

Protein Engineering of *Escherichia coli* β -glucuronidase

A thesis submitted for the degree of Doctor of Philosophy of
The Australian National University

By

Jo Leen Lim

Research School of Chemistry



**Australian
National
University**

November 2017

© Copyright by Jo Leen Lim 2017

All Right Reserved

Declaration

This thesis presents original research that I have performed under the supervision of Professor David Ollis at the Research School of Chemistry. I have made an effort to acknowledge the contributions of individuals to this thesis as listed below as well as in the relevant parts of the thesis:

- Plasmid of wild-type β -GUS and in pET28a was provided by the McLeod research group
- pJ401 backbone vector was obtained from Dr. Bradley Stevenson from the McLeod research group
- Substrates for glycosylation reaction of glucuronylsynthase, α -glucosyl fluoride and acceptor steroids were synthesized by Mr. Andy Pranata from the McLeod research group

To the best of my knowledge, the work presented in this thesis does not contain material that has been previously published, except where due acknowledgement has been made.

This thesis conforms to the Australian National University guidelines and regulations of Higher Degree Research award and thesis production.

.....
Jo Leen Lim

Acknowledgements

Neither this thesis nor my continuing sanity would have been possible without the support and assistance of a great many people. While some contributions were larger than others, all were equally essential.

I wish to express my sincere thanks and gratitude to my supervisor Professor David Ollis for giving me the opportunity to work on such an interesting and challenging project. His support, guidance and help from the initial to final stages of my candidature have helped me to develop a solid understanding and passion for project research. He has been a brilliant mentor, colleague and friend whose knowledge knows no boundaries. He would always be thinking about the project and continually developing new angles to take the research.

My research involving organic chemistry could have not been done if not for the assistance of Associate Professor Malcolm McLeod and his research group. A huge thank you to him for his patience when introducing me to the glucuronylsynthase system and for sharing his experience and vast knowledge in organic chemistry. I would also like to thank Andy Pranata who has kindly synthesized the substrates that were used for the glucuronylsynthase/ glucosynthase enzymes. I am also grateful to him for his assistance with NMR and mass spectroscopy analyses.

I thank Dr. Bradley Stevenson for his invaluable input and assistance in various biochemical techniques and troubleshooting. He has always found the time to offer insightful advice whenever I came across unusual results or half-baked ideas. His enthusiasm, creativity and acumen never cease to amaze me. I am grateful to Tracy Murray for providing a highly efficient and supportive work environment and giving me a good grounding in the practical aspects of molecular biology. Despite the minor contribution of protein crystallisation to this project, many thanks to Dr. Paul Carr for his guidance and help with experiments in X-ray crystallography. Thanks go out to all the members of the Biological Chemistry Groups. Being part of such a community make the bad times bearable and the good times better. Thank you for your support on a personal and professional level. Your queries, discussions and advice have helped make this thesis what it is.

Additional thanks go out to Rebecca Buckland and Jonathan Wu for the time and effort spent proof reading this thesis.

Acknowledgements

The DNA sequencing in this thesis was carried out through the services of the Biomolecular Research Facility based at the John Curtin School of Medical Research, ANU.

Personal thanks to my Buddhist friends and friends in Toad Hall, past and present, who have shown me the true meaning of friendship and; difference between living and being alive.

I am grateful to Research School of Chemistry, the Australian National University and Australian Government Research Training Program (RTP) Scholarship for financially supporting my PhD candidature.

Finally, to my family and to Vee Cheun thank you for your unconditional support. I couldn't have done this without you. This thesis happened because my family and good friends believe in me.

Abstract

This thesis describes engineering studies with the *Escherichia coli* β -glucuronidase enzyme (*E. coli* β -GUS) that catalyzes the hydrolysis of D-glucuronic acids (glycone) that are conjugated through a β -O-glycosidic linkage to an aglycone. The enzyme is specific for the glucuronic acid component and will tolerate a variety of aglycones. A point mutation was known to convert β -GUS into a glucuronylsynthase, that is an enzyme capable of the synthesis of glucuronide conjugates. The long-term aim of the present research was to change the donor sugar specificity of the glucuronylsynthase from a glucuronyl donor to a glucosyl donor allowing the synthesis of glucosides. The approach taken to achieve the long-term aim was to first alter the specificity of β -GUS and to then convert the variants to a synthase. There were two approaches taken to altering the substrate specificity of β -GUS. The first approached was to use site saturation mutagenesis to alter key residues that are located at the active site. A more effective approach involved using directed evolution to generate variants with altered specificity. A selection of these variants were converted to (putative) synthetic enzymes and tested for activity.

The structure-guided site saturation mutagenesis of 9 sites was carried out in an attempt to alter the substrate specificity of β -GUS. The choice of the residues to be altered was made with the aid of the structure of β -GUS with a bound substrate analogue. Nine codons in *the* β -GUS gene were randomised to create libraries that contained all possible amino acid of residues in and / or near the active site. Mutants were assayed using substrates presenting 5 different glycones. Of the positions randomised, most glycosyl binding residues were found not to tolerate amino acid substitutions, suggesting they are essential for β -GUS function while majority of non-glycosyl binding residues were found to tolerate amino acid substitutions – but not with good activity

One of the common dogmas of directed evolution is the idea that evolvability is related to stability. We set out to test this idea while evolving substrate specificity. Other workers had generated a more thermostable variant of β -GUS. In parallel, we evolved 1) the native enzyme (GUS-WT) and 2) the thermostable (GUS-TR3337) variant of β -GUS. Mutant libraries of both GUS-WT and GUS-TR3337 were created under identical conditions and had the same distributions of mutations. After five rounds of evolution, the catalytic efficiency (k_{cat}/K_m) of the best mutant of the wild type parent for *p*NP-glucoside was increased \sim 307-fold while the best mutant of the thermophilic parent demonstrates a \sim 4-fold increased k_{cat}/K_m over the best mutant of wild type parent. Selected mutants from both libraries were

Abstract

characterised with regard to conformational properties and stability and these investigations, combined with kinetic data, provided valuable information about how thermostability promotes the ease of protein evolvability.

The initial characterisation of β -GUS variants was done with crude lysates. It was observed that the GUS-WT and GUS-TR3337 variants lost their newly evolved activity after purification. It was eventually determined that the BugBuster™ reagent used to lyse cells prior to screening, had affected the directed evolution campaign. The *n*-octyl β -D-thioglucoopyranoside (OTG) presents in the BugBuster™ reagent, was very similar in structure to *p*NP-glucoside and was identified as a competitive inhibitor that suppresses glucosidase activity of wild-type β -GUS, indicating that it can be bound in the active site. In response to the addition of OTG in the screening assay, OTG enhanced the glucosidase activity of selected mutants. This observation highlights the potential pitfall in the use of commercial reagents to lyse cells for enzymes with glycosidase activities. However, the evolution in the presence of OTG gives some insight into how an enzyme might evolve to be regulated by an effector molecule – OTG, in this case.

Finally, improved cell lysis variants were converted to glucuronylsynthase variants by introducing the site specific mutation. Their glycosynthase ability was tested using a similar protocol developed by McLeod Group for assaying glucuronylsynthase activity. Unfortunately, glycosynthase activity was not observed with α -D-glucuronyl fluoride donor and steroid accepters. Time constraints did not allow other substrate to be tested.

Table of Contents

Declaration	ii
Acknowledgements	iii
Abstract	v
Table of Contents	vii
List of Abbreviations	xii
Chapter 1 – Introduction	1
1.1. Background	2
1.2. β -glucuronidase enzyme	3
1.2.1. The mechanism of β -GUS	4
1.3. Glucuronylsynthase	5
1.4. The glycosylation of steroids	6
1.5. Specific aims of the project	7
1.6. A brief review on protein engineering	11
1.6.1. Rational design	12
1.6.2. Directed evolution	12
1.6.3. Semi-rational approaches	15
1.7. Overview of the thesis	16
1.8. References	17
Chapter 2 – Experimental Methods	22
2.1. Cell strains and expression vectors	22
2.1.1. <i>E. coli</i> strain BW25141	22
2.1.2. <i>E. coli</i> strain GMS407	22
2.1.3. <i>E. coli</i> strain DH5 α	22
2.1.4. Expression vector pJWL1030	23
2.1.5. Expression vector pET-28a(+)	23
2.1.6. Expression vector pJexpress 401	23
2.2. Computer software programs	23
2.3. Molecular biology	24

Table of Contents

2.3.1. Directed evolution	24
2.3.1.1. Introducing genetic diversity	24
2.3.1.1.1. Site-saturation mutagenesis (SSM)	24
2.3.1.1.2. Site-directed mutagenesis (SDM)	25
2.3.1.1.3. Error-prone polymerase chain reaction (ePCR)	26
2.3.1.1.4. Staggered extension process (StEP)	27
2.3.1.1.5. StEP-ePCR	27
2.3.1.2. Preparing libraries of variants for screening	27
2.3.1.2.1. <i>In vitro</i> screening of site-directed libraries with β -glycoside substrates	27
2.3.1.2.2. Colony screening	28
2.3.1.2.2.1. Primary screening of colonies	28
2.3.1.2.2.1.1. Agar plate screening	29
2.3.1.2.2.1.2. Top agar screening	29
2.3.1.2.2.2. Secondary screening of β -GUS activity	29
2.3.2. Protein expression	29
2.3.3. Protein purification	30
2.3.3.1. Immobilised metal affinity chromatography (IMAC).....	30
2.3.3.2. Size-exclusion chromatography (SEC)	30
2.3.4. Analysis of GUS-WT and GUS-TR3337 variants	33
2.3.4.1. Kinetic analysis	34
2.3.4.1.1. Non-cooperative β -GUS enzymes	34
2.3.4.1.2. Negatively cooperative β -GUS enzymes	35
2.3.4.2. Thermal optima	35
2.3.4.3. Thermal denaturation	35
2.3.4.4. Thermal shift assay	36
2.3.4.5. Comparison of protein solubility	36
2.3.4.6. Discontinuous native protein gel electrophoresis analysis	36
2.3.4.7. Bis(sulfosuccinimidyl)suberate (BS3) crosslinking	36
2.3.4.8. Effects of additives	37

2.3.4.9. Determination of the modes of enzyme inhibition and the inhibitory constant (K_i) values	37
2.3.4.10. Enzyme unfolding	37
2.3.5. Enzyme crystallisation screens	37
2.3.6. Steroid glycosynthase reactions	38
2.4. References	38
Chapter 3 – Identification of Residues Critical to Substrate Binding in β-GUS	40
3.1. Introduction	40
3.2. Measurement of activities in crude cell lysates	41
3.3. Choice of sites for site-saturation mutagenesis	42
3.4. Sequence comparison using the ConSurf program	44
3.5. Codon randomisation and screening	46
3.6. Glycosyl binding residues	47
3.7. Non-glycosyl binding residues	49
3.8. Concluding remarks	51
3.9. References	52
Chapter 4 – Stability and the Evolution of New Substrate Specificity	54
4.1. Introduction	54
4.2. Thermostable β -GUS	55
4.3. Directed evolution for glucosidase activity	56
4.3.1. Error rate	57
4.3.2. Generation of GUS-WT and GUS-TR3337 in directed evolution	59
4.3.2.1. GUS-WT and GUS-TR3337 sequence pedigree	60
4.3.2.2. Improvement in GUS-WT and GUS-TR3337 over five generations	64
4.3.2.3. Shuffling of best mutants from each generation	65
4.3.2.4. Analysis of mutation type	65
4.3.2.5. Distribution of mutations in the gene and the protein	66
4.4. Characterisation of GUS-WT and GUS-TR3337 variants	69
4.4.1. Expression and solubility of GUS-WT and GUS-TR3337 variants	69
4.4.2. Kinetics characterisation of GUS-WT and GUS-TR3337 variants	71

Table of Contents

4.4.3. Optimal catalytic temperature	73
4.4.4. Thermal and chemical stability	75
4.4.4.1. Thermal denaturation	75
4.4.4.2. Chemical induced denaturation	78
4.4.5. Summary of observations	81
4.5. Mutation location and effect	82
4.5.1. WT5P26A7 variant	83
4.5.2. THERMO4P11F2 variant	87
4.6. Summary	90
4.7. References	92
Chapter 5 – Kinetic Analysis of GUS-WT and GUS-TR3337 Variants in Directed Evolution	94
5.1. Introduction	94
5.2. Loss of enzyme activity on purification	95
5.3. Effects of 10 X BugBuster™ protein extraction reagent	96
5.3.1. Effect of SDS	101
5.3.2. Effect of OG	101
5.3.3. Effect of OTG	102
5.3.3.1. Inhibitory and activating effects of OTG	103
5.3.3.2. Kinetic analysis of GUS-WT and GUS-TR3337 variants	105
5.3.3.2.1. Assessing negative cooperativity in evolved variants using the Hill equation	111
5.4. Structure – Implications of mutations	115
5.5. Conclusions	118
5.6. References	120
Chapter 6 – Synthesis of Steroidal Glucosides using Glucuronylsynthase	122
6.1. Introduction	122
6.2. Glycosylation reaction of glucuronylsynthase	122
6.3. Extraction for glucuronides from the reaction mix	124
6.4. The glycosylation reaction	125
6.5. Detection of steroidal glucosides from the glycosylation reaction	126

6.6. Concluding remarks	127
6.7. References	127
Chapter 7 – Conclusion and Future Directions	128
7.1. References	134
Appendices	135
A. Chapter 1	135
B. Chapter 2	141
C. Chapter 4	142
D. Chapter 5	145

List of Abbreviations

β -GUS	β -glucuronidase
λ_{\max}	Fluorescence emission maximum
BS ³	Bis(sulfosuccinimidyl)suberate
CMO-DHEA	Dehydroepiandrosterone O-(carboxymethyl)oxime
DHEA	Dehydroepiandrosterone
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
<i>E. coli</i>	<i>Escherichia coli</i>
ePCR	Error-prone polymerase chain reaction
ER	Endoplasmic reticulum
StEP-ePCR	Staggered extension polymerase chain reaction – error-prone polymerase chain reaction
GDL	Glucaro- δ -lactam
GdnHCl	Guanidinium hydrochloride
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β -D-1-thiogalactopyranoside
LB	Luria-Bertani media
LBK	Luria-Bertani-Kanamycin
MS-ESI	Electron-spray ionization mass spectrum
m/z	Mass to charge ratio
OG	Octyl- β -D-glucoside
OTG	Octyl- β -D-thioglucoside
pDNA	Plasmid Deoxyribonucleic acid
pNP	<i>Para</i> -nitrophenol
RMSD	Root mean square deviation
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
StEP	Staggered extension process

List of Abbreviations

SDM	Site-directed mutagenesis
SSM	Site-saturation mutagenesis
Trp	Tryptophan
TX100	Triton-X-100
UDP	Uridine diphosphate
UV	Ultraviolet spectrum
WT	Wild-type
X-glucoside	5-bromo-4-chloro-3-indolyl- β -D-glucoside
YENB	Yeast extract nutrient broth

Chapter 1

Introduction

This thesis describes engineering studies pertaining to *Escherichia coli* β -glucuronidase (*E. coli* β -GUS) (EC 3.2.1.31) enzyme (Doyle *et al.* 1955). This enzyme catalyses the hydrolysis of glycosidic bonds in glucuronides as shown below in Figure 1.1.

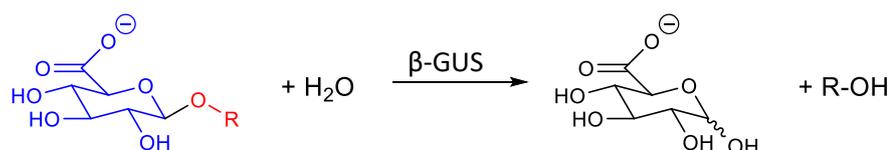


Figure 1.1 General β -GUS hydrolysis reaction for glucuronide

A glucuronide consists a glucuronyl group (blue) and an aglycone component (red).

A brief description of the naming conventions for the glycone and aglycone groups in glycosides is provided in the Appendix A. This appendix gives a description of the relevant nomenclature and is not intended to be a detailed description of carbohydrate chemistry. It should be noted that glucuronides figure prominently in Phase II metabolism of xenobiotics that include drugs. Glucuronic acid is used to tag xenobiotics that are then excreted via the urinary and digestive tracts. *E. coli* is found in the digestive tract of humans and encounters a variety glucuronides being excreted as part of detoxification. β -GUS primarily acts to de-conjugate glucuronides to obtain glucuronic acid, that are then used as a carbon source (Wilson *et al.* 1992). The enzyme is highly specific for the glucuronic acid, but not for the aglycone.

This project was part of a research program aimed at converting β -GUS into a synthetic enzyme capable of forming glycosides. An explanation as to how the catalytic reaction is altered is given later in this chapter. Given that the β -GUS is highly specific for glucuronic acid and not for the aglycone, the synthase formed from it has the potential of producing a variety of glucuronides. The aim of the present thesis was firstly to convert β -GUS into a glucosidase – that is, to change its substrate specificity. It was intended that these evolved glucosidases would then be converted to synthetic enzymes.

This introductory chapter has a brief description of β -GUS that includes its structure and mechanism. Also described are the changes used to bring about the synthetic activity along within the principal target for synthesis – the glycosylation of steroids. A brief

introduction to the methods used to alter the substrate specificity is given. The emphasis of this latter section is on directed evolution that was the main tool used to alter substrate specificity.

1.1. Background

Exogenous (xenobiotics) and endogenous (products of metabolism) compounds are eliminated from the body principally through excretion in urine. Hydrophilic drug molecules can be excreted in urine. However, hydrophobic compounds, such as steroids, are less likely to enter into the aqueous medium and will remain “stored” in the body. These molecules require metabolic transformation in the liver to convert them into more hydrophilic derivatives. An effective means of increasing their hydrophilicity consists in their conjugation with hydrophilic molecular species, like glucuronic acid. For example, in phase II detoxification, isomers of glucuronosyltransferase in the lumen of the endoplasmic reticulum (ER) catalyse conjugation with uridine diphosphate (UDP) glucuronic acid to form hydrophilic glucuronide conjugates as shown schematically in Figure 1.2 (Burchell & Coughtrie 1989). Once synthesized, those glucuronides are then usually excreted from the body via the bile ducts into the intestine, through apocrine secretions, and through the bladder into the urinary tract.

Excreted glucuronides in the intestinal lumen may undergo de-conjugation by commensal microorganisms, for example, *E. coli*. Most *E. coli* strains living in natural environments possess β -GUS and are able to utilize glucuronides as nutrients (Chang *et al.* 1989). The diversity of glucuronide compounds that reach the intestine has resulted in *E. coli* β -GUS evolving to cleave almost any aglycone at its β -linkage to glucuronic acid (Wilson *et al.* 1995). The released glycone can then be used as a carbon source and the apolar aglycone can frequently be reabsorbed by the host (Hazenberg *et al.* 1988). Due to its capacity to act on a broad range of β -glucuronide conjugates, *E. coli* β -GUS has widespread application in the field of analytical chemistry for the de-conjugation of glucuronide metabolites.

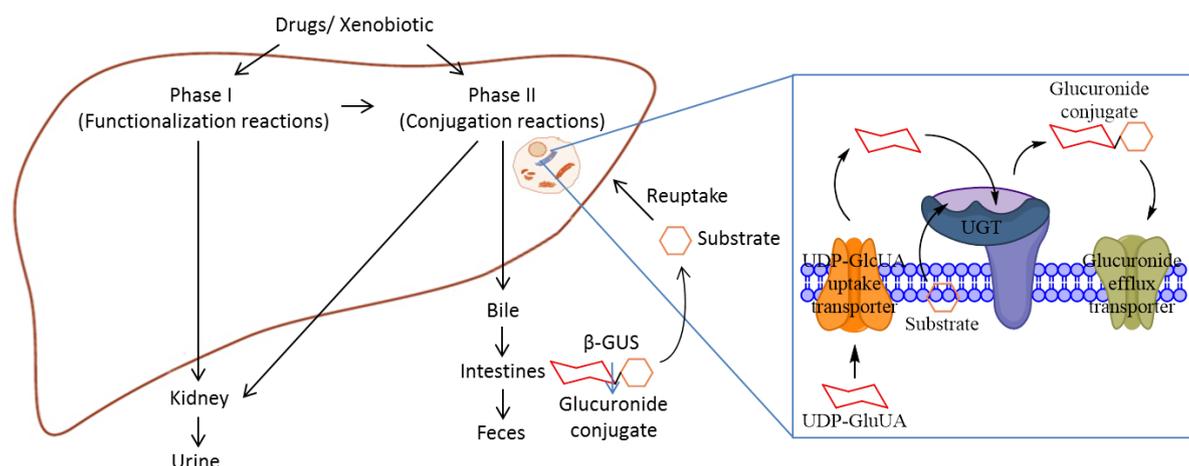


Figure 1.2 Detoxification pathway

Xenobiotic molecules may undergo phase I functionalization reactions alone or phase II biosynthetic conjugation reactions alone while some undergo phase I reactions followed by phase II reactions. A phase I reaction introduces polar functional groups, i.e. hydroxyl or amino, onto lipophilic toxicant molecules. Phase II metabolism involves the introduction of hydrophilic endogenous species, i.e. glucuronic acid or sulfate, to the drug molecule. In the Phase II glucuronylation reaction, glucuronides are synthesized by UDP-glucuronosyltransferase (UGT) in the lumen of the liver ER and eliminated via the kidney or bile. The glucuronides excreted in the bile may be hydrolysed in the intestines by β -GUS and the unbound toxin can then be reabsorbed via the enterohepatic recirculation.

1.2. β -glucuronidase enzyme

As mentioned above, β -GUS is an exoglycosidase and is a member of the glycosidase enzyme family that catalyses the hydrolysis of *O*-glycosidic bonds. They exist in a range of animals, i.e. humans (Islam *et al.* 1999), dogs (Ray *et al.* 1998), rats (Nishimura *et al.* 1986); plants, i.e. rye (Schulz & Weissenböck 1987), rhubarb (Hodal *et al.* 1992), skullcap (Levy 1954); and microorganisms, i.e. *Escherichia coli* (Jefferson *et al.* 1986), *Clostridium* (Fujisawa *et al.* 2001), *Staphylococcus* (Arul *et al.* 2008). The β -GUS of interest in this study is from *E. coli*.

E. coli β -GUS is a member of the retaining β -glycosidase family 2 (Henrissat 1991). It exists as a homo-tetramer with subunits related by three perpendicular twofold axes. The subunits consist of 603 amino acids that have a molecular weight of 68 kDa (Aiba *et al.* 1996). The bacterial enzyme folds into three domains: i) the N-terminal resembling a sugar-binding domain that consists of 180 residues, ii) an immunoglobulin-like- β -sandwich domain that consists of 96 residues and iii) the C-terminal domain that folds into a $\alpha\beta$ -barrel (TIM-barrel) that consists of 327 residues (residues 276 to 603). Domain I and II are important structural elements of the enzyme while the domain III is the catalytic domain that contains the active-site residues E413 and E504 (Figure 1.3). It is worthwhile mentioning that a small section of the loop from domain I (residues 161 to 164) also forms part of the active site.

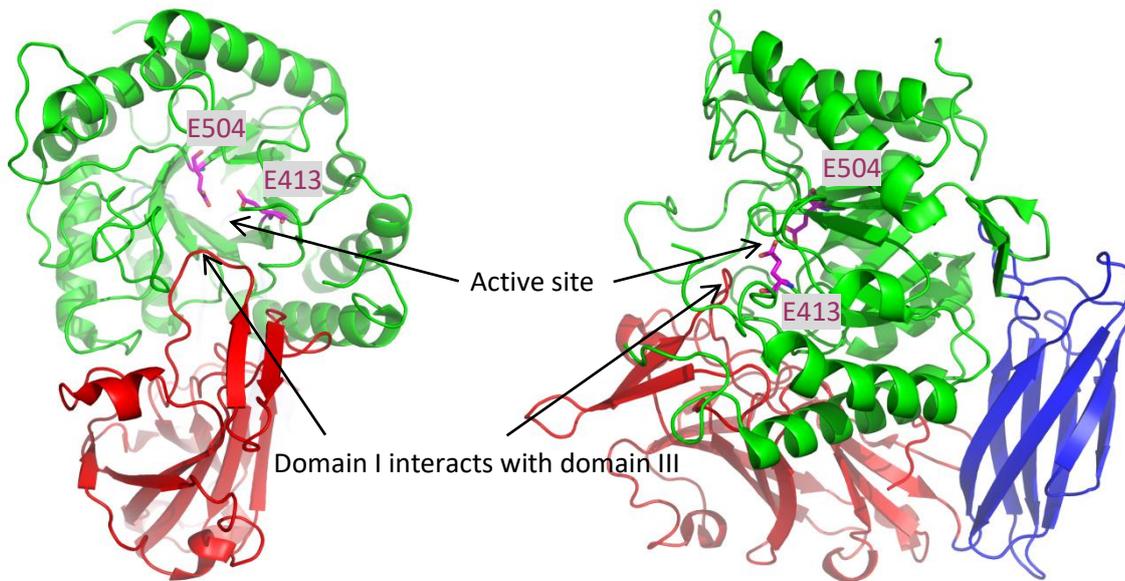


Figure 1.3 Structure of *E. coli* β -GUS (PDB ID: 3K4D)

Domain I is coloured red, domain II coloured blue and domain III coloured green. Catalytic residues E413 and E504 are represented as sticks. The left side of the monomer in the top view and the right side of the monomer in the side view.

1.2.1. The mechanism of β -GUS

The mechanism of action for β -GUS has been well characterised (Zechel & Withers 2000) and is depicted in Figure 1.4. β -GUS catalyses the hydrolysis of glycosidic linkages by a double displacement mechanism with retention of the stereochemical configuration of the anomeric carbon (Collins *et al.* 2005). Its active site consists of two strategically positioned glutamic acid residues: E413 and E504 that provide a general acid/base and nucleophile, respectively (Sinnott 1990) as shown in the figure 1.4 below. The first step in catalysis involves a nucleophilic attack by carboxylate anion of E504 on the anomeric carbon of the glucuronide to form a covalent that is broken down by the action of E413 that functions as a general acid (Figure 1.4) (Liang *et al.* 2005).

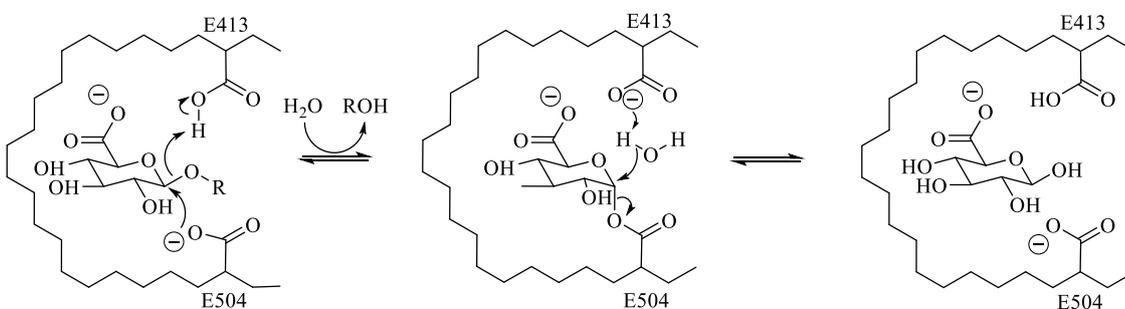


Figure 1.4 β -GUS mechanism of action

If an alcohol is the attacking species instead of water then a β -trans-glycosylated product results. This was known to occur and can be exploited to synthesize oligosaccharides.

However, the trans-glycosylated product remains a substrate for hydrolysis, especially under aqueous conditions, so that poor yields occur unless careful substrate selection or perturbation of the enzymatic equilibrium occurs (i.e. with organic co-solvents) (Johansson *et al.* 1986).

1.3. Glucuronylsynthase

Glucuronylation activity can be acquired by mutating the nucleophilic residue of β -GUS to a non-nucleophilic residue as shown below in Figure 1.5. The three glucuronylsynthase enzymes utilized in this project, were prepared by mutating the glutamate at position 504 (E504) to either an alanine (E504A), a glycine (E504G), or a serine (E504S) residue. This renders the residue incapable of nucleophilic attack at the C1 position, thereby preventing degradation of the glucuronide product (Figure 1.5). However, the rest of the active site remains intact and formation of glucuronide product can be catalysed in the presence of a synthetically-derived α -D-glucuronyl fluoride and an acceptor alcohol substrate. The α -D-glucuronyl fluoride donor with anomeric configuration opposite to that of the original substrate (β -glucuronide), mimicking the enzyme-linked intermediate in the β -GUS. The fluoride substituent at the anomeric carbon acts as a good leaving group for general acid-base catalysis by the E413 residue. The leaving group is displaced by an alcohol leading to the formation of a new glycosidic bond. The reverse mechanism (fluorination) is essentially absent since the fluoride ion is not bound strongly to the enzyme and will be dispersed into the medium following glucuronide synthesis. Additionally, the newly formed glucuronide has a poor anomeric leaving group and won't be displaced by poor nucleophiles such as a fluoride ion.

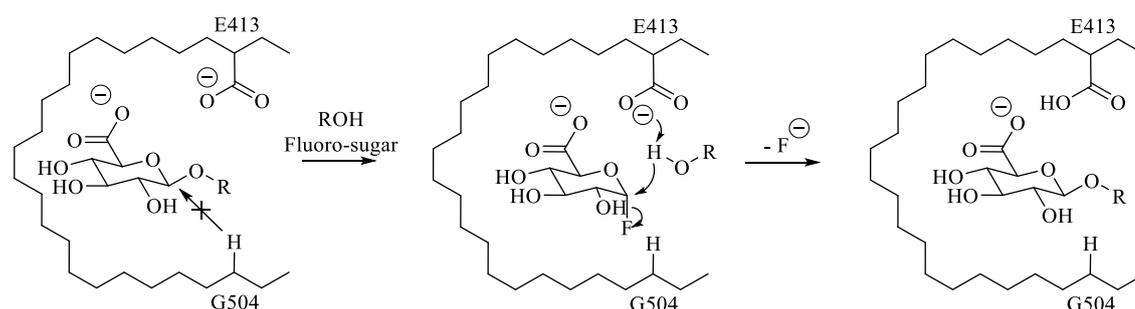


Figure 1.5 General mechanism for the *E. coli* β -glucuronylsynthase

Most glycosynthases are specific for the acceptor sugar (Perugino *et al.* 2004, Perugino *et al.* 2003, Trincone *et al.* 2000). Interestingly, after screening which aglycons would be susceptible to glucuronylation by the glucuronylsynthase enzyme, it was found that non-carbohydrate acceptors such as primary, secondary and aryl alcohols were better candidates

for glucuronylation than carbohydrate-based acceptors (Wilkinson *et al.* 2008). While a few steroid substrates were successfully used to synthesise steroid glucuronides, a screen of a suitably diverse steroid library was performed by Ma *et al.* (2014). Out of eighteen different steroidal alcohols, fifteen substrates providing glucuronide conjugate in high purity and on a scale suitable for ^1H NMR analysis.

1.4. The glycosylation of steroids

Among the vast array of glycol-conjugates, steroidal glycosides have drawn much attention in the last few decades as economically important raw materials for the pharmaceutical industry in the production of various steroidal hormones (Görög 2011, Fernandes *et al.* 2003, Hwang *et al.* 1999). The compounds are often biologically active (Francis *et al.* 2002, Bedir *et al.* 2002, Yokosuka 2015) and as such are also used as ingredients for cosmetics (Kashibuchi *et al.* 1996). It should come as no surprise that there has been extensive work on the chemical syntheses of oligosaccharides and steroids (Ma *et al.* 2001, Tang *et al.* 2013, Fernandez-Herrera *et al.* 2012) as well as conjugates of the two compound classes. The key to synthesising steroidal glycosides is the construction of the glycosidic bond between a sterol with the sugar anomeric carbon (Figure 1.6).

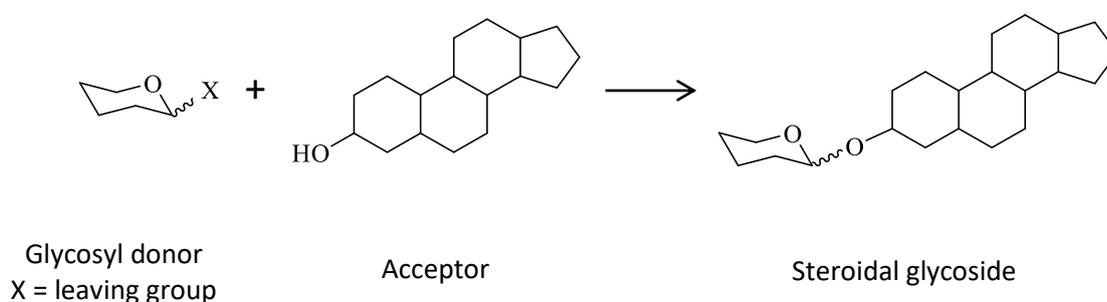


Figure 1.6 Glycosylation of steroids

Despite the significant advances observed in carbohydrate chemistry, the stereoselective formation of *O*-glycosidic bonds between carbohydrates and steroids remains time-consuming due to the requirement of multiple protection and deprotection steps that have relatively low yields due to the low reactivity glycosyl donor and glycosyl acceptor (such as the secondary alcohol functions in the steroid moiety). The need for more efficient approaches has stimulated the development of enzyme-mediated glycosylation. Such an approach avoids the need for stringent reaction conditions which allow synthesizing unstable glycosides and any protecting groups that allow one-step glycosylation (Crout & Vic 1998).

Both glycosyltransferases and glycosidases have been vigorously studied for synthetic purposes (Shaikh & Withers 2008, Hancock *et al.* 2006, Bennett & Wong 2007). Glycosyltransferases are the natural enzymes for catalysing the formation of glycosidic bonds in high yields with excellent selectivity, i.e. UGTs, were employed by Werschkun *et al.* (1998) to synthesize β -D-glucuronides from uridine UDP- α -D-glucuronic acid and a wide variety of endo- as well as xenobiotic aglycons. While the inherent nature of glycosidases is to break glycosidic linkages in carbohydrate metabolism, there have been increasing interest in exploring their transglycosylation activity for making complex carbohydrates. A β -galactosidase from *Aspergillus oryzae*, showed a high transglycosylation activity and as such it has been involved in the enzymatic synthesis of various chemically unstable cardiac glycosides (Ooi *et al.* 1984). Although glycosyltransferase-catalyzed reaction provides a method for mild and stereospecific single-step synthesis without the need for protected glycosyl donors, its application has been limited by two major reasons. Firstly, glycosyltransferases require a sugar nucleotide as glycosylation donor and this is an unstable and expensive substrate. Secondly, they are substrate-specific to the acceptors, which limits the range of substrates that can be used as starting materials (Williams *et al.* 2007). Glycosidases, on the other hand, have the advantage of being readily available commercial enzymes and they also use simple glycosyl donors and have a relaxed substrate specificity for acceptors (Trincone *et al.* 2003). Nevertheless, glycosidases are still hydrolytically active, the product of this transglycosylation reaction can be hydrolysed, thus reducing yields of the product obtained (McCarter & Withers 1994).

The drawbacks mentioned above, were circumvented by the invention of glycosynthases (Mackenzie *et al.* 1998). As demonstrated for a range of other retaining β -glycosidase enzymes, single point mutation of the nucleophilic residue to a non-nucleophilic residue disables the hydrolytic pathway (Mackenzie *et al.* 1998). This concept had been pioneered by the McLeod group to design a feasible glucuronylsynthase from β -glucuronidase (β -GUS) that could be used in chemoenzymatic glucuronylation of steroid substrates (Ma *et al.* 2014). This thesis sought to extend these early investigations.

1.5. Specific aims of the project

To the author's knowledge, glucuronylsynthase was the only reported glycosynthase that catalysed glycosidic bond formation with a wide range of alcohol and steroidal acceptors while most of the engineered glycosynthase catalysed the formation of oligosaccharides. The main aim of this project was to convert the sugar donor specificity of *E. coli* glucuronylsynthase from

glucuronic acid to glucose. If successful, this will open up a new path to flavonoid glycoside synthesis (Figure 1.7) where the new enzyme biocatalysts will grant synthetic access to advanced organic materials that are in high demand in the fields of pharmaceutical development and chemical analysis. Moreover, a great deal about enzymes can be learnt from the process. For example, what type of mutation is required for changing substrate specificity? How do proteins evolve – are stabilizing mutations required before active site mutations are observed? Are mutations that affect the dynamics of the enzyme important? These are the type of questions we wanted to address when mutant proteins produced in this study were characterised.

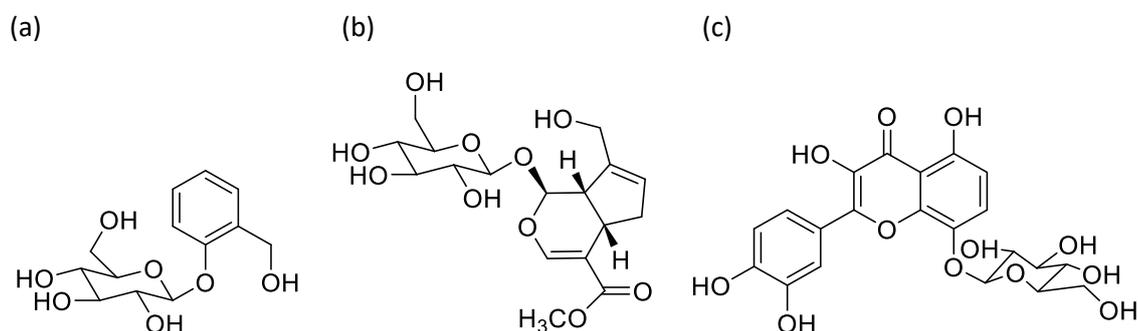


Figure 1.7 flavonoid glycosides
(a) Salicin (b) geniposide (c) gossypin.

To attain the stated goals of changing substrate specificity, directed evolution was thought to be the method of choice. An essential requirement for the success of these experiments is the ability to screen large libraries for the desired property. In the case of glycosynthases, the development of a screening assay is challenging, because the reaction products do not produce a distinct, screenable signal, such as absorbance or fluorescence. Directed evolution is not uncommon amongst glycosynthases. To date, there have been two elegant approaches of high throughput screening applied to glycosynthases. These include chemical complementation and enzyme-coupled spectrophotometry that are described briefly in the following paragraphs.

Chemical complementation was developed by the Cornish group and involves transcriptional modification brought about by glycosynthase activity (Lin *et al.* 2004). Using a yeast three-hybrid system, dihydrofolate reductase (DHFR) binds to the methotrexate (MTX) moiety and the glucocorticoid receptor (GR) binds to the dexamethasone moiety (DEX) (Figure 1.8). The DHFR is fused to a DNA binding domain (DBD), which binds to a specific DNA binding site upstream from the *LEU2* gene. The GR is fused to a transcription activation domain which up-regulates the transcription of the *LEU2* gene. The formation of oligosaccharide linkage via

a glycosynthase reaction results in the reconstitution of a transcriptional activator that can promote the transcription of the *LEU2* gene. The *LEU2* gene is an essential gene in the biosynthesis of leucine and its up-regulation allows cells to grow in leucine-deficient media. A novel selection criterion is achieved by relating glycosynthase activity to cell survival. This makes this assay a high throughput selection as opposed to a screen as only mutants with appreciable glycosynthase activity will survive. Chemical complementation is an ingenious and complex model as a high throughput selection for glycosynthase activity. Unfortunately, the system relies on oligosaccharide formation through the use of a small molecule disaccharide acceptor substrate and a small molecule disaccharide α -fluoro donor substrate, the synthesis of which is a multistep endeavour (Tao *et al.* 2006). As a result, the glucuronosyltransferase needs the small molecule steroid acceptor substrate to drive the glycosidic bond formation. This requires the design and synthesis of the new acceptor substrate. In addition, the small molecules attached both to the donor and the acceptor might be too bulky to enter the active sites due to the fact that β -GUS having an active site that possesses pocket topology. This makes chemical complementation unsuitable for the glucuronosyltransferase high throughput screen.

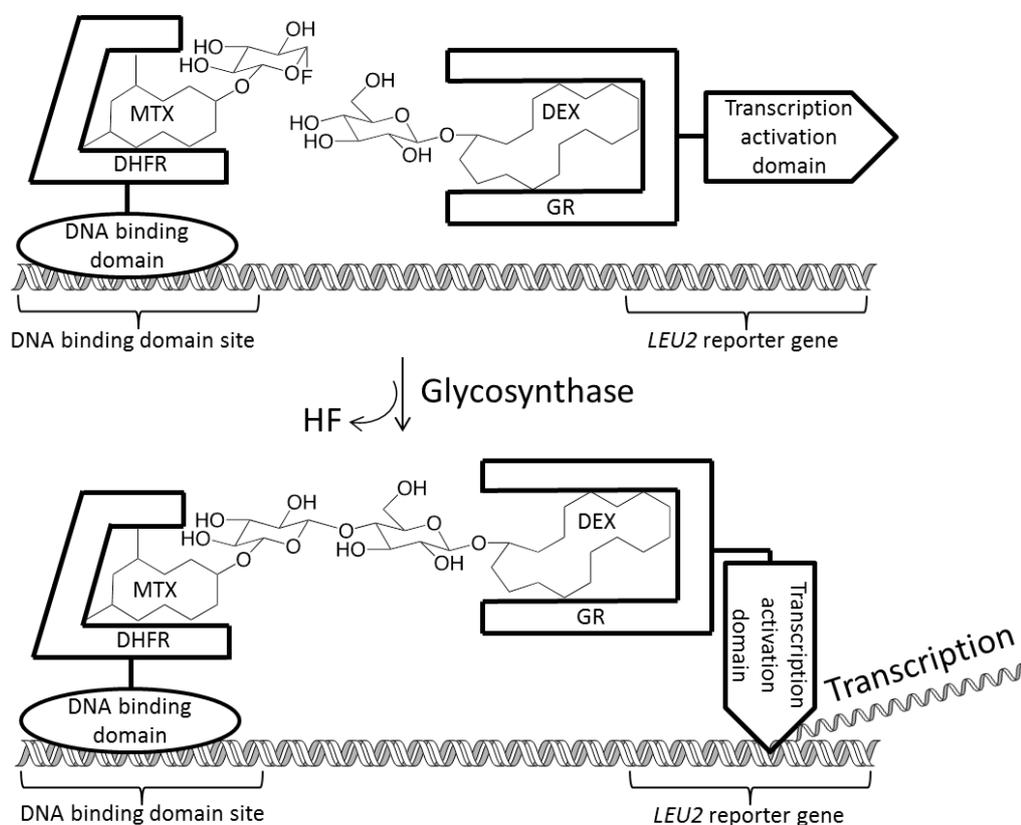


Figure 1.8 Cornish's chemical complementation model

Formation of the glycosidic bond via glycosynthase reaction up-regulates *LEU2* transcription.

The second approach, enzyme-coupled spectrophotometric assay, was developed in the Withers group and is based on the hydrolysis of a fluorescent or colourimetric tag from the glycosynthase product (Kim *et al.* 2004). It involves the reaction of a glycosyl fluoride donor with a glycoside acceptor bearing a chromophore. An *endo*-glycosidase, that recognises the newly formed oligosaccharide, hydrolyses the glycosidic bond linking the chromophore which can be measured spectrophotometrically (Figure 1.9). The hydrolysis step is usually of many magnitudes faster than the glycosynthase step, so the glycosynthesis is the rate-determining step. The rate of fluorescence or colour formation is therefore directly related to the glycosynthase activity. This assay, however, is unsuitable for the glucuronylsynthase high throughput screen because a new specific chromophore-containing steroid acceptor has to be synthesized, and an endoglycosidase that cleaves the linkage between steroids and chromophores, if it exists at all, should be found.

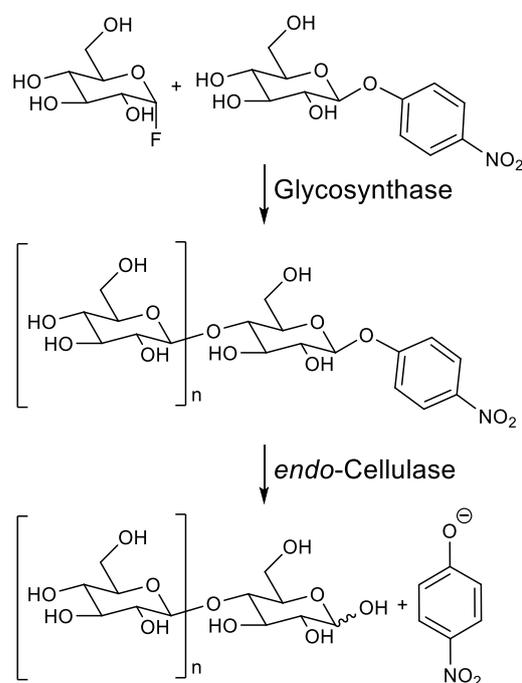


Figure 1.9 An example of a glycosynthase reaction coupled with *endo*-cellulase to hydrolyse the 4-nitrophenolate chromophore in an enzyme-coupled high throughput screen $n = 1$ or 2 (Mayer *et al.* 2001).

It should be clear from the previous two paragraphs that the two major high throughput screening assays designed for glycosynthases are not applicable for glucuronylsynthase. For this reason, a two-step strategy was employed in the present research. This approach involves i) evolving the wild-type β -GUS to have higher glucosidase activity and ii) converting improved β -GUS enzymes to glucuronylsynthases by incorporating a mutation at position 504. The screening assay applies *para*-nitrophenyl- β -D-glucuronide (*p*NP-glucuronide) that liberates

para-nitrophenolate on hydrolysis which absorbs strongly in the visible light spectrum (405 nm) and can be monitored over time by spectrophotometry (Figure 1.10). The major advantage of this approach is that the substrate required for the assay *p*NP-glucuronide is commercially available and the assay is rapid and sensitive. This two-step strategy that had been proposed is a unique approach and has never been reported before. More importantly, it offers an alternative that circumvents the need for the development of a high-throughput assay for glucuronylsynthase.

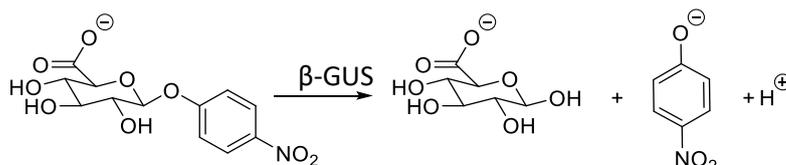


Figure 1.10 The hydrolysis of *p*NP-glucuronide by β -GUS

1.6. A brief review on protein engineering

To achieve the aims, protein engineering technologies will be applied. Many of these techniques have been described in the literature. This section below gives a review of the technologies that are available and details how they will be adopted to design suitable engineering approaches.

Enzymes evolution takes place over times scales of millions years. It is a very slow process for which natural selection as the sole driving force. It has been suggested that during the course of this enormous length of time that the resultant enzymes should be near perfect. However, because enzyme evolution is driven by selection, enzymes only evolve to the point where they satisfy requirements of the host organism, and no more. This state of affairs usually means that enzymes may not ideally suit for industrial applications. For example, enzymes may often be unsuitable because of stability issues. In many cases, there may be no enzyme to catalyse the target reaction and even if an enzyme is available, a great deal of work is often required to optimize its activity for practical applications. Hence the need for protein engineering.

There are a number of ways to engineer or modify proteins; it can be achieved by chemical techniques, but the process is not specific and usually results in a considerable loss of protein (Means & Feeney 1990). Direct synthesis can also be used to produce proteins, including proteins with modified residues, but the process is expensive and is not simple to execute. Neither of the above approaches is a suitable method to produce proteins or modified proteins for industrial applications. The idea that proteins could be modified in a

specific way didn't become a realistic proposition until the development of molecular biology – the ability to manipulate DNA. Initially this tool led to studies that were aimed at understanding protein structure function. Residues were mutated site specifically and changes in activity were monitored. The process of engineering proteins for practical applications was accelerated by the discovery of the polymerase chain reaction (PCR). This technology allowed large mutant libraries with a specified average error rate to be generated in the laboratory (Cadwell & Joyce 1994). This technical development meant that natural evolution could be mimicked on a laboratory time-scale and with the intent of selecting for traits desired for practical applications; this became known as directed evolution. It should be noted that the technology and knowledge in the field of genetic engineering has advanced significantly in the last few years. In addition to the evolutionary approach (Bornscheuer & Pohl 2001), one can use rational design (Steiner & Schwab 2012) to enhance protein properties. One often views different techniques as competitors, but this is not the case with the different approaches to engineering proteins as predictions made by rational methods can be incorporated into libraries to be screened in directed evolution.

1.6.1. Rational design

Rational design uses prior knowledge of the candidate protein and molecular modelling to predict amino acid substitutions that may yield the desired properties. It requires a thorough knowledge of protein structure and function to enable computational predictions of molecular dynamics and behaviour (Pikkemaat *et al.* 2002). Specific amino acids substitutions will then be introduced by site-directed mutagenesis techniques. This was the approach used to convert an improved β -GUS into a glucuronylsynthase, where glutamate is replaced by alanine/glycine/serine specifically at position 504.

Although rational approaches to protein engineering have yielded some impressive results, it does have its limitations as only a small number of modifications can usually be made and there are many aspects of protein function that cannot be predicted (Ness *et al.* 2000).

1.6.2. Directed evolution

As noted above, directed evolution or *in vitro* evolution is a strategy that mimics natural evolution on a laboratory timescale. As with natural evolution, the process usually involves iteration of two steps – the generation of a genetically diverse, yet homologous population, and then application of an assay to screen or select for improved clones (Figure 1.11).

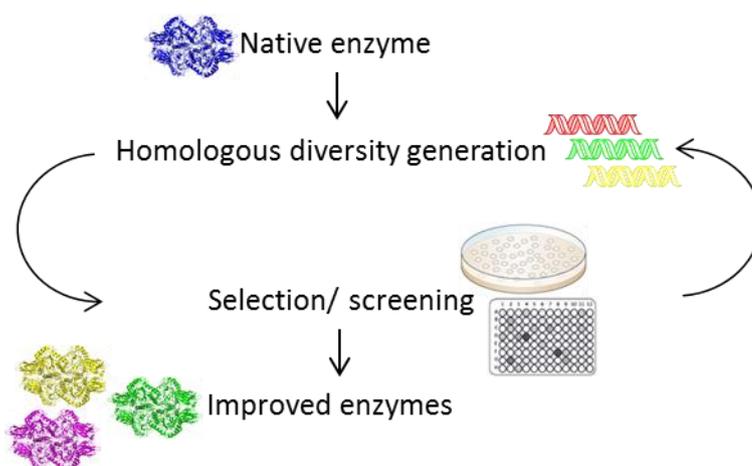


Figure 1.11 General strategies for the directed evolution of enzymes

Population diversity can be generated *in vitro* either by random mutagenesis or gene recombination. However, by far the most popular *in vitro* mutagenesis strategy is error-prone PCR (ePCR), that relies on the misincorporation of nucleotides by DNA polymerase to generate point mutations in a gene sequence (Cadwell & Joyce 1992) (Figure 1.12a). The accuracy of DNA polymerase (*Taq* polymerase) can be adjusted *in vitro* by addition of manganese ion (Mn^{2+}) into the PCR reaction mixture (Zhao *et al.* 1999) and the mutations can be restricted to a specific gene or even a section of the gene – both important points for the construction of mutant libraries. Gene recombination involves the fragmentation and subsequent recombination of multiple DNA sequences with the aim of combining beneficial mutations (Stemmer 1994b). Recombination also makes it possible to remove neutral or deleterious mutations, which accumulate during random mutagenesis, by backcrossing progeny with excess parental or wild-type DNA (Stemmer 1994b). The first gene recombination method that mimic genetic recombination, DNA shuffling, was introduced by Stemmer (1994a). Shuffling is performed on a set of naturally occurring homologous sequences using Deoxyribonuclease I to fragment the genes, which are then reassembled in a self-priming PCR reaction according to their sequence homology (Figure 1.12b). However, this method requires relatively high homology between the parental genes yet crossovers tend to aggregate in regions of high sequence identity due to the annealing-based reassembly. Fortunately, a group of homologous gene recombination methods that do not involve DNA fragmentation but require addition of primers were developed, and staggered extension process (StEP) was among them (Zhao & Zha 2006). This technique consists of many cycles of denaturation and extremely short extension. This effectively generates polynucleotide fragments that repeatedly switch template at every denaturation step and grow very slowly at each extension stage until the full-length is reached (Figure 1.12c). To further increase the efficiency of

introducing homologous diversity, an elegant one-step approach that integrates ePCR and StEP, was developed in our group. StEP-ePCR consists of a modification of the normal ePCR protocol to allow for the recombination of generated mutants during the process of mutation introduction as shown in the figure below (Figure 1.12c).

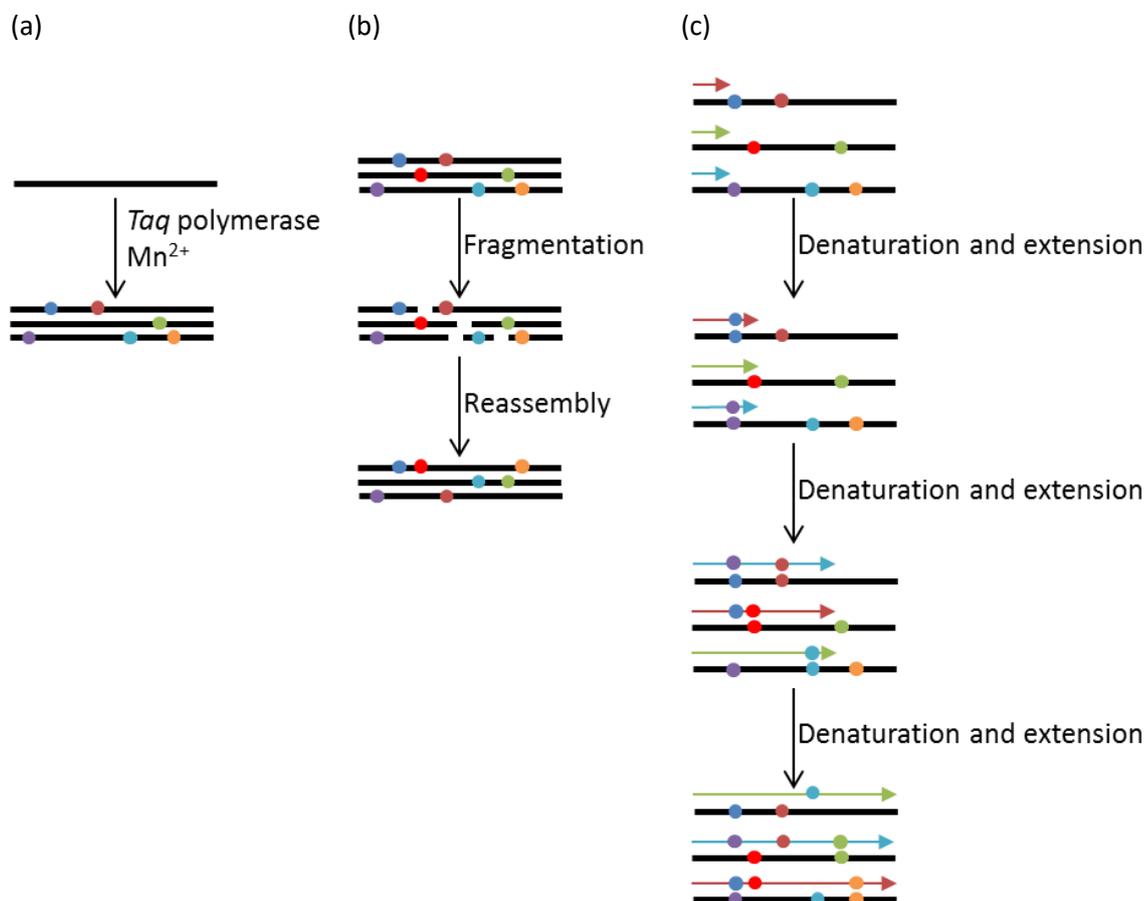


Figure 1.12 Common methods of library generation for directed evolution

(a) ePCR (b) DNA shuffling by fragmentation and reassembly (c) DNA shuffling by StEP.

The second step in directed evolution is the identification of the improved variants by either screening or selection. Selection and screening are very enzyme dependent and are frequently the limiting steps in directed evolution. Selection techniques are agar plate based and rely on a direct correlation between cell survival and the desired enzyme function. Cells are transformed with an enzyme library and plated on selective media where only cells containing a viable variant will survive. Selections are typically limited to antibiotic resistant enzymes (i.e. β -lactamases) or enzymes involved in the synthesis of nutrients essential for cell growth (Watson *et al.* 2007). Selection methods are generally less laborious and allow larger libraries of the order of 10^5 to be evaluated. However, it is not always possible to employ selection as the evaluation method, as not all enzymes can be linked to cell survival. Also, selection methods may not give a good quantitative response, in which case a secondary

screen will be required. A screen is more versatile than selection, involving a direct measurement of enzymatic activity by a high-throughput assay and can be carried out using bacteria growing in the wells of microtiter plates or on agar plates. In both cases, the enzyme assays are carried out inside the cell and can be followed by visual inspection if the reaction involves the formation (or degradation) of a coloured or fluorescent substrate. In the agar plate screening method, the desired biochemical activity is often tied to a phenotype. Thus, it can be applied to large mutant libraries of the order of $10^4 - 10^5$. Similarly to selection methods, agar plate screening methods are not suitable for quantifying catalytic activities of individual variants in the library, making them ideal as pre-screening methods. In contrast, microtiter plate screening enables quantitative activity measurements of each screened variant thus providing a more complete dataset for hit evaluation. A large diversity of enzymatic assays can be ported to the microtiter plate format. Correspondingly microtiter plate screening is significantly more labour intensive and is therefore generally limited in scope to the order of 10^4 individuals for screening (Arnold & Georgiou 2003). Both agar and microtiter plate screenings were used during the course of the research described herein and will be described in greater detail in Chapter 2.

For over two decades, directed evolution has proven to be an effective strategy for tailoring enzyme properties to the needs of industrial, research and therapeutic applications (Farinas *et al.* 2001, Reetz 2006, Toscano *et al.* 2007). It allows the analysis of very large numbers of mutants, each containing many different sets of concerted changes all of which contribute to a marked enhancement in the property – often several orders of magnitude. In addition, valuable insights can be obtained into the structure-function relationships of recently isolated and less studied enzymes (Tee & Schwaneberg 2007, Wong *et al.* 2006). The main advantage of directed evolution, compared with rational approaches to engineering, is that it does not require prior knowledge of enzyme structure and mechanism. However, several rounds of evolution have to be applied and a large number of variants have to be screened in directed evolution. This is especially true while evolving large enzymes, i.e. β -GUS. For a protein of 600 amino acids (i.e. β -GUS) with random changes in just 1, 2 or 3 amino acids in the whole protein, the numbers of variants screened (the library size) required to achieved 95% expected coverage are 1.08×10^4 , 1.94×10^7 and 2.32×10^{10} , respectively (the working for these calculations are given in Appendix A.2.).

1.6.3. Semi-rational approaches

To avoid some of the tedium of directed evolution, semi-rational approaches are often considered as they combine the advantages of rational and evolutionary methods (Chica *et al.*

2005). An example for a semi-rational approach is site-saturation mutagenesis, that uses the information derived from structural data to identify amino acids in interesting regions (i.e. active site), which are then mutated randomly, in either an individual or combinatorial fashion. This method utilizes degenerate oligonucleotides as primers to create a library with all 20 possible amino acids introduced at the residue positions of interest. The mutagenic primers used for site saturation mutagenesis experiments in this thesis utilized the NNK codon (where N = A, C, G or T and K = G or T) representing 32 possible codons code for all 20 amino acids.

1.7. Overview of the thesis

The aim of the thesis is to alter the substrate specificity of β -GUS so that it can act on glucoside substrates. Mutants with altered specificity will be further mutated to produce potential synthetic enzymes that can be tested for activity. The remaining chapters have information as follows:

Chapter 2. Description of routine experimental procedures used in this work. These procedures crop up in a number of chapters and are referred to in these chapters. The procedures are routine and range from molecular biology to protein purification and enzymology. These are routine procedures that reader may want to initially skip and come back to later when referenced in subsequent chapters.

Chapter 3. The substrate profile of β -GUS on various glycosides was investigated and site-saturation experiments aimed at altering the substrate specificity were conducted.

Chapter 4. Directed evolution was applied to β -GUS and laboratory-evolved thermostable β -GUS (GUS-TR3337) to produce enzyme with altered substrate specificity and test the idea that stability further enhances evolution of new function.

Chapter 5. Although directed evolution has been successfully used to improve the glycosidase activity of β -GUS and GUS-TR3337, the glucosidase activities obtained for purified mutants were significantly lower than activities measured in crude extracts. This chapter sought to understand the possible reason for loss of activity in purified mutants.

Chapter 6. Mutations E504A/G/S were introduced into β -GUS, GUS-TR3337 and 3 improved variants obtained in Chapter 4, to produce the glucuronylsynthase variants. These variants were tested for glycosynthase activity.

Chapter 7. Brief summary and concluding remarks

1.8. References

- Aiba, H., T. Baba, K. Hayashi, T. Inada, K. Isono, T. Itoh, H. Kasai, K. Kashimoto, S. Kimura, M. Kitakawa, M. Kitagawa, K. Makino, T. Miki, K. Mizobuchi, H. Mori, T. Mori, K. Motomura, S. Nakade, Y. Nakamura, H. Nashimoto, Y. Nishio, T. Oshima, N. Saito, G.-i. Sampei, Y. Seki, S. Sivasundaram, H. Tagami, J.-i. Takeda, K. Takemoto, Y. Takeuchi, C. Wada, Y. Yamamoto & T. Horiuchi (1996) A 570-kb DNA Sequence of the *Escherichia coli* K-12 Genome Corresponding to the 28.0–40.1 min Region on the Linkage Map. *DNA Research*, 3, 363-377.
- Arnold, F. H. & G. Georgiou. 2003. *Directed enzyme evolution: Screening and selection methods*. Totowa, New Jersey: Humana Press.
- Arul, L., G. Benita & P. Balasubramanian (2008) Functional insight for beta-glucuronidase in *Escherichia coli* and *Staphylococcus* sp. RLH1. *Bioinformation*, 2, 339-43.
- Bedir, E., I. A. Khan & L. A. Walker (2002) Biologically active steroidal glycosides from *Tribulus terrestris*. *Pharmazie*, 57, 491-3.
- Bennett, C. S. & C. H. Wong (2007) Chemoenzymatic approaches to glycoprotein synthesis. *Chem Soc Rev*, 36, 1227-38.
- Bornscheuer, U. T. & M. Pohl (2001) Improved biocatalysts by directed evolution and rational protein design. *Current Opinion in Chemical Biology*, 5, 137-143.
- Burchell, B. & M. W. Coughtrie (1989) UDP-glucuronosyltransferases. *Pharmacol Ther*, 43, 261-89.
- Cadwell, R. C. & G. F. Joyce (1992) Randomisation of genes by PCR mutagenesis. *PCR Methods Appl*, 2, 28-33.
- Cadwell, R. C. & G. F. Joyce (1994) Mutagenic PCR. *Genome research*, 3, S136-S140.
- Chang, G. W., J. Brill & R. Lum (1989) Proportion of beta-D-glucuronidase-negative *Escherichia coli* in human fecal samples. *Appl Environ Microbiol*, 55, 335-9.
- Chica, R. A., N. Doucet & J. N. Pelletier (2005) Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr Opin Biotechnol*, 16, 378-84.
- Collins, T., C. Gerday & G. Feller (2005) Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiol Rev*, 29, 3-23.
- Crout, D. H. & G. Vic (1998) Glycosidases and glycosyl transferases in glycoside and oligosaccharide synthesis. *Curr Opin Chem Biol*, 2, 98-111.
- Doyle, M. L., P. A. Katzman & E. A. Doisy (1955) Production and properties of bacterial beta-glucuronidase. *J Biol Chem*, 217, 921-30.
- Farinas, E. T., T. Bulter & F. H. Arnold (2001) Directed enzyme evolution. *Curr Opin Biotechnol*, 12, 545-51.
- Fernandes, P., A. Cruz, B. Angelova, H. M. Pinheiro & J. M. S. Cabral (2003) Microbial conversion of steroid compounds: recent developments. *Enzyme and Microbial Technology*, 32, 688-705.

