) were separately preheated for 15 min at 45°C, after which the reaction was initiated by merging both fractions. Samples (25 µL) were taken at different time points (each 15-30 s for 90-180 s) and set to pH 2.0 to by addition of 5 µL of 0.1 N HCl for inhibition. After boiling samples for 6 min in a 110 °C dry bath, neutralization was achieved with equimolar amounts of 0.1 N NaOH. Then, released glucose was measured by addition of 200 µl of GOD-POD solution (composition see chapter VI 2.3.2) to 25 µL of quenched and neutralized sample. This final reaction was incubated for 30 min at 37 °C before measuring OD_{405nm}. The standard series, containing 0.05-1.00 mM of UDP-Glc in 5 mM Tris-HCl buffer pH 7.5 and UDP-Gal to a final concentration of 2 mM of UDP-sugar, was treated the same way.

Optimal temperature was determined in 5 mM Tris-HCl pH 8.6 with varying temperatures covering a 45-100 °C range. Thermostability studies were performed in the same way, except that the purified enzyme solution was incubated at 45 °C or 60 °C for different time periods before initiating the reaction to measure initial reaction rate.

For the determination of the pH optimum, activity assays were performed at 45°C in 5 mM of the following buffers: Tris-HCl + 0.1 N HCl pH 2, acetate pH 4.0 and 5.0, citrate pH 6, GlyGly pH 7.6, Tris-HCl pH 7.0, 7.5, 8.0, 8.6, 9.0, glycine pH 8.6, 9.6, 10.6, hereby covering a range of pH 2 to pH 10.6.

Kinetic studies were performed in 5 mM Tris-HCl pH 7.5 at 45 °C, with 8 different substrate concentrations varying from 300 µM to 3 mM. Each time, a series of different ratios of UDP-Glc and UDP-Gal, with a total concentration of 2 mM UDP-sugar, was equally treated to serve a standard. For kinetic studies 2 standard series were used, one containing a total of 2 mM UDP-sugar for higher concentrations and one with 0.5 mM UDP-sugar for the lower range.

2.5.2. Activity on UDP-Glc

The formation of UDP-Gal from UDP-Glc was measured in a similar fashion. First, acid hydrolysis was achieved like described above, namely by addition of 5 µL of 0.1 N HCl to 25 µL of sample followed by boiling the samples during 6 min. Secondly, after neutralization of the sample, released galactose content in the mixture was measured using the commercially available Lactose/D-Galactose Assay Kit (Megazyme). The suppliers instructions were followed; however, the used volumes were ten times lower to ensure processing in micro titer plates. As such, 252 µL of assay solution was added to 20 µL of quenched and neutralized samples. The standard containing 0.05-1.00 mM of UDP-Gal in 5 mM Tris-HCl buffer pH 7.5 , plus UDP-Glc to a final concentration of 2 mM of UDP-sugar, was treated the same way.

2.5.3. Conversion of N-acetylated UDP-sugars

Conversion of the N-acetylated UDP-sugars, UDP-GalNAc and UDP-GlcNAc, was performed under the same conditions as used for non-acetylated UDP-sugars, namely in 5 mM Tris-HCl buffer pH 7.5, at 45 °C and with 50 µg/mL purified protein. Initial activity tests were performed using 2 mM of UDP-GlcNAc and the same 8 substrate concentrations as for UDP-Gal were used in the kinetic experiment (300 µM – 3 mM). Samples were taken for 2.5 min with 30 s intervals and immediately inactivated by placing them in an 110 °C dry bath (FB15103, Fisher Scientific). After inactivation, samples were cooled down on ice and diluted (6-30x) to a total concentration of 50-100 µM of N-acetylated UDP-sugars. Samples (10 µL) were then analyzed using a HPAEC-PAD system (ICS-3000 system, Dionex) with a CarboPac PA20 column (Dionex) at 30°C. Therefore, a sodium acetate (NaOAc) gradient from 0.50 M to 0.68 M was applied over a time span of 12 min. Afterwards NaOAc concentration was brought back to 0.50 M in 1 min and the column was re-equilibrated during 1 min. Flow rate and NaOH concentration were kept constant during analysis and re-equilibration at 0.5 mL/min and 100 mM, respectively. For both N-acetylated UDP-sugars a standard series was run between a concentration of 1 µM and 100 µM.

3 Results and Discussion

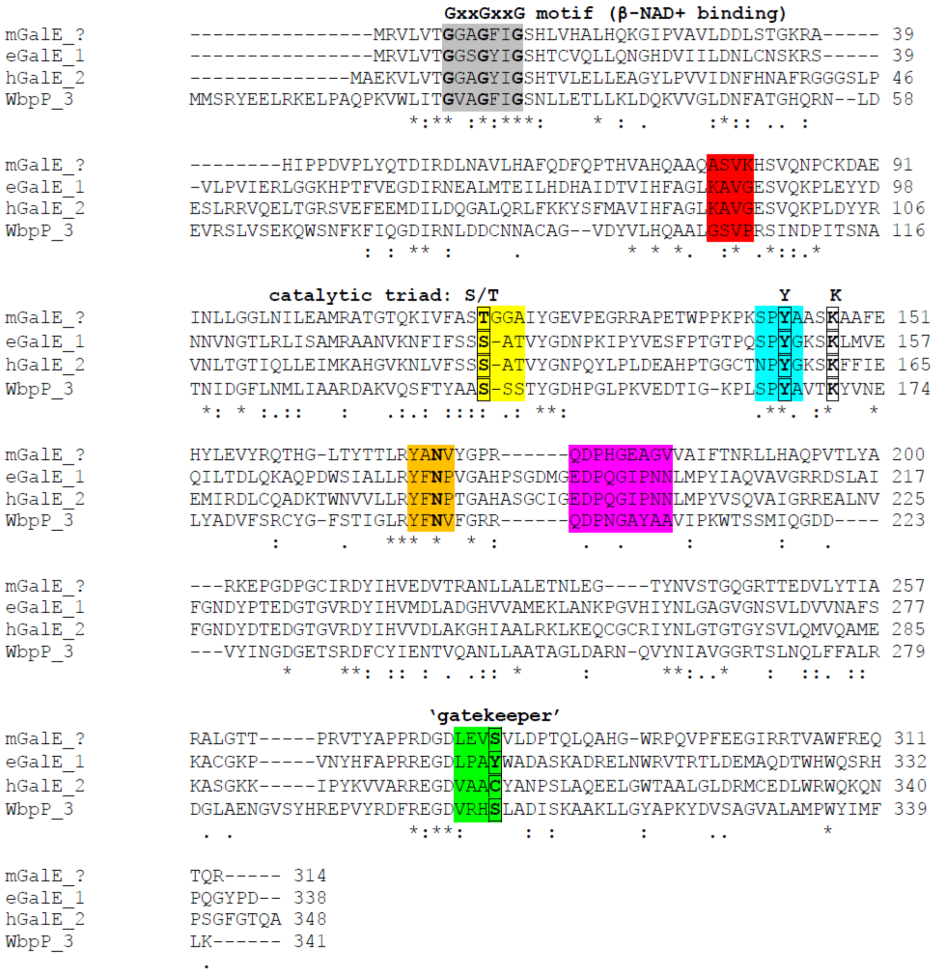
3.1. Sequence analysis and homology models

A multiple alignment (Figure VII.2) of UDP-hexose 4-epimerases revealed that the UDP-hexose 4-epimerase from *Marinithermus hydrothermalis* shares 30 % amino acid identity with the *E. coli* GalE, 27 % with the human GalE, and 30 % with the WbpP enzyme from *Pseudomonas aeruginosa*, which are typical representatives of the three different types of substrate specificities [71, 110]. A higher identity of 66 % was found with the UDP-hexose 4-epimerases from *Thermus thermophilis* and *T. aquaticus* (tGalE), which are active on UDP-GalNAc as well as on UDP-Gal [187].

Furthermore, both important conserved regions are present in this enzyme, namely the GxxGxxG motif in the N-terminal domain which is likely to be involved in the binding of β-NAD$^+$ and the catalytic triade Sx_nYx_3K. Based upon sequence alignments, S116 was postulated to serve as catalytic amino acid in the *Thermus thermophilis* and *T. aquaticus* [187]. However, based upon our alignments, both S116 and T117 could be the catalytic amino acid, whereas the positioning of both residues in the homology model has revealed that T117 is positioned close to the substrate's C4-OH, whereas S116 is pointing away from the active site (Figure VII.3). It is known that in some cases a threonine can replace a serine in a catalytic function, as has been proven for some SDR family members [35, 85, 181]; however, to our knowledge it would be the first GalE bearing a Tx_nYx_3K triad. The activity test of the following two enzyme variants will provide final proof: S116A and T117A (see chapter VIII).

Another reason for this uncertainty is the presence of two consecutive glycine residues (G118-G119) close to the catalytic triad. This unique feature is a notable difference between GalE from *Thermus* species and other UDP-hexose 4-epimerases, in which a single alanine or serine is found at that position. Furthermore, the effect of this double glycine feature will also be examined by mutating them towards a single alanine or serine, as found in other GalE enzymes. (see chapter VIII).

More importantly, a serine residue (S279) is observed at the gatekeeper position in both mGalE and tGalE, pointing towards a similar substrate preference. This suggests that mGalE is likely to be active on UDP-GalNAc in addition to UDP-Gal, as was also found to be the case for the highly similar *T. thermophilis* and *T. aquaticus* GalEs [187]. Activity on non-acetylated and acetylated substrates was investigated (see 3.4).



```
                                      GxxGxxG motif (β-NAD+ binding)
mGalE_?         ----------------MRVLVTGGAGFIGSHLVHALHQKGIPVAVLDDLSTGKRA----- 39
eGalE_1         ----------------MRVLVTGGSGYIGSHTCVQLLQNGHDVIILDNLCNSKRS----- 39
hGalE_2         --------------MAEKVLVTGGAGYIGSHTVLELLEAGYLPVVIDNFHNAFRGGGSLP 46
WbpP_3          MMSRYEELRKELPAQPKVWLITGVAGFIGSNLLETLLKLDQKVVGLDNFATGHQRN--LD 58
                                   *:** :*:***:    * : .     :*:: .. :

mGalE_?         --------HIPPDVPLYQTDIRDLNAVLHAFQDFQPTHVAHQAAQASVKHSVQNPCKDAE 91
eGalE_1         -VLPVIERLGGKHPTFVEGDIRNEALMTEILHDHAIDTVIHFAGLKAVGESVQKPLEYYD 98
hGalE_2         ESLRRVQELTGRSVEFEEMDILDQGALQRLFKKYSFMAVIHFAGLKAVGESVQKPLDYYR 106
WbpP_3          EVRSLVSEKQWSNFKFIQGDIRNLDDCNNACAG--VDYVLHQAALGSVPRSINDPITSNA 116
                               : : ** :     .         * * *.  :* .*::.*

                      catalytic triad: S/T                         Y   K
mGalE_?         INLLGGLNILEAMRATGTQKIVFASTGGAIYGEVPEGRRAPETWPPKPKSPYAASKAAFE 151
eGalE_1         NNVNGTLRLISAMRAANVKNFIFSSS-ATVYGDNPKIPYVESFPTGTPQSPYGKSKLMVE 157
hGalE_2         VNLTGTIQLLEIMKAHGVKNLVFSSS-ATVYGNPQYLPLDEAHPTGGCTNPYGKSKFFIE 165
WbpP_3          TNIDGFLNMLIAARDAKVQSFTYAAS-SSTYGDHPGLPKVEDTIG-KPLSPYAVTKYVNE 174
                 *: * :.::   :   .:.: :::: .: **:                .**. :*   *

mGalE_?         HYLEVYRQTHG-LTYTTLRYANVYGPR------QDPHGEAGVVAIFTNRLLHAQPVTLYA 200
eGalE_1         QILTDLQKAQPDWSIALLRYFNPVGAHPSGDMGEDPQGIPNNLMPYIAQVAVGRRDSLAI 217
hGalE_2         EMIRDLCQADKTWNVVLLRYFNPTGAHASGCIGEDPQGIPNNLMPYVSQVAIGRREALNV 225
WbpP_3          LYADVFSRCYG-FSTIGLRYFNVFGRR------QDPNGAYAAVIPKWTSSMIQGDD---- 223
                       :     .   *** *  * :       .  .     :       :   .

mGalE_?         ---RKEPGDPGCIRDYIHVEDVTRANLLALETNLEG----TYNVSTGQGRTTEDVLYTIA 257
eGalE_1         FGNDYPTEDGTGVRDYIHVMDLADGHVVAMEKLANKPGVHIYNLGAGVGNSVLDVVNAFS 277
hGalE_2         FGNDYDTEDGTGVRDYIHVVDLAKGHIAALRKLKEQCGCRIYNLGTGTGYSVLQMVQAME 285
WbpP_3          ---VYINGDGETSRDFCYIENTVQANLLAATAGLDARN-QVYNIAVGGRTSLNQLFFALR 279
                        *    **: :: : . .:: *     :      **:..*   :  ::. ::

                                    'gatekeeper'
mGalE_?         RALGTT-----PRVTYAPPRDGDLEVSVLDPTQLQAHG-WRPQVPFEEGIRRTVAWFREQ 311
eGalE_1         KACGKP-----VNYHFAPRREGDLPAYWADASKADRELNWRVTRTLDEMAQDTWHWQSRH 332
hGalE_2         KASGKK-----IPYKVVARREGDVAACYANPSLAQEELGWTAALGLDRMCEDLWRWQKQN 340
WbpP_3          DGLAENGVSYHREPVYRDFREGDVRHSLADISKAAKLLGYAPKYDVSAGVALAMPWYIMF 339
                 . .               *:**:     : :       :     ..        *

mGalE_?         TQR----- 314
eGalE_1         PQGYPD-- 338
hGalE_2         PSGFGTQA 348
WbpP_3          LK------ 341
                 .
```

Figure VII.2 Multiple alignment of UDP-hexose 4-epimerases. mGalE, UDP-Glc(NAc) 4-epimerase from *Marinithermus hydrothermalis*; eGalE, UDP-Glc 4-epimerase from *E. coli*, hGalE, human UDP-Glc(NAc) 4-epimerase; WbpP, UDP-GlcNAc 4-epimerase from *Pseudomonas aeruginosa*. Different walls of the hexagonal box model are enlightened in the colors in accordance with the figure in [71] and Figure VII.3. Yellow, blue and orange correspond to the three conserved walls, green indicates the wall containing the gatekeeper, red and purple the other variable walls. In grey the GxxGxxG motif involved in β-NAD+ binding

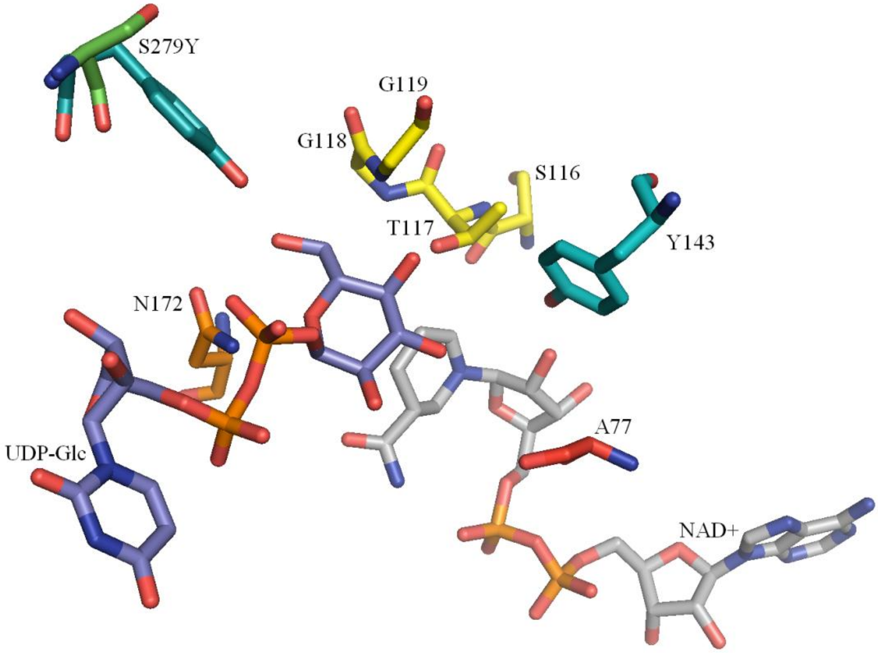


Figure VII.3 Homology model of mGalE. The position of S116 and T117 indicates that the latter is most likely the catalytic residue. The NAD$^+$ forms the bottom of the 'hexagonal box', whereas T117 (yellow), Y143 (blue) and N172 (orange) are the three conserved walls. The gatekeeper S279 (green) and its tyrosine mutant are both shown to emphasize their impact on the size of the active site. Of the two other variable walls, only A77 (red) is shown, since Q178-V186 (purple in Figure VII.2) would be blocking the view

Homology models had been created for the wild-type mGalE as well as for the six different mutants (see chapter VIII). Superimposing the created homology models with the structures available from the different types of UDP-hexose 4-epimerases revealed no major differences in global structure, except for a greater variation in the positioning of the G118-P139 loop or coil in the wild-type and a 3_{10}-helix being formed in the two double glycine mutants (GG→A, GG→S) at A/S118-Y121. The created homology models will be used to interpret the activity measurements and kinetic determinations.

3.2. Construction of the expression vectors

The pIXhPtrc-mGalE vector was successfully created by ligation of the codon optimized *mGalE* gene that was cut from the by GenScript supplied vector into the pIXhPtrc expression vector, as described above. Sequencing of the plasmid confirmed correct insertion in the plasmid and the plasmid could serve as a template for mutagenesis. Furthermore, recombinant expression of the enzyme was successfully achieved in BL21 cells, and His-tag purification gave rather pure, active and highly concentrated epimerase useful for the activity assays.

3.3. Determination of optimal conditions

3.3.1. Temperature optimum and range

Optimal conditions for the *Marinithermus* epimerase were determined using UDP-Gal as substrate. As expected for an enzyme from a thermophilic organism, the optimal reaction temperature is rather high, at about 70 °C (Figure VII.4). This corresponds to the optimal growth temperature of *M. hydrothermalis* [222]. At 60 °C, around 55% of activity is found compared with activity at 71°C whereas 25 % of activity is retained at 45 °C. At 100 °C, no epimerase activity is observed most likely due to fast heat inactivation of the enzyme, rendering heat inactivation possible when taking samples. The optimal temperature is higher than that of its homologues from *Thermus* species [187] but 10 °C lower than that of the GalE from the hyperthermophilic archaeon *Pyrobaculum calidifontis* [223].

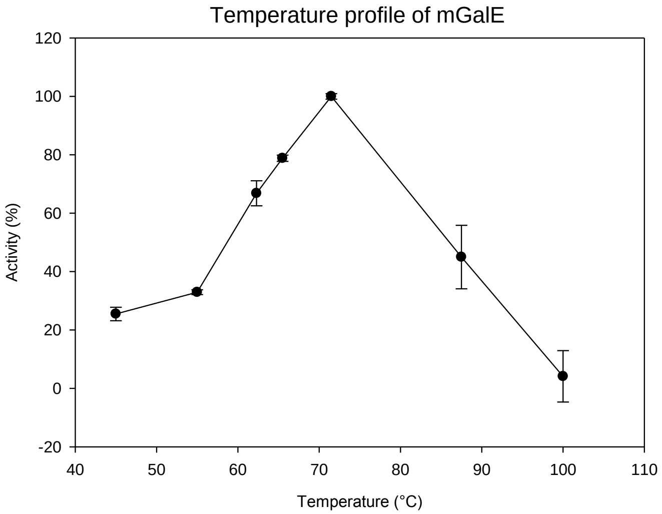


Figure VII.4 Temperature profile of mGalE activity

3.3.2. pH optimum and range

The mGalE was found to be active in a broad pH range (pH 6-9), which is rather typical for this enzyme class (Figure VII.5). However, the highest activity was measured at pH 7-7.5, whereas the optimum for the related GalE enzymes from *T. thermophilus* and *T. aquaticus* was reported to be pH 8.6 [187]. Less than 10% activity is found at pH value ≤5 or ≥10.6 and no activity is found at pH 2. This latter result, allows us to easily stop the reaction by addition of acid, which is also a necessary step for acid hydrolysis of the formed UDP-Glc/UDP-Gal for further analysis.

