University of Nottingham
UK | CHINA | MALAYSIA

Engineering Novel S-Glycosidase Activity into Extremo-Adapted β-Glucosidase by Rational Design

Nourah Almulhim

Supervisor:

Prof. Francesca Paradisi

Thesis submitted to the University of Nottingham for the degree of

Doctor of Philosophy

December 2020

Abstract

Biocatalysts have proven their efficiency and superiority in industry for the last few decades. Enzymes are made by nature to carry out specific functions required by the host organisms, therefore the practical applications of enzymes can be limited to their natural functions and this has encouraged efforts in the development of these natural catalysts to work in different conditions and with a significantly broader range of substrates.

Enzymes from extremophile organisms have significantly higher tolerance than mesophilic counterparts to temperature and/or pH, making them attractive for industrial applications. In industrial biotechnology, hydrolases, which are one of the six classes of enzymes, are the most commonly used biocatalysis.

Engineering extremo-adapted glycoside hydrolases to broaden their substrate scope towards β-thioglycosidase activity could significantly increase their potential applications. The breakdown of sulphur glycosidic bonds by β-thioglycosidases can produce isothiocyanate, a chemoprotective agent linked to the prevention of cancers, however only a handful of enzymes have been identified that are known to catalyse this reaction. Structural studies of the myrosinase enzyme, which is the only example among natural catalysts capable of hydrolysing the thioglycosidic bond, has identified residues that may play important roles in sulphate group recognition.

Protein engineering techniques can be used to introduce new amino acid residues into enzymes to improve their properties. By using rational design, two extremo-adapted β-glycosidases from the species *Thermus nonproteolyticus* (*Tno*GH1) and *Halothermothrix orenii* (*Hor*GH1) were engineered in this study towards thioglycosidic substrates. Twelve variants,

six for *Tno*GH1and six for *Hor*GH1, were assayed for activity. Remarkable enhancement of the specificity (k_{cat}/K_M) of *Tno*GH1 and *Hor*GH1 towards β-thioglycoside was observed in the single mutants *Tno*GH1-V287R (2500 $M^{-1}s^{-1}$) and *Hor*GH1-M229R, (13480 $M^{-1}s^{-1}$) which showed a 3-fold increase with no loss in turnover rate when compared to the WT enzymes. Thus, the role of arginine is key to induce β-thioglycosidase activity. Thorough kinetic investigation of the different mutants has shed light on the mechanism of β-glycosidases when acting on the native substrate.

Table of contents

Abbreviation

AMP	Ampicillin
ATP	Adenosine triphosphate
*Bb*MYR	Cabbage aphid *Brevicoryne brassicae* myrosinase
CAZy	Carbohydrate-Active Enzymes database
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
ε	extinction coefficient
EC	Enzyme classification number
E. coli	*Escherichia coli* bacterial strain
EtOH	Ethanol
GH	Glycoside hydrolase
GH1	Glycoside hydrolase family 1
HEPES	4-2-hydroxyethyl)-1-piperazineethanesulfonic acid
*Hor*GH1	*Halothermothrix orenii* glycoside hydrolase
iPrOH	Isopropanol
ITC	Isothiocyanate
IPTG	Isopropyl β-d-1-thiogalactopyranoside
GOS	Galacto-oligosaccharide synthesis
GSL	Thioglycosides
K_{cat}	Turnover number
K_M	Michaelis constant
LB	Lysogeny broth medium
MeOH	Methanol
MYR	Myrosinase
MW	Molecular weight
OD_{600}	Optical density (λ=600 nm)