- Fernandez-Herrera, M. A., H. Lopez-Munoz, J. M. Hernandez-Vazquez, L. Sanchez-Sanchez, M. L. Escobar-Sanchez, B. M. Pinto & J. Sandoval-Ramirez (2012) Synthesis and selective anticancer activity of steroidal glycoconjugates. *Eur J Med Chem*, 54, 721-7.
- Francis, G., Z. Kerem, H. P. Makkar & K. Becker (2002) The biological action of saponins in animal systems: a review. *Br J Nutr*, 88, 587-605.
- Fujisawa, T., K. Aikawa, T. Takahashi, S. Yamai, K. Watanabe, Y. Kubota & M. Miyaoka (2001) Influence of butyric and lactic acids on the beta-glucuronidase activity of *Clostridium perfringens*. *Lett Appl Microbiol*, 32, 123-5.
- Görög, S. (2011) Advances in the analysis of steroid hormone drugs in pharmaceuticals and environmental samples (2004–2010). *Journal of Pharmaceutical and Biomedical Analysis*, 55, 728-743.
- Hancock, S. M., M. D. Vaughan & S. G. Withers (2006) Engineering of glycosidases and glycosyltransferases. *Curr Opin Chem Biol*, 10, 509-19.
- Hazenberg, M. P., W. W. de Herder & T. J. Visser (1988) Hydrolysis of iodothyronine conjugates by intestinal bacteria. *FEMS Microbiol Rev*, 4, 9-16.
- Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J*, 280 (Pt 2), 309-16.
- Hodal, L., A. Bocharde, J. E. Nielsen, O. Mattsson & F. T. Okkels (1992) Detection, expression and specific elimination of endogenous β -glucuronidase activity in transgenic and non-transgenic plants. *Plant Science*, 87, 115-122.
- Hwang, B. Y., S. E. Kim, Y. H. Kim, H. S. Kim, Y.-S. Hong, J. S. Ro, K. S. Lee & J. J. Lee (1999) Pregnane Glycoside Multidrug-Resistance Modulators from *Cynanchum wilfordii*. *Journal of Natural Products*, 62, 640-643.
- Islam, M. R., S. Tomatsu, G. N. Shah, J. H. Grubb, S. Jain & W. S. Sly (1999) Active site residues of human beta-glucuronidase. Evidence for Glu(540) as the nucleophile and Glu(451) as the acid-base residue. *J Biol Chem*, 274, 23451-5.
- Jefferson, R. A., S. M. Burgess & D. Hirsh (1986) beta-Glucuronidase from *Escherichia coli* as a gene-fusion marker. *Proceedings of the National Academy of Sciences*, 83, 8447-8451.
- Johansson, E., L. Hedby, P.-O. Larsson, K. Mosbach, A. Gunnarsson & S. Svensson (1986) Synthesis of mannose oligosaccharides via reversal of the α -mannosidase reaction. *Biotechnology Letters*, 8, 421-424.
- Kashibuchi, N., K. Matsubara, Y. Kitada & H. Suzuki. 1996. Scalp moisturizer and external skin preparation. Google Patents.
- Kim, Y. W., S. S. Lee, R. A. Warren & S. G. Withers (2004) Directed evolution of a glycosynthase from *Agrobacterium* sp. increases its catalytic activity dramatically and expands its substrate repertoire. *J Biol Chem*, 279, 42787-93.
- Levy, G. A. (1954) Baicalinase, a plant β -glucuronidase. *Biochemical Journal*, 58, 462-469.
- Liang, W. J., K. J. Wilson, H. Xie, J. Knol, S. Suzuki, N. G. Rutherford, P. J. Henderson & R. A. Jefferson (2005) The gusBC genes of *Escherichia coli* encode a glucuronide transport system. *J Bacteriol*, 187, 2377-85.

- Lin, H., H. Tao & V. W. Cornish (2004) Directed Evolution of a Glycosynthase via Chemical Complementation. *Journal of the American Chemical Society*, 126, 15051-15059.
- Ma, P., N. Kanizaj, S. A. Chan, D. L. Ollis & M. D. McLeod (2014) The Escherichia coli glucuronylsynthase promoted synthesis of steroid glucuronides: improved practicality and broader scope. *Org Biomol Chem*, 12, 6208-14.
- Ma, X., B. Yu, Y. Hui, Z. Miao & J. Ding (2001) Synthesis of steroidal glycosides bearing the disaccharide moiety of OSW-1 and their antitumor activities. *Carbohydrate Research*, 334, 159-164.
- Mackenzie, L. F., Q. Wang, R. A. J. Warren & S. G. Withers (1998) Glycosynthases: Mutant Glycosidases for Oligosaccharide Synthesis. *Journal of the American Chemical Society*, 120, 5583-5584.
- Mayer, C., D. L. Jakeman, M. Mah, G. Karjala, L. Gal, R. A. Warren & S. G. Withers (2001) Directed evolution of new glycosynthases from Agrobacterium beta-glucosidase: a general screen to detect enzymes for oligosaccharide synthesis. *Chem Biol*, 8, 437-43.
- McCarter, J. D. & S. G. Withers (1994) Mechanisms of enzymatic glycoside hydrolysis. *Curr Opin Struct Biol*, 4, 885-92.
- Means, G. E. & R. E. Feeney (1990) Chemical modifications of proteins: history and applications. *Bioconjug Chem*, 1, 2-12.
- Ness, J. E., S. B. Del Cardayre, J. Minshull & W. P. Stemmer (2000) Molecular breeding: the natural approach to protein design. *Adv Protein Chem*, 55, 261-92.
- Nishimura, Y., M. G. Rosenfeld, G. Kreibich, U. Gubler, D. D. Sabatini, M. Adesnik & R. Andy (1986) Nucleotide sequence of rat preputial gland beta-glucuronidase cDNA and in vitro insertion of its encoded polypeptide into microsomal membranes. *Proc Natl Acad Sci U S A*, 83, 7292-6.
- Ooi, Y., T. Hashimoto, N. Mitsuo & T. Satoh (1984) Enzymic synthesis of chemically unstable cardiac glycosides by β -galactosidase from *Aspergillus oryzae*. *Tetrahedron Letters*, 25, 2241-2244.
- Perugino, G., A. Trincone, A. Giordano, J. van der Oost, T. Kaper, M. Rossi & M. Moracci (2003) Activity of Hyperthermophilic Glycosynthases Is Significantly Enhanced at Acidic pH. *Biochemistry*, 42, 8484-8493.
- Perugino, G., A. Trincone, M. Rossi & M. Moracci (2004) Oligosaccharide synthesis by glycosynthases. *Trends Biotechnol*, 22, 31-7.
- Pikkemaat, M. G., A. B. M. Linssen, H. J. C. Berendsen & D. B. Janssen (2002) Molecular dynamics simulations as a tool for improving protein stability. *Protein Engineering*, 15, 185-192.
- Ray, J., A. Bouvet, C. DeSanto, J. C. Fyfe, D. Xu, J. H. Wolfe, G. D. Aguirre, D. F. Patterson, M. E. Haskins & P. S. Henthorn (1998) Cloning of the canine beta-glucuronidase cDNA, mutation identification in canine MPS VII, and retroviral vector-mediated correction of MPS VII cells. *Genomics*, 48, 248-53.
- Reetz, M. T. 2006. Directed Evolution of Enantioselective Enzymes as Catalysts for Organic Synthesis. In *Advances in Catalysis*, eds. C. G. Bruce & K. Helmut, 1-69. Academic Press.

- Schulz, M. & G. Weissenböck (1987) Partial purification and characterisation of a luteolin-triglucuronide-specific β -glucuronidase from rye primary leaves (*Secale cereale*). *Phytochemistry*, 26, 933-937.
- Shaikh, F. A. & S. G. Withers (2008) Teaching old enzymes new tricks: engineering and evolution of glycosidases and glycosyl transferases for improved glycoside synthesis. *Biochem Cell Biol*, 86, 169-77.
- Sinnott, M. L. (1990) Catalytic mechanism of enzymic glycosyl transfer. *Chemical Reviews*, 90, 1171-1202.
- Steiner, K. & H. Schwab (2012) Recent advances in rational approaches for enzyme engineering. *Comput Struct Biotechnol J*, 2, e201209010.
- Stemmer, W. P. (1994a) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci U S A*, 91, 10747-51.
- Stemmer, W. P. C. (1994b) Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, 370, 389-391.
- Tang, Y., N. Li, J.-a. Duan & W. Tao (2013) Structure, Bioactivity, and Chemical Synthesis of OSW-1 and Other Steroidal Glycosides in the Genus *Ornithogalum*. *Chemical Reviews*, 113, 5480-5514.
- Tao, H., P. Peralta-Yahya, H. Lin & V. W. Cornish (2006) Optimized design and synthesis of chemical dimerizer substrates for detection of glycosynthase activity via chemical complementation. *Bioorg Med Chem*, 14, 6940-53.
- Tee, K. L. & U. Schwaneberg (2007) Directed evolution of oxygenases: screening systems, success stories and challenges. *Comb Chem High Throughput Screen*, 10, 197-217.
- Toscano, M. D., K. J. Woycechowsky & D. Hilvert (2007) Minimalist active-site redesign: teaching old enzymes new tricks. *Angew Chem Int Ed Engl*, 46, 3212-36.
- Trincone, A., E. Pagnotta, A. Giordano, G. Perugino, M. Rossi & M. Moracci (2003) Enzymatic Synthesis of 2-Deoxyglycosides Using the β -Glycosidase of the Archaeon *Sulfolobus solfataricus*. *Biocatalysis and Biotransformation*, 21, 17-24.
- Trincone, A., G. Perugino, M. Rossi & M. Moracci (2000) A novel thermophilic Glycosynthase that effects branching glycosylation. *Bioorganic & Medicinal Chemistry Letters*, 10, 365-368.
- Watson, M., J. W. Liu & D. Ollis (2007) Directed evolution of trimethoprim resistance in *Escherichia coli*. *FEBS J*, 274, 2661-71.
- Werschkun, B., A. Wendt & J. Thiem (1998) Enzymic synthesis of β -glucuronides of estradiol, ethynylestradiol and other phenolic substrates on a preparative scale employing UDP-glucuronyl transferase. *Journal of the Chemical Society - Perkin Transactions 1*, 3021-3023.
- Wilkinson, S. M., C. W. Liew, J. P. Mackay, H. M. Salleh, S. G. Withers & M. D. McLeod (2008) *Escherichia coli* glucuronylsynthase: an engineered enzyme for the synthesis of beta-glucuronides. *Org Lett*, 10, 1585-8.
- Williams, G. J., C. Zhang & J. S. Thorson (2007) Expanding the promiscuity of a natural-product glycosyltransferase by directed evolution. *Nat Chem Biol*, 3, 657-62.

- Wilson, K. J., S. G. Hughes & R. A. Jefferson. 1992. The Escherichia coli gus Operon: Induction and Expression of the gus Operon in E. coli and the Occurrence and Use of GUS in Other Bacteria. In *Gus Protocols: Using the GUS Gene as a Reporter of Gene Expression*, ed. S. R. Gallagher, 7-22. San Diego: Academic Press.
- Wilson, K. J., A. Sessitsch, J. C. Corbo, K. E. Giller, A. D. Akkermans & R. A. Jefferson (1995) beta-Glucuronidase (GUS) transposons for ecological and genetic studies of rhizobia and other gram-negative bacteria. *Microbiology*, 141 (Pt 7), 1691-705.
- Wong, T. S., D. Roccatano, M. Zacharias & U. Schwaneberg (2006) A statistical analysis of random mutagenesis methods used for directed protein evolution. *J Mol Biol*, 355, 858-71.
- Yokosuka, A. (2015) [Discovery of Novel Biologically Active Compounds of Natural Origin, with a Focus on Anti-tumor Activity]. *Yakugaku Zasshi*, 135, 1109-14.
- Zechel, D. L. & S. G. Withers (2000) Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc Chem Res*, 33, 11-8.
- Zhao, H., J. C. Moore, A. A. Volkov & F. H. Arnold. 1999. *Methods for optimizing industrial enzymes by directed evolution*. Washington DC: ASM Press.
- Zhao, H. & W. Zha (2006) In vitro 'sexual' evolution through the PCR-based staggered extension process (StEP). *Nat Protoc*, 1, 1865-71.

Chapter 2

Experimental Methods

This chapter is intended to provide details on the routine experimental procedures used throughout the course of this research. Information on specific protocols developed during this study is given in the relevant chapters.

There are a few general points covering basic chemicals: All chemicals were purchased from Sigma-Aldrich unless otherwise specified and water was purified using a Milli-Q purification system (Millipore). The composition of solutions, buffers and growth media are given in Appendix B.

2.1. Cell strains and expression vectors

Throughout the course of this work, enzymes were synthesised using the cellular machinery of bacteria. The requirements for cell based protein expression are a vector incorporating the DNA sequence for the protein of interest and a host cell (*Escherichia coli*). The bacterial strains and plasmids are listed below.

2.1.1. *E. coli* strain BW25141

BW25141 is a strain of *E. coli* that does not contain the β -galactosidase and β -glucuronidase genes and was used to generate site-saturation libraries (Wanner 1983). It is usually designated BW25141 (*lacI*^q *rrnB*_{T14} Δ *lacZ*_{WJ16} Δ *phoBR580* *hsdR514* Δ *araBAD*_{AH33} Δ *rhaBAD*_{LD78} *galU95* *endA*_{BT333} *uidA*(Δ *MluI*)::*pir*⁺ *recA1*).

2.1.2. *E. coli* strain GMS407

The *E. coli* strain of choice for directed evolution was the GMS407 (*lacY1* *tsx-29* *glnX44* *galk2* λ *manA4* *uidA1* *mtl-1* *argE3*). This β -glucuronidase-deficient *E. coli* lysogenic cell line was used to ensure protein expression that was free of wild-type activity (Novel & Novel 1973).

2.1.3. *E. coli* strain DH5 α

The *E. coli* strain DH5 α was used for storage of plasmids. It is a commonly used laboratory K12 strain. It has a high transformation efficiency, simple preparation and is easily stored. Detailed genotype of DH5 α strain – *supE44* Δ *lac169* ϕ 80 *lacZ* Δ M15 *hsdR17* *recA1* *endA1* *gyrA96* *thi-1* *relA1*. It lacks *recA*, a protein responsible for recombination between

homologous DNA sequences, so there was minimum chance of recombination between homologous sequences of plasmid gene and bacteria genome. The *endA1* implies the deletion of the normally present non-specific endonucleases that could possibly digests the foreign DNA, thus avoids difficulties when preparing plasmid. The designation *hsdR17* indicates that a native restriction enzyme that is normally present in *E. coli* has been eliminated, preventing degradation of target plasmid transformed into the cell.

2.1.4. Expression vector pJWL1030

Vector pJWL1030 (*par*⁺, *bla*⁺, *lacZpro*, *T7*Ø10^{tir+}) was derived from a high-copy number vector pCY76 (*par*⁺ *bla*⁺ *lacZpro* *T7*Ø10^{tir+}) and pJJKmf (Liu *et al.* 2006). The constitutive expression cassette was isolated from pCY76 and ligated to the backbone of pJJKmf (Kirschman & Cramer 1988). A leaky *lac* promoter allows constitutive expression of target gene cloned into pJWL1030. pJWL1030 confers kanamycin resistance to the host.

2.1.5. Expression vector pET-28a(+)

The *gus-wt* and *gus-tr3337* genes encoding β-glucuronidase (β-GUS) and thermostable β-GUS, respectively, were cloned into a Novagen® pET-28a(+) vector at the *NdeI/XhoI* cloning sites (Wilkinson *et al.* 2008). The pET-28a(+) vector also carries the kanamycin resistance gene, *T7lac* promoter, and an N-terminal His₆-tag extension.

2.1.6. Expression vector pJexpress 401

The pJexpress 401 expression vector with N-terminal His₆-tag was obtained from DNA 2.0 (Menlo Park, CA, USA). The pJExpress 401 is an inducible protein expression system under the control of the Isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible T5 promoter. It also contains a kanamycin resistance marker. The *gus-wt* and *gus-tr3337* genes were cloned into the multi-cloning site of pJExpress 401 using *XbaI* and *EcoRI* unique restriction sites.

2.2. Computer software programs

Protein sequences were analysed with the program CONSURF (<http://consurf.tau.ac.il/>), that gave the conserved and non-conserved region of a protein (Ashkenazy *et al.* 2010). Structural diagrams presented in this thesis were generated using the PYMOL program (Delano 2002). Molecular docking of *para*-nitrophenyl-β-D-glucuronide (*pNP*-glucuronide) and glucuronic acid was performed using the pair-fit command in Pymol 1.3. The Kolmogorov-Smirnov test, (<http://www.physics.csbsju.edu/>) was used to analyse the potential difference in two libraries. The null hypothesis was H₀: both samples come from a population with the same distribution.

The H_0 was rejected at the 0.05 significance level. Protein extinction coefficient was estimated by PROTPARAM tool (<http://web.expasy.org/protparam/>). It allowed the computation of various physical and chemical properties of a protein given its sequence (Gasteiger *et al.* 2005). Analysis of scanned photos was possible with IMAGEJ, a program available from the National Institutes of Health (USA) website (<http://rsb.info.nih.gov/ij/>). SWISS-MODEL (<http://www.swissmodel.expasy.org/>) (Schwede *et al.* 2003) was used to generate model structures of GUS-WT and GUS-TR3337 mutants by using GUS-WT as template. The solvent accessible surface area (SASA) for each residue in β -GUS was calculated with ASA (<http://cib.cf.ocha.ac.jp/bitool/ASA/>), using a 1.4 Å radius probe for water. The SASA calculations were based on the method of Shrake & Rupley (1973).

2.3. Molecular biology

2.3.1. Directed evolution

There were two primary steps in directed evolution – library generation followed by screening and/ or selection. The polymerase chain reaction (PCR) was used in the first stage to introduce genetic diversity. The methods used have been published and will not be described in detail here. The introduction of random errors in a gene (error-prone PCR) was described by Cadwell & Joyce (1992) while the staggered extension process (StEP) was described by members of the Arnold laboratory (Zhao *et al.* 1998). The combination ePCR and StEP, that is staggered extension process-error prone polymerase chain reaction was described by Stevenson (2006) who also detailed protocols for targeted [site-saturation mutagenesis or site-directed mutagenesis (SDM)]. The second stage is screening and/ or selection for the desired traits. Much of this methodology is standard molecular biology and is described briefly below. Detailed information is provided in the relevant chapters that follow.

2.3.1.1. Introducing genetic diversity

The *gus-wt* gene (GUS-WT) was manipulated using the construct created by cloning it into the pET28a(+) vector using the restriction sites *NdeI* and *XbaI*. Five methods were used to create variants of *gus-wt* and *gus-tr3337* (designated *gus-wt** and *gus-tr3337**, respectively) and are outlined below.

2.3.1.1.1. Site-saturation mutagenesis (SSM)

β -GUS residues E413, N412, D163, R562, Y472, N566, T509, Y469, S557 and M447 were subjected to randomisation with SSM (Zheng *et al.* 2004). Degenerate primers containing the complete combination of 32 codons at the mutation points are designed to generate the SSM

libraries. Sets of offset NNK mutagenic primers (Table 2.1) were used to PCR amplify the entire gene and plasmid sequence. The product was digested with *DpnI* then purified and transformed into DH5 α *E. coli*.

Table 2.1 Mutagenic NNK primers for SSM mutagenesis

Primer	Sequence (5' – 3')
GUS E413 NNK F	tat tgc caa cnn kcc gga tac ccg tc
GUS N412 NNK F	gag tat tgc cnn k ga acc gga tac cc
GUS D163 NNK F	agt taa aga ann kat gga agt aag ac
GUS R562 NNK F	aag gca tat tgn nkg ttg gcg gta ac
GUS Y472 NNK F	tta cgg atg gnn kgt cca aag cgg cg
GUS N566 NNK F	ctc gca agg cnn ktt gcg cgt tg
GUS T509 NNK F	gcg tgg atn nkt tag ccg ggc tgc a
GUS Y469 NNK F	gaa ccg tta tnn kgg atg gta tgt cca aag c
GUS S557 NNK F	ttt tgc gac cnn kca agg cat att gcg c
GUS M447 NNK F	gcg tca atg tan nkt tct gcg acg ct
pJWL1030GUSHindIII R	aag ctt ttg cca ttc tca ccg gat tc

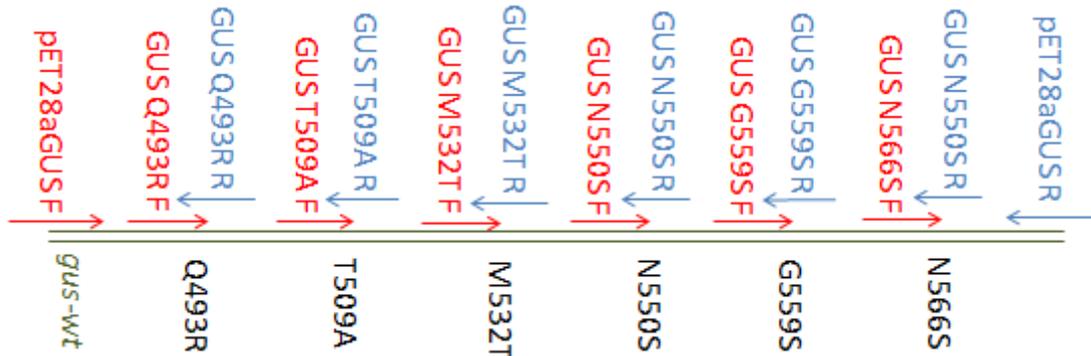
SSM-PCR reactions were set up with 10 ng template pJWL1030-*gus*, 0.2 mM dNTP, 2.5U *Pfu* DNA polymerase, 1 μ M forward and reverse primers, and 5 μ L 10x *Pfu* reaction buffer and made up to 50 μ L with deionised water. The SSM-PCR reaction mixtures were subjected to single 2-stage whole-plasmid PCR: 1 cycle at 95 °C for 3 min, followed by 5 cycles of 95 °C for 30 s, 55 °C for 1 min, 72 °C for 1 min/kb of mega-primer length, and 20 cycles of 95 °C for 30 s, 68 °C for 1 min/kb of template length, lastly, 1 cycle at 68 °C for 10 min.

2.3.1.1.2. Site-directed mutagenesis (SDM)

gus-tr3337, *wt3p24e11-E504A*, *wt3p24e11-E504G*, *wt3p24e11-E504S*, *wt5p26a7-E504A*, *wt5p26a7-E504G*, *wt5p26a7-E504S*, *thermo4p11f2-E504A*, *thermo4p11f2-E504G* and *thermo4p11f2-E504G* were constructed using SDM that was performed by the overlap extension PCR method with *gus-wt* as the DNA template and the primers listed in Table 2.2 (Xiong *et al.* 2004). The mutation site was introduced into the first PCR fragments, and the second fragment was amplified by annealing the overlapping ends of the first fragments. Final products were ligated into the pET28a(+) vector and sequenced. The strategy for the SDM was shown in Figure 2.1.

Table 2.2 Primers used for point mutations in site-directed mutagenesis

Primer	Sequence (5' – 3')
GUS Q493R F	tgg cct ggc ggg aga aac tgc atc ag
GUS Q493R R	atg cag ttt ctc ccg cca ggc cag aag ttc
GUS T509A F	ccc ggc taa cgc atc cac gcc gta tt
GUS T509A R	ggc gtg gat gcg tta gcc ggg ctg
GUS M532T F	tgg ctg gat acg tat cac cgc gtc ttt g
GUS M532T R	gac gcg gtg ata cgt atc cag cca tgc ac
GUS N550S F	cag gta tgg agt ttc gcc gat ttt gc
GUS N550S R	aaa atc ggc gaa act cca tac ctg ttc ac
GUS G559S F	gac ctc gca aag cat att gcg cgt tgg
GUS G559S R	aac gcg caa tat gct ttg cga ggt cgc aa
GUS N566S F	gtt ggc ggt agc aag aaa ggg atc ttc
GUS N566S R	gat ccc ttt ctt gct acc gcc aac gcg
pET28aGUS F	aag aag gag ata tac cat ggg cag c
pET28aGUS R	tgg tgg tgc tcg agt cat tgt ttg cc

**Figure 2.1 SDM from *gus-wt* and *gus-tr3337* primers employed in this study**

Red: forward; blue: reverse.

2.3.1.1.3. Error-prone polymerase chain reaction (ePCR)

ePCR was based on the protocol developed by Leung *et al.* (1989), with reactions consisting of: 16 ng of the parent gene, 10 μ M of forward and reverse primers, 0.2 mM of dNTPs, 5 mM $MgCl_2$, different concentrations of Mn^{2+} ranging from 0.05 - 0,5 mM, 5 μ L of 10x polymerase buffer without Mg^{2+} , 2.5 U *Taq* DNA polymerase and made up to 50 μ L with deionised water.

Thermocycling was performed with 30 cycles of 94 °C for 10 s, 50 °C for 10 s and 2 min at 72 °C. The ePCR products were purified with PCR purification kit and digested in one reaction with *NdeI*, *XbaI* and *DpnI*. The latter was required to selectively digest the parent genes since they were synthesised *in vivo* and have methylated *DpnI* restriction sites (Sambrook & Russell 2001). Finally, DNA with the correct size was isolated by gel purification with Promega® PCR purification kit and ligated into *pET28a(+)* with T4 ligase. The constructs

used in verifying the error-rate by DNA sequencing were purified from randomly selected GMS407 transformants.

2.3.1.1.4. Staggered extension process (StEP)

The StEP reaction consisted of 16 ng of template, 10 μ M of each forward and reverse primer, 0.2 mM dNTP, 5 mM MgCl₂ and 1 U of Phusion high fidelity polymerase. The cycling conditions were 100 cycles of 94 °C for 10 s, 50 °C for 10 s and 72 °C for 2 s.

2.3.1.1.5. StEP-ePCR

Recombination of gene variants was simplified by the use of the StEP-ePCR. StEP-ePCR was set up in 50 μ L with: 16 ng of parent genes in equal proportions, 10 μ M of forward and reverse primers, 0.2 mM of dNTPs, 5 mM MgCl₂, 1.5 mM MnCl₂, 5 μ L of 10x polymerase buffer without Mg²⁺ and 2.5 U of *Taq* polymerase. StEP-ePCR thermocycling required 30 cycles of: 94 °C for 10 s, 50 °C for 10 s and 72 °C for 10 s.

2.3.1.2. Preparing libraries of variants for screening

A population of *gus-wt*, prepared by SSM, was ligated into pJWL1030 via *Nde*I and *Xba*I restriction sites. The variants of *gus-wt* were created and resulting in pJWL1030-*guswt** constructs. In contrast, the random mutagenesis libraries of *gus-wt* and *gus-tr3337* were digested at *Xba*I and *Xho*I restriction sites and ligated into the pET28a(+) expression vector, resulting in pET28a-*guswt** and pET28a-*gustr3337**, respectively. The ligation reaction products were isolated with a Promega® PCR purification kit and eluted in 30 μ L of water. 5 μ L of the purified ligation products was transformed into 50 μ L of *E. coli* strain GMS407 and recovered with 1 mL yeast extract nutrient broth (YENB) (Appendix B) at 37 °C. 50 μ L of the culture was plated out on Luria-Bertani-Kanamycin (LBK) agar (Appendix B) and incubated overnight at 37 °C to estimate the cell density. After determining the cell density, the transformants were diluted such that 50 μ L of culture that contained about 300 viable cells. Library culture aliquots of 50 μ L were applied to each LB-Kanamycin agar and spread out with sterile ball bearings (5 mm diameter, steel) to provide an even distribution of colonies. 10 colonies were sequenced to examine the randomness of the point mutations; that is the average number of mutations.

2.3.1.2.1. In vitro screening of site-directed libraries with β -glycoside substrates

In each SSM library, 96 single colonies were picked manually from the LBK agar plates and grown individually in 96-well round bottom culture plates with 200 μ L of LBK per well. The plates were incubated at 37 °C overnight. 120 μ L of culture from each well was transferred to a new sterile 96-well microplates and 20 μ L of 7x BugBuster was added to each well to lyse the

cells. The proteins were assayed with 200 μ M *para*-nitrophenyl- β -D-glucuronide (*p*NP-glucuronide), 3 mM *p*NP-glucopyranoside, 3 mM *p*NP-galactopyranoside, 3 mM *p*NP-mannopyranoside and 3 mM *p*NP-xylopyranoside in 50 mM phosphate buffer, pH 7.4. The release of *p*NP was monitored at 405 nm (A_{405}) on a Labsystems Multiskan Ascent 96-well microplate ultraviolet/visible spectrometer. The activities were compared to those of wild-type. Clones with enhanced, moderate and decreased activities toward five substrates were selected. Plasmids were isolated with miniprep kit and sequenced.

2.3.1.2.2. Colony screening

E. coli expressing β -GUS can moderately cleave externally supplied 5-bromo-4-chloro-3-indolyl- β -D-glucoside (X-glucoside) to produce glucose and an intense blue precipitate of chloro-bromoindigo. Another artificial chromogenic substrate, *p*NP-glucoside, can also be moderately hydrolysed by β -GUS resulting in glucose and the yellow stained *p*NP.

This formed the primary screening and secondary phases. The primary screening can be done on solid medium that contains X-glucoside whereas the liquid medium base for secondary screening contained *p*NP-glucoside. The primary screening narrows hundreds of thousands of β -GUS versions down to less than three thousand.

2.3.1.2.2.1. Primary screening of colonies

The first two generation examined colonies via agar plate screening while generations 3, 4 and 5 examined colonies via top agar screening. An overview of this scheme is presented in the Figure 2.2.

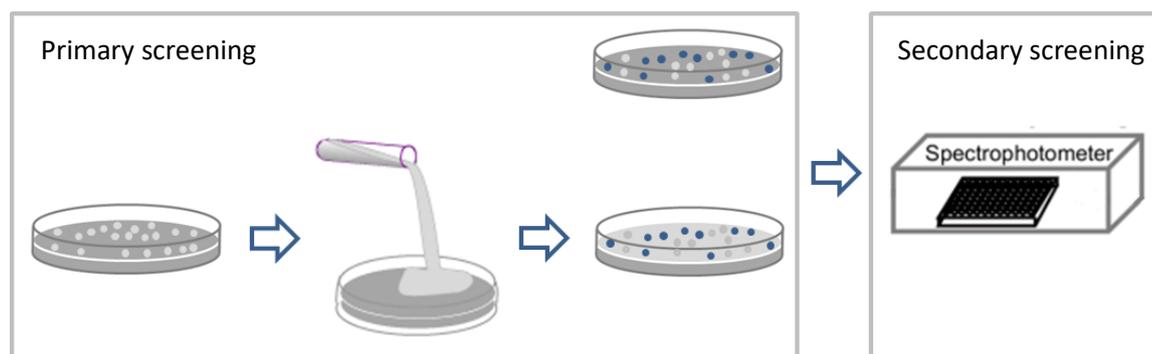


Figure 2.2 Screening strategies

The two screening strategies used in directed evolution of GUS-WT and GUS-TR3337 are presented in a schematic form with agar plates, top agars plates and 96-well plates are illustrated at the top, bottom and right, respectively. Agar plate screening is described in Section 2.3.1.2.2.1.1., top agar screening is described in Section 2.3.1.2.2.1.2. and culture screening is described in Section 2.3.1.2.2.2.

2.3.1.2.2.1.1. Agar plate screening

50 μ L of library culture aliquots were applied to each agar plate. The agar plates were prepared with 10 mL of LBK agar supplemented with 0.15 mM IPTG and 0.1 mM X-glucoside. After overnight incubation at 37 °C, thirty most blue colonies on each plate were picked and inoculated in individual 200 μ L aliquots of sterile LBK into a 96-well culture plate.

Each screening batch, typically a 100 agar plates, resulted in about 3000 inoculants with additional inoculants as controls. The negative control was *E. coli* GMS407 transformed with empty pET28a(+) and the positive control was the same strain expressing GUS-WT, GUS-TR3337 or the best variant from each generation.

2.3.1.2.2.1.2. Top agar screening

15 mL of molten glucosidase activity indicator agar (Appendix A) was carefully poured onto each agar plate, covering all the colonies. Colonies expressing active β -GUS would start to turn darker blue after 10 min with the top agar. The two most blue and distinct colonies on each plate were identified within 30 min. Each of the identified colonies was used to inoculate individual 200 μ L aliquots of sterile LBK (Appendix B).

2.3.1.2.2.2. Secondary screening of β -GUS activity

After the 96-well culture plates were incubated overnight, 120 μ L of each culture was transferred to a 96-well plate. The cells of each sample were lysed by adding 20 μ L of 7x BugBuster and incubating for 30 min at room temperature. Addition of 100 μ L of assay buffer produced a 220 μ L reaction with: 50 mM phosphate buffer at pH 7.4, 800 μ M of pNP-glucoside. The reactions were performed at room temperature (25 °C) and the hydrolysis of pNP-glucoside was measured by monitoring the rate of the release of pNP at A₄₀₅ over 20 min.

2.3.2. Protein expression

*pET28a-guswt**, *pET28a-gustr3337**, *pJexpress401-guswt** and *pJexpress401-gustr3337** plasmid DNA (pDNA) samples were purified with the Qiagen Miniprep® kit as follows. 1 μ L of the pDNA was used to transform BW25141 or GMS407 cells, that were recovered with 1 mL YENB medium after electroporation. 50 μ L of the recovery culture was streaked onto an LBK agar plate that was then incubated at 37 °C overnight. The following day, a single colony was inoculated into 20 mL 2x YT-Kanamycin broth and incubated at 37 °C overnight. The culture was then transferred to 1 L of same medium and grown at 37 °C overnight. Cells were induced with a final concentration of 0.5 mM IPTG when. The culture was further incubated aerobically by orbital shaking for 3 hours until an OD₆₀₀ of 0.8 was reached. Cells were

harvested by centrifugation at 3700 g in a Sorvall SLA 3000 for 20 minutes. The supernatant was removed and the cells were used immediately for lysis or stored at -80 °C for later use.

2.3.3. Protein purification

Prior to purification, the theoretical values of molecular weight (monomer), isoelectric point and extinction coefficient of every β -GUS protein were calculated with PROTPARAM.

All purification steps were carried out at 4 °C to minimise loss due to denaturation. Two protocols were used to purify GUS* in this study. All buffers and milliQ water (mQH₂O) used in purification were filtered with vacuum driven 0.45 μ m nitrocellulose membrane filters. AKTATM FPLC system was used to automatically control the elution in chromatography steps.

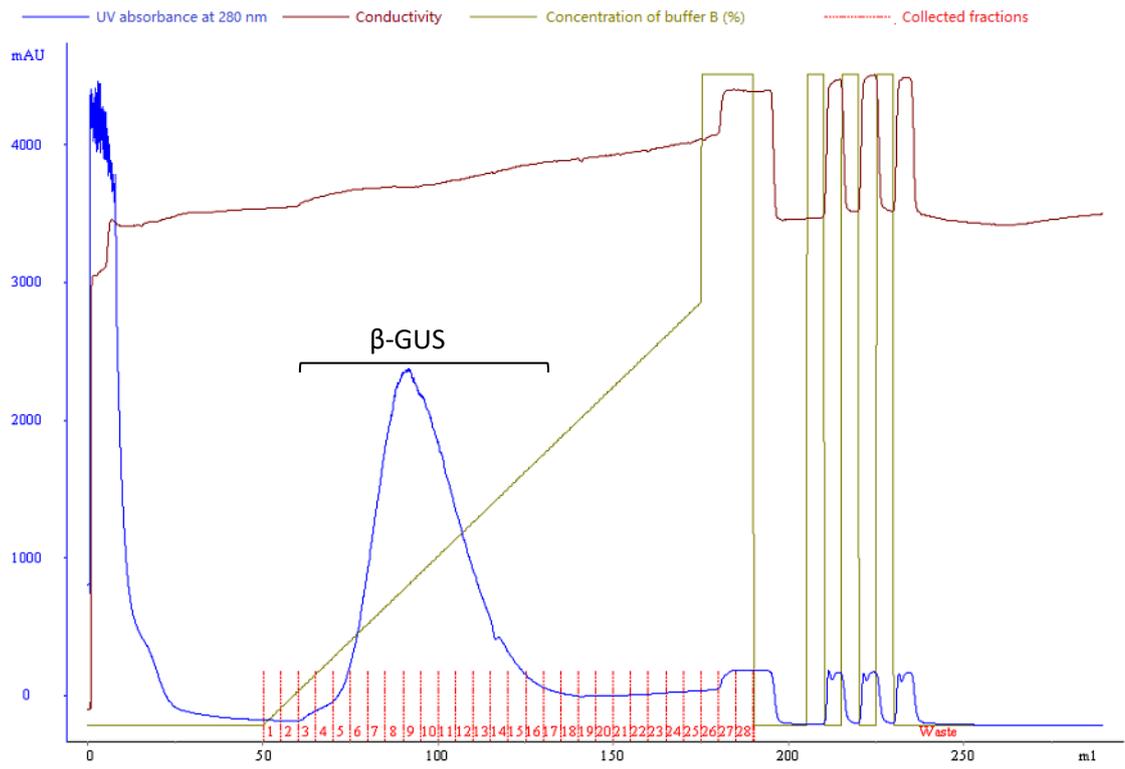
2.3.3.1. Immobilised metal affinity chromatography (IMAC)

The harvested cells were resuspended in 25 mL of buffer A (Appendix B). The cells were lysed with a French Press and the lysate was centrifuged at 3700 g for 40 min at 4 °C to separate the cell debris. The supernatant was loaded onto 5 mL HisTrap FF (GE Healthcare) column that eluted with 0-0.5 M imidazole gradient at a flow rate of 3 mL/min. Fractions that displayed the highest activity were analysed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Fractions with the highest amounts of β -GUS protein were concentrated to 0.5-1 mL for the next purification step.

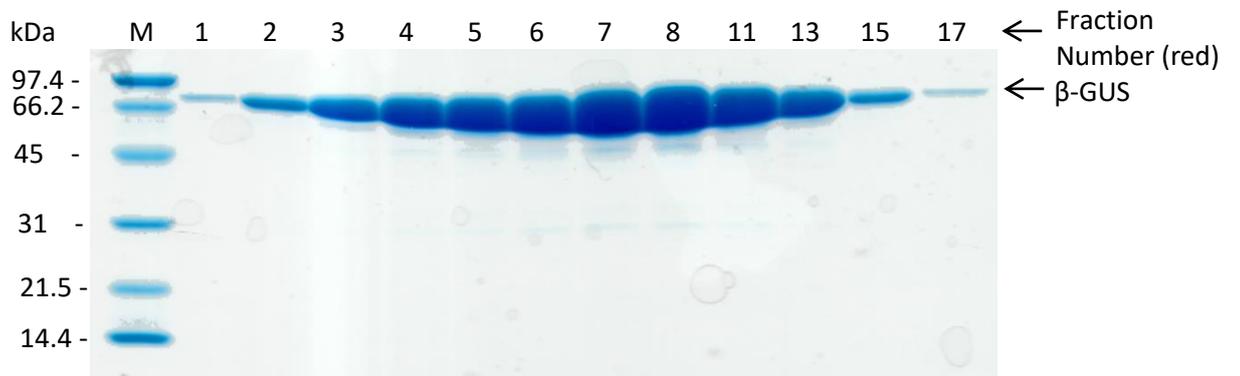
2.3.3.2. Size-exclusion chromatography (SEC)

SEC or gel-filtration chromatography was used to separate proteins according to their molecular weight. The concentrated protein solution was then loaded onto a Superdex 200 gel filtration column. The sizing column had a column volume of 120 mL and had been calibrated with protein standards. The protein was eluted at 0.3 mL/min to ensure good resolution. Fractions with the highest activity were confirmed via SDS-PAGE and pooled. Examples of the elution profiles and SDS-PAGE of column fractions are given below for the IMAC and SEC (Figure 2.3). Fractions judged to be at least 95% pure were dialysed against buffer 50 mM phosphate, pH 7.4 overnight. Purified β -GUS proteins were stored at 4 °C. A description of the procedure used to calibrate the SEC columns and the determination of protein molecular weights with SEC are given below.

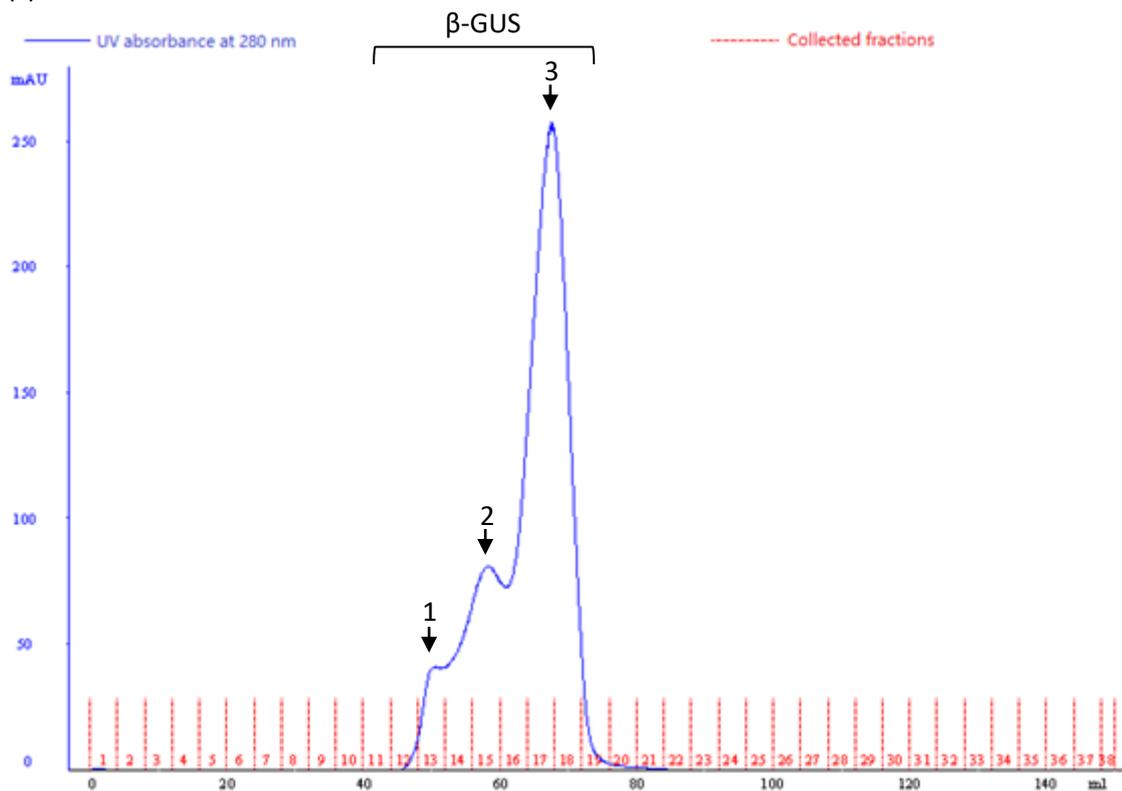
(a)



(b)



(c)



(d)

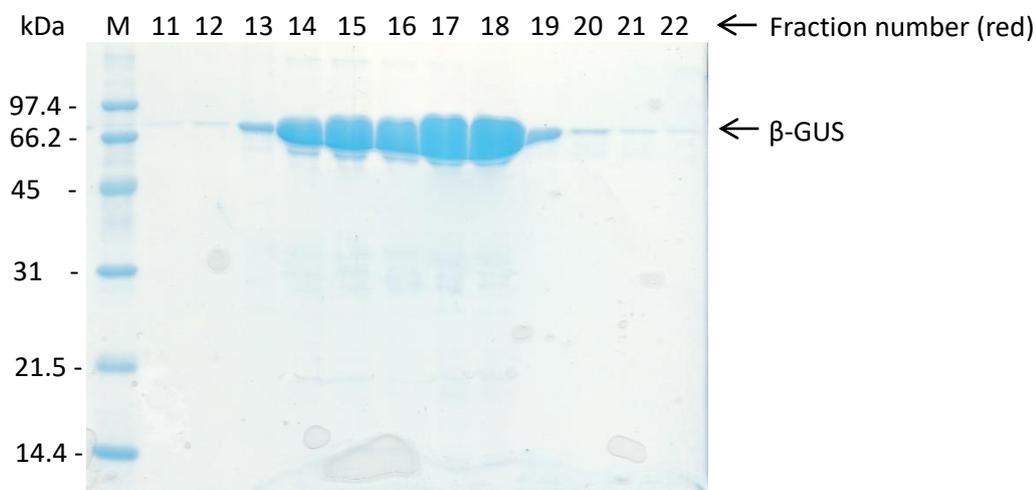


Figure 2.3 Purification of GUS-WT

(a) Elution profile using IMAC (b) SDS-PAGE gel of selected fractions from IMAC. Fraction numbers are indicated above each lane (c) Elution profile using SEC. Peak 1, tetramer (279 kDa); peak 2, dimer (139 kDa) and peak 3, monomer (69 kDa). (d) SDS-PAGE gel of selected fractions from SEC. Fraction numbers are indicated above each lane.

The size exclusion column used in this study was pre-calibrated with protein standards whose molecular weights ranged between 43 – 669 kilo-Daltons (kDa) (Figure 2.4). A linear graph of logarithm of protein sizes in Da versus their elution volumes was plotted and its

function is derived from linear regression (Figure 2.5). β -GUS protein sizes were subsequently calculated based on the equation.

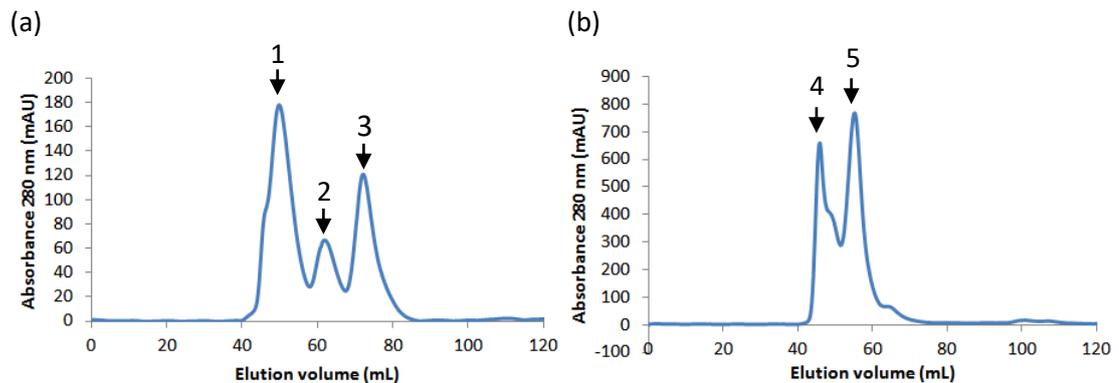


Figure 2.4 Elution profiles of molecular weight markers from Superdex 200 size exclusion column

(a) Peak 1: Ferritin, 440 kDa; 2: Conalbumin, 75 kDa; 3: Ovalbumin, 43 kDa (b) Peak 4: Thyroglobulin, 669 kDa; 5: Aldolase, 150 kDa.

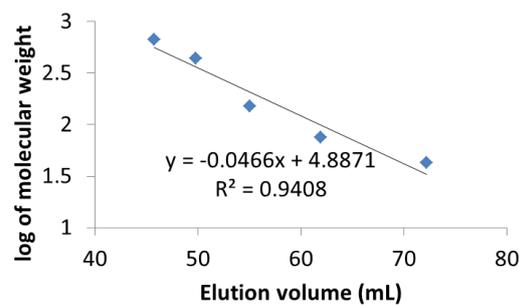


Figure 2.5 Calibration curve obtained with molecular weight markers eluted from Superdex 200 size exclusion column

The linear equation derived from the curve fit is given.

2.3.4. Analysis of GUS-WT and GUS-TR3337 variants

The GUS-WT and GUS-TR3337 variants were characterised in several ways. The pure GUS-WT and GUS-TR3337 variants were subject to kinetic investigations with respect to glucuronide, glucoside concentrations and temperature. Finally, an attempt was made to rationalize these results in terms of the β -GUS structure.

Each of the variants selected as progeny throughout the directed evolution of GUS-WT and GUS-TR3337 were named according to the generation number (1-5) and an identification number. The first digit refers to the batch number (1-5), the second digit refers to the 96-well plate number and is followed by the well position. For example, the variant WT1P17G2, from the *gus-wt* library, is derived from generation 1 and isolated from 96-well plate number 17 at well position G2.

2.3.4.1. Kinetic analysis

Purified and dialyzed GUS-WT or GUS-TR3337 variants proteins were assayed for glucuronide and glucosidase activity with *p*NP-glucuronide and *p*NP-glucoside, respectively, as the substrate. The rate of hydrolysis was monitored through the release of *p*NP at 405 nm (extinction coefficient $14730.38 \text{ M}^{-1}\text{cm}^{-1}$ and path length 0.52 cm) with Cary IE UV-Vis spectrophotometer (Figure 2.6). The standard curve of *p*NP were constructed to determine the extinction coefficients of *p*NP by using the Beer-Lambert equation that is given in Equation 2.1 (Grimsley & Pace 2001):

$$A = \epsilon/c \quad (\text{Equation 2.1})$$

where A is the absorbance, ϵ is the molar absorption coefficient ($\text{M}^{-1}\text{cm}^{-1}$), l is the cell path length (cm) and c is the molar concentration (M).

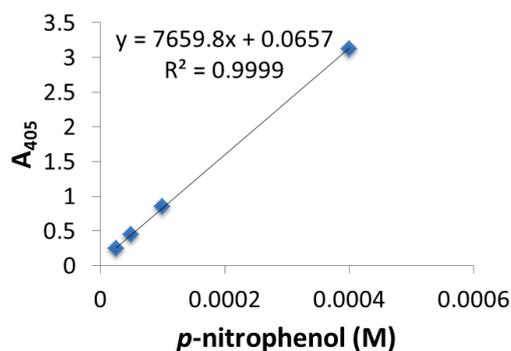


Figure 2.6 The standard curve used to calculate the extinction coefficient value

The standard reaction mixture contained (in a total volume of 200 μL): 50 mM *phosphate* buffer pH 7.4, various concentration *p*NP-glucuronide or *p*NP-glucoside and 20 μL of diluted enzyme sample. The two assay conditions were the standard reaction mixture and the standard reaction mixture with 5 mM octyl- β -D-thioglucopyranoside (OTG). The production rate of *p*NP was monitored spectrophotometrically at A_{405} by various concentrations of *p*NP-glucuronide and *p*NP-glucoside. *p*NP-glucuronide concentrations ranged from 2.5 μM to 20 mM while *p*NP-glucoside concentrations ranged from 2.5 μM to 90 mM.

2.3.4.1.1. Non-cooperative β -GUS enzymes

For β -GUS enzymes that showed no cooperativity, the kinetic parameters were determined by fitting the data to the Michaelis-Menten Equation 2.2 using the curve-fitting software Kaleidagraph Version 3.5 (Synergy Software, USA).

$$v = V_{\max} [S] / (K_m + [S]) \quad (\text{Equation 2.2})$$

where v is the initial velocity, V_{\max} is the maximum velocity, K_m is the Michaelis constant, and $[S]$ is the substrate concentration. K_{cat} was calculated by dividing V_{\max} by the enzyme concentration ($V_{\max} = k_{\text{cat}}[E]$).

2.3.4.1.2. Negatively cooperative β -GUS enzymes

For β -GUS enzymes that showed negative cooperativity, the Eadie-Hofstee is a linear form of the Michaelis-Menten equation that can be rearranged as follows:

$$v = -K_m (v / [S]) + V_{\max} \quad (\text{Equation 2.3})$$

In an Eadie-Hofstee plot, the velocity v is plotted against v divided by $[S]$. These plots were used to obtain the apparent values of V_{\max} and K_m as described in Chapter 5. Microsoft Excel was used for Eadie-Hofstee curve fitting.

2.3.4.2. Thermal optima

The catalytic rate of GUS-WT* and GUS-TR3337* was analysed in reactions at various temperatures, ranging from 25 to 85 °C. The phosphate buffer, containing *p*NP-glucuronide or *p*NP-glucopyranoside, was pre-incubated for 20 min in the micro plate reader. 20 μ L of enzyme was added to give a final reaction volume of 200 μ L in a 96-well microplate.

The activity was determined from the slope during 10 s to 60 s after the reaction was started; the first 10 s of reaction were required for thorough mixing.

2.3.4.3. Thermal denaturation

This trait was measured by incubating 25 μ L of 20 μ M GUS-WT* or GUS-TR3337* sample at one of several temperatures for 5 min. Before and after incubation, the enzyme samples were stored on ice. 20 μ L of 20 μ M GUS-WT* or GUS-TR3337* sample was used in a final reaction volume of 200 μ L. The activity of these GUS-WT* and GUS-TR3337* samples was assayed using 800 μ M of *p*NP-glucoside at 25 °C and pH 7.4. All samples were assayed simultaneously using 96-well plate format to monitor the release of *p*NP in A_{405} over 1 min. The results were analysed by non-linear regression (Kaleidagraph) with the following equation:

$$a = a_0 - a_0 \cdot T^h / (T^h + T_{1/2}^h) \quad (\text{Equation 2.4})$$

where the activity (a) after exposure to temperature (T) is described in terms of original activity (a_0), the temperature at which half the GUS-WT or GUS-TR3337 mutants will be irreversibly inactivated after 5 min ($T_{1/2}$) with a cooperativity term h . Equation was based on

the assumptions that residual activity is proportional to the amount of GUS-WT or GUS-TR3337 mutants that was not irreversibly denatured, and that irreversible denaturation is a cooperative process. Cooperativity is involved because the overall integrity of a protein depends upon the stability of individual secondary structures, each of which depends upon the integrity of neighbouring structures.

2.3.4.4. Thermal shift assay

Overall protein unfolding was monitored using the change in fluorescence caused by binding of the fluorophore SYPRO Orange to hydrophobic residues exposed upon protein denaturation. Fluorescence was measured at 605 – 610 nm with a 7900HT Fast Real-Time PCR system with the temperature increased linearly from 10 – 90 °C at the rate of 1 °C min⁻¹. Samples were run in triplicate, consisting of 5 µM enzyme and 5 x SYPRO orange dye with 5% v/v DMSO (dimethylsulfoxide) 100 mM NaCl in 50 mM phosphate buffer at pH 7.4. Melting temperatures were determined by fitting the data to the Boltzmann equation (Niesen *et al.* 2007) using SciDAVis version 0.2.4 where errors are the standard deviation.

2.3.4.5. Comparison of protein solubility

Cell growth was standardized by growing them aerobically to an optical density of 1 at 600 nm (OD₆₀₀). 5 mL were sedimented at 3700 g for 5 min. The pellet was resuspended in 120 µL of 50 mM phosphate buffer (pH 7.4). 20 µL 7x Bugbuster lysis solution was added and incubated for 30 min at ambient temperature. 70 µL of the lysate was centrifuged. The soluble fraction was aspirated and the insoluble protein was resuspended in 70 µL of 9 M urea. Equal volumes of the lysate, soluble fraction and resuspended insoluble protein were analysed by SDS-PAGE. These SDS-PAGE gels are shown in Chapter 4 (Section 4.4.1.).

2.3.4.6. Discontinuous native protein gel electrophoresis analysis

The discontinuous native gel electrophoresis analysis was carried out as described by Niepmann & Zheng (2006). Briefly, purified proteins were mixed with sample buffer (100 mM Tris-Cl pH 8, 40% glycerol, 0.05% Serva Blue G-250) and incubated for 10 min at room temperature. Protein samples were loaded onto 8% native PAGE gels and separated at 4 °C in a cathode buffer containing 100 mM histidine (adjusted to pH 8 using Tris base) and 0.002% Serva Blue G-250.

2.3.4.7. Bis(sulfosuccinimidyl)suberate (BS³) crosslinking

To covalently cross-link GUS-WT and WT3P24E11, a water soluble BS³ was used. Around 10 µg of purified protein was mixed with 0.15 mM of freshly prepared BS³, then incubated at room temperature for varying time points (0 min, 1 min, 5 min, 20 min). The reactions were

quenched by adding 5 μ L of 1 M Tris. The protein samples were then boiled at 95 °C for 5 min, and separated in the 12% SDS-PAGE. Proteins including the crosslinked proteins were detected by Coomassie Blue staining.

2.3.4.8. Effects of additives

The effects of additives were evaluated by measuring glucosidase activity in the presence of 10 x BugBuster™ protein extraction reagent, octyl- β -D-glucoside, sodium dodecyl sulphate, octyl- β -D-thiogluconide, Tween 20 and Triton-X-100. The reaction mix with respective additives, but without enzyme, served as blank. The reaction mix without any additives was taken as a control (100%).

2.3.4.9. Determination of the modes of enzyme inhibition and the inhibitory constant (K_i) values

The mode of enzyme inhibition was graphically determined from the Lineweaver-Burk double reciprocal plots of the reciprocal of velocity versus the reciprocal of substrate concentration. The inhibitory constant (K_i) value was calculated with the secondary plot representing the slope rate of the Lineweaver-Burk plot versus the OTG concentrations (Ramasamy *et al.* 2014).

2.3.4.10. Enzyme unfolding

The process of enzyme unfolding was monitored by assaying the initial rate of the residual enzyme activity and the fluorescence spectrum. In the enzyme activity assays, GUS-WT and GUS-TR3337 mutants were incubated with various concentrations of urea and GdnHCl in phosphate buffer (50 mM, pH 7.4) at 25°C. After 10 min, the substrate hydrolysis was measured by following the absorbance at 405 nm for 10 min at 6 s intervals. The rate of hydrolysis in the absence of urea was taken as 100%. In the intrinsic fluorescence spectroscopic experiments, the spectra were recorded in the range 300 – 380 nm after exciting the sample at 295 nm, at a protein concentration of 0.6 μ M.

2.3.5. Enzyme crystallisation screens

High through-put screening for appropriate protein crystallisation conditions were performed using the Hampton Research® crystal screens including Index™, Crystal Screen™, PEGRx™, PEG/Ion™ and SaltRx™. Crystal screens were performed as microbatches in 96-well crystal plates. WT5P26A7 was concentrated to 20 mg/mL in 20 mM HEPES, pH 7.4. Hanging drop method was used for crystallisation and 1 μ L of protein sample was mixed with 1 μ L of each screen buffer. Plates were stored at 4 °C and examined under microscope for the presence of crystallisation.

2.3.6. Steroid glycosynthase reactions

0.4 mg/mL GUSWT-E504A, GUSWT-E504G, GUSWT-E504S, WT3P24E11-E504A, WT3P24E11-E504G, WT3P24E11-E504S, WT5P26A7-E504A, WT5P26A7-E504G, WT5P26A7-E504S, GUSTR3337-E504A, GUSTR3337-E504G, GUSTR3337-E504S, THERMO4P11F2-E504A, THERMO4P11F2-E504G or THERMO4P11F2-E504G was added to a solution containing 2 mM DHEA *O*-(carboxymethyl)oxime (CMO-DHEA) or 5 mM testosterone, 10% v/v *tert*-butanol and 4 mM α -D-glucosyl fluoride in 100 mM sodium phosphate buffer (pH 7.5). The reaction was incubated at 37 °C for 2 days. The reaction was then subjected to solid-phase extraction (SPE). The Oasis® WAX SPE cartridge (3 mL, Waters Co.) was pre-conditioned with 1 mL methanol, then 3 mL mQH₂O. The crude reaction was loaded into the cartridge and washed with 3 mL of 2% formic acid, 3 mL mQH₂O, 3 mL methanol, then ammonium hydroxide in methanol (5% v/v, 9 mL). The solvent from the appropriate fraction was evaporated to dryness by rotary evaporation. The dried residue was resuspended in an appropriate deuterated solvent or mQH₂O and subjected to ¹H nuclear magnetic resonance or electron-spray ionisation mass spectrum analysis, respectively. This work did not constitute a large portion of the experimental work of this thesis and was done with the assistance of Mr. Andy Pranata of the McLeod lab. The experimental details of this work have been published (Ma *et al.* 2014).

2.4. References

- Ashkenazy, H., E. Erez, E. Martz, T. Pupko & N. Ben-Tal (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res*, 38, W529-33.
- Cadwell, R. C. & G. F. Joyce (1992) Randomisation of genes by PCR mutagenesis. *PCR Methods Appl*, 2, 28-33.
- Delano, W. L. 2002. The PyMOL Molecular Graphics System.
- Gasteiger, E., C. Hoogland, A. Gattiker, S. e. Duvaud, M. R. Wilkins, R. D. Appel & A. Bairoch. 2005. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, ed. J. M. Walker, 571-607. Totowa, NJ: Humana Press.
- Grimsley, G. R. & C. N. Pace. 2001. Spectrophotometric Determination of Protein Concentration. In *Current Protocols in Protein Science*. John Wiley & Sons, Inc.
- Kirschman, J. A. & J. H. Cramer (1988) Two new tools: multi-purpose cloning vectors that carry kanamycin or spectinomycin/streptomycin resistance markers. *Gene*, 68, 163-5.
- Leung, D. W., E. Chen & D. V. Goeddel (1989) A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique*, 1, 11-15.