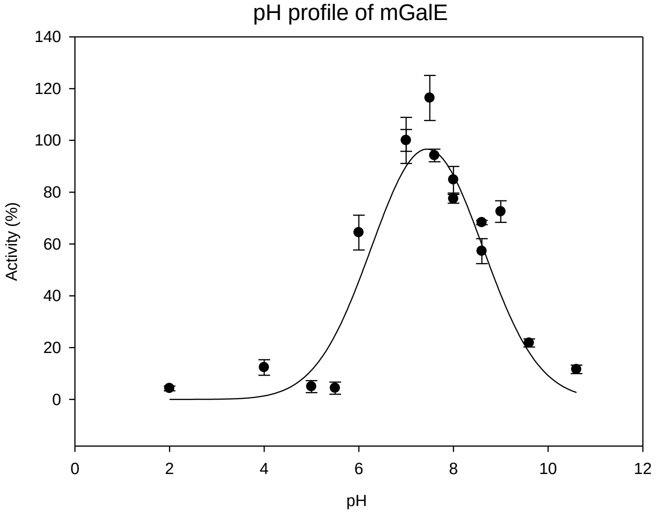


Figure VII.5 pH profile of mGalE activity, including Gaussian curve fit (3 parameters)

3.3.3. Thermostability at 45 °C and 60 °C: half-life (t_{50})

As previously mentioned, fast heat inactivation was observed at higher temperatures. As the conversion of free and/or phosphorylated monosaccharides is expected to be rather low, longer incubation times are favorable and therefore half-life values (t_{50}) at 60 °C and 45 °C were determined. Residual activity was plotted against pretreatment time and with Sigma Plot the t_{50} value was determined (Figure VII.6). The half-life of mGalE at 60 °C and 45 °C is 23 min and 13.5 h, respectively, and almost no activity was lost after 3h incubation at 45 °C. Therefore, it was chosen to perform the reactions at 45 °C, especially for the reactions on free and phosphorylated monosaccharides. The *Marinithermus* enzyme, however, does not display the same thermostability as these counterparts, since it is quickly inactivated at temperatures above 75°C. In contrast, the *Thermus* epimerases retain most of their activity after 10 min incubation at 80 °C [187], whereas the *P. calidifontis* epimerase remains fully active even after 10 min at 90 °C [223].

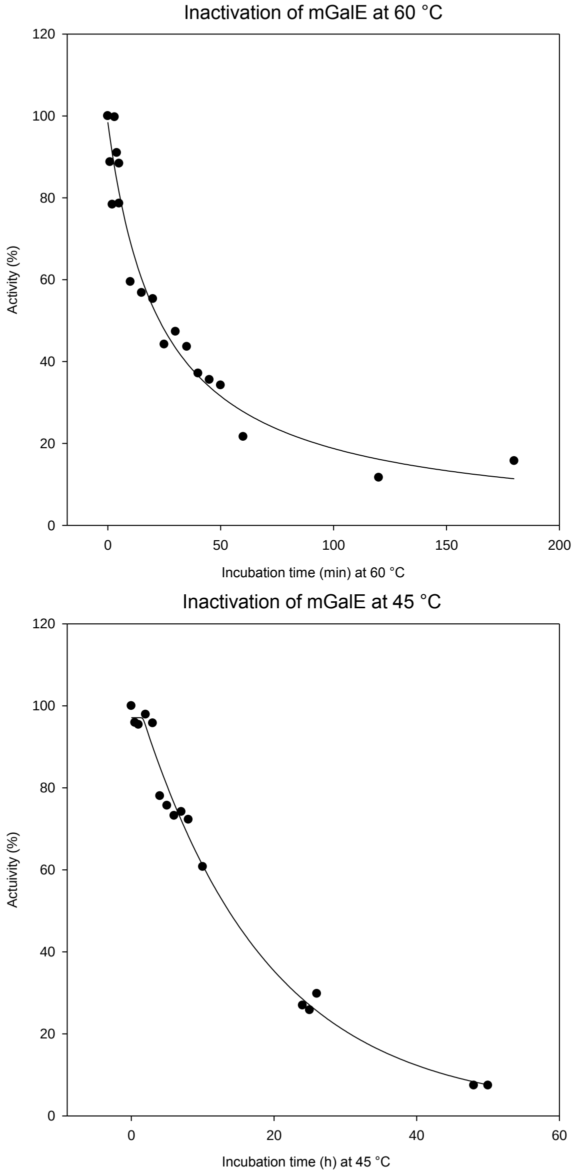


Figure VII.6 Inactivation of mGalE at 60 °C (top) and 45 °C (bottom)

3.4. Substrate preference and kinetic parameters

The mGalE was found to display activity on acetylated as well as non-acetylated UDP-hexoses and can thus be classified as type 2 epimerase. However, a small preference for the latter substrates could be detected, with a specific activity on UDP-Gal and UDP-Glc of 3.0 and 6.7 U/mg, respectively, compared to 1.2 and 0.3 U/mg on their respective acetylated counterparts. The determination of kinetic parameters (Hanes-Woolf linearization) confirmed this observed difference in activity (k_{cat}), whereas the enzyme's affinity (K_m) is much less affected by the presence of an *N*-acetyl group. The obtained kinetic values (K_m, k_{cat} and k_{cat}/K_m) are given in Table VII.1.

Table VII.1 Kinetic parameters from mGalE wild-type

	k_{cat} (s^{-1})	K_m (µM)	k_{cat}/K_m ($s^{-1}mM^{-1}$)
UDP-Gal	2.62 ± 0.49	362 ± 38	6.61 ± 2.39
UDP-GlcNAc	0.58 ± 0.09	519 ± 35	1.11 ± 0.17

4 Conclusion

A novel UDP-Glc(NAc) 4-epimerase from the thermophilic bacterium *Marinithermus hydrothermalis* is described in this chapter. The enzyme was recombinantly expressed in *E. coli* and thoroughly characterized with respect to its substrate specificity and thermal behavior.

Since it is active on both non-acetylated and acetylated UDP-sugars, it belongs to the type 2 GalE epimerases. Nonetheless, it has been found to have a slight preference for non-acetylated UDP-hexoses. This can be seen both in its specific activity and the kinetic parameters. The specific activity on UDP-Gal and UDP-Glc (3.0 and 6.7 U/mg, respectively) was found to be higher than those for their respective acetylated counterparts (1.2 and 0.3 U/mg, respectively). On the other hand, the determined kinetic parameters validate this observation as higher activity (k_{cat}) is detected and higher affinity (lower K_m) in the absence of an *N*-acetyl group.

The optimal conditions for the *Marinithermus* epimerase were pH 7-7.5 and temperature of around 71 °C. Like for other GalE enzymes, the mGalE was also active in a broad pH range (pH 6-9). At temperatures above 75 °C, fast inactivation was observed, whereas at 60 °C and 45 °C around 55 % and 25 % of activity is retained, respectively. At these temperatures, the epimerase shows prolonged stability, namely half-life values (t_{50}) at 60 °C and 45 °C were 23 min and 13.5 h, respectively. Since long incubation times might be necessary for measurements with alternative substrates and/or mutant enzymes, it was decided to perform all other activity tests at 45 °C, to avoid the risk of protein denaturation during enzyme assays.

Furthermore, analysis of its sequence and structure generated new insights in the mechanism of GalE enzymes. These new insights are examined by site-directed mutagenesis, which is discussed in the next chapter.

VIII. MUTATIONAL ANALYSIS OF UDP-HEXOSE 4-EPIMERASE

A part of this chapter has been published as:

Beerens, K., Desmet, T., Soetaert, W. (Epub ahaed of print, 2012). Characterization and mutational analysis of the UDP-Glc(NAc) 4-epimerase from *Marinithermus hydrothermalis*. *Applied Microbiology and Biotechnology*, DOI: 10.1007/s00253-012-4635-6

1 Introduction

After the successful cloning and expression of the *Marinithermus* UDP-Glc(NAc) 4-epimerase, it was thoroughly characterized concerning optimal conditions, substrate specificity and thermal stability. Now, mutational analysis can be applied to further examine the *Marinithermus* epimerase and some of its features.

In this chapter, the determinants of its specificity are analyzed by means of sequence alignments, homology modeling and site-directed mutagenesis. Mutational analysis is focused on 4 different parts of the epimerase, namely the catalytic triad (S116A, T117A, T117S), the unique double glycine motif found in *Thermus* GalE (GG→A, GG→S), the gatekeeper residue for (S279Y) and the 77-82 loop (Loop1_HRDD, Loop2_ARDD). Furthermore, the effects of these mutations on activity and affinity will be examined and explained with the help of homology models for each of the different mutants.

Finally, the activity of mGalE (as well as *E. coli* GalE) on the free monosaccharides Glc/Gal and Fru/Tag as well as the phosphorylated α-Glc-1-P will be tested. Monosaccharide epimerization will be tested with and without addition of UDP(-like) components as so-called ‘cofactors’, in order to test whether the activating effect of the UDP-group (see chapter II 3.3.3) can improve the epimerization reaction on monosaccharides.

2 Material & Methods

2.1. Bacterial strains, plasmids, growth conditions and chemicals

The pIXhPtrc-mGalE expression plasmid containing the *E. coli* codon optimized gene for mGalE with an N-terminal His-tag was created earlier (see chapter III). Kits for gel extractions, PCR purifications and plasmid isolation were obtained from Qiagen. PfuUltra^{TM} High-Fidelity DNA polymerase was purchased from Stratagene, restriction enzyme was obtained from New England Biolabs. UDP-sugars, UDP, UMP and uridine were obtained from Carbosynth. Tagatose was a kind gift from Nitrilab. All other chemicals were obtained from Sigma Aldrich unless otherwise stated.

E. coli BL21 (DE3) was used for expression of the GalE variants as described in chapter VII. On the other hand, large protein expression experiments were performed in a Biostat-B5L fermentor (B. Braun) in double LB, supplemented with 20 g/l glucose. Cells were grown until OD_{600nm} reached 0.6, then IPTG was added to a final concentration of 150 µM to induce recombinant protein expression and growth was continued until maximal growth had been reached.

For the cloning of the *E. coli* GalE (eGalE), the gene was amplified from *E. coli* K-12 MG1655 gDNA with a high fidelity PCR and the following primers (Fw: 5'- CTTAAGATGAG-AGTTCTGGTTACCGGT-'3, Rv: 5'- ACTAGTCATGGTCGTTCCTTAATCG-'3, AflII and SpeI restriction sites are underlined, respectively). Like before, cloning in the pIXhPtrc vector was achieved with AflII and SpeI restriction enzyme and T4 DNA ligase (see chapter III 2.2.2).

2.2. Rational mutagenesis: Site directed mutagenesis (SDM)

For single side directed amino acid mutation, the Sanchis protocol was used [227]. Briefly summarized, reaction mixtures contained 50–100 ng of wild-type plasmid as template, 5 pmol of mutagenic primer and a non-mutagenic reverse primer (for primer sequence see Table VIII.1), 0.2 mM dNTP mix, 2.5 U of PfuUltra^{TM} High-Fidelity DNA polymerase and the supplied buffer (Stratagene) in a final volume of 50 µl. The PCR program consists of the following steps: initial denaturation during 3 min at 95 °C, followed by 5 cycles of 30 s denaturation at 95 °C, annealing 45 s at 53 °C, extension 1 min/kb according to the megaprimer size at 72 °C. The second stage contained the whole plasmid amplification starting from the created megaprimer by 20-25 cycles of 30 s at 95 °C and extension at 2 min/kb of template at 68 °C The program was ended by a final extension of 2 min/kb of template at 68 °C. PCR mixtures where then digested with DpnI to remove template DNA, purified using the QIAquick PCR purifcation kit (Qiagen) and transformed into electrocompetent BL21 cells as described above. Plasmids were then extracted and mutagenesis was checked by sequencing.

Reactions for the loop exchange mutant (Loop1_HRDD) had the same composition, except that both a forward and reverse mutagenic primer were added (Table VIII.1). This PCR program consists of an initial denaturation for 3 min at 95 °C, followed by 30 cycles of 30 s denaturation at 95 °C, annealing for 45 s at 53 °C and an extension of 2 min/kb at 68 °C, and ended with a final extension of 7 min at 68 °C. The second loop mutant (Loop2_ARDD) was created using the Sanchis protocol and the Loop_HRDD plasmid as template. Again, the PCR mixtures were DpnI digested, purified and transformed into electrocompetent BL21 cells as described earlier. Mutagenesis was checked by sequencing after plasmid extraction.

Table VIII.1 Primers used for site directed mutagenesis and loop exchange of the mGalE gene. Target codon are underlined, codon deletions are marked with a * per deleted codon

Primer	**Sequence (5’→3’)**
Reverse primer	GAGCTCTCCCATATGGTCGAC
S116A_Fw	CAGAAAATTGTTTTTGCAGCCACCGGTGGCGCAATTTATGGTGAAGTTC
T117A_Fw	GTTTTTGCAAGCGCCGGTGGCGCAATTTATG
T117S_Fw	GTTTTTGCAAGCAGCGGTGGCGCAATTTATG
S279Y_Fw	GTGATGGTGACCTGGAAGTTTATGTTCTGGATCCGACACAGC
GG118/119A_Fw	GAAAATTGTTTTTGCAAGCACC*GCTGCAATTTATGGTGAAGT
GG118/119S_Fw	GAAAATTGTTTTTGCAAGCACC*AGTGCAATTTATGGTGAAGT
Loop_HRDD_Fw	CCCATGTTGCACATCAGGCAGCACAGCAT**AGGGATGATGTTCAGAATCCGTGTAAAGATGCCGA
Loop_HRDD_Rv	TCGGCATCTTTACACGGATTCTGAACATCATCCCT**ATGCTGTGCCTGCCTGATGTGCAACATGGG
Loop2_ARDD_Fw	GCACATCAGGCAGCACAGGCTAGGGATGATGTTCAGAATCCG

2.3. Expression, purification and activity assays

Expression and purification of the created mutants, and the wild-type as reference, was achieved as described in chapter VII. The same counts for activity assays of these mutants on UDP-Gal and UDP-GlcNAc, whereas activity on free monosaccharides and α-Glc-1-P is described in the next paragraph.

2.4. mGalE activity on free monosaccharides and α-Glc-1-P

To check whether the epimerase was able to interconvert the free monosaccharides glucose/galactose (Glc/Gal) or fructose/tagatose (Fru/Tag) as well as convert α-Glc-1-P into α-Gal-1-P, coupled enzyme assays were checked for their appliance. Glucose detection in the presence of excess of galactose was checked using the GOD-POD method described above (see chapter VI 2.3.2). The other way around, Gal in excess of Glc, was checked using the Lactose/D-Galactose Assay Kit (Megazyme) according to the suppliers instructions. α-Gal-1-P in the presence of α-Glc-1-P was checked with this commercial kit as well; however, the phosphate moiety was first removed by a treatment with alkaline phosphatase (AP). Therefore, 15 µl of AP solution (50 U/ml AP, 1 mg/ml $MgCl_2$, 1 mg/ml BSA in 250 mM Tris-HCl pH 7.5) was added to 15 µl of sample and incubated at 37 °C for 1 h.

Fructose formation from tagatose was checked using the GOD-POD method which was preceded this time by an isomerization step using soluble glucose isomerase (SGI, Genencor). In the first step of the SGI-GOD-POD method, 10 µL of SGI-solution (30 U/ml SGI, 100 mM $MgSO_4$, 5 mM $CoCl_2$ in 250 mM Tris pH 9.5) are added to 40 µl of sample and then incubated for 1.5 h at 60 °C, afterwards 50 µl of the isomerized sample is checked with 200 µl of GOD-POD solution.

Next to these coupled enzyme assays, the samples were also diluted to a maximum of 5 mM substrate concentration and then analysed using the previously mentioned HPAEC-PAD system. Free monosaccharides (Fru, Tag, Glc, Gal) were separated using an isocratic flow of 10 mM NaOH during 25 min on a PA20 column at 30°C (0.5 ml/min). A gradient was applied for the separation of the phosphorylated sugars, α-Gal-1-P and α-Glc-1-P. While NaOH concentration was kept constant at 100 mM, the NaOAc concentration was gradually increased from 100 mM to 175 mM in 7.5 min at a 0.5 ml/min flow rate.

The wild-type mGalE and the S279Y mutant were tested on their ability to epimerize free monosaccharides and α-Glc-1-P. The reaction was performed in 5 mM Tris pH 7.5 and at 45°C to ensure prolonged incubation times. High enzyme concentrations (finally 1 mg/ml) were achieved by concentration of the purified enzyme solutions with the Centricon YM-30 (Millipore) before adding them to the substrate (500 mM). Uridine, UMP and UDP (12.5 mM) was added to the mixtures as a so-called ‘cofactor’ to check their activating effect on the epimerase, whereas double distilled water was added as a negative control. Samples were taken after 0 h, 2 h, 4 h, 6 h and 21 h.

3 Results and Discussion

3.1. Site directed mutagenesis (SDM)

All previously mentioned mutants were successfully created, as was confirmed by sequencing. Afterwards all 6 mutants and the wild-type enzyme were expressed in BL21 cells, subsequently His-tag purified and the buffer was changed to 5 mM Tris-HCl pH 7.5 as described before. The purified mutants were then tested for activity on both non-acetylated and acetylated UDP-sugars and kinetic experiments were performed on all active variants (results are given in the next paragraphs).

3.2. Mutational analysis of mGalE

3.2.1. Catalytic triad Tx_nYx_3K (S116A, T117A and T117S)

As was previously mentioned (see chapter VII 3.1), sequence alignment of the mGalE with other GalE members is insufficient to designate the S116 or T117 as part of the catalytic triad. This is particularly due to the presence of 2 consecutive glycine residues (G118-G119) directly next to these candidates, impeding the alignment. Small errors in the alignment could lead to the hasty and incorrect conclusion of appointing the serine as catalytic residue. According to the homology model we created based on multiple GalE crystal structures, the hydroxyl group from the serine is positioned away from the active site and that of the threonine is pointing towards the 4'-hydroxyl group of the substrate (Figure VII.3), suggesting the threonine to possess the catalytic function. Since both S116 and T117 carry a hydroxyl group that could be part of the catalytic triad, both residues were mutated to alanine. Variant S116A was found to possess activity in the same range as the wild-type enzyme on UDP-Gal, whereas no activity could be detected with variant T117A (Figure VIII.1). Activity could not be restored by increasing enzyme concentration to 0.5 mg/ml and using much longer incubation times. Interestingly, the activity of the latter could be restored by a substitution with serine (T117S) (Figure VIII.1), confirming that both types of residues can function as catalytic amino acid in GalE enzymes. Both S116A and T117S were also tested on UDP-GlcNAc and found to be more or less equally active as the wild-type enzyme (Figure VIII.1). Interestingly, the determination of kinetic parameters has revealed that the catalytic threonine's methyl group is involved in a specific interaction with the substrate's N-acetyl group (Table VIII.2). Indeed, the affinity of T117S for UDP-GlcNAc decreased significantly but that is much less the case for UDG-Gal.

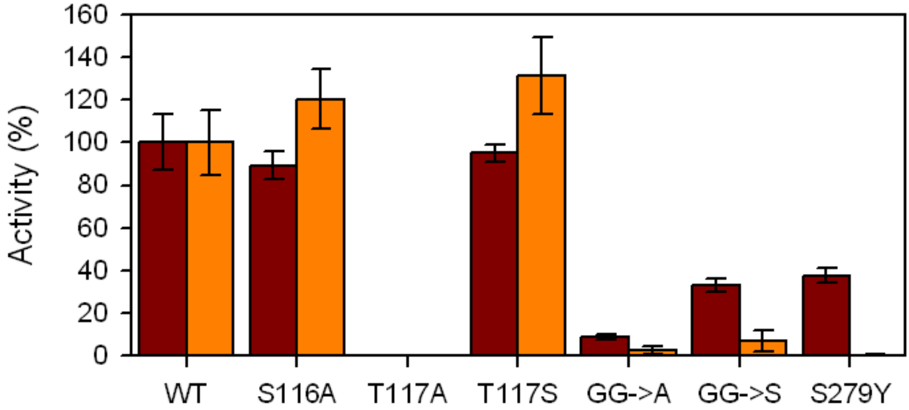


Figure VIII.1 Specific activity of different mGalE variants. The activity is reported relative to the wild-type enzyme, and was measured on UDP-Gal (brown) as well as UDP-GlcNAc (orange). Standard deviation = 1σ (P = 68 %)

3.2.2. Two consecutive glycine residues (G118-G119)

As previously mentioned, the mGalE and the other GalE enzymes from *Thermus* species [187] exhibit the unique feature of having two consecutive glycine residues directly next to the threonine from the catalytic triad, whereas the others possess a single alanine or single serine. Introducing the corresponding mutations in mGalE reduces the specific activity on UDP-Gal to approximately 10-30% of that of the wild-type enzyme (Figure VIII.1). The activity of both mutants on UDP-GlcNAc was reduced even more, to about 3-7% of the wild-type activity. Thus, exchanging this double glycine motif by a single alanine or serine influences the activity on acetylated UDP-hexoses more than on the non-acetylated counterparts, thus pointing to a role in determining substrate specificity. These effects were found to involve changes in k_{cat} as well as in K_m (Table VIII.2). Indeed, the affinity of both mutants for UDP-Gal has clearly increased, whereas the turnover number has significantly decreased. A possible explanation might be found in the homology models of these variants, which suggest the formation of a 3_{10}-helix between residues 118 and 121 instead of the coil or loop in the wild-type enzyme (Figure VIII.2). As a consequence, A120 is pushed into the active site, thereby decreasing the cavity size and limiting the rotational freedom of acetylated substrates. On the other hand, a smaller cavity could improve the interactions with non-acetylated substrates, which is observed in the form of an increased affinity. In turn, the 3_{10}-helix could explain the variants' decreased turnover number as this more stable structure would limit the flexibility of the catalytic threonine residue.