O/N	Overnight
PCR	Polymerase chain reaction
pNP	p-nitrophenol
pNP-Fuc	p-nitrophenyl-β-D-fucopyranoside
pNP-Gal	p-nitrophenyl-β-D galactoside
pNP-Glc	p-nitrophenyl-β-D-glucopyranoside
pNP-Xyl	p-nitrophenyl-β-D-xylopyranoside
pNT	p-nitrothiophenol
pNT-Glc	p-nitrothiophenol- β-D-thioglucopyranoside
RT	Room temperature
*Sa*MYR	*Sinapis alba* myrosinase
SDM	Site-directed mutagenesis
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
TB	Terrific broth
THF	Tetrahydrofuran
TIM	Triose phosphate isomerase
*Tno*GH1	*Thermus nonproteolyticus* glycoside hydrolase
*Tte*GH1	*Thermobaculum terrenum* glycoside hydrolase
WT	Wild type

Amino acids

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

DNA and RNA nucleotides

A	Adenine
C	Cytosine
G	Guanine
T	Thymine (DNA)
U	Uracil (RNA)

Units of measure

Å	Angstrom
g	Gram
h	Hour
kDa	Kilodalton
L	Liter
M	Molarity (mol/L)
mA	Milliampere
mg	Milligram
min	Minute
µL	Microliter
mL	Millilitre
µm	Micrometre
mM	Millimolar
nm	Nanometre
mol	Mole
mo	Month
rpm	Rotation per Minute
s	Second
V	volt
yr	Year

Chapter 1*:* Introduction

1.1 Enzymes as biocatalysts

In the last century, the rapid growth of applications, production methods, and manufacturing were the main concerns of scientists and researchers in the chemical industry, with little consideration for environmental safety. This cemented a strong link between chemical industry and environmental problems which caused concern in communities for general health and quality of life. An increase in risk of breast cancer was linked to organochlorine chemical exposure. Similarly, hormonal equilibria in the body can be altered by a chemical endocrine disruptor.[1]

This lead to important improvements in the chemical industry and, more in general, synthetic process developments in the 21st century, with a pivotal point linked to when Paul Anastas and John Warner introduced the Twelve Principles of Green Chemistry[2,3] (Figure 1.1).[4] The 9th principle focuses on using catalytic amounts of reagents to enhance a chemical reaction rather than stoichiometric ones and/or catalysts to increase the reaction efficiency and decrease the formation of chemical waste, feedstock, and required energy.[2,5]

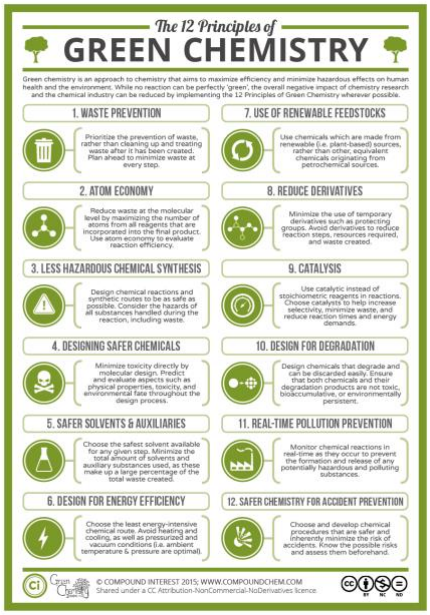
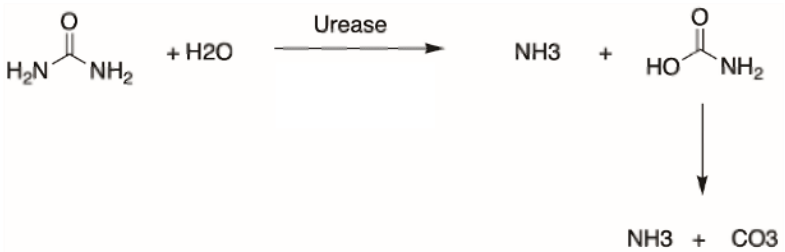


Figure 1.1. The Twelve Principles of Green Chemistry. Figure taken from Compound Interest.[4]