- Liu, J. W., Y. Boucher, H. W. Stokes & D. L. Ollis (2006) Improving protein solubility: the use of the *Escherichia coli* dihydrofolate reductase gene as a fusion reporter. *Protein Expr Purif*, 47, 258-63.
- Ma, P., N. Kanizaj, S. A. Chan, D. L. Ollis & M. D. McLeod (2014) The *Escherichia coli* glucuronylsynthase promoted synthesis of steroid glucuronides: improved practicality and broader scope. *Org Biomol Chem*, 12, 6208-14.
- Niepmann, M. & J. Zheng (2006) Discontinuous native protein gel electrophoresis. *Electrophoresis*, 27, 3949-51.
- Niesen, F. H., H. Berglund & M. Vedadi (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protocols*, 2, 2212-2221.
- Novel, G. & M. Novel (1973) [Mutants of *E. coli* K 12 unable to grow on methyl-beta-D-glucuronide: map location of uid A. locus of the structural gene of beta-D-glucuronidase]. *Mol Gen Genet*, 120, 319-35.
- Ramasamy, S., L. V. Kiew & L. Y. Chung (2014) Inhibition of human cytochrome P450 enzymes by *Bacopa monnieri* standardized extract and constituents. *Molecules*, 19, 2588-601.
- Sambrook, J. & D. Russell. 2001. *Molecular Cloning: A Laboratory Manual*. 2001. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Schwede, T., J. Kopp, N. Guex & M. C. Peitsch (2003) SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res*, 31, 3381-5.
- Shrake, A. & J. A. Rupley (1973) Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J Mol Biol*, 79, 351-71.
- Stevenson, B. J. (2006) Directed evolution of pyruvate decarboxylase for *in vitro* glycolysis.
- Wanner, B. L. (1983) Overlapping and separate controls on the phosphate regulon in *Escherichia coli* K12. *J Mol Biol*, 166, 283-308.
- Wilkinson, S. M., C. W. Liew, J. P. Mackay, H. M. Salleh, S. G. Withers & M. D. McLeod (2008) *Escherichia coli* glucuronylsynthase: an engineered enzyme for the synthesis of beta-glucuronides. *Org Lett*, 10, 1585-8.
- Xiong, A. S., Q. H. Yao, R. H. Peng, X. Li, H. Q. Fan, Z. M. Cheng & Y. Li (2004) A simple, rapid, high-fidelity and cost-effective PCR-based two-step DNA synthesis method for long gene sequences. *Nucleic Acids Res*, 32, e98.
- Zhao, H., L. Giver, Z. Shao, J. A. Affholter & F. H. Arnold (1998) Molecular evolution by staggered extension process (StEP) *in vitro* recombination. *Nat Biotechnol*, 16, 258-61.
- Zheng, L., U. Baumann & J. L. Reymond (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res*, 32, e115.

Chapter 3

Identification of Residues Critical to Substrate Binding in β -GUS

3.1. Introduction

As noted in Chapter 1, the initial objective of this research was to alter the substrate specificity of β -GUS. The possible target substrates were conjugates of monosaccharide sugars that were like β -glucuronic acid (Figure 3.1). It was thought that a reasonable place to start was to examine the ability of native β -GUS to act on potential target substrates. The objective was to identify those substrates that gave rise to low levels of activity with β -GUS as they would be suitable targets for future directed evolution studies. β -GUS was known to be very specific for the glycone, but not the aglycone. This meant that it was reasonable to choose an aglycone that was easy to detect in an enzyme assay – *para*-nitrophenol (*pNP*). The term “reasonable”, was meant to imply that mutant forms of β -GUS that act on *pNP* derivatives were also likely to act on substrate with a variety of aglycones, as does the native enzyme. The compounds to be tested were: *pNP*-glucuronide, a control and native substrate, and *pNP*-glucoside, *pNP*-galactoside, *pNP*-mannoside, *pNP*-xyloside, the possible target substrates.

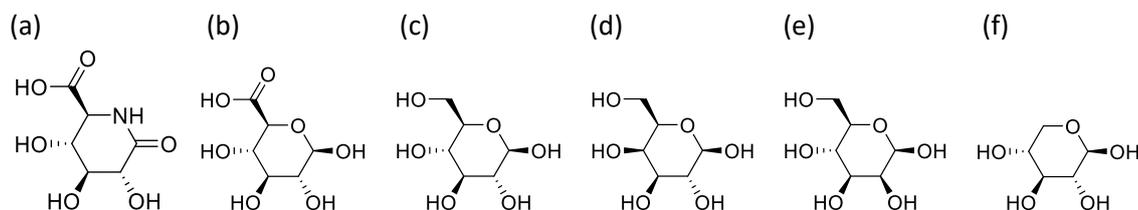


Figure 3.1 Chemical structures of the inhibitor and sugars

(a) GDL (b) β -glucuronic acid (c) β -glucose (d) β -galactose (e) β -mannose (f) β -xylose.

Of the possible alternative substrates, glucuronic acid was most similar glucose. These two compounds were different in that carbon six of glucose is oxidized to a carboxylic acid in glucuronic acid (Figure 3.1b & c). The structure had been determined of β -GUS with a bound inhibitor that mimics of the transition state of the native substrate during catalysis (Wallace *et al.* 2010). This meant that residues that interacted with the C6 carboxylate could be identified and mutated. The question arose: could simply altering active site residues, known to interact with the C6 carboxylate, alter the substrate specificity of β -GUS from that of a glucuronide degrading enzyme to that of a glucoside degrading enzyme? Testing this idea was relatively

straightforward. As outlined in Chapter 2, protocols for site-saturation mutagenesis were available and the substrate could be tested in a high-throughput assay. Although it was felt that the probability of success was low, the resulting mutant genes could be used (along with that of the native enzyme) to generate libraries for directed evolution. Active site changes often destabilize an enzyme, but this destabilization can be compensated for by second site changes, as would occur in the library generation step of directed evolution.

Similar arguments to those presented in the previous paragraph, can be made regarding a few active site residues. For example, does changing the residues that interact with the C4 hydroxide enhance the activity of variant β -GUS enzymes towards β -galactose derivative? The choice of residues that were changed is described below. For now, it is sufficient to note that the worst-case outcome for the proposed experiments was thought to be improved libraries for directed evolution and a better understanding of the importance of active site residues for catalysis; i.e. a better understanding of how β -GUS functions, specifically the importance of specific residues in binding substrate.

3.2. Measurement of activities in crude cell lysates

The β -GUS activities towards the native and related substrates are shown below in Figure 3.2 with the assay conditions given in Table 3.1. The assay conditions required to get a reasonable estimate of the activity towards non-physiological substrates were quite different to that used to measure the activity with the physiological substrate. As expected, the activity towards non-physiological substrates was significantly weaker than that observed for the native substrate. As shown in Figure 3.2, the native substrate, β -glucuronide, was rapidly hydrolysed by β -GUS with an observed activity of 73 mAU/min, but *p*NP-glucoside, *p*NP-galactoside, *p*NP-mannoside, *p*NP-xyloside substrates were hydrolysed relatively poorly. Notably, the *p*NP-glucoside substrate was down from the native substrate by more than a factor of ten, but it was at least seven-fold better substrate than the other glycosides. As evident from Figure 3.2, the concentrations used to detect activity with the non-physiological substrates were much higher than that used to detect activity with the physiological substrate. The activity of β -GUS towards *p*NP-glucoside was detectable in high throughput assays, provided high substrate concentration was used, so that altering the substrate specificity towards glucoside derivatives was a realistic proposition. Both enhanced (due to desirable changes) and diminished (due to detrimental changes) activities could be detected in high throughput screens.

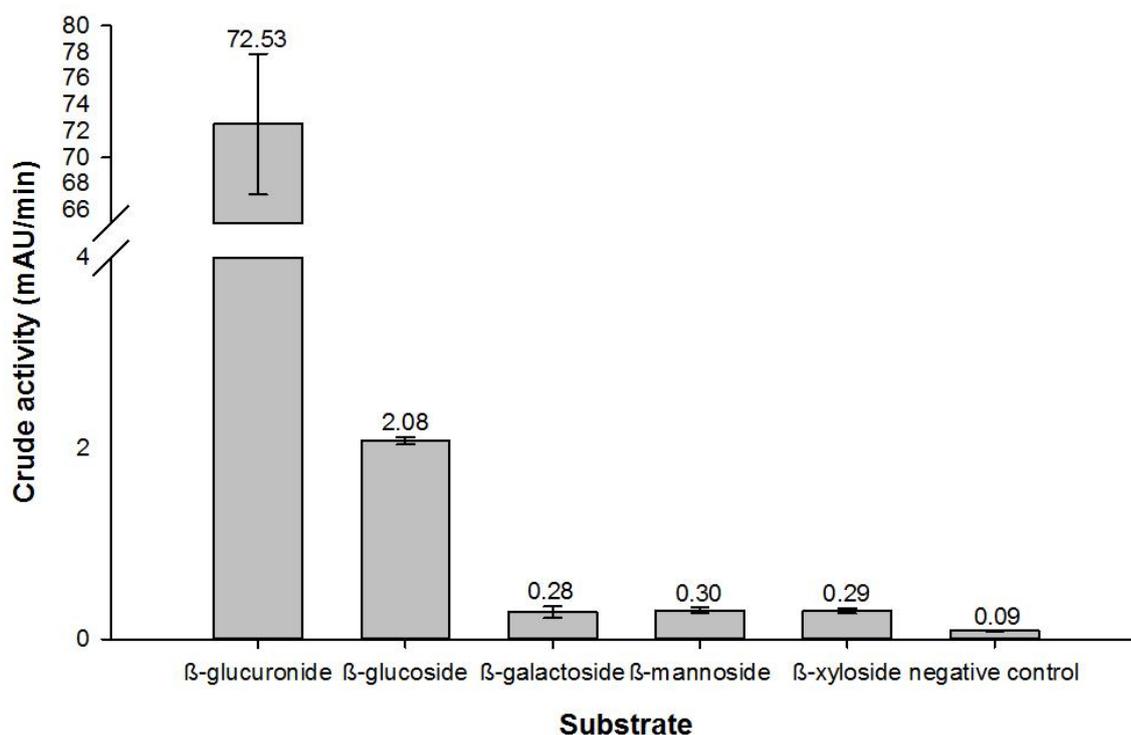


Figure 3.2 Activities of native β -GUS hydrolysing *pNP*-glycoside substrates are presented in the colorimetric assay units (mAU/min)

Error bars represent standard deviation from the mean of triplicate measurements.

Table 3.1 Assay conditions with different substrate concentrations and assay time

Substrate	Assay time	Substrate concentration (mM)
<i>pNP</i> -glucuronide	10 min	0.2
<i>pNP</i> -glucoside	1 day	3
<i>pNP</i> -galactoside	1 day	3
<i>pNP</i> -mannoside	1 day	3
<i>pNP</i> -xyloside	1 day	3

3.3. Choice of sites for site-saturation mutagenesis

The choice of residues to be altered was initially made by examining the structure of *E. coli* β -GUS with a low-affinity inhibitor glucaro- δ -lactam (GDL) (PDB ID: 3K4D) (Wallace *et al.* 2010). The structure of this inhibitor is very similar to the glucuronic acid (Figure 3.1a & b), and serves as a suitable enzyme-glycosyl complex model to deduce important glycosyl binding sites. *pNP*-glucuronide was docked into the active site by pair-fitting *pNP*-glucuronide with the inhibitor GDL. The pair-fit root mean square deviation (RMSD) value was 0.182 (Figure 3.3). Since β -GUS is glycosyl-specific but not specific for the aglycone component, key binding residues were taken to be those that interacted with the glycosyl rings. It was thought that by targeting these sites or nearby sites that changes in the binding behaviour would occur and

that changes in the specificity might occur. The ten residues that were chosen for SSM are listed in Table 3.2.

Table 3.2 The sites chosen for randomisation and their roles

Sites	Comments
D163	Glycosyl C4 interaction
R562	Glycosyl C6 interaction
N412	Basic residue adjacent to catalytic glutamate E413 participating in general acid/base hydrolysis
E413	General acid/base catalyst
M447	Leaving group interaction
Y469	Flipped residue creating room within the active site
Y472	Glycosyl C6 interaction upon flipping into the active site
T509	Important site based on literature not involved in binding
S557	Important site based on literature not involved in binding

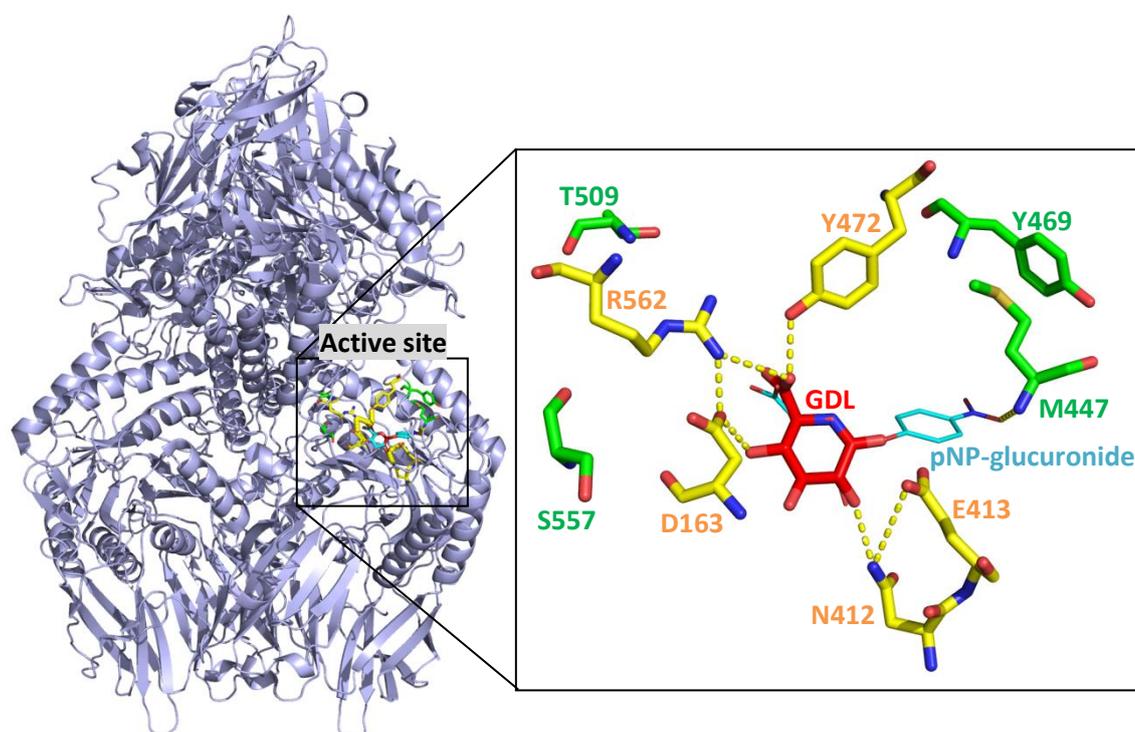


Figure 3.3 Location of residues selected for SSM in β -GUS

Key residues that are colored yellow interact with the compound. Non-glycosyl binding residues are coloured green.

This structural study indicated that the primary contacts between the enzyme and the inhibitor involved D163, Y472, R562, N412 and E413. All of these residues were found in the TIM-barrel domain, except D163 that was located on the loop of Domain 1 that protruded into the active site (see Chapter 1). Residue D163 also forms salt-bridge interaction with R562, that was also implicated in glycosyl binding. Y472 and R562 were observed to interact with the –

COOH group of the inhibitor and were within 3.5 Å of the C6-glycosyl. Residue N412 was a polar residue next to the catalytic residue E413 and was capable of forming hydrogen bond with it. As noted in the introduction E413 was known to be a catalytic residue that was a general base responsible the second step of the double displacement mechanism of the β -GUS. Consequentially, changing N412 was likely to alter the interaction and disturb the configuration in the active site.

Four other residues (M447, Y469, T509 and S557), that were not implicated directly in glycosyl binding, were also investigated as they were in locations that it was thought might allow them to modulate catalysis. Residue M447 forms hydrophobic interaction with the *para*-nitrophenolate (*p*NP) leaving group (aglycone) in the model structure shown in Figure 3.3. It was positioned at the mouth of the active site formed where it could facilitate departure of the leaving group product. In addition, a comparison of the native and the GDL-bound crystal structures of β -GUS suggested that side chain of Y469 flipped away from the active site making room for the carboxylic functional group of the inhibitor. This conformational change was thought to have consequences for substrate binding. In addition to residues chosen on the basis of the β -GUS structure, two residues were selected based on literature reports. Mutants of these two residues, T509A and S557P were reported to have altered substrate specificity compared with native β -GUS (Matsumura & Ellington 2001).

3.4. Sequence comparisons using the ConSurf program

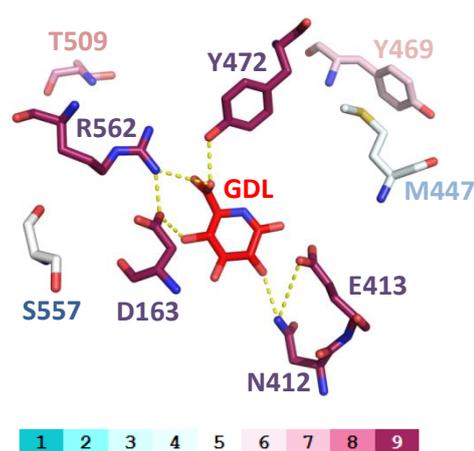
Apart from an inspection of relevant crystal structures, there were other methods of identifying residues of functional or structural significance. Sequence comparisons were frequently used to identify highly conserved residues that were taken to be important for either structure or function. This type of approach can be combined with known structural data has been incorporated into the ConSurf program. The β -GUS amino acid sequence was submitted for ConSurf analysis to quantify the degree of conservation for each residue based on the evolutionary relatedness between the β -GUS and its homologues (Mayrose *et al.* 2004).

A total of 360 homologs were collected from the SWISS-PROT database using CS-BLAST and 123 unique amino acid sequences were selected for analysis (Landau *et al.* 2005). Out of these unique homologous sequences, 89 of them are β -GUS found in different species and 34 other unique sequences are unknown proteins that have significant sequence identity between 35% - 95% with *E. coli* β -GUS. The ConSurf analysis revealed that the catalytic residue (E413) and 4 key glycosyl binding residues (N412, D163, R562 and Y472) have a high

evolutionary conservation score of 9, which indicated a high level of conservation (Table 3.3 & Figure 3.4). A high conservation score indicated the significance of the sites either for function or structure (Draghi *et al.* 2010, Soskine & Tawfik 2010, Tokuriki & Tawfik 2009). These may appear through direct interaction with the substrate (i.e. N412, D163, R562 and Y472) or they may be spatially oriented such that they are involved in the dynamics that will affect important binding sites (e.g. Y469 and T509). Therefore, one expects that highly conserved residues to appear less tolerant towards substitutions. However, highly conserved sites that are involved in substrate binding may also hold the potential for the acquisition of a novel function. Despite the fact that some of the chosen sites were invariant, they were selected for saturation mutagenesis in hope of altering the specificity of β -GUS. In among the chosen residues, M447 had a relatively low evolutionary conservation score of 5 that indicated a high variability of this residue.

Table 3.3 Conservation score and residues variety of targeted residues provided by ConSurf

Amino acid residue	Conservation score 9-conserved, 1-variable	Residues variety
E413	9	E
N412	9	N
D163	9	D
R562	9	R, L
Y472	9	F, Y
T509	8	S, A, T, N
Y469	7	S, F, W, N, P, E, Y, R, L
S557	6	A, S, W, T, N, K, V, H, Q, D, I
M447	5	A, S, F, T, K, E, V, Q, M, R, G



Variable Average Conserved

Figure 3.4 A ConSurf analysis for the β -GUS

The amino acids are colored by their conservation grades using the color-coding bar, with turquoise-through-maroon indicating variable-through-conserved. The run was carried out using PDB code 3K4D and the figure was generated using the PyMol (Sayle & Milner-White 1995) script output by ConSurf.

3.5. Codon randomisation and screening

SSM libraries for each position were made using degenerate oligonucleotides so that each would randomly introduce one of the twenty amino acids. SSM has been performed with a variety of different protocols developed during the past two decades (Hogrefe *et al.* 2002). At the time of writing, the most widely used protocol was the traditional protocol of whole plasmid amplification as implemented by Stratagene QuikChange™. For the present study, the individual codons of interest in the β -*gus* gene were randomised using a single two-stage whole-plasmid PCR as described by Sanchis *et al.* (2008) and shown in Figure 3.5. In the first stage of the PCR, a mutagenic primer contained a single degenerated “NNK” codon (that had 32 triplets, encoding 20 amino acids) and an anti-primer (a non-mutagenic primer used to complete the complementary extension and help in uncoiling the DNA) were used for only 5 cycles to generate a small amount of mega-primer as excess mega-primer promotes self-annealing. Once the mega-primer was generated, the second stage began in which the annealing temperature was increased to avoid the priming of oligonucleotide primers, and a further 20 cycles were carried out to amplify the mutated plasmid.



Figure 3.5 Reaction scheme illustrates a single site randomisation experiment

The gene is represented in red, the vector backbone in blue, the position randomised in green and the formed megaprimer in black.

To avoid contamination with the wild-type enzyme, it was necessary to identify an *E. coli* strain that was free of endogenous glucuronidase activity, GMS407 (see Chapter 2, Section 2.1.2.). Each of the 9 random libraries was introduced into *E. coli* strain GMS407 (Datsenko & Wanner 2000).

An ideal PCR protocol for performing SSM should yield i) a uniform mutational spectrum in terms of quality and ii) a sufficient number of colonies to achieve 95% library coverage, so that there was a 95% probability of observing a given variant within the pool. To ensure that libraries were not biased toward a certain sequence, an average of ten random mutants from each library were sequenced to verify that the codon of interest had been randomly mutagenized. The amino acid residues of the random mutants from each library are shown in Table 3.4. In terms of achieving 95% library coverage, a library comprised of 100%

variants would require screening 96 variants (NNK = 32 codons, $32 \times 3 = 96$) (Patrick *et al.* 2003). To cover the theoretical library size of SSM libraries, two 96-well plates were used for each library to screen for one substrate (total 5 β -glycoside substrates). Each of these plates contained 84 library transformants and 12 positive control wells. Triplicate sets of screens were performed for each library in the presence of 200 μ M *p*NP-glucuronide, 3 mM *p*NP-glucoside, 3 mM *p*NP-galactoside, 3 mM *p*NP-mannoside and 3 mM *p*NP-xyloside, using the assay method described in Section 3.2.

Table 3.4 Summary of sequencing results for the β -GUS random libraries

Stop codons are represented by an asterisk, the number of occurrences of the mutant among sequenced clones is presented by the numbers in superscript.

Amino acid position	D163	N412	E413	M447	Y469	Y472	T509	R562	S557
Amino acids found among the random mutants	G ³ E C ² Q R V *	G ⁴ V ² F D ² *	E G ² V ³ Y W ² R	G ⁴ S L ² A P D	G ⁴ R ² C A V *	G ³ R Q N K ² L D	T G ⁴ Q S A E P	R G ⁷ T D	G ² V ² A ² C ² I L

The activity exhibited by the mutants in a SSM library was classified into four types of phenotype:

- (i) Inactive mutants (relative activity < 10% of the positive control)
- (ii) Partially active mutants (relative activity between 10 and 80%)
- (iii) Neutral mutants (relative activity between 80 to 120%)
- (iv) Improved mutants (relative activity > 120%).

Since only 10 mutants from each category were sequenced, it was not expected that all 19 canonical amino acids would be found but the screen did enable different type of mutations to be identified. For the purposes of describing the results of the activity characterisation, the variants were divided into two categories – those that contacted the substrate and other residues in or near the active site.

3.6. Glycosyl binding residues

The glycosyl binding residues investigated were E413, Y472, D163, R562 and N412. The results of these experiments are given in Table 3.5 below. As was expected the E413 variants

exhibited no activity because of the key role it played in catalysis. The residues Y472 and R562 both interacted with the acid group at the C6 position. No viable variants were identified for Y472, but active variants were found for R562. In the presence of four non-native substrates, mutations at position R562 appeared to retain partial β -glycosidase activity even though they were not active in the presence of native substrate β -glucuronide (Table 3.5). The residue D163 interacted with both R562 and the OH attached the C4 carbon. Again no viable variants were identified for any substrate. There were observations of N412 variants with low levels of activity towards the physiological substrate as well as some other substrates.

Table 3.5 Mutants isolated from the SSM libraries targeting E413, Y472, D163, R562 and N412 residues

The numbers indicate the relative activity of the mutant to β -GUS on 5 substrates: blue if below 10%, red if between 10 and 80%, green if between 80% and 120% and purple if above 120%. The relative activity of each clone was determined in quantitative whole cell glycosidase assays. The tested substrates included *p*NP-glucuronide (*p*NP-glucu), *p*NP-glucoside (*p*NP-glucu), *p*NP-galactoside (*p*NP-gal), *p*NP-mannoside (*p*NP-man) and *p*NP-xyloside (*p*NP-xyl). A stop codon is coded with an asterisk (*).

Position	Amino acid substitution	Relative activity (%)				
		<i>p</i> NP-glucu	<i>p</i> NP-glucu	<i>p</i> NP-gal	<i>p</i> NP-man	<i>p</i> NP-xyl
E413	Y, R, G, W, A, V, S, D, T	n.d.	n.d.	n.d.	n.d.	n.d.
Y472	Q, D, N, K, L, R, F, G, M, *	n.d.	n.d.	n.d.	n.d.	n.d.
D163	V, E, G, Q, C, A, W, S, R, *	n.d.	n.d.	n.d.	n.d.	n.d.
R562	C, G, T, L, K, V	n.d.	~ 30 - 71	~ 22 - 58	~ 31 - 79	~ 22 - 30
	T, M	n.d.	~ 30, ~ 10	~ 40, ~ 23	~ 60, ~ 31	~ 8, ~ 3
	D, W	n.d.	n.d.	n.d.	n.d.	n.d.
N412	D	~ 15	~ 14	~ 13	n.d.	n.d.
	F, Y, A	n.d.	~ 10 - 17	~ 12 - 17	~ 1	~ 11 - 16
	C, E, G, H, I, K, L, R, S, T, V, *	n.d.	n.d.	n.d.	n.d.	n.d.

n.d.: not detected.

To explain the results observed in Table 3.5, at least some mutants should be subjected to much more detailed kinetic studies to determine k_{cat} and K_m values. In addition, structural studies of at least some mutants should be carried out to determine the possible effects of structural change. However, a few points can be made with the available data. The retention of partial activity exhibited by some variants implies that in these variants the substrate binding site must be reasonably similar to that of the native protein. Furthermore, these variants must form a sufficiently strong affinity for the substrates to allow catalysis to occur – that is the K_m cannot be elevated to the point where the activity could not be observed with the substrate concentration used in the assay. Residues R562 and N412 fall into this category. The results observed with R562 are not surprising as this residue does not appear to be essential for binding non-physiological substrates. N412, on the other hand appears to make a contact that would be common to all the substrates tested, except mannose. A

significant drop in activity could be best explained by a small drop in affinity for the substrates, but with retention of the active structure.

Variants that show no observed activity cannot be explained as easily with the available data. The Both D163 and Y472 make interactions with the native substrate and would be expected to make similar interactions with the other substrates. If this were indeed the case, then one would expect diminished activity as was observed for N412. Thus it appears that the most likely cause for the lack of activity of the D163 and Y472 variants would be an altered active site. A superimposition of the *E. coli* β -GUS crystal structure (PDB ID: 3K46) and the inhibitor bound crystal structure (PDB ID: 3K4D) is given below in Figure 3.6. From this Figure it is apparent that Y472 flips into the active site so that it can interact with the inhibitor/substrate. Y469, on the other hand, would flip away from the active site. As noted below, all variants of Y469 are inactive with all substrates. It is not just “structure” that is important; the ability to undergo an appropriate conformational change is critical for the maintenance of activity.

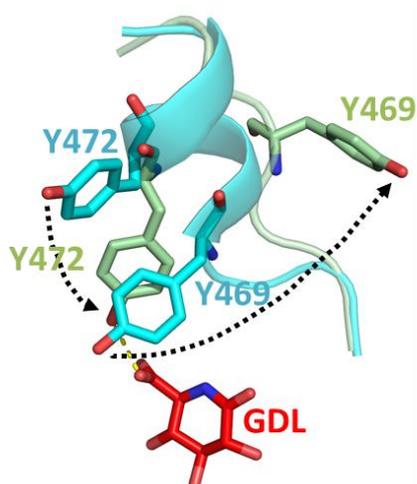


Figure 3.6 Adapted from Wallace *et al.* (2010) *E. coli* β -GUS crystal structure (PDB ID: 3K46; shown in blue) superimposed on the inhibitor bound crystal structure (PDB ID: 3K4D; shown in green)

Y472 flipped into the active site and formed interaction with the carboxylate group at C6, and Y469 flipped out.

3.7. Non-glycosyl binding residues

The results for the experiments with non-glycosyl binding residues are given below in Table 3.6. Although these residues did not participate in binding the glycosyl moiety, they did have other interesting features. As previously noted, M447 appeared to interact with the aglycone moiety (Figure 3.3). Like Y472, Y469 was part of a section of peptide that appeared to undergo a conformational change on binding substrate. Both T509A/S and S557P variants were found

in multiple directed evolution experiments. While the T509 and S557 sites did not participate in substrate binding, the various experiments reported in the literature showed that they can alter the function of β -GUS (Rowe *et al.* 2003, Geddie & Matsumura 2004, Smith *et al.* 2011).

M447 was observed to have a low conservation score residue (conservation score of 5) and showed moderate activity with most β -glycoside substrates (Table 3.6). Nevertheless, it did not produce any mutants that could improve the glycosidase activity. An explanation for this is that its interaction with *p*NP leaving group was not a crucial interaction; this was consistent with the fact that β -GUS was not aglycone-specific. Therefore, mutagenesis at this site would have been unlikely to alter its preference in accommodating the different glycosyl ring in β -glucoside.

Table 3.6 Mutants isolated from the SSM libraries targeting M447, Y469, T5095 and S557 residues

The numbers indicate the relative activity of the mutant to β -GUS on 5 substrates: blue if below 10%, red if between 10% and 80%, green if between 80% and 120% and purple if above 120%. The relative activity of each clone was determined in quantitative whole cell glycosidase assays. The tested substrates included *p*NP-glucuronide (*p*NP-glucu), *p*NP-glucoside (*p*NP-gluco), *p*NP-galactoside (*p*NP-gal), *p*NP-mannoside (*p*NP-man) and *p*NP-xyloside (*p*NP-xyl). A stop codon is coded with an asterisk (*).

Position	Amino acid substitution	Relative activity (%)				
		<i>p</i> NP-glucu	<i>p</i> NP-gluco	<i>p</i> NP-gal	<i>p</i> NP-man	<i>p</i> NP-xyl
M447	Q, H	n.d.	~ 96, ~ 83	~ 68, ~ 15	~ 58, ~ 58	~ 42, ~ 10
	G, S	~ 33, ~ 27	~ 9, ~ 1	~ 12, ~ 17	~ 81, ~ 80	n.d.
	D, A, C, L, P, R, V, W, *	n.d.	n.d.	n.d.	n.d.	n.d.
Y469	A, R, C, V, G, *	n.d.	n.d.	n.d.	n.d.	n.d.
T509	A, S	~ 91, ~ 115	~ 306, ~ 204	~ 186, ~ 173	~ 96, ~ 94	~ 106, ~ 81
	E, Q	~ 74, ~ 15	~ 87, ~ 80	~ 45, ~ 59	~ 32, ~ 41	~ 57, ~ 69
	G	~ 15	~ 107	~ 10	~ 34	~ 23
	L	~ 19	~ 6.6	~ 76	~ 16	~ 85
	P	n.d.	~ 9.4	~ 88	~ 70	~ 68
	C, V, W	~ 22 - 69	~ 27 - 74	~ 23 - 34	~ 25 - 49	~ 22 - 52
	S557	P	~ 75	~ 119	~ 131	~ 63
	V, A, I	~ 27 - 68	~ 83 - 115	~ 81 - 103	~ 35 - 77	~ 11 - 79
	C, L	~ 41, ~ 44	~ 70, ~ 79	~ 10, ~ 11	~ 51, ~ 55	~ 30, ~ 33
	G, W	n.d.	n.d.	n.d.	n.d.	n.d.

n.d.: not detected.

The residue Y469 showed moderate conservation score of 7 but its variants were completely inactivated. As noted above, this observation suggests that the flip of Y469 was important for catalytic activity and that altering Y469 may have significant impact on the dynamics of Y472 and the subsequent interactions of Y472.

Even though residue T509 exhibited a high conservation score of 8 (Table 3.3), it was moderately tolerant and yielded several clones with increased β -glucosidase and β -

galactosidase activities. The sequences of the improved clones were determined to be T509A and T509S. The T509A variant had 3 and 2 times higher β -glucosidase and β -galactosidase activities compared to wild-type while the β -glucosidase and β -galactosidase activities of T509S variant are 2 and 1.7 times greater. Even though this residue does not interact with the substrate or with substrate binding residues, it may modulate the flexibility of the loop that contains the catalytic nucleophile E504, thereby enabling the accommodation of the non-native β -glucoside substrate.

Mutation at S557 also gave variants with moderate glycosidase activity. The S557P variant identified by Matsumura *et al.* (2001), showed 1.3 times higher than the wild-type with β -galactoside as substrate. It was proposed that the S557P mutations might destabilize the active-site loop, thereby decreasing the specificity of the binding pocket.

3.8. Concluding remarks

The experiments described in this chapter have the potential of providing information relating to the structure and function of β -GUS. The inability to find viable mutants of Y469 was unexpected as it was not a highly conserved residue. The low levels of activity exhibited by variants of N412 were surprising as it was adjacent to a key catalytic residue that figures prominently in catalysis. Although the activities of these variants were very low, it would be interesting and informative to examine the structures of these variants and to determine their kinetic properties; was there a small structural change that caused weaker binding (as manifested by a higher K_m) or did the structural change make the location of the nucleophile suboptimal so that a decrease k_{cat} resulted. These questions are left for future investigators.

The principal aim of this work was to alter the substrate specificity. The approach taken in this chapter was to observe the effects of changes to active site residues – those that appear to interact with the substrate and those that sit of the periphery of the active site. Based on the results of site-saturation mutagenesis experiments, the chosen sites can be classified into four groups:

- (i) Critical, only the wild-type amino acid had activity towards substrates
- (ii) Restrictive, substitutions had 10% to 80% activity towards substrates
- (iii) Partially tolerant, 1 - 3 substitutions had at least 80% activity towards substrates
- (iv) Tolerant, 3 - 6 substitutions had at least 80% activity towards substrates.

The critical set of residues included the catalytic residue (E413), 2 glycosyl binding residues (Y472 and D163) and a non-glycosyl binding residue (Y469). Another 2 main glycosyl binding residues (R562 and N412) were restrictive sites. Only 2 amino acid types were consistent with wild-type levels of hydrolysis at position M447. Finally, variants with changes at positions 509 and 557 showed consistent activity with a broad range of substrates; some with significantly more activity than the wild-type enzyme. Notably, variants with the following changes, T509A/S, showed about 3- and 2-fold higher β -glucosidase activity than that of the wild-type β -GUS, respectively. T509A and T509S were also exhibiting an average of 0.6 and 0.4-fold improvement, respectively, in galactosidase activity. Hence, this implied that position T509 may affect the substrate specificity of the enzyme. It is also important to highlight that even though assays were performed with triplicate culture of each mutant, β -galactosidase, β -mannosidase and β -xylosidase activities in crude lysates were too low to be measured accurately. Therefore, unless a mutant exhibits a significant higher glycosidase activity than of wild-type β -GUS, it is thought to be unworthy of further investigation.

In total, 6 of the 9 mutated residue positions are highly conserved positions and important for glycosidase activities (E413, Y472, Y469, D163, R562 and N412). Five of this set of essential residue were thought to be important for the binding and positioning of substrate. The highly variable positions are S557 and T509, which are slightly further away from the inhibitor. T509 obtained a high conservation score of 8 but 2 different mutants were found to have increased activity against substrates with β -glucoside and β -galactoside while most of the isolated mutants are active against the β -glycoside. Single site mutants T509A, T509S and S557P were found to have enhanced glucosidase and galactosidase activities. The effect of T509A and S557P mutation on the glucosidase activity will also be discussed in Chapter 4 along with a description of experiments to more effectively alter substrate specificity.

3.9. References

- Datsenko, K. A. & B. L. Wanner (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A*, 97, 6640-5.
- Draghi, J. A., T. L. Parsons, G. P. Wagner & J. B. Plotkin (2010) Mutational robustness can facilitate adaptation. *Nature*, 463, 353-355.
- Geddie, M. L. & I. Matsumura (2004) Rapid evolution of beta-glucuronidase specificity by saturation mutagenesis of an active site loop. *J Biol Chem*, 279.
- Hogrefe, H. H., J. Cline, G. L. Youngblood & R. M. Allen (2002) Creating randomised amino acid libraries with the QuikChange Multi Site-Directed Mutagenesis Kit. *Biotechniques*, 33, 1158-60, 1162, 1164-5.

- Landau, M., I. Mayrose, Y. Rosenberg, F. Glaser, E. Martz, T. Pupko & N. Ben-Tal (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res*, 33, W299-302.
- Matsumura, I. & A. D. Ellington (2001) In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. *J Mol Biol*, 305, 331-9.
- Mayrose, I., D. Graur, N. Ben-Tal & T. Pupko (2004) Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior. *Mol Biol Evol*, 21, 1781-91.
- Patrick, W. M., A. E. Firth & J. M. Blackburn (2003) User-friendly algorithms for estimating completeness and diversity in randomised protein-encoding libraries. *Protein Eng*, 16, 451-7.
- Rowe, L. A., M. L. Geddie, O. B. Alexander & I. Matsumura (2003) A comparison of directed evolution approaches using the beta-glucuronidase model system. *J Mol Biol*, 332.
- Sanchis, J., L. Fernandez, J. D. Carballeira, J. Drone, Y. Gumulya, H. Hobenreich, D. Kahakeaw, S. Kille, R. Lohmer, J. J. Peyralans, J. Podtetenieff, S. Prasad, P. Soni, A. Taglieber, S. Wu, F. E. Zilly & M. T. Reetz (2008) Improved PCR method for the creation of saturation mutagenesis libraries in directed evolution: application to difficult-to-amplify templates. *Appl Microbiol Biotechnol*, 81, 387-97.
- Sayle, R. A. & E. J. Milner-White (1995) RASMOL: biomolecular graphics for all. *Trends in Biochemical Sciences*, 20, 374-376.
- Smith, W. S., J. R. Hale & C. Neylon (2011) Applying neutral drift to the directed molecular evolution of a β -glucuronidase into a β -galactosidase: Two different evolutionary pathways lead to the same variant. *BMC Research Notes*, 4, 1-10.
- Soskine, M. & D. S. Tawfik (2010) Mutational effects and the evolution of new protein functions. *Nat Rev Genet*, 11, 572-82.
- Tokuriki, N. & D. S. Tawfik (2009) Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol*, 19, 596-604.
- Wallace, B. D., H. Wang, K. T. Lane, J. E. Scott, J. Orans, J. S. Koo, M. Venkatesh, C. Jobin, L. A. Yeh, S. Mani & M. R. Redinbo (2010) Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science*, 330, 831-5.

Chapter 4

Stability and the Evolution of New Substrate Specificity

4.1. Introduction

One of the key aims of this thesis was to alter the substrate specificity of GUS. In the previous chapter, site-saturation mutagenesis was used to probe the role and importance of active site residues. At the time this seemed like a reasonable place to start a study to alter the substrate specificity, but it soon became apparent that simply changing one residue did not provide sufficiently useful insights into how the enzyme could be rationally engineered to achieve the project aims. Furthermore, it was suspected that directed evolution was a far more useful approach. However, the choice of a starting point for such a study was not clear. One could use the gene for the native protein or one could use the gene for a mutated form of the protein that had been evolved with selection for enhanced stability (Xiong *et al.* 2007).

It should be noted that the choice of whether to choose the native or the more stable form of β -GUS was not straightforward. The native protein was more active, but there were precedents in the literature that suggested the stable form of the protein was the better choice. Enhancement of β -glucosidase activity in β -GUS could be achieved through introduction of changes at key positions. However, changes that modulate enzyme function frequently hamper interactions that ensure the stability of the protein. For example, Tokuriki *et al.* (2008) observed that changes that confer new functions give rise to a large stability burden and were found to be more destabilizing than neutral mutations. Hence, as changes accumulate, protein fitness declines. To compensate for the destabilizing effect of the crucial function-altering changes, proteins often accommodate changes that exert stabilizing effects so that the protein stability can be restored. In other words, there appears to be a relationship between stability and evolvability as noted by Bloom *et al.* (2006) that enhanced stability allows a protein to withstand the destabilizing effect of functionally beneficial changes and thereby increasing its ability to adopt new functions. This makes the more stable protein an attractive starting point for the proposed study.

After some consideration, it was decided to evolve both the native and stabilized forms of GUS in parallel. This would test the idea that stability promotes evolvability and allow

the effects of evolution to be monitored for both the native and the stable forms of GUS. For example, would the stable form of the protein remain stable or would its stability drop in response to selection for altered substrate specificity? Similarly, would the native form of the protein gain or lose stability to accommodate altered substrate specificity?

In the remainder of this chapter, there is a description of the thermostable enzyme, as well as a general description of the methods used in directed evolution. This section places emphasis on the special needs of the present work; that is the need to evolve the native and stable enzymes in an essentially identical manner. Following this there is a description of the course of the evolution that includes a comparison of the results of the two sets of experiments. These evolution experiments must, by necessity, be carried out by monitoring the activities of proteins as found in crude lysates. However, crude lysates can conceal a great deal of useful information, and a more thorough investigation of multiple factors was required to gain an understanding of the evolutionary process. The chapter concludes with a description of the experiments to characterise the possible effects of evolution; these include changes in the expression and solubility as well as changes in the stability and kinetic properties.

4.2. Thermostable GUS

The thermostable mutant, GUS-TR3337, was chosen as the stable form of GUS. This variant was evolved in the laboratory by Xiong *et al.* (2007) and found to have six changes that gave rise to increased thermostability. As shown towards the end of this chapter, the GUS-TR3337 protein was significantly more stable than the wild-type protein, but displays less activity towards glucuronide substrates when compared with the native protein. Both the native and GUS-TR3337 proteins have extremely high levels of activity towards glucuronide substrates, and low (but detectable) levels of activity towards glucoside substrates. The afore-mentioned facts make the test of “stability enhances evolvability” a reasonable objective.

The effects of the mutations that give rise to the enhanced stability of the GUS-TR3337 are not concentrated in one area of the protein. One of the residues that was changed was S559 that lies at the tetramer interface. Other residues that were changed were R493 and T532 that lie in the TIM-barrel domain near the tetramer surface. Finally, the residue of T509, N566 and S550 were found near the active site of GUS as shown in Figure 4.1. The *gus-tr3337* gene was obtained using site-directed mutagenesis of the *gus-wt* via an overlap extension polymerase chain reaction (PCR) method as described in Chapter 2 (Section 2.3.1.1.2.).

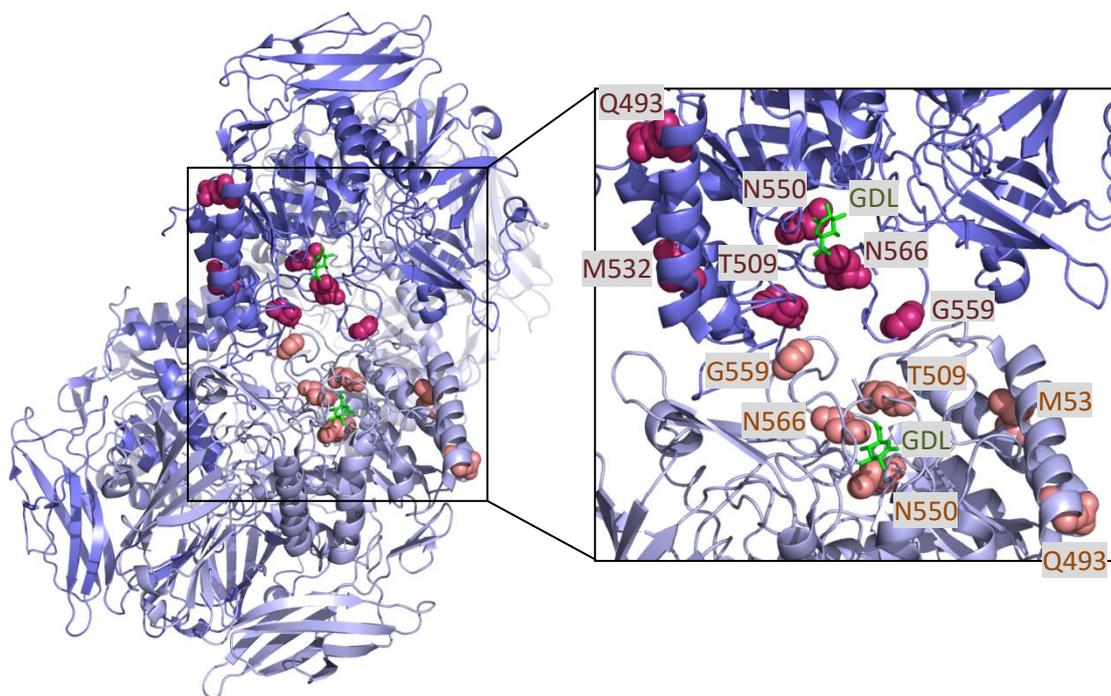


Figure 4.1 The positions of the six amino acid residues Q493, T509, M532, N550, G559 and N566 in the WT-GUS

The labelled residues are shown as colour balls.

4.3. Directed evolution for glucosidase activity

A schematic depiction of the steps involved in directed experiments is given in Figure 4.2. To reasonably compare the evolvability of GUS-WT and GUS-TR3337, it was important to ensure that both sets of mutant libraries were generated under identical conditions with two checkpoints:

1. An error rate that was sufficiently high for changes to be made, but low enough to retain functionality in most of the variant population. In effect, this required library sizes that had a reasonable number of active mutants yet manageable with respect to the screening effort.
2. A distribution of mutations that was random, since all residues in GUS-WT and GUS-TR3337 have potential to influence the protein's properties.

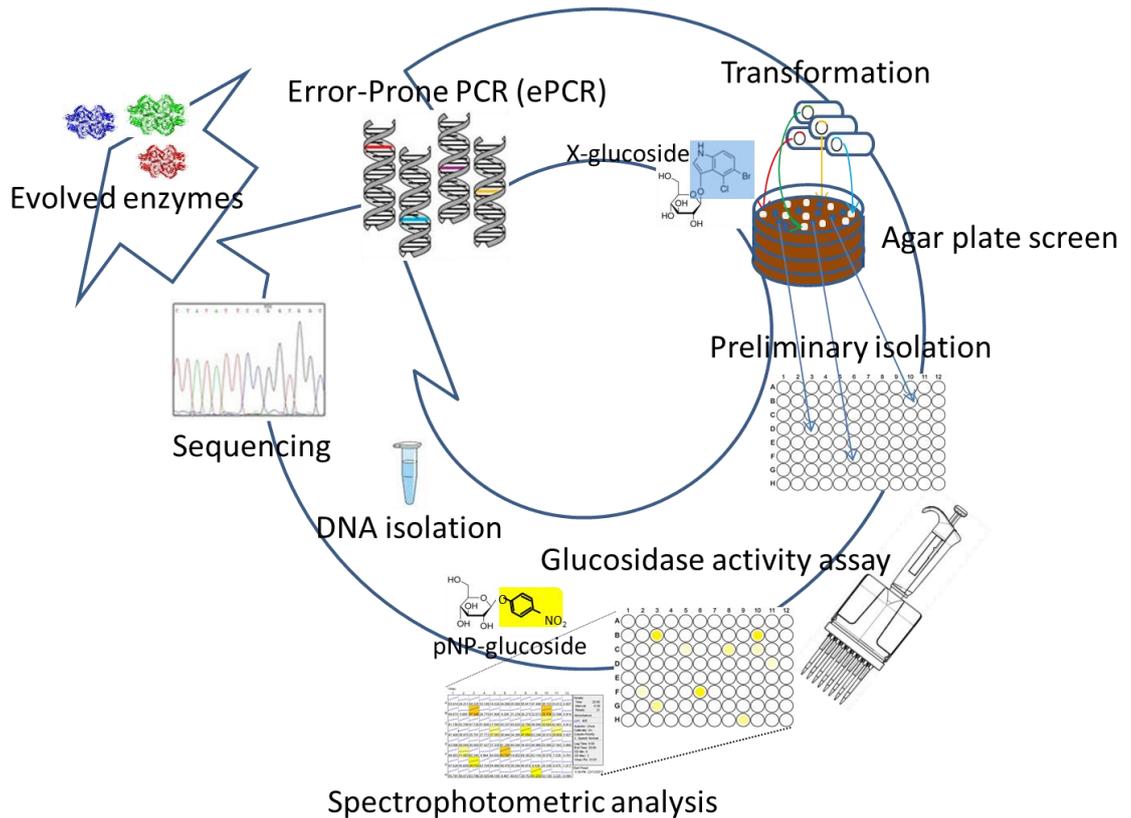


Figure 4.2 Screening strategies

The screening strategy used in directed evolution of GUS-WT and GUS-TR3337 is presented in a schematic form with agar plates and 96-well plates illustrated.

4.3.1. Error rate

The error-rate used to create libraries for directed evolution must maintain a balance between diversity and function. Other workers have investigated this balance and it usually involves a low mutation rate of between 1 to 2 mutations per gene (Moore & Arnold 1996). Part of the rationale for this strategy is that a high mutation rate usually results in the incorporation of detrimental changes, so that very large libraries are required to find the few variants that only have positive or neutral mutations.

To measure the mutation rate, GUS-WT libraries were prepared with different MnCl_2 concentrations. For each MnCl_2 concentration, 20 clones were randomly selected for sequencing. Varying concentration of MnCl_2 from 0.05 to 0.3 mM yielded mutation frequencies between 0.4 to 4.5 amino acid substitutions within a region of the gene that consisted of 165 residues (Figure 4.3). To reduce screening efforts, in the first two rounds, only 165 residues that coded for the domain that had the active site were subjected to random mutagenesis. Justification for this strategy has been provided by workers who demonstrated that active site substitutions can rapidly alter substrate specificity (Geddie & Matsumura 2004,

Chica *et al.* 2005, Parikh & Matsumura 2005). A MnCl_2 concentration of 0.1 mM was chosen to create both libraries as it generated a library containing approximately 2 amino acid substitutions per gene.

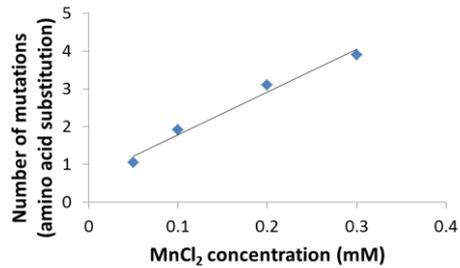
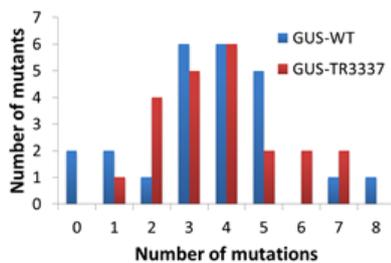


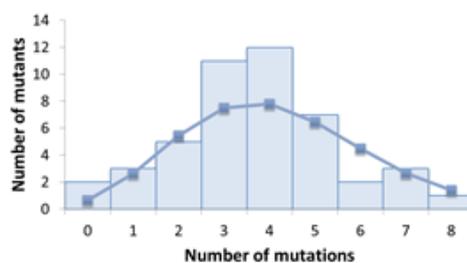
Figure 4.3 Landscape for four mutant libraries using different MnCl_2 concentration

As noted previously, the GUS-WT and GUS-TR3337 libraries were generated under identical conditions. The p-values from the Kolmogorov-Smirnov test (Sachs 1982) suggest that these two libraries had the similar mutation rates. Figure 4.4a, b and c show that the mutational distributions obtained from each library were statistically indistinguishable ($P = 0.98$), the mutational distribution of combined libraries is Poisson distributed ($P = 0.98$) and the mutations are uniformly distributed along the gene ($P = 0.98$), respectively.

(a)



(b)



(c)

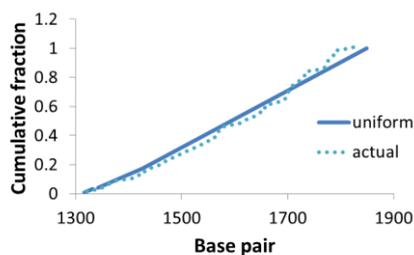


Figure 4.4 Distribution of mutations in the two GUS libraries

(a) The distribution of mutations among 20 randomly chosen GUS-WT mutants and 20 randomly chosen GUS-TR3337 mutants (b) The distribution of mutations (histogram) among all the 40 distribution of sequenced mutants along with a theoretical Poisson distribution (solid line) (c) The cumulative fraction of mutations that occur at that of position in the gene – this demonstrates that the mutations are uniformly distributed along genes.

4.3.2. Generations of GUS-WT and GUS-TR3337 in directed evolution

The directed evolution of GUS-WT and GUS-TR3337 occurred over six generations. Each generation began with the preparation of a variant library followed primary and secondary screens. From the results of the secondary screens, the ten best variants were identified and used to generate the progeny for the next round of evolution. Table 4.1 gives a summary of the experimental details. In all, six rounds of evolution were carried out with GUS-WT and GUS-TR3337; the objective being to isolate variants with improved glucosidase activity.

In generations 1 and 2, the yield of supercoiled plasmid DNA (pDNA) was the limiting factor in determining the library sizes. The diversity available in the library was estimated by the number of GMS407 transformants created during library preparation. However, from the third generation onwards, the low-copy-number pET28a(+) was substituted by a high-copy-number pJexpress 401 that produced sufficient supercoiled pDNA to give about 100,000 GMS407 transformants. Estimates of the number of genotypes examined in each generation are presented in Table 4.1.

Table 4.1 Summary of generation in GUS-WT and GUS-TR3337 directed evolution

The genetic diversity for each generation was provided by ePCR, StEP-ePCR or DNA shuffling. Each generation was obtained through primary screening (colony) and secondary screening (culture) methods. 10 unique genes were used as parents in each generation.

Generation	Diversity	Selection	Randomised region	Genotypes examined	
				GUS-WT	GUS-TR3337
1	ePCR	Colony	T438-Q603	10000	10000
		Culture		3000	3000
2	StEP-ePCR	Colony	T438-Q603	32000	32000
		Culture		3000	3000
3	StEP-ePCR	Colony	F293-Q603	65000	63000
		Culture		3000	3000
4	StEP-ePCR	Colony	M1-Q603	78000	80000
		Culture		3000	3000
5	StEP-ePCR	Colony	M1-Q603	90000	90000
		Culture		3000	3000
6	DNA shuffling	Culture	M1-Q603	3000	3000

In generation 1, about 10,000 cultures were analysed for the primary selection and 3,000 progenies for the secondary selection in both GUS-WT and GUS-TR3337 libraries. Generation 2 began with StEP-ePCR to recombine the changes present in the 10 parents and introduce more random mutations into the libraries. Almost ten times as many colonies were examined in the primary selection before 3,000 cultures were assessed in the secondary

screen. All the selected variants from the second round (i.e. the parents for the next round) were unique.

A major change in the primary screen was necessary for round three of the evolution. The agar plate primary screen (Chapter 2 Section 2.3.1.2.2.1.1.), used in round one and two, could no longer identify improved variants. Evolution had progressed and as a result, as most of the colonies now turned uniformly blue after overnight incubation. In response, top agar screening was used in generations 3, 4 and 5 (Chapter 2 Section 2.3.1.2.2.1.2.). The X-glucoside was applied as a top agar layer after the colonies had grown. The colonies responsible for colour production could then be identified within 30 min.

To increase the probability of finding a beneficial mutation, the length of the mutagenized region of the gene was increased so that in generations 4 and 5 the entire gene was mutated. Along with increasing the length of gene that was mutated, the library size was also increased from generations 3 through 5.

4.3.2.1. GUS-WT and GUS-TR3337 sequence pedigree

For convenience, all variants of *gus-wt* and *gus-tr3337* produced by mutagenesis will be designated *gus-wt** and *gus-tr3337**, respectively. During the directed evolution process, selected *gus-wt** and *gus-tr3337** were sequenced to ensure that they were unique. A summary of the observed changes in the 10 variants from each generation is presented in Table 4.2 while a full list of all changes present in the 50 variants from each library is presented in Table 4.3. The sequences are grouped according their generation so that it is easy to observe the evolutionary process. The variants 1P17G2 implies that it is a first generation (1P) and was found in tray 17 and well G2. In any generation, some new changes appear – a few maybe functional, but most are likely not to be beneficial, but simply occur in the gene in addition to a beneficial mutation (Table 4.3). Changes that enhance activity to a significant (i.e. detectable) degree are retained while the DNA shuffling tends to result in the elimination of neutral mutation. DNA shuffling also accelerated the accumulation of positive mutations. Thus, in generation 5, a number of positive changes were found in all variants – that is, the evolutionary process has converged for these mutations. This extensive analysis of the sequences has provided an interesting example of the progression of directed evolution.

In GUS-WT library, only one of the eight changes that arose in the first generation was maintained in the variant population through to generation 5. This mutation, T509A, reached saturation by generation 3. Another mutation that reached saturation in the finalist population is from generation 2, K568Q. K568Q must be highly beneficial for activity in the

selection conditions since it became saturating within one generation whereas T509A required two generations to achieve saturation.

Similarly, in the GUS-TR3337 library, only one of the eight mutations, S550N, appeared in the first generation and was maintained through to generation 5. Another three mutations that reached saturation in a final population were Q498, S566N and K568E that all appeared in generation 2. Among these four mutations, positions 550 and 566 were reversions to wild-type residues. The reason S550N took three generations to achieve saturation was probably due to the competition between this change and the combination of F551L, S566N and K568E changes. The S550N and F551L arose from changes that were relatively close in the nucleotide sequence so that a recombination event to bring them together would be rare.

Table 4.2 Comparison of GUS-WT and GUS-TR3337 variants over 5 generations

The 10 selected parents from each generation are summarized with the mutations in primary sequence indicated. The asterisks (*) indicate residues that were mutated back to their native residues. The hash (#) symbol is associated to mutated residues that were found in both libraries. Residues that confer enhanced thermostability are bordered in black. Shading has been used to highlight the significant differences. The number of occurrences of the mutation is presented by the numbers in superscript.