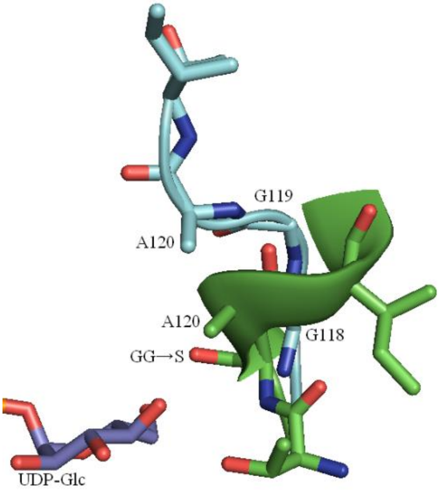


Figure VIII.2 Close-up view of residues 118-121 in mGalE (blue) and its GG→S variant (green). The homology models suggest that a 3_{10}-helix is formed in the variant, which pushes A120 towards the substrate, hereby creating a more narrow active site cleft

3.2.3. The gatekeeper residue (S279Y)

The mGalE can be classified as a type 2 UDP-hexose 4-epimerases since it displays activity on both non-acetylated and acetylated UDP-hexoses, albeit with a slight preference for the former. The serine at position 279 constitutes the so-called gatekeeper residue in the hexagonal box model [110] (Figure II.20). Depending on the type of residue found here, the epimerase does or does not have activity on N-acetylated UDP-sugars. In agreement with the proposed model, variant S279Y (Figure VII.3) was found to lose nearly all activity on UDP-GlcNAc whereas nearly 40 % of its activity on UDP-Gal is retained (Figure VIII.1). More specifically, only a small decrease in K_m value for the latter substrate can be observed, whereas the k_{cat} shows a much larger decrease (Table VIII.2). Thus, mutating its so-called gatekeeper residue S279 to a larger tyrosine made the enzyme specific for non-acetylated substrates. Similar observations were reported for the C307Y and C297Y variants of human GalE [101] and *Yersinia enterocolitica* Gne [26], respectively. Similarly, the reverse mutation in the *E. coli* enzyme (Y299C) lowered the activity on UDP-Gal almost 5-fold, while the activity on UDP-GalNAc increased more than 230-fold [95]. As such, it can be stated that the gatekeeper is indeed a prime determinant for substrate specificity in UDP-hexose 4-epimerases, especially in type 1 and type 2 enzymes. However, their possible conversion into type 3 epimerases is complicated by the need for a hydrophobic '297-308 belt' in the latter enzymes [32]. Furthermore, introducing a mutation in the active site can also affect other parameters of the protein. For example, the substitution of serine by tyrosine in mGalE resulted in complete inactivation after overnight storage at 4 °C, whereas the wild-type retained most of it activity under those conditions.

Table VIII.2 Kinetic parameters from mGalE variants on UDP-Gal and UDP-GlcNAc. Standard deviation = 1σ (P = 68 %) (n.d. = not determined)

	UDP-Gal			**UDP-GlcNAc**		
	k_{cat} (s^{-1})	**K_m (μM)**	**k_{cat}/K_m ($s^{-1}mM^{-1}$)**	**k_{cat} (s^{-1})**	**K_m (μM)**	**k_{cat}/K_m ($s^{-1}mM^{-1}$)**
WT	2.62 ± 0.49	362 ± 38	6.61 ± 2.39	0.58 ± 0.09	519 ± 35	1.11 ± 0.17
S116A	2.32 ± 0.07	437 ± 75	5.39 ± 0.77	1.00 ± 0.07	578 ± 17	1.73 ± 0.07
T117S	2.50 ± 0.16	426 ± 65	5.92 ± 0.54	0.83 ± 0.08	887 ± 35	0.93 ± 0.16
S279Y	0.99 ± 0.02	316 ± 9	3.13 ± 0.07	n.d.	n.d.	n.d.
GG→A	0.37 ± 0.02	243 ± 42	1.55 ± 0.21	n.d.	n.d.	n.d.
GG→S	0.78 ± 0.04	146 ± 25	5.43 ± 0.90	n.d.	n.d.	n.d.

3.2.4. Loop exchange mutants

For the first loop mutant, loop77-82 (ASVKHS) was exchanged by the loop that is encountered in genuine UDP-GlcNAc 4-epimerases, namely a HRDD loop. This mutant was made to check whether the epimerase could be made more specific towards N-acetylated UDP-sugars. However, no activity could be detected for the loop mutant, not even with higher enzyme concentrations and prolonged incubation times. According to the homology model that was made for this loop exchange mutant, the Gln76 present at the beginning of the loop was positioned into the active site, pushing away the catalytic tyrosine (Tyr143) and hereby disrupting the mechanism (Figure VIII.3). Therefore, a second loop mutant (ARDD) was created in which this histidine was replaced by an alanine residue hoping that this smaller residue's side chain would allow a better positioning of Gln76 and hereby not disrupt the delicate chemistry of the mechanism. However, no activity could be found for this loop mutant either. The homology model of this mutant reveals that the positioning of the catalytic tyrosine is better but still not ideal (Figure VIII.3). According to the homology models, the tyrosine's hydroxylic group had dislocated by 5.5 Å in the first loop mutant, while for the second loop mutant the dislocation was only 1.8 Å but it also pushed the catalytic threonine away from the substrate (Figure VIII.3).

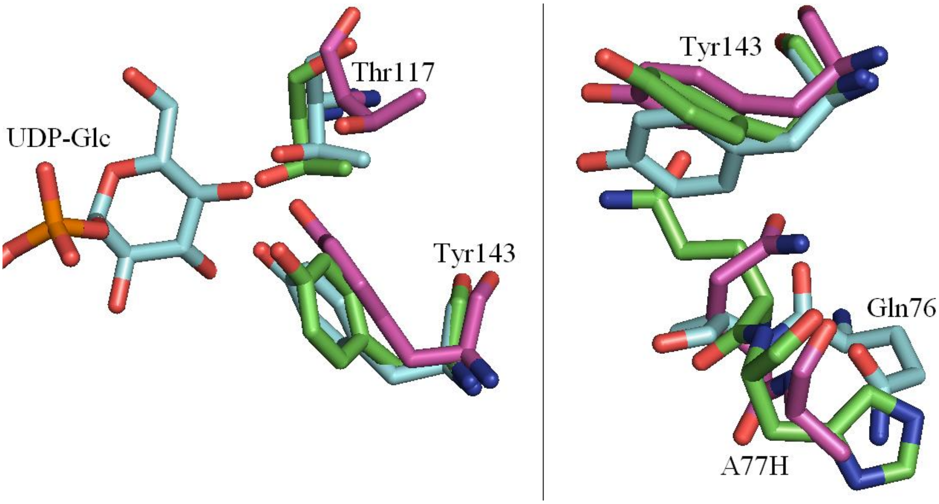


Figure VIII.3 Overlapping homology models of the loop exchange mutants and wild-type. The catalytic tyrosine (Tyr143) is pushed away by both loop exchanges, hereby disrupting the delicate chemistry of the mechanism. Left: positioning of Tyr143 and Thr117 near the substrate. Right: Positioning of Gln76 has a big effect on the dislocation of Tyr143. Wild-type, Loop1_HRDD and Loop2_ARDD are given in blue, green and purple, respectively

3.3. Free monosaccharides detection assays

From the previously mentioned enzyme assays for free monosaccharides (and α-Glc-1-P) detection, the assays for detection of galactose in excess of glucose (Lactose/D-Galactose Assay Kit, Megazyme) and the previously developed detection of fructose in access of tagatose (see chapter VI) are useful as a nice standard curve could be obtained without big interference of excess amounts of the substrate. The GOD-POD method as such was insufficient to detect glucose in the presence of high concentration of galactose. Nevertheless, for each of the desired reactions, one of the screening directions was available: Fru in Tag, Gal in Glc, α-Gal-1-P in α-Glc-1-P.

Furthermore, chromatographic separation (HPAEC-PAD) of the four monosaccharides and 2 phosphorylated sugars was achieved. However, due to high substrate concentration, samples had to be diluted much, which impedes the detection since product peaks are diluted equally. On the other hand, high concentrations may lead to peak broadening and overlapping of the product peak.

3.4. GalE activity on free monosaccharides and α-Glc-1-P

Although very low activity on free monosaccharides has been reported for the *E. coli* epimerase [130], a similar observation could not be made with mGalE even when high enzyme concentrations and long incubation times were used. In order to trigger the activating effect of the nucleotide group [226], either uridine, UMP or UDP was added to the reaction but activity could still not be detected. Since eGalE is a type 1 epimerase that harbors a small active site and large gatekeeper, the S279Y mutant of mGalE was also checked for activity on free monosaccharides but no such activity could be detected for the mutant. Finally, α-glucose 1-phosphate was tried as substrate, since this compound resembles UDP-Glc more closely than a free monosaccharide. However, activity could again not be detected, neither with wild-type mGalE, nor with its S279Y variant.

In contrast to previous observations with the GalE from *E. coli* [130], no activity on free monosaccharides could be detected with the *E. coli* K-12 MG1655 GalE we had cloned. Furthermore, it is well known that the presence of the UDP-group is required for tight binding of the substrate as well as to increase the reactivity of the enzyme-bound NAD^+ cofactor [272, 275]. As such, GalE would only be able to oxidize free sugars to the C4-ketose intermediate, leaving behind an inactivated enzyme in its NADH-bound form [30, 34].

4 Conclusion

In this chapter, the mutational analysis of the *Marinithermus* UDP-hexose 4-epimerase is described. Via site directed mutagenesis, 8 mutants were created and tested for their activity on both non-acetylated and acetylated UDP-sugars.

At first, Thr117 is identified as part of the catalytic triad instead of Ser 116 as was previously postulated [187], confirming the correct positioning of both residues in the homology model. However, not only a threonine can act as part of the catalytic triad as also the T117S mutant shows a similarly active epimerase, clearly indicating this residue as crucial for activity. Secondly, the consecutive glycine residues next to the catalytic threonine (G118-G119) have been shown to be of importance for activity and affinity, with a bigger effect on activity. Furthermore, they seem to be more important for the enzyme's activity on N-acetylated substrates since loss of the double glycine motif lowered the specific activity more drastically.

Furthermore, mutating the gatekeeper towards a bigger residue like found in type 1 epimerases, clearly converted the mGalE from a type 2 towards a type 1 epimerase. This confirms the hexagonal box model [110] that states that type 1 and type 2 epimerases are easily switched into one another by mutating a single residue, namely this gatekeeper residue [232]. However, the connection with type 3 epimerases is likely to be less simple as was also demonstrated by the importance of extra interactions like the hydrophobic '297-308 belt' [32]. The loop exchange mutants that were created to make the mGalE more active on and/or specific for N-acetylated UDP-sugars were inactive. This in accordance with the created homology models that suggest a dispositioning of the catalytic triad's tyrosine (Tyr143).

At last, the enzyme's epimerization capacity on glucose/galactose, fructose/tagatose and α-Glc-1-P/α-Gal-1-P was examined. However, no such activity could be detected under any of the conditions tested. No free monosaccharide 4-epimerization was observed when using the GalE from *E. coli* either, contrary to what was previously reported [130]. However, free monosaccharides epimerization is not obvious since after the (slow) oxidation step, the enzyme is most likely incapable of holding the C4-ketose intermediate long enough to reduce it again, leaving behind a NADH-bound and inactivated enzyme [30, 34].

IX. GENERAL DISCUSSION & FUTURE PERSPECTIVES

1 Introduction

Rare Sugars are defined as monosaccharides and their derivatives that are scarce in nature [91]. Only a small set of monosaccharides is considered to be naturally present in significant amounts (Glc, Gal, Man, Fru, Xyl, Rib and L-Ara), whereas twenty hexoses and nine pentoses were described as rare sugars. Furthermore, deoxygenated, aminated or methylated monosaccharides form another group of rare sugars, which often play a crucial role as recognition elements in bioactive molecules [87, 160, 264]. Notwithstanding their low natural abundance, rare sugars are (potentially) useful in a wide range of applications, as can be seen in Table II.1.

Since rare sugars are scarce in nature, they cannot be extracted from natural resources and thus have to be produced by (bio)chemical reactions. The challenge lies in the optimization and expansion of these production routes in order to synthesize rare sugars with higher yields and at lower costs. More efficient production routes will increase their availability for research purposes, resulting in the discovery of new applications and/or yet unidentified characteristics [91]. Furthermore, increased production efficiency will reduce their cost and increase their use.

The current biochemical production routes have been summarized in the Izumoring [91]. Nevertheless, a large set of potential biocatalysts is missing in this schematic representation, namely the epimerases. Only one type of epimerization is mentioned in the Izumoring, that are ketose 3-epimerizations. More recently, some other epimerases have been found to possess activity on free monosaccharides or hold this potential, see chapter II section 1 (Biocatalytic production routes for rare sugars) [20]. In this work, two different C4-epimerases were investigated for their potential as rare sugar producing biocatalysts, namely a L-ribulose-5-phosphate 4-epimerase and a UDP-hexose 4-epimerase.

In summary, the major achievements of this PhD thesis are

1) **The construction of the inducible expression vector pIXhPtrc**, which is fully complementary with the 4 constitutive expression vectors (pCXhPxx) that were earlier created at InBio.be [1] (Chapter III).

2) **The construction of an inducible and constitutive expression system for L-ribulose-5-phosphate 4-epimerase from *G. thermodenitrificans***, from which the inducible expression system gave higher expression levels (Chapter III).

3) **The construction of an inducible expression system for L-ribulokinase from *E. coli***, which was then heterologously expressed and His-tag purified for the **production of L-ribulose-5-phosphate** starting from L-ribulose and ATP (Chapter III).

4) **Characterization of the *G. thermodenitrificans* L-ribulose-5-phosphate 4-epimerase** with respect to affinity for L-Ru5-P, metal ion activation and stability at 37 °C (Chapter III).

5) **The development of 2 selection strains (SelTag2 and SelTag3) for the detection of tagatose 4-epimerase activity** (Chapter IV), which were then transformed with mutant libraries created by both random and site saturation mutagenesis for selection towards improved variants (Chapter V).

6) **The development of a screening assay that can be applied in search for tagatose 4-epimerase activity** that was afterwards applied on libraries of random and site saturation mutagenesis; however, no improved variants have been detected (Chapter VI).

7) **The construction of an inducible expression system for the UDP-hexose 4-epimerases from *Marinithermus hydrothermalis*** (Chapter VII) **and *E. coli*** (Chapter VIII), which have successfully been overexpressed in *E. coli*.

8) **The thorough characterization of the *Marinithermus* UDP-hexose 4-epimerase** with respect to its substrate specificity and thermal behavior (Chapter VII).

9) **New insights were found for the mechanistical aspects of the *Marinithermus* UDP-hexose 4-epimerase**, as such it was found to be the first UDP-hexose 4-epimerase for which a Tx_nYx_3K catalytic triad instead of the usual Sx_nYx_3K is evidenced (Chapter VIII).

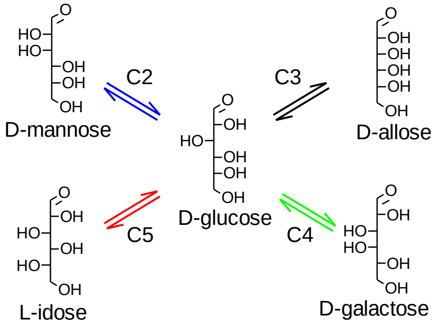
10) **Mutational analysis also demonstrates a new substrate determinant for UDP-GlcNAc 4-epimerase activity**, namely the double glycine motif found in *Thermus* GalEs that seems to be of bigger importance for activity on N-acetylated substrates and **confirmed the interconversion of type 1 and type 2 epimerases by simply mutating a single residue, namely the gatekeeper** (Chapter VIII).

2 General Discussion & Future perspectives

2.1. Rare sugars and potential of epimerases: ‘Epimering’

As can be deduced from Table II.1, rare sugars have a wide range of applications and thus hold tremendous economical value. Nonetheless, due to their low natural abundance they have to be produced from more abundant substrates. To date, the biochemical production of all rare, unmodified monosaccharides is summarized in the Izumoring [91]. Nevertheless, this schematic overview only contains one type of epimerases, i.e. the ketose 3-epimerases like D-tagatose 3-epimerase (D-TE). Since other epimerases can create shortcuts in the current production routes, the discovery or creation of such enzymes is both scientifically and economically interesting. Scientifically because the discovery of new enzymes will give new insights in natural processes and pathways, while redesign of existing epimerases will provide more knowledge on structure-function relationships and enzyme-substrate interactions.

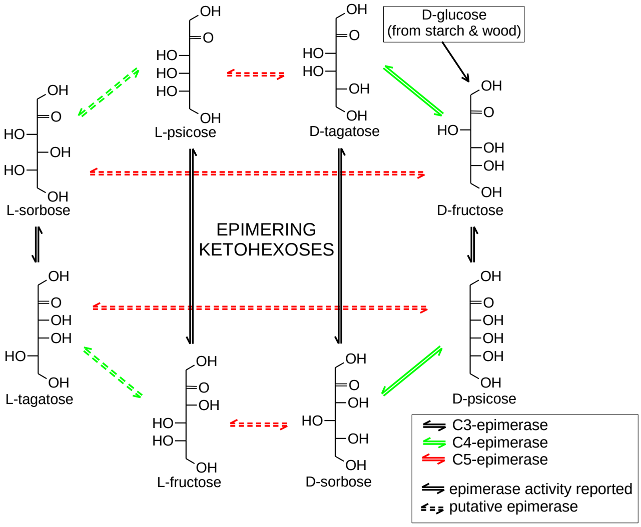
On the other hand, new and/or more efficient production routes will increase the availability of rare sugars for both research purposes and industrial applications. For the application of rare sugars, both the price and availability can be an obstacle. For example, for food applications, the price needs to be as low as possible and availability should be high (*cfr.* bulk) in order to compete with refined sugars and corn-derived sweeteners. On the other hand, when rare sugars are applied in ‘high-value’ products like pharmaceuticals, a higher price is less of a bottleneck but availability should still be high enough (*cfr.* specialty chemicals). New research will result in the discovery of new applications and/or yet unidentified characteristics of rare sugars, as was also stated by Granström et al. [91]. The rare sugar psicose is a clear example that production is the key for application. Only since the discovery of D-tagatose 3-epimerase, it can easily be made from the widely available fructose and since then several applications have been reported (Table II.1 and references herein) and still research is performed in order to further investigate this sugar. As such, one of the four sessions of the latest Rare Sugar Congress (2011 in Kagawa, Japan) was fully dedicated to psicose [111]. In total, over 10 oral presentations were given and over 10 posters presented research on psicose.



D-mannose
C2
C3
D-allose
D-glucose
L-idose
C5
C4
D-galactose

Figure IX.1 Overview of the theoretical C2-, C3-, C4- and C5-epimerase that can use glucose as substrate for the synthesis of mannose, allose, galactose and L-idose, respectively

To point out the potential of other new epimerases for rare sugar production, I have made an 'Epimering' (Figure IX.2 and Figure IX.3). These ring-like schemes that are similar to the Izumoring [91] demonstrate what different epimerases can mean for rare sugar synthesis. In Figure IX.1, the theoretical options are given for the different epimerization reactions starting from glucose, the most abundant sugar. Glucose could be used to produce mannose, allose, galactose and L-idose if a C2-, C3-, C4- and C5-epimerase would be available, respectively. All theoretical epimerizations for ketohexoses and aldohexoses are given in the Epimerings in Figure IX.2 and Figure IX.3, respectively.