Catalysts, such as acids or bases, organo-catalysts and metals, are used in laboratories to speed up chemical reactions. Certainly, this is not the case for the biochemical reactions inside our bodies that are strongly regulated and, in many cases, would be rapidly inactivated at certain temperatures and pHs. On the other hand, most biochemical reactions inside living cells use enzymes as (bio)catalysts to enhance their reaction rate. An example of this, among many, is pyruvate kinase which catalyzes the last step in the metabolic pathway which breaks down glucose (glycolysis) herein a phosphate group from phosphoenolpyruvate is transferred to adenosine diphosphate (ADP) yielding pyruvate and energy (ATP).[6]

In 1959, enzymes were defined by John M. Reiner as "compounds, which cause chemical reactions to proceed at higher rates than would be the case in the absence of enzymes".[7] The significance of these biological compounds was reflected in the 1947 Noble Prize, which was awarded to James B. Sumner for his achievement in the isolation and crystallization of urease, a species that catalyzes the hydrolysis of urea into carbon dioxide and ammonia as shown in Scheme 1.1.[8] Together with Sumner, the Nobel Prize was shared with John H. Northrop and Wendell M. Stanley for the isolation of the digestive enzyme pepsin.[9]



Scheme 1.1. The hydrolysis of urea catalyzed by ureases.[8]

The fundamental role of enzymes is clearly seen in their central role as catalysts in almost all metabolic processes in our daily biochemical activities that sustain life which could not be achieved at the speed they are carried out under spontaneous conditions. These activities

include cell reproduction, the metabolism of food to produce energy, building or replacing tissues, and disposing of waste. For instance: the amount of time required to break down a protein in the body is ≤4 h. In contrast, 24 h in the laboratory is required to complete the same reaction. Of course, while the natural process takes place under normal temperatures and pH values, special conditions are needed to achieve the same result in the laboratory, such as boiling the protein in a 20% HCl solution.[9]

Moreover, biocatalysts are usually natural or modified enzymes that work in mild reaction conditions such as atmospheric pressure and ambient temperature. Enzymes are considered to be non-toxic, chemo-, regio- and stereoselective catalyst, and the common solvent in biotransformations is an aqueous medium rather than organic solvents.

Importantly, unlike classical chemical production strategies, biocatalysis play a central role in the improvement of the continuous growth of the chemical industry mitigating the environmental impact.[10–12]

Therefore, biocatalysis is a great example of green chemistry, defined as the "design of chemical products and processes to reduce or eliminate the use and generation of hazardous substances".[2]

Recently, enzymes such as epoxide hydrolases, reductases, ammonia lyases, transaminases, and dehalogenases have been also used in synthetic chemistry because of their advantages over chemical approaches such as high selectivity, specificity, and ability to work under mild conditions.[13–15]

Enzymes from extremophilic organisms, which are able to survive in a diversity of extreme environmental conditions such as temperature, pH, salt, pressure, metal, and radiation are attractive for industrial applications because of their unparalleled properties.[16] A lipase from

Bacillus thermocatenulatus is a thermostable enzyme (highest activity at 60-70°C at pH 8.0-9.0) which showed a great stability in a variety of detergents and organic solvents. More examples can be seen in Table 1.1. In contrast, mesophilic enzymes are usually less stable in harsh reaction conditions typical of many industrial processes and not attractive for practical applications.[17]

Table 1.1. Industrial applications of enzymes isolated from extremophiles.[17]

Extremophile	Habitat	Enzymes	Representative applications
Thermophile	High temperature	Proteases	Baking, brewing, detergents
Psychrophile	Low temperature	Proteases	Cheese maturation, dairy production
Alkalophile	High pH	Cellulases	Polymer degradation in detergents
Acidophile	Low pH	Sulfur oxidation	Desulfurization of coal
Halophile	High salt concentration	Halophilic amylase[18]	fish sauce and soy sauce preparations[18]
Piezophile	High pressure	Whole microorganism	Formation of gels and starch granules
Metalophile	High metal concentration	Whole microorganism	Biomineralization
Radiophile	High radiation levels	Whole microorganism	Bioremediation