Generation	64	279	350	469	474	480	493	498	509	516	532	550	557	559	561	566	568	597
GUS-WT	A	E	D	Y	Q	T	Q	Q	T	M	M	N	S	G	L	N	K	P
WT1	A	E	D	Y	Q	T	Q	Q	A ⁶	M	M	N	P ²	G	L	N	K	Q
WT2	A	E	D	Y	Q	T	Q	L ²	A ⁷	M	K ² /V	N	S	G	L	N	Q ²	Q ³ /L
WT3	A	E	D	Y	Q	T	Q	Q	A ¹⁰	V ²	K	N	P ³	G	L	N	Q ¹⁰	Q ⁸
WT4	A	E	D	N	Q	T	Q	Q	A ¹⁰	V ⁵	M	N	P ³	G	L	N	Q ¹⁰	Q ⁶
WT5	E ³	D ³	G ⁵	N ⁶	Q	T	Q	Q	A ¹⁰	M	M	N	P ⁶	G	L	N	Q ¹⁰	Q ⁹
GUS-TR3337	A	E	D	Y	Q	T	R	Q	A	M	T	S	S	S	L	S	K	P
THERMO1	A	E	D	Y	K	M	R	Q	A	R	T	G/N ²	P ⁴	S	L	N	K	P
THERMO2	A	E	D	H ³	K ³	A	R	K	A	M	M	N ³	P ⁵	S	S	N ²	E ²	P
THERMO3	A	E	D	H ²	H	M	R	K	A	M	M ²	N ⁵	P ³	S	S	N ⁷	E ⁷	P
THERMO4	A	E	D	Y	Q	M ⁸	R	K ¹⁰	A	M	T	N ¹⁰	S	S	L	N ¹⁰	E ¹⁰	P
THERMO5	A	E	D	Y	Q	M ⁸	R	K ¹⁰	A	M	T	N ¹⁰	S	S	L	N ¹⁰	E ¹⁰	P
												↑				↑		
												Reversion				Reversion		

Table 4.3 Changes in GUS-WT and GUS-TR3337 sequences over 5 generations

Residues that are different to the native amino acid are shaded grey and bordered in black at the point where they first appeared. Only the unique variants selected to be parents for future generation are presented. The asterisks (*) indicate residues that were mutated back to their native residues. The hash (#) symbol is associated to mutated residues that were found in both libraries. A schematic representation of the protein domains of β -GUS is illustrated below the table. Domain I, II, III comprising amino acid residue 1-180, 181-274, 275-603 correspond to sugar-binding domain, immunoglobulin domain and TIM barrel, respectively. A schematic representation of the actual domain size is presented at the bottom. The scale bar corresponds to the length in amino acids of β -GUS. (a) GUS-WT library (b) GUS-TR3337 library

(a)

Residues	30	64	115	193	193	151	174	199	234	253	279	319	350	369	372	382	379	427	444	451	469	475	479	484	486	502	503	511	516	522	540	540	557	560	568	582	590	591	592	593	594	596	597	Relative activity
Native	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	T	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	1.00	
1P17G2	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	5.50		
1P22H7	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	5.29		
1P21A3	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	5.15		
1P2F3	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	5.01		
1P7B11	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	4.98		
1P23E12	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	T	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	4.94	
1P21A3	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	4.88		
1P22A7	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	T	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	4.78	
1P27D11	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	4.70		
1P25F12	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	T	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	4.65	
2P14E7	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	T	A	M	K	V	S	I	K	F	G	M	N	F	G	K	P	14.80	
2P4E5	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	I	S	I	K	F	G	M	N	F	G	K	P	14.48		
2P15H11	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	T	A	M	K	V	S	I	K	F	G	M	N	F	G	K	P	14.33	
2P2B7	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	14.51		
2P6G9	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	14.26		
2P4C2	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	T	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	14.73	
2P6E9	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	V	S	I	K	F	G	M	N	F	G	K	P	14.53			
2P5B9	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	14.51		
2P2A7	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	14.49		
2P6F6	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	14.31		
3P22H2	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	94.23		
3P21A4	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	V	P	I	K	F	G	M	N	F	G	K	P	90.26			
3P22F5	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	96.35		
3P25D2	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	93.81		
3P20H2	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	I	S	I	K	F	G	M	N	F	G	K	P	97.52		
3P24E11	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	94.21		
3P20B5	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	P	I	K	F	G	M	N	F	G	K	P	63.07		
3P20H10	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	K	V	S	I	K	F	G	M	N	F	G	K	P	60.43		
3P21G2	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	P	I	K	F	G	M	N	F	G	K	P	67.68		
3P25F3	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	64.04		
4P25G5	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	156.37		
4P26E7	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	V	P	I	K	F	G	M	N	F	G	K	P	156.33			
4P26C8	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	156.25		
4P25E9	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	156.08		
4P25G4	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	I	S	I	K	F	G	M	N	F	G	K	P	155.94		
4P26F8	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	V	S	I	K	F	G	M	N	F	G	K	P	155.87			
4P11H6	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	V	P	I	K	F	G	M	N	F	G	K	P	155.81			
4P25B9	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	155.77		
4P26D7	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	V	S	I	K	F	G	M	N	F	G	K	P	155.70			
4P25H6	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	V	P	I	K	F	G	M	N	F	G	K	P	155.64			
5P1G3	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	P	I	K	F	G	M	N	F	G	K	P	250.19		
5P7H5	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	250.13		
5P26A7	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	P	I	K	F	G	M	N	F	G	K	P	250.00		
5P12C8	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	P	I	K	F	G	M	N	F	G	K	P	243.72		
5P13D8	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	P	I	K	F	G	M	N	F	G	K	P	243.51		
5P3B6	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	I	S	I	K	F	G	M	N	F	G	K	P	243.40		
5P4F3	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	S	I	K	F	G	M	N	F	G	K	P	243.29		
5P15G8	I	A	E	C	N	D	V	H	V	C	E	D	D	N	K	G	A	P	V	A	Y	S	E	V	Q	I	A	M	M	V	P	I	K	F	G	M	N	F	G	K				

(b)

Residues	M	51	81	185	237	260	318	365	369	386	444	452	456	469	474	480	483	487	498	511	510	516	511	522	531	532	533	537	541	550	551	552	557	560	561	565	566	568	570	572	593	594	601	Relative activity						
																																													#	#	#	#	#	#
Native	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	F	A	S	I	L	G	S	K	I	T	F	G	G	1.00						
1P10H11	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	F	A	S	I	L	G	S	K	I	T	F	G	G	6.15						
1P9G5	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	F	A	S	I	L	G	S	K	I	T	F	G	G	5.73						
1P9D7	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	F	A	S	I	L	G	S	K	I	T	F	G	G	5.16						
1P10D12	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	F	A	S	I	L	G	S	K	I	T	F	G	G	5.60						
1P8A5	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	F	A	S	I	L	G	S	K	I	T	F	G	G	5.36						
1P9E6	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	F	A	S	I	L	G	S	K	I	T	F	G	G	5.28						
1P8C11	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	F	A	S	I	L	G	S	K	I	T	F	G	G	5.08						
1P4E9	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	F	A	S	I	L	G	S	K	I	T	F	G	G	5.04						
1P10E4	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	F	A	S	I	L	G	S	K	I	T	F	G	G	5.49						
1P6B3	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	F	A	S	I	L	G	S	K	I	T	F	G	G	5.82						
2P16A8	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	L	A	S	I	L	G	S	K	I	T	F	G	G	23.51						
2P21B10	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	L	A	S	I	L	G	S	K	I	T	F	G	G	23.50						
2P16B9	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	L	A	S	I	L	G	S	K	I	T	F	G	G	23.86						
2P25A6	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	F	T	P	I	L	G	S	K	I	T	F	G	G	23.63						
2P23C4	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	F	A	S	I	L	G	S	K	I	T	F	G	G	20.38						
2P27D3	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	F	A	S	I	L	G	S	K	I	T	F	G	G	22.36						
2P23H5	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	21.24					
2P23C6	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	F	A	S	I	L	G	S	K	I	T	F	G	G	20.61						
2P26A2	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	F	A	S	I	L	G	S	K	I	T	F	G	G	20.33						
2P23H9	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	20.96					
3P19E4	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	L	A	S	I	L	G	S	K	I	T	F	G	G	138.77						
3P24F7	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	138.54					
3P16B10	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	L	A	S	I	L	G	S	K	I	T	F	G	G	102.23						
3P1F6	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	L	A	S	I	L	G	S	K	I	T	F	G	G	108.52						
3P1G6	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	D	M	Y	F	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	108.53
3P11G8	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	D	M	Y	F	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	105.11
3P9B6	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	D	S	L	A	S	I	L	G	S	K	I	T	F	G	G	104.93				
3P9B2	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	D	T	Y	F	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	104.93
3P5E11	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	L	A	S	I	L	G	S	K	I	T	F	G	G	101.40						
3P5E5	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	101.32					
4P11F2	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	785.67					
4P11G8	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	784.62					
4P22F2	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	784.31					
4P26C7	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	784.34					
4P11D2	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	784.22					
4P22A9	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	784.25					
4P5A5	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	784.13					
4P20D10	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	784.10					
4P27B6	K	Y	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	784.05					
4P22F6	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	E	T	Y	F	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	784.01
5P10A5	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	E	T	Y	F	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	781.31
5P16H6	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	781.22					
5P16D9	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	781.13					
5P16H10	K	F	G	D	V	T	M	F	N	Q	V	H	I	Y	Q	T	K	K	Q	A	L	M	Y	S	D	T	Y	F	S	S	N	F	A	S	I	L	G	S	K	I	T	F	G	G	781.18					
5P16A5	M	F	G	D	V	T	M	F	N	Q																																								

4.3.2.2. Improvement in GUS-WT and GUS-TR3337 over five generations

Activity of GUS-WT and GUS-TR3337 improved in each generation under the selection conditions with the greatest improvements occurring in the first and third generation where new screening method was applied (Table 4.4). The improvement in the activity of GUS-TR3337 was much better than that of the wild-type protein (Table 4.4b). Further improvement was not observed in the final generation, suggesting that GUS-TR3337 variants had reached optimal activity under the selection conditions. GUS-TR3337 library also had fewer changes and mutational saturation occurred in generation 4 compared to GUS-WT where mutational saturation occurred in generation 5.

There were two distinct phases in the evolution project: the first phase was completed with the agar plate screening method. This method was replaced by the top agar screening method in the second phase, from generation 3 onwards, however, the selection conditions remained identical to those in generations 1 & 2. That is selection essentially involved observation of increased activity. However, it appeared that the screen used in the second phase was much more sensitive than that used in the first phase. This was supported by the observation that variants from generation 3 in both libraries had the greatest increment in glucosidase activity compared to variants from other generations, as shown in Table 4.4.

Table 4.4 Optimal GUS-WT and GUS-TR3337 versions for each generation

The data have been abbreviated to show only the best variant for each of the five generations. The first number in the variant name refers to the generation. The relative activity refers to the activity of a variant divided by the activity of best variant from the previous generation. Activity was determined with the crude cell lysates and 1 mM *p*NP-glucoside buffered at pH 7.4. (a) GUS-WT (b) GUS-TR3337

(a)

Variant	64	279	350	451	469	498	509	516	532	557	568	597	Relative activity	Compared to:
GUS-WT	A	E	D	A	Y	Q	T	M	M	S	K	P	1.0	-
WT1P17G2	A	E	D	A	Y	Q	A	M	M	S	K	P	5.5	Native
WT2P14E7	A	E	D	P	Y	L	T	M	K	S	Q	Q	2.6	WT1P17G2
WT3P24E11	A	E	D	A	Y	Q	A	M	M	S	Q	Q	6.4	WT2P14E7
WT4P26F8	A	E	D	A	Y	Q	A	V	M	S	Q	Q	1.7	WT3P24E11
WT5P26A7	E	D	G	A	N	Q	A	M	M	P	Q	Q	1.6	WT4P26F8

(b)

Variant	474	480	498	550	557	561	566	568	Relative activity	Compared to:
GUS-TR3337	Q	T	Q	S	S	L	S	K	1.0	-
THERMO1P6B9	Q	T	Q	S	P	L	S	K	5.8	Native
THERMO2P23C4	K	T	Q	S	P	S	S	K	3.5	THERMO2P23C4
THERMO3P24F7	Q	T	Q	N	S	L	N	E	6.8	THERMO3P24F7
THERMO4P11F2	Q	M	K	N	S	L	N	E	5.7	THERMO4P11F2
THERMO5P16H6	Q	M	K	N	S	L	N	E	1.0	THERMO5P16H6

4.3.2.3. Shuffling of best mutants from each generation

To identify a “best” combination of changes, a final shuffling experiment was performed. Three shuffled libraries were generated by using StEP (Zhao & Zha 2006). Unique mutants from GUS-WT library, GUS-TR3337 library and both libraries, were used as parents in shuffled GUS-WT, GUS-TR3337 and combination libraries, respectively (Table 4.5). Including the wild-type variant in the shuffling procedure would allow us to use it as a control for the screening process – the confidence in the screen was higher if none of the selected variants returns as wild-type even though it is included in the parent mix. This objective of the shuffling process was to find the best combination of mutations. 20 clones from each of the three libraries, that were examined for cross-over events. The shuffled GUS-WT, GUS-TR3337 and combination libraries gave an average of 2.5, 2.1 and 2.4 cross-overs per gene. The shuffled mutants were screened like the method used previously for library screening – for further enhancement in glucosidase activity. However, after screening 3,000 mutants from each library, no variant was identified with better activity than WT5P26A7 and THERMO4P11F2, suggesting that, for the screening conditions, the evolution had converged.

Table 4.5 Parental sequences used for shuffled GUS-WT, GUS-TR3337 and combination libraries

Library	Parental genes
Shuffled GUS-WT	<i>gus-wt</i> , <i>wt2p14e7</i> , <i>wt4p26f8</i> and <i>wt5p26a7</i>
Shuffled GUS-TR3337	<i>gus-tr3337</i> , <i>thermo2p23c4</i> , <i>thermo3p24f7</i> and <i>thermo4p11f2</i>
Shuffled combination	<i>gus-wt</i> , <i>wt2p14e7</i> , <i>wt4p26f8</i> , <i>wt5p26a7</i> , <i>gus-tr3337</i> , <i>thermo2p23c4</i> , <i>thermo3p24f7</i> and <i>thermo4p11f2</i>

4.3.2.4. Analysis of mutation type

For generation 1, ePCR was used to produce libraries with an average of 4.3 and 4.2 nucleotide base substitutions per *gus-wt* and *gus-tr3337* gene, respectively. In GUS-WT and GUS-TR3337 libraries, transitions (purine to purine or pyrimidine to pyrimidine mutations) accounted for 62% and 67% of all mutations whereas transversions (purine mutated to pyrimidine or *vice versa*) accounted for the remaining 38% and 33%, respectively (Figure 4.5). In both libraries, up until generation 4, the mutation density of progeny was increased gradually due to the use of StEP-ePCR where this method generates new mutations and propagates existing mutations in new combinations. The escalation of mutation density in generation 3 and 4 was also due to more changes present in the parent genes while the mutagenized region was expanded.

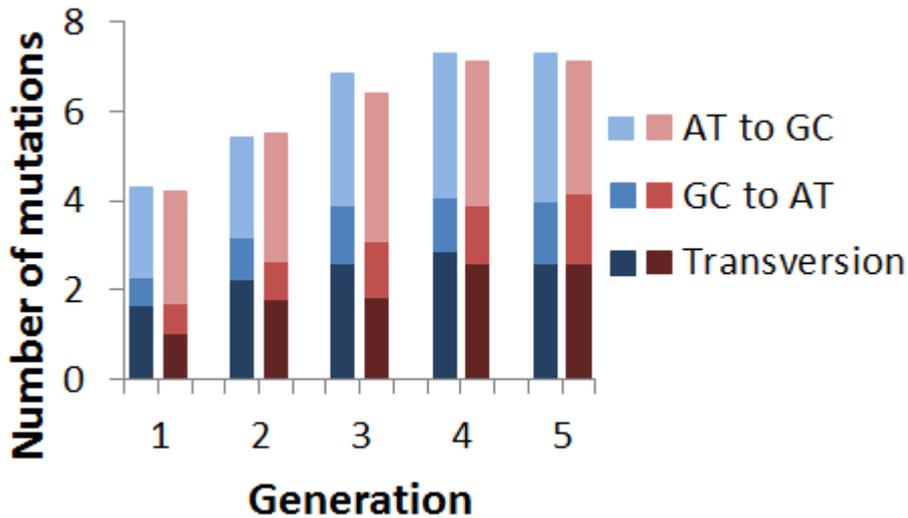


Figure 4.5 Mutation density in progeny from each generation in GUS-WT and GUS-TR3337 libraries

The average number of mutations per *gus-wt* and *gus-tr3337* is presented for the unique progeny of each generation. The chart columns are divided into regions indicating particular types of base substitutions. Blue: GUS-WT library; red: GUS-TR3337 library.

4.3.2.5. Distribution of mutations in the gene and the protein

This project was based on the assumption that beneficial changes could occur at any point in GUS-WT and GUS-TR3337. For this reason, it was necessary for mutations to occur randomly throughout the *gus-wt* and *gus-tr3337* gene. Although the number of mutations was reduced by the selection for active GUS-WT and GUS-TR3337, the location of silent mutations was not affected by selection. Silent mutations, therefore, represent the distribution of mutations in the libraries used for directed evolution. Since the silent mutations observed over the last 2 generations (generation 4 and 5) were evenly distributed along the *gus-wt* and *gus-tr3337* gene (Figure 4.6), it was concluded that the mutagenesis methods did not influence where mutations would occur. The random mutagenesis targeted the entire *gus-wt* and *gus-tr3337* genes in generations 4 and 5, therefore, only the silent mutations in the progeny from these 2 generations were analysed for this purpose.

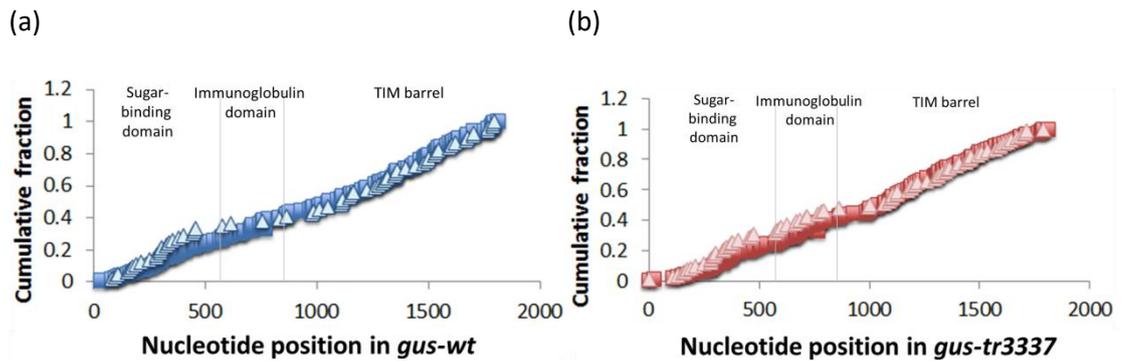


Figure 4.6 Distribution of all mutations points over *gus-wt* and *gus-tr3337*

The positions of silent mutations are plotted as squares and the missense mutations as triangles with respect to the cumulative number of mutations from the start codon. The horizontal axis is divided into regions to indicate each of the three domains in GUS. (a) *gus-wt* (the r^2 value is 0.969 for silent mutations and 0.9778 for missense mutations) (b) *gus-tr3337* (the r^2 value is 0.9931 for silent mutations and 0.9923 for missense mutations).

In contrast to silent mutations, missense mutations (resulting in a change in an amino acid) were not randomly distributed in the GUS-WT or GUS-TR3337 progeny: missense mutations were more likely to occur within the domains rather than at the linker regions. As the mutagenesis process was not responsible for biasing mutation location, it was concluded that the selection for active GUS-WT and GUS-TR3337 was biasing mutations towards regions that can accommodate change and give rise to enhanced activity. Regional bias of missense mutations was observed in one of the first examples of enzyme directed evolution (Moore *et al.* 1996). There were two regions in both GUS-WT and GUS-TR3337 that were susceptible to change. One region corresponded to the end of the sugar-binding domain and the beginning of the immunoglobulin domain (residues 159-190) while the other corresponded to the beginning of the TIM barrel domain (residues 288-331). The location of these two regions within the structure of β -GUS (Figure 4.7) influenced the interaction of the TIM domain with the sugar binding and immunoglobulin domains. In addition, most parts of both regions were buried, and buried residues were known to be less commonly mutated than surface residues (Espinosa *et al.* 2014).

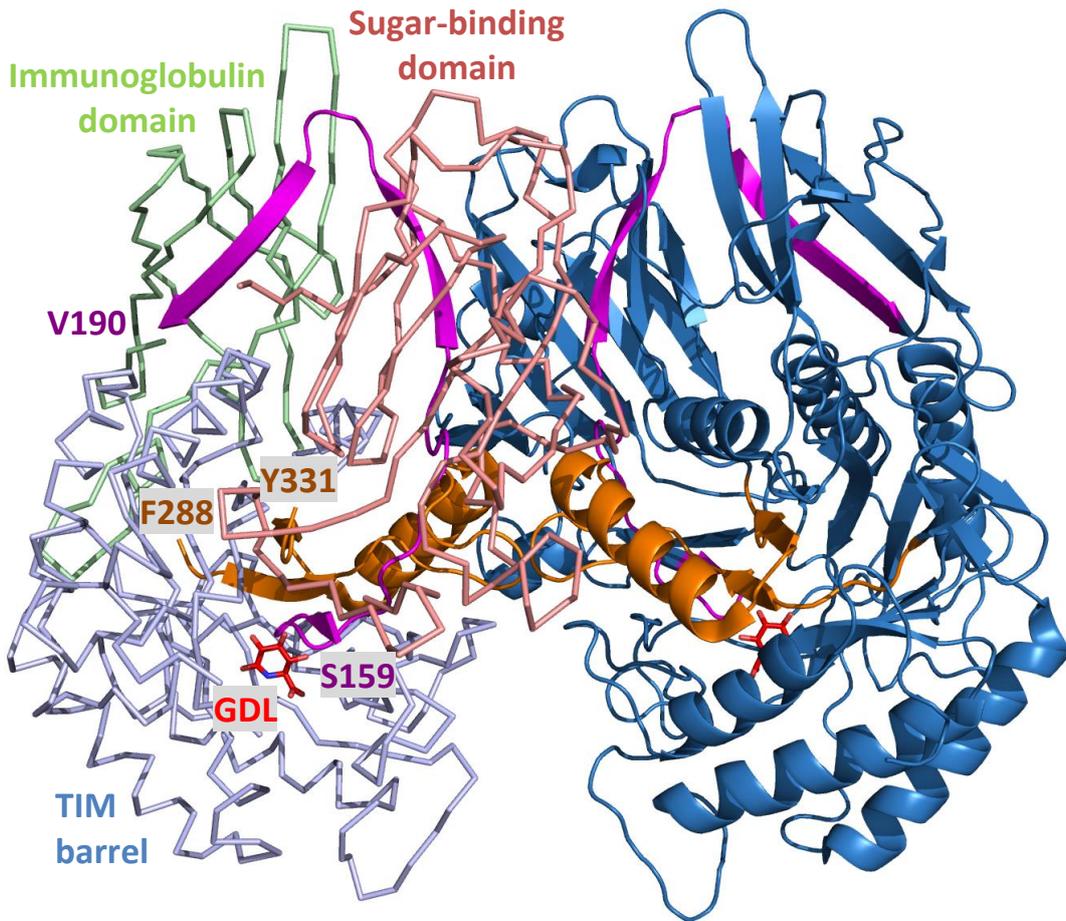


Figure 4.7 GUS-WT regions 159 to 190 and 288 to 331

The dimer of GUS-WT solved by Wallace *et al.* (2010) is illustrated with the established low-affinity inhibitor, glucaro- δ -lactam (GDL) (Niwa *et al.* 1972) molecule in red. The majority of the protein is illustrated as ribbon, whereas the two regions that are vulnerable to missense changes are presented as cartoons. The right monomer is coloured in blue and the left monomer has the three domains labelled and the two vulnerable regions are coloured in purple and orange with the end residues labelled.

The sugar-binding and TIM barrel domains not only tolerated mutations but also allowed changes to improve the performance of GUS-WT and GUS-TR3337 under the selection conditions. Of the eight missense mutations present in the best variant of the WT5P26A7 and five missense mutations present in the best variant of the THERMO5P16H6, one occurred in the sugar-binding domain and the rest occurred in the TIM barrel domain. Both the number and location of beneficial changes in the TIM barrel domain suggested that this domain was robust and that it was a suitable target to alter the substrate specificity.

Overall, GUS-WT was less tolerant of mutations than GUS-TR3337, particularly in the immunoglobulin domain. Mutational analysis of both libraries revealed that GUS-TR3337 had a higher compromised resistance and carried a higher fraction of missense mutations (55%)

than GUS-WT library (45%) (Table 4.6). This suggested that the stabilized variant, GUS-TR3337, showed increased tolerance to missense mutations. The property of resilience to mutations was known to be important to protein engineers as it allowed a larger number of beneficial mutations to be introduced and screened (Elena & Sanjuan 2008).

Table 4.6 Mutational analyses of nucleotide changes

DNA library	GUS-WT	GUS-TR3337
Average number of mutations per clone in generation 4*	7.28	7.12
Average number of mutations per clone in generation 5*	7.29	7.12
Silent mutations (%)	54.79	45
Missense mutations (%)	45.21	55

* Mutations identified in clones sequenced from each library (20 clones per library)

4.4. Characterisation of GUS-WT and GUS-TR3337 variants

Directed evolution resulted in GUS-WT and GUS-TR3337 variants that could be expressed in *E. coli* with improved activity toward 1 mM *p*NP-glucoside at pH 7.4. The analysis of bacterial lysate during the directed evolution process indicated that the best variant for *p*NP-glucoside hydrolysis from GUS-WT library was WT5P26A7 and GUS-TR3337 library was THERMO4P11F2. Both variants were selected for a more thorough characterisation alongside parents GUS-WT and GUS-TR3337. Another six variants from GUS-WT and GUS-TR3337 libraries were selected on the basis that they could help elucidate the effect of mutations at residues 480, 498, 509, 516, 550, 557, 566, 568 and 597. These mutants were: WT1P17G2, WT3P24E11, WT4P26F8, THERMO1P6B9, THERMO2P23C4 and THERMO3P24F7. The final preparations of GUS-WT and GUS-TR3337 variants were dialysed into 50 mM phosphate buffer at pH 7.4 with 1 mM DTT.

4.4.1. Expression and solubility of GUS-WT and GUS-TR3337 variants

The primary goal of this directed evolution was to improve the glucosidase activity of β -GUS. However, that activity was determined with the protein in crude lysates. The level of activity in crude lysates depends upon more than the kinetic properties of the enzyme; it will also be increased by enhanced expression, solubility and stability of the protein. Each of these factors was investigated to determine the observed increase in activity.

To determine changes in the level of expression and solubility due to evolution, GUS-WT and GUS-TR3337 along with a selection of variants were grown under identical conditions in 5 mL test tubes and the results analysed with SDS-PAGE. The overall production, soluble and insoluble forms of GUS-WT and GUS-TR3337 variants, were calculated as the proportion of density attributed to them out of the entire protein density in both the supernatant and pellet analyses (Chapter 2 Section 2.2.). This software quantifies the production and expression level

for each of the GUS-WT and GUS-TR3337 along with their variants. As indicated in Table 4.7, there were no significant improvement in solubility from the GUS-TW and GUS-TR3337 variants, except for WT4P26F8 displayed ~ 7% increase in protein solubility. These results suggest that some of the changes, such as M516V, improved the solubility of β -GUS during expression in *E. coli*. However, the improvement was small. Since the error associated with estimating the expression level was difficult to estimate, it is very likely that they were the same as the differences observed for expression. An inspection of the gels shown in Figure 4.8, gave little indication of changes due to increased expression of the variant proteins.

Overall, the changes in the expression and solubility of GUS-WT and GUS-TR3337 variants were minimal and certainly not enough to account for the differences observed changes in the levels of activity. Analysis of the kinetic properties and stability of these purified variants revealed the more significant effects from this collection of mutations.

Table 4.7 Expression of GUS-WT and GUS-TR3337 variants

Each variant is presented with the residues that vary between them: mutated residues are shaded. The solubility refers to the proportion of GUS-WT and GUS-TR3337 variants that remains in the soluble fraction of cell lysate and was determined by densitometry of the images in Figure 4.11. (a) GUS-WT variants (b) GUS-TR3337 variants

(a)

Variant	64	279	350	469	509	516	557	568	597	Solubility (%)
GUS-WT	A	E	D	Y	T	M	S	K	P	62
WT1P17G2	A	E	D	Y	A	M	S	K	P	65
WT3P24E11	A	E	D	Y	A	M	S	Q	Q	67
WT4P26F8	A	E	D	Y	A	V	S	Q	Q	69
WT5P26A7	E	D	G	N	A	M	P	Q	Q	67

(b)

Variant	474	480	498	550	557	561	566	568	Solubility (%)
GUS-TR3337	Q	T	Q	S	S	L	S	K	69
THERMO1P6B9	Q	T	Q	S	P	L	S	K	68
THERMO2P23C4	K	T	Q	S	P	S	S	K	66
THERMO3P24F7	Q	T	Q	N	S	L	N	E	66
THERMO4P11F2	Q	M	K	N	S	L	N	E	66

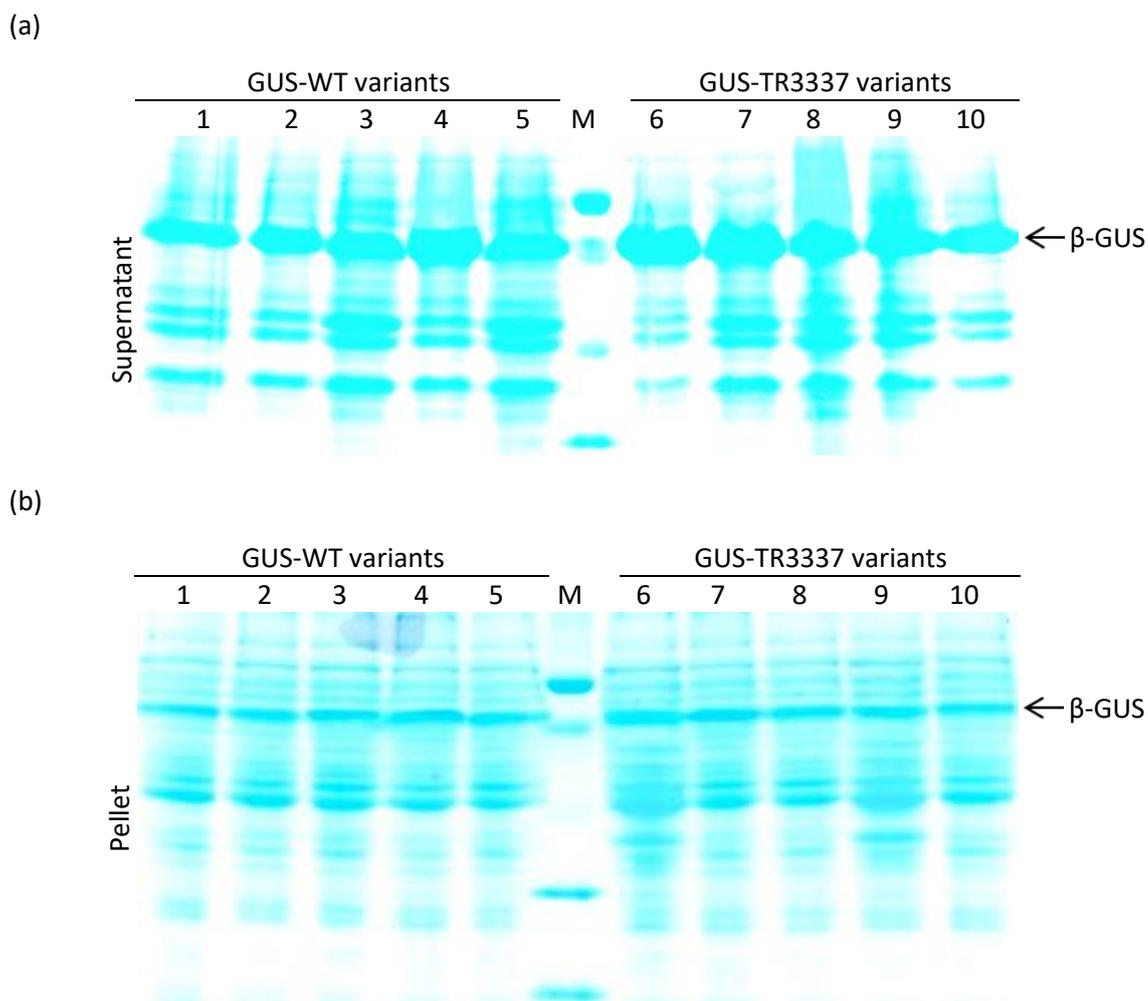


Figure 4.8 SDS-PAGE analysis of GUS-WT and GUS-TR3337 variants expression

During the preparation of GUS-WT and GUS-TR3337 variants, samples were taken of the cell lysate, supernatant and pellet and analysed by SDS-PAGE. The samples were analysed on three 12% gels run simultaneously. The prominent band in supernatant and pellet proteins is GUS (labeled on the right side of gel images). Lane 1: GUS-WT; 2: WT1P17G2; 3: WT3P24E11; 4: WT4P26F8; 5: WT5P26A7; M: Protein marker; 6: GUS-TR3337; 7: THERMO1P6B9; 8: THERMO2P23C4; 9: THERMO3P24F7; 10: THERMO4P11F2 (a) Results for supernatant samples. (b) Results for pellet samples.

4.4.2. Kinetics characterisation of GUS-WT and GUS-TR3337 variants

GUS-WT and GUS-TR3337 variants were characterised against *p*NP-glucuronide and *p*NP-glucoside and the kinetic parameters can be found in Table 4.8. A casual inspection of Table 4.8 reveals a few trends. The glucuronidase activity of both sets of mutants decreases during the evolution. The change is principally caused by a drop in the k_{cat} of the enzymes while there is little change in the K_m . All the variants had higher activities towards *p*NP-glucoside when compared with their parent enzymes. The increases in activities were evident in the k_{cat}/K_m values that ranged from 7 to 1299-fold, and were primarily due to reductions in K_m values. It was expected that improved variants would show lower K_m values as the screening was

performed at a significantly lower substrate concentration, 1 mM, than the observed K_m of GUS-WT (44 mM) and GUS-TR3337 (40 mM). As observed in nature, the K_m of the variant enzymes evolved to the point where the K_m was close to the substrate concentration in their environment. In addition to decreases in K_m value, the variants showed modest enhancement in k_{cat} but GUS-TR3337 variants had greater k_{cat} values than GUS-WT mutants when compared to their respective parent enzymes.

Table 4.8 Kinetic parameters for the hydrolysis of *p*NP-glucuronide and *p*NP-glucoside by selected mutants

Substrate	Variant	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($M^{-1}s^{-1}$)	Relative k_{cat}/K_m	
<i>p</i> NP-glucuronide	GUS-WT	93.09 ± 4.0900	0.17 ± 0.0018	540256.69	1.0000	
	WT1P17G2	70.14 ± 1.1824	0.15 ± 0.0134	465439.85	0.9000	
	WT3P24E11	1.29 ± 0.0256	0.19 ± 0.0166	7203.20	0.0130	
	WT4P26F8	0.53 ± 0.0071	0.59 ± 0.0405	897.94	0.0017	
	WT5P26A7	0.39 ± 0.0060	0.14 ± 0.0101	2852.49	0.0053	
	GUS-TR3337	14.25 ± 0.1798	0.81 ± 0.0402	17608.21	1.0000	
	THERMO2P23C4	17.63 ± 0.3465	1.77 ± 0.1131	9968.59	0.5661	
	THERMO3P24F7	1.06 ± 0.0402	0.90 ± 0.1293	1170.91	0.0665	
	THERMO4P11F2	0.22 ± 0.0062	0.30 ± 0.0734	373.78	0.0212	
	<i>p</i> NP-glucoside	GUS-WT	0.136 ± 0.0040	43.52 ± 2.3900	3.14	1.0000
		WT1P17G2	0.177 ± 0.0030	40.69 ± 2.5100	4.35	1.3853
		WT3P24E11	0.183 ± 0.0020	0.33 ± 0.0230	547.90	174.8000
WT4P26F8		0.453 ± 0.0140	0.96 ± 0.0210	470.96	150.2000	
WT5P26A7		0.751 ± 0.0070	0.78 ± 0.0430	962.61	307.1000	
GUS-TR3337		0.049 ± 0.0008	40.38 ± 1.2700	1.22	1.0000	
THERMO1P6B9		0.39 ± 0.0090	48.53 ± 1.9600	8.04	6.6000	
THERMO2P23C4		0.724 ± 0.0100	24.74 ± 0.7600	29.27	24.1000	
THERMO3P24F7		0.314 ± 0.0060	0.43 ± 0.0620	723.88	595.3000	
THERMO4P11F2		0.89 ± 0.0100	0.56 ± 0.0430	1579.38	1298.8000	

It should be noted that the kinetic constants for glucosidase activity were assayed in the presence of 5 mM octyl- β -D-thioglucoside (this will be discussed in further detail in Chapter 5). The K568Q and P597P mutations in WT3P24E11 further increased affinity to the *p*NP-glucoside substrate, judging from the lower K_m value; but accounted for the significantly reduced k_{cat} of *p*NP-glucuronide substrate. The k_{cat} value of S557P in THERMO1P6B9 increased 6.6-fold, suggesting the significant influence of this mutation on the enzyme activity in substrate binding and product release. The M516V mutation in WT4P26F8 caused a slight decrease in glucosidase activity. This reflects the increased catalytic activity in the isolated clone during the screening process is mainly due to increased protein solubility and that the enzymatic properties of WT4P26F8 were not significantly altered.

Overall, the best variant, THERMO4P11F2, was obtained from GUS-TR3337 library and exhibited 1.6-fold greater k_{cat}/K_m than WT5P26A7, which was the best variant from the GUS-WT library (Figure 4.9).

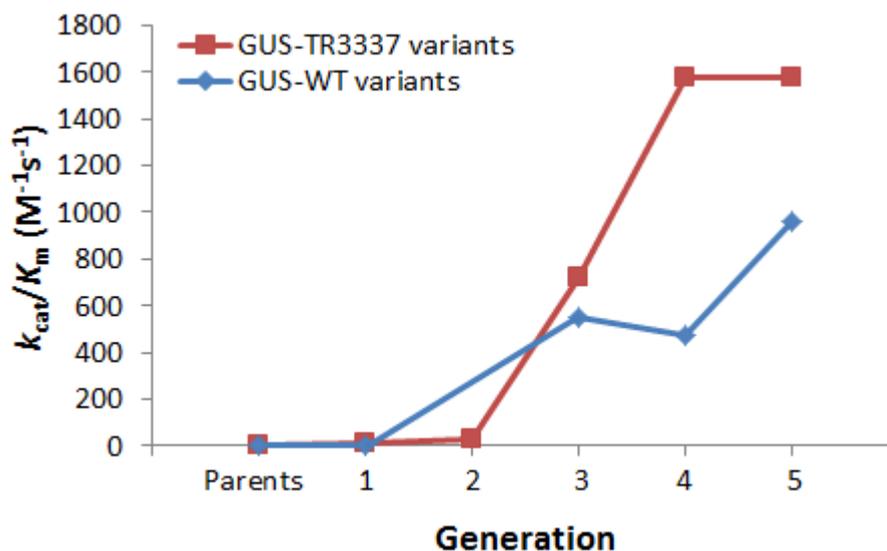


Figure 4.9 Glucosidase specificity constant of GUS-WT and GUS-TR3337 variants through the progress of directed evolution

4.4.3. Optimal catalytic temperature

The temperature optima extracted for GUS-WT and GUS-TR3337 variants are displayed graphically in Figure 4.10c. This figure showed the optimal activities for the starting enzymes and those of the variants with the greatest change from each generation. The trend in the changes in the optimal temperatures is easily seen in Figure 4.10. A more complete display of all the variants is given in Appendix C.1. The results from assays with 800 μ M *p*NP-glucoside at pH 7.4 from 25 °C to 85 °C revealed that the optimal temperatures of glucosidase activity were different, being 45 °C and 60 °C for GUS-WT and GUS-TR3337, respectively. In GUS-WT library, the variants from generation 1-4 have an optimal temperature of 45 °C and 50 °C but the variant from generation 5, WT5P26A7, has increased optima at 55 °C. In GUS-TR3337 library, compared to the thermophilic parent, a shift of the optimal temperature for the variant from generation 4, from 60 °C to 40 °C, was observed. The directed evolution was carried out at 25 °C and therefore variants isolated from the screening process tend to perform better at a lower temperature. Furthermore, the optimal temperature range gradually becomes wider with later generation of directed evolution (Figure 4.10a & b). The variant from generation 4 (THERMOP11F4) displays a wide optimal temperature range between 35 °C to 55 °C. The mutations T480M and Q498K seem to be important residues that enhance glucosidase activity at lower temperatures and thus sustain a wider operating temperature range.

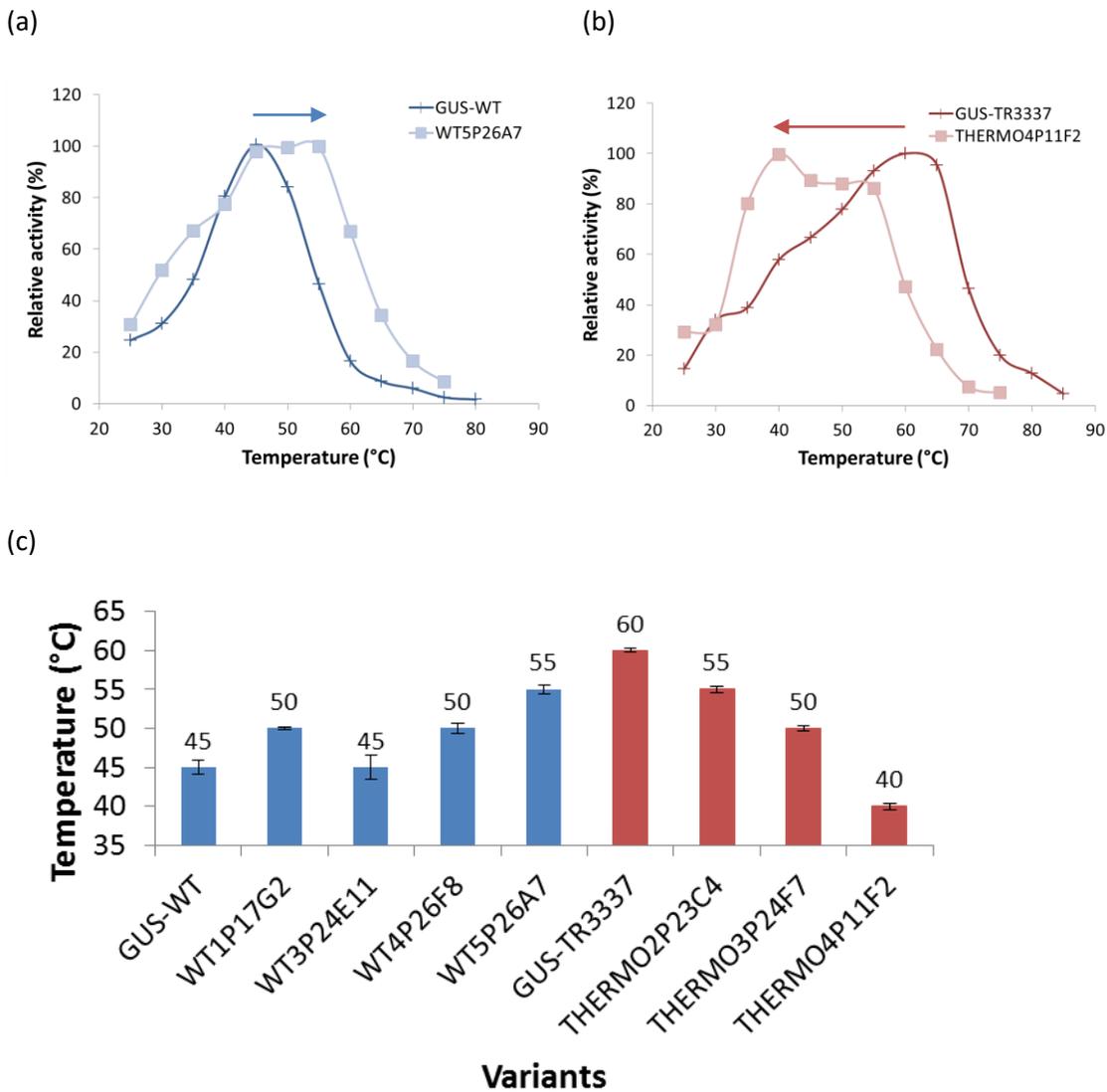


Figure 4.10 Optimal temperature with 800 μ M pNP-glucoiside

The relative activity for each version of GUS in the presence of 800 μ M at pH 7.4 was determined for each GUS variant at a series of temperatures. (a) GUS-WT and WT5P26A7 (b) GUS-TR3337 and THERMO4P11F2 (c) GUS-WT and GUS-TR3337 variants

It was widely accepted that thermostable enzymes tend to display greater conformational rigidity (Vieille & Zeikus 2001). However, protein function depends on both structure and dynamics; making a protein more stable by rigidifying it does not necessarily lead to enhanced functionality as proteins require optimized flexibility to perform their biological roles (Teilum *et al.* 2011). This explains why thermostable enzymes generally function optimally at a higher temperature. At that temperature, the enzymes have increased thermal motion so that they are flexible enough to show full activity. Even though it has been proposed that both proteins exhibit the same flexibility when compared at their respective optimal working temperature (Wolf-Watz *et al.* 2004), the glucuronidase activity of GUS-WT at

its optimum temperature (45 °C) was 1.6-fold higher than the activity of GUS-TR3337 at 60 °C (Figure 4.11a). In contrast, GUS-TR3337 has a higher glucuronidase activity at its optimal temperature compare to GUS-WT which might be mainly due to the presence of T509A mutations that conferred the glucosidase activity (Figure 4.11b). At the screening temperature (25 °C), GUS-WT displayed 8% higher glucosidase activity relative to GUS-TR3337. Even so, the fourth generation variant derived from thermophilic parent, THERMO4P11F2, showed higher glucosidase activity compared to that of fifth generation variant, WT5P26A7, that was derived from the wild-type library. Furthermore, it was evident from figure 4.11b that THERMO4P11F2 was active at the broad temperature range of 35 – 55 °C compared to WT5P26A7 that was active at 45 – 55 °C. Kinetic parameters in Table 4.8 were carried out at 25 °C. Therefore, higher improvement rates can be expected when the kinetic analysis was carried out at their respective optimum temperature.

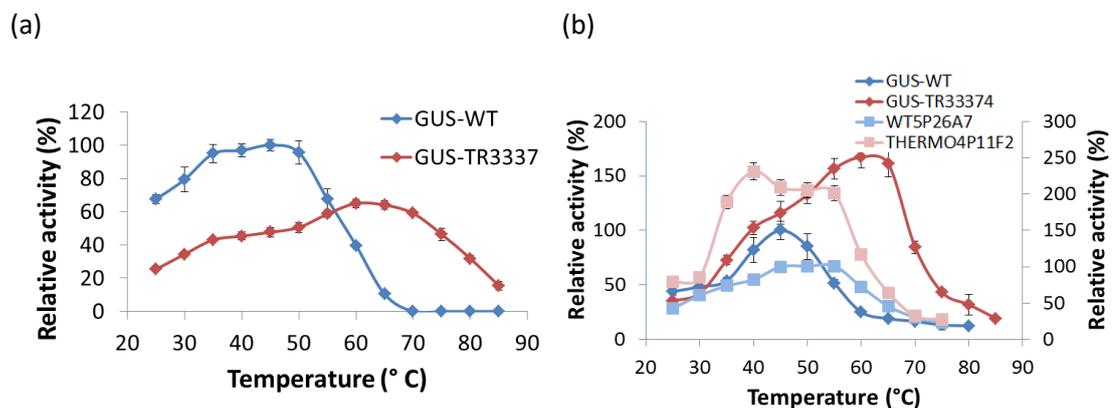


Figure 4.11 Temperature dependence of enzymatic activity

(a) Glucuronidase activities were expressed as relative to that of the highest activity in GUS-WT (b) Glucosidase activities of GUS-TR3337 and THERMO4P11F2 were expressed as relative to that of the highest activity in GUS-WT (right y-axis) and WT5P26A7 (left y-axis), respectively.

4.4.4. Thermal and chemical stability

The measurement of protein stability involves disrupting the protein structure by either physical or chemical means. Temperature is one of the most widely used physical denaturation tools while urea and guanidine hydrochloride (GdnHCl) are the two most common chemical denaturants (Privalov & Khechinashvili 1974).

4.4.4.1. Thermal denaturation

Thermal stability was determined in two ways. The first involved monitoring the deterioration of the activity by measuring residual activity after heat treatment while the second involved the use of the fluorophore SYPRO orange to monitor the overall unfolding of the enzymes. In the first set of experiments, the loss of residual activity reflects a deterioration of the active

site whereas the second set of experiments gives a picture of overall changes in the protein structures. The experimental details for these experiments can be found in Chapter 2 (Section 2.3.4.3. and 2.3.4.4.) and the results are given below in Figure 4.12. The inflection points in the curves for the residues activity as a function of temperature were 45.2 ± 0.02 °C and 74.7 ± 1.2 °C for the GUS-WT and GUS-TR3337 enzymes, respectively. In the GUS-WT directed evolution, variants with the mutations at 509 had an increase in $T_{1/2}$ of 14 °C. However, in GUS-TR3337 library the variant from generation 2 had a decrease in $T_{1/2}$ of 4 °C, whereas variants from generation 3 and 4 had an additional decline of 11-12 °C. The total decrease in $T_{1/2}$ for the variant from generation 4, compared to the thermophilic parent, was 16 °C. Even though this variant had a lower optimal temperature (40 °C) compared to the variant from generation 3 (50 °C), the presence of the two mutations at 480 and 498 did not decrease its $T_{1/2}$.

When measured using the increase in fluorescence caused by binding of SYPRO orange to expose hydrophobic upon protein denaturation, similar trends were observed in both libraries. The melting temperatures of the same sample differed since the two methods used to determine stability measured slightly different unfolding events. The SPYRO orange method gave slightly higher melting temperatures, as it monitors the overall structural stabilities during 1 minute of unfolding; while the residual activities gave lower melting temperatures, as it monitors the conformational stability of the active site during 5 minute of elevated temperatures.

The melting temperature of GUS-TR3337* decreases from the first generation to the fifth generation, however, the melting temperature of THERMO4P11F2 as determined by the individual residual activity plot was 2.4°C higher than WT5P26A7 while the value obtained with fluorescence monitored thermal denaturation was 5.4 °C higher than WT5P26A7.

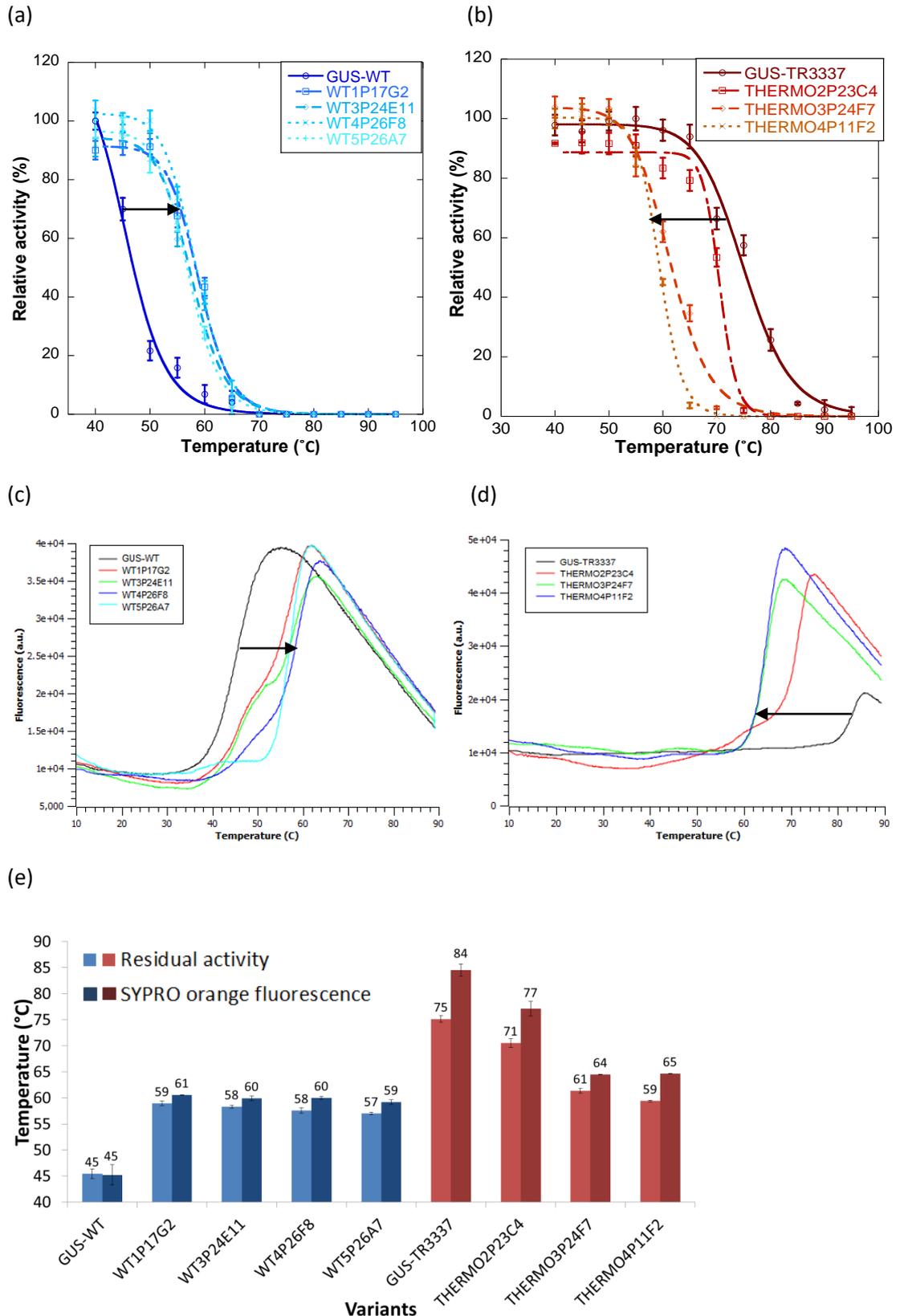


Figure 4.12 Thermal denaturation

(a) Residual activity plot of GUS-WT variants (b) Residual activity plot of GUS-TR3337 variants (c) Change in SYPRO orange fluorescence in GUS-WT variants (d) Change in SYPRO orange fluorescence in GUS-TR3337 (e) Corresponding melting temperatures of GUS-WT and GUS-TR4447 variants, extracted from the residual activity and SYPRO orange fluorescence data.

4.4.4.2. Chemical induced denaturation

The effects of chemical denaturants on GUS-WT, GUS-TR3337 and their variants were also measured. The chemical denaturants were those typically used to study denaturation; they were urea and guanidinium hydrochloride (GdnHCl). The residual activity in the presence of denaturants was used to determine if the active site was intact. In addition, intrinsic fluorescence emission spectra were used; this technique gave information about the environment of the tryptophan (Trp) residues that gave rise to the intrinsic fluorescence. Spectra were recorded from 300 to 380 nm; the red shift of the intrinsic Trp fluorescence emission maximum (λ_{max}) was monitored as a function of urea and GdnHCl concentrations. The experimental details for these experiments can be found in Chapter 2 (Section 2.3.4.10.). A sample of the results is given in the Figure 4.13 and 4.14 below. It was difficult to discern a trend when the data of GUS-WT and its variants were superimposed, so only GUS-WT and the relevant variant are shown in the figures below; similar plots are given with all the sets in Appendix C.2. and C.3. Similar comments apply to the measurements with GUS-TR3337 and its variants.

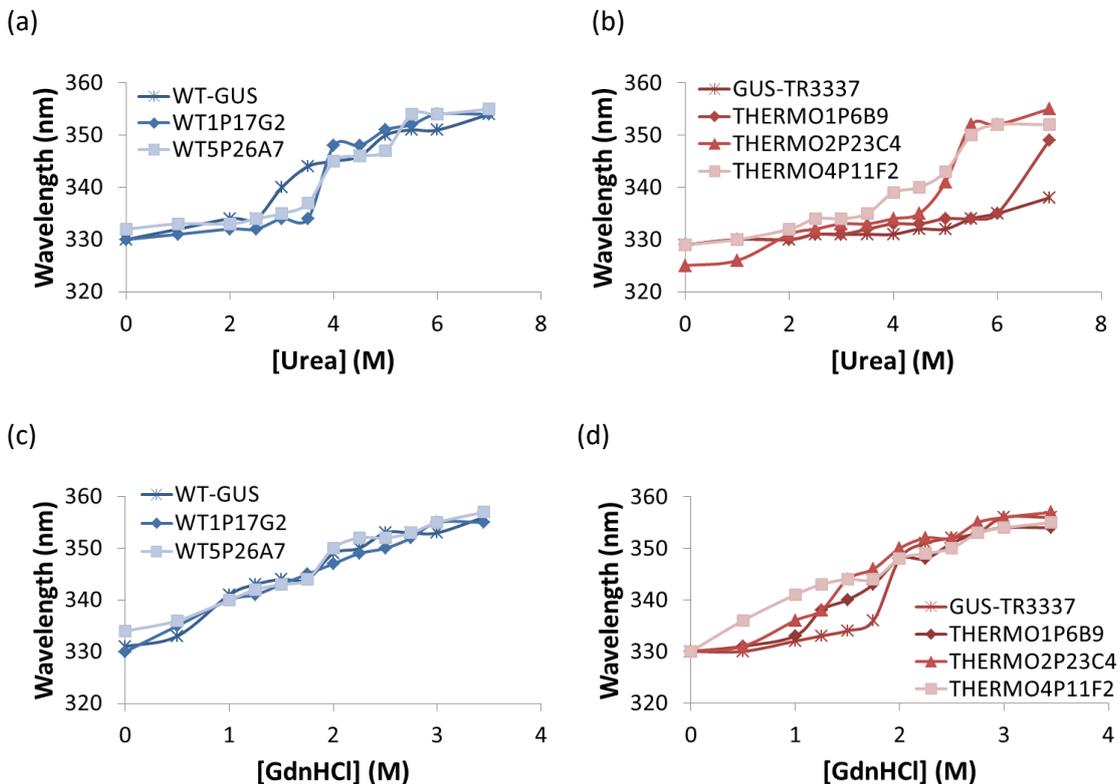


Figure 4.13 Chemical induced unfolding is shown by changes in λ_{max}

(a) Urea induced unfolding in GUS-WT* (b) Urea induced unfolding in GUS-TR3337* (c) GdnHCl induced in GUS-WT* (d) GdnHCl induced in GUS-TR3337.

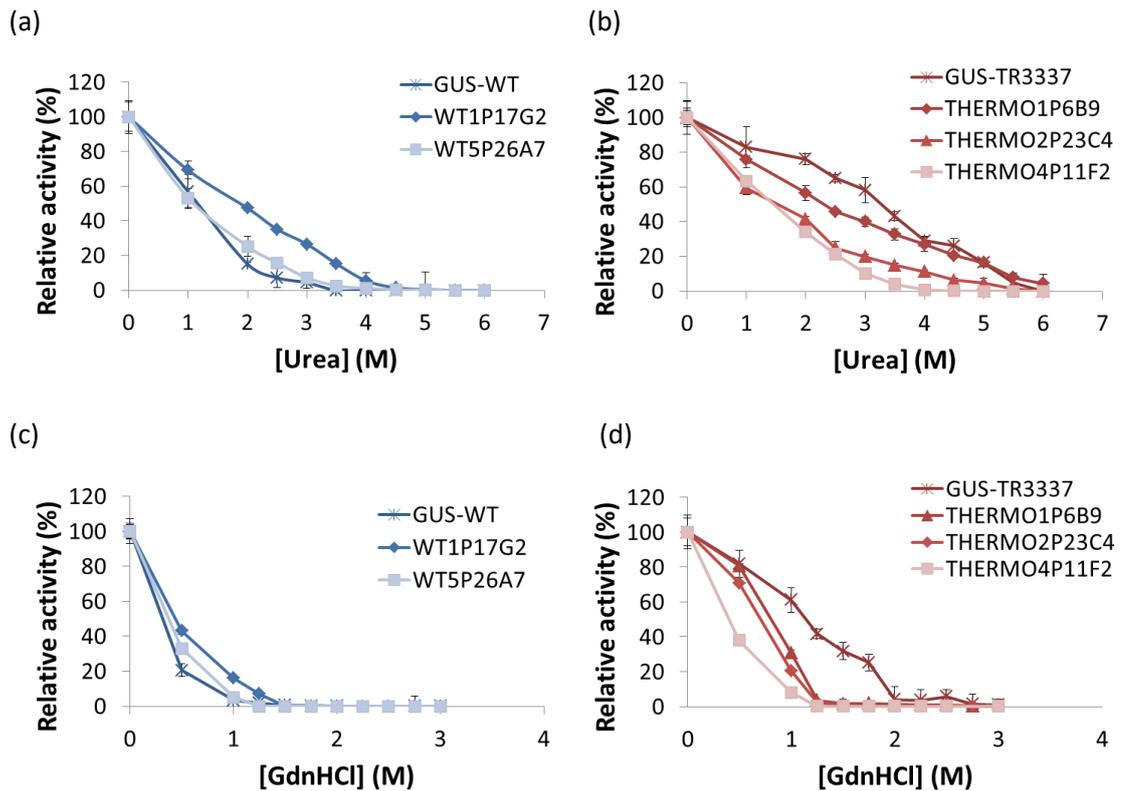


Figure 4.14 Impact of the presence of different concentrations of urea and GdnHCl on the glucosidase activity of β -GUS variants

With 4 M urea and 1 M GdnHCl, the λ_{\max} of GUS-WT was shifted to 348 nm and 341 nm, respectively, indicating the complete unfolding of the protein (Figure 4.13a & c). In contrast, the fluorescence emission spectra of GUS-TR3337 showed no significant change in λ_{\max} up to 4.5 M urea and 1.75 M GdnHCl, implying that urea concentrations as high as 4.5 M and GdnHCl concentrations as high as 1.75 M do not significantly unfold GUS-TR3337 (Figure 4.13b & d). These observations are well supported by the effect of urea upon the catalytic activity of GUS-WT and GUS-TR3337. As presented in Figure 4.14, glucosidase activity of GUS-WT was completely lost in 3.5 M urea and 1.25 M GdnHCl while GUS-TR3337 was only 50% inactivated under the same conditions.

With increasing urea and GdnHCl concentrations the λ_{\max} of GUS-WT variants and GUS-TR3337 variants were red-shifted and their glucosidase activities were observed to be decreased gradually. GUS-WT variants had a similar intrinsic fluorescence emission spectrum and behaviour in the presence of chemical denaturants, compared to GUS-WT. Notably, the variant WT1P17G2 showed the highest tolerance to GdnHCl and urea, suggesting the T509A increased stability toward chemical denaturation (Figure 4.14a & c). The stability of the enzyme was first increased, creating the capacity to accept the functionally necessary

destabilizing mutations in later generations. In GUS-TR3337 library, the gradual increase in glucosidase activity with each generation was accompanied by decreases in stability. This was expected as stability was not included as a fitness parameter in the screening assay. Another explanation for this trade-off was that during evolution, enzymes adjust the strength and number of stabilizing interactions to optimize the balance between rigidity for stability and flexibility for activity at the selection temperature (Fontana *et al.* 1998). The K474Q and S561L mutations in THERMO2P23C4 reduce the stability. This observation was consistent with the observation that the λ_{max} of THERMO2P23C4 red-shifted to 352 nm as the urea concentration was increased to 5.5 M, indicating exposure of Trp residues upon complete unfolding (Subbarao & MacDonald 1994), while at the same urea concentration the λ_{max} of GUS-TR3337 and THERMO1P6B9 only appeared at 334 nm suggesting only a slight relaxation of enzyme structure. In addition, the THERMO2P23C4 mutant lost only 85% of its initial activity in the presence of 3.5 M urea (Figure 4.14b) while the GUS-TR3337 retained approximately 43% activity in the same concentration of urea.

Overall, GUS-TR3337 variants were more stable than GUS-WT variants. The λ_{max} of WT5P26A7 was red-shifted to 350 nm at 2 M GdnHCl and 5 M urea whereas λ_{max} of THERMO4P11F2 red-shifted to 350 nm at 2.5 M GdnHCl and 5.5 M urea (Figure 4.15a). In addition, THERMO4P11F2 had a relative activity that was 10% higher than WT5P26A7 at 0.5 M GdnHCl and urea concentration ranging from 1.0 to 2.5 M (Figure 4.15b).

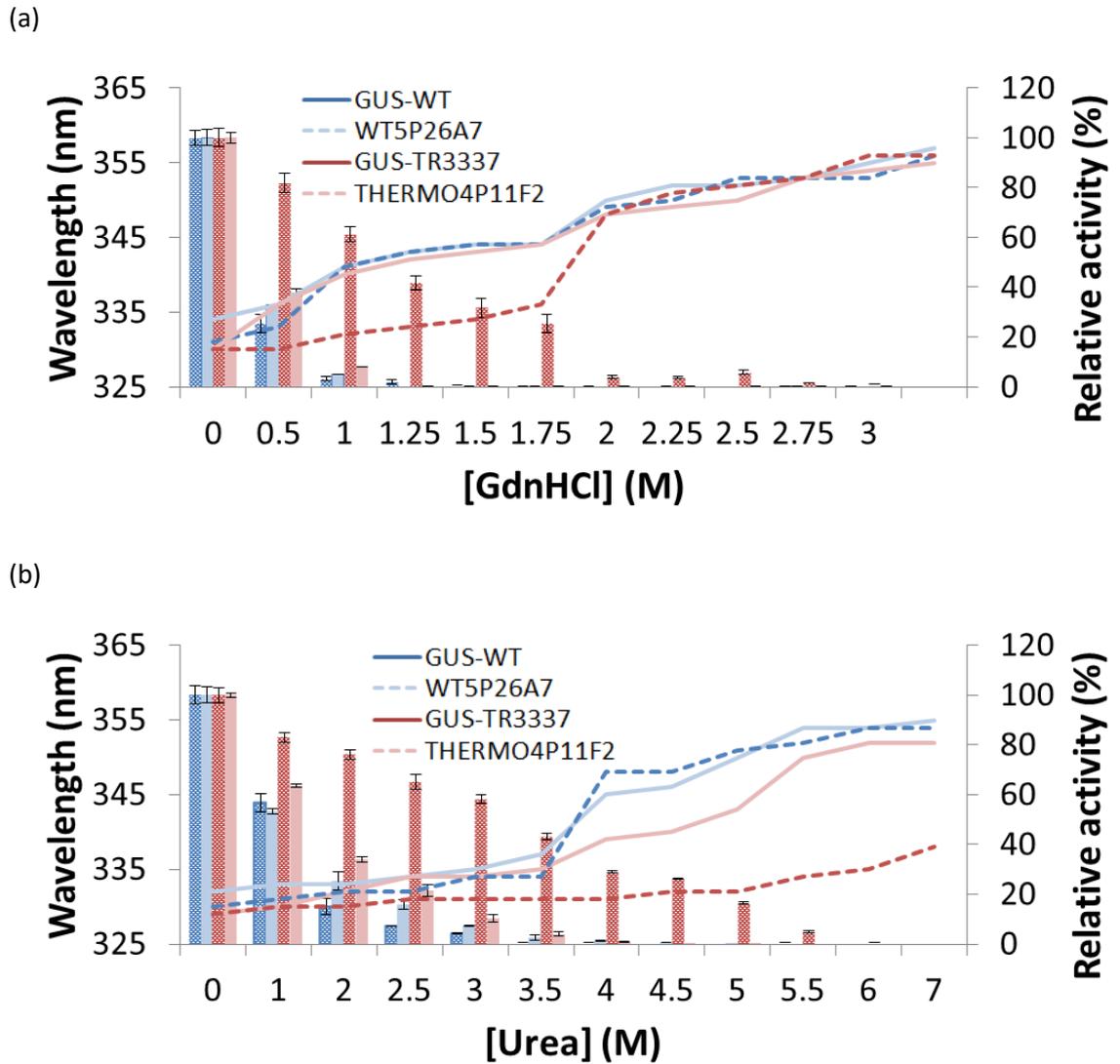


Figure 4.15 Intrinsic fluorescence spectra (line, left y-axis) and relative glucosidase activities (histogram, right y-axis) of GUS-WT, WT5P26A7, GUS-TR3337 and THERMO4P11F2 in the presence of different concentrations of chemical denaturants (a) GdnHCl (b) Urea.