D-glucose
(from starch & wood)
L-psicose
D-tagatose
L-sorbose
D-fructose
EPIMERING
KETOHEXOSES
L-tagatose
D-psicose
L-fructose
D-sorbose
C3-epimerase
C4-epimerase
C5-epimerase
epimerase activity reported
putative epimerase

Figure IX.2 The 'Epimering' emphasizes the potential of epimerases for rare sugar synthesis: ketohexoses

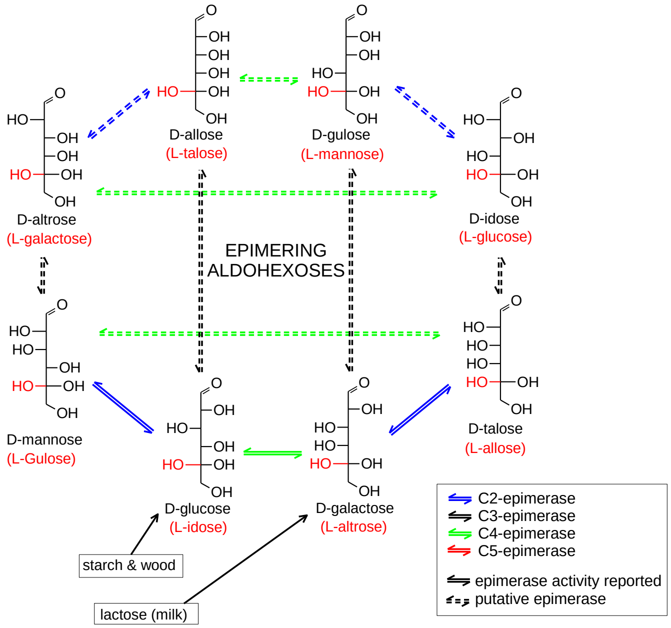

D-allose
(L-talose)
D-gulose
(L-mannose)
D-altrose
(L-galactose)
D-idose
(L-glucose)
EPIMERING
ALDOHEXOSES
D-mannose
(L-Gulose)
D-glucose
(L-idose)
D-galactose
(L-altrose)
D-talose
(L-allose)
starch & wood
lactose (milk)
C2-epimerase
C3-epimerase
C4-epimerase
C5-epimerase
epimerase activity reported
putative epimerase

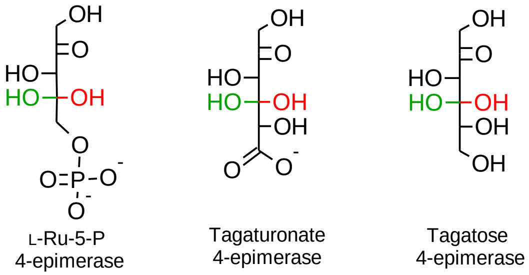
Figure IX.3 The 'Epimering' emphasizes the potential of epimerases for rare sugar synthesis: aldohexoses

2.2. Evaluation of the present project and used epimerases

In this project two epimerases have been subjected to enzyme engineering in order to redesign them to rare sugar producing biocatalysts. Despite the fact that the engineering of these two epimerases did not result in an enzyme harboring tagatose 4-epimerase activity, the choice to start with these enzymes was not a thoughtless idea. For both epimerases there were reasons to believe that the engineering thereof could lead to the desired activity. An evaluation of both enzymes as starting point for the creation of a tagatose 4-epimerase is given.

2.2.1. L-Ru-5-P 4-epimerase: a bad choice as starting point or not?

The attempt to convert a L-Ru-5-P 4-epimerase into a tagatose 4-epimerase was not an unrealistic idea since a mutant of the related L-rhamnulose-1-phosphate aldolase was found to show L-rhamnulose aldolase activity [244], in other words, it had become active on the non-phosphorylated form of the substrate. Also other enzymes have been found active on substrates lacking a phosphate group, as such the fructose-1,6-bisphosphate aldolase is used for the condensation of dihydroxyacetone phosphate onto non-phosphorylated aldehydes, instead of on the natural substrate glyceraldehyde-3-phosphate [8]. Furthermore, a transaldolase could be mutated towards a fructose aldolase, or accept non-phosphorylated glyceraldehyde as substrate rather than glyceraldehyde-3-phosphate [230, 231]. On the other hand, very recently *Thermotoga maritime* was found to possess a tagaturonate 4-epimerase [217]. This enzyme is predicted to be a distant homologue of class II aldolases, utilizing a divalent metal ion for catalysis. Thus, resemblance with the L-Ru-5-P 4-epimerase is striking, showing that such an aldolase-type epimerase can be active on a hexose-like substrate. However, both L-Ru-5-P 4-epimerase and the tagaturonate 4-epimerase are active on substrates with a negatively charged group, whereas in our attempt the substrate would be uncharged (Figure IX.4). This negatively charged group (the phosphate moiety or the carboxylic group, respectively) could be important to hold the intermediate bound in the active site, while the second intermediate is also retained via interactions between its negative charge and the positive charge from the metal cation. Loss of the first interaction might lead to loss of the intermediate, hereby precluding the second step needed for epimerization. The importance of the phosphate group for epimerase activity is also evidenced by mutagenesis studies that have been performed on the *E. coli* epimerase. The loss of the lysine in the phosphate binding pocket resulted in a much lower affinity, the K_m-value of the K42M mutant had drastically increased up to more than 2 mM, while that of the wild-type was only 0.047 mM. On the other hand, the related L-fuculose-1-phosphate aldolase has a K_m-value of 2.2 mM for the wild-type [225]. This could also indicate that the salt bridge between the K42 and the phosphate moiety is necessary to retain the intermediate in order to complete the epimerization reaction, while for the aldolase retention of the intermediate is not needed and therefore no salt bridge is needed/present. Furthermore, the loss of a stronger salt bridge (12.5-17 kJ/mol, and up to 30 kJ/mol) [204] between new substrate (tagatose/fructose) and the binding pocket, as a result of the lack of a charged (phosphate) group, is bigger than the new hydrogen bond (2-6 kJ/mol) [204] that can be formed between the extra hydroxymethyl group and the binding pocket (wild-type or mutant).



L-Ru-5-P 4-epimerase

Tagaturonate 4-epimerase

Tagatose 4-epimerase

Figure IX.4 Comparison of the L-Ru-5-P 4-epimerase and tagaturonate 4-epimerase with the desired reaction (tagatose 4-epimerase). Substrate and product are given in black-green and black-red

Secondly, the L-Ru-5-P 4-epimerase has only a few related enzymes (the 6-deoxysugar-1-phosphate aldolases) and thus little natural variation is known among this enzyme family. Since the desired substrates are hexoses instead of phosphorylated pentoses, the majority of the targeted residues is located in the phosphate binding pocket of the epimerase. Since all AraD-like aldolase/epimerase family members are active on phosphorylated carbohydrates, it is not unlikely that (semi-)conserved residues were targeted. Due to their (semi-)conservation, these residues might not only be of importance for interactions with the phosphate, but also for other enzyme features, like folding or stability. If so, mutation of such residues are likely to be detrimental for the enzyme, lowering chances that mutants are still active ánd will be active on non-phosphorylated sugars.

Due to problems with the selection strains, it is likely that the mutants found via selection are not real hits. Indeed, none of the potential hits were confirmed to possess tagatose 4-epimerase activity by activity measurements using the developed screening assay and HPLC analysis. A possible explanation why these specific mutants were found enriched in the selection cultures could be that they have fewer negative effects on growth, while the minimal growth observed was not a result of little activity but of growth on Luria broth remainders. These traces had been inoculated with the library and thus provide nutrients. For example, the mutant bearing the stop codon will result in a truncated form of the enzyme, producing less useless protein and thus leaving more building blocks and energy for other proteins or functions. Loss of the conserved lysine (K42S/A) or the introduction of positively charged residues (N28R and S73R) might destabilize the recombinant protein, which might then be degraded and recycled faster, returning energy and building blocks for growth. Destabilizing effects can result from the loss of interactions with the hydrophobic part (-$(CH_2)_4$-) of the lysine's side chain and its charged head (-NH_3^+) or from repulsion between the lysine and the introduced arginines (Figure IX.5).

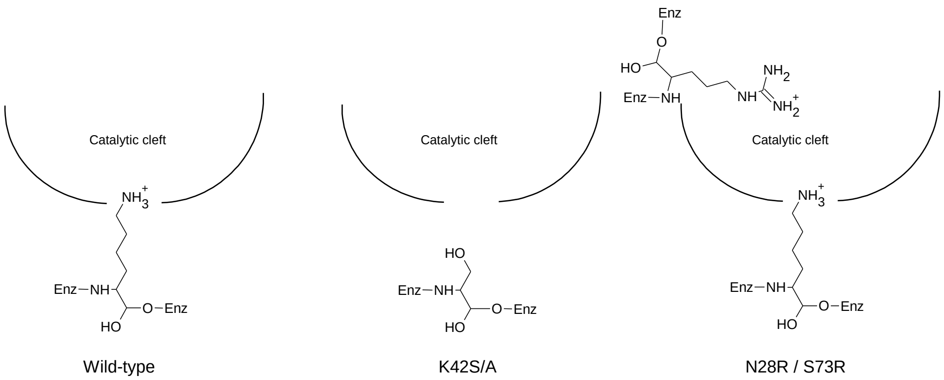


Figure IX.5 The conserved lysine can interact with multiple neighboring residues and these interactions might have a stabilizing effect on the enzyme, loss of these interactions (K42S/A) might lead to an instable variant. Introduction of a second charged residue (N28R and S73R) might destabilize the enzyme due to repulsion between the two charges

Since only a limited number of residues was targeted, the right residue(s) or changes needed for activity on tagatose might not have been targeted. Therefore, more mutagenesis and screening would therefore be advisable. Furthermore, the L-Ru-5-P epimerase might need substantial changes to be converted into a tagatose 4-epimerase. Perhaps multiple residues should be mutated at the same time or insertions and deletions might be needed. Multiple residues have been targeted simultaneously but due to the 'numbers problem' in enzyme engineering [212], these libraries of simultaneous saturating or Combinatorial Active-site Saturation Test (CASTing) become very large quickly, making it very labor-intensive to do all the screening via the developed screening assay. Saturating two residues at the same time requires the screening of around 3000 mutants to have a 95 % chance of full library coverage [212]. For triple and quadruple saturation the amount of mutants increases to almost 100000 and over a million, respectively. Using a NDT codon could decrease library size and still cover most of the amino acids' functionalities, but eight amino acids will not be tested [212]. Library size for single, double, triple and quadruple saturation using NNS or NDT, respectively, is given in Table IX.1. The introduction of insertions and/or deletions is much more complicated since the amount of residues inserted or deleted can be different and in this case a very thorough computational analysis is needed to maximize the chance of success. In fact, in this case *de novo* enzyme design might be a better option (see also below). Furthermore, the detection of desired mutants should also be optimized by improvements in the selection and screening system, which is discussed below as well.

Table IX.1 Screening effort of CASTing libraries using NNS and NDT codons

SSM	NNS			NDT		
# positions	# codons	# CFU	# MTPs	# codons	# CFU	# MTPs
1	32	94	1	12	34	0.4
2	1024	3066	32	144	430	4.5
3	32768	98163	1023	1728	5175	54
4	1048576	3141251	32721	20736	62118	647

2.2.2. UDP-hexose 4-epimerase: epimerase for free monosaccharides?

According to our activity measurements on glucose/galactose, fructose/tagatose and α-Glc-1-P/α-Gal-1-P, C4-epimerase activity could not be observed for the *Marinithermus* enzyme or its S279Y mutant. No C4-epimerization of free (or phosphorylated) monosaccharides was observed when using the GalE from *E. coli* either, unlike what was previously reported [130]. Indeed, epimerization of free monosaccharides is not evident, since the enzyme is most likely incapable of holding the C4-ketose intermediate long enough to reduce it again, leaving behind a NADH-bound and inactivated enzyme [30, 34]. The UDP-hexose 4-epimerases need to undergo rather substantial changes in order to become active on free monosaccharides because the nucleotide group of the substrate is of importance for both substrate binding and activation of the enzyme by promotion of the conformational shift [272, 275]. Mutations would be needed to induce this conformational shift without the UDP-moiety or permanently activate the enzyme as well as to increase affinity for the free monosaccharides in order to bind them and hold the intermediate properly in the active site. The *E. coli* epimerase was previously tested using a fructose dehydrogenase assay and HPLC analysis after derivatization [130], whereas we used our developed screening assay (chapter VI). Nonetheless, according to the specific activity reported, similar amounts of activity should be able to be detected by this screening assay. However, even with longer incubation times, increased enzyme and substrate concentrations, no epimerase activity was observed, nor could the addition of UDP, UMP or uridine as cofactor trigger epimerase activity.

2.3. Where to find or how to create new epimerases?

As was discussed above, the chance of obtaining a tagatose 4-epimerase by directed evolution of a L-Ru-5-P 4-epimerase or a UDP-hexose 4-epimerase seems rather small. In order to obtain such an enzyme, different aspects could (or should) be changed. Among these aspects are starting from another enzyme (fold), a rational approach using *de novo* design, and optimization of the selection and screening strategies.

Thus, how to create or where to find this new biocatalyst (or for the production of other rare sugars) remains an important question to be answered. Nature might hide some of these biocatalysts in the genetic code of one of the millions of creatures inhabiting earth. Genome sequencing has provided scientists with massive amounts of yet uncoded DNA, perhaps somewhere hiding a new epimerase active on free monosaccharides, like the ketose 3-epimerases (e.g. D-tagatose 3-epimerase) [114] or the recently discovered tagaturonate 4-epimerase [217]. Many genes can be given a 'putative' function based on homology with known/existing enzymes but for totally new enzymes and/or enzyme folds this cannot be done. On the other hand, galactose 4-epimerization has been reported in cell extract from *Kluyveromyces* species induced with xylose [73]. Nevertheless, the responsible enzyme has not been identified, nor is it known whether one or more enzymes are responsible for this reaction. Identification of this (these) enzyme(s) can be achieved by separation of different enzyme fractions of the *Kluyveromyces* cell extract by for example ammonium sulfate or solvent precipitation or separation based on protein size. The enzymes in the fraction containing the activity can further be separated and analyzed. Finally, protein sequencing can be applied to determine the primary structure of the responsible enzyme. However, when multiple enzymes are involved, these different biocatalysts must always be found in the same fraction. Another option to identify the responsible enzyme(s) is by comparing cell extract of xylose-induced *Kluyveromyces* with that of normally grown *Kluyveromyces* on a 2D-PAGE to look for differently expressed proteins. Spots that are only observed in the xylose-induced cell extract can then be tested for activity and/or identified by protein sequencing.

However, repetitions of the described experiments on a *Kluyveromyces marxianus* strain (MUCL 30062) failed to confirm this galactose 4-epimerization activity (data not given). The *K. marxianus* strain that we have used might be different from the one that was found to have galactose 4-epimerase activity, explaining why experiments could not be repeated. Nevertheless, there were no new papers published dealing with this very interesting topic, although it was stated in the article (in 1996) that 'it is the beginning of a series of investigations on the mentioned enzyme'. This might also mean that the 'weak epimerase activity' found in the 'preliminary assays' had been a hasty conclusion or that they were unable to repeat and/or isolate the enzyme responsible for the galactose 4-epimerization. On the other hand, xylose can induce the UDP-hexose 4-epimerase in the yeast *Pachysolen tannophilus* [239], so it might have done the same in the *Kluyveromyces* strain. However, it would be unlikely that the UDP-hexose 4-epimerase from *Kluyveromyces* is active on only 1 g/l (5.55 mM) of galactose.

Another option to look for tagatose converting enzymes are *Mucoraceae* fungi. They can be applied in a two step production route towards tagatose starting from fructose, in combination with tagatose 3-epimerase [114]. At first, the fructose is converted to psicose, which can then be converted to tagatose via biotransformation with the help of *Mucoraceae* fungi (Figure IX.5). Various *Mucoraceae* strains have been reported to be able to carry out this reaction [289]. Neither the availability, nor the price of psicose are limiting factors here, as psicose mass production has become industrially feasible since the discovery of ketohexose 3-epimerases [129, 250]. The two-step tagatose production from fructose over psicose is a good alternative method; however, it still requires further intensive investigation [191]. Similar experiments as described for the galactose 4-epimerase identification in the *Kluyveromyces* strain could be applied for the identification of the responsible enzymes for psicose to tagatose conversion in *Mucoraceae* fungi. Most likely, one or two oxidoreductases are involved and conversion proceeds over D-talitol.

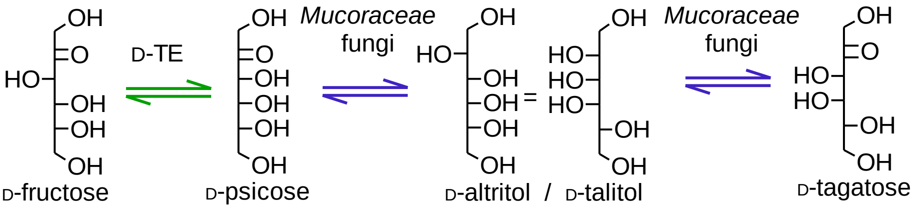


Figure IX.6 Conversion of fructose into tagatose using D-tagatose 3-epimerase (D-TE) and Mucoraceae fungi

Another place to look for new epimerases active on rare sugars are niches in which they are found. For example, since tagatose is found in the gum of *Sterculia setigera* [80, 191], this tree needs to have a pathway to produce tagatose. Nonetheless, synthesis could also proceed using aldolases or an isomerase instead of an epimerase.

In order to overcome this issue of screening in nature, enzyme engineering remains an interesting and promising field. By mutagenesis of an existing biocatalyst, new biocatalysts might be created. More recently, also *de novo* enzyme design has become an option. As the name suggests, *de novo* enzyme design is making an enzyme from scratch, using computational tools and complicated algorithms [119, 216, 218, 219, 221]. An example hereof is the computational design of enzymes to catalyze the Kemp elimination, a reaction not found in nature, that was made by the research group of David Baker [221]. The Kemp elimination enzyme that was created is far from optimal but it can be further improved by enzyme engineering, like via random or site saturation mutagenesis.

An important starting point is the choice of enzyme (fold) to use as template for rational design. It can be based on binding of the substrate or product instead of the performed reaction. Several enzymes would then be eligible, like the tagatose 3-epimerase [114] but also L-arabinose isomerase, tagatose-6-phosphate isomerase, and other monosaccharide isomerases, since these enzymes are known to bind free monosaccharides. Starting from these enzymes has the advantage that the enzyme is already able to bind the substrate but its catalytic machinery should be repositioned or changed. Their natural reaction could easily be inactivated by mutation of the catalytic residues, while their ability to bind the monosaccharides is (most likely) maintained. The big challenge is then to introduce new catalytic residues that catalyze a different reaction. For the tagatose 3-epimerase, it might also be worth looking for a way to move the catalytic machinery to the next hydroxyl group or to make the enzyme bind the substrate in a slightly rotated position. Since the catalytic residues need to be positioned correctly in order to make the enzyme active, this should be thoroughly analyzed by using computational models. It is necessary to have as much information as possible on how the current fold binds the substrate and what other interactions are present within the enzyme fold itself since mutations can disrupt this fine equilibrium. However, these interactions can also be monitored using bioinformatics and visualization programs like the YASARA program [139] or the YASARA/WHATIF twinset [268].

The easiest ways for the inactivated enzymes to become active again by directed evolution is by reintroduction of its original catalytic residues. Since this is of course not the purpose, it would be advisable to engineer the enzyme folds via a rational approach. This rational approach would also overcome the disadvantages of directed evolution, for example the insufficiency when multiple mutations are needed [22], narrow range [21], limits in introducing new function [84]. In order to minimize lab work, but maximize the chance of finding a positive mutant, the technique of *de novo* design is recommend. Using computer tools like the Rossetta algorithms [216, 219, 221] a small set of potential enzymes could be designed. These computer algorithms are based on all 'restrictions and rules' that have been found to occur in natural protein folding and catalysis. This would also decrease the mutagenesis and screening work in the 'wet lab'. As mentioned above, the more information that is available, the more accurate the algorithms can predict correct folds. Computer scientists need to have information on how nature and its building blocks work in order to translate it to algorithms and programs that can be used to create and analyze computational models.

Not only computational tools are needed but also easy and efficient screening assays to test these theoretical catalysts in a 'wet lab'. The methods to check whether these theoretical enzymes are truly active enzymes harboring the desired activity need to be – preferably – simple and easy. These techniques should also be applicable in random or site saturation mutagenesis experiments as the *in silico* designed enzymes will need further optimization. In this way new information can also be gained and later be translated into new or better algorithms to improve the current bioinformatics tools. The screening assay that was developed during this doctoral research could be used to test a set of theoretical biocatalysts. A set of 96 proteins could be tested in a single microtiter plate. However, the bigger the *in silico* library, the more work that has to be done on the bench and the higher the costs. It would for instance demand the chemical synthesis of all these genes, which becomes expensive, and transformation into an expression system, while expression of each of these enzymes should be optimized.