Thermophilic glycoside hydrolases, which will be the focus of the next section, are reported to be one of the most well studied enzymes from extremophile organisms. These studies provide plenty of information about their functions and characteristics which makes them great candidates to work when high temperature conditions are required. One of the advantages of using these enzymes at high temperatures in biotechnological applications is the concomitant achievement of a decreased risk of (bacterial) contamination in the food and dairy industry.[19]

In paper manufacturing, xylanases, part of the glycoside hydrolases superfamily, from *Thermotoga* species were reported to be a perfect alternative to chlorine in the pre-

treatment of the wood pulp bleaching step which is used to increase the brightness of the paper by conversion xylan on the surface into xylose.[20]

1.2 Glycoside hydrolases

Carbohydrates such as starch, cellulose, and glycogen are abundant organic compounds in nature. They are fundamental sources of metabolic energy for plant, animal, fungi, and bacteria. More than being energy storages, carbohydrates are prime molecules in living organisms in the structure of nucleic acids, cells walls and metabolic intermediates. Moreover, in biological processes such as food storage and utilisation, carbohydrates were found to play a fundamental role, possibly more important than their peptides or nucleic acids analogous. In nature, carbohydrates exist as mono-, di-, oligo- and polysaccharides which are linked together by a bond, namely the *glycosidic bond*, that occurs in three types: O-, S-, and N- glycosidic linkages (Figure 1.2).

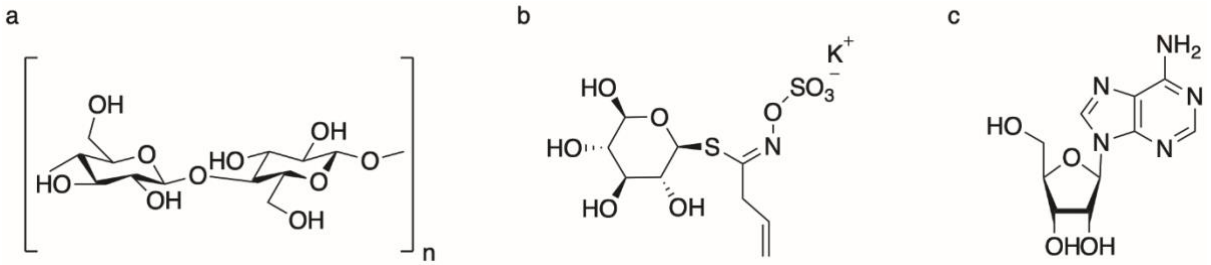
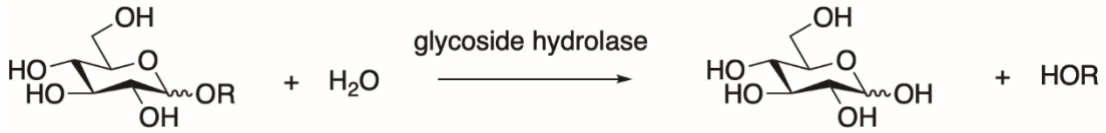


Figure 1.2. Structures of the three types of glycosidic bonds in **a** cellulose, **b** sinigrin, and **c** adenosine**.**