4.4.5. Summary of observations

The variants obtained with directed evolution were studied to monitor the changes responsible for increased activity and to determine the changes in kinetic parameters that gave rise to increased activity. The results of these experiments are summarized in the Table 4.9 below. The level of expression and the solubility of variant proteins was found to be close to that of the starting proteins; these factors did not have an effect on activity, as observed in crude lysates. This investigation also looked at the kinetic parameters and stability; two factors that are often linked for proteins. The kinetic properties were as expected; drops in the K_m values were observed. There were clear trends in the experiments in which stability

was measured as a function of temperature. The variants that evolved from GUS-WT showed small increases in stabilities as the evolution proceeded. The opposite trend was observed with variants that were evolved from the GUS-TR3337 protein. However, the magnitudes of the changes in the melting temperatures were far greater for the variants that evolved from GUS-TR3337 compared with those that evolved from GUS-WT. The result obtained with the chemical denaturants followed a similar trend to those obtained with variations in temperature.

Table 4.9 Comparison of GUS-WT and GUS-TR3337 variants over all traits

The eight versions of GUS-WT and GUS-TR3337 are presented with the mutations in primary sequence indicated. The GUS-WT and GUS-TR3337 versions have been ranked for each trait, with the best denoted as 1. Shading has been used to highlight the significant differences.

Variant	64	279	350	469	474	480	493	498	509	516	532	550	557	559	561	566	568	597	Glucuronidase activity k_{cat}/K_m ($M^{-1}s^{-1}$)	Glucosidase activity k_{cat}/K_m ($M^{-1}s^{-1}$)	Thermal stability Melting temperature ($^{\circ}C$)	Chemical stability Residual activity at 2 M urea (%)
GUS-WT	A	E	D	Y	Q	T	Q	Q	T	M	M	N	S	G	L	N	K	P	540257	3	45	15
WT1P17G2	A	E	D	Y	Q	T	Q	Q	A	M	M	N	S	G	L	N	K	P	465440	4	59	47
WT3P24E11	A	E	D	Y	Q	T	Q	Q	A	M	M	N	S	G	L	N	Q	Q	7203	548	58	40
WT4P26F8	A	E	D	Y	Q	T	Q	Q	A	V	M	N	S	G	L	N	Q	Q	898	471	58	27
WT5P26A7	E	D	G	N	Q	T	Q	Q	A	M	M	N	P	G	L	N	Q	Q	2852	963	57	25
GUS-TR3337	A	E	D	Y	Q	T	R	Q	A	M	T	S	S	S	L	S	K	P	17608	1	75	76
THERMO1P6B9	A	E	D	Y	Q	T	R	Q	A	M	T	S	P	S	L	N	K	P	ND	8	ND	57
THERMO2P23C4	A	E	D	Y	K	T	R	Q	A	M	T	S	P	S	S	S	K	P	9969	29	71	41
THERMO3P24F7	A	E	D	Y	Q	T	R	Q	A	M	T	N	S	S	L	N	E	P	1171	724	61	36
THERMO4P11F2	A	E	D	Y	Q	M	R	K	A	M	T	N	S	S	L	N	E	P	374	1579	59	34

ND – Not determined

4.5. Mutation location and effect

At the time of writing the structure of GUS-WT had been reported (Wallace *et al.* 2010). Attempts to reproduce the GUS-WT crystals were made without success. In addition, crystal screens were employed to identify new crystal forms – again without success. The inability to reproduce published crystallisation procedures has been encountered in the Ollis laboratory on previous occasions and has been attributed to a lack of detail in the published experimental conditions. A few mutant forms of both the GUS-WT and GUS-TR3337 were tested with commercial crystallisation screens – again without success. Details of these experiments will be given in Chapter 5 Section 5.4.

The effects of mutations have been interpreted in terms of the native β -GUS structure. The effects of changes in sequence on the structure were not known. In other studies, the effects of changes produced by directed evolution have been found to be relatively modest (Porter *et al.* 2015, Hawwa *et al.* 2009). However, the possibility that structural change may occur must be considered when interpreting the effects of sequence changes. The initial location of a mutated residue can be found in the native structure and in many cases this can reveal useful information and provide a basis for speculation of possible structural changes. However, given the scattered distribution of mutations throughout the protein, the specific role of every mutation cannot be assessed with absolute confidence. Only the location of important mutations within the best mutants isolated from GUS-WT (WT5P26A7) and GUS-TR3337 (THERMO4P11F2) libraries, will be discussed in the following section. This preliminary analysis has led to several hypotheses relating to residues responsible for the activity and stability.

4.5.1. WT5P26A7 variant

An attempt was made to understand the effects of changes on the activity of the variant proteins. Apart from the location of the changes in the native structure, it was thought that some insight could be obtained from the calculated values of the solvent accessible surface area (SASA). Figure 4.16 shows the SASA calculated for the residues in the GUS-WT structure. The residues that were changed in the five rounds of directed evolution are shown on this figure. The SASA values for the changed residues are also given in Table 4.10 as in the % of the SASA found for that type of residue in a Gly-X-Gly tripeptide (Miller *et al.* 1987, Lee & Richards 1971).

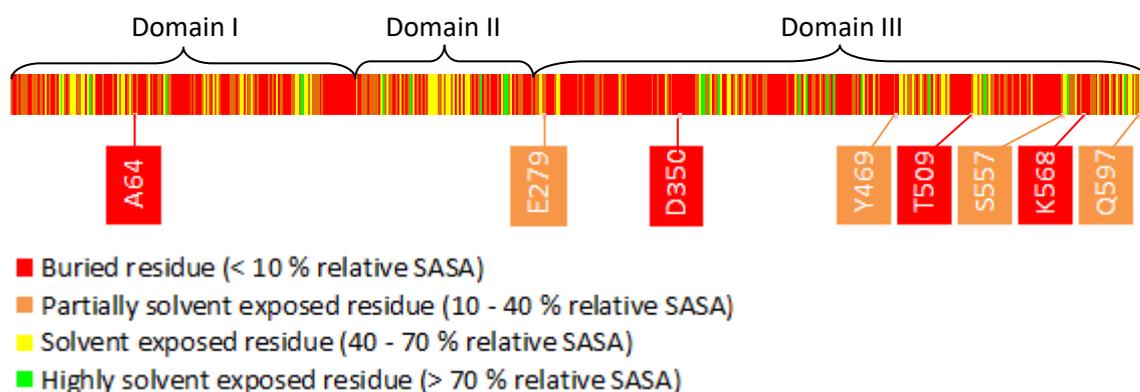


Figure 4.16 Predicted relative SASA of mutation sites in WT5P26A7

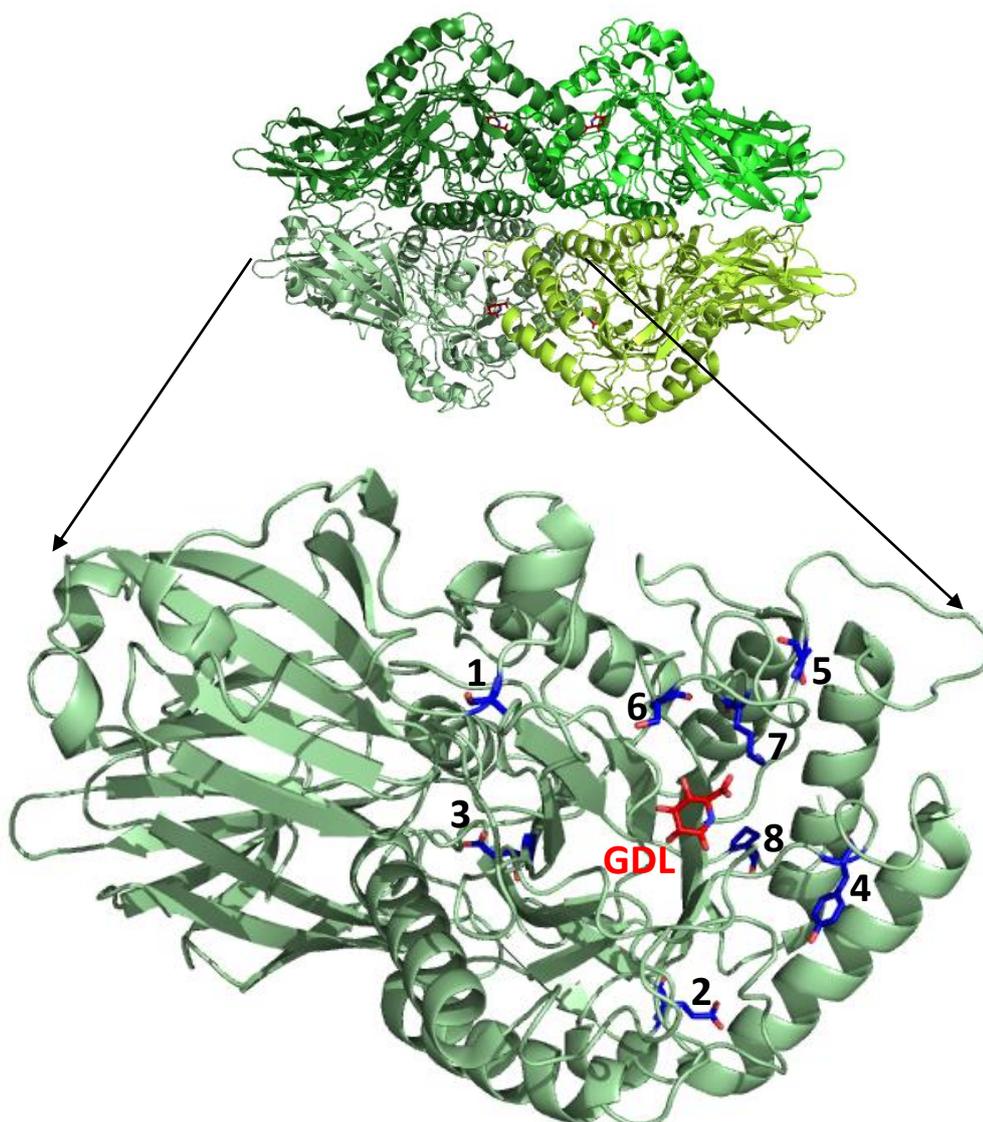
Out of 8 mutations in WT5P26A7, 4 residues (E279, Y469, S557 and Q597 – as in the native sequence) are partially exposed to solvent and 4 residues (A64, D350, T509 and K568)

are buried (Table 4.10). Of these, T509A, S557P and K568Q are in the active site (Figure 4.17) and were found in the early round of evolution while A64E, E279D, D350G and Y469N changes were found in the fourth or fifth generation variants. The change at position 597 did not occur in isolation (it only occurs with other changes) so that its effects are difficult to predict.

Table 4.10 Predicted SASA and their relative values of mutation sites in WT5P26A7

Residue	SASA (\AA^2)	Relative (%)
A64	11.44	9.5
E279	43.93	23.3
D350	0.00	0.0
Y469	67.93	28.1
T509	0.46	0.3
S557	27.17	21.2
K568	1.29	0.6
Q597	20.31	13.1

(a)



(b)

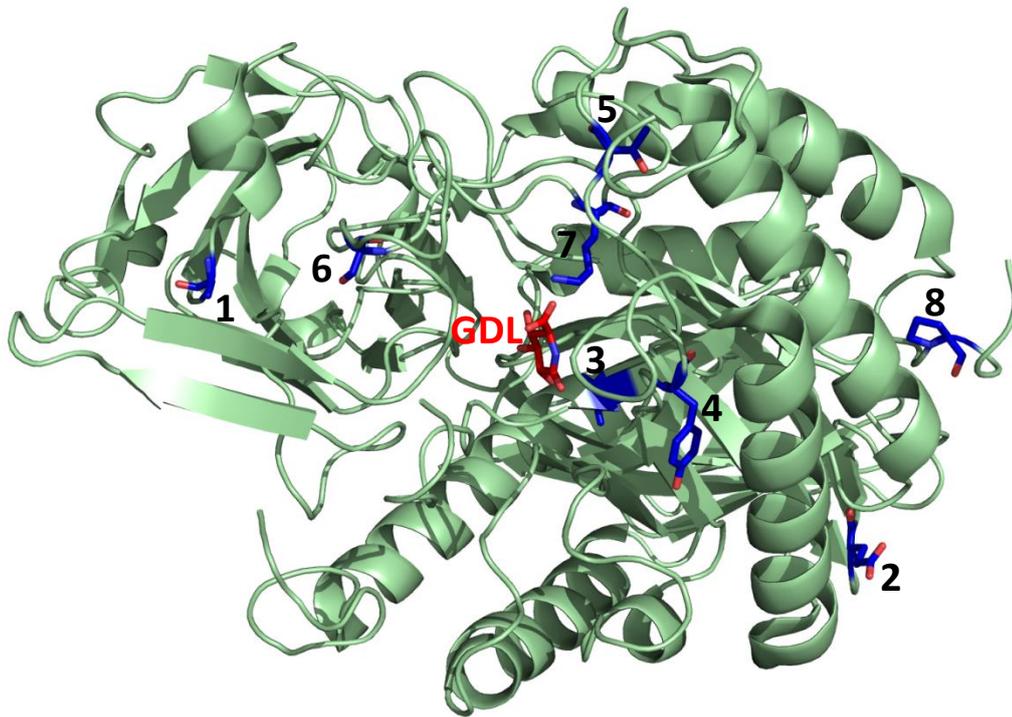
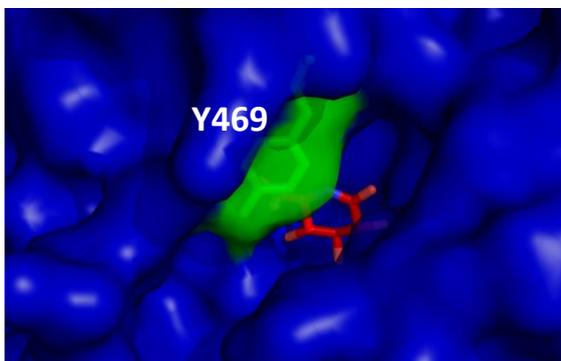


Figure 4.17 Mutation locations of WT5P26A7

The crystal structure of the GUS tetramer (Wallace *et al.* 2010) is illustrated at the top. One monomer is drawn (a) with each mutation site labelled: 1; A64, 2; E279, 3; D350, 4; Y469, 5; T509, 6; S557, 7; K568, 8; P597. The GDL molecule is shown in red. This diagram is also shown viewed from the right side (b).

Y469 is one of three tyrosine residues (Y472 and Y468) that occluded the entrance of the active site pocket (Wallace *et al.* 2010). The change brought about by the Y469N mutation could widen the substrate entrance and facilitate easier access to the active site (Figure 4.18).

(a)



(b)

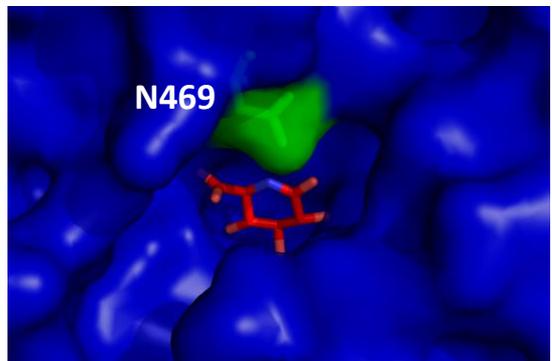


Figure 4.18 Surface model of the active sites of GUS-WT and WT5P26A7

The surface of WT5P26A7 was calculated using Swiss PDB viewer. The GDL is overlaid in both structures to show the entrance of the active site. The residue Y469 and N469 are labelled. (a) GUS-WT (b) WT5P26A7.

In the GDL (an established low-affinity inhibitor)-bound wild-type crystal structure, the negatively charged C-6 carboxylate group of GDL was shown to interact with the positively charged residue K568, suggesting that K568 was involved in substrate recognition and binding (Matsumura & Ellington 2001). This is consistent with the results obtained in the present study since K568Q change gave a k_{cat}/K_m for *p*NP-glucoside that was increased 223-fold while activity with *p*NP-glucuronide was reduced 65-fold. The replacement of lysine 568 with an uncharged with a polar amino acid glutamine, allowing glucose to be better accommodated in the active site (Figure 4.19b). In addition, this lysine was also thought to be important for the quaternary structure of the protein (Matsumura *et al.* 1999). Mutations to the residue 568 have already been shown in experiments by Matsumura *et al.* (2001) to be deleterious to enzyme function in isolation. As a result, this change was only found in variants with the T509A mutation, whereby T509A introduced stability into the protein to compensate for loss of stability brought about by changes at position 568. The $T_{1/2}$ of WT1P17G2 was 14 °C higher compared with the parent. In addition, WT1P17G2 retained 27% of its activity in 3 M urea while all selected variants from the GUS-WT library lost virtually all of their activities when measured under the same conditions. These observations can explain why the change of residue 568 was found in GUS-TR3337. It was suggested that alanine residues are the best helix-forming residues (Argos *et al.* 1979). Facchiano *et al.* (1999) observed that helices of thermophilic proteins were generally more stable than those of mesophilic proteins. In our studies, the T509A mutation occurred in the first round of directed evolution while K568Q mutation was only identified later in the third round. Apart from providing higher stability, T509A was also the key residue involved in enhanced glucosidase activity. The two mobile loops (green and blue in Figure 4.19) above the active site, are connected through hydrogen bonds between T509, Q526, K567, K568 and K578. However, the hydrogen-bonding interactions between the hydroxyl group of T509 and residues K567, K578 and Q526 were lost upon T509A mutation, thus adding to the mobility of the active-site loops.

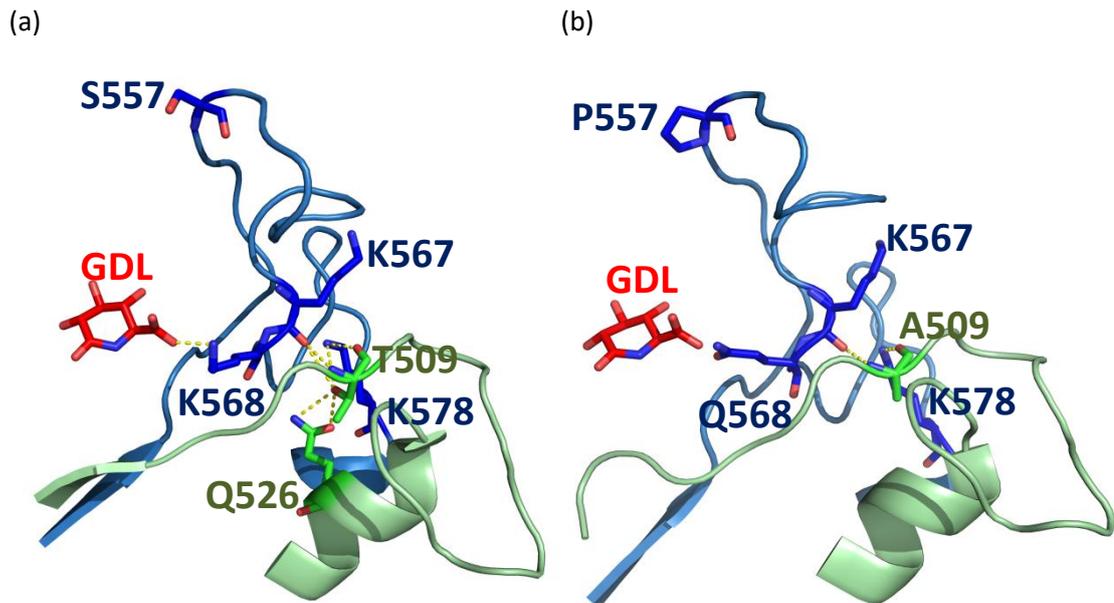


Figure 4.19 Connectivity between the loop (blue) bearing residues 557, 567, 568 and 578; and loop bearing residues 509 and 526 in WT5P26A7 and wild-type β -GUS

GDL is presented in red. The two loops (green and blue) are interconnected via T509. Atoms capable of hydrogen bonding are connected by yellow dash lines and intermolecular bonding distances are all less than 3 Å. (a) The environment around active-site loops is displayed from the crystal structure by Wallace *et al.* (2010) (b) The mutations were modelled with SWISS-MODEL using GUS-WT (PDB ID: 3K4D) as template.

The S557P change of a serine to a proline is unusual. Proline is unique amongst the amino acids in being cyclic. As a result, this amino acid can change the direction of the backbone. The introduction of the cyclic structure of proline forces a directional change to the loop. This change affects the local shape of the binding pocket and changes the hydrogen bonding ability. This change is most likely to affect the binding of the new substrate.

4.5.2. THERMO4P11F2 variant

This variant had five amino acid substitutions. S550, S566 and K568 were at least partially solvent accessible and two sites – T480 and Q498 – had less than 10% relative SASA (Figure 4.20) (Table 4.11). These were plotted onto the crystal structure of GUS to determine their position relative to the active site and other mutation sites (Figure 4.21).

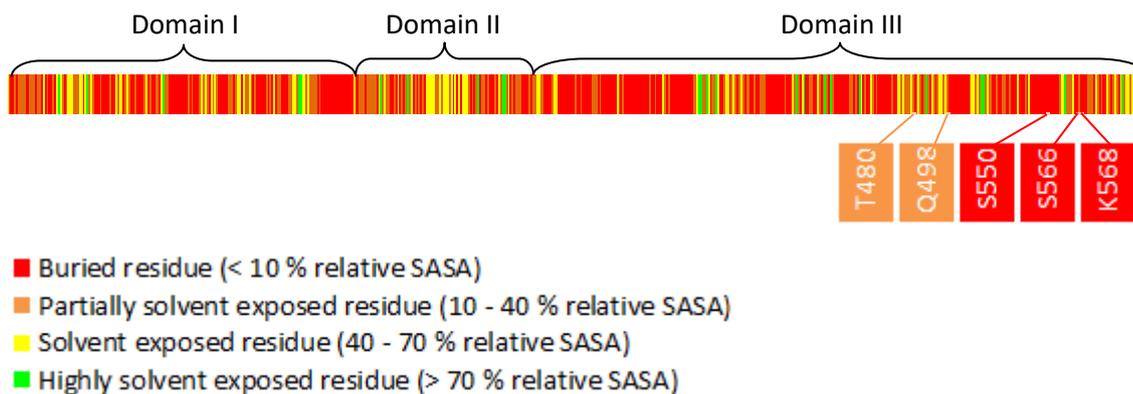


Figure 4.20 Relative SASA of mutation sites in THERMO4P11F2

Table 4.11 Predicted SASA and their relative values of mutation sites in THERMO4P11F2

Residue	SASA (\AA^2)	Relative (%)
T480	48.38	34.9
Q498	21.23	11.4
N550	1.69	0.01
N566	2.44	0.02
K568	1.29	0.01

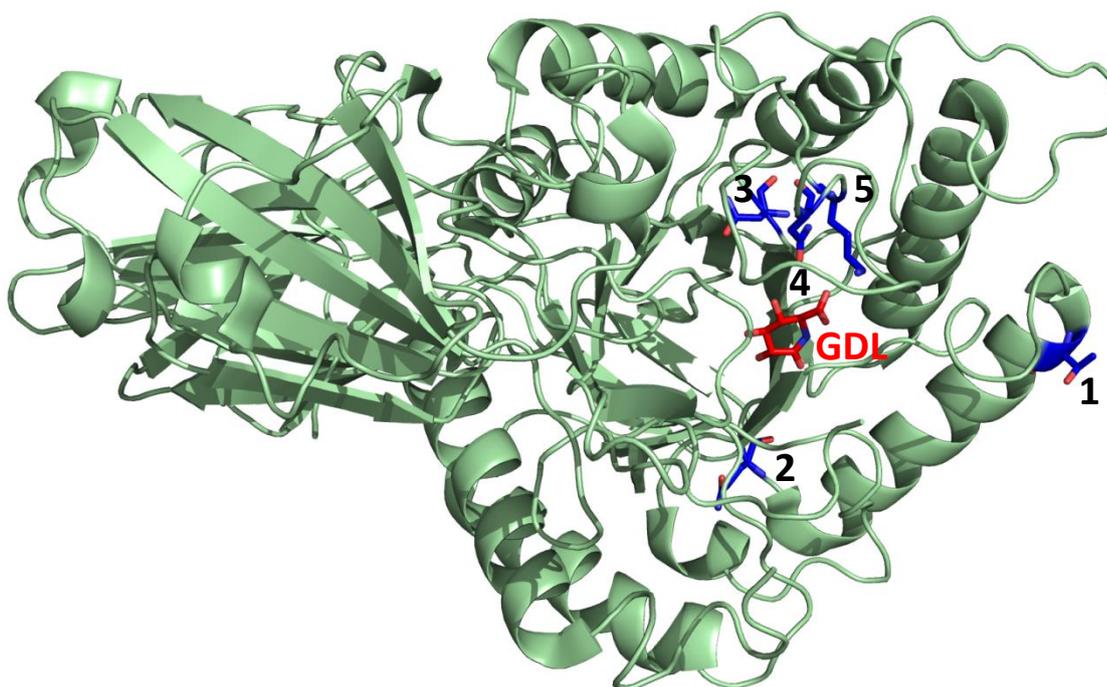


Figure 4.21 Mutation locations of THERMO4P11F2

The mutated residues have been colored blue to highlight their position relative to other mutations. Each mutation site labelled: 1; T480, 2; Q498, 3; N550, 4; N566, 5; K568. GDL is presented in red.

The two mutations on the THERMO4P11F2 surface (Q498K and T480M), displayed about a 2-fold higher k_{cat}/K_m as compared to THERMO3P24F7, which does not include these

two mutations. T480 was at the interface with another monomer – well removed from the active site. However, substituting it with methionine could involve in stabilizing the quaternary structure of the enzyme by independently increasing the affinity between the subunits.

Two key amino acids (S550 and S566) that conferred thermostability in GUS-TR3337 had reverted to wild type amino acids (N550 and N566). These two reversions accompany a significant decrease in $T_{1/2}$ of ~ 10 °C. N566 was known to be essential for the substrate binding due to the hydrogen bond with the C4 OH of GDL (Figure 4.22). In the predicted GUS-TR3337 structure, the hydroxyl side chain of S550 is hydrogen-bonded to the backbone of F551 and the hydrogen of G294, the backbone of S550 is hydrogen-bonded to the backbones of I570 and K568 (Figure 4.23a). The hydrogen bonds of G294, F551 and K567 with S550 are eradicated upon mutation of S550 to asparagine, resulting in a significant increase in mobility of the residues adjacent to 568 (Figure 4.23b). As discuss earlier, K568 was thought to be important for the substrate specificity as its positive charge resulted in an interaction with the negatively charged C6 carboxylate of *p*NP-glucuronide when bound to the enzyme. As a result, point mutations of K568 either by a change to glutamine as in WT5P26A7 or to a glutamic acid as in THERMO4P11F2 leads to an enzyme with higher glucosidase activity.

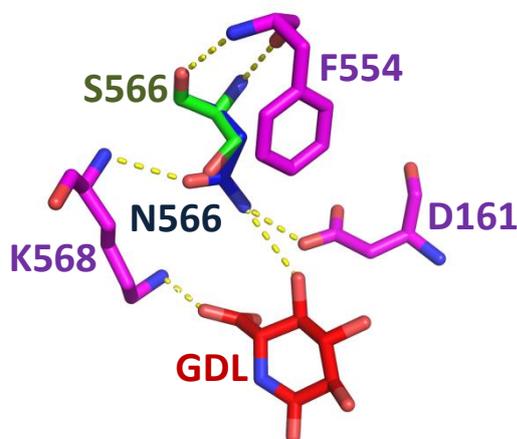


Figure 4.22 Residue 566 and GDL

S566 in GUS-TR3337, N566 in THERMO4P11F2 and the environment around these residues are shown in green, blue and magenta, respectively. They were modelled with SWISS-MODEL using GUS-WT as template. The GDL is presented in red. Atoms capable of hydrogen bonding are connected by yellow dashes.

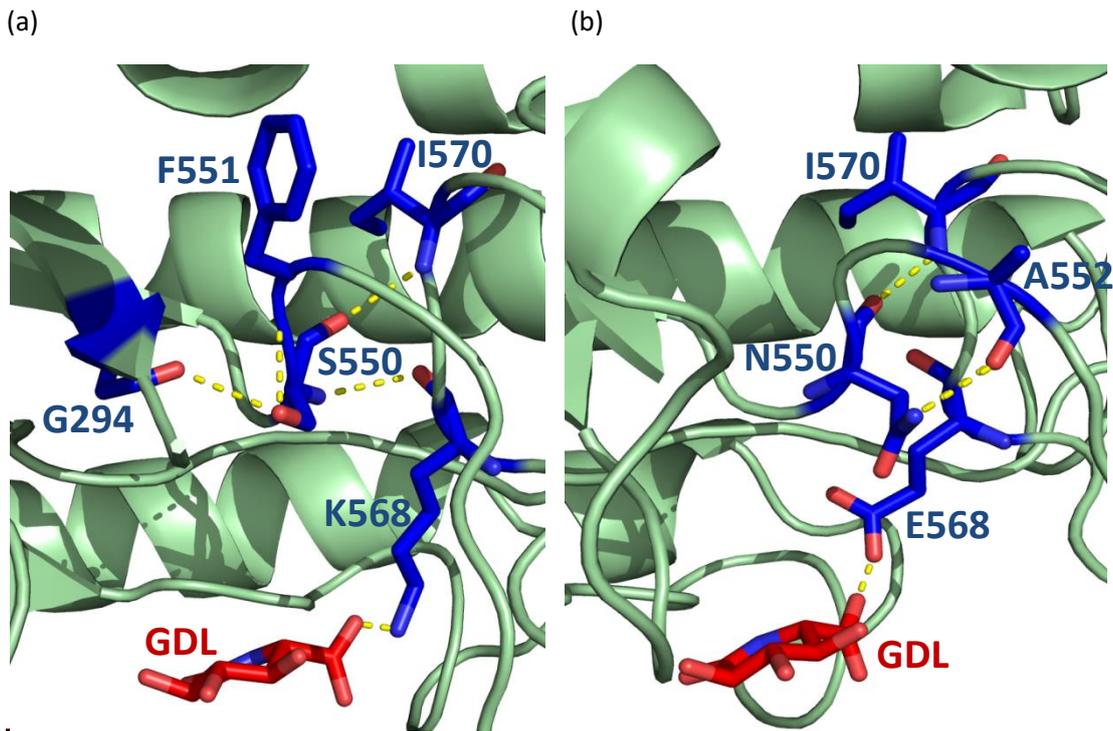


Figure 4.23 Hydrogen bonding network of residue 550

(a) The hydroxyl functionality of K568 in the predicted GUS-TR3337 structure is hydrogen bonded to the side chain of G294 and the backbone of F551, the backbone of S550 is hydrogen bonded to the backbone of I570 and K568 (b) Upon S550N mutation, the hydrogen bonding network illustrated in (a), except the hydrogen bond between S550 and I570, is eradicated. This results in increasing the flexibility of the loop where residue E568 resides.

4.6. Summary

Directed evolution was used to explore in some detail the relationship between evolvability and stability. This technique was applied to a native as well as a previously stabilized variant of GUS. It was found that evolution of the native protein towards a non-physiological substrate (i.e. increased glucosidase activity) resulted in changes that increase stability as well as activity. That is, increased stability was most easily achieved by mutations stabilized the overall protein structure so that destabilizing active site mutations could be made to increase activity. Conversely, the same experimental conditions resulted in a drop in the overall stability of the variant that initially had been evolved to increase stability. These results are consistent with those obtained by other workers who worked with different proteins. To the author's knowledge the present study is unique in that it monitored the evolution of similar proteins with different stabilities.

It should be noted that the activity of GUS-WT and GUS-TR3337 towards the glucosides had a very elevated K_m values – that is, substrate binding was poor in both cases. Evolution was carried out by monitoring changes in activity at concentrations well below these

K_m values. As expected, most of the shift in the activity of the variant enzymes was achieved by decreases in the K_m values of the enzymes. These brought about significant increases in activity as reflected in greatly increased values in k_{cat}/K_m , i.e. WT3P24E11 and THERMO3P24F7. Had more cycles of evolution been carried out, an increase in k_{cat} would be expected. As expected and as has been noted by others (Smith *et al.* 2011, Rowe *et al.* 2003, Geddie *et al.* 2004), the activity of GUS towards its physiological substrate decreased as activity towards the non-physiological substrate increased.

The best variant derived from the GUS-WT library was the WT5P26A7 variant that encompassed eight mutations – A64E, D279E, D350G, Y469N, T509A, S557P, K568Q and P597Q whereas the best variant derived from the GUS-TR3337 library was the THERMO4P11F2 variant that encompassed five mutations - T480M, Q498K, S550N, S566N and K568E. WT5P26A7 and THERMO4P11F2 had 307-fold and 1300-fold catalytic efficiency (k_{cat}/K_m) improvement towards β -glucoside, respectively. The fourth-generation mutant from the GUS-TR3337 library, THERMO4P11F2, displayed about 1.6-fold greater k_{cat}/K_m than the fifth generation mutant from the GUS-WT library, WTP26A7. Furthermore, GUS-TR3337 also showed greater tolerance for amino acid substitution than GUS-WT as mutants from the GUS-TR3337 library exhibited 10% more missense mutations. A higher level of mutational tolerance could afford a higher error-rate in libraries. It could also increase the chance of finding synergistic beneficial mutations that would not be identified individually.

Apart from significant changes to glucosidase activity, GUS-TR3337 variants were found to exhibit higher stabilities both thermally and chemically compared to WT5P26A7: a 2 °C increase in melting temperature was observed while the difference in denaturing urea concentration was 0.5 M in favour of THERMO4P11F2. In addition, the range of temperatures at which THERMO4P11F2 was active was broader than those of WT5P26A7; THERMO4P11F2 retained over 80% of relative activity in the range from 35 – 55 °C as opposed to the range from 45 – 55 °C for WT5P26A7.

Although the directed evolution approach was successfully used to enhance the glucosidase activity of GUS, the activity levels of purified proteins were significantly lower than those of cell lysate. The following chapter presents a series of experiments to investigate the factor behind their loss of enzyme activity. The kinetic parameters of glucosidase activities presented in this chapter, are described in more detail in Chapter 5.

4.7. References

- Argos, P., M. G. Rossmann, U. M. Grau, H. Zuber, G. Frank & J. D. Tratschin (1979) Thermal stability and protein structure. *Biochemistry*, 18, 5698-5703.
- Bloom, J. D., S. T. Labthavikul, C. R. Otey & F. H. Arnold (2006) Protein stability promotes evolvability. *Proc Natl Acad Sci U S A*, 103, 5869-74.
- Chica, R. A., N. Doucet & J. N. Pelletier (2005) Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Current Opinion in Biotechnology*, 16, 378-384.
- Elena, S. F. & R. Sanjuan (2008) The effect of genetic robustness on evolvability in digital organisms. *BMC Evol Biol*, 8, 284.
- Espinosa, O., K. Mitsopoulos, J. Hakas, F. Pearl & M. Zvelebil (2014) Deriving a Mutation Index of Carcinogenicity Using Protein Structure and Protein Interfaces. *PLoS ONE*, 9, e84598.
- Facchiano, A. M., G. Colonna & R. Ragone (1999) Helix-stabilizing factors and stabilization of thermophilic proteins: an X-ray based study. *Protein Eng*, 12, 893.
- Fontana, A., V. De Filippis, P. P. de Laureto, E. Scaramella & M. Zambonin. 1998. Rigidity of Thermophilic Enzymes. In *Progress in Biotechnology*, eds. F. J. P. J. L. I. A. Ballesteros & P. J. Halling, 277-294. Elsevier.
- Geddie, M. L. & I. Matsumura (2004) Rapid evolution of beta-glucuronidase specificity by saturation mutagenesis of an active site loop. *J Biol Chem*, 279, 26462-8.
- Hawwa, R., S. D. Larsen, K. Ratia & A. D. Mesecar (2009) Structure-based and random mutagenesis approaches increase the organophosphate-degrading activity of a phosphotriesterase homologue from *Deinococcus radiodurans*. *J Mol Biol*, 393, 36-57.
- Lee, B. & F. M. Richards (1971) The interpretation of protein structures: estimation of static accessibility. *J Mol Biol*, 55, 379-400.
- Matsumura, I. & A. D. Ellington (2001) In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. *J Mol Biol*, 305, 331-9.
- Matsumura, I., J. B. Wallingford, N. K. Surana, P. D. Vize & A. D. Ellington (1999) Directed evolution of the surface chemistry of the reporter enzyme β -glucuronidase. *Nat Biotech*, 17, 696-701.
- Miller, S., J. Janin, A. M. Lesk & C. Chothia (1987) Interior and surface of monomeric proteins. *J Mol Biol*, 196, 641-56.
- Moore, J. C. & F. H. Arnold (1996) Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nat Biotechnol*, 14, 458-67.
- Niwa, T., T. Tsuruoka, S. Inoue, Y. Naito & T. Koeda (1972) A new potent β -glucuronidase inhibitor, D-glucaro- δ -lactam derived from nojirimycin. *J Biochem*, 72, 207-11.
- Parikh, M. R. & I. Matsumura (2005) Site-saturation Mutagenesis is more Efficient than DNA Shuffling for the Directed Evolution of β -Fucosidase from β -Galactosidase. *Journal of Molecular Biology*, 352, 621-628.

- Porter, J. L., P. L. S. Boon, T. P. Murray, T. Huber, C. A. Collyer & D. L. Ollis (2015) Directed Evolution of New and Improved Enzyme Functions Using an Evolutionary Intermediate and Multidirectional Search. *ACS Chemical Biology*, 10, 611-621.
- Privalov, P. L. & N. N. Khechinashvili (1974) A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. *J Mol Biol*, 86, 665-84.
- Rowe, L. A., M. L. Geddie, O. B. Alexander & I. Matsumura (2003) A comparison of directed evolution approaches using the beta-glucuronidase model system. *J Mol Biol*, 332.
- Sachs, L. 1982. *Applied Statistics. A Handbook of Techniques*. New York: Springer-Verlag.
- Smith, W. S., J. R. Hale & C. Neylon (2011) Applying neutral drift to the directed molecular evolution of a β -glucuronidase into a β -galactosidase: Two different evolutionary pathways lead to the same variant. *BMC Research Notes*, 4, 1-10.
- Subbarao, N. & R. MacDonald (1994) Fluorescence studies of spectrin and its subunits. *Cell Motil Cytoskeleton*, 29, 72-81.
- Teilum, K., J. G. Olsen & B. B. Kragelund (2011) Protein stability, flexibility and function. *Biochim Biophys Acta*, 1814, 969-76.
- Tokuriki, N., F. Stricher, L. Serrano & D. S. Tawfik (2008) How Protein Stability and New Functions Trade Off. *PLoS Computational Biology*, 4, e1000002.
- Vieille, C. & G. J. Zeikus (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev*, 65, 1-43.
- Wallace, B. D., H. Wang, K. T. Lane, J. E. Scott, J. Orans, J. S. Koo, M. Venkatesh, C. Jobin, L. A. Yeh, S. Mani & M. R. Redinbo (2010) Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science*, 330, 831-5.
- Wolf-Watz, M., V. Thai, K. Henzler-Wildman, G. Hadjipavlou, E. Z. Eisenmesser & D. Kern (2004) Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat Struct Mol Biol*, 11, 945-949.
- Xiong, A. S., R. H. Peng, Z. M. Cheng, Y. Li, J. G. Liu, J. Zhuang, F. Gao, F. Xu, Y. S. Qiao, Z. Zhang, J. M. Chen & Q. H. Yao (2007) Concurrent mutations in six amino acids in beta-glucuronidase improve its thermostability. *Protein Eng Des Sel*, 20, 319-25.
- Zhao, H. & W. Zha (2006) In vitro 'sexual' evolution through the PCR-based staggered extension process (StEP). *Nat Protoc*, 1, 1865-71.

Chapter 5

Kinetic Analysis of GUS-WT and GUS-TR3337 Variants in Directed Evolution

5.1. Introduction

The previous chapter described directed evolution experiments that resulted in β -GUS variants that had enhanced glucosidase activity. However, the “enhanced activity” was obtained with crude lysates of cells expressing β -GUS and was not observed with purified enzymes. In fact, the activity of later generation variants was like that of the native protein. It appeared that the changes brought about by evolution did not increase the catalytic properties of the purified enzymes.

Initially, it was thought that that the disparity between results obtained with crude lysates and purified protein was due to incorrect variants being selected. That is the variants assayed in crude lysate were not those used to produce purified protein. The variants genes were sequenced (again) to verify that the correct mutants had been selected – they were correct.

Another possibility for the differences between the results obtained with crude lysate and purified protein was due to differences in protein concentration. In other words, the protein evolved so that variants were expressed at higher concentrations. That is, the enhanced activity of the variants was simply due to a more of the variant proteins being expressed. As noted in the previous chapter, the expression of native and variant proteins was very similar and it could not account for the different levels of catalytic activity. The results of these experiments can be found in Section 4.4.1.

The effects of temperature were considered as a possible explanation for the above anomalies. This was not thought likely, but it was also thought prudent to check. It was found that the optimal activities of the eight selected variants and parent strains were above 40 °C (Chapter 4, Section 4.4.3.) while the activities were measured at 25 °C. Thus, the assays were repeated at 45 °C (Figure 5.1b). At this temperature, all the purified variants exhibited increased activities, but the increases were small compared with the increased activities obtained with crude lysates.

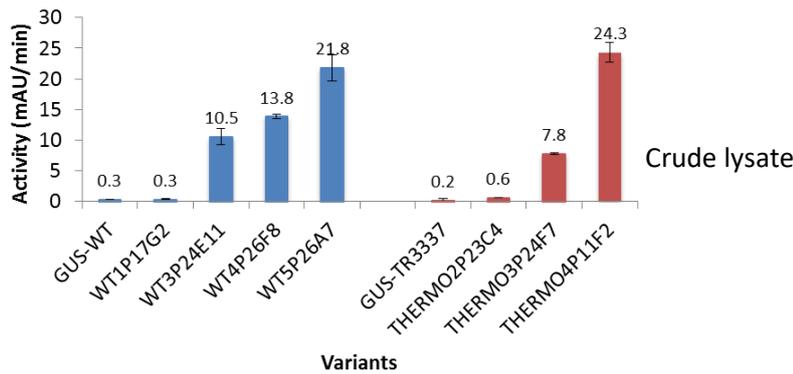
Some insight into the puzzling observations that are described above, was gained when it was realized that β -GUS had been evolved in the presence of a factor in the crude lysates: BugBuster™, the reagent used to break open the cells to give the crude lysates. This reagent was not used in the purification protocol. It should be noted that the manufacturer of BugBuster™ reagent would not disclose the ingredients of the product and attempts to identify the chemical composition by mass spectrometry were not initially successful. The reagent had an unknown composition. It was a proprietary mixture of non-ionic detergents that the manufacturer said was “capable of cell wall perforation without denaturing soluble protein”. Non-ionic detergents contain uncharged, hydrophilic head groups that consist of either polyoxyethylene moieties as in Tween, Triton, and the Brij series or glycosidic groups, i.e. octyl glucoside and dodecyl maltoside.

This chapter describes early experiments in which purified forms of evolved GUS were compared with the similarly purified native enzyme. The experiments showed that the BugBuster™ reagent was the cause of the anomalies in the directed evolution. The identification of the probable component of BugBuster™ reagent, octyl- β -D-thiogluco-side (OTG), that interacts with β -GUS is also described along with the characterisation of the native and mutant enzymes, in the presence and absence of this factor. The chapter concludes with the implications of the directed evolution in the presence of OTG – that is the evolution of an activating factor.

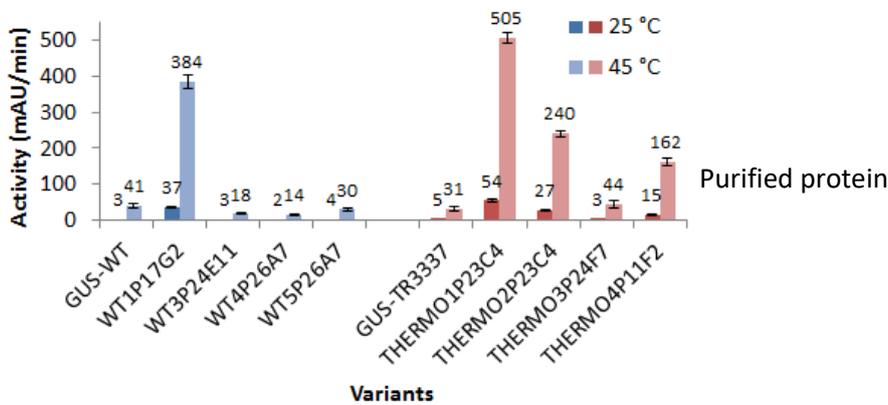
5.2. Loss of enzyme activity on purification

A selection of mutants of GUS-WT and GUS-TR3337 were chosen for purification. Mutants were chosen from the earlier rounds as well as the final round. By choosing variants from each round some idea of the progress made with evolution could be obtained. In most cases these variants had the highest activity for their generation. These variants along with GUS-WT and GUS-TR3337 were expressed and crude lysates used to measure their activities (Figure 5.1a). The results were as expected in that the activities increased as evolution proceeded. These native and variant proteins were then purified, as described in Chapter 2 (section 2.3.3.1.) and their activities measured again (Figure 5.1b). However, most of the GUS-WT and GUS-TR3337 variants displayed a dramatic loss of activities after purification, especially those in the later rounds of directed evolution (Figure 5.1b). The trend in the loss of activity was particularly disturbing – later generations appeared to have less activity than earlier generations. However, the expected trend in activity with the progress of evolution was obtained by the addition of BugBuster™ reagent to the purified protein (Figure 5.1c).

(a)



(b)



(c)

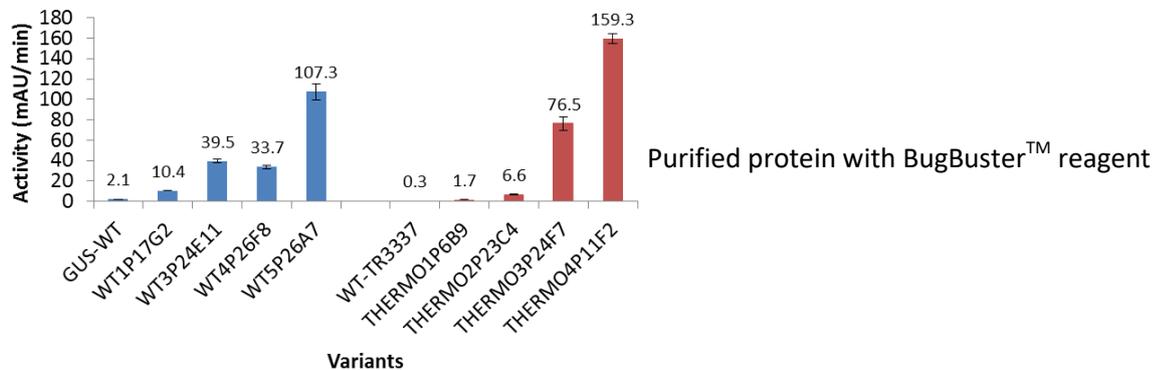


Figure 5.1 Activity data of variants with 0.8 mM pNP-glucoside substrate under the default assay conditions of 50 mM phosphate buffer pH 7.4 at different temperatures

(a) Activity of 100 μ L GUS-WT* and GUS-TR3337* crude lysates at 25 °C (b) Activity of 60 nM purified GUS-WT* and GUS-TR3337* at 25 °C and 45 °C (c) Activity of 60 nM purified GUS-WT* and GUS-TR3337* in the presence of 0.05 X BugBuster™ reagent at 25 °C.

5.3. Effects of 10 X BugBuster™ protein extraction reagent

It appeared that anomalous behaviour of variant enzymes was due to the presence of BugBuster™ reagent. In the screening procedure, the lysate was prepared with 1 X

BugBuster™ reagent and assayed for hydrolysis of *p*NP-glucoside; a non-physiological substrate. The question arose as to whether the enhanced activity stimulated by BugBuster™ reagent would also be observed with a glucuronide substrate, a physiological substrate?

The influence of BugBuster™ reagent upon glucuronidase and glucosidase activities was monitored by assaying GUS-WT and WT3P24E11 in the presence of 1 X BugBuster™ reagent. The results of these experiments are shown in Figure 5.2. As can be seen in this figure, glucosidase activity of GUS-WT was inhibited by BugBuster™ reagent while that of WT3P24E11 was stimulated by a factor of about 10. However, the glucuronide activities of GUS-WT and WT3P24E11 were inhibited. At this point it seemed that directed evolution for increased glucosidase activity had resulted in variants with increased activity in the presence of BugBuster™ reagent. It was clear that the presence of this reagent greatly affected the course of the evolution.

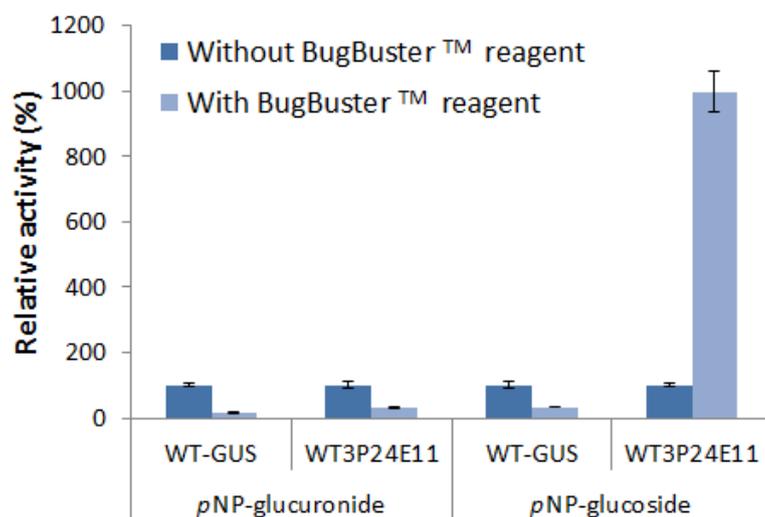


Figure 5.2 Effect of BugBuster™ reagent on glucuronidase and glucosidase activity of GUS-WT and WT3P24E11

Activity was calculated as relative (%) considering control as 100%. All assays were performed in triplicate. The data points and error bars indicate the average values and standard errors.

Bugbuster™ is not the only reagent on the market that is used to break open cells. It was thought that other commercial reagents probably contain similar ingredients. A test was made of the B-PER® reagent with the results presented below in Figure 5.3. It should be clear from this Figure that B-PER® reagent has a very similar effect to BugBuster™ reagent.

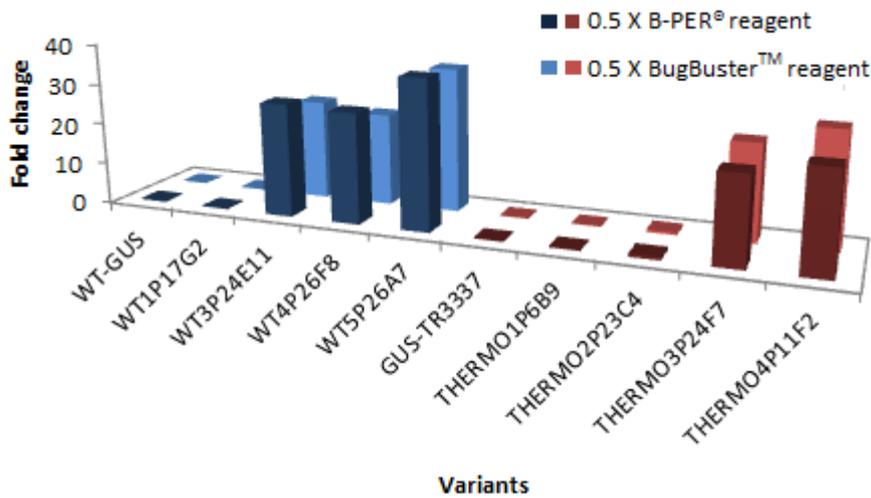


Figure 5.3 Effect of 0.5 X B-PER® and 0.5 X BugBuster™ reagentson glucosidase activity of GUS-WT and GUS-TR3337 variants

The BugBuster™ reagent probably exerted its effect in the evolution experiments by binding to GUS-WT and GUS-TR3337. If this were the case, then the activities of all the proteins would depend upon the concentration of BugBuster™ reagent. To confirm that this was the case, the glucosidase activity of GUS-WT, evolved variants along with GUS-TR3337 and its evolved variants were measured with increasing concentrations of BugBuster™ reagent. In these experiments the rate of hydrolysis of *p*NP-glucoside was used as the substrate. For GUS-WT variants, WT1P17G2, WT3P24E11, WT4P26F8 and WT5P26A7 were chosen; while for GUS-TR3337 variants, THERMO1P6B9, THERMO2P23C4 THERMO3P24F7 and THERMO4P11F2 were chosen, these being either the best or close to the best within their generation. The BugBuster™ reagent was titrated into the assay solution over the concentration range 0.005 to 0.5 X. The results of these experiments are shown below in Figures 5.4 & 5.5.

An inspection of Figure 5.4 reveals that GUS-WT and the first generation variant are inhibited by the BugBuster™ reagent. At the end of the titration, activity is only slightly above the error level of the assay. The later generations (3, 4 and 5) of GUS-WT all show enhanced activity that saturates at quite low concentrations of the BugBuster™ reagent. A similar trend is observed with GUS-TR3337 and its variants, except that this series of reactions includes a member of generation 2 – this enzyme appears to show very little effect due to BugBuster™ reagent (Figure 5.5). Overall, the evolution firstly abolishes inhibition and then evolves so that the BugBuster™ reagent appears to enhance activity. This way of describing these experiments is a little misleading. The BugBuster™ reagent is not “enhancing” the activity of variants, rather evolution results in the enhancement of β -GUS-BugBuster complex activity.

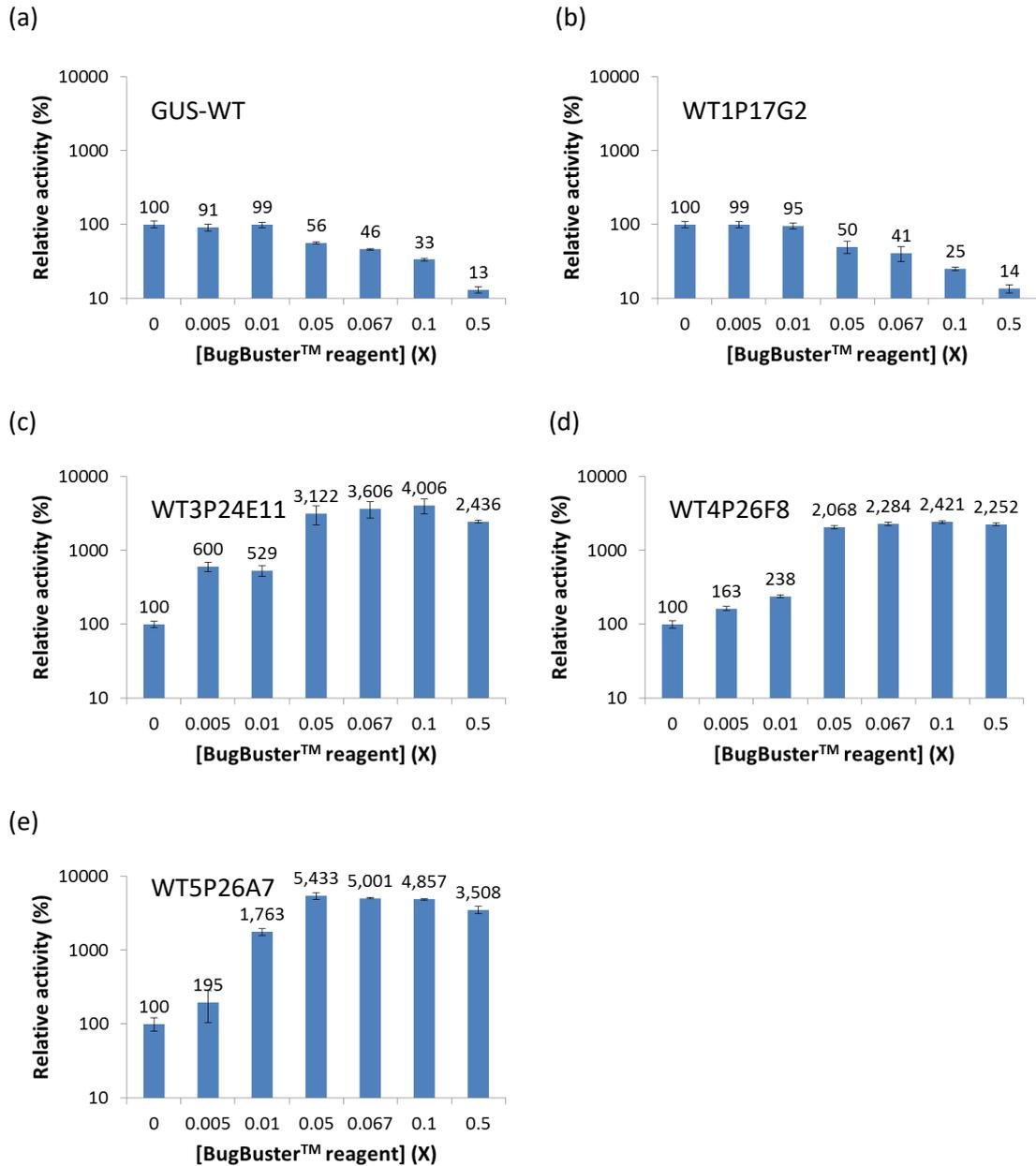


Figure 5.4 Effect of BugBuster™ reagent on glucosidase activity of GUS-WT mutants
Control activity refers to the activity of the individual enzyme in the absence of added BugBuster™ reagent (a) GUS-WT (b) WT1P17G2 (c) WT3P24E11 (d) WT4P26F8 (e) WT5P26A7.

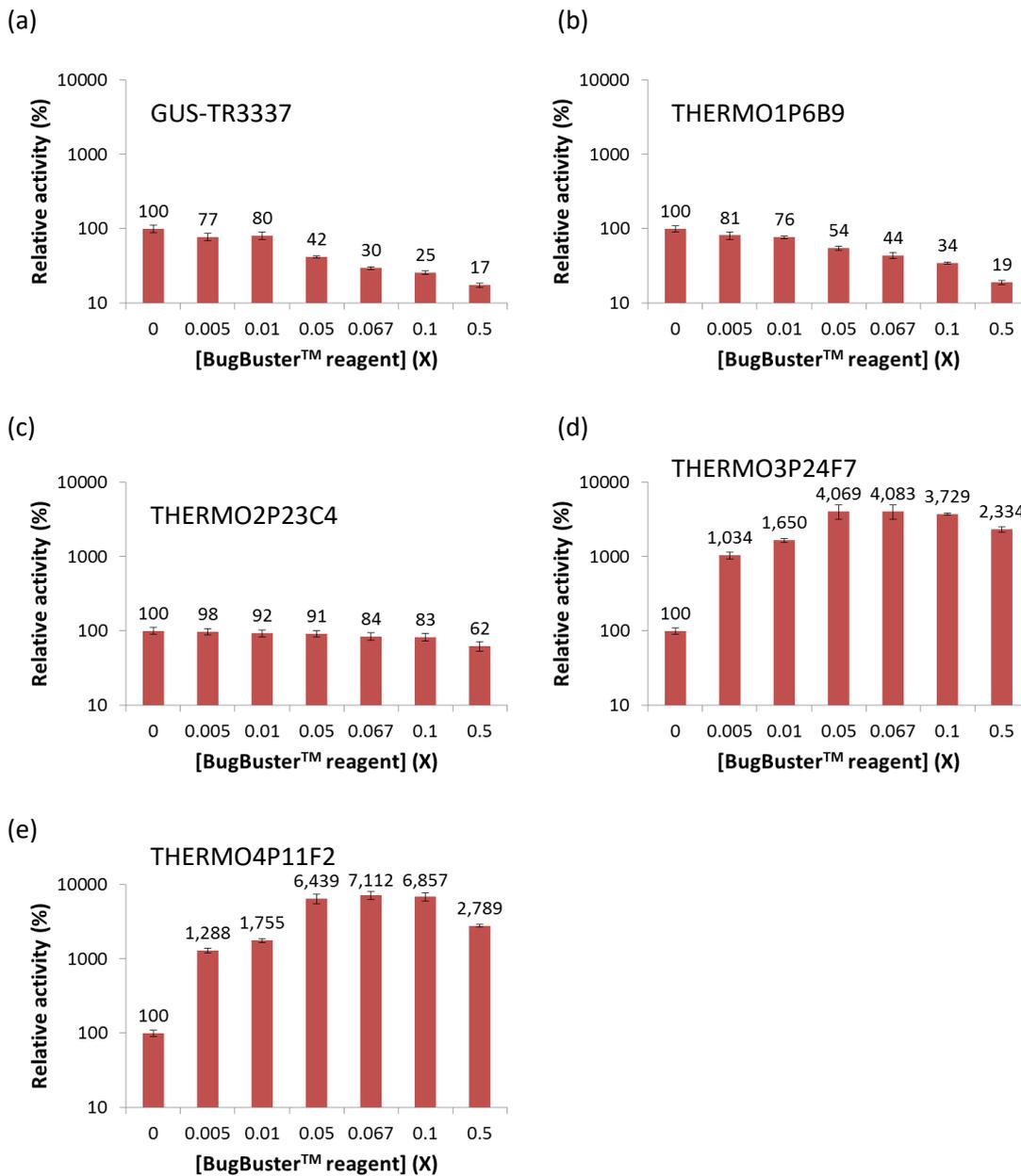


Figure 5.5 Effect of BugBuster™ reagent on glucosidase activity of GUS-TR3337 mutants
Control activity refers to the activity of the individual enzyme in the absence of added BugBuster™ reagent (a) GUS-TR3337 (b) THERMO1P6B9 (c) THERMO2P23C4 (d) THERMO3P24F7 (e) THERMO4P11F2.