Despite having developed an easy and accurate screening assay, further optimization can be achieved for screening in order to increase the chance of success. Furthermore, the selection system can also be improved since the one developed in this project was not perfect. For the selection strain, potential problems that could occur are the lack of tagatose transport systems, too low concentration of tagatose in the cells and problems downstream of the epimerization reaction. Since *E. coli* K-12 MG1655 is able to absorb tagatose via the methylgalactoside transport system, *mglABC* [132], uptake of tagatose should not be the problem. Even though it has a higher flux transporting tagatose out than in, due to high concentration of tagatose in the medium, the *mlgABC* transport system should be able to transport it inside the cells [132]. Nonetheless, intracellular levels of tagatose might not reach high levels which could be a problem. This potential problem could also be solved by overexpression of the methylgalactoside transporter *mglB.* Similarly, the overexpression of the *galP* transporter could restore growth on glucose in strains lacking the PTS system [98].

Furthermore, it can be assumed that a second bottleneck, downstream of the epimerization reaction, was present in the selection strains, that is the first step in fructose utilization (Figure IX.7). Due to this second bottleneck, the growth and consequently the selection is hampered. In the growth tests, fructose was imported as fructose-1-phosphate that is easily further broken down, resulting in proper growth, whereas in the selection system, fructose would be present inside the cells instead of fructose-1-P. A consequence hereof is that the fructose has to be phosphorylated first in order to further metabolize it [136]. The responsible enzyme, a manno(fructo)kinase (MAK) is only slightly active in *E. coli* K-12 MG1655 and thus the phosphorylation of fructose is a problem in the selection strains based on *E. coli* K-12 MG1655. A perfect selection strain would only have one bottleneck and that is the target enzyme. The constitutive overexpression of an active fructokinase could overcome this second bottleneck and improve the selection strain. Nonetheless, the choice of the right fructokinase and promoter is not an easy task and would be very time-consuming (personal communication Joeri Beauprez). Since also a screening assay was developed to detect tagatose 4-epimerase activity, this optimization of the selection strain was omitted.

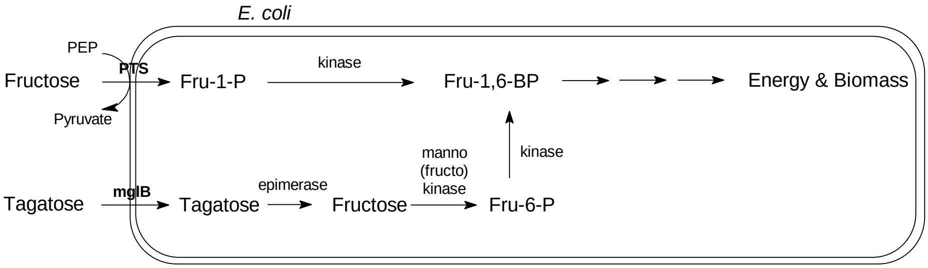


Figure IX.7 Normal fructose utilization goes via the PTS system, immediately phosphorylating it. However, in the selection strains fructose would be formed inside the cell, needing prior phosphorylating to fructose-6-phosphate by manno(fructo)kinase. Since only low fructokinase activity is found in *E. coli*, growth and selection are hampered

Another option to overcome the fructose phosphorylation problem as well as the possibly low substrate levels is by expressing the epimerase in the periplasm. Tagatose concentration in the periplasm will be higher than that measured in the cytoplasm since the inner membrane is much more disciminating than the outer membrane [169]. Periplasmatic expression can be achieved by expressing the target enzyme with a N-terminal signal peptide in order to transport the (pro)enzyme from the cytoplasm to the periplasm. Two systems are available in *E. coli* namely the secretion pathway (Sec) and the twin-arginine translocation system (Tat) [173, 265, 271]. Nevertheless, selection of the optimal signal peptide is crucial and is mostly performed by trial-and-error [54]. Furthermore, informatics tools are available to help with the selection of the appropriate signal peptide, for example SignalP [24, 25, 79, 186, 203]. Periplasmatic expression would also facilitate further uptake of the formed fructose since the fructose present in the periplasm will easily be taken up by the cell via its PTS system and in this way avoid the need of cytosolic MAK activity.

On the other hand, the screening assay should be able to screen as much mutants as possible in a minimum of time. In the setup of the presented research (search for a tagatose 4-epimerase), the application of Fluorescence-Activated Cell Sorting (FACS) could be an option [99]. Using FACS, 10^8 mutants can be tested per day [281]. Nevertheless, as the technique FACS needs fluorescence to be able to detect and separate cells, the desired activity should be directly coupled to a fluorescent outcome. If the product of the reaction can be used to induce protein expression, it can be used to induce for instance green fluorescent protein (GFP) [57]. Fluorescence can then be used to sort cells via FACS. Despite the fact that this idea is written down in one sentence, it is not that simple. At first, one must have a protein expression system that is only triggered when the target enzyme is active and thus the product is formed. For a tagatose 4-epimerase, this would mean that the formed fructose may not be consumed but must trigger expression. As such, there is the need for a fructose induced promoter, a fructose sensitive repressor or another fructose sensitive biomarker. When no fructose is present, expression is not induced or repressed by binding of the repressor, meaning that no GFP (fluorescence) is present in the cells. If active epimerase would be expressed, fructose would be formed that can in turn induce GFP expression or bind to the fructose sensitive repressor enabling GFP expression. Again, the question is where or how to find such a ‘fructose sensitive promoter/repressor system’. Also here, protein engineering and bio-informatics can be used as tools to redesign existing promoter/repressor systems. Starting points could be the lac repressor system or the L-arabinose-inducible araBAD promoter (P_{BAD}) that induce protein expression in the presence of lactose (and IPTG) or L-arabinose, respectively. Nevertheless, it will take a lot of work and time to convert such a system into a desired and useful system.

REFERENCES

1. Aerts D, Verhaeghe T, De Mey M, Desmet T, Soetaert W (2011) A constitutive expression system for high-throughput screening. Eng Life Sci 11:10-19

2. Ahmed Z (2001) Production of natural and rare pentoses using microorganisms and their enzymes. Electron J Biotechnol 4:103-111

3. Airoldi C, Merlo S, Nicotra F (2010) Synthesis of 3-deoxy-d-threopentofuranose 5-phosphate, a substrate of arabinose 5-phosphate isomerase. J Carbohydr Chem 29:30-38

4. Alcalde M, Ferrer M, Plou FJ, Ballesteros A (2006) Environmental biocatalysis: from remediation with enzymes to novel green processes. Trends Biotechnol 24:281-287

5. Allard STM, Giraud MF, Naismith JH (2001) Epimerases: structure, function and mechanism. Cell Mol Life Sci 58:1650-1665

6. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410

7. Anderson RL (1966) L-ribulose-5-phosphate. Methods Enzymol 9:48-51

8. Andre C, Demuynck C, Gefflaut T, Guerard C, Hecquet L, Lemaire M, Bolte J (1998) Fructose-1,6-bisphosphate aldolase and transketolase: Complementary tools for the de novo syntheses of monosaccharides and analogues. J Mol Catal B-Enzym 5:113-118

9. Armstrong LM, Luecke KJ, Bell LN (2009) Consumer evaluation of bakery product flavour as affected by incorporating the prebiotic tagatose. Int J Food Sci Technol 44:815-819

10. Arnold EC, Silady PJ (1997) Use of D-allose as an immunosuppressive agent. US Patent 5620960

11. Arnold FH (2001) Combinatorial and computational challenges for biocatalyst design. Nature 409:253-257

12. Arnold FH, Georgiou G (2003) Directed Evolution: Library Creation. Methods in Molecular Biology

13. Austin WC, Humoller FL (1934) The preparation of L-ribose. J Am Chem Soc 56:1152-1153

14. Austin WC, Humoller FL (1934) The preparation of two new crystalline aldohexoses, L-allose and L-altrose, from L-ribose by the cyanohydrin reaction. J Am Chem Soc 56:1153-1155

15. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2:

16. Baek SH, Park SJ, Lee HG (2010) D-psicose, a sweet monosaccharide, ameliorate hyperglycemia, and dyslipidemia in C57BL/6J db/db mice. J Food Sci 75:H49-H53

17. BCC Research (2007) Sugars and Sweeteners in Processed Foods and Beverages (Report highlights of Report FOD018C). October, 24 2012 on http://www.bccresearch.com

18. BCC Research (2012) 2011 Food and Beverage Research Review (Report highlights of Report FOD048A). 24 October 2012 on http://www.bccresearch.com

19. Beadle JR, Saunders JP, Wayda TJJ (1992) Process for manufacturing tagatose. US Patent 5078796

20. Beerens K, Desmet T, Soetaert W (2012) Enzymes for the biocatalytic production of rare sugars. J Ind Microbiol Biotechnol 39:823-834

21. Behe MJ (2010) Experimental evolution, loss-of-function mutations, and "the first rule of adaptive evolution". Q Rev Biol 85:419-445

22. Behe MJ, Snoke DW (2004) Simulating evolution by gene duplication of protein features that require multiple amino acid residues. Protein Sci 13:2651-2664

23. Belanger M, Burrows LL, Lam JS (1999) Functional analysis of genes responsible for the synthesis of the B-band O antigen of *Pseudomonas aeruginosa* serotype O6 lipopolysaccharide. Microbiology-Uk 145:3505-3521

24. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783-795

25. Bendtsen JD, Nielsen H, Widdick D, Palmer T, Brunak S (2005) Prediction of twin-arginine signal peptides. BMC Bioinformatics 6:167

26. Bengoechea JA, Pinta E, Salminen T, Oertelt C, Holst O, Radziejewska-Lebrecht J, Piotrowska-Seget Z, Venho R, Skurnik M (2002) Functional characterization of gne (UDP-N-acetylglucosamine-4-epimerase), Wzz (chain length determinant), and Wzy (O-antigen polymerase) of *Yersinia enterocolitica* serotype O : 8. J Bacteriol 184:4277-4287

27. Bernatchez S, Szymanski CM, Ishiyama N, Li JJ, Jarrell HC, Lau PC, Berghuis AM, Young NM, Wakarchuk WW (2005) Single bifunctional UDP-GlcNAc/Glc 4-epimerase supports the synthesis of three cell surface glycoconjugates in *Campylobacter jejuni*. J Biol Chem 280:4792-4802

28. Bertelsen H, Andersen H, Tvede M (2001) Fermentation of D-tagatose by human intestinal bacteria and dairy lactic acid bacteria. Microb Ecol Health Dis 13:87-95

29. Bertelsen H, Jensen BB, Buemann B (1999) D-tagatose - A novel low-calorie bulk sweetener with prebiotic properties. Low-Calorie Sweeteners: Present and Future 85:98-109
30. Bertland A, Bugge B, Kalckar HM (1966) Fluorescence enhancement of uridine diphosphogalactose 4-epimerase induced by specific sugars. Arch Biochem Biophys 116:280-283
31. Bhatt VS, Guan W, Xue M, Yuan H, Wang PG (2011) Insights into role of the hydrogen bond networks in substrate recognition by UDP-GaINAc 4-epimerases. Biochem Biophys Res Commun 412:232-237
32. Bhatt VS, Guo C-y, Guan W, Zhao G, Yi W, Liu Z-j, Wang PG (2011) Altered architecture of substrate binding region defines the unique specificity of UDP-GalNAc 4-epimerases. Protein Sci 20:856-866
33. Bhuiyan SH, Itami Y, Takada G, Izumori K (1999) Preparation of L-talose and D-gulose from L-tagatose and D-sorbose, respectively, using immobilized L-rhamnose isomerase. J Biosci Bioeng 88:567-570
34. Blackburn P, Ferdinand W (1976) The concerted inactivation of *Escherichia coli* uridine diphosphate galactose 4-epimerase by sugar nucleotide together with a free sugar. Biochem J 155:225-229
35. Blankenfeldt W, Kerr ID, Giraud MF, McMiken HJ, Leonard G, Whitfield C, Messner P, Graninger M, Naismith JH (2002) Variation on a theme of SDR: dTDP-6-deoxy-L-lyxo-4-hexulose reductase (RmID) shows a new Mg^{2+}-dependent dimerization mode. Structure 10:773-786
36. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, ColladoVides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277:1453-1462
37. Bochner BR, Savageau MA (1977) Generalized indicator plate for genetic, metabolic, and taxonomic studies with microorganisms. Appl Environ Microbiol 33:434-444
38. Bommarius AS, Blum JK, Abrahamson MJ (2011) Status of protein engineering for biocatalysts: how to design an industrially useful biocatalyst. Curr Opin Chem Biol 15:194-200
39. Bornscheuer UT (1998) Directed evolution of enzymes. Angew Chem, Int Ed 37:3105-3108
40. Bornscheuer UT, Pohl M (2001) Improved biocatalysts by directed evolution and rational protein design. Curr Opin Chem Biol 5:137-143
41. Buemann B, Toubro S, Raben A, Blundell J, Astrup A (2000) The acute effect of D-tagatose on food intake in human subjects. Br J Nutr 84:227-231
42. Burke JR, Frey PA (1993) The importance of binding energy in catalysis of hydride transfer by UDP-galactose 4-epimerase: a 13C and 15N NMR and kinetic study. Biochemistry 32:13220-13230
43. Cabirol FL, Tan PL, Tay B, Cheng S, Hanefeld U, Sheldon RA (2008) *Linum usitatissimum* hydroxynitrile lyase cross-linked enzyme aggregates: a recyclable enantioselective catalyst. Adv Synth Catal 350:2329-2338
44. Cerdobbel A, De Winter K, Aerts D, Kuipers R, Joosten H-J, Soetaert W, Desmet T (2011) Increasing the thermostability of sucrose phosphorylase by a combination of sequence- and structure-based mutagenesis. Protein Eng Des Sel 24:829-834
45. Cerdobbel A, De Winter K, Desmet T, Soetaert W (2010) Sucrose phosphorylase as cross-linked enzyme aggregate: Improved thermal stability for industrial applications. Biotechnol J 5:1192-1197
46. Cerdobbel A, Desmet T, De Winter K, Maertens J, Soetaert W (2010) Increasing the thermostability of sucrose phosphorylase by multipoint covalent immobilization. J Biotechnol 150:125-130
47. Chelain E, Floch O, Czernecki S (1995) New synthesis of L-ribofuranose from L-xylose. J Carbohydr Chem 14:1251-1256
48. Chen CY, Georgiev I, Anderson AC, Donald BR (2009) Computational structure-based redesign of enzyme activity. Proc Natl Acad Sci U S A 106:3764-3769
49. Cheng LF, Mu WM, Jiang B (2010) Thermostable L-arabinose isomerase from Bacillus stearothermophilus IAM 11001 for D-tagatose production: gene cloning, purification and characterisation. J Sci Food Agric 90:1327-1333
50. Cheng LF, Mu WM, Zhang T, Jiang B (2010) An L-arabinose isomerase from Acidothermus cellulolytics ATCC 43068: cloning, expression, purification, and characterization. Appl Microbiol Biotechnol 86:1089-1097
51. Chevrier C, Le Nouen D, Defoin A, Tarnus C (2006) Synthesis of amino-L-lyxose phosphonates as fucosyl-phosphate mimics. Eur J Org Chem 2384-2392
52. Cho EA, Lee DW, Cha YH, Lee SJ, Jung HC, Pan JG, Pyun YR (2007) Characterization of a novel D-lyxose isomerase from Cohnella laevoribosii RI-39 sp nov. J Bacteriol 189:1655-1663
53. Choi JG, Ju YH, Yeom SJ, Oh DK (2011) Improvement in the thermostability of D-psicose 3-epimerase from Agrobacterium tumefaciens by random and site-directed mutagenesis. Appl Environ Microbiol 77:7316-7320

54. Choi JH, Lee SY (2004) Secretory and extracellular production of recombinant proteins using *Escherichia coli*. Appl Microbiol Biotechnol 64:625-635

55. Chouayekh H, Bejar W, Rhimi M, Jelleli K, Mseddi M, Bejar S (2007) Characterization of an L-arabinose isomerase from the Lactobacillus plantarum NC8 strain showing pronounced stability at acidic pH. FEMS Microbiol Lett 277:260-267

56. Cisar JO, Kolenbrander PE, McIntire FC (1979) Specificity of coaggregation reactions between human oral streptococci and strains of *Actinomyces viscosus* or *Actinomyces naeslundii*. Infect Immun 24:742-752

57. Cormack BP, Valdivia RH, Falkow S (1996) FACS-optimized mutants of the green fluorescent protein (GFP). Gene 173:33-38

58. Creuzenet C, Belanger M, Wakarchuk WW, Lam JS (2000) Expression, purification, and biochemical characterization of WbpP, a new UDP-GlcNAc C4 epimerase from *Pseudomonas aeruginosa* serotype O6. J Biol Chem 275:19060-19067

59. Dalby PA (2003) Optimising enzyme function by directed evolution. Curr Opin Struct Biol 13:500-505

60. Darwin C (1859) On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. London: John Murray, 502 p.

61. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640-6645

62. Davis JE, Nolan LD, Frey PA (1974) UMP-dependent reduction of UDP-galactose 4-epimerase-NAD+ complex by sodium cyanoborohydride. Biochim Biophys Acta 334:442-447

63. Davis L, Lee N, Glaser L (1972) Mechanism of pentose phosphate epimerases. J Biol Chem 247:5862-5866

64. De Groeve MRM, De Baere M, Hoflack L, Desmet T, Vandamme EJ, Soetaert W (2009) Creating lactose phosphorylase enzymes by directed evolution of cellobiose phosphorylase. Protein Eng Des Sel 22:393-399

65. De Mey M, Maertens J, Lequeux GJ, Soetaert WK, Vandamme EJ (2007) Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. BMC Biotechnol 7:34

66. De Muynck C, Pereira C, Soetaert W, Vandamme E (2006) Dehydrogenation of ribitol with *Gluconobacter oxydans*: Production and stability of L-ribulose. J Biotechnol 125:408-415

67. De Muynck C, Pereira CSS, Naessens M, Parmentier S, Soetaert W, Vandamme EJ (2007) The genus *Gluconobacter oxydans*: Comprehensive overview of biochemistry and biotechnological applications. Crit Rev Biotechnol 27:147-171

68. De Muynck C, Van der Borght J, De Mey M, De Maeseneire SL, Van Bogaert INA, Beauprez J, Soetaert W, Vandamme E (2007) Development of a selection system for the detection of L-ribose isomerase expressing mutants of *Escherichia coli*. Appl Microbiol Biotechnol 76:1051-1057

69. De Winter K, Soetaert W, Desmet T (2012) An Imprinted Cross-Linked Enzyme Aggregate (iCLEA) of Sucrose Phosphorylase: Combining Improved Stability with Altered Specificity. Int J Mol Sci 13:11333-11342

70. DeLano, WL (2002) The PyMOL Molecular Graphics System. 20 September 2012 on http://www.pymol.org

71. Demendi M, Ishiyama N, Lam JS, Berghuis AM, Creuzenet C (2005) Towards a better understanding of the substrate specificity of the UDP-N-acetylglucosamine C4 epimerase WbpP. Biochem J 389:173-180

72. Demir AS, Talpur FN, Sopaci SB, Kohring GW, Celik A (2011) Selective oxidation and reduction reactions with cofactor regeneration mediated by galactitol-, lactate-, and formate dehydrogenases immobilized on magnetic nanoparticles. J Biotechnol 152:176-183

73. deOrellano MEE, deAscheri AMP, deTosetti MIS, Segovia R (1996) An epimerasic activity on galactose induced by xylose in *Kluyveromyces* sp. Appl Biochem Biotechnol 60:107-113

74. Deupree JD, Wood WA (1972) L-Ribulose 5-phosphate 4-epimerase from Aerobacter aerogenes. Evidence for a role of divalent metal ions in the epimerization reaction. J Biol Chem 247:3093-3097

75. Dobbs CM, Bell LN (2010) Storage stability of tagatose in buffer solutions of various compositions. Food Res Int 43:382-386

76. Dondoni A, Marra A, Massi A (1997) Carbohydrate homologation by the use of 2-(trimethylsilyl)thiazole. Preparative scale synthesis of rare sugars: L-gulose, L-idose, and the disaccharide subunit of bleomycin A(2). J Org Chem 62:6261-6267

77. Donner TW, Wilber JF, Ostrowski D (1999) D-tagatose, a novel hexose: acute effects on carbohydrate tolerance in subjects with and without type 2 diabetes. Diabetes Obes Metab 1:285-291

78. Dreyer MK, Schulz GE (1996) Refined high-resolution structure of the metal-ion dependent L-fuculose-1-phosphate aldolase (class II) from *Escherichia coli*. Acta Crystallogr Sect D Biol Crystallogr 52:1082-1091

79. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2:953-971

80. Espinosa I, Fogelfeld L (2010) Tagatose: from a sweetener to a new diabetic medication? Expert Opin Invest Drugs 19:285-294

81. Fechter MH, Stutz AE (1999) Synthetic applications of glucose isomerase: isomerisation of C-5-modified (2R,3R,4R)-configured hexoses into the corresponding 2-ketoses. Carbohydr Res 319:55-62

82. Fessner WD, Gosse C, Jaeschke G, Eyrisch O (2000) Enzymes in organic synthesis, 15 - Short enzymatic synthesis of L-fucose analogs. Eur J Org Chem 125-132