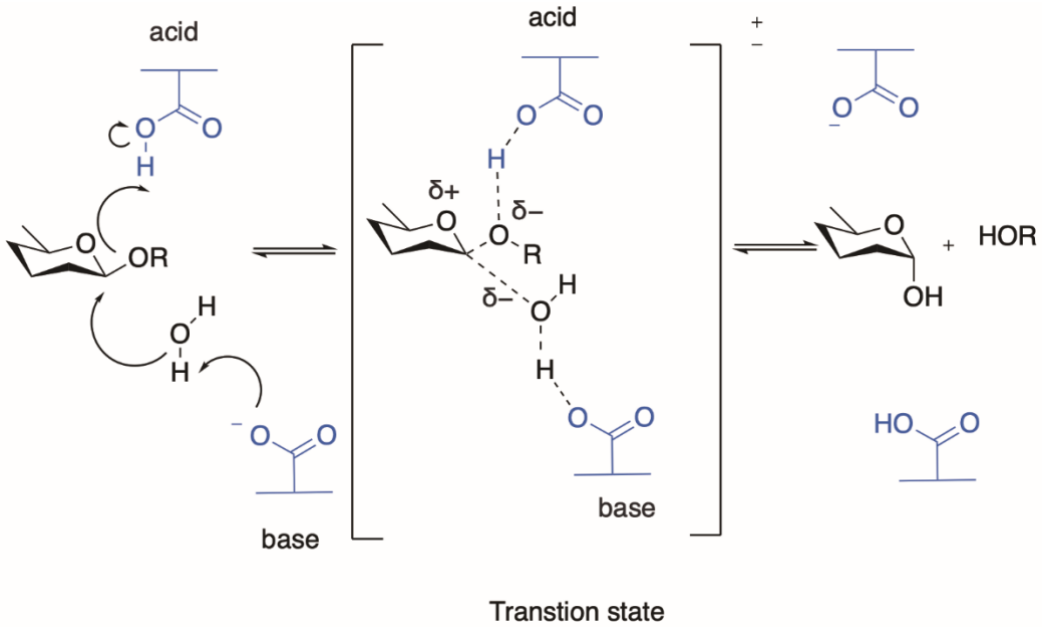
Glycoside hydrolases (GHs; EC 3.2.1.) are a heterogeneous group of enzymes characterised by the ability to catalyse the hydrolysis of glycoside linkages in a variety of carbohydrates (oligosaccharides, polysaccharides) or glycosides, between carbohydrate (glycone) and a non-carbohydrate moiety (aglycone).[21,22] These enzymes are found in almost all living

organisms from unicellular species (bacteria) to the most advanced ones (mammals) and show an exceptional variety of enzymatic activities due to the great diversity of their natural substrates.[23–25] The hydrolysis of O-glycosidic bond by GH is shown in Scheme 1.2.



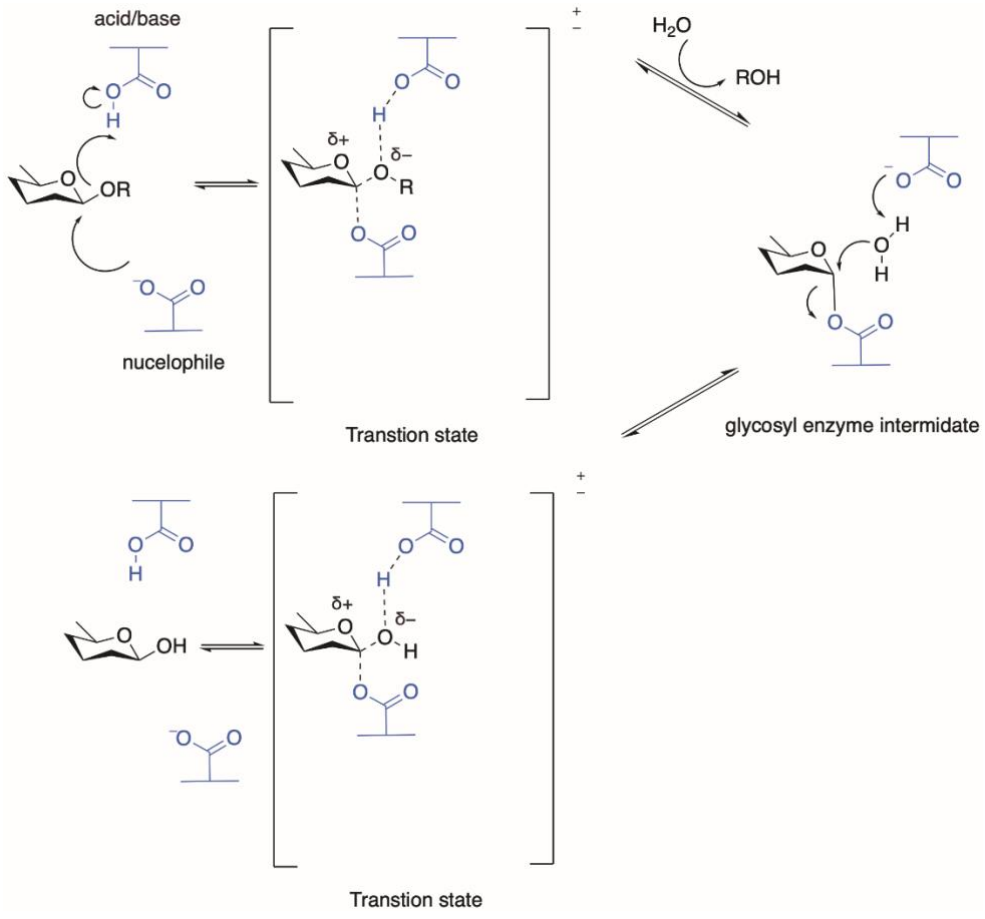
Scheme 1.2. General equation of glycoside hydrolyases.