To determine which components affect the glucosidase activity of β -GUS variants, it was decided that the effect of octyl- β -D-glucoside (OG), octyl- β -D-thioglucoside (OTG), Triton-X-100 (TX100) and Tween 20 would be investigated. Since the glucosidase activities of GUS-WT and its variants as well those of GUS-TR3337 and its variants were completely inhibited by 0.05% Tween 20 and 0.05% TX100 (data not shown), the effects of these two surfactants were not investigated further. In addition to OG and OTG, it was also thought useful to examine the

effects of sodium dodecyl sulphate (SDS) on the enzymes as it has some structure similarity to non-ionic detergents (Figure 5.6).

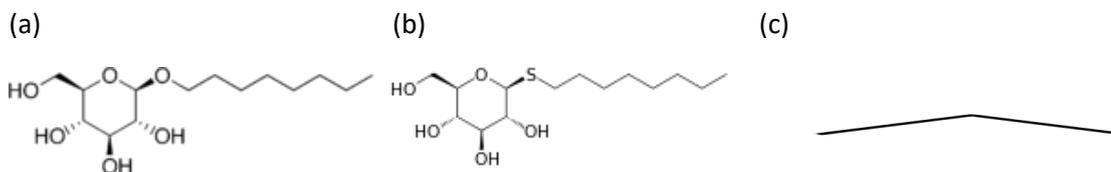


Figure 5.6 Chemical structures
(a) OG (b) OTG (c) SDS.

5.3.1. Effect of SDS

The results of the experiment with SDS are shown below in Figure 5.7. Like BugBuster™, SDS reagent inhibits the starting enzymes and early generation variants, but stimulates the activity of later generation variants. These later generation variants showed approximately 3- to 4-fold increased activity with 17.34 mM (0.5%) SDS but lost activity as the concentration was increased to 34.68 mM (1%) SDS. However, the enhanced glucosidase activity of mutants in the presence of SDS, while significant, was much smaller than observed with BugBuster™.

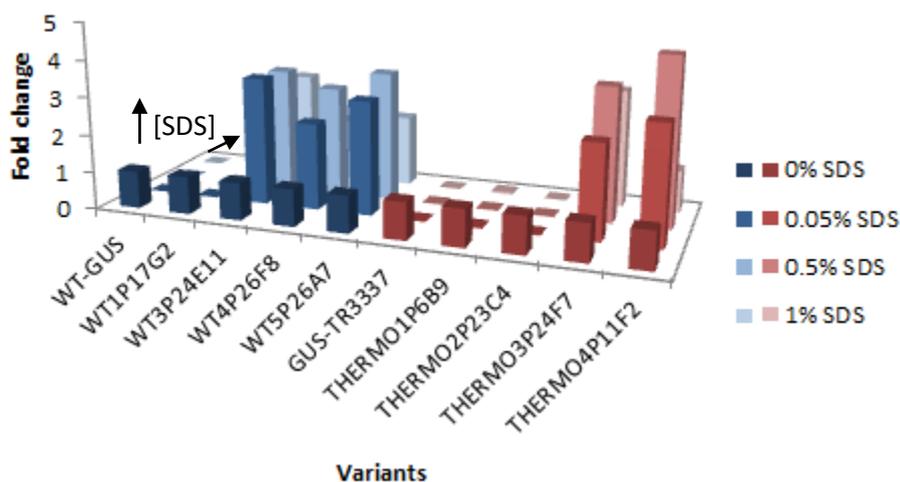


Figure 5.7 Effect of SDS concentrations on the glucosidase activity of the GUS-WT and GUS-TR3337 variants

5.3.2. Effect of OG

OG is one of the common examples of non-ionic detergents that is used for cell lysis. The influence of OG upon the glucosidase activity was monitored by assaying the enzyme in the presence of increasing concentrations of OG as shown in Figure 5.8. This Figure showed that 1.2 mM OG inhibited the activity of WT-GUS, WT1P17G2, GUS-TR3337 as well as THERMO1P6B9 by about 50%. In a similar manner to BugBuster™ reagent, the addition of OG up to a concentration of 5 mM resulted in increased glucosidase activity for WT3P24E11,

WT4P26F8, WT5P26A7, THERMO3P24F7 and THERMO4P11F2. Of these, the glucosidase activities of GUS-WT variants were greater than those of GUS-TR3337 variants. Furthermore, these improvements in activities of variant enzymes due to OG were relatively modest compared with those observed with BugBuster™.

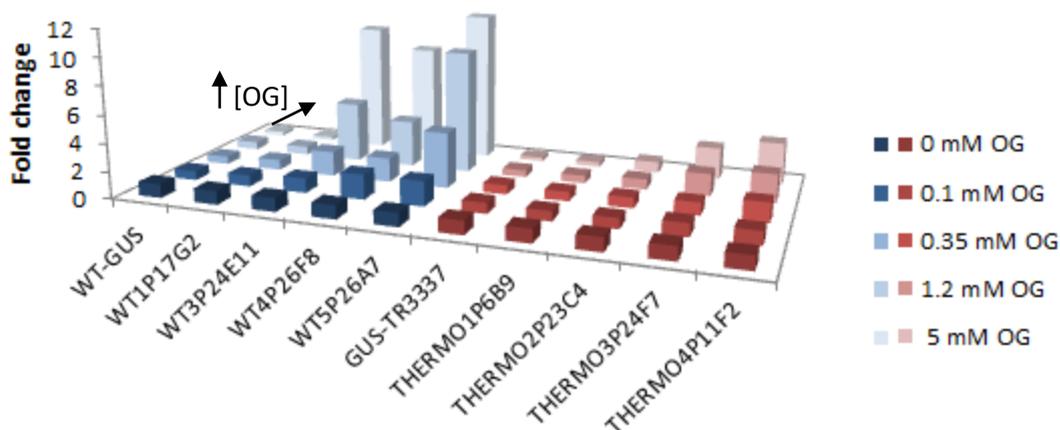


Figure 5.8 Glucosidase activity of GUS-WT and GUS-TR3337 mutants as a function of OG concentrations

The activity was assayed in the presence of various OG concentrations. The whole procedure was carried out at 25 °C.

It should be noted that it is probable that OG is a potential substrate of GUS and that the inhibition of starting and early generation mutants may just be due to competitive inhibition.

5.3.3. Effect of OTG

The effects of OTG on glucosidase activity was tested and the results shown below in Figure 5.9. A similar activity pattern was observed with OG, 0.5 mM OTG inhibited glycosidase activity of GUS-WT, WT1P17G2 and GUS-TR3337 by about 70%. However, maximum activity of WT3P24E11, WT4P26F8, WT5P26A7 and THERMO4P11F2 was achieved with the addition of at least 5 mM OTG. These improvements in the glucosidase activities are very similar to the increases observed by the addition of BugBuster™ to later generation variants. Therefore, it was concluded that OTG was a reasonable candidate for the component of BugBuster™ reagent that caused the major increase in glucosidase activity of later generation variants of GUS-WT and GUS-TR3337. The presence of OTG in the BugBuster™ reagent was later confirmed by nuclear magnetic resonance spectroscopy and mass spectrometry (Appendix D.1.).

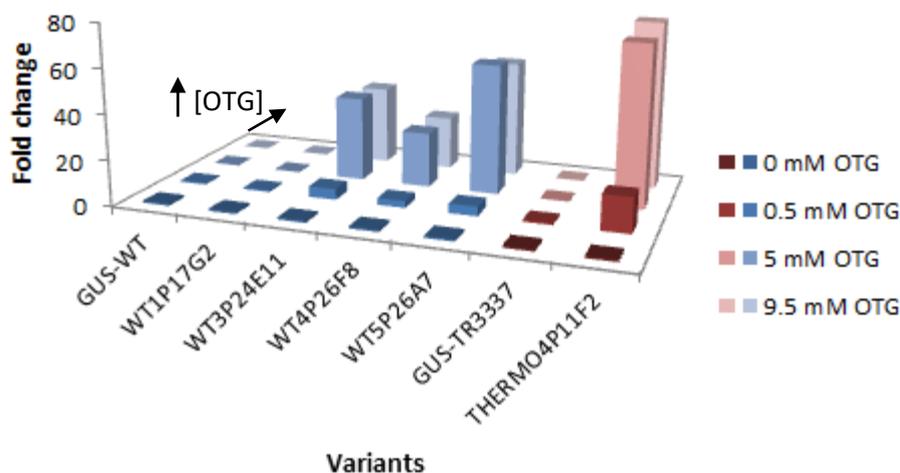


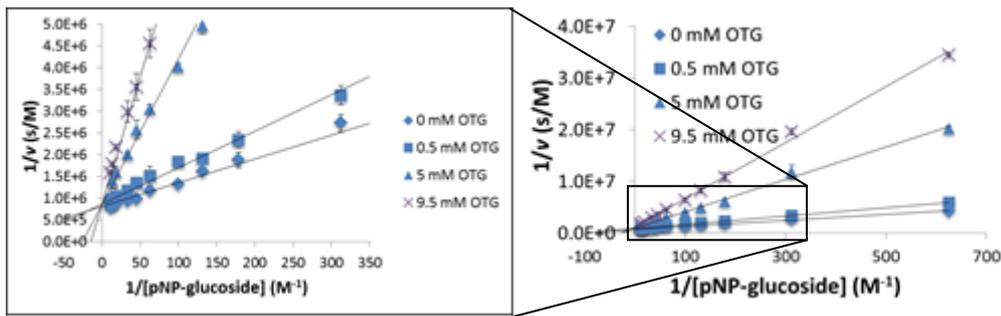
Figure 5.9 Effects of different OTG concentrations on GUS-WT and GUS-TR3337 mutants

5.3.3.1. Inhibitory and activating effects of OTG

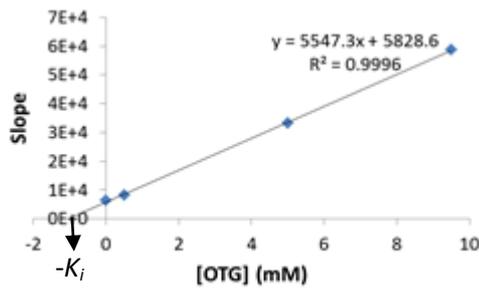
At this stage it was clear that OTG bound to GUS-WT and GUS-TR3337 and inhibited the activity of these enzymes. However, the nature of the inhibition was uncertain. Given the structural similarity of OTG and *p*NP-glucoside, it was possible that OTG was a competitive inhibitor. Consequently, OTG was tested for its mode of inhibition. The inhibition constant (K_i) is the concentration of inhibitor required to reduce the rate to half of the uninhibited value. The higher the K_i value the weaker the inhibitor binds to the enzyme and *vice versa*. The K_i variable is similar in concept to K_m and gives an indication of the binding constant for the substrate binding to the enzyme.

The K_i value was evaluated by using Microsoft Excel® (Chapter 2.3.4.10.). The graphical analysis with the Lineweaver-Bulk and the consequent secondary plots are given below in Figures 5.10. The figures 5.10b & f indicated that OTG binds to GUS-WT and GUS-TR3337 and gives rise to competitive inhibition with a K_i value that could be described as weak (1 mM). Competitive inhibition of OTG implies that OTG competes with a substrate (*p*NP-glucoside used in this study) to bind to the active site of β -GUS. The observed high K_m values given in Chapter 4 (Section 4.4.2, Table 4.7) for GUS-WT, WT1P17G2, GUS-TR3337 and THERMO1P6B9 were determined in the presence of 5 mM OTG while their true K_m values (i.e. with no OTG) are given in Table 5.1. In the case of the WT1P17G2 variant; it was also inhibited by OTG, but it was with a K_i of 1.5 mM (Figure 5.10d). It appears that the evolution has decreased the affinity between the enzyme and OTG – at least, as determined by completion studies with a simple substrate. It was possible that OTG still bound to the variant enzymes via the long aliphatic chain and that this could give rise to the enhanced activity in later generation variants.

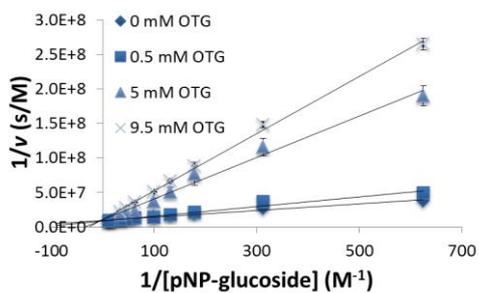
(a)



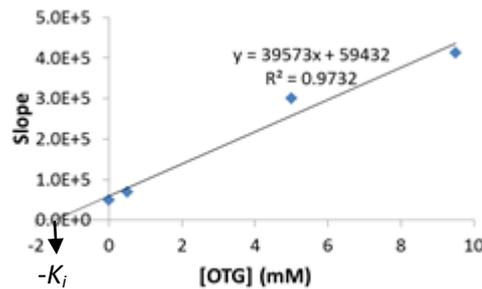
(b)



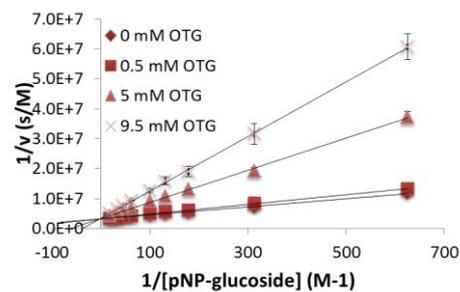
(c)



(d)



(e)



(f)

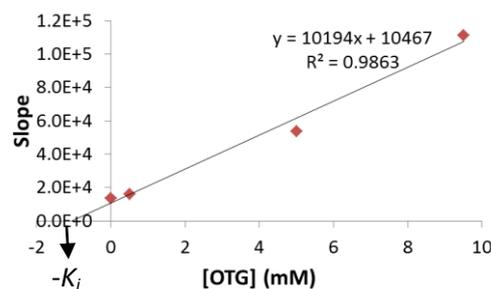


Figure 5.10 OTG inhibition of GUS-WT, WT1P17G2 and GUS-TR3337

(a) Double reciprocal Lineweaver-Burk plots of the GUS-WT kinetic assays for hydrolysis reaction of pNP-glucoside in the presence of different fixed concentrations of inhibitor compounds (b) the secondary plots with an abscissa intercept of $-K_i$ (c) Lineweaver-Burk plots for WT1P17G2 (d) Secondary plots for WT1P17G2 (e) Lineweaver-Burk plots for GUS-TR3337 (f) Secondary plots for GUS-TR3337.

By way of contrast with the plots shown above, the Michaelis-Menten plots for WT3P24E11, WT5P26A7 and THERMO4P11F2, shown below in Figure 5.11, indicate that there was no inhibition with OTG, but rather OTG acts as an activator. While WT3P24E11 was maximally activated at above 10 mM OTG, lower concentrations of WT5P26A7 and THERMO4P11F2 achieved maximum activation of glucosidase activity at OTG concentrations of 6mM (Figure 5.11). The kinetics data obtained for glucoside activity given in Chapter 4 (Section 4.4.2., Table 4.7), were measured in the presence of 5 mM OTG. A description is given in the next section of a more detailed investigation of the reaction kinetics of GUS-WT and its variants as well as of GUS-TR3337 and its variants.

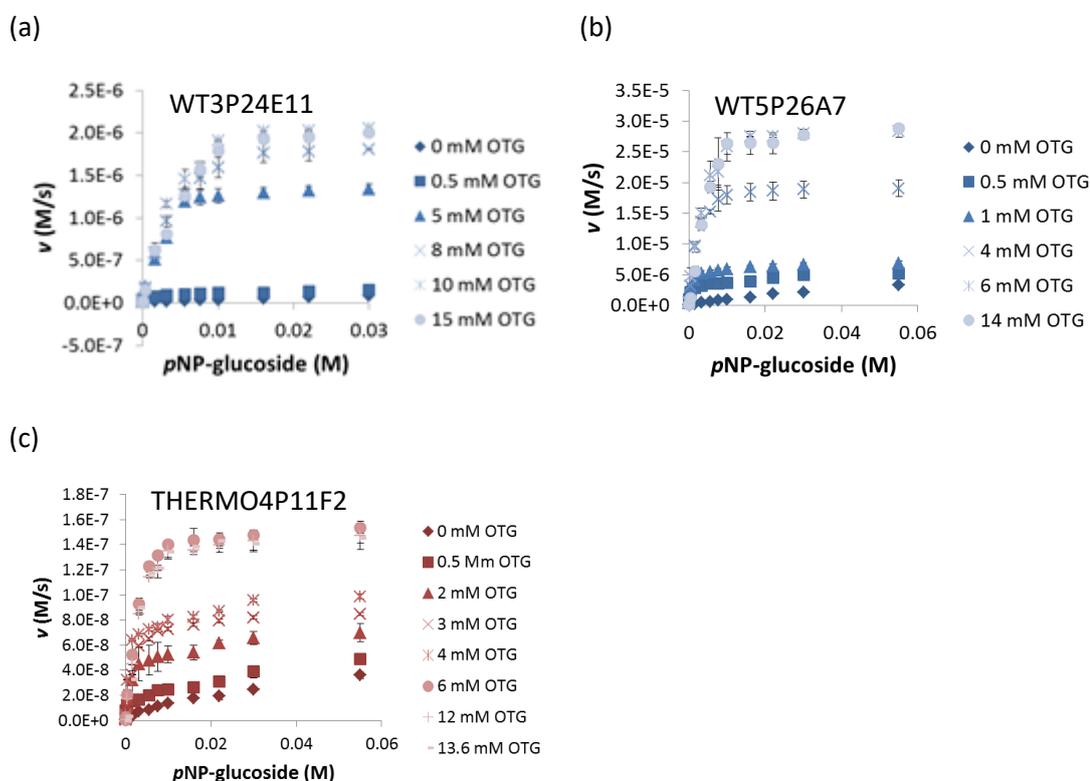


Figure 5.11 Michaelis-Menten plots for the activation of glucosidase activity by OTG as the variable substrate

(a) 2.33 μ M of WT3P24E11 (b) 1.47 μ M WT5P26A7 (c) 0.7 μ M THERMO4P11F2.

5.3.3.2. Kinetic analysis of GUS-WT and GUS-TR3337 variants

To better show the differences between the purified native enzymes (GUS-WT and GUS-TR3337) and later generation variants (WT5P26A7 and THERMO4P11F2), the enzymes were subjected to kinetic analysis with both pNP -glucuronide and pNP -glucoside substrates at pH 7.4 and 25 °C. The points to be made can be adequately illustrated with plots for the above enzymes, but a more complete collection of variant data sets can be found in Appendix D.2., D.3., D.4., D.5., D.6. & D.7. Figure 5.12 showed Michaelis-Menten plots for GUS-WT and

WT5P26A7 with the corresponding Eadie-Hofstee plots given in Figure 5.13. The corresponding plots for GUS-TR3337 and THERMO4P11F2 are given in Figure 5.14 & 5.15.

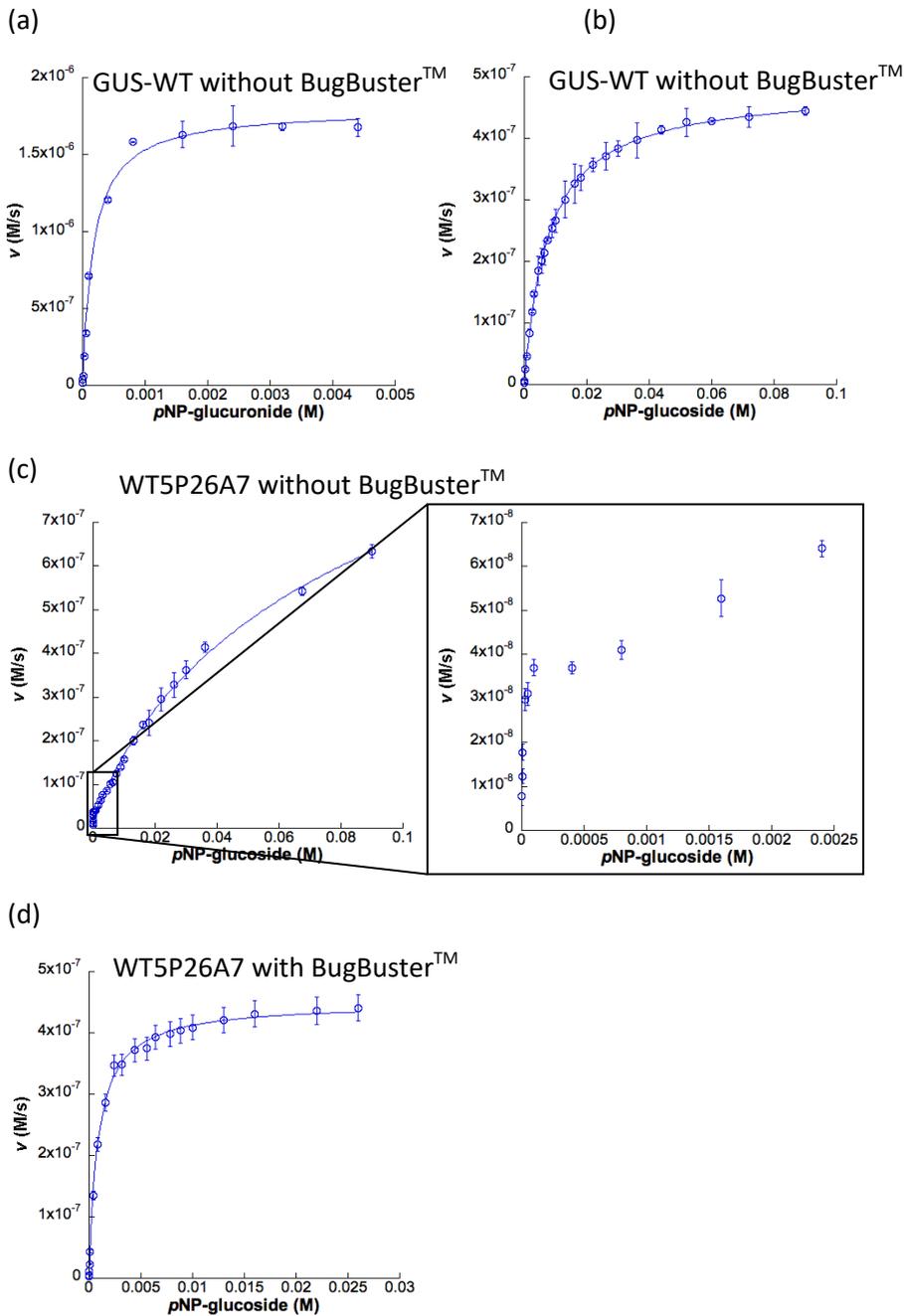


Figure 5.12 Michaelis-Menten plots of GUS-WT and WT5P26A7

The points show the measured values with their standard deviation. (a) Hydrolysis of p NP-glucuronide by GUS-WT (b) Hydrolysis of p NP-glucoside by GUS-WT (c) Hydrolysis of p NP-glucoside by WT5P26A7 (d) Hydrolysis of p NP-glucoside by WT5P26A7 in the presence of 5 mM OTG.

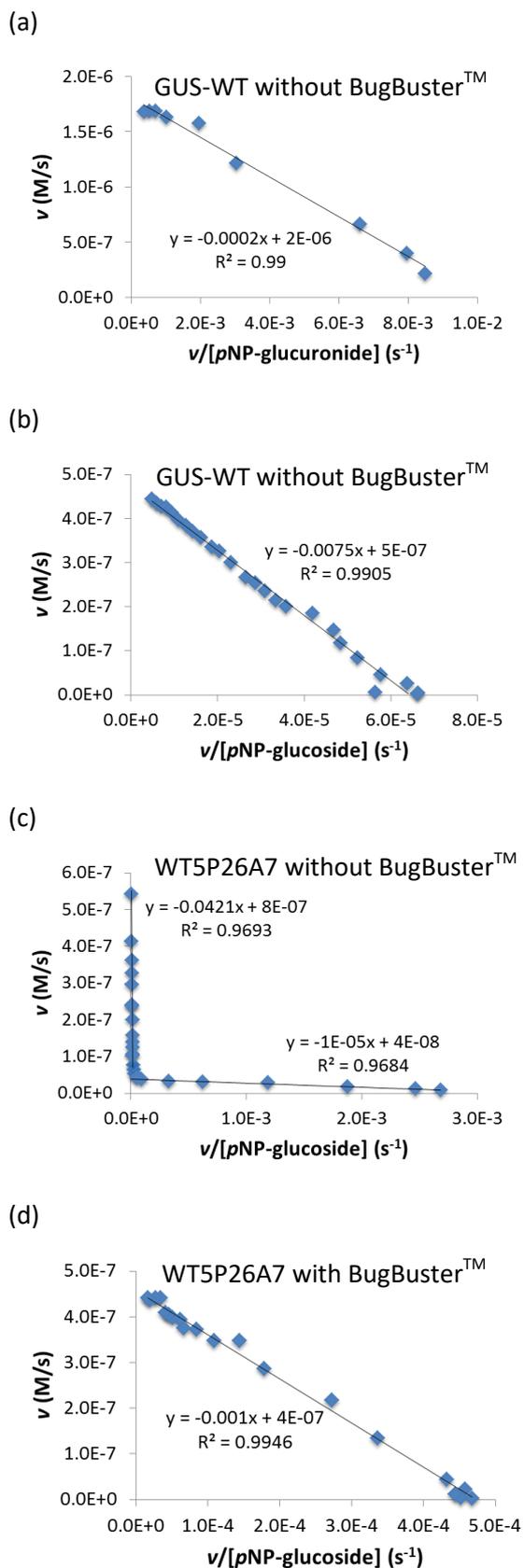


Figure 5.13 Eadie-Hofstee plots of GUS-WT and WT5P26A7

Each plot is the mean of triplicates. (a) Hydrolysis of *p*NP-glucuronide by GUS-WT (b) Hydrolysis of *p*NP-glucoside by GUS-WT (c) Hydrolysis of *p*NP-glucoside by WT5P26A7 (d) Hydrolysis of *p*NP-glucoside by WT5P26A7 in the presence of 5 mM OTG.

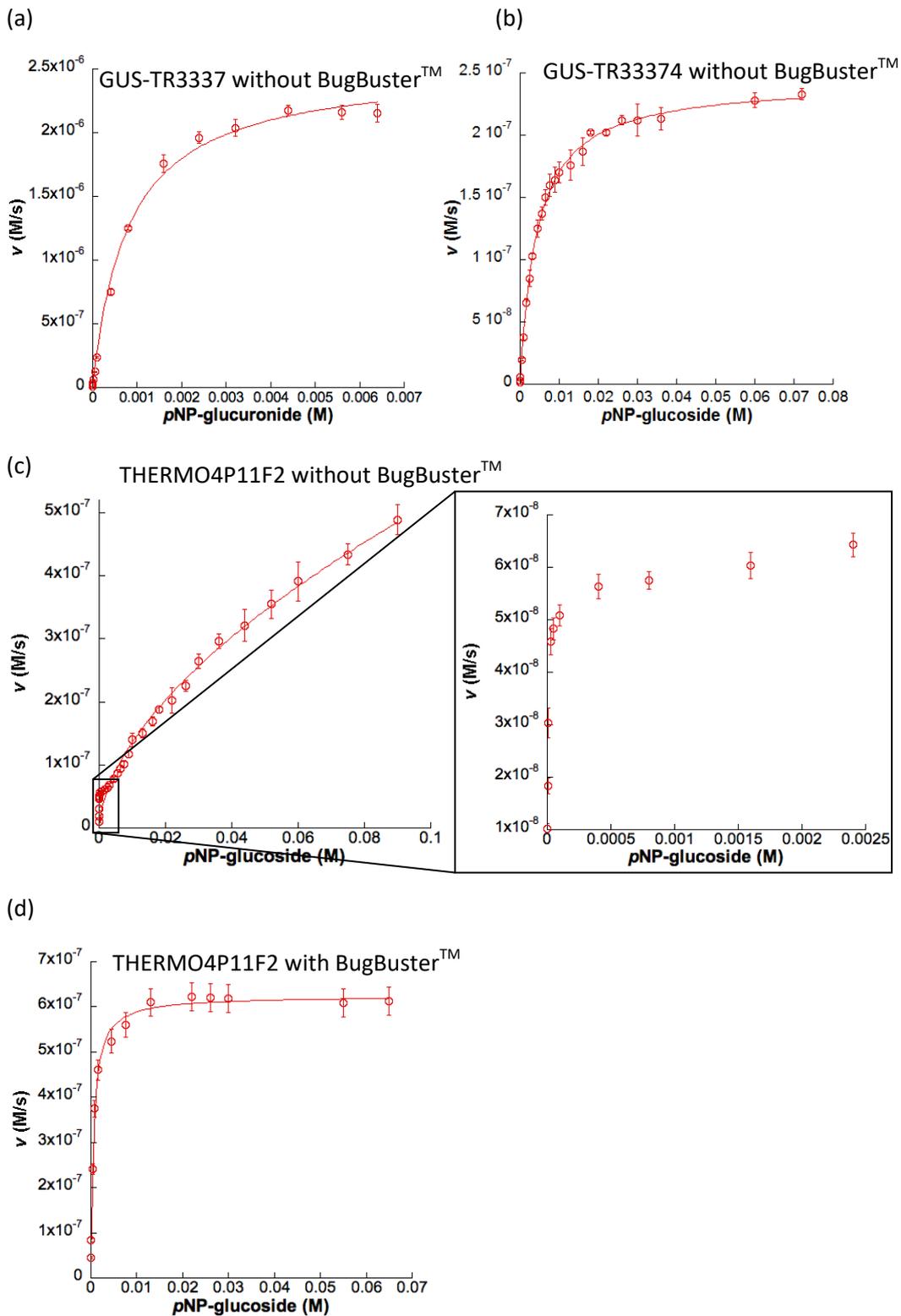


Figure 5.14 Michaelis-Menten plots of GUS-TR3337 and THERMO4P11F2

The points show the measured values with their standard deviation. (a) Hydrolysis of pNP -glucuronide by GUS-TR3337 (b) Hydrolysis of pNP -glucoside by GUS-TR3337 (c) Hydrolysis of pNP -glucoside by THERMO4P11F2 (d) Hydrolysis of pNP -glucoside by THERMO4P11F2 in the presence of 5 mM OTG.

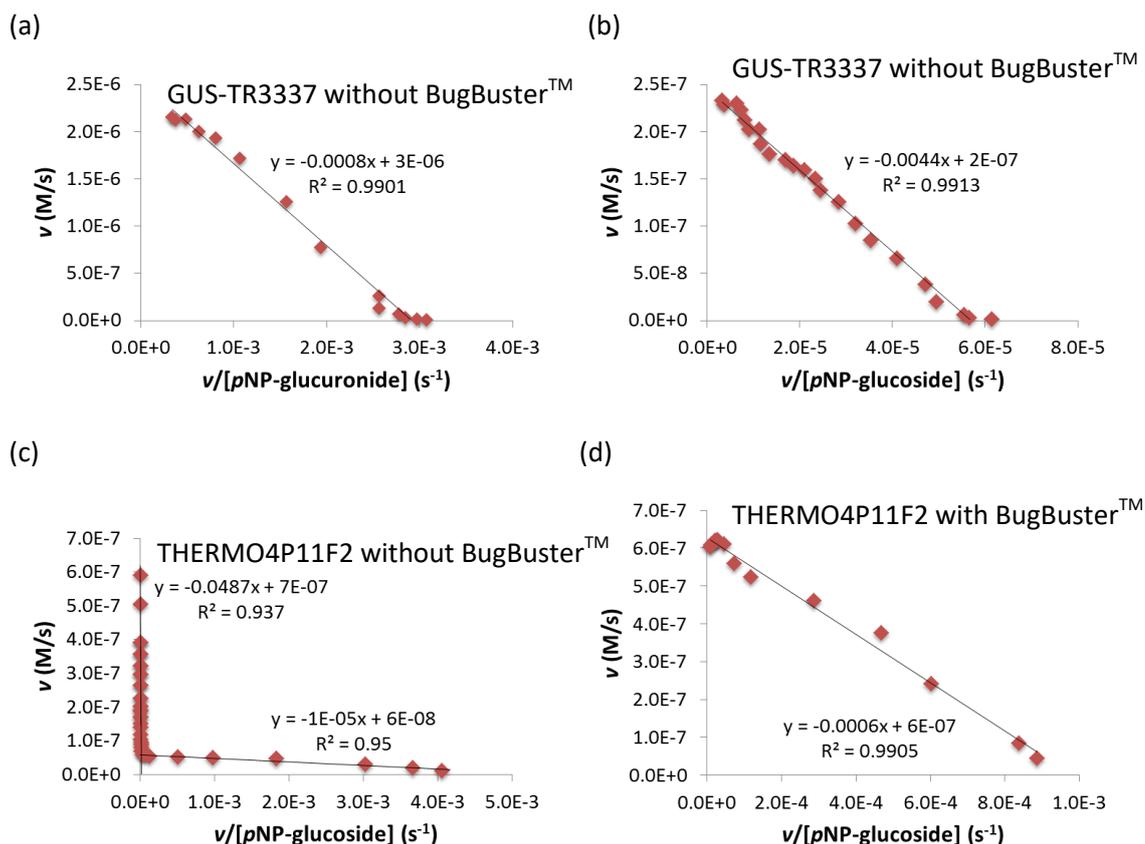


Figure 5.15 Eadie-Hofstee plots of GUS-TR3337 and THERMO4P11F2

Each plot is the mean of triplicates. (a) Hydrolysis of *p*NP-glucuronide by GUS-TR3337 (b) Hydrolysis of *p*NP-glucoside by GUS-TR3337 (c) Hydrolysis of *p*NP-glucoside by THERMO4P11F2 (d) Hydrolysis of *p*NP-glucoside by THERMO4P11F2 in the presence of 5 mM OTG.

In Figures 5.12 and 5.14, the (a) and (b) Michaelis-Menten plots show the native enzymes with glucuronide and glucoside substrates, respectively. The hyperbolic natures of Michaelis-Menten plots are confirmed by the linear Eadie-Hofstee plots given in the (b) and (d) of Figures 5.13 for wild-type enzyme and its variants; and along with the (b) and (d) of Figures 5.15 for the thermostable enzyme and its variants. These are quite different to the Michaelis-Menten plots for the later generation mutants given in part (c) of Figure 5.12 and 5.14. These plots appear to be biphasic and this is confirmed by the Eadie-Hofstee plots given in Figures 5.13c & 5.15c. These plots can be used to calculate kinetic parameters for the low and high affinity substrate ranges (Martin-Nieto *et al.* 1992). Table 5.1 contained the kinetic parameters obtained with Eadie-Hofstee plots for native enzymes (GUS-WT and GUS-TR3337) and representative variant enzymes from each round of evolution. The validity of values found for variants exhibiting biphasic kinetics can be questioned since in neither the low or high substrate ranges is it possible to see a saturation of the binding event. However, from these calculations it was evident that the kinetic parameters obtained at high substrate differ

dramatically from those obtained at low substrate concentrations suggesting that substrate binding brings about a change in the enzyme in the later generation variants. The addition of saturating concentrations of OTG to later generation variants gives rise to a Michaelis-Menten curves (Figure 5.12d & 5.14d) that appear to be hyperbolic and Eadie-Hofstee plots that are straight lines (Figure 5.13d & 5.15d). In the presence of OTG, substrate concentration has little effect on the nature of catalysis of the variant enzymes – as is the case for a typical non-cooperative enzyme.

In the previous section (5.4.2.1.) it was found that OTG was a competitive inhibitor for the native enzymes; that is the hyperbolic shape of the Michaelis-Menten curves did not change with increasing concentrations of OTG. This is in sharp contrast to the variant enzymes that show a change in their response to substrate as the substrate concentration increases. Furthermore, OTG decreases the velocity of the native enzymes, it effectively increases the catalytic rate of variants at the same substrate concentration.

Table 5.1 Apparent values of k_{cat} , K_m and k_{cat}/K_m of GUS-WT* and GUS-TR3337* with pNP-glucoside at different temperatures

(a) 25 °C (b) 45 °C.

(a)

Enzyme	Activity	App.* k_{cat} (s^{-1})	App.* K_m (mM)	App.* k_{cat}/K_m ($M^{-1}s^{-1}$)
WT-GUS		0.200 ± 0.0260	7.600 ± 0.492	26.28
WT1P17G2		0.590 ± 0.0610	4.486 ± 0.084	109.75
WT3P24E11	High activity, low affinity (At high substrate concentrations)	0.409	96.6	4.23
	Low activity, high affinity (At low substrate concentrations)	0.018	0.007	2503.34
WT4P26F8	High activity, low affinity (At high substrate concentrations)	0.072	78.2	0.92
	Low activity, high affinity (At low substrate concentrations)	0.0036	0.03	120.34
WT5P26A7	High activity, low affinity (At high substrate concentrations)	0.33	42.1	7.95
	Low activity, high affinity (At low substrate concentrations)	0.017	0.01	1673.64
GUS-TR3337		0.083 ± 0.0079	4.400 ± 0.016	18.71
THERMO1P6B9		0.400 ± 0.031	8.300 ± 0.200	47.59
THERMO2P23C4	High activity, low affinity (At high substrate concentrations)	0.62	10.5	58.90
	Low activity, high affinity (At low substrate concentrations)	0.0062	0.05	123.69
THERMO3P24F7	High activity, low affinity (At high substrate concentrations)	0.16	47.7	3.36
	Low activity, high affinity (At low substrate concentrations)	0.0092	0.01	917.01
THERMO4P11F2	High activity, low affinity (At high substrate concentrations)	0.88	48.7	18.00
	Low activity, high affinity (At low substrate concentrations)	0.075	0.01	7517.20

App. – apparent

(b)

Enzyme	Activity	App.* k_{cat} (s^{-1})	App.* K_m (mM)	App.* k_{cat}/K_m ($M^{-1}s^{-1}$)
WT-GUS		0.33 ± 0.0041	9.79 ± 0.38	33.56
WT1P17G2		1.74 ± 0.0480	11.80 ± 0.00098	147.10
WT3P24E11	High activity, low affinity (At high substrate concentrations)	0.44	37.3	11.89
	Low activity, high affinity (At low substrate concentrations)	0.02	0.005	3992.90
WT4P26F8	High activity, low affinity (At high substrate concentrations)	0.14	13.1	10.53
	Low activity, high affinity (At low substrate concentrations)	0.01	0.02	517.42
WT5P26A7	High activity, low affinity (At high substrate concentrations)	1	54	18.62
	Low activity, high affinity (At low substrate concentrations)	0.025	0.005	5027.65
GUS-TR3337		0.51 ± 0.0054	8.00 ± 0.28	64.38
THERMO2P23C4	High activity, low affinity (At high substrate concentrations)	2.33	8.5	274.28
	Low activity, high affinity (At low substrate concentrations)	0.0058	0.01	582.85
THERMO3P24F7	High activity, low affinity (At high substrate concentrations)	1.6	93.2	17.17
	Low activity, high affinity (At low substrate concentrations)	0.048	0.04	1200.00
THERMO4P11F2	High activity, low affinity (At high substrate concentrations)	4.79	58.5	81.92
	Low activity, high affinity (At low substrate concentrations)	0.32	0.04	7987.22

App. – apparent

5.3.3.2.1. Assessing negative cooperativity in evolved variants using the Hill equation

The biphasic kinetics obtained with the variant enzymes were clearly not indicative of Michaelis-Menten enzymes. However, not all enzymes obey Michaelis-Menten kinetics – the most common deviation is due to cooperativity that can usually be described with Hill equation; a modification of the Michaelis-Menten equation, as shown below:

$$v = (V_{max} [S]^n) / (K_{0.5}^n + [S]^n) \quad (\text{Equation 5.2})$$

Where v is initial velocity; V_{max} , maximum velocity; n , Hill coefficient; $[S]$, substrate concentration and $K_{0.5}$ is the substrate concentration at half-maximal velocity. When $n = 1$, Equation 5.2 is exactly the same as the Michaelis-Menten equation and $K_{0.5} = K_m$.

The Hill equation can be used to describe cooperative binding of ligands to macromolecules was proposed by Hill (1910), based on his observation of the oxygen binding nature to haemoglobin. Cooperativity in proteins and enzymes is often due to interactions between subunits or subsites so that binding of the first ligand molecule increases (positive cooperativity) or decreases (negative cooperativity) the binding affinity of the other(s) (Atkins

2005). In the aforementioned case, cooperativity involves multiple subunits so that each molecule has multiple substrate binding sites. Cooperativity also gives rise to a non-Michaelis-Menten kinetic response in certain enzymes that consist of a single subunit with a single substrate-binding site (Rose *et al.* 1993, Ferdinand 1966, Rabin 1967).

For the Hill equation, if $n = 1$, there is no cooperativity, $n < 1$ indicates negative cooperativity while $n > 1$ indicates positive cooperativity (Cornish-Bowden 2004). A cursory glance at the plots in Figures 5.11C & 5.13C are clearly not indicative of positive cooperativity, but are suggestive of negative cooperativity. Proving negative cooperativity using the Hill equation can be a rather ambiguous process (Goodrich & Kugel 2007). The Hill equation is more easily applied to ligand binding studies in receptors and cytochrome P450, where the binding capacity, which is equivalent to V_{\max} is known (Sakai 1994). The Hill coefficients were determined by using KaleidaGraph software to perform a non-linear least-squares fit of all the data points to the Hill equation (Equation 5.2). The Hill plots for all the data sets are given in Appendix D.3. & D.4. while the Hill constants (n values) for the data sets are given below in Table 5.2. The calculated values of n demonstrate the non-cooperativity ($n = 1$) of both parents and first generation mutants; and suggest negative cooperativity ($n < 1$) of mutants in later generations in the absence of OTG. Hill coefficients of 1 indicate that there was no cooperativity in *p*NP-glucoside hydrolysis after treating the mutants with 0.5 mM OTG. Even though fitting the data to the a Hill equation yields values for n and parameters corresponding to V_{\max} and K_m , the errors associated with the parameters were high for the mutants, making the computation of the Hill coefficient meaningless (Appendix D.3.). The latter generation enzymes cannot be said to exhibit negative cooperativity in the absence of OTG.

Table 5.2 Values of the Hill constant (n) of wild-type and evolved variants as derived from the plots in Appendix D.3. & D.4. and their corresponding correlation coefficient (R^2) values

Enzyme	In the absence of 5 mM OTG		In the presence of 5 mM OTG	
	n	R^2	n	R^2
GUS-WT	1.00	1.00	1.01	1.00
WT1P17G2	1.03	1.00	1.00	1.00
WT3P24E11	0.65	0.99	1.00	1.00
WT4P26F8	0.51	1.00	1.01	1.00
WT5P26A7	0.90	1.00	1.00	1.00
GUS-TR3337	1.01	1.00	1.00	1.00
THERMO1P6B9	1.01	1.00	1.00	1.00
THERMO2P23C4	0.74	1.00	1.00	1.00
THERMO3P24F7	0.52	1.00	1.00	1.00
THERMO4P11F2	0.59	0.99	1.00	1.00

The kinetics curves of WT5P26A7 and THERMO4P11F2 were dissected into 2 regions (one operating at low and another at high p NP-glucoside concentrations). Data were fitted to the Hill equation, one for concentrations up to 0.8 mM and another one using only concentrations above 3.2 mM. Dissection of the kinetic data reveals that at substrate concentrations below 0.8 mM, glucosidase activity conforms to Michaelis-Menten kinetics with the Hill coefficient equal to 1 (Figure 5.16a & c); at elevated substrate concentration (<0.8 mM) both mutants exhibited greater activity but a decrease in cooperativity (Figure 5.16b & d). Random addition of OTG abolished cooperativity at higher substrate concentrations, as evidenced by the hyperbolic curve (Figure 5.17).

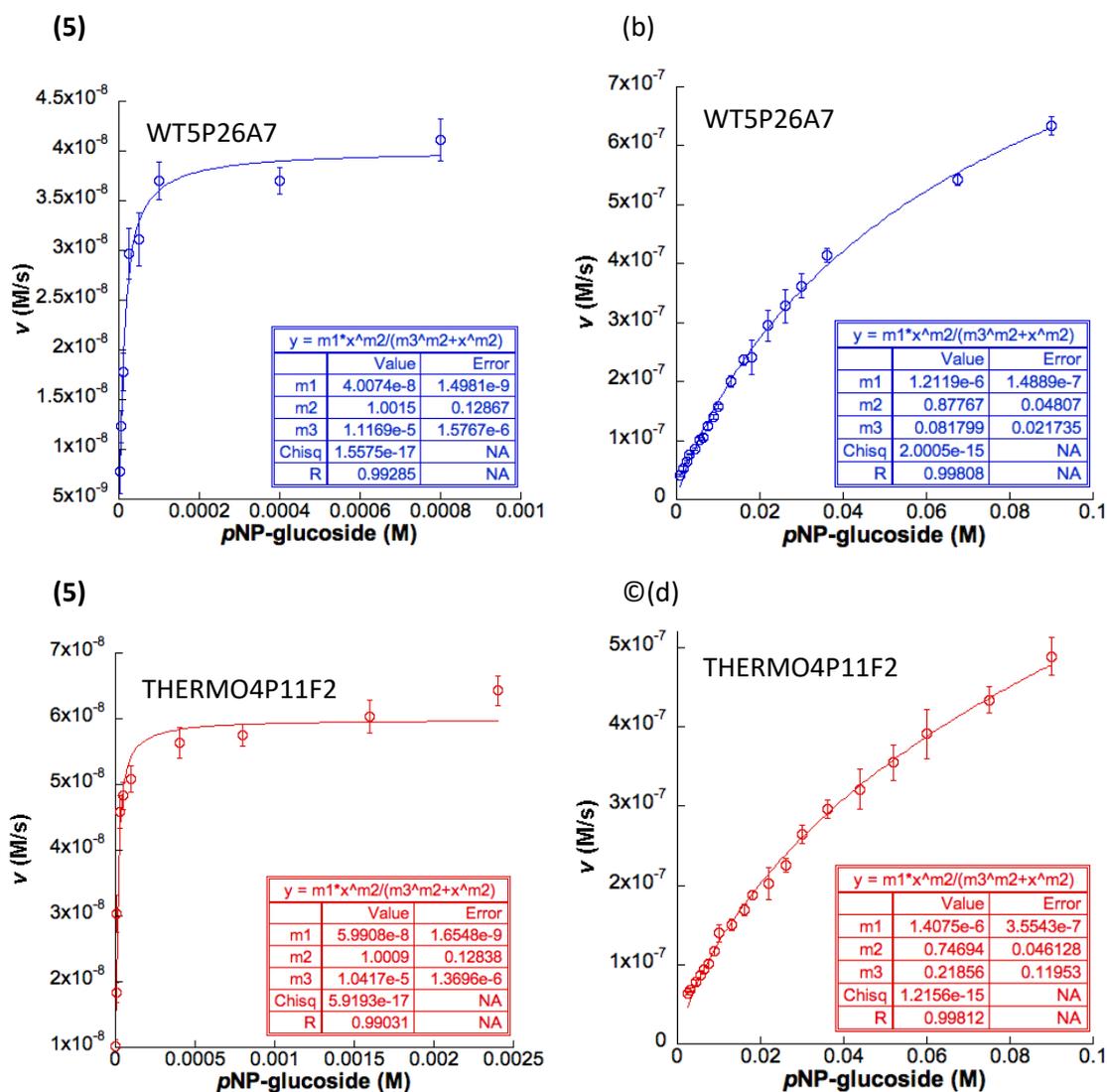


Figure 5.16 Velocity-substrate curves of improved variants at low (micromolar range) and high (millimolar range) substrate concentrations

(5) Plots of WT5P26A7 at the substrate concentration range of 2.5 – 800 μ M (b) Plots of WT5P26A7 at the substrate concentration range of 0.8 – 90 mM (c) Plots of THERMO4P11F2 at the substrate concentration range of 2.5 – 2400 μ M (d) Plots of THERMO4P11F2 at the substrate concentration range of 2.4 – 90 mM.

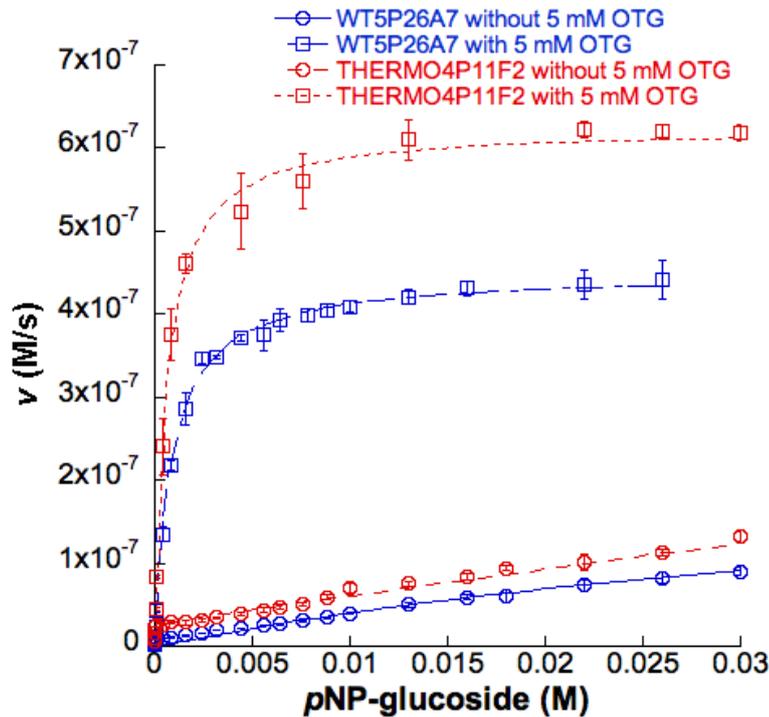


Figure 5.17 Comparison of substrate saturation curves for 0.6 μm WT6P26A7 and THERMO4P11F2 in the presence and absence of 0.5 mM OTG

The initial turnover at 25 °C in pH 7.4 solutions with several concentrations of *p*NP-glucoside up to 30 mM. Data collection and curve fitting are described in Chapter 2 (Section 2.3.4.1. & 2.3.4.1.1.).

Since the kinetics of variants starts with a hyperbolic-like pattern followed by a near linear portion at later stage, the biphasic curves represent data also fitted to a double Michaelis-Menten curve:

$$v = V_{\max 1} [S] / (K_{m1} + [S]) + V_{\max 2} [S] / (K_{m2} + [S]) \quad (\text{Equation 5.3})$$

in which v is the initial velocity, V_{\max} is the maximum velocity, K_m is the Michaelis-Menten constant and $[S]$ is the substrate concentration. This model assumes $V_{\max 2} > V_{\max 1}$ and $K_{m2} > K_{m1}$. Although this equation provides good fit to the experimental data, the values of the constants estimated from the equations can vary dramatically when $[S]$ approaches K_{m2} . In this study, the apparent $V_{\max 2}$ was unable to be accurately predicted and thus apparent K_{m2} in which saturation is unobtainable up to 0.1 M *p*NP-glucoside (Appendix D.8.). The value of apparent K_m and V_{\max} returned from the fit were different from the values calculated from Eadie-Hofstee transformation (Hofstee 1959) (Table 5.3). It is apparent that the double Michealis-Menten equation gives a much more precise estimation of K_m and V_{\max} as indicated by a higher R^2 as compared to Eadie-Hofstee transformation.

Table 5.3 K_m values obtained from the linear fittings of Eadie-Hofstee transformation and double Michealis-Menten equationGoodness of fit was evaluated from R^2 .

Enzyme	Eadie-Hofstee			Double Michealis-Menten			
	K_{m1} (mM)	R^2	K_{m2} (mM)	R^2	K_{m1} (mM)	K_{m2} (mM)	R^2
WT3P24E11	0.007	0.99	96.6	0.98	0.0045 ± 0.0012	150.26 ± 8.35	1
WT4P26F8	0.030	0.91	78.2	0.90	0.1249 ± 0.0442	128.48 ± 23.52	1
WT5P26A7	0.010	0.97	42.1	0.97	0.0038 ± 0.0033	66.12 ± 4.08	1
THERMO2P23C4	0.050	0.92	10.5	0.91	2.4276 ± 1.5005	60.43 ± 20.55	1
THERMO3P24F7	0.010	0.99	47.7	0.93	0.0186 ± 0.0072	70.99 ± 7.83	1
THERMO4P11F2	0.010	0.95	48.7	0.94	0.0057 ± 0.0019	114.08 ± 9.61	1

5.4. Structure – Implications of mutations

Structures often provide a basis for understanding complex biological phenomena and for this reason some effort was made to obtain some structural information. Cooperativity is often associated with inter-subunit interactions where the binding of a ligand to one subunit may affect ligand binding to other subunits (Koshland *et al.* 1966). The enhanced activity in the presence of OTG was first observed for WT3P24E11 (T509A, K568Q, P597Q) and THERMO3P24F7 (S550N, S566N and K568E). Given T509A has no significant effect on the activation of WT1P17G2 and P597Q is located far away from the active site, it was likely that K568Q was responsible for the activation. This agrees well with the similar mutations observed in THERMO3P24F7 where the serine at position 550 and 566, were both converted back to asparagine, the original residues in GUS-WT; and the fact that K568E affects activation directly. Furthermore, it has been noted in other studies that evolution has resulted in changes at the interface of subunit boundaries and that this could account for enhanced stability of the variant enzymes. For these reasons, it was thought useful to obtain some indication of the oligomeric structure of the native and variant enzymes. Size exclusion chromatography and discontinuous native protein gel electrophoresis analysis were performed to monitor the different oligomeric forms of the native and variant enzymes. The data are presented in the Appendix D.9.1. & D.9.2., respectively. These are not approaches that could lead to the identification of minor changes in the oligomeric state of the proteins. In fact, the elution and native gel profiles of the selected variants and parents look very similar leading to the conclusion that there are no significant differences between the natives and their variant forms. However, neither physical technique could detect small differences in the equilibrium between the different oligomeric forms and that there were small differences in this equilibrium. Our results also ruled out the possibilities that it might be caused by: i) differences in the ratio of dimer/tetramer that gave rise to different performance; ii) protein

concentrations and iii) end-product inhibition and activation. The results and discussions of these experiments are found in Appendix D.9.3.

A more definitive approach to structure involves the use of X-ray crystallography. Although this is a very powerful tool, it requires the production of crystals for diffraction experiments. The crystal structures of *E. coli* β -GUS alone and in combination with a low-affinity inhibitor have been solved (Wallace *et al.* 2010). An attempt to grow a crystal of WT5P26A7 was performed using the published conditions (2 mg/mL protein in 17% polyethylene glycol 3350, 250 mM magnesium acetate, and 0.02% sodium azide at 16°C) with varying polyethylene glycol (PEG) 3350 concentration from 14% to 19% versus magnesium acetate concentration from 150 mM to 300 mM. Unfortunately, no crystals were observed. Second attempt was performed with a library of known crystallising conditions. A series of microbatch crystallisations in 96-well plates were performed using Hampton Research Crystallisation kits including Index™, Crystal Screen™, PEGRx™, PEG/Ion™ and SaltRx™. Out of 480 crystallising conditions trialled, only six conditions displayed possible crystal formation (Figure 5.18).

The growth of crystals large enough for x-ray diffraction was attempted using the hanging-drop method in 24-well plates. The crystallisation conditions were optimized by varying the pH values and PEG concentrations. No crystals were observed from any of the scaled up crystallisations. Repeating the conditions using batch crystallisation (as used in the screens) might be more successful. However, the appropriate crystallisation conditions to grow sizable crystals for x-ray crystal analysis are still under investigation.

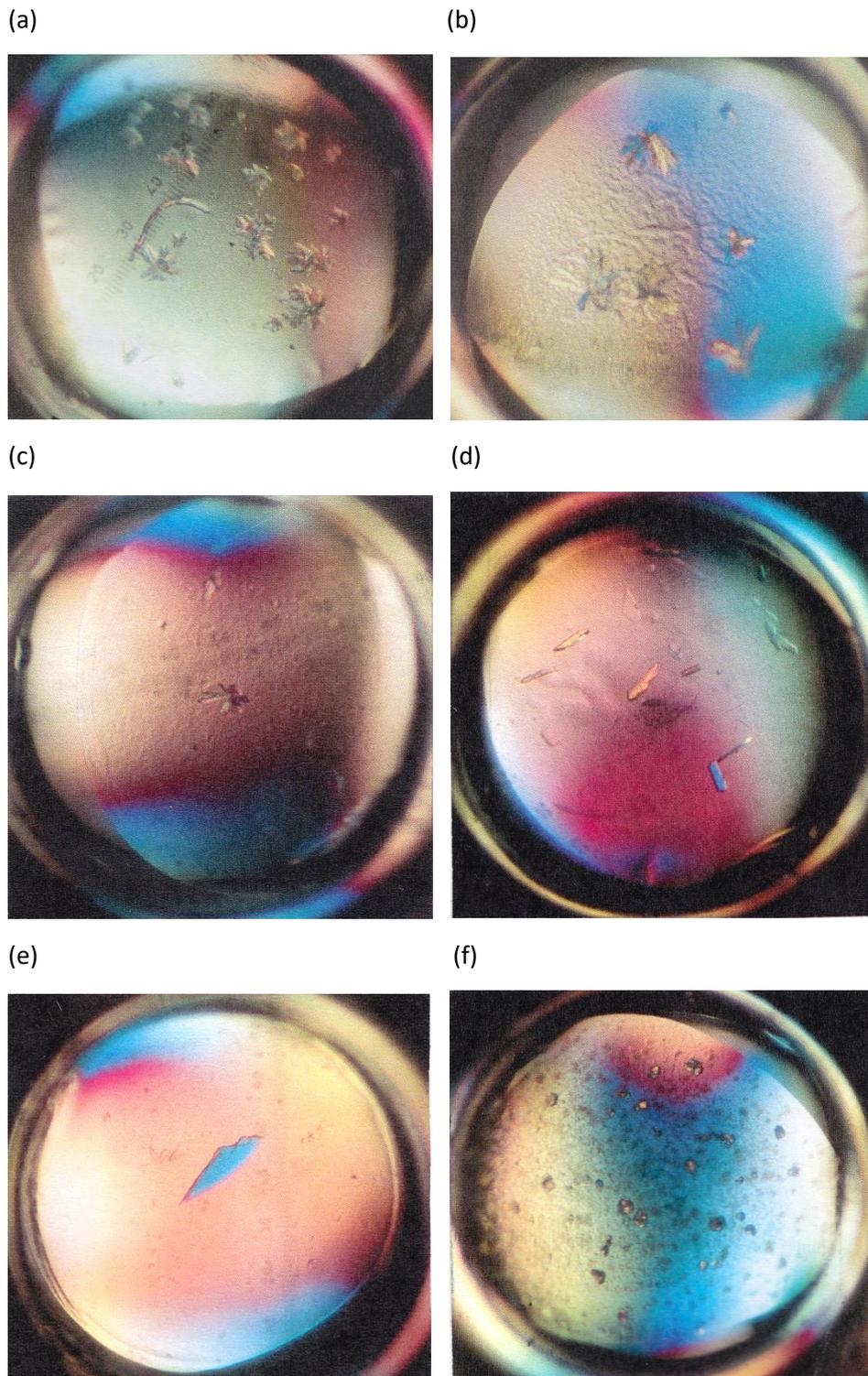


Figure 5.18 WT5P26A7 crystals grown under different conditions

(a) 0.2 M ammonium tartrate dibasic, 20% w/v PEG 3350, pH 6.6 (b) 0.2 M ammonium tartrate dibasic (pH 7), 20% w/v PEG 3350 (c) 0.2 M ammonium sulfate (pH 6), 20% w/v PEG 3350 (d) 0.1 M sodium citrate tribasic dehydrate (pH 5.5), 22% w/v PEG 3350, 0.1% w/v OG (e) 0.1 M sodium citrate tribasic dehydrate (pH 5.5), 18% w/v PEG 3350 (f) 1.8 M ammonium sulfate, 0.1 M BIS-TRIS (pH 6.5) 2% v/v PEG monomethyl ether 550.

5.5. Conclusions

The contents of this chapter were, for the most part, not anticipated in the initial stages of the project. The work came about as a result of the use of the BugBuster™ reagent. As noted at the start of this chapter, the manufacturer advertises that the reagent disrupts membranes but does not interfere with the cellular proteins. In addition, the manufacturer did not specify what was in the BugBuster™ reagent so that it was difficult for the investigator to determine if the reagent was likely to affect a protein of interest without resorting to experimental tests. BugBuster™ did not completely obliterate the activity of β -GUS and the treatment devised for the assays gave rise to low levels of activity so that it appeared that this was an ideal protocol to detect the increased activity of variants. It now appears that a test of alternative methods of lysis, such as lysozyme (and not some other surfactant), should have been carried out.

Despite the problems introduced by the BugBuster™ reagent, this work serves as an example of how enzyme evolves to increase new activity in response to a competitive inhibitor. This study also reveals the presence of OTG in most of the commercial non-ionic detergents making them not an ideal choice for cell lysis in directed evolution where the aim is to enhance glucosidase activity. Consequently, the directed evolution of GUS-WT and GUS-TR3337 has confirmed that “you can’t always get what you want”, but you do “get what you select for” (Jagger & Richards 1969, Giver & Arnold 1998) – that is, variants that have enhanced activity in the presence of a component of BugBuster™.

The early experiments (Section 5.3) of this chapter identify the BugBuster™ reagent as the cause of the anomalies in the directed evolution. The evidence presented was clear-cut. However, the process by which OTG affects β -GUS, was not so clear-cut. The BugBuster™ reagent is a mixture of compounds and analysis by chemical means was not straightforward. Given the time constraints of an Australian PhD, it seemed that testing likely components of the BugBuster™ reagent was a more realistic option. OTG mimics the effects of Bugbuster™ reagent on purified proteins, but it was possible that the BugBuster™ reagent contains more than one compound that affects β -GUS. If there was more than one compound or if there were different compounds in the Bugbuster™ reagent that was (or were) responsible for the effects on β -GUS, then it was very likely that it was (or they were) very similar to OTG.