83. Frirdich E, Whitfield C (2005) Characterization of Gla(KP), a UDP-galacturonic acid C4-epimerase from *Klebsiella pneumoniae* with extended substrate specificity. J Bacteriol 187:4104-4115

84. Gauger AK, Axe DD (2011) The evolutionary accessibility of new enzyme functions a case study from the biotin pathway. BIO-Complexity 2011:1-17

85. Gerratana B, Cleland WW, Frey PA (2001) Mechanistic roles of Thr134, Tyr160, and Lys 164 in the reaction catalyzed by dTDP-glucose 4,6-dehydratase. Biochemistry 40:9187-9195

86. Goodwin TE, Cousins KR, Crane HM, Eason PO, Freyaldenhoven TE, Harmon CC, King BK, LaRocca CD, Lile RL, Orlicek SG, Pelton RW, Shedd OL, Swanson JS, Thompson JW (1998) Synthesis of two new maytansinoid model compounds from carbohydrate precursors. J Carbohydr Chem 17:323-339

87. Graber M, Morin A, Duchiron F, Monsan PF (1988) Microbial polysaccharides containing 6-deoxysugars. Enzyme Microb Technol 10:198-206

88. Granstrom TB, Izumori K, Leisola M (2007) A rare sugar xylitol. Part I: the biochemistry and biosynthesis of xylitol. Appl Microbiol Biotechnol 74:277-281

89. Granstrom TB, Izumori K, Leisola M (2007) A rare sugar xylitol. Part II: biotechnological production and future applications of xylitol. Appl Microbiol Biotechnol 74:273-276

90. Granstrom TB, Takata G, Morimoto K, Leisola M, Izumori K (2005) L-xylose and L-lyxose production from xylitol using *Alcaligenes* 701B strain and immobilized L-rhamnose isomerase enzyme. Enzyme Microb Technol 36:976-981

91. Granstrom TB, Takata G, Tokuda M, Izumori K (2004) Izumoring: A novel and complete strategy for bioproduction of rare sugars. J Biosci Bioeng 97:89-94

92. Gullapalli P, Shiji T, Rao D, Yoshihara A, Morimoto K, Takata G, Fleet GWJ, Izumori K (2007) Bioproduction of a novel sugar 1-deoxy-L-fructose by Enterobacter aerogenes IK7; isomerization of a 6-deoxyhexose to a 1-deoxyhexose. Tetrahedron: Asymmetry 18:1995-2000

93. Gullapalli P, Yoshihara A, Morimoto K, Rao D, Akimitsu K, Jenkinson SF, Fleet GWJ, Izumori K (2010) Conversion of L-rhamnose into ten of the sixteen 1-and 6-deoxyketohexoses in water with three reagents: D-tagatose-3-epimerase equilibrates C3 epimers of deoxyketoses. Tetrahedron Lett 51:895-898

94. Gumina G, Song GY, Chu CK (2001) L-nucleosides as chemotherapeutic agents. FEMS Microbiol Lett 202:9-15

95. Guo HJ, Li L, Wang PG (2006) Biochemical characterization of UDP-GlcNAc/Glc 4-epimerase from *Escherichia coli* O86:B7. Biochemistry 45:13760-13768

96. Hausler H, Stutz AE (2001) D-xylose (D-glucose) isomerase and related enzymes in carbohydrate synthesis. Top Curr Chem 215:77-114

97. Hayashi N, Iida T, Yamada T, Okuma K, Takehara I, Yamamoto T, Yamada K, Tokuda M (2010) Study on the postprandial blood glucose suppression effect of D-psicose in borderline diabetes and the safety of long-term ingestion by normal human subjects. Biosci Biotechnol Biochem 74:510-519

98. Hernandez-Montalvo V, Martinez A, Hernandez-Chavez G, Bolivar F, Valle F, Gosset G (2003) Expression of galP and glk in a *Escherichia coli* PTS mutant restores glucose transport and increases glycolytic flux to fermentation products. Biotechnol Bioeng 83:687-694

99. Herzenberg LA, Parks D, Sahaf B, Perez O, Roederer M (2002) The history and future of the fluorescence activated cell sorter and flow cytometry: A view from Stanford. Clin Chem 48:1819-1827

100. Hofmann C, Boll R, Heitmann B, Hauser G, Durr C, Frerich A, Weitnauer G, Glaser SJ, Bechthold A (2005) Genes encoding enzymes responsible for biosynthesis of L-lyxose and attachment of eurekanate during avilamycin biosynthesis. Chem Biol 12:1137-1143

101. Holden HM, Rayment I, Thoden JB (2003) Structure and function of enzymes of the Leloir pathway for galactose metabolism. J Biol Chem 278:43885-43888

102. Hoshikawa H, Indo K, Mori T, Mori N (2011) Enhancement of the radiation effects by D-allose in head and neck cancer cells. Cancer Lett 306:60-66

103. Hossain MA, Izuishi K, Tokuda M, Izumori K, Maeta H (2004) D-Allose has a strong suppressive effect against ischemia/reperfusion injury: a comparative study with allopurinol and superoxide dismutase. J Hepatobiliary Pancreat Surg 11:181-189

104. Hossain MA, Wakabayashi H, Goda F, Kobayashi S, Maeba T, Maeta H (2000) Effect of the immunosuppressants FK506 and D-allose on allogenic orthotopic liver transplantation in rats. Transplant Proc 32:2021-2023

105. Hricoviniova Z, Hricovini M, Petrus L (2000) Molybdic acid-catalysed isomerization of D-ribulose and D-xylulose to the corresponding 2-C-(hydroxymethyl)-D-tetroses. J Carbohydr Chem 19:827-836

106. Hung SC, Wang CC, Thopate SR (2000) Efficient synthesis of L-altrose and L-mannose. Tetrahedron Lett 41:3119-3122

107. Huwig A, Emmel S, Giffhorn F (1996) Preparation of D-sorbose from L-glucitol by bioconversion with *Pseudomonas* sp Ac. Carbohydr Res 281:183-186

108. Huwig A, Emmel S, Jakel G, Giffhorn F (1997) Enzymatic synthesis of L-tagatose from galactitol with galactitol dehydrogenase from Rhodobacter sphaeroides D. Carbohydr Res 305:337-339

109. Ishida Y, Kamiya T, Itoh H, Kimura Y, Izumori K (1997) Cloning and characterization of the D-tagatose 3-epimerase gene from *Pseudomonas cichorii* ST-24. J Ferment Bioeng 83:529-534

110. Ishiyama N, Creuzenet C, Lam JS, Berghuis AM (2004) Crystal structure of WbpP, a genuine UDP-N-acetylglucosamine 4-epimerase from *Pseudomonas aeruginosa* - Substrate specificity in UDP-hexose 4-epimerases. J Biol Chem 279:22635-22642

111. International Society of Rare Sugars (2011) Rare Sugar Congress in Kagawa, Japan

112. Ito S (2009) Features and applications of microbial sugar epimerases. Appl Microbiol Biotechnol 84:1053-1060

113. Itoh H, Izumori K (1996) Enzymatic production of L-tagatose and L-fructose from L-sorbose and L-psicose, respectively. J Ferment Bioeng 81:351-353

114. Itoh H, Okaya H, Khan AR, Tajima S, Hayakawa S, Izumori K (1994) Purification and charachterization of D-tagatose 3-epimerase from *Pseudomonas* sp. ST-24. Biosci Biotechnol Biochem 58:2168-2171

115. Itoh H, Sato T, Takeuchi T, Khan AR, Izumori K (1995) Preparation of D-sorbose from D-tagatose by immobilized D-tagatose 3-epimerase. J Ferment Bioeng 79:184-185

116. Izumori K (2006) Izumoring: A strategy for bioproduction of all hexoses. J Biotechnol 124:717-722

117. Izumori K, Miyoshi T, Tokuda S, Yamabe K (1984) Production of d-tagatose from dulcitol by *Arthrobacter globiformis*. Appl Environ Microbiol 48:1055-1057

118. Izumori K, Tsuzaki K (1988) Production of d-tagatose from d-galactitol by *Mycobacterium smegmatis*. J Ferment Technol 66:225-227

119. Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF, Hilvert D, Houk KN, Stoddard BL, Baker D (2008) De novo computational design of retro-aldol enzymes. Science 319:1387-1391

120. Johnson AE, Tanner ME (1998) Epimerization via carbon-carbon bond cleavage. L-ribulose-5-phosphate 4-epimerase as a masked class II aldolase. Biochemistry 37:5746-5754

121. Jokela J, Pastinen O, Leisola M (2002) Isomerization of pentose and hexose sugars by an enzyme reactor packed with cross-linked xylose isomerase crystals. Enzyme Microb Technol 31:67-76

122. Jones NA, Rao D, Yoshihara A, Gullapalli P, Morimoto K, Takata G, Hunter SJ, Wormald MR, Dwek RA, Izumori K, Fleet GWJ (2008) Green syntheses of new 2-C-methyl aldohexoses and 5-C-methyl ketohexoses: D-tagatose-3-epimerase (DTE)-a promiscuous enzyme. Tetrahedron: Asymmetry 19:1904-1918

123. Ju YH, Oh DK (2010) Characterization of a recombinant l-fucose isomerase from Caldicellulosiruptor saccharolyticus that isomerizes l-fucose, d-arabinose, d-altrose, and l-galactose. Biotechnol Lett 32:299-304

124. Jumppanen J, Nurmi J, Pastinen O (2000) Process for the continuous production of high purity L-ribose. US Patent Number 6140498

125. Kallnik V, Schultz C, Schweiger P, Deppenmeier U (2011) Properties of recombinant Strep-tagged and untagged hyperthermophilic D-arabitol dehydrogenase from Thermotoga maritima. Appl Microbiol Biotechnol 90:1285-1293

126. Karp PD, Riley M, Saier M, Paulsen IT, Paley SM, Pellegrini-Toole A (2000) The EcoCyc and MetaCyc databases. Nucleic Acids Res 28:56-59

127. Kazlauskas RJ, Bornscheuer UT (2009) Finding better protein engineering strategies. Nat Chem Biol 5:526-529

128. Keseler IM, Collado-Vides J, Santos-Zavaleta A, Peralta-Gil M, Gama-Castro S, Muniz-Rascado L, Bonavides-Martinez C, Paley S, Krummenacker M, Altman T, Kaipa P, Spaulding A, Pacheco J, Latendresse

M, Fulcher C, Sarker M, Shearer AG, Mackie A, Paulsen I, Gunsalus RP, Karp PD (2011) EcoCyc: a comprehensive database of *Escherichia coli* biology. Nucleic Acids Res 39:D583-D590

129. Kim HJ, Hyun EK, Kim YS, Lee YJ, Oh DK (2006) Characterization of an *Agrobacterium tumefaciens* D-psicose 3-epimerase that converts D-fructose to D-psicose. Appl Environ Microbiol 72:981-985

130. Kim HJ, Kang SY, Park JJ, Kim P (2011) Novel activity of UDP-galactose-4-epimerase for free monosaccharide and activity improvement by active site-saturation mutagenesis. Appl Biochem Biotechnol 163:444-451

131. Kim HJ, Uhm TG, Kim SB, Kim P (2010) Escherichia coli arabinose isomerase and Staphylococcus aureus tagatose-6-phosphate isomerase: Which is a better template for directed evolution of non-natural substrate isomerization? J Microbiol Biotechnol 20:1018-1021

132. Kim JH, Lim BC, Yeom SJ, Kim YS, Kim FJ, Lee JK, Lee SH, Kim SW, Oh DK (2008) Differential selectivity of the *Escherichia coli* cell membrane shifts the equilibrium for the enzyme-catalyzed isomerization of galactose to tagatose. Appl Environ Microbiol 74:2307-2313

133. Kim N-H, Kim H-J, Kang D-IL, Jeong K-W, Lee J-K, Kim Y, Oh D-K (2008) Conversion shift of D-fructose to D-psicose for enzyme-catalyzed epimerization by addition of borate. Appl Environ Microbiol 74:3008-3013

134. Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM (2006) MUSTANG: A multiple structural alignment algorithm. Proteins Struct Funct Bioinformat 64:559-574

135. Konopka JM, Halkides CJ, Vanhooke JL, Gorenstein DG, Frey PA (1989) UDP-galactose 4-epimerase. Phosphorus-31 nuclear magnetic resonance analysis of NAD+ and NADH bound at the active site. Biochemistry 28:2645-2654

136. Kornberg HL (2001) Routes for fructose utilization by *Escherichia coli*. J Mol Microbiol Biotechnol 3:355-359

137. Kotake T, Takata R, Verma R, Takaba M, Yamaguchi D, Orita T, Kaneko S, Matsuoka K, Koyama T, Reiter W-D, Tsumuraya Y (2009) Bifunctional cytosolic UDP-glucose 4-epimerases catalyse the interconversion between UDP-D-xylose and UDP-L-arabinose in plants. Biochem J 424:169-177

138. Kovacs J, Pinter I, Koll P (1995) Direct transformation of D-idose and D-altrose with potassium cyanate into cyclic carbamates of derived glycosylamines. Carbohydr Res 272:255-262

139. Krieger E, Koraimann G, Vriend G (2002) Increasing the precision of comparative models with YASARA NOVA - a self-parameterizing force field. Proteins: Struct, Funct, Genet 47:393-402

140. Kroemer M, Merkel I, Schulz GE (2003) Structure and catalytic mechanism of L-rhamnulose-1-phosphate aldolase. Biochemistry 42:10560-10568

141. Kroemer M, Schulz GE (2002) The structure of L-rhamnulose-1-phosphate aldolase (class II) solved by low-resolution SIR phasing and 20-fold NCS averaging. Acta Crystallogr Sect D Biol Crystallogr 58:824-832

142. Kwon HJ, Yeom SJ, Park CS, Oh DK (2010) Substrate specificity of a recombinant D-lyxose isomerase from Providencia stuartii for monosaccharides. J Biosci Bioeng 110:26-31

143. Kwon SC, Cho CS, Shim SC, Kim TJ (1999) Catalytic formation of cyclic carbonates and carbamates by [Cu(1)](BF4)(2) (1 = 2,5,19,22-tetraaza[6,6](1,1 ')ferrocenophane-1,5-diene). Bull Korean Chem Soc 20:103-105

144. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and clustal X version 2.0. Bioinformatics 23:2947-2948

145. Leang K, Sultana I, Takada G, Izumori K (2003) A novel bioconversion of L-fructose to L-glucose by Klebsiella pneumoniae. J Biosci Bioeng 95:310-312

146. Leang K, Takada G, Fukai Y, Morimoto K, Granstrom TB, Izumori K (2004) Novel reactions of L-rhanmose isomerase from Pseudomonas stutzeri and its relation with D-xylose isomerase via substrate specificity. Biochim Biophys Acta Gen Subj 1674:68-77

147. Leang K, Takada G, Ishimura A, Okita M, Izumori K (2004) Cloning, nucleotide sequence, and overexpression of the L-rhamnose isomerase gene from Pseudomonas stutzeri in Escherichia coli. Appl Environ Microbiol 70:3298-3304

148. Lee LV, Poyner RR, Vu MV, Cleland WW (2000) Role of metal ions in the reaction catalyzed by L-ribulose-5-phosphate 4-epimerase. Biochemistry 39:4821-4830

149. Lee LV, Vu MV, Cleland WW (2000) C-13 and deuterium isotope effects suggest an aldol cleavage mechanism for L-ribulose-5-phosphate 4-epimerase. Biochemistry 39:4808-4820

150. Leemhuis H, Kelly RM, Dijkhuizen L (2009) Directed Evolution of Enzymes: Library Screening Strategies. IUBMB Life 61:222-228

151. Lerner LM, Mennitt G (1994) A new synthesis of L-talose and preparation of its adenine nucleosides. Carbohydr Res 259:191-200

152. Lesk AM (1995) NAD-binding domains of dehydrogenases. Curr Opin Struct Biol 5:775-783

153. Levin GV (2000) Increased fertility and improved fetal development. US Patent 6225452

154. Levin GV (2002) Tagatose, the new GRAS sweetener and health product. J Med Food 5:23-36

155. Levin GV, Zehner LR, Saunders JP, Beadle JR (1995) Sugar substitutes - their energy values, bulk charachteristics, and potential health benefits. Am J Clin Nutr 62:S1161-S1168

156. Li ZJ, Cai L, Qi QS, Styslinger TJ, Zhao GH, Wang PG (2011) Synthesis of rare sugars with L-fuculose-1-phosphate aldolase (FucA) from Thermus thermophilus HB8. Bioorg Med Chem Lett 21:5084-5087

157. Lichenstein HS, Hamilton EP, Lee N (1987) Repression and catabolite gene activation in the araBAD operon. J Bacteriol 169:811-823

158. Lim BC, Kim HJ, Oh DK (2007) High production of D-tagatose by the addition of boric acid. Biotechnol Prog 23:824-828

159. Lim YR, Oh DK (2011) Microbial metabolism and biotechnological production of d-allose. Appl Microbiol Biotechnol 91:229-235

160. Liu HW, Thorson JS (1994) Pathways and mechanisms in the biogenesis of novel deoxysugars by bacteria. Annu Rev Microbiol 48:223-256

161. Liu YJ, Arabshahi A, Frey PA (2000) Rate enhancements brought about by uridine nucleotides in the reduction of NAD(+) at the active site of UDP-galactose 4-epimerase. Bioorg Chem 28:29-37

162. Liu YJ, Thoden JB, Kim J, Berger E, Gulick AM, Ruzicka FJ, Holden HM, Frey PA (1997) Mechanistic roles of tyrosine 149 and serine 124 in UDP-galactose 4-epimerase from *Escherichia coli*. Biochemistry 36:10675-10684

163. Loewus FA, Saito K, Suto RK, Maring E (1995) Conversion of D-arabinose to D-erythroascorbic acid and oxalic acid in *Sclerotinia sclerotiorum*. Biochem Biophys Res Commun 212:196-203

164. Lu Y, Levin GV, Donner TW (2008) Tagatose, a new antidiabetic and obesity control drug. Diabetes Obes Metab 10:109-134

165. Luecke KJ, Bell LN (2010) Thermal Stability of Tagatose in Solution. J Food Sci 75:C346-C351

166. Luo Y, Samuel J, Mosimann SC, Lee JE, Tanner ME, Strynadka NCJ (2001) The structure of L-ribulose-5-phosphate 4-epimerase: An aldolase-like platform for epimerization. Biochemistry 40:14763-14771

167. Lutz S (2010) Beyond directed evolution -semi-rational protein engineering and design. Curr Opin Biotechnol 21:734-743

168. Ma TW, Lin JS, Newton MG, Cheng YC, Chu CK (1997) Synthesis and anti-hepatitis B virus activity of 9-(2-deoxy-2-fluoro-beta-L-arabinofuranosyl)purine nucleosides. J Med Chem 40:2750-2754

169. Madigan MT, Martinko JM, Dunlap PV, Clark DP (2009) Biology of Microorganisms. Pearson Benjamin Cummings, 1061 p.

170. Manzoni M, Rollini M, Bergomi S (2001) Biotransformation of D-galactitol to tagatose by acetic acid bacteria. Process Biochem 36:971-977

171. Mathe C, Gosselin G (2006) L-Nucleoside enantiomers as antivirals drugs: A mini-review. Antivir Res 71:276-281

172. Menavuvu BT, Poonperm W, Takeda K, Morimoto K, Granstrom TB, Takada G, Izumori K (2006) Novel substrate specificity of D-arabinose isomerase from Klebsiella pneumoniae and its application to production of D-altrose from D-psicose. J Biosci Bioeng 102:436-441

173. Mergulhao FJM, Summers DK, Monteiro GA (2005) Recombinant protein secretion in *Escherichia coli*. Biotechnol Adv 23:177-202

174. Miller BG, Raines RT (2005) Reconstitution of a defunct glycolytic pathway via recruitment of ambiguous sugar kinases. Biochemistry 44:10776-10783

175. Mitani T, Hoshikawa H, Mori T, Hosokawa T, Tsukamoto I, Yamaguchi F, Kamitori K, Tokuda M, Mori N (2009) Growth inhibition of head and neck carcinomas by D-allose. Head Neck 31:1049-1055

176. Mizanur RM, Takata G, Izumori K (2001) Cloning and characterization of a novel gene encoding L-ribose isomerase from Acinetobacter sp strain DL-28 in Escherichia coli. Biochim Biophys Acta Gene Struct Expression 1521:141-145

177. Moore MC (2006) Drug evaluation: Tagatose in the treatment of type 2 diabetes and obesity. Curr Opin Investig Drugs 7:924-935