Two different mechanisms, inverting and retaining of the anomeric configuration, are known. **The inverting mechanism** (Scheme 1.3) is a one-step mechanism that depends on acid/base amino acid residues. The first residue plays an important role in the glycosylation step which is the activation of the glycosidic oxygen. On the other hand, the "base" residue activates the water molecule resulting the departure of aglycon and the formation of sugar.[26]



Scheme 1.3. The inverting mechanism of glycoside hydrolase. Scheme is adapted from Park.[26]

The retaining mechanism (Scheme 1.4) consists of two steps. It starts with the protonation of the glycosidic oxygen by an amino acid residue which acts as a proton donor/acceptor (generally a glutamate). This is followed by a nucleophilic attach of a second glutamate. This protonation leads to the formation of glycosyl enzyme intermediate and separation of aglycon at the same time. In the next step, the proton donor/acceptor works as a base to activate the attack of a water molecule and produce sugar and aglycon. The proton/donor catalytic glutamate is located in in motif - T(F/L/M)NE(P/L/I). On the other hand, the nucleophilic one is in motif - (I/V)TENG. In fact, these key residues are only 5.5 Å apart which is an appropriate distance for the catalytic function of β-glucosidases.[26]



Scheme 1.4. The retaining mechanism of glycoside hydrolase. Scheme is adapted from Park.[26]

Within GHs families, the catalytic mechanism is either inverting or retaining. However, Masayuki and his group at Hokkaido University in Japan, noticed an exception in the GHs of family 97 which have both inverting or retaining enzymes and three catalytic residues: Glu439, Glu508, and Glu532. In the inverting members, the catalytic acid is Glu532 and the catalytic base is either Glu439 or Glu508 and in the retaining members the catalytic acid is Glu532 and the catalytic nucleophile Asp415.[27]

Since the discovery of GHs in 1989, three methods for their classification appeared in literatures, and they are based either on substrate specificity, amino acid sequence similarities, or three-dimensional structure similarities.[28]

Substrate specificity classification is the simplest one where the enzymes are divided into groups based on their substrates. In this way the substrate specificity of an enzyme can be clearly identified from its enzyme classification number (EC). The EC of O-glycoside hydrolyases is (EC 3.2.1.X), where X is a number representing a specific substrate. For instance, the β-D-glucoside substrates are number 21 and β-D-galactosides are number 23. Therefore, (EC 3.2.1.21) stands for a β-glucosidase and (EC 3.2.1.23) is the enzyme classification number of a β-galactosidase (Figure 1.3).[21]

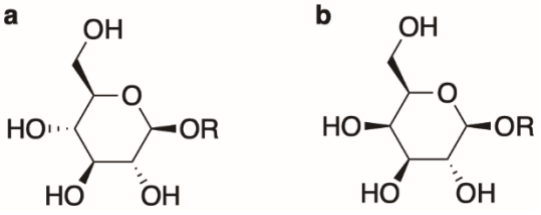


Figure 1.3. Glycoside hydrolases substrates. **a** is β-D-glucoside and **b** is β-D-galactoside.