Structurally, OG was a glycoside derived from glucose and octanol (Figure 5.6a). Because it was less stable and can be hydrolysed by β -glucosidase, Saito & Tsuchiya (1984) had first synthesized OTG to offer an alternative to OG. The presence of the thioether linkage made OTG resistant to degradation by β -glucosidase (Figure 5.6b). Given the structures of OG

and OTG it was not a surprise they inhibited the glucosidase activities of wild-type and the variants generated in the first 2 rounds of directed evolution. Similar behaviour was observed with SDS suggests that the aliphatic chain of OTG was in large part responsible for the affects observed in the experiments described in this chapter.

The fact that OTG was a competitive inhibitor of GUS-WT and GUS-Tr3337 was consistent with the idea that it binds in the active site. This is also consistent with the observation that OTG has a sugar component that is similar to the physiological substrate and the substrate being selected for in the directed evolution experiments. One would expect that the glycosyl group of OTG caused it to bind in the active site and this gave rise to the observed competitive inhibition. The K_i obtained in these experiments indicates that it binds with very low affinity. Variants have evolved such that OTG activates glucosidase activity, in a concentration dependent manner, is surprising and has important implications.

Activation by OTG of later generation variants was surprising (at least to the candidature) because evolution could simply have produced variants that did not bind OTG. Clearly, the glycosyl group of OTG did not interfere with the activity of the later generation variants suggesting that mutations prevented the glycosyl group of OTG from binding in the sugar binding site, but how did this happen? Perhaps, mutations were introduced that prevented the glycosyl group from binding in the active site, but which allowed retention of binding via the aliphatic group. This suggestion was speculative and further proof is needed if it is to be regarded as factual. However, the facts were that 1) OTG enhances the activity and 2) it did so in a concentration dependent manner. This second fact strongly suggests that OTG acts on the later generation variants by binding to them. Experiments to verify and quantitate OTG binding to GUS are best done with an isothermal titration calorimetric study, however, a suitable instrument could not be accessed in the time needed to complete the experiment. It is intended that this experiment be done in the future.

Further evidence that OTG binds to and causes significant changes to β -GUS is provided by a comparison of the kinetics of the β -GUS and β -GUS in the presence of OTG. OTG causes β -GUS to take on Michaelic-Menten kinetics while the activity profile of the enzyme in the absence of OTG is clearly not Michaelis-Menten. The saturation curves of OTG free enzyme variants are best fit with a double Michealis-Menten model but obey Michaelis-Menten kinetics upon addition of OTG. This suggests that a structural change is involved and that the structural change is brought about by OTG. This change may be due to OTG having a direct effect on the active site or it may cause an alteration in the subunit interactions that

then causes a change in the active site. This later case was explored by trying to use negative cooperativity to explain the reaction profile. However, the error associated with the calculated K_m is significant indicated a poor fitting model. A more definitive technique was required to settle this issue; ideally x-ray crystallography could have been used to clarify this issue if suitable crystals had been obtained.

The use of Bugbuster™ cost the PhD candidature a great deal of time and effort, but its use also gave some insight into how enzymes might evolve in nature. At this stage some speculation is in order. A component of Bugbuster™ (OTG) binds to β -GUS and it seems likely that it induces a structural change in the protein. Variants were selected based on the activity of the enzyme / OTG complex. In the absence of OTG, these same variants displayed much less activity suggesting that OTG also binds to the variants, but not in the same manner as the occurs in native β -GUS (and thermostable GUS-TR3337). In the absence of OTG, the structure of the variant enzymes is not optimal for activity; hence the appearance of OTG activates the variant forms of β -GUS. In the experiments described in this chapter, OTG is an experimental artefact. However, in the cell, it is possible that there is selective advantage in turning on (or activating) an enzyme response to the presence of a non-substrate reagent (call it factor X). Under these circumstances, there will be some tendency for the enzyme to evolve in such a way that later generation variants bind X and require it for activity. This seems to be what has occurred in the evolution of β -GUS with OTG.

5.6. References

- Atkins, W. M. (2005) Non-Michaelis-Menten kinetics in cytochrome P450-catalyzed reactions. *Annu Rev Pharmacol Toxicol*, 45, 291-310.
- Cornish-Bowden, A. 2004. *Fundamental of enzyme kinetics*. Portland Press.
- Ferdinand, W. (1966) The interpretation of non-hyperbolic rate curves for two-substrate enzymes. A possible mechanism for phosphofructokinase. *Biochem J*, 98, 278-83.
- Giver, L. & F. H. Arnold (1998) Combinatorial protein design by in vitro recombination. *Curr Opin Chem Biol*, 2, 335-8.
- Goodrich, J. A. & J. F. Kugel. 2007. *Binding and kinetics for molecular biologists*. Cold Spring Harbor Laboratory Press.
- Hill, A. V. (1910) The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *Journal of Physiology*, 40, iv-vii.
- Hofstee, B. H. (1959) Non-inverted versus inverted plots in enzyme kinetics. *Nature*, 184, 1296-8.

- Jagger, M. & K. Richards. 1969. You can't always get what you want. In *Let it bleed*. ABKCO.
- Koshland, D. E., Jr., G. Nemethy & D. Filmer (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, 5, 365-85.
- Martin-Nieto, J., E. Flores & A. Herrero (1992) Biphasic Kinetic Behavior of Nitrate Reductase from Heterocystous, Nitrogen-Fixing Cyanobacteria. *Plant Physiology*, 100, 157-163.
- Rabin, B. R. (1967) Co-operative effects in enzyme catalysis: a possible kinetic model based on substrate-induced conformation isomerization. *Biochemical Journal*, 102, 22C-23C.
- Rose, I. A., J. V. Warms & R. G. Yuan (1993) Role of conformational change in the fumarase reaction cycle. *Biochemistry*, 32, 8504-11.
- Saito, S. & T. Tsuchiya (1984) Characteristics of n-octyl beta-D-thioglucopyranoside, a new non-ionic detergent useful for membrane biochemistry. *Biochem J*, 222, 829-32.
- Sakai, S. (1994) Negative cooperativity in the prolactin-receptor interaction in the rabbit mammary gland: action of high prolactin concentration. *J Dairy Sci*, 77, 433-8.
- Wallace, B. D., H. Wang, K. T. Lane, J. E. Scott, J. Orans, J. S. Koo, M. Venkatesh, C. Jobin, L. A. Yeh, S. Mani & M. R. Redinbo (2010) Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science*, 330, 831-5.

Chapter 6

Synthesis of steroidal glucosides using glucuronylsynthase

6.1. Introduction

One of the aims of the present study was to convert the evolved β -GUS variants with glucosidase activity into enzymes that could catalyse the synthesis of steroid glucosides, enzymes that were denoted glucosynthases. Similar experiments had been carried out with β -GUS mutants; denoted as glucuronylsynthases. Background for this work was given in Chapter 1. The main impediment to the experiments described in this chapter was time. The experiments to evolve GUS variants with altered substrate specificity had taken most of the time allowed for an Australian PhD and there was only a short interval of time available to make the necessary changes that might bring about synthetic activity. Very little time was left to optimize experimental conditions for the synthesis reaction.

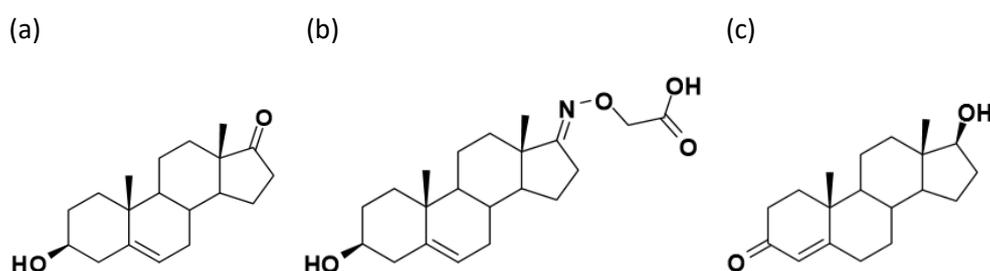
6.2. Glycosylation reaction of glucuronylsynthase

In late 2005, three glucuronylsynthases (E504A, E504G, and E504S mutants of GUS-WT) were successfully expressed and isolated at the University of Sydney (Wilkinson *et al.* 2008). From McLeod group's small scale glucuronylation reactions, it was found that the glycine variant (E504G) was more active. This observation was thought to be due to the absence of a side chain on G504. Similar experimental trials were carried out with some of the best variants described in Chapter 4. That is, the nucleophilic glutamate residue (E504) was also replaced by either an alanine (E504A), glycine (E504G), or serine (E504S) residue, substitutions known to be effective in other glycosynthases (Williams & Withers 2002). Both GUS-WT and GUS-TR3337 along with evolved mutants (WT3P24E11, WT5P26A7 and THERMO4P11F2) were subjected to mutagenesis at position 504. The mutants are shown in Table 6.1 below.

Table 6.1 The glucuronylsynthase enzymes prepared through the site-directed mutagenesis of the native *E. coli gus*

Hydrolase	Glucuronylsynthase		
	E504A	E504G	E504S
GUS-WT	GUSWT-E504A	GUSWT-E504G	GUSWT-E504S
WT3P24E11	WT3P24E11-E504A	WT3P24E11-E504G	WT3P24E11-E504S
WT5P26A7	WT5P26A7-E504A	WT5P26A7-E504G	WT5P26A7-E504S
GUS-TR3337	GUSTR3337-E504A	GUSTR3337-E504G	GUSTR3337-E504S
THERMO4P11F2	THERMO4P11F2-E504A	THERMO4P11F2-E504G	THERMO4P11F2-E504S

The present work follows protocols developed in the study of glucuronylsynthases so that a brief description of that work will be given here. The first steroid glucuronide isolated from the glucuronylsynthase reaction was derived from dehydroepiandrosterone (DHEA) (Figure 6.1a) (Wilkinson *et al.* 2008). DHEA was glucuronylated in a low yield of 26%, largely caused by the very hydrophobic nature of the DHEA (Wilkinson *et al.* 2008). The low aqueous solubility of the DHEA was overcome by the synthesis of DHEA *O*-(carboxymethyl)oxime (CMO-DHEA) (Figure 6.1b) (Wilkinson *et al.* 2008). The carboxylic acid attached to the end of this oxime has a pK_a of $\sim 2-3$ meaning the carboxylate anion would and provide prevail enhanced water solubility in the glucuronylsynthase reactions performed at pH 7.5. The higher aqueous solubility afforded an impressive 76% yield of CMO-DHEA β -glucuronide from the glucuronylsynthase reaction (Wilkinson *et al.* 2008). Of particular interest, the glucuronylation of testosterone afforded the desired glucuronide in 48 – 50% yield (Ma *et al.* 2014). Since CMO-DHEA and testosterone were two examples used to demonstrate the successful glucuronylsynthesis of steroidal glucuronides, these two steroids were used as steroidal acceptors to test the glycosylation capability of the glucuronylsynthase.

**Figure 6.1 Chemical structures of steroids**
(a) DHEA (b) CMO-DHEA (c) Testosterone.

The glucuronylation protocol developed had three main components: the glucuronylsynthase enzyme, the synthetically derived α -D-glucuronyl fluoride donor and a suitable acceptor alcohol (Figure 6.2) (Wilkinson *et al.* 2011). In this study, the glucuronylsynthase enzymes are given in Table 6.1. The donor and acceptor substrates were α -D-glucosyl fluoride and steroids (CMO-DHEA and testosterone), respectively. They were

prepared by Mr. Andy Pranata from the McLeod research group (ANU). Similarly, the glycosylation reactions were carried out in a mixed 10% v/v *tert*-butanol-sodium phosphate buffer at pH 7.5 and 37 °C for 2 days. The addition of *tert*-butanol aids the solubility of hydrophobic steroids and enhances the reaction rate for glucuronylsynthase (Wilkinson *et al.* 2011).

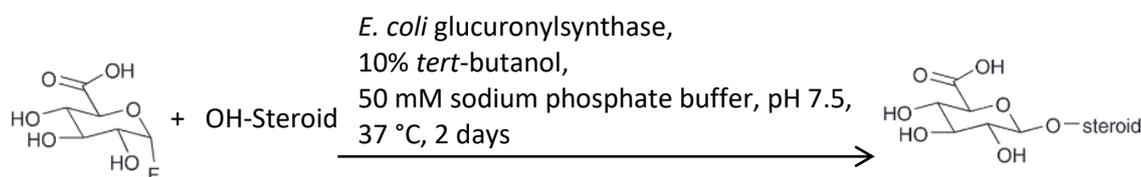


Figure 6.2 The glucuronylsynthase protocol

6.3. Extraction for glucuronides from the reaction mix

The solid-phase extraction (SPE) cartridges were used to separate the components of the glycosylation reactions. The SPE cartridges elute one or more components from a small amount of sorbent using different types of mobile phases. The sorbent can be normal-phase, reverse-phase, ion-exchange or a combination of each media. The protocol developed by the McLeod group for SPE purification employed Oasis® WAX (weak anion exchange) branded SPE cartridges containing polar and weak anion exchange residues (Figure 6.3). The carboxylic acid function of the glucuronide conjugates, allow02 isolation of these conjugates using mixed phase anionic SPE cartridges. Figure 6.4 gives an example of how SPE works; it illustrates the extraction of testosterone β -D-glucuronide directly from a reaction mixture by the use of an Oasis® WAX cartridge. The reaction mixture contains the parent steroid (testosterone), α -D-glucuronyl fluoride, the glucuronide product (testosterone β -D-glucuronide) and the glucuronylsynthase enzyme.

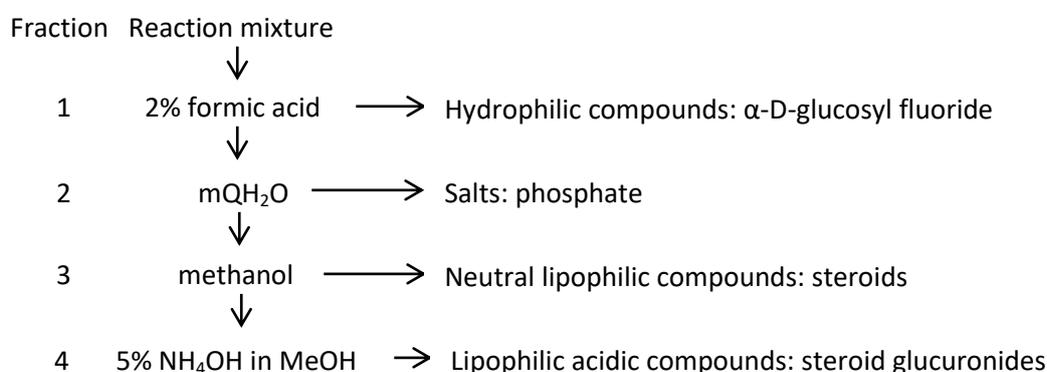


Figure 6.3 Schematic diagram of the procedure used for WAX SPE cartridges

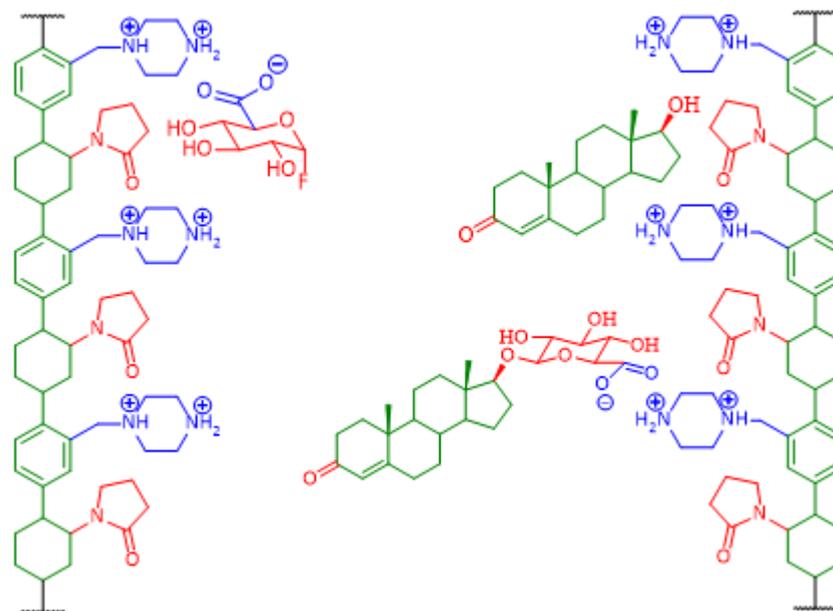


Figure 6.4 The interactions of the components from the testosterone glucuronylsynthase reaction with the Waters Oasis® WAX sorbent

Lipophilic residues are coloured green, hydrophilic residues are red and ionic residues are coloured blue.

There were three key components present in the sorbent of the cartridge: i) the polymeric hydrocarbon functionality (green) provided the lipophilic interaction with the similarly lipophilic backbone of the steroid; ii) the hydrophilic functionality (red) of γ -lactam provided the polar interactions with the hydroxyl groups of the sugar and iii) the ionic functionality (blue) of piperazine provided the weak anion exchange.

The first mobile phase is 2 % v/v aqueous formic acid, that serves to elute the sugars. Then water was passed through the cartridge to protonate piperazine (blue) and ionize the carboxylate on the glucuronide at about pH 6-7. Elution with methanol was then performed. The unconjugated testosterone was readily soluble in methanol and eluted while the glucuronide remained bound to the resin through the charged interaction between the carboxylate anion and the piperazinium cation. The final elution step was performed with 5 % v/v ammonium hydroxide in methanol where ammonium hydroxide deprotonates the piperazinium cations, removing the charged interaction and allowing the glucuronide to dissolve into methanol and elute from the SPE cartridge in the form of an ammonium salt.

6.4. The glycosylation reaction

In this study, the target glucoside conjugate was purified by SPE using the same protocol. Unlike glucuronide conjugates, steroidal glucosides were isolated based on the chemical

properties of steroids. Thus, steroidal glucosides will be eluted with steroid in the same fraction. For example, CMO-DHEA β -glucosides produced in the glycosynthase reactions would co-elute with CMO-DHEA in fraction 4 (5 % NH_4OH in MeOH) due to the carboxylic acid of the oxime and glucoside; testosterone β -D-glucuronides would co-elute with testosterone in fraction 3 (methanol). The resulting fractions 3 and 4 from each glycosylation reaction were subjected to the analysis by ^1H NMR and electron-spray ionization mass spectrum (MS-ESI) analyses.

6.5. Detection of steroidal glucosides from the glycosylation reaction

After a 2 day incubation of 6 reactions containing CMO-DHEA, α -D-glucuronyl fluoride and six synthetic enzymes (GUSWT-E504A, GUSWT-E504G, GUSWT-E504S, GUSTR3337-E504A, GUSTR3337-E504G and GUSTR3337-E504S), the crude reactions were loaded onto Oasis[®] WAX SPE cartridges. The ^1H NMR spectrum of fraction 4 from the reactions with CMO-DHEA, gave evidence for only the CMO-DHEA – no product. Among these reactions, visible precipitation was observed in the fraction 4 of the reaction catalyzed by GUSWT-E504G. The solids were isolated and evaporated to dryness by rotary evaporation. The dried residue was suspended in approximately 600 μL deuterated water (D_2O) prior to ^1H -NMR analysis. The ^1H -NMR spectrum showed some evidence for the presence of the CMO-DHEA β -glucoside (data not shown). Unfortunately the result could not be further confirmed by MS-ESI as the sample was accidentally discarded.

Following the failure to detect the glucoside compounds from the glycosylation reaction using wild-type synthases, the reactions were tested against CMO-DHEA and testosterone using mutant enzymes listed in Table 6.1. From the experience gained in previous reactions, fractions 3 and 4 from each extraction of the reactions, were dried and dissolved in D_2O for ^1H - NMR analysis and subsequent MS-ESI. However, no glucoside products were detected in all of the 18 glucuronylsynthase reactions.

Since octyl- β -D-thioglucoside (OTG) was found to have a positive effect on the glucosidase activity of improved mutants, it was added into a new set of synthetic reactions using mutant listed in Table 6.1, in the hope of providing similar effects on the glycosylation activity. OTG was eluted in fraction 3 and identified by ^1H -NMR and MS-ESI. Other than OTG, no glucoside products were detected in all reactions.

6.6. Concluding remarks

The GUS-WT and GUS-TR3337 and their variants were successfully converted to the corresponding synthases. The methods, which involved glucuronylsynthase reaction and subsequent SPE purification of the products (steroidal glucoside), were adopted from Ma *et al.* (2014). No activity was detected with the substrates used to produce steroidal glucuronides. There was insufficient time to optimize the protocols for the synthesis and purification of steroidal glucosides. Furthermore, although no activity was detected with steroids, it is possible that the enzymes could have worked with other substrates. Again, there was not time to test this hypothesis.

It was envisaged that the mutations obtained while evolving GUS-WT and GUS-TR3337 for a high level of glucosidase activity could provide insight into the glycosylation reaction. Of particular interest was the correlation between hydrolysis and glycosylation, i.e. would the evolved variants also possess higher glycosylation activity? It was hoped that this might be possible in the future after refinements had been made to the reaction protocols so that useful levels of product could be obtained.

6.7. References

- Ma, P., N. Kanizaj, S. A. Chan, D. L. Ollis & M. D. McLeod (2014) The *Escherichia coli* glucuronylsynthase promoted synthesis of steroid glucuronides: improved practicality and broader scope. *Org Biomol Chem*, 12, 6208-14.
- Wilkinson, S. M., C. W. Liew, J. P. Mackay, H. M. Salleh, S. G. Withers & M. D. McLeod (2008) *Escherichia coli* glucuronylsynthase: an engineered enzyme for the synthesis of beta-glucuronides. *Org Lett*, 10, 1585-8.
- Wilkinson, S. M., M. A. Watson, A. C. Willis & M. D. McLeod (2011) Experimental and Kinetic Studies of the *Escherichia coli* Glucuronylsynthase: An Engineered Enzyme for the Synthesis of Glucuronide Conjugates. *The Journal of Organic Chemistry*, 76, 1992-2000.
- Williams, S. J. & S. G. Withers (2002) Glycosynthases: Mutant Glycosidases for Glycoside Synthesis. *Australian Journal of Chemistry*, 55, 3-12.

Chapter 7

Conclusion and Future Directions

Chemists have known for many years that there are advantages to be gained by using enzymes. However, there are not many enzymes for reactions of interest to chemists. Part of the problem in finding a suitable enzyme for a given reaction is that enzymes are usually very specific – one of the features that make enzymes very useful in a biological context. So, the question arises, can protein engineers alter the substrate specificity of an enzyme to make it a useful catalyst for some practical application, as would be encountered by a synthetic chemist? This is the question addressed in this thesis. There is no general answer to this question, rather answers can be provided on a case-by-case basis. So, it is important to choose an enzyme and reaction that are of some practical utility.

One of the more common steps in many synthetic reactions is to fuse a sugar ring, such a glucose or glucuronic acid, to a variety of substrates. It was a class of reactions that was the target of the present work. The reaction catalysed by *E. coli* β -GUS was unusual in that it is only specific for glucuronic acid and not the aglycone. A simple mutation converts this hydrolytic enzyme into glucuronosynthase – one that will fuse a glucuronyl group onto a variety of substrates. This is potentially a very useful enzyme, but the kinetic properties of this glucuronosynthase cannot be described as spectacular. Improving the properties of this enzyme through directed evolution is the project of another student in the laboratory. My project was to identify β -GUS variants with altered substrate specificity and to convert the best of these to glycosynthases, via the same mutation that converted β -GUS into a glucuronosynthase.

The initial approach to altering the substrate specificity of β -GUS was to change selected residues surrounding the active site and to monitor the effects on a number of substrates. These residues were subjected to site saturation mutagenesis to identify variants that had reasonable activity on non-physiological substrates. Residues were selected by examination of the structure of β -GUS with bound glucaro- δ -lactam – a reasonably good model for bound substrate. Two sets of residues were investigated, primary (glycosyl binding residues) and secondary (non-glycosyl binding residues). It had been observed that most of the selected glycosyl binding residues are not suitable for our purposes as the activities of these variants toward *p*NP-glucuronide, *p*NP-glucoside, *p*NP-galactoside, *p*NP-mannoside and

*p*NP-xyloside substrates were quite low (< 80 % of the activity of the starting enzyme). In contrast, different substitutions of non-glycosyl binding residues were observed to enhance the catalytic activity with particular substrates, i.e. T509A and T509S mutations demonstrating increased glucosidase and galactosidase activities while mutant S557P showed higher galactosidase activities. Even though this approach met with a small amount of success, there was still considerable room for improvement in the hydrolytic activity with all of the tested substrates. None of the variants seemed to be suitable starting points for the anticipated directed evolution studies.

The variants produced in the above studies did give some insight into the structure function relationship of β -GUS. Residues that contact the substrate seemed to have a vital importance for the enzyme and substitutions gave rise to variants with little activity. These residues included those that made contact with the acid group of glucuronic acid that is not present in any of the other substrates that were tested. The reasons underlying this observation were never entirely clear and considerable time and effort would have been required to address this issue. It is likely that active site residues are important for binding substrate and maintaining the integrity of the active site. The appropriate tool to test this hypothesis was X-ray crystallography, but unfortunately a procedure for producing suitable crystals could not be found; a published protocol (Wallace *et al.* 2010) for producing diffraction quality crystals gave only amorphous precipitates.

Residues close to the active site, but which did not appear to make contact with the substrate, produce reasonable activity with non-physiological substrates and some of these same variants were later identified in the directed evolution studies described in Chapter 4. Some of the variants identified in Chapter 3 were also found in directed evolution studies carried out in other laboratories (Matsumura & Ellington 2001). These later studies were directed towards finding variants that were more stable than the native enzyme. The topic of stability often arises in directed evolution studies and not just for those studies that are aimed at improving stability. Increased stability often occurs in studies where the aim is to improve activity; leading to the idea that stability enhances evolvability. The word “evolvability” is usually used to describe the ability of an organism to adapt to changing conditions. When applied to the evolution of an enzyme, the changing condition implies the selection condition – usually an increase in activity. If evolvability is taken to mean a process that gives increased activity, then the link between evolvability and stability is really a link between activity and stability. Chapter 4 of this thesis was intended to be a more systematic attempt to study the link between stability and activity. In this case, the activity was not with the physiological

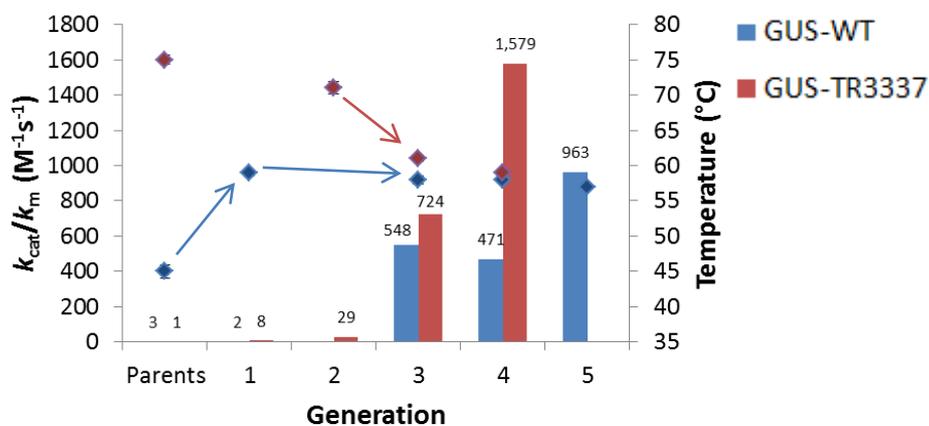
substrate of the enzyme as β -GUS had already evolved to the point that it was close to the diffusion limit. In addition, the presence of OTG in crude extracts meant that the evolution had to overcome the presence of a weakly bound inhibitor. This later aspect of the experiments was unintended and took some effort to identify, as outlined in Chapter 5. However, the presence of OTG meant that the laboratory evolution might mimic the way enzyme regulations evolve in nature. Specifically, β -GUS evolved such that the final variants show much reduced activity in the absence of OTG; this reagent has become an activator of the variant enzymes.

Proteins from thermophilic organisms generally show not only greater thermal stability, but also stability towards chemical denaturants and proteolysis (Vieille & Zeikus 2001, Leuschner & Antranikian 1995, Jaenicke 1991). Since not all enzymes are presently found in thermophiles and thermophilic proteins sometimes do not fold into the native conformation when expressed in *E. coli*, the overall stability of a target enzyme can be raised with directed evolution. However, it had previously been observed that the emergence of a desirable trait in an enzyme is accompanied by an evolutionary cost in some other properties, such as stability, activity or specificity (Antikainen *et al.* 2003, Celenza *et al.* 2008, Valderrama *et al.* 2002). Even though it is virtually impossible to predict these costs, it is widely believed that improvements in enzyme activity through protein engineering often come at the cost of reduced stability. For this reason, a stable enzyme was thought to evolve a function more efficiently than a similar, but less stable enzyme. This hypothesis has been demonstrated experimentally, by comparing the evolvability of marginally stable and thermostable variants of cytochrome P450 BM3 (Bloom *et al.* 2006). However, variants were only examined through one round of directed evolution. In general, directed evolution uses multiple generations of mutation and selection to improve or alter biochemical functions of proteins. In our study, we compare the evolvability of wild-type and thermostable variants of β -GUS for up to 5 rounds of directed evolution. The laboratory evolved version of the GUS-WT, that is GUS-TR3337, not only exhibited higher thermostability (Xiong *et al.* 2007) but also showed better resistance to chemical denaturants and mutations than wild type β -GUS. These comparative experiments show that, even though the stable GUS-TR3337 protein started with a lower glucosidase activity at room temperature relative to GUS-WT, the best variants obtained during each generation from GUS-TR3337 library achieved higher glucosidase activity and tolerance towards OTG than those from GUS-WT library.

By measuring the stability of the evolved enzyme and their precursors, it was demonstrated that the effect of a mutation was depended on the stability of the parent, i.e.

the stability of GUS-WT was first raised by introducing the stabilizing T509A mutation in WT1P17G2 to allow for the subsequent discovery of the functionally beneficial but destabilizing mutation, K568Q, in WT3P24E11 (Figure 7.1). Mutations that improve both properties simultaneously are rare, however, the mutation T509A was found not only to increase the stability of β -GUS, but also increase glucosidase activity. Similar to the results obtained with GUS-WT, the highly thermostable THERMO2P23C4 could tolerate the destabilizing K568E mutation even though this was accompanied with a loss of 2 thermostabilizing mutations (S550N and S566N) in THERMO3P24F7.

(a)



(b)

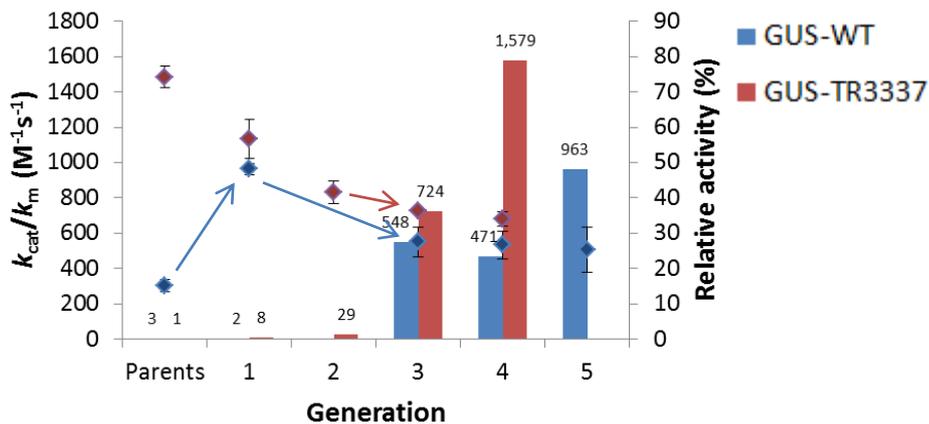


Figure 7.1 Activity and stability changes during the directed evolution of GUS-WT and GUS-TR3337 for activity on β -glucoside

(a) The changes in glucosidase activity and protein thermostability (b) The changes in glucosidase activity and chemical stability at 2 M urea.

The stabilities of evolved variants were measured before the effects of OTG were established. Clearly, the stability of native and variant proteins needs to be measured in the

presence of OTG. The time constraints of an Australian PhD did not allow these measurements to be repeated before this thesis was written. It is intended that these measurements will be made after thesis submission so that a manuscript can be submitted for publication.

In most directed evolution experiments, cell lysis is an important step for releasing enzyme into the supernatant for the subsequent assays to determine protein content and enzyme activity. Common cell disruption methods include mechanical (i.e. sonication, French press, bead milling), chemical (i.e. alkali, detergents), and enzymatic (i.e. lysozyme) methods (Chisti & Moo-Young 1986). Physical methods are not suitable for high throughput experiments while enzymatic methods are expensive. Since detergent treatment is a relatively inexpensive and effective method; and suited for high-throughput screening assays, lysis with the BugBuster™ reagent was applied in our study. However, it had led to unexpected impacts on the output of the screening process due to the presence of molecules that are structurally similar to the native substrate of β -GUS. One lesson that was learnt from the use of BugBuster™ reagent, was that reagents of unknown composition should not be used in experiments. If they are to be used a test for negative effects of the reagent should be made. In the experiments described in Chapter 4, a comparison of the activity obtained with BugBuster™ reagent and R-lysozyme would have revealed a drop in the activity due to BugBuster™ reagent.

During screening process, OTG was added to assays unintentionally as part of the BugBuster™ reagent. This eventually resulted in variants with glucosidase activity that was activated by OTG. By examining the effects of mutations, the mutation at K568 residue in particular appears responsible for restoring the kinetic model from double Michaelis-Menten kinetics to single Michaelis-Menten kinetics (non-cooperativity) in the presence of excess OTG. To our knowledge, this is the first time such kinetic behavior was reported for a tetrameric enzyme with a single ligand-binding site one each subunit. The sum of two Michaelis-Menten model often represents two independent enzymes operating in the range of assayed substrate concentrations and it therefore prompted us to launch an investigation into whether the variants contain two kinetic forms due to the existence of different oligomeric forms.

Unfortunately, no direct link between the biphasic kinetics and different oligomeric forms of enzymes was observed. However, future work should determine if OTG modulates the oligomeric stage and thus the activity of evolved enzymes. Much effort has been directed towards investigating the reason for the loss of glucosidase activity of evolved variants in the

absence of OTG. There was insufficient time available in this project to investigate various effects of OTG on evolved variants. For example, it seems likely that evolution abolished the inhibitory effect of OTG, but it still has an effect on the later generation variants. This observation suggests that OTG is bound to the later generation variants. Binding of OTG to variants can be verified and quantitated the OTG with isothermal titration calorimetry.

It is also worth considering whether variants can use OTG as a substrate. This is not likely as OTG instead of OG as the latter reagent is subject to enzymatic hydrolysis while the former does not. However, one way of dealing with the inhibitory effects of OTG might be to produce variants that can break the β -thioglucoside linkage. This may explain the abolition of competitive binding by OTG. That is, the cleavage of β -thioglucoside linkage by evolved variant would release D-glucose from the active site. Again, these experiments should be done in the future.

Despite the need to perform additional experiments, the results obtained to-date are generally positive regarding the alteration of the substrate specificity of β -GUS. Furthermore, the evolution of the thermostable variant was proceeding more rapidly than the native β -GUS; as expected. The evolution was particularly impressive given that it overcame the presence of OTG. Although the presence and effects of OTG were not anticipated, it was of considerable interest. The evolution introduced the capacity to regulate the β -GUS. This is possible similar to the route by which enzyme regulation evolves in nature.

Lastly, the E504G mutation was introduced into evolved variants in order to test their synthetic capability. As it turned out, steroidal glucosides were not detected after the enzymatic reaction. This was a negative result and unfortunately there was insufficient time to optimize the experimental conditions, nor was there time to test other substrates. The intent of this thesis was first to improve the glucosidase activity of the β -GUS followed by converting the hydrolysis activity into the glycosylation activity. While the evolved glucosidase activity remained much lower than the native enzyme's activity on its physiological substrate, the catalytic efficiency of GUS-WT for glucuronidase activity was 342 fold higher than THERMO4P11F2 for glucosidase activity, making it possible that the converted glycosylation activity was too low to be detected. Further evolution of the glucosidase activity of the best variants could be carried out in the future. Alternatively, attempts could be made to apply directed evolution on the glucuronosynthase directly. These experiments may form the subject of other research projects.

7.1. References

- Antikainen, N. M., P. J. Hergenrother, M. M. Harris, W. Corbett & S. F. Martin (2003) Altering substrate specificity of phosphatidylcholine-preferring phospholipase C of *Bacillus cereus* by random mutagenesis of the headgroup binding site. *Biochemistry*, 42, 1603-10.
- Bloom, J. D., S. T. Labthavikul, C. R. Otey & F. H. Arnold (2006) Protein stability promotes evolvability. *Proc Natl Acad Sci U S A*, 103, 5869-74.
- Celenza, G., C. Luzi, M. Aschi, B. Segatore, D. Setacci, C. Pellegrini, C. Forcella, G. Amicosante & M. Perilli (2008) Natural D240G Toho-1 mutant conferring resistance to ceftazidime: biochemical characterization of CTX-M-43. *J Antimicrob Chemother*, 62, 991-7.
- Chisti, Y. & M. Moo-Young (1986) Disruption of microbial cells for intracellular products. *Enzyme and Microbial Technology*, 8, 194-204.
- Jaenicke, R. (1991) Protein stability and molecular adaptation to extreme conditions. *Eur J Biochem*, 202, 715-28.
- Leuschner, C. & G. Antranikian (1995) Heat-stable enzymes from extremely thermophilic and hyperthermophilic microorganisms. *World J Microbiol Biotechnol*, 11, 95-114.
- Matsumura, I. & A. D. Ellington (2001) In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. *J Mol Biol*, 305, 331-9.
- Valderrama, B., M. Ayala & R. Vazquez-Duhalt (2002) Suicide inactivation of peroxidases and the challenge of engineering more robust enzymes. *Chem Biol*, 9, 555-65.
- Vieille, C. & G. J. Zeikus (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev*, 65, 1-43.
- Wallace, B. D., H. Wang, K. T. Lane, J. E. Scott, J. Orans, J. S. Koo, M. Venkatesh, C. Jobin, L. A. Yeh, S. Mani & M. R. Redinbo (2010) Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science*, 330, 831-5.
- Xiong, A. S., R. H. Peng, Z. M. Cheng, Y. Li, J. G. Liu, J. Zhuang, F. Gao, F. Xu, Y. S. Qiao, Z. Zhang, J. M. Chen & Q. H. Yao (2007) Concurrent mutations in six amino acids in beta-glucuronidase improve its thermostability. *Protein Eng Des Sel*, 20, 319-25.

Appendices

A. Chapter 1

A.1. Nomenclature

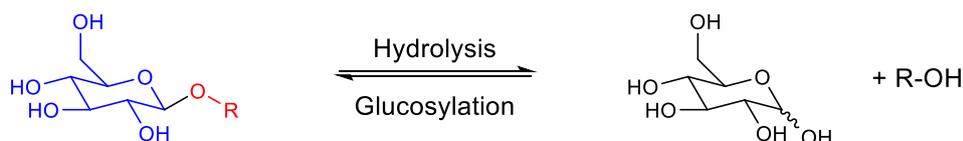
A summary of the nomenclature used for the chemical compounds and the main reactions that will be encountered throughout this thesis is provided below for convenience of the reader. The substrates used in this thesis are structurally complex and are related to a number of stereoisomers. This thesis follows the nomenclature outlined in the International Union of Pure and Applied Chemistry (IUPAC) guidelines, though trivial names will be used where possible for brevity.

A.1.1. Reactions

Glycosides consist of a sugar residue (glycone) covalently bound to a different structure called the aglycone. The sugar moiety can be joined to the aglycone via an oxygen atom (*O*-glycoside), sulphur (*S*-glycoside), nitrogen (*N*-glycoside) or carbon (*C*-glycoside). Specifically in this thesis, glycosides consist of a glycosyl and a general aglycone component, that are linked by an *O*-glycosidic bond. In naming of glycosides, the “ose” suffix of the sugar name is replaced by “oside”. For example, if the glycone group of a glycoside is glucose, then the molecule is a glucoside; if it is glucuronic acid, then the molecule is a glucuronide; etc.

The hydrolysis of glycosides will generate a free sugar moiety and an aglycone moiety. The reverse process, conjugation or glycosylation, results in the synthesis of a glycosidic bond. Among various glycosylation reactions, the main focus in this thesis is on glucosylation and glucuronylation, which involve addition of a glucose or glucuronic acid to an aglycone. The reactions are shown in Figure A.1 and the nomenclature for the two components in a glycoside is detailed in the next two sections.

(a)



(b)

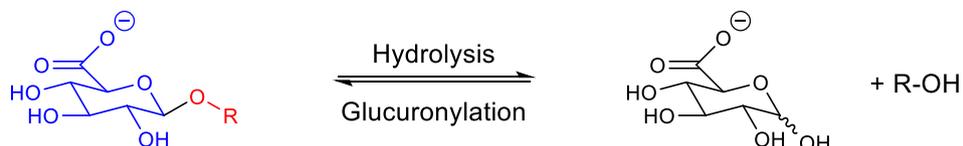


Figure A.1 Reactions that are discussed in this thesis are the hydrolysis and synthesis of glycosides

A glycoside consists a glycosyl group (blue) and an aglycone component (red). (a) Glucosylation (b) Glucuronylation.

A.1.2. Glycosyl groups

Following the guidelines set out by McNaught (1997), the numbering for the monosaccharides begins at the aldehyde chain in the Fisher projection (Figure A.2). The configuration descriptor, D or L, is determined by the stereochemistry at the “highest-numbered center of chirality” which exists at C5. Therefore, with the hydroxyl group (red) positioned to the right, this is a “D” sugar; conversely, with the hydroxyl group (blue) positioned to the left this is an “L” sugar. They are enantiomerically related but the “D” form is the more common of the two in nature. It is the form that is applicable to this thesis. Therefore, only “D” sugars were represented in Figure A.1.

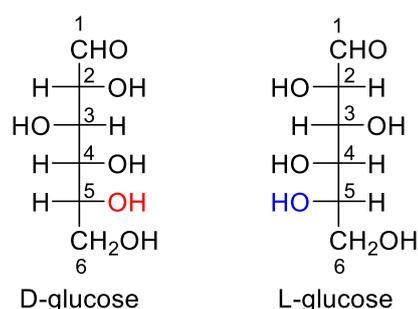


Figure A.2 The Fischer project of D-glucose and L-glucose

Monosaccharides usually occur in aqueous solution as cyclic structures, in which the C5 hydroxyl group forms an intramolecular bond with the C1 aldehyde. The newly-produced center of chirality at C1 is designated as the anomeric carbon, as shown in Figure A.3. The stereochemistry at the C1 is defined by prefixes “ α ” or “ β ” depending on its relative

stereochemistry to C5. In the β -anomer, the ring oxygen at C5 (red) is formally *trans* to the C1 oxygen (blue), while in the α -anomer the relationship is formally *cis*.

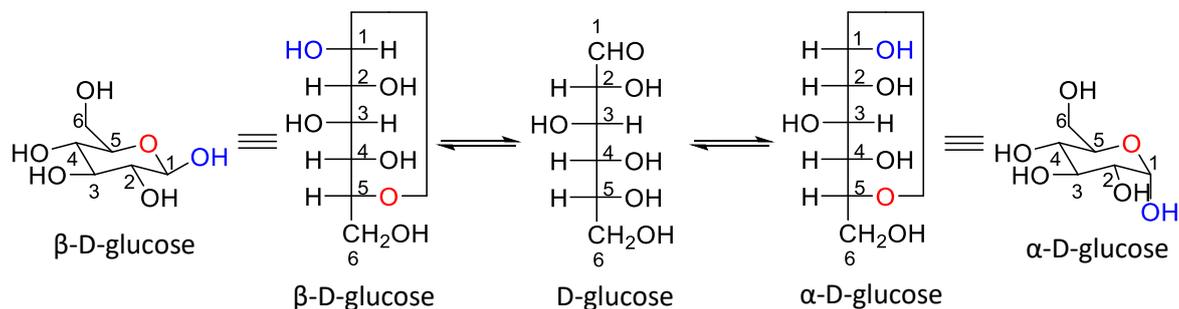


Figure A.3 Glycosyl groups: The glucuronyl acid is essentially a glucose oxidized at the C6 position

The glycosyl groups on glycosides that relates to this study in a hydrolysis reaction, are β -D-glucuronic acid and β -D-glucose. These are hydrolysis products that will be frequently presented throughout this thesis. Most glycosynthases derived from β -retaining glycoside hydrolases, i.e. β -glucuronidase (β -GUS), prefer α -glycosyl-fluorides donors and proceed through a single-displacement mechanism ($\alpha \rightarrow \beta$) (Perugino *et al.* 2004). Therefore, the glycosyl donors used for this work (Chapter 6) in a glycosylation reaction, are α -D-glycosyl fluoride and α -D-glucuronyl fluoride, and their glycone groups are α -D-glucose and α -D-glucuronic acid, respectively (Figure A.4). Their C1 carbon carries a fluorine atom instead of a hydroxyl group. The reaction mechanisms of β -GUS and glucuronylsynthase will be discussed in detail in Chapter 1.

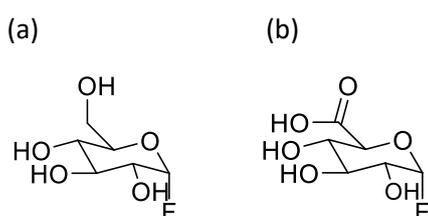


Figure A.4 β -D-glycosyl fluoride
(a) β -D-glucosyl fluoride (b) β -D-glucuronyl fluoride.

With regards to ring size, the term “pyran” refers to the six-membered cyclic sugars, collectively known as the pyranose. The sugars discussed in this thesis will only cover such sugars and no other ring sizes will be examined. For the purpose of this thesis, the trivial names for the common monosaccharides were used, i.e. glucose and glucuronic acid.

A.1.3. Aglycone groups

The other component in the β -D-glucuronide is the aglycone. This is either the leaving group in the hydrolysis reaction or an acceptor substrate for the conjugation reaction. *Para*-nitrophenol is the leaving group in the substrate used to monitor hydrolysis in this thesis. In glucuronylation, the two main acceptors used for this work are steroids where the center for the conjugation reaction happens at a hydroxyl group. They are dehydroepiandrosterone *O*-(carboxymethyl)oxime (CMO-DHEA) and testosterone (Figure A.5). Since the systematic names are tedious to present, the common names and abbreviations will be used.

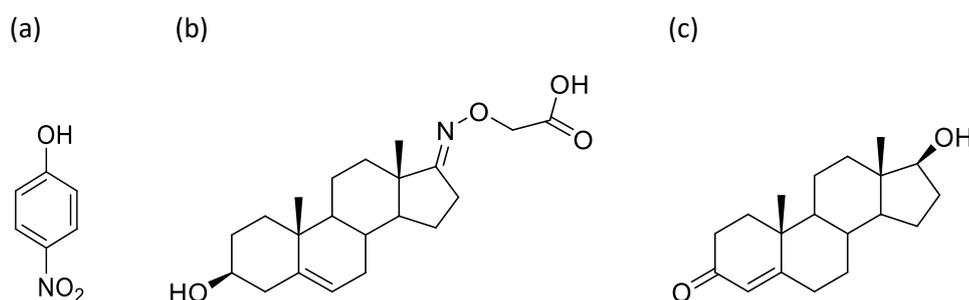


Figure A.5 Aglycones discussed

Acceptors of the leaving group in the hydrolysis reaction or glucuronyl donors in glucuronylation reactions. (a) *Para*-nitrophenol (b) CMO-DHEA (c) Testosterone.

A.2.

Considering ePCR introduces point mutations at the DNA level, on average just 5.7 of the potential 19 alternative amino-acids are accessible by a single nucleotide change (Miyazaki & Arnold 1999). Therefore, theoretical library size is calculated by equation below which was discussed by (Patrick *et al.* 2003):

$$V = 5.7^x N! / [x! (N-x)!]$$

Where V is the total number of possible variants, N is the length of template sequence and x is the average mutation frequency.

To achieve 95 % library coverage, the sampling factor is about three. This is calculated by:

$$F = 1 - e^{-L/V}$$

Where F is the fraction of completeness of the library, L is the library size and V is the total number of possible sequence variants. To construct a library expected to contain 95 % of all possible sequence variants, then $F = 0.95$, therefore:

$$0.95 = 1 - e^{-L/V}$$

$$L = -V \ln 0.05$$

$$\approx 3 V$$

Which means three fold excess of transformants needs to be screened.

Table B.1 Library sizes and the oversampling size required to achieve 95 % library coverage of a 600 amino acid in directed evolution experiments.

Average mutation frequency	Theoretical library size (number of variants)	95 % coverage (number of variants)
1	3.42×10^3	1.02×10^4
2	5.84×10^6	1.75×10^7
3	6.63×10^9	1.98×10^{10}

A.3. References

- McNaught, A. D. (1997) Nomenclature of carbohydrates. *Carbohydrate Research*, 297, 1-92.
- Miyazaki, K. & F. H. Arnold (1999) Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. *J Mol Evol*, 49, 716-20.
- Patrick, W. M., A. E. Firth & J. M. Blackburn (2003) User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Eng*, 16, 451-7.
- Perugino, G., A. Trincone, M. Rossi & M. Moracci (2004) Oligosaccharide synthesis by glycosynthases. *Trends Biotechnol*, 22, 31-7.

B. Chapter 2

B.1. Recipes and Reagents

The recipes and chemical preparations used throughout this project are described below.

Buffer A

Solution of 0.5 M NaCl, 20 mM imidazole and 20 mM phosphate buffer adjusted to pH 7.4.

Buffer B

Solution of 0.5 M NaCl, 500 mM imidazole and 20 mM phosphate buffer adjusted to pH 7.4.

β -Glucuronidase (β -GUS) Activity Indicator Agar

2 g of agar (Difco) is dissolved in 500 mL of water at 100 °C, this solution is then cooled to 70 °C before 2.75 g of disodium phosphate (Ajax) and 0.68 g of monosodium phosphate (Ajax) is added and dissolved to provide a buffer of 50 mM buffer at pH 7.4. This is allowed to cool further, to 60 °C before adding 0.2 g of 5-bromo-4-chloro-3-indolyl- β -D-glucoside (X-glucoside) for 1 mM final concentration.

Luria-Bertani (LB)

For one litre of this media, dissolve 10 g of tryptone (Difco), 5 g of yeast extract (Difco) and 10 g of NaCl in 0.9 L of water and adjust to pH 7 with 250 μ L of 10 M NaOH make up to volume and sterilize by autoclaving.

LB-Kanamycin (LBK) Media

Prepare LB as above and add filter-sterilised ampicillin (Astral) at 50 mg/mL (stored in aliquots at -20 °C) for a final concentration of 50 μ g/mL in the sterilized media.

LB(K) Agar

Prepare LB(K) as above, but include 15 g/L of agar (Difco) before autoclaving.

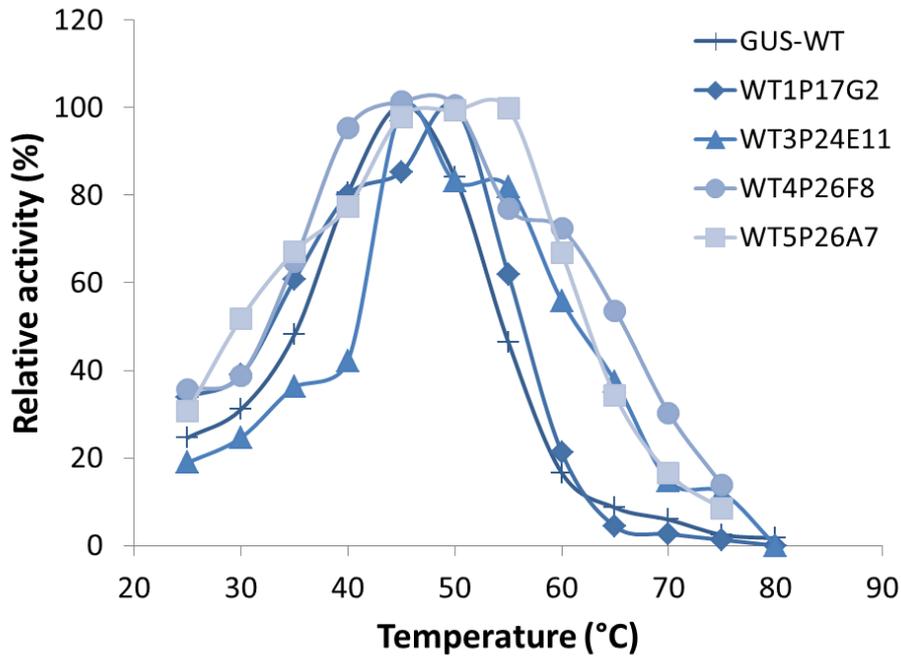
Yeast Extract Nutrient Broth (YENB)

Solution of 7.5 g/L bacto yeast extract (Difco) and 8 g/L bacto nutrient broth (Difco) sterilized by autoclaving.

C. Chapter 4

C.1.

(a)



(b)

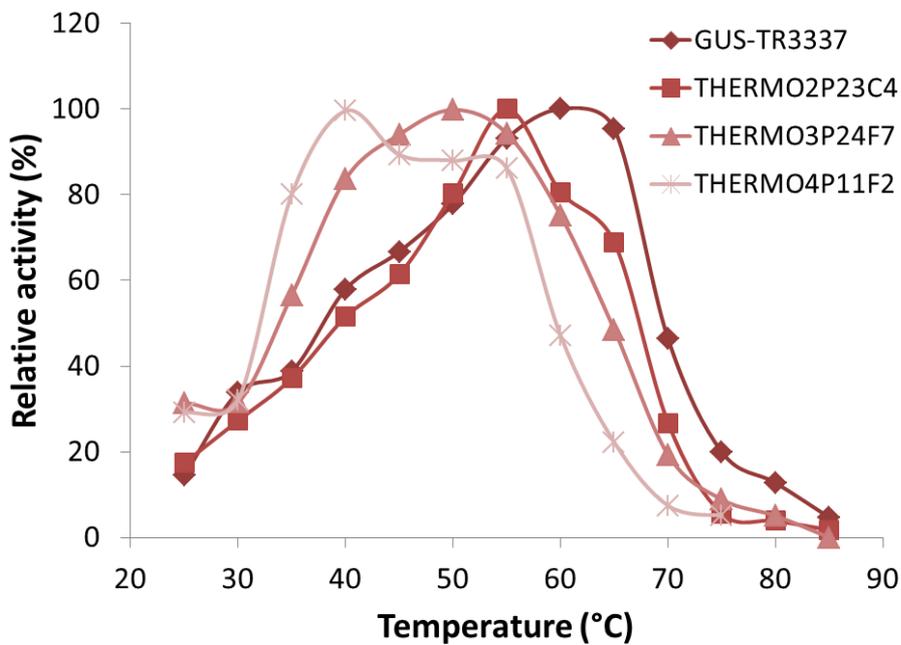
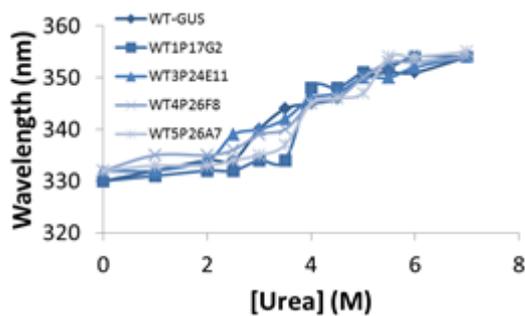


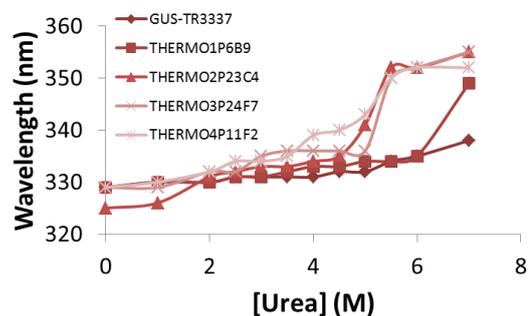
Figure C.1 Temperature optima of GUS-WT and GUS-TR3337 variants
 (a) GUS-WT (b) GUS-TR3337.

C.2.

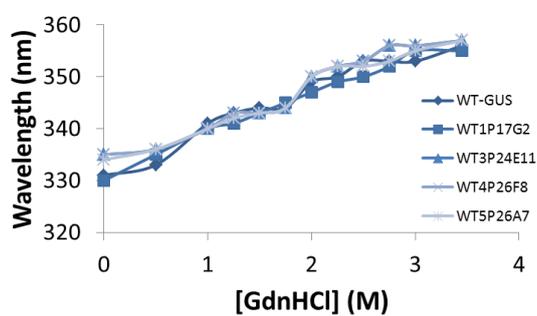
(a)



(b)



(c)



(d)

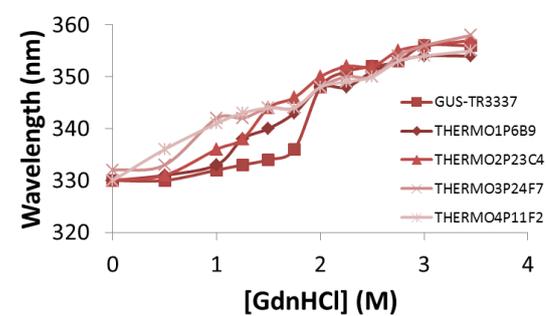
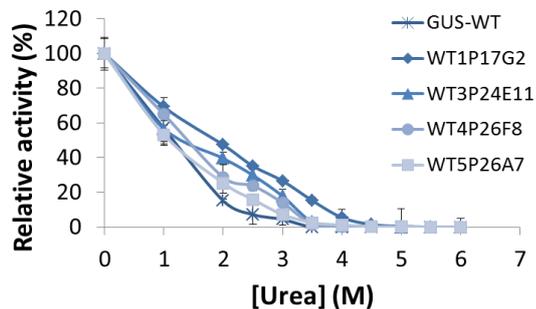


Figure C.2 Chemical induced unfolding is shown by changes in λ_{max}

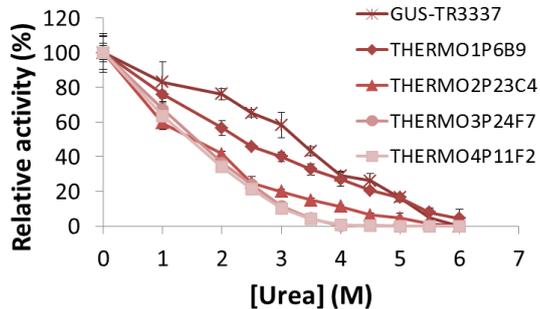
(a) Urea induced unfolding in GUS-WT* (b) Urea induced unfolding in GUS-TR3337* (c) GdnHCl induced in GUS-WT* (d) GdnHCl induced in GUS-TR3337.

C.3.