178. Moran EJ, Tellew JE, Zhao ZC, Armstrong RW (1993) Dehydroamino acid derivatives from D-arabinose and L-serine: synthesis of models for the azinomycin antitumor antibiotics. J Org Chem 58:7848-7859
179. Moreno F, Rodicio R, Herrero P (1981) A new colorimetric assay for UDP-glucose 4-epimerase activity. Cell Mol Biol 27:589-592
180. Morita M, Sawa E, Yamaji K, Sakai T, Natori T, Koezuka Y, Fukushima H, Akimoto K (1996) Practical total synthesis of (2S,3S,4R)-1-O-(alpha-D-galactopyranosyl)-N-hexacosanoyl-2-amino-1,3,4-octadecanetriol, the antitumorial and immunostimulatory alpha-galactosylceramide, KRN7000. Biosci Biotechnol Biochem 60:288-292
181. Mulichak AM, Theisen MJ, Essigmann B, Benning C, Garavito RM (1999) Crystal structure of SQD1, an enzyme involved in the biosynthesis of the plant sulfolipid headgroup donor UDP-sulfoquinovose. Proc Natl Acad Sci U S A 96:13097-13102
182. Naha N, Lee HY, Jo MJ, Chung BC, Kim SH, Kim MO (2008) Rare sugar D-allose induces programmed cell death in hormone refractory prostate cancer cells. Apoptosis 13:1121-1134
183. Nahon DF, Roozen JP, de Graaf C (1998) Sensory evaluation of mixtures of maltitol or aspartame, sucrose and an orange aroma. Chem Senses 23:59-66
184. Nakamura T, Miyamoto O, Lu F, Okabe N, Kawai N, Tamiya T, Itano T (2009) Neuroprotective effects of d-allose, rare sugar, against cerebral ischemia/reperfusion injury in rats. J Cereb Blood Flow Metab 29:S240-S241
185. Neylon C (2004) Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. Nucleic Acids Res 32:1448-1459
186. Nielsen H, Engelbrecht J, Brunak S, vonHeijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng 10:1-6
187. Niou Y-K, Wu W-L, Lin L-C, Yu M-S, Shu H-Y, Yang H-H, Lin G-H (2009) Role of galE on biofilm formation by *Thermus* spp. Biochem Biophys Res Commun 390:313-318
188. Novozymes (2012) Novozymes - Biosolutions: Giving you cost savings and product improvements. 24 October 2012 on http://www.novozymes.com/en/solutions/Pages/default.aspx
189. NUTRA Ingredients USA (2003) Tagatose on the fast track. 24 October 2012 on www.nutraingredients-usa.com
190. Oberg CT, Blanchard H, Leffler H, Nilsson UJ (2008) Protein subtype-targeting through ligand epimerization: Talose-selectivity of galectin-4 and galectin-8. Bioorg Med Chem Lett 18:3691-3694
191. Oh DK (2007) Tagatose: properties, applications, and biotechnological processes. Appl Microbiol Biotechnol 76:1-8
192. Oh DK, Kim HJ, Ryu SA, Rho HJ, Kim P (2001) Development of an immobilization method of L-arabinose isomerase for industrial production of tagatose. Biotechnol Lett 23:1859-1862
193. Oh HJ, Kim HJ, Oh DK (2006) Increase in D-tagatose production rate by site-directed mutagenesis of L-arabinose isomerase from *Geobacillus thermodenitrificans*. Biotechnol Lett 28:145-149
194. Okano K (2009) Synthesis and pharmaceutical application of L-ribose. Tetrahedron 65:1937-1949
195. Osbourn A, Goss RJM, Field RA (2011) The saponins - polar isoprenoids with important and diverse biological activities. Nat Prod Rep 28:1261-1268
196. Otten LG, Hollmann F, Arends IWCE (2010) Enzyme engineering for enantioselectivity: from trial-and-error to rational design? Trends Biotechnol 28:46-54
197. Otten LG, Quax WJ (2005) Directed evolution: selecting today's biocatalysts. Biomol Eng 22:1-9
198. Park CS, Kim JE, Choi JG, Oh DK (2011) Characterization of a recombinant cellobiose 2-epimerase from Caldicellulosiruptor saccharolyticus and its application in the production of mannose from glucose. Appl Microbiol Biotechnol. Doi: 10.1007/s00253-011-3403-3
199. Park CS, Yeom SJ, Lim YR, Kim YS, Oh DK (2011) Substrate specificity of a recombinant ribose-5-phosphate isomerase from Streptococcus pneumoniae and its application in the production of L-lyxose and L-tagatose. World J Microbiol Biotechnol 27:743-750
200. Park HY, Park CS, Kim HJ, Oh DK (2007) Substrate specificity of a galactose 6-phosphate isomerase from Lactococcus lactis that produces D-allose from D-psicose. J Biotechnol 132:88-95
201. Pastinen O, Visuri K, Schoemaker HE, Leisola M (1999) Novel reactions of xylose isomerase from *Streptomyces rubiginosus*. Enzyme Microb Technol 25:695-700
202. Paterna JC, Boess F, Staubli A, Boelsterli UA (1998) Antioxidant and cytoprotective properties of D-tagatose in cultured murine hepatocytes. Toxicol Appl Pharmacol 148:117-125
203. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth 8:785-786

204. Petsko GA, Ringe D (2004) Protein Structure and Function. New Science Press, 195 p.

205. Poonperm W, Takata G, Morimoto K, Granstrom TB, Izumori K (2007) Production of L-xylulose from xylitol by a newly isolated strain of Bacillus pallidus Y25 and characterization of its relevant enzyme xylitol dehydrogenase. Enzyme Microb Technol 40:1206-1212

206. Poonperm W, Takata G, Okada H, Morimoto K, Granstrom TB, Izumori K (2007) Cloning, sequencing, overexpression and characterization of L-rhamnose isomerase from Bacillus pallidus Y25 for rare sugar production. Appl Microbiol Biotechnol 76:1297-1307

207. Prabhu P, Tiwari MK, Jeya M, Gunasekaran P, Kim IW, Lee JK (2008) Cloning and characterization of a novel L-arabinose isomerase from Bacillus licheniformis. Appl Microbiol Biotechnol 81:283-290

208. Pritchard L, Corne D, Kell D, Rowland J, Winson M (2005) A general model of error-prone PCR. J Theor Biol 234:497-509

209. Rao D, Best D, Yoshihara A, Gullapalli P, Morimoto K, Wormald MR, Wilson FX, Izumori K, Fleet GWJ (2009) A concise approach to the synthesis of all twelve 5-deoxyhexoses: D-tagatose-3-epimerase-a reagent that is both specific and general. Tetrahedron Lett 50:3559-3563

210. Rao D, Gullapalli P, Yoshihara A, Jenkinson SF, Morimoto K, Takata G, Akimitsu K, Tajima S, Fleet GWJ, Izumori K (2008) Direct production of L-tagatose from L-psicose by *Enterobacter aerogenes* 230S. J Biosci Bioeng 106:473-480

211. Rao D, Yoshihara A, Gullapalli P, Morinioto K, Takata G, da Cruz FP, Jenkinson SF, Wormald MR, Dwek RA, Fleet GWJ, Izumori K (2008) Towards the biotechnological isomerization of branched sugars: D-tagatose-3-epimerase equilibrates both enantiomers of 4-C-methyl-ribulose with both enantiomers of 4-C-methyl-xylulose. Tetrahedron Lett 49:3316-3321

212. Reetz MT, Kahakeaw D, Lohmer R (2008) Addressing the numbers problem in directed evolution. ChemBioChem 9:1797-1804

213. Rhimi M, Aghajari N, Juy M, Chouayekh H, Maguin E, Haser R, Bejar S (2009) Rational design of Bacillus stearothermophilus US100 L-arabinose isomerase: Potential applications for D-tagatose production. Biochimie 91:650-653

214. Rhimi M, Chouayekh H, Gouillouard I, Maguin E, Bejar S (2011) Production of D-tagatose, a low caloric sweetener during milk fermentation using L-arabinose isomerase. Bioresour Technol 102:3309-3315

215. Rhimi M, Ilhammami R, Bajic G, Boudebbouze S, Maguin E, Haser R, Aghajari N (2010) The acid tolerant L-arabinose isomerase from the food grade Lactobacillus sakei 23K is an attractive D-tagatose producer. Bioresour Technol 101:9171-9177

216. Richter F, Leaver-Fay A, Khare SD, Bjelic S, Baker D (2011) De Novo Enzyme Design Using Rosetta3. PLoS One 6:(5):e19230

217. Rodionova IA, Scott DA, Grishin NV, Osterman AL, Rodionov DA (2012) Tagaturonate-fructuronate epimerase UxaE, a novel enzyme in the hexuronate catabolic network in *Thermotoga maritima*. Environ Microbiol 14:2920-2934

218. Rohl CA, Baker D (2002) De novo determination of protein backbone structure from residual dipolar couplings using rosetta. J Am Chem Soc 124:2723-2729

219. Rohl CA, Strauss CEM, Misura KMS, Baker D (2004) Protein structure prediction using Rosetta. Methods Enzymol 383:66-93

220. Rollini M, Manzoni M (2005) Bioconversion of D-galactitol to tagatose and dehydrogenase activity induction in *Gluconobacter oxydans*. Process Biochem 40:437-444

221. Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik DS, Baker D (2008) Kemp elimination catalysts by computational enzyme design. Nature 453:190-194

222. Sako Y, Nakagawa S, Takai K, Horikoshi K (2003) *Marinithermus hydrothermalis* gen. nov., sp nov., a strictly aerobic, thermophilic bacterium from a deep-sea hydrothermal vent chimney. Int J Syst Evol Microbiol 53:59-65

223. Sakuraba H, Kawai T, Yoneda K, Ohshima T (2011) Crystal structure of UDP-galactose 4-epimerase from the hyperthermophilic archaeon *Pyrobaculum calidifontis*. Arch Biochem Biophys 512:126-134

224. Sambrook J, Russell DW (2001) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press,

225. Samuel J, Luo Y, Morgan PM, Strynadka NCJ, Tanner ME (2001) Catalysis and binding in L-ribulose-5-phosphate 4-epimerase: A comparison with L-fuculose-1-phosphate aldolase. Biochemistry 40:14772-14780

226. Samuel J, Tanner ME (2002) Mechanistic aspects of enzymatic carbohydrate epimerization. Nat Prod Rep 19:261-277

227. Sanchis J, Fernandez L, Carballeira J, Drone J, Gumulya Y, Hobenreich H, Kahakeaw D, Kille S, Lohmer R, Peyralans J, Podtetenieff J, Prasad S, Soni P, Taglieber A, Wu S, Zilly F, Reetz M (2008) Improved PCR method for the creation of saturation mutagenesis libraries in directed evolution: application to difficult-to-amplify templates. Appl Microbiol Biotechnol 81:387-397
228. Sasahara H, Izumori K (2005) Production of L-talitol from L-psicose by *Metschnikowia koreensis* LA1 isolated from soy sauce mash. J Biosci Bioeng 100:335-338
229. Sato M, Kurose H, Yamasaki T, Izumori K (2008) Potential anthelmintic: D-psicose inhibits motility, growth and reproductive maturity of L1 larvae of *Caenorhabditis elegans*. J Nat Med 62:244-246
230. Schneider S, Gutierrez M, Sandalova T, Schneider G, Clapes P, Sprenger GA, Samland AK (2010) Redesigning the Active Site of Transaldolase TalB from *Escherichia coli*: New Variants with Improved Affinity towards Nonphosphorylated Substrates. ChemBioChem 11:681-690
231. Schneider S, Sandalova T, Schneider G, Sprenger GA, Samland AK (2008) Replacement of a Phenylalanine by a Tyrosine in the Active Site Confers Fructose-6-phosphate Aldolase Activity to the Transaldolase of *Escherichia coli* and Human Origin. J Biol Chem 283:30064-30072
232. Schulz JM, Watson AL, Sanders R, Ross KL, Thoden JB, Holden HM, Fridovich-Keil JL (2004) Determinants of function and substrate specificity in human UDP-galactose 4 '-epimerase. J Biol Chem 279:32796-32803
233. Senoura T, Ito S, Taguchi H, Higa M, Hamada S, Matsui H, Ozawa T, Jin S, Watanabe J, Wasaki J, Ito S (2011) New microbial mannan catabolic pathway that involves a novel mannosylglucose phosphorylase. Biochem Biophys Res Commun 408:701-706
234. Sheldon RA (2007) Enzyme immobilization: The quest for optimum performance. Adv Synth Catal 349:1289-1307
235. Sheldon RA (2011) Cross-Linked Enzyme Aggregates as Industrial Biocatalysts. Org Process Res Dev 15:213-223
236. Shimonishi T, Izumori K (1996) A new enzyme, L-ribose isomerase from Acinetobacter sp strain DL-28. J Ferment Bioeng 81:493-497
237. Sigrist-Nelson K, Hopfer U (1974) A distinct D-fructose transport system in isolated brush border membrane. Biochim Biophys Acta 367:247-254
238. Simpson FJ, Wood WA (1956) L-ribulose-5-phosphate: formation by purified kinase from *Aerobacter aerogenes*. J Am Chem Soc 78:5452-5453
239. Skrzypek M, Maleszka R (1994) A gene homologous to that encoding UDP galactose-4-epimerase is inducible by xylose in the yeast *Pachysolen tannophilus*. Gene 140:127-129
240. Sproul AA, Lambourne LTM, Jean-Jacques J, Kornberg HL (2001) Genetic control of manno(fructo)kinase activity in *Escherichia coli*. Proc Natl Acad Sci U S A 98:15257-15259
241. Stemmer WPC (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc Natl Acad Sci U S A 91:10747-10751
242. Stemmer WPC (1994) Rapid evolution of a protein in vitro by DNA shuffling. Nature 370:389-391
243. Stoop JMH, Chilton WS, Pharr DM (1996) Substrate stereospecificity of the NAD-dependent mannitol dehydrogenase from celery. Phytochemistry 43:1145-1150
244. Sugiyama M, Hong ZY, Greenberg WA, Wong CH (2007) In vivo selection for the directed evolution Of L-rhamnulose aldolase from L-rhamnulose-1-phosphate aldolase (RhaD). Bioorg Med Chem 15:5905-5911
245. Sui L, Nomura R, Youyi D, Yamaguchi F, Izumori K, Tokuda M (2007) Cryoprotective effects of D-allose on mammalian cells. Cryobiology 55:87-92
246. Sultana I, Mizanur RMD, Takeshita K, Takada G, Izumori K (2003) Direct production of D-arabinose from D-xylose by a coupling reaction using D-xylose isomerase, D-tagatose 3-epimerase and D-arabinose isomerase. J Biosci Bioeng 95:342-347
247. Takagi Y, Nakai K, Tsuchiya T, Takeuchi T (1996) A 5'-(trifluoromethyl)anthracycline glycoside: Synthesis of antitumor-active 7-0-(2,6-dideoxy-6,6,6-trifluoro-alpha-L-lyxo-hexopyranosyl)adriamycinon e. J Med Chem 39:1582-1588
248. Takata G, Poonperm W, Morimoto K, Izumori K (2010) Cloning and overexpression of the xylitol dehydrogenase gene from *Bacillus pallidus* and its application to L-xylulose production. Biosci Biotechnol Biochem 74:1807-1813
249. Takeshita K, Ishida Y, Takada G, Izumori K (2000) Direct production of allitol from D-fructose by a coupling reaction using D-tagatose 3-epimerase, ribitol dehydrogenase and formate dehydrogenase. J Biosci Bioeng 90:545-548

250. Takeshita K, Suga A, Takada G, Izumori K (2000) Mass production of D-psicose from D-fructose by a continuous bioreactor system using immobilized D-tagatose 3-epimerase. J Biosci Bioeng 90:453-455

251. Tanner ME (2002) Understanding nature's strategies for enzyme-catalyzed racemization and epimerization. Acc Chem Res 35:237-246

252. Tatai J, Osztrovszky G, Kajtar-Peredy M, Fugedi P (2008) An efficient synthesis of L-idose and L-iduronic acid thioglycosides and their use for the synthesis of heparin oligosaccharides. Carbohydr Res 343:596-606

253. Taylor TP, Fasina O, Bell LN (2008) Physical properties and consumer liking of cookies prepared by replacing sucrose with tagatose. J Food Sci 73:S145-S151

254. Thoden JB, Frey PA, Holden HM (1996) Crystal structures of the oxidized and reduced forms of UDP-galactose 4-epimerase isolated from *Escherichia coli*. Biochemistry 35:2557-2566

255. Thoden JB, Frey PA, Holden HM (1996) High-resolution X-ray structure of UDP-galactose 4-epimerase complexed with UDP-phenol. Protein Sci 5:2149-2161

256. Thoden JB, Frey PA, Holden HM (1996) Molecular structure of the NADH/UDP-glucose abortive complex of UDP-galactose 4-epimerase from *Escherichia coli*: Implications for the catalytic mechanism. Biochemistry 35:5137-5144

257. Thoden JB, Hegeman AD, Wesenberg G, Chapeau MC, Frey PA, Holden HM (1997) Structural analysis of UDP-sugar binding to UDP-galactose 4-epimerase from *Escherichia coli*. Biochemistry 36:6294-6304

258. Thoden JB, Henderson JM, Fridovich-Keil JL, Holden HM (2002) Structural analysis of the Y299C mutant of *Escherichia coli* UDP-galactose 4-epimerase - Teaching an old dog new tricks. J Biol Chem 277:27528-27534

259. Thoden JB, Holden HM (1998) Dramatic differences in the binding of UDP-galactose and UDP-glucose to UDP-galactose 4-epimerase from *Escherichia coli*. Biochemistry 37:11469-11477

260. Thoden JB, Wohlers TM, Fridovich-Keil JL, Holden HM (2000) Crystallographic evidence for Tyr 157 functioning as the active site base in human UDP-galactose 4-epimerase. Biochemistry 39:5691-5701

261. Thoden JB, Wohlers TM, Fridovich-Keil JL, Holden HM (2001) Human UDP-galactose 4-epimerase - Accommodation of UDP-N-acetylglucosamine within the active site. J Biol Chem 276:15131-15136

262. Tianwei M, Balakrishna PS, Lian ZY, Sheng LJ, Kirupa S, Jinfa D, Chunguang W, Hongbum K, Gary MN (1996) Structure-activity relationships of 1-(2-deoxy-2-fluro-beta-L-arabino-furanosyl) pyrimidine nucleosides as anti-hepatitis B virus agents. J Med Chem 39:2835-2843

263. Tokuda M, Tsukamoto I, Konishi R, Kubota Y, Izumori K (2010) Method of controlling the proliferation of vascular endothelial cells and inhibiting lumen formation. US Patent 0222285 A1

264. Trefzer A, Salas JA, Bechthold A (1999) Genes and enzymes involved in deoxysugar biosynthesis in bacteria. Nat Prod Rep 16:283-299

265. Tseng T-T, Tyler BM, Setubal JC (2009) Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. BMC Microbiol 9:S2

266. Tukel SS, Alagoz D (2008) Catalytic efficiency of immobilized glucose isomerase in isomerization of glucose to fructose. Food Chemistry 111:658-662

267. Usvalampi A, Kiviharju K, Leisola M, Nyyssola A (2009) Factors affecting the production of l-xylulose by resting cells of recombinant Escherichia coli. J Ind Microbiol Biotechnol 36:1323-1330

268. Vriend G (1990) WHAT IF - A molecular modeling and drug design program. J Mol Graph 8:52-56

269. Vuolanto A, Pastinen O, Schoemaker HE, Leisola M (2002) C-2 epimer formation of tetrose, pentose and hexose sugars by xylose isomerase. Biocatal Biotransform 20:235-240

270. Vuolanto A, Uotila S, Leisola M, Visuri K (2003) Solubility and crystallization of xylose isomerase from *Streptomyces rubiginosus*. J Cryst Growth 257:403-411

271. Waegeman H, Soetaert W (2011) Increasing recombinant protein production in *Escherichia coli* through metabolic and genetic engineering. J Ind Microbiol Biotechnol 38:1891-1910

272. Wee TG, Frey PA (1973) Studies on mechanism of action of uridine diphosphate-galactose-4-epimerase. 2. Substrate-dependent reduction by sodium-borohydride. J Biol Chem 248:33-40

273. Wee TG, Frey PA, Davis J (1972) Studies on mechanism of action of uridine diphosphate-galactose-4-epimerase. 1. Ambiguity in chemical trapping of a proposed keto-intermediate by NaB3H4. J Biol Chem 247:1339-&

274. Werner W, Rey HG, Wielinge.H (1970) Properties of a new chromogen for determination of glucose in blood according to GOD/POD-method. Z Anal Chem Fresenius 252:224-228

275. Wong SS, Frey PA (1977) Fluorescence and nucleotide binding properties of *Escherichia coli* uridine diphosphate galactose 4-epimerase: support for a model for nonstereospedific action. Biochemistry 16:298-305