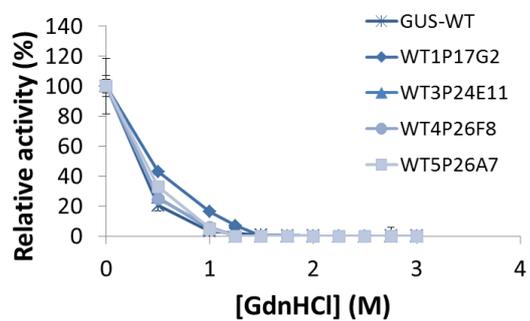
(a)



(b)



(c)



(d)

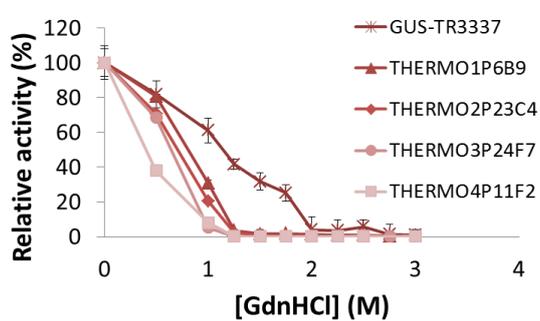
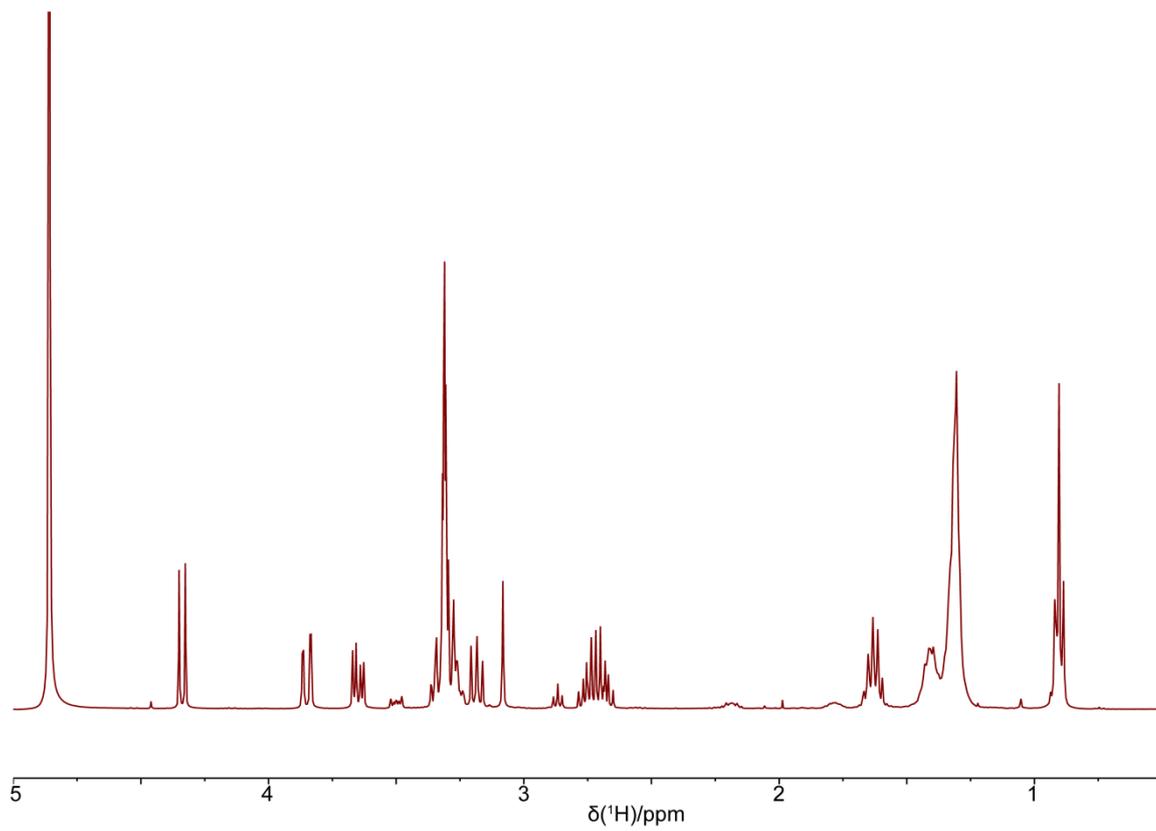


Figure C.3 Impact of the presence of different concentrations of urea and GdnHCl on the glucosidase activity of GUS variants

D. Chapter 5

D.1. Spectra of octyl thioglucoside

(a)



(b)

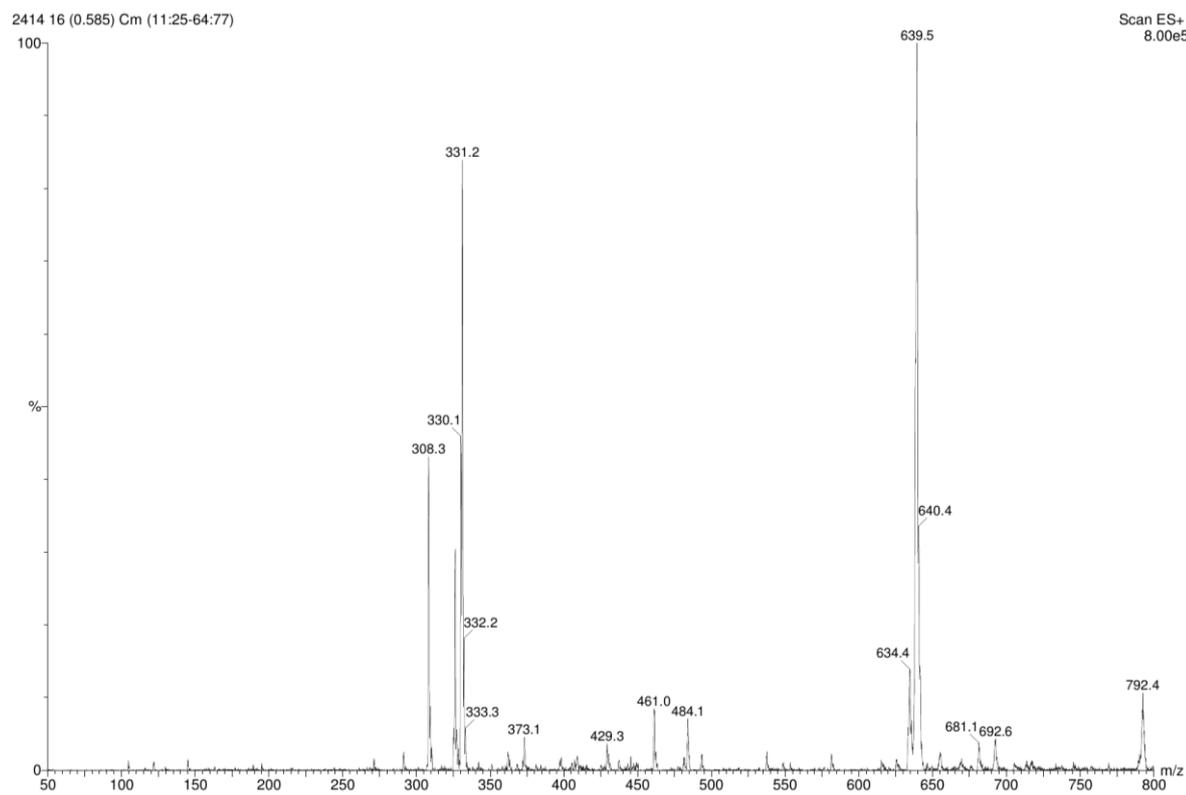
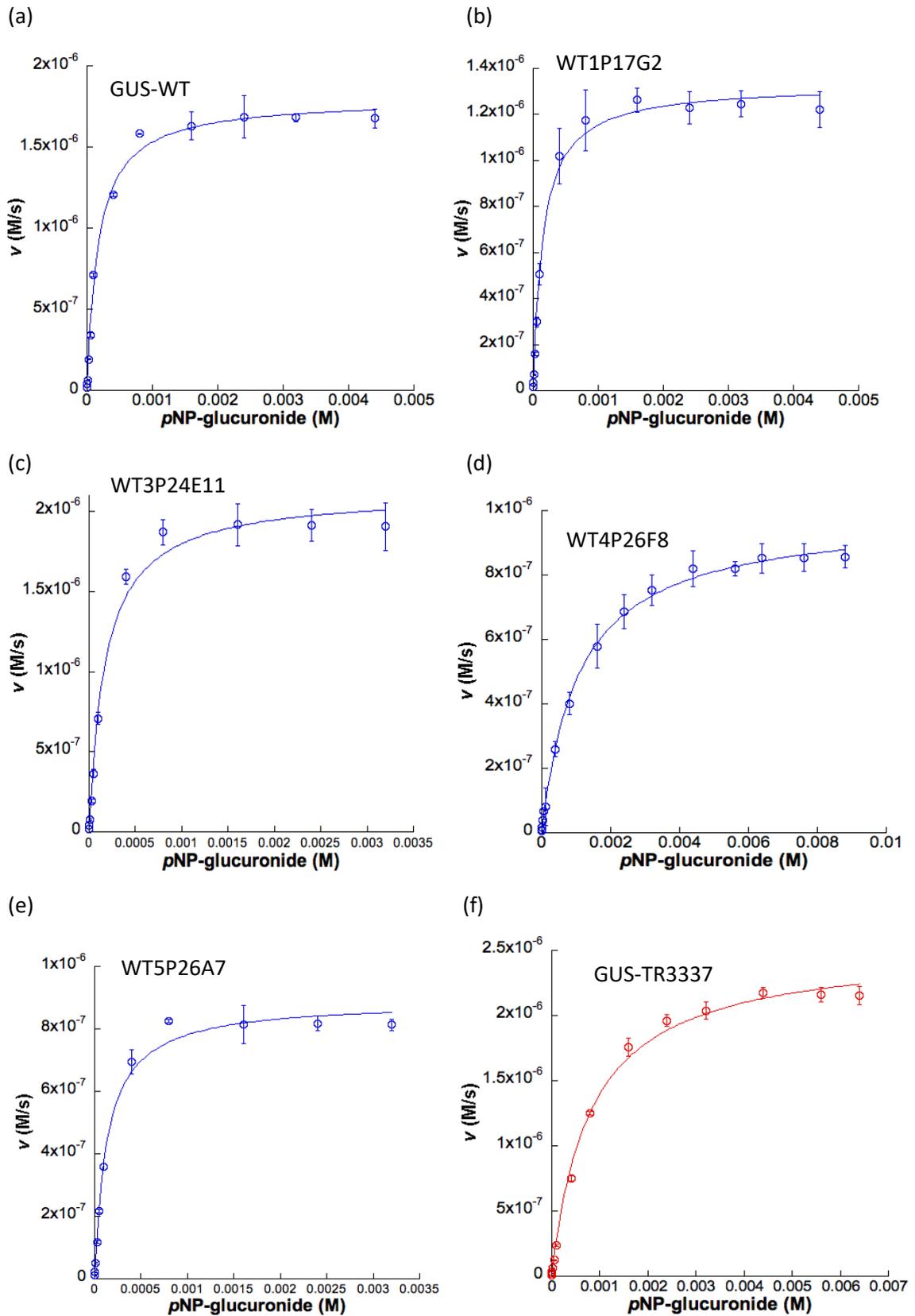


Figure D.1. Spectra of octyl thioglucoside in BugBuster™ reagent (methanol fraction)
(a) NMR spectrum (b) MS spectrum.

The BugBuster™ reagent was subjected to solid-phase purification. A 3 mL Oasis WAX SPE cartridge was pre-conditioned with 1 mL methanol and 3 mL milliQ water. The BugBuster™ reagent was loaded onto the cartridge and washed with 3 mL aqueous formic acid, 3 mL mQH₂O, 3 mL methanol and finally with 9 mL saturated aqueous ammonium hydroxide in methanol.

¹H NMR (400 MHz, CD₃OD): 4.34 (1H, d, *J* 9.7 Hz), 3.85 (1H, dd, *J* 12.1, 2.0 Hz), 3.65 (1H, dd, *J* 11.9, 5.3 Hz), 3.36-3.32 (1H, m), 3.30-3.24 (2H, m), 3.18 (1H, m), 2.72 (2H, m), 2.72 (2H, m), 1.63 (2H, m), 1.43-1.31 (10H, m), 0.90 (3H, m); **LRMS (+ESI)** *m/z* 640 ([2M+Na]⁺), 331 ([M+Na]⁺); **HRMS (+ESI)** *m/z* calculated for C₁₄H₂₈O₅NaS ([M+Na]⁺) 331.1555, found 331.1548.

D.2. Michaelis-Menten plots for the hydrolysis of *p*NP-glucuronide by wild-type and evolved variants



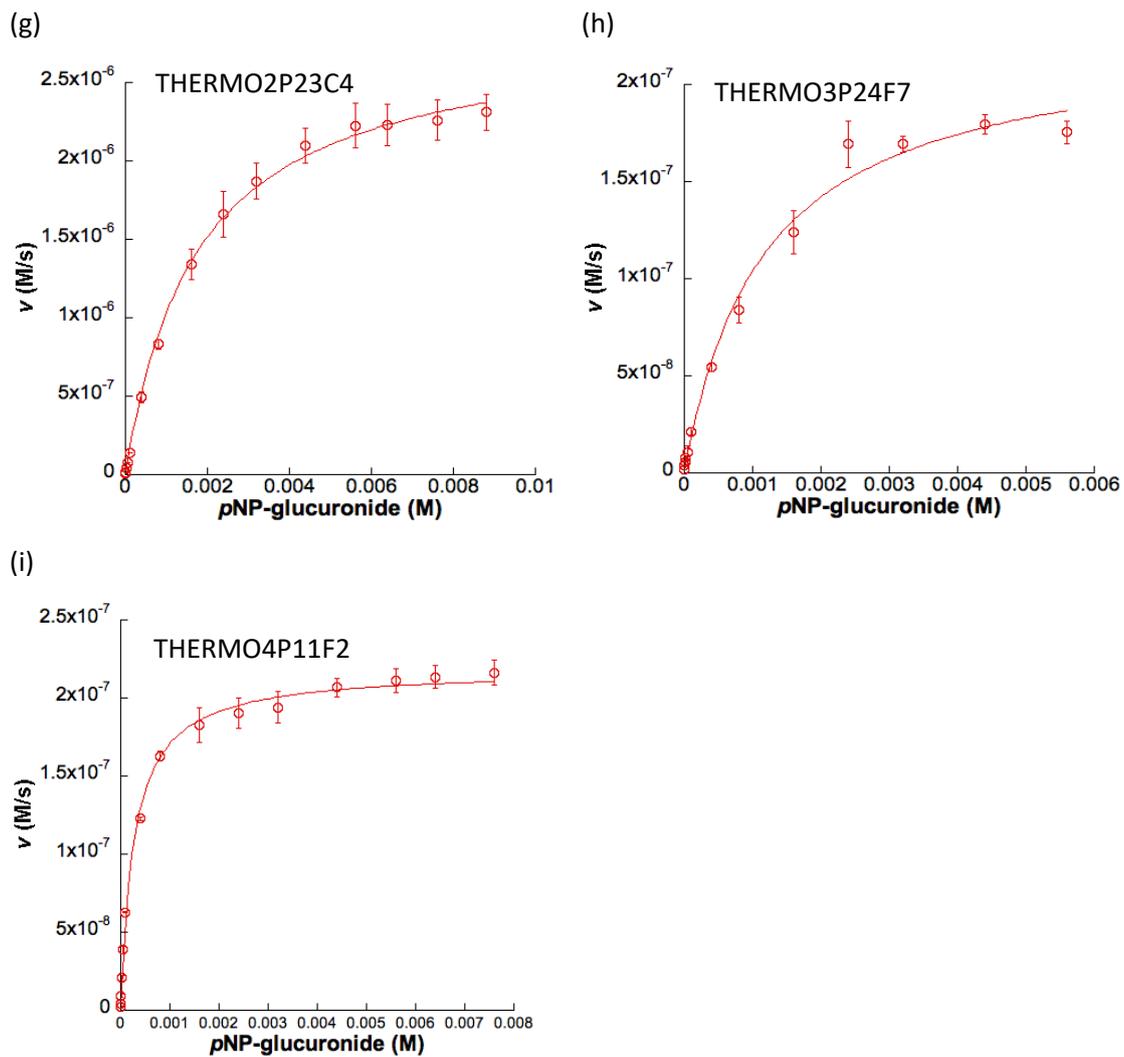
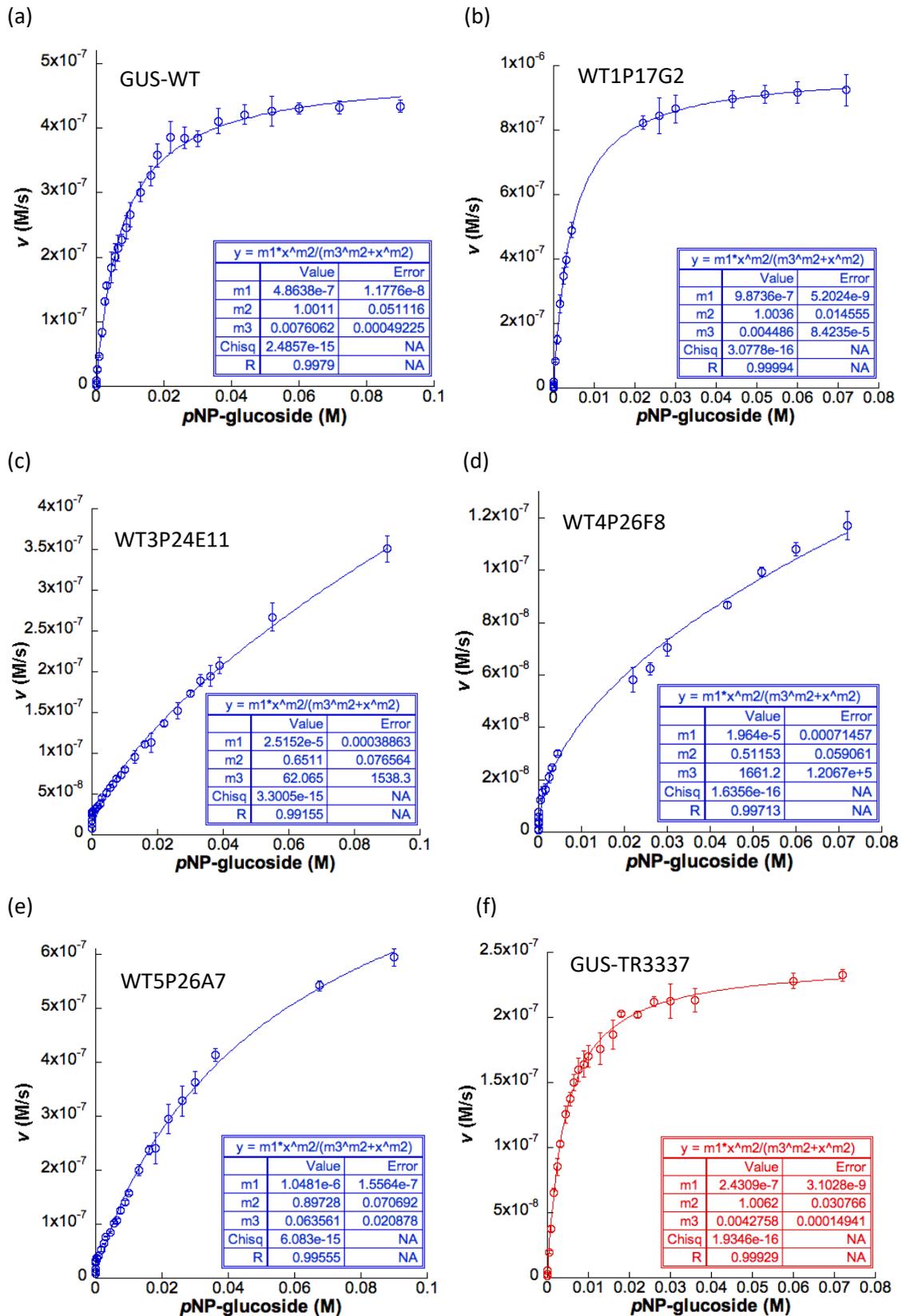


Figure D.2 Michaelis-Menten plots for the hydrolysis of pNP-glucuronide by wild-type and evolved variants

(a) GUS-WT (b) WT1P17G2 (c) WT3P24E11 (d) WT4P26F8 (e) WT5P26A7 (f) GUS-TR3337 (g) THERMO2P23C4 (h) THERMO3P24F7 (i) THERMO4P11F2.

D.3. Michaelis-Menten plots for the hydrolysis of pNP-glucoside by wild-type and evolved variants



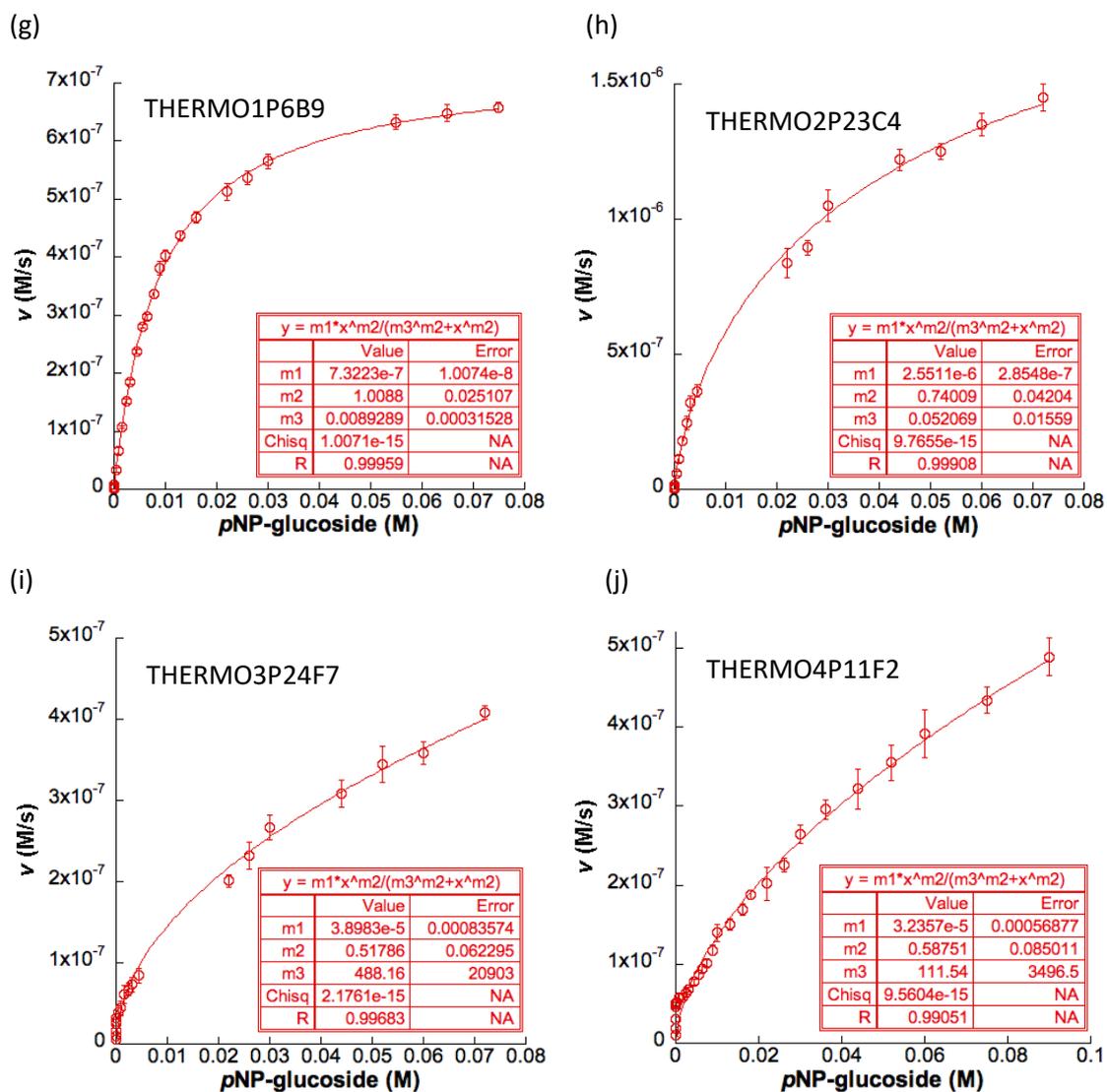
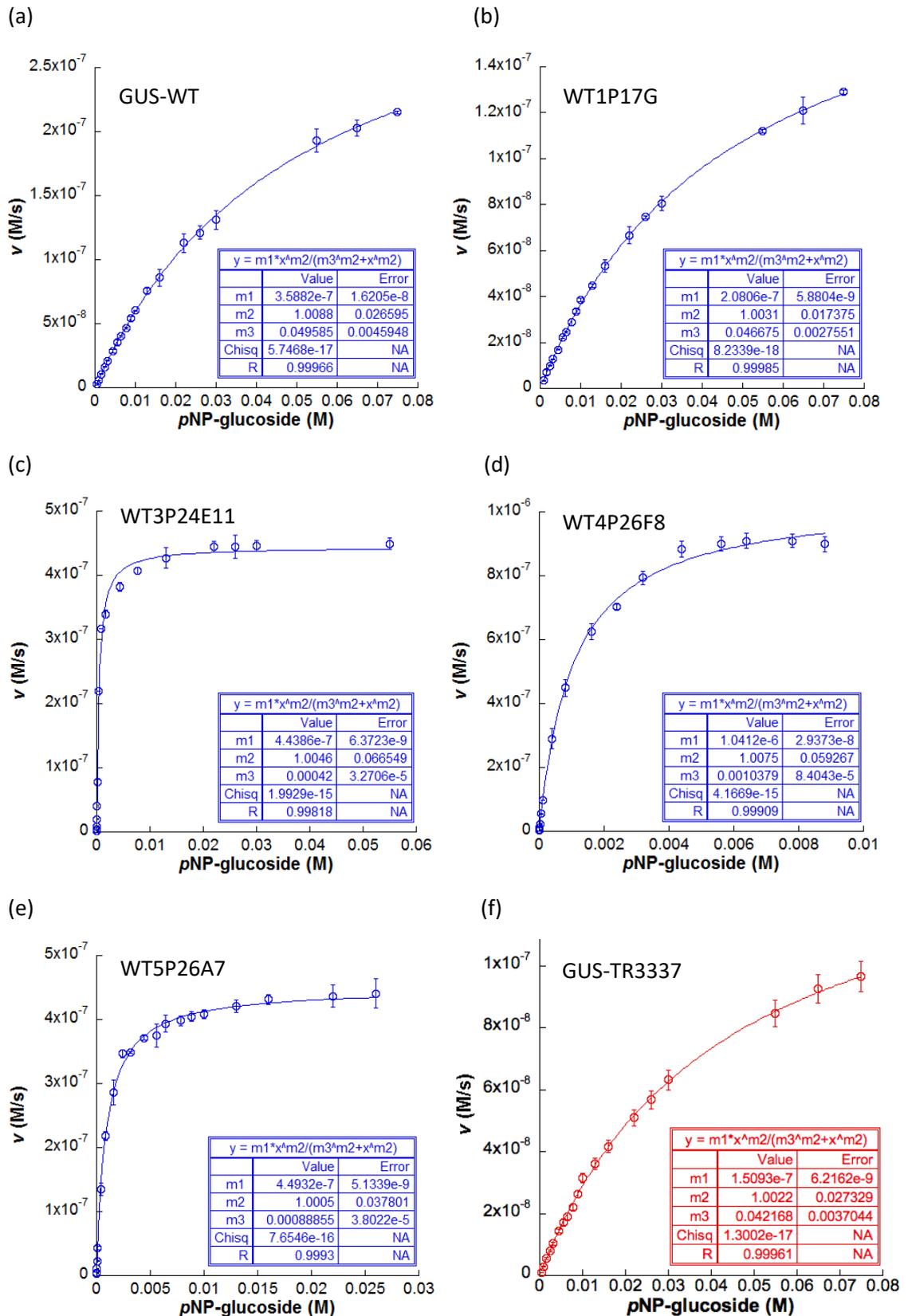


Figure D.3 Michaelis-Menten plots for the hydrolysis of pNP-glucoside by wild-type and evolved variants

Data were fitted to the Hill equation using Kaleidagraph version 4.1 (Synergy software). Parameters for the fitted curves are given (m1: V_{max} ; m2: n ; m3: $K_{0.5}$); R is the correlation coefficient. (a) GUS-WT (b) WT1P17G2 (c) WT3P24E11 (d) WT4P26F8 (e) WT5P26A7 (f) GUS-TR3337 (g) THERMO1P6B9 (h) THERMO2P23C4 (i) THERMO3P24F7 (j) THERMO4P11F2.

D.4. Michaelis-Menten plots for the hydrolysis of pNP-glucoside by wild-type and evolved variants in the presence of 5 mM OTG



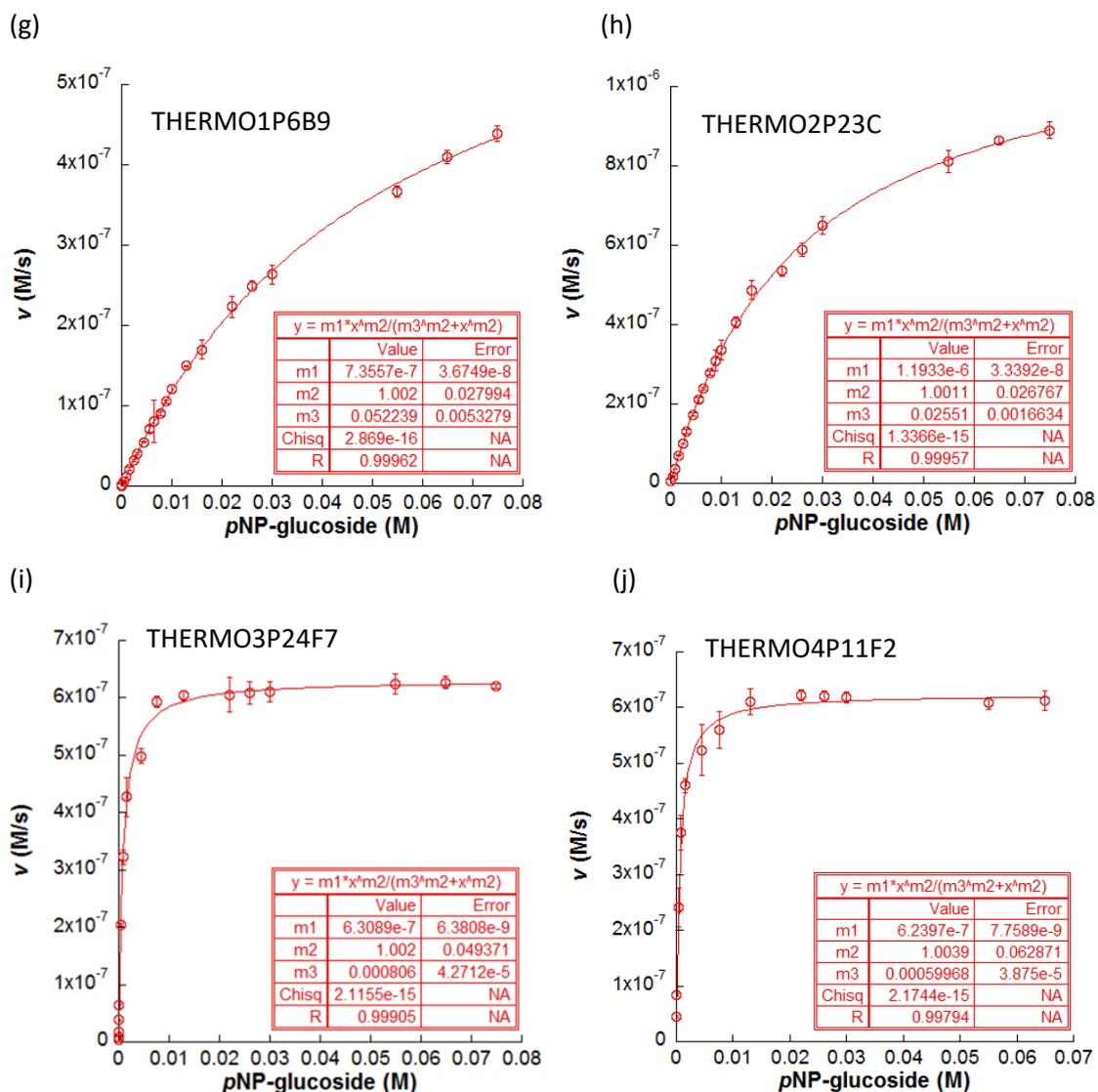
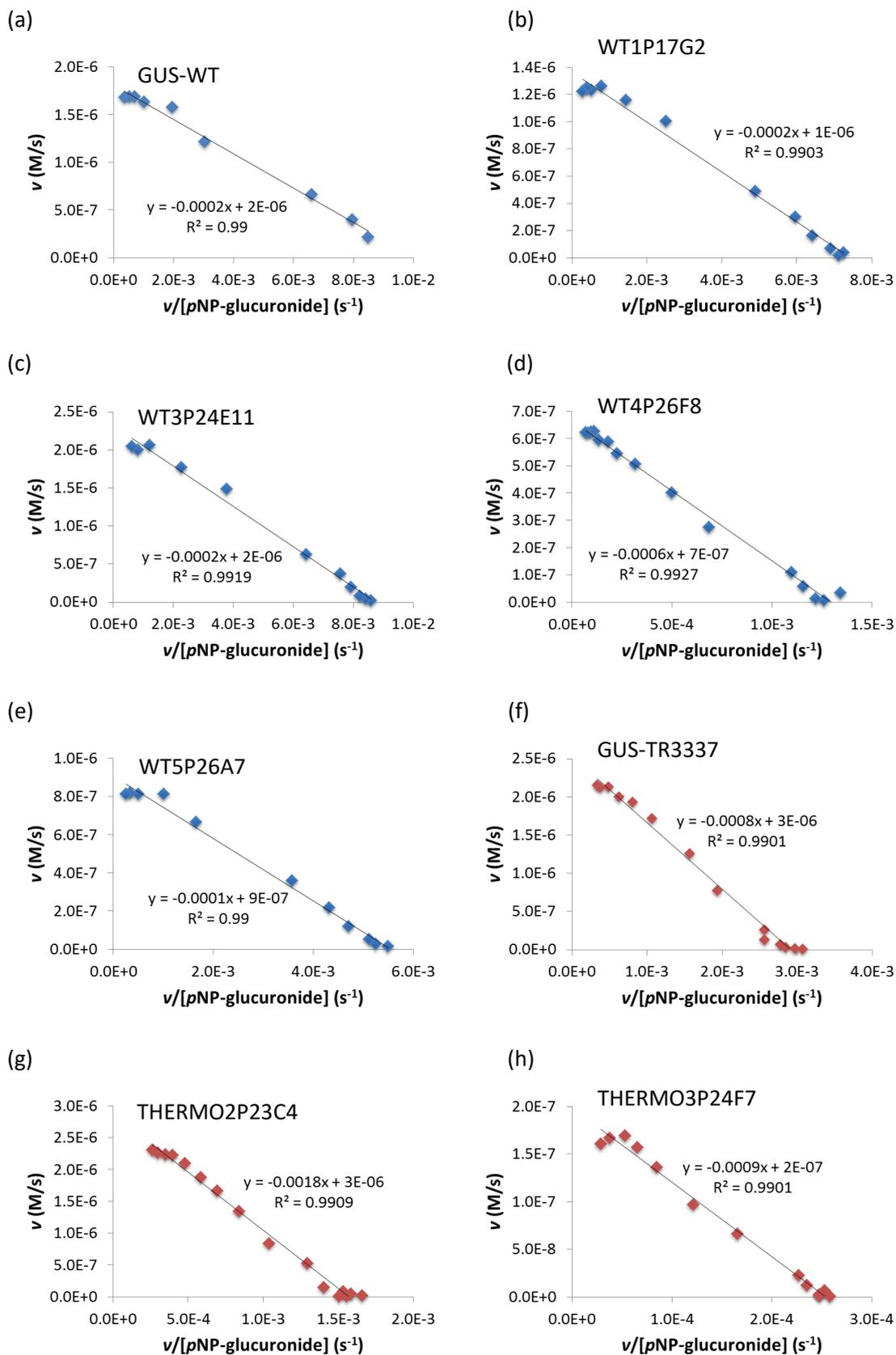


Figure D.4 Michaelis-Menten plots for the hydrolysis of pNP-glucoside by wild-type and evolved variants in the presence of 5 mM OTG

Data were fitted to the Hill equation using Kaleidagraph version 4.1 (Synergy software). Parameters for the fitted curves are given (m1: V_{max} ; m2: n ; m3: $K_{0.5}$); R is the correlation coefficient. (a) GUS-WT (b) WT1P17G2 (c) WT3P24E11 (d) WT4P26F8 (e) WT5P26A7 (f) GUS-TR3337 (g) THERMO1P6B9 (h) THERMO2P23C4 (i) THERMO3P24F7 (j) THERMO4P11F2.

D.5. Eadie-Hofstee plots of the *p*NP-glucuronide hydrolysis by wild-type and evolved variants



(i)

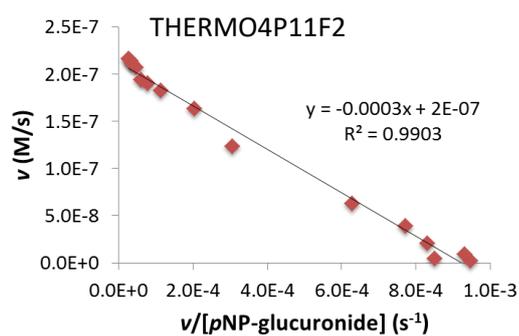
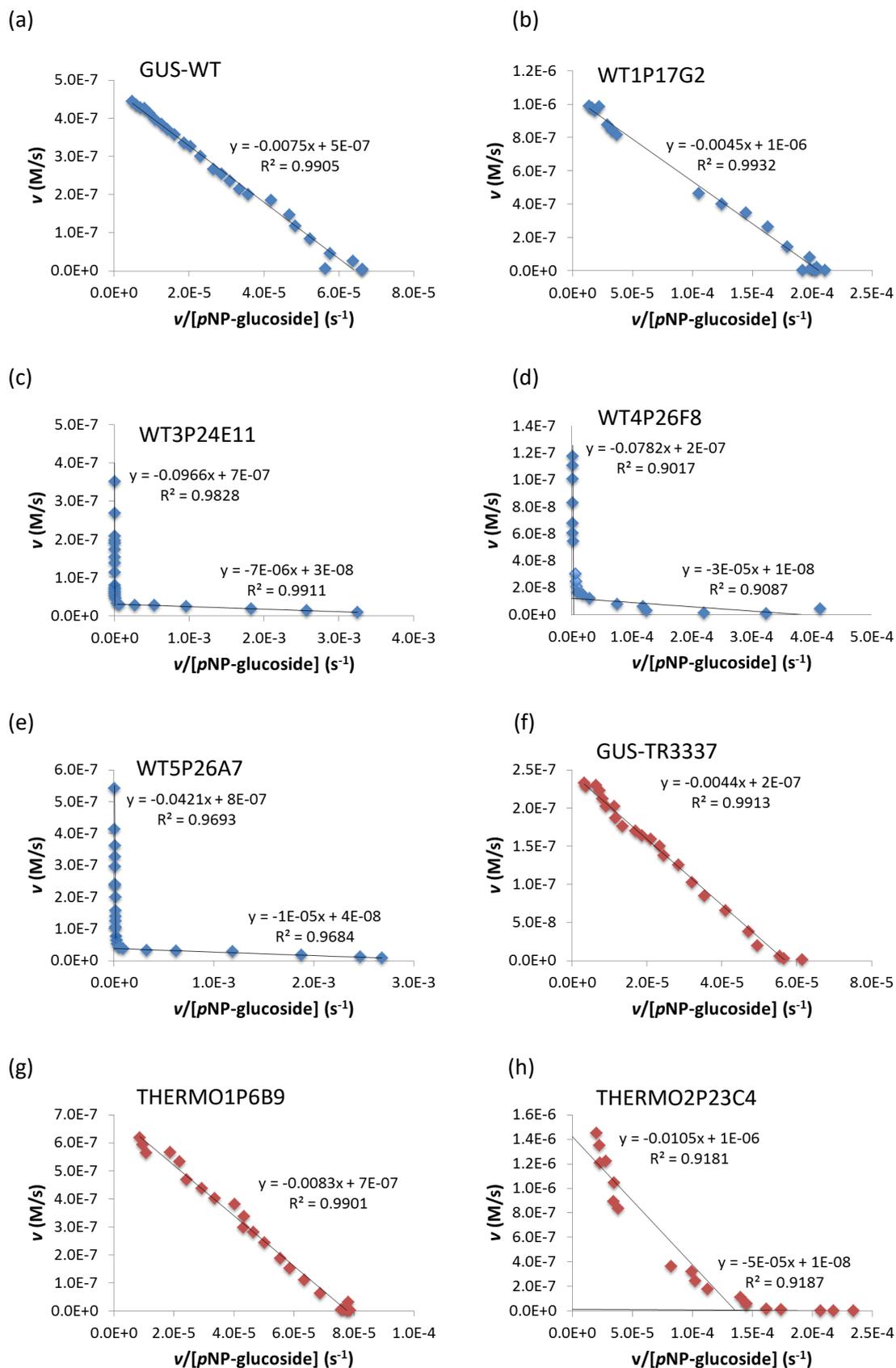


Figure D.5 Eadie-Hofstee plots of the *p*NP-glucuronide hydrolysis by wild-type and evolved variants

Each plot is the mean of triplicates

(a) GUS-WT (b) WT1P17G2 (c) WT3P24E11 (d) WT4P26F8 (e) WT5P26A7 (f) GUS-TR3337 (g) THERMO2P23C4 (h) THERMO3P24F7 (i) THERMO4P11F2.

D.6. Eadie-Hofstee plots of the *p*NP-glucoside hydrolysis by wild-type and evolved variants



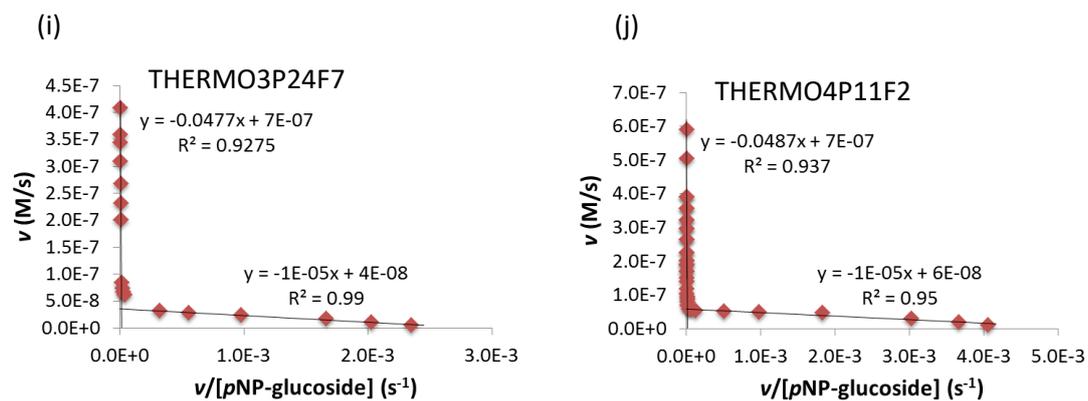
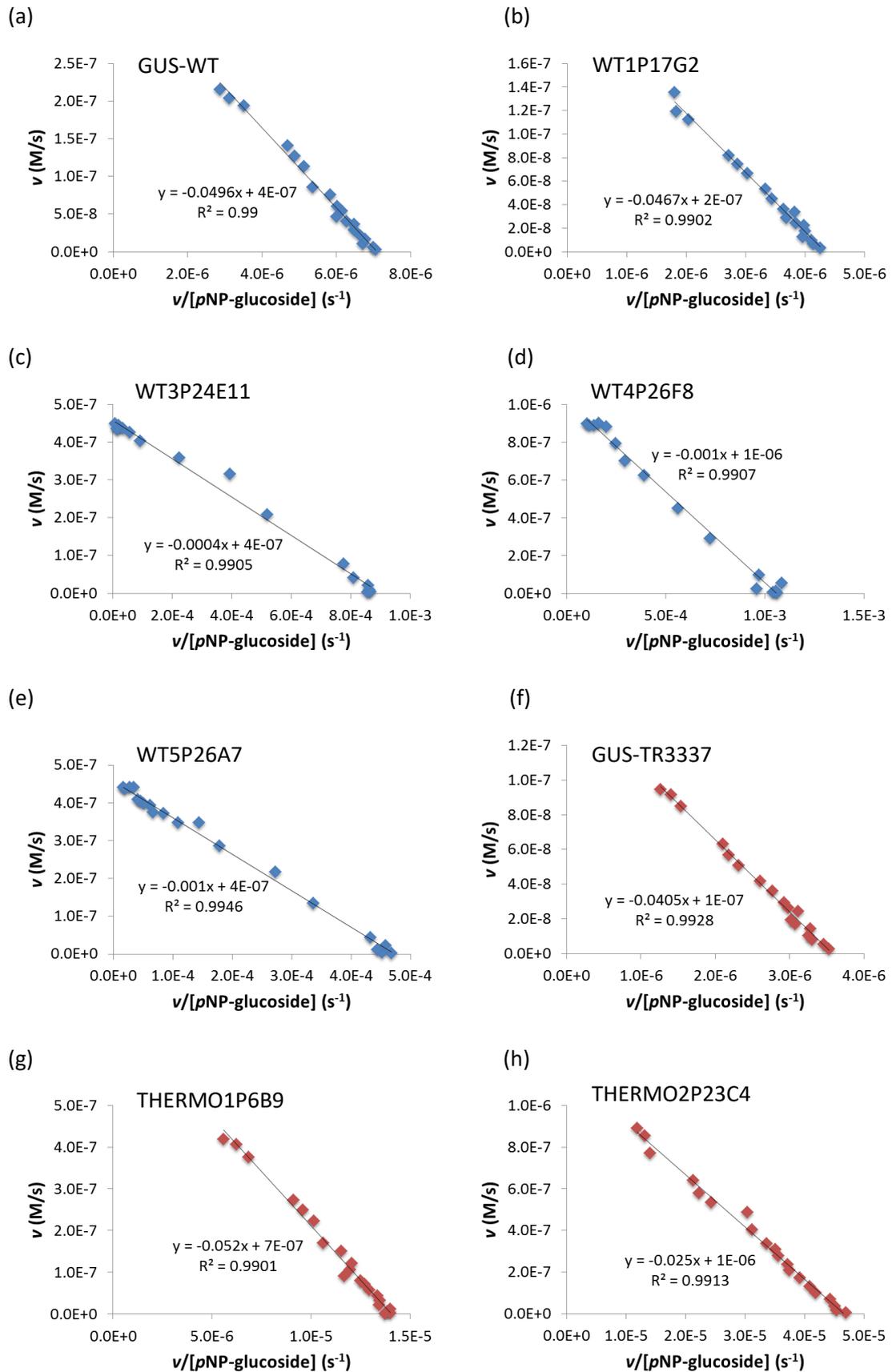


Figure D.6 Eadie-Hofstee plots of the *p*NP-glucoside hydrolysis by wild-type and evolved variants

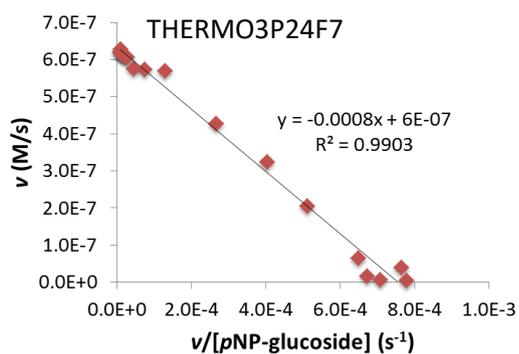
Each plot is the mean of triplicates

(a) GUS-WT (b) WT1P17G2 (c) WT3P24E11 (d) WT4P26F8 (e) WT5P26A7 (f) GUS-TR3337 (g) THERMO1P6B9 (h) THERMO2P23C4 (i) THERMO3P24F7 (j) THERMO4P11F2.

D.7. Eadie-Hofstee plots of the *p*NP-glucoside hydrolysis by wild-type and evolved variants in the presence of 5 mM OTG



(i)



(j)

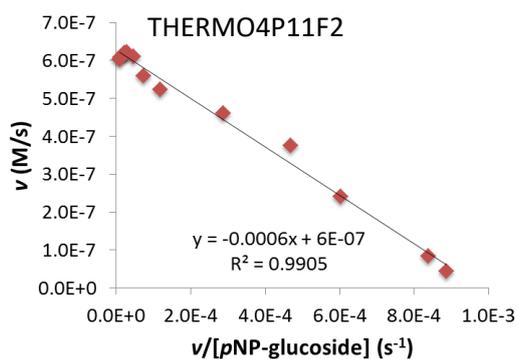


Figure D.7 Eadie-Hofstee plots of the *p*NP-glucoside hydrolysis by wild-type and evolved variants in the presence of 5 mM OTG

Each plot is the mean of triplicates

(a) GUS-WT (b) WT1P17G2 (c) WT3P24E11 (d) WT4P26F8 (e) WT5P26A7 (f) GUS-TR3337 (g) THERMO1P6B9 (h) THERMO2P23C4 (i) THERMO3P24F7 (j) THERMO4P11F2.

D.8. Hydrolysis rate by evolved variants as a function of pNP-glucoside concentration, fitted to the double Michealis-Menten equation

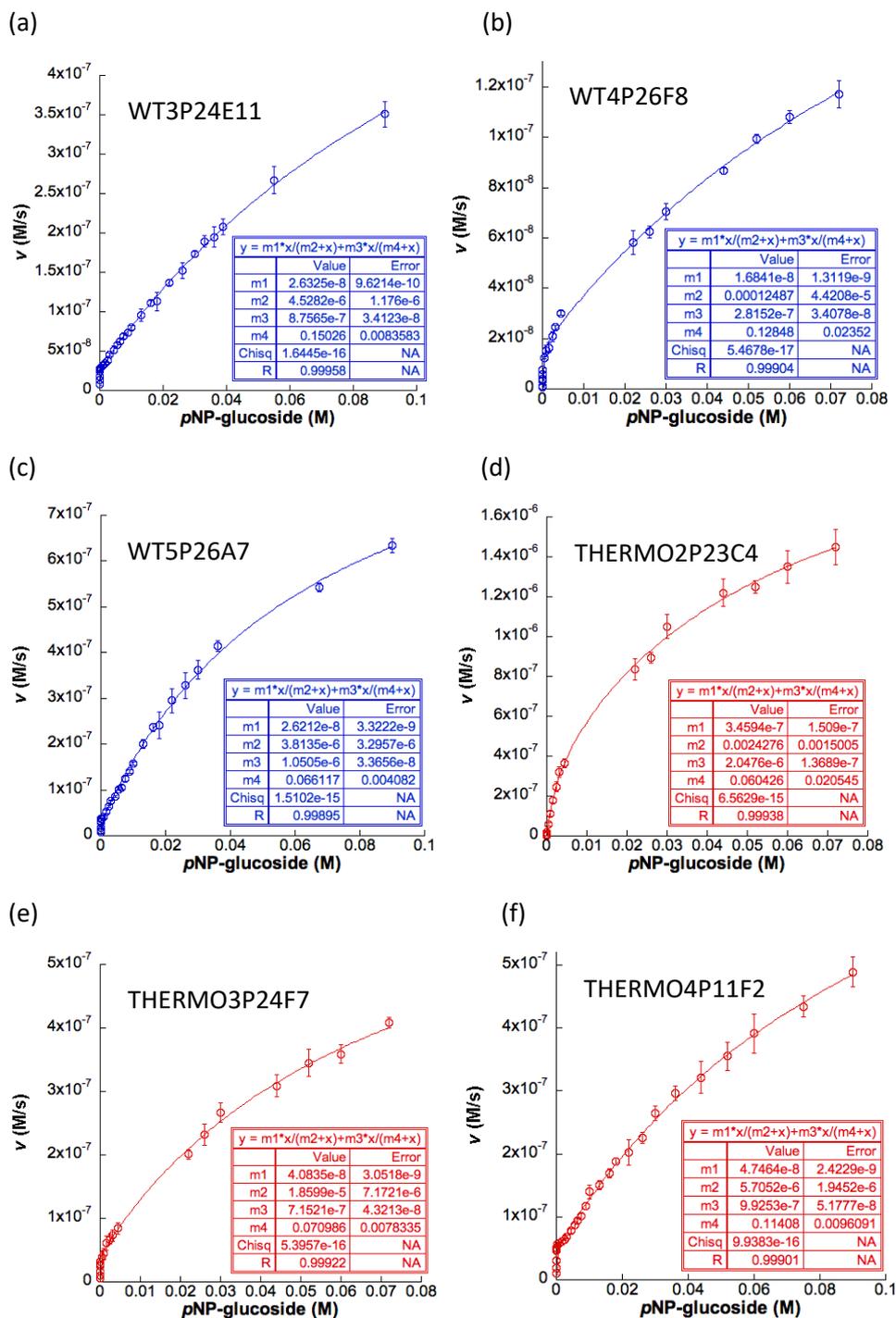


Figure D.8 Hydrolysis rate by evolved variants as a function of pNP-glucoside concentration, fitted to the double Michealis-Menten equation

Parameters for the fitted curves are given (m1: V_{max1} ; m2: K_{m1} ; m3: V_{max2} ; m4: K_{m2}); R is the correlation coefficient. (a) WT3P24E11 (b) WT4P26F8 (c) WT5P26A7 (d) THERMO2P23C4 (e) THERMO3P24F7 (f) THERMO4P11F2.

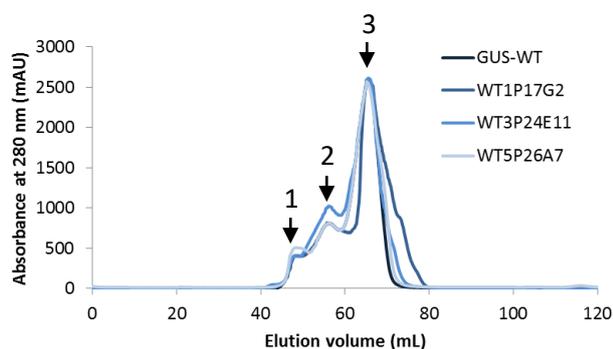
D.9.

D.9.1. Determination of protein molecular weight

The Superdex 200 size exclusion column (120 mL) used to estimate the size of GUS-WT* and GUS-TR3337* were pre-calibrated with five protein standards (ferritin, conalbumin, ovalbumin, thyroglobulin and aldolase) purchase from GE Healthcare Life Sciences (Section 2.3.4.5.).

The GUS-WT, WT1P17G2, WT3P24E11, WT5P26A7, GUS-TR3337, THERMO4P11F2 were loaded on the size-exclusion chromatography column. For all samples (20 mg of wild-type and mutants), three peaks were observed corresponding to the potential tetrameric, dimeric and monomeric states, which had apparent molecular masses of 69, 139 and 279 kDa, respectively (Figure D.9). Analysis by size exclusion chromatography demonstrated the presence of predominantly monomer species with smaller fractions of dimer and tetramer. This might be due to the sample dilution since diffusion occurs as sample passes through the column.

(a)



(b)

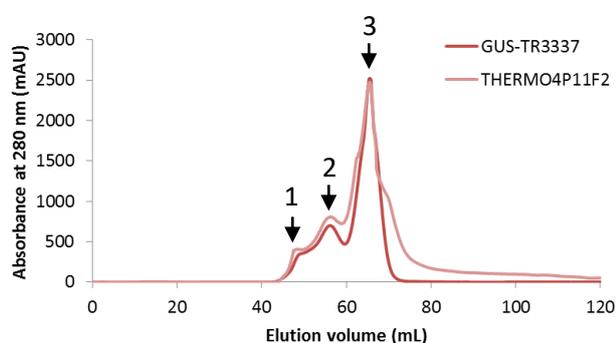


Figure D.9 Oligomeric GUS-WT and GUS-TR3337 variants are observed during size exclusion chromatography

Peak 1, tetramer (279 kDa); peak 2, dimer (139 kDa) and peak 3, monomer (69 kDa). Protein samples were analysed on a Superdex 200 column equilibrated with 50 mM phosphate buffer (pH 7.4), 100 mM NaCl at a flow rate of 0.3 mL/min. (a) GUS-WT* (b) GUS-TR3337*.

D.9.2. Discontinuous native protein gel electrophoresis analysis

To further investigate the oligomerization state of GUS-WT* and GUS-TR3337*, discontinuous native protein gel electrophoresis analysis was used, which allows the separation of the protein complex in its native quaternary structure according to its size and shape (Niepmann & Zheng 2006). As shown in Figure D.10, GUS-WT and GUS-TR3337 variants form dimeric and tetrameric bands with a molecular size of 139 and 279 kDa, respectively. Judging from the intensities of the bands, the monomeric form is more abundant than the tetrameric form at a protein concentration of 1 mg/mL. This observation was consistent with all tested mutants except THERMO4P11F2 migrated on the discontinuous native gel as a broad smear in addition to a weak 139 kDa band.

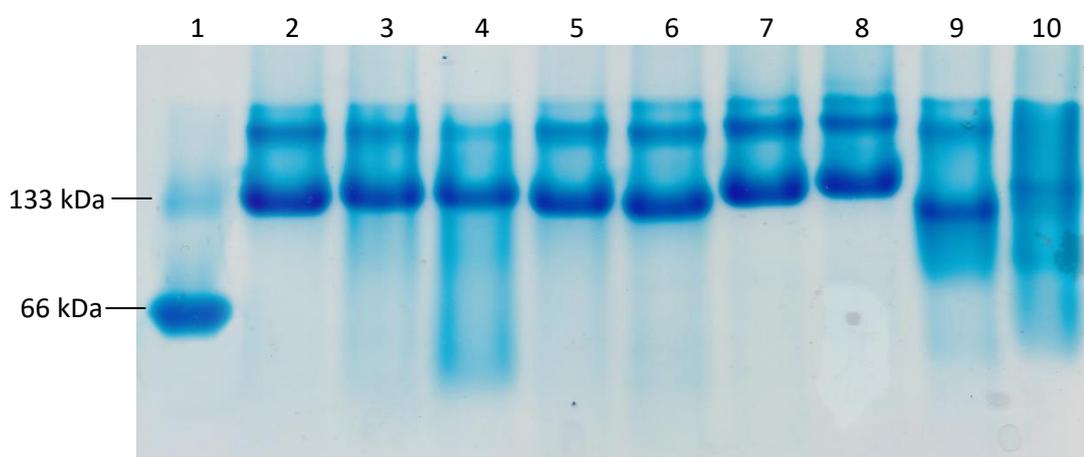


Figure D.10 Discontinuous native gel electrophoresis of GUS-WT and GUS-TR3337 variants

Lane 1: Bovine serum albumin (BSA), lane 2: GUS-WT, lane 3: WT1P17G2, lane 4: WT3P24E11, lane 5: WT4P26F8, lane 6: WT5P26A7, lane 7: THERMO1P6B9, lane 8: THERMO2P23C4, lane 9: THERMO3P24F7, lane 10: THERMO4P11F2.

D.9.3. Assessment of biphasic kinetics in WT3P24E11

Based on results obtained by using size exclusion chromatography and discontinuous native gel electrophoresis, GUS-WT and GUS-TR3337 mutants did not show significant differences in the oligomerization state. Since GUS-WT and GUS-TR3337 mutants did not exhibit biphasic but standard Michaelis-Menten kinetics with *p*NP-glucuronide, a series of experiments was conducted with *p*NP-glucopyranoside in order to explore factors that might affect the negative cooperativity in the mutant enzymes. WT3P24E11 was selected for further analysis as it carried fewest mutations (T509A, K568Q and P597Q) yet showed biphasic kinetics.

D.9.3.1. Crosslinking

The size-exclusion chromatography and discontinuous native gel experiments showed GUS-WT* and GUS-TR3337* may exist as an equilibrium mixture of monomers, dimers and

tetramers, dependent on protein concentration. In order to investigate further whether the biphasic curves were caused by differences in the enzymatic activity of different oligomers, GUS-WT and WT3P24E11 were incubated with bis(sulfosuccinimidyl)suberate (BS³) for various times to produce different ratios of covalent dimers to covalent tetramers and subsequently examined their glucosidase activity. SDS-PAGE analysis of GUS-WT incubated with BS³ indicated a fraction of the monomers were cross-linked into dimers, as evident from the decrease and increase in the intensity of the monomer and dimer bands, respectively (Figure D.11). The cross-linking reaction also resulted in the formation of a band with a molecular weight higher than that of cross-linked dimers (~130 kDa). It was presumably the result of intramolecular cross-links within un-cross-linked dimers preventing complete unfolding during SDS-PAGE. The same cross-linking patterns were observed for WT3P24E11. However, GUS-WT displayed significantly more dimer formation compared to the WT3P24E11 protein. Since the samples were not heat-treated before loading onto the gel, it was surprising to find that the GUS-WT dimer was more resistant to SDS than the WT3P24E11 dimer as one additional band with molecular weight estimates of 130 kDa indicated the GUS-WT dimer was observed before incubation with BS³. The assessment of glucosidase activity was performed on GUS-WT and WT3P24E11 with the 20 min incubation period. However, the cross-linking product of GUS-WT was inactive while the cross-linked sample of WT3P24E11 was less active than the uncross-linked WT3P24E11 (Figure D.12).

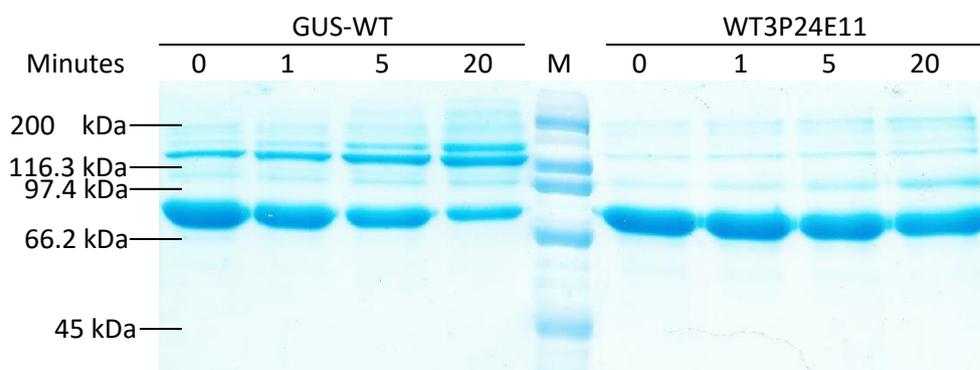


Figure D.11 Crosslinking of GUS-WT and WT3P24E11 with BS³, and subjected to SDS-PAGE

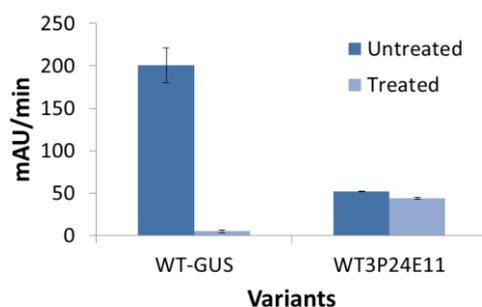


Figure D.12 Glucuronidase activity of native and cross-linked GUS-WT and WT3P24E11

Untreated: Samples were prepared in the absence of BS³. Treated: Samples were prepared by incubating BS³ for 20 minutes.

D.9.3.1. Effect of protein concentration on the biphasic kinetics

If the biphasic kinetics result from two distinct forms of GUS-WT* and GUS-TR3337*, performing assays with *p*NP-glucoopyranoside at different protein concentrations will give distinct kinetic patterns as the ratio of dimer/tetramer is depending on the protein concentration. The glucosidase activity was measured using four different concentrations (0.1 mg/mL, 0.05 mg/mL, 0.025 mg/mL and 0.01 mg/mL) of WT3P24E11. Similar biphasic curve was observed under the range of 0.1-0.01 mg/mL protein, suggesting that the kinetic behaviour does not caused by different enzymatic species as the kinetic profile does not change with the protein concentration (Figure D.13).

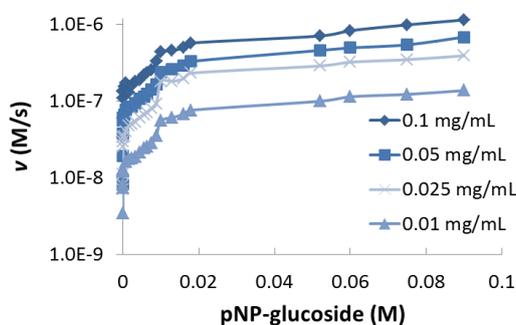


Figure D.13 The glucosidase activity is measured at different WT3P24E11 concentration

D.9.3.2. End-product inhibition and activation

Biphasic kinetics can also result from product activation, even without binding cooperativity. In order to investigate end-product inhibition, glucosidase activities of WT3P24E11 were obtained in the presence of different glucose concentrations (0.009 to 75 mM). A control experiment was performed in the absence of glucose. Figure D.14 exhibits substantially no activation nor inhibition of activity with the presence of high glucose concentration at 75 mM. The mutant retains its activity, both in the presence and absence of glucose. This property has great biotechnological merit as most β -glucosidases are inhibited by low concentrations of glucose (Saha & Bothast 1996).

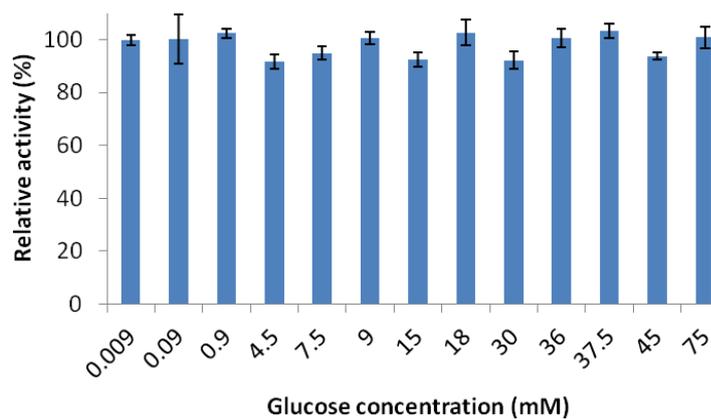


Figure D.14 Effects of glucose on the relative activity of WT3P27F8

D.10. References

Niepmann, M. & J. Zheng (2006) Discontinuous native protein gel electrophoresis. *Electrophoresis*, 27, 3949-51.

Saha, B. C. & R. J. Bothast (1996) Production, purification, and characterization of a highly glucose-tolerant novel beta-glucosidase from *Candida peltata*. *Appl Environ Microbiol*, 62, 3165-70.