276. Woodyer RD, Christ TN, Deweese KA (2010) Single-step bioconversion for the preparation of L-gulose and L-galactose. Carbohydr Res 345:363-368

277. Woodyer RD, Wymer NJ, Racine FM, Khan SN, Saha BC (2008) Efficient production of L-ribose with a recombinant Escherichia coli biocatalyst. Appl Environ Microbiol 74:2967-2975

278. Xiao H, Wang G, Wang P, Li Y (2010) Convenient synthesis of D-talose from D-galactose. Chin J Chem 28:1229-1232

279. Xu Z, Li S, Fu F, Li G, Feng X, Xu H, Ouyang P (2012) Production of D-tagatose, a Functional Sweetener, Utilizing Alginate Immobilized *Lactobacillus fermentum* CGMCC2921 Cells. Appl Biochem Biotechnol 166:961-973

280. Yamaguchi F, Kazuyo K, Dong YY, Hirata Y, Tokuda M (2009) Rare sugar D-allose enhances anti-tumor effect of 5-fluorouracil on the human hepatocellular carcinoma cell line HuH-7. J Physiol Sci 59:304-304

281. Yang G, Withers SG (2009) Ultrahigh-Throughput FACS-Based Screening for Directed Enzyme Evolution. ChemBioChem 10:2704-2715

282. Yeom SJ, Kim NH, Park CS, Oh DK (2009) L-ribose production from L-arabinose by using purified L-arabinose isomerase and mannose-6-phosphate isomerase from Geobacillus thermodenitrificans. Appl Environ Microbiol 75:6941-6943

283. Yeom SJ, Kim NH, Yoon RY, Kwon HJ, Park CS, Oh DK (2009) Characterization of a mannose-6-phosphate isomerase from Geobacillus thermodenitrificans that converts monosaccharides. Biotechnol Lett 31:1273-1278

284. Yeom SJ, Seo ES, Kim BN, Kim YS, Oh DK (2011) Characterization of a mannose-6-phosphate isomerase from Thermus thermophilus and increased L-ribose production by its R142N mutant. Appl Environ Microbiol 77:762-767

285. Yeom SJ, Seo ES, Kim YS, Oh DK (2011) Increased D-allose production by the R132E mutant of ribose-5-phosphate isomerase from Clostridium thermocellum. Appl Microbiol Biotechnol 89:1859-1866

286. Yoon RY, Yeom SJ, Park CS, Oh DK (2009) Substrate specificity of a glucose-6-phosphate isomerase from Pyrococcus furiosus for monosaccharides. Appl Microbiol Biotechnol 83:295-303

287. Yoshida H, Yamada M, Nishitani T, Takada G, Izumori K, Kamitori S (2007) Purification, crystallization and preliminary X-ray diffraction studies of D-tagatose 3-epimerase from *Pseudomonas cichorii*. Acta Crystallogr, Sect F: Struct Biol Cryst Commun 63:123-125

288. Yoshihara A, Haraguchi S, Gullapalli P, Rao D, Morimoto K, Takata G, Jones N, Jenkinson SF, Wormald MR, Dwek RA, Fleet GWJ, Izumori K (2008) Isomerization of deoxyhexoses: green bioproduction of 1-deoxy-D-tagatose from L-fucose and of 6-deoxy-D-tagatose from D-fucose using Enterobacter agglomerans strain 221e. Tetrahedron: Asymmetry 19:739-745

289. Yoshihara K, Shinohara Y, Hirotsu T, Izumori K (2006) Bioconversion of D-psicose to D-tagatose and D-talitol by *Mucoraceae* fungi. J Biosci Bioeng 101:219-222

290. Yoshikawa M, Murakami N, Inoue Y, Hatakeyama S, Kitagawa I (1993) A new approach to the synthesis of optically active pseudo-sugar and pseudo-nucleoside: Synthesis of pseudo-D-arabinofuranose, (dextro)-cyclaradine, and (dextro)-1-pseudo-D-arabinofuranosyluracil from D-arabinose. Chem Pharm Bull 41:636-638

291. Yun M, Moon HR, Kim HO, Choi WJ, Kim YC, Park CS, Jeong LS (2005) A highly efficient synthesis of unnatural L-sugars from D-ribose. Tetrahedron Lett 46:5903-5905

292. Zhang Y-W, Tiwari MK, Jeya M, Lee J-K (2011) Covalent immobilization of recombinant Rhizobium etli CFN42 xylitol dehydrogenase onto modified silica nanoparticles. Appl Microbiol Biotechnol 90:499-507

SUMMARY & SAMENVATTING

Summary

Tagatose is a rare sugar that can be applied for multiple reasons in different industries, for instance as a low-caloric sweetener in dietary food, as well as additive in detergents, cosmetics, and pharmaceutical formulations, but also as drug molecule itself in diabetes treatment. As for other rare sugars, tagatose is not abundantly present in nature and therefore it has to be made from more available sugars in order to use it. It is currently being produced starting from galactose. However, to be able to compete with the current predominant sweeteners (like sucrose and high fructose corn syrup), it should be produced in much higher quantities than is possible when starting from galactose. In order to overcome this issue, tagatose production should start from a more widely available (and cheaper) substrate. Fructose and glucose are two such very abundant substrates; however, no (bio)catalysts are available to convert fructose or glucose into tagatose. Nevertheless, some enzymes were found to perform similar reactions and therefore are promising to one day become biocatalysts for tagatose production.

In this work, the cloning and analysis of two totally different C4-epimerases is described in respect to their capability of tagatose production. The first enzyme is L-ribulose-5-phosphate 4-epimerase from *Geobacillus thermodenitrificans,* an aldolase-related epimerase that would require an adaptation of the substrate binding site around the phosphate moiety of the substrate. The second C4-epimerase is naturally active on nucleotide activated sugars, namely the UDP-Glc(NAc) 4-epimerase from *Marinithermus hydrothermalis*. The major challenge in the engineering of this UDP-hexose 4-epimerase is trying to get rid of the necessity of the UDP-group of the substrate and making it active on free monosaccharides.

At first, the *Geobacillus* L-ribulose-5-phosphate 4-epimerase gene was cloned in an appropriate expression vector and expressed in *E. coli*. The recombinant enzyme was first characterized with respect to affinity for L-ribulose-5-phosphate, metal ion activation and stability at 37 °C. To that end, its natural substrate had to be produced first, which was accomplished by phosphorylation of L-ribulose using ATP as phosphate donor and recombinantly expressed L-ribulokinase as biocatalyst.

After characterization, mutagenesis was achieved both randomly and semi-rationally by error-prone PCR and site saturation mutagenesis, respectively. To be able to detect enzyme variants harboring (improved) tagatose 4-epimerase activity among the thousands mutant enzymes, two ‘identification’ systems were developed. Two selection strains were developed that can be used for the Darwinian selection of improved enzyme variants, while also a colorimetric screening assay has been created. Although several millions and thousands of mutants were analyzed using the selection strains and screening assay, respectively, no variants were confirmed to possess (improved) tagatose 4-epimerase activity.

Secondly, the UDP-hexose 4-epimerase from *Marinithermus hydrothermalis* was also cloned and heterologously expressed in *E. coli*. A thorough characterization of this second epimerase was performed, revealing that it belongs to the type 2 UDP-hexose 4-epimerases. As expected for a type 2 epimerase, its substrate specificity could easily be altered by mutagenesis of a single residue, namely the so-called gatekeeper. This also confirms the previously reported hypothesis about substrate specificity in type 1 and type 2 epimerases.

Mutational analysis of the UDP-hexose 4-epimerase uncovered two new features that can be found in these epimerases. The *Marinithermus* enzyme was found to possess a Tx_nYx_3K catalytic triad, rather than the usual serine containing triad (Sx_nYx_3K). The presence of the threonine’s methyl function was found to be of more importance for the enzyme’s affinity for N-acetylated UDP-sugars than for non-acetylated substrates. As such, the Tx_nYx_3K triad might be a new substrate specificity determinant for type 2 UDP-hexose 4-epimerases. The second new feature was the presence of two consecutive glycine residues next to the catalytic threonine, which were found to be important for activity of the enzyme with non-acetylated and even bigger importance for activity on N-acetylated substrates. In an attempt to identify new determinants for specificity towards UDP-GlcNAc, two loop mutants were created but they were found to be inactive, most likely due to dispositioning of the catalytic tyrosine, which results in the disruption of the subtle catalytic chemistry.

In addition, the *Marinithermus* UDP-hexose 4-epimerase was also tested for its ability to convert the free monosaccharides fructose/tagatose, glucose/galactose and the phosphorylated α-Glc-1-P. Furthermore, also the *E. coli* UDP-hexose 4-epimerase was cloned and also here no epimerase activity could be detected on free monosaccharides, in contrast to what has previously been reported.

Samenvatting

Tagatose is een zeldzame suiker die om meerdere redenen kan worden toegepast in verschillende industrietakken, bv. als lage-energie suiker in dieetvoeding, maar ook als additief in detergent, cosmetica en farmaceutische formuleringen, en zelfs als medicatie op zichzelf in de behandeling van diabetes. Net als de andere zeldzame suikers is tagatose niet overvloedig aanwezig in de natuur en daarom moet het (bio)chemisch worden aangemaakt startende van meer abundante suikers vooraleer het gebruikt kan worden voor een van zijn toepassingen. Momenteel wordt het gemaakt uit galactose, maar om de competitie met de huidige zoetstoffen (zoals sucrose en *high fructose corn syrup*) aan te gaan, zou het in grotere hoeveelheden moeten kunnen worden aangemaakt dan mogelijk is uit galactose. Om dit euvel te overwinnen, zou tagatose productie moeten starten van meer abundante suikers, zoals fructose and glucose. Er zijn echter geen (bio)katalysatoren beschikbaar die deze suikers kunnen omzetten in tagatose. Desalniettemin zijn er enzymen die gelijkaardige reacties katalyseren en die mogelijks interessant zijn voor de synthese van tagatose.

In dit proefschrift wordt de klonering en analyse van twee totaal verschillende C4-epimerasen besproken in relatie tot hun mogelijkheid of potentieel voor tagatose productie. Het eerste enzym is het L-ribulose-5-fosfaat 4-epimerase van *Geobacillus thermodenitrificans* en dit aan aldolasen verwante enzym vereist een aanpassing ter hoogte van de fosfaatbindende regio. Het tweede C4-epimerase is van nature actief op nucleotide geactiveerde suikers, namelijk het UDP-Glc(NAc) 4-epimerase van *Marinithermus hydrothermalis*. De grote uitdaging voor dit epimerase is om komaf te maken met de nood aan de UDP-groep voor activiteit en het actief te krijgen op niet-gemodificeerde suikers.

Ten eerste werd het L-ribulose-5-fosfaat 4-epimerase gen van *Geobacillus* gekloneerd in een geschikt expressieplasmide en vervolgens tot expressie gebracht in *E. coli*. Het recombinant enzym werd gekarakteriseerd met zijn natuurlijk substraat, L-ribulose-5-fosfaat, om meer te weten te komen over de affiniteit hiervoor, activatie door metaalionen alsook thermostabiliteit. Hiervoor werd eerst L-Ru-5-P aangemaakt startende van L-ribulose en ATP en gebruikmakend van L-ribulokinase als biokatalysator.

Na karakterisatie, werden twee verschillende mutagenesetechnieken toegepast, namelijk random mutagenese via error-prone PCR en een semi-rationele aanpak door site saturation mutagenesis. Om enzymvarianten te identificeren die (verbeterde) tagatose 4-epimerase activiteit bezitten, werd enerzijds een selectiestam en anderzijds een screeningsmethode ontwikkeld. De selectiestam maakt gebruik van Darwins ‘survival of the fittest’ theorie, terwijl met de screening alle mutanten afzonderlijk worden geanalyseerd. Ondanks de miljoenen en duizenden mutanten die via respectievelijk selectie en screening werden geanalyseerd, werden geen (verbeterde) tagatose 4-epimerase varianten gevonden.

Ten tweede werd ook het UDP-hexose 4-epimerase van *Marinithermus hydrothermalis* gekloneerd en tot overexpressie gebracht in *E. coli*. Een grondige karakterisatie van dit epimerase toonde aan dat het behoort tot de type 2 UDP-hexose 4-epimerasen. Zoals verwacht voor dit type epimerasen kon de substraatspecificiteit simpelweg worden aangepast naar een type 1 door middel van mutatie van één enkel residue, de gatekeeper. Dit bevestigt eveneens oudere hypothesen rond de substraatspecificiteitsdeterminanten van de type 1 en type 2 epimerasen.

Mutationele analyses van het UDP-hexose 4-epimerase brachten ook twee nieuwe kenmerken aan het licht. Het enzyme blijkt een Tx_nYx_3K katalytische triade te bezitten in plaats van de gewoonlijke Sx_nYx_3K. De extra methylgroep van de threonine is belangrijker voor de activiteit op N-geacetyleerde UDP-suikers dan voor niet-geacetyleerde substraten. Alsdusdanig zou een Tx_nYx_3K katalytische triade een nieuwe determinant kunnen zijn voor substraatspecificiteit van type 2 UDP-hexose 4-epimerasen. Daarnaast werden ook twee opeenvolgende glycines gevonden vlak naast de katalytische threonine, dewelke belangrijk blijken te zijn voor de activiteit van het enzym, met een grotere invloed voor de geacetyleerde substraten. In een poging om meer te weten te komen over de substraatspecificiteitsdeterminanten voor UDP-GlcNAc werden twee loopmutanten gemaakt, maar deze bleken inactief. Volgens de homologiemodellen wordt door deze mutaties de delicate chemie van het katalytisch centrum verstoord door verplaatsing van de katalytische tyrosine.

Bovendien werd het *Marinithermus* UDP-hexose 4-epimerase ook getest op zijn vermogen voor het omzetten van de vrije monosachariden fructose/tagatose, glucose/galactose en het gefosforyleerde α-Glc-1-P. Daarnaast werd ook het *E. coli* UDP-hexose 4-epimerase gekloneerd, maar voor beide enzymen kon geen C4-epimerase activiteit op vrije monosachariden worden waargenomen, in tegenstelling tot wat eerder gemeld was voor het *E. coli* epimerase.

CURRICULUM VITAE

Curriculum Vitae

Personalia

Koen Beerens

°03/02/1985, Turnhout

koen_beerens@hotmail.com
+32(0)498/70.84.37

Contact

Centre for Industrial Biotechnology and Biocatalysis
Department of Biochemical and Microbial Technology
Faculty of Bioscience Engineering, Ghent University
Coupure Links 653, B-9000 Gent, Belgium
www.inbio.be
+32 (0)9/264.99.21
koen.beerens@ugent.be

Professional career

2009-2013	Doctoral research funded by the Agency for Innovation by Science and Technology in Flanders (IWT-Vlaanderen) Thesis: “Characterization and engineering of epimerases for the production of rare sugars” Promotors: Prof. dr. ir. Wim Soetaert & Prof. dr. Tom Desmet Centre for Industrial Biotechnology and Biocatalysis Department of Biochemical and Microbial Technology Faculty of Bioscience Engineering, Ghent University, Gent, Belgium
2008	Scientific researcher (IOF funding) Research project: Redesigning a cellobiose phosphorylase towards a lactose phosphorylase Centre for Industrial Biotechnology and Biocatalysis Department of Biochemical and Microbial Technology Faculty of Bioscience Engineering, Ghent University, Gent, Belgium

Education

2009-2013	Doctoral schools training program Doctoral school of Bioscience engineering, Ghent University, Gent, Belgium
2005-2007	Master in Biochemistry Faculty of Sciences, Ghent University, Gent, Belgium Master dissertation: “Cloning, recombinant expression and immune recognition of candidate venom (gland) components showing homology to known components of snake and scorpion venom.” Promotors: Prof. dr. F. J. Jacobs & dr. D. de Graaf Student member of the Educational Commission ‘Biochemistry’
2003-2005	Bachelor in Biology Faculty of Sciences, Ghent University, Gent, Belgium
1997-2003	Secondary school (Science & Mathematics) Klein Seminarie, Hoogstraten, Belgium

Student supervision

Master thesis supervision: Koen De Blieck (2011-2012)
Directed evolution of an epimerase for the production of rare sugars
Promotors: Prof. dr. ir. Wim Soetaert & Prof. dr. Tom Desmet

Master thesis supervision: Simon Vandermersch (2009-2010)
Engineering of an isomerase for the production of rare sugars
Promotor: Prof. dr. ir. Wim Soetaert

Practical exercises “Biocatalysis & Enzyme Technology” (2010-2012)

Practical exercises “Microbiology” (2008-2010)

Advanced courses

Ghent Bio-Economy summer school (August 2012)
Faculty of Bioscience Engineering, Ghent University, Gent, Belgium

Biocatalysis & Enzyme Technology (2009-2010)
Faculty of Bioscience Engineering, Ghent University, Gent, Belgium

Advanced Course on Downstream Processing (June 2009)
Biotechnology Studies Delft Leiden, Delft, The Netherlands
Awarded with the “Golden Award for Best Downstream Process”

Scientific output

Peer-reviewed publications

Beerens, K., Desmet, T., Soetaert, W. (2012). Enzymes for the biocatalytic production of rare sugars. *Journal of Industrial Microbiology and Biotechnology*, 39 (6), 823-834, DOI: 10.1007/s10295-012-1089-x

Beerens, K., Soetaert, W., Desmet, T. (Epub ahaed of print, 2012). Characterization and mutational analysis of the UDP-Glc(NAc) 4-epimerase from Marinithermus hydrothermalis. *Applied Microbiology and Biotechnology*, DOI: 10.1007/s00253-012-4635-6

Conference Contributions (oral and poster contributions)

Journée scientifique SFM-BBM: 'Protein structures: from computer to the application' (2012)
Beerens, K.
Oral presentation: "Characterization and mutational analysis of the UDP-hexose 4-epimerase from *Marinithermus hydrothermalis*"
Life Science Institute (ISV), Université Catholique de Louvain, Belgium, 16 November 2012

Cambridge Protein Engineering Meeting (CamProEng, 2011)
Beerens, K.
Oral presentation: "Selection and screening for isomerase activity: pitfalls and stepping stones"
Department of Biochemistry, University of Cambridge, Cambridge, UK, 26-29 August 2011

Cambridge Protein Engineering Meeting (CamProEng, 2010)
Beerens, K.
Oral presentation: "Engineering the Phosphate Binding Site"
Department of Biochemistry, University of Cambridge, Cambridge, UK, 28-30 August 2010

Ghent Bio-Economy summer school (2012)
Beerens, K., Desmet T. and Soetaert,W.
Poster presentation: " Redesigning epimerases towards rare sugar producing biocatalysts"
Faculty of Bioscience Engineering, Ghent University, Gent, Belgium, 7-10 August 2012

8th Carbohydrate Bioengineering Meeting (CBM8, 2009)
De Groeve, M. R. M., De Baere, M., **Beerens, K.**, Cerdobbel, A., Desmet T. and Soetaert,W.
Poster presentation: "Novel lactose phosphorylase enzymes by protein engineering of cellobiose phosphorylase"
Ischia, Italy, 10-13 May 2009

Conference attendance (without poster or oral contributions)

Eurocarb 16 (2011), Sorrento, Italy
Bee Lateral Symposium (2006), Leuven, Belgium

APPENDICES

Appendix I: Nucleotides (IUPAC-IUB code)

Nucleotide	Name	Complement
A	Adenine	T
C	Cytosine	G
G	Guanine	C
T	Thymine	A
U	Uracyl (RNA)	A
R	Purine (A or G)	Y
Y	Pyrimidine (C or T)	R
W	Weak (A or T)	S
S	Strong (C or G)	W
M	Amino (A or C)	K
K	Keto (G or T)	M
V	A, C or G	B
B	C, G or T	V
H	A, C or T	D
D	A, G or T	H
N	A, C, G or T	N

Appendix II: The standard genetic code

Second position

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

Appendix III: SSM – Library size (top) and screening effort (bottom)

Amino acid	NNN	NNK/NNS	NDT
Ala	4	2	0
Arg	6	3	1
Asp	2	1	1
Asn	2	1	1
Cys	2	1	1
Glu	2	1	0
Gln	2	1	0
Gly	4	2	1
His	2	1	1
Ile	3	1	1
Leu	6	3	1
Lys	2	1	0
Met	1	1	0
Phe	2	1	1
Pro	4	2	0
Ser	6	3	1
Thr	4	2	0
Trp	1	1	0
Tyr	2	1	1
Val	4	2	1
STOP	3	1	0
Total codons	64	32	12
Total amino acids	20	20	12
Screening effort	190 CFU	94 CFU	34 CFU

# NNS	# codons	# CFU	# MTPs
1	32	94	1
2	1024	3066	32
3	32768	98163	1023
4	1048576	3141251	32721

Appendix IV: Amino acid terminology

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

Every story has an end.

But in life, every ending is just a new beginning.

- Ray, Uptown girls (2003)