The rapid increase in the number of new GHs gene and protein sequences, as well as three-dimensional structures of GHs, leads to the next classification method which is based on the

amino acid sequence similarities. It also includes molecular mechanism and the protein folding. Depending on the homology of their amino acid sequences and the similarity of their structures, GHs are classified into 168 families of enzymes as reported in the last update in the Carbohydrate-Active Enzymes database (CAZy).[28–30]

For examples, GH1 and GH13 have retaining mechanism and both are $(\beta/\alpha)_8$ barrel structure enzymes, but they differ in their catalytic residues which are two Glu in GH1 and Glu and Asp in GH13.[29]

More than one family of GHs can be classified in a one larger group called a clan if they have the same tertiary structure, molecular mechanism, and catalytic residues.[31] For example, GH-C clan has two families; GH11 and GH12 which are both β-jelly roll folded, retaining enzymes, and have Glu as the catalytic residues.[29] However, the amino acid sequence alignment of GH11 and GH12 shows only 19% similarity between these two GHs families. The continuously updated database (CAZy) reported 14 clans of GH as shown in Table 1.2.

Table 1.2. Glycoside hydrolase clans.[29]

Clan	Families	Mechanism	Structure
GH-A	1,2,5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113, 128, 147, 148, 157, 158, 167	Retaining	$(\beta/\alpha)_8$-barrel
GH-B	7, 16	Retaining	β-jelly roll
GH-C	11, 12	Retaining	β-jelly roll
GH-D	27, 31, 36	Retaining	$(\beta/\alpha)_8$-barrel
GH-E	33, 34, 83, 93	Retaining	6-fold β-propeller
GH-F	43, 62, 117	Inverting	5-fold β-propeller
GH-G	37, 63, 100, 125	Inverting	$(\alpha/\alpha)_6$
GH-H	13, 70, 77	Retaining	$(\beta/\alpha)_8$-barrel
GH-I	24, 46, 80	Inverting	α+β
GH-J	32, 68	Retaining	5-fold β-propeller
GH-K	18, 20, 85	Retaining	$(\beta/\alpha)_8$-barrel
GH-L	15, 65, 125	Inverting	$(\alpha/\alpha)_6$
GH-M	8, 48	Inverting	$(\alpha/\alpha)_6$
GH-N	28, 49	Inverting	β-helix
GH-O	52, 116	Retaining	$(\alpha/\alpha)_6$
GH-P	127, 146	Retaining	$(\alpha/\alpha)_6$
GH-Q	94, 149, 161	Inverting	$(\alpha/\alpha)_6$
GH-R	29, 107	Retaining	$(\beta/\alpha)_8$-barrel

GHs are able to catalyse the hydrolysis of both α-glycosides and β-glycosides. GH acting on α-glucosidic linkage are called α-glucosidases and glucoamylases which are found in GH families

4, 13, 15, 31, 63, 97, and 122. GH13 has a great variety of enzymatic activities and considered to be one of the largest GH families which, to date, has 102422 members.[32–34]

The majority of GH families are, however, β-glucosidases. They are huge group of enzymes widely spread in the living world and play fundamental role in the biological processes.[35]

Depending on the reaction condition, β-glucosidases are reported to be able to either hydrolyse β-glucosidic bonds in carbohydrates or to synthesis the same bond in a process called transglycosylation.[23,36–39] Recently, β-glucosidases have been attractive in industrial chemical production and employed in a range of biotechnological applications.[40] One of the most common β-glycosides families is Glycoside Hydrolase Family 1 (GH1) which will be the main focus of this study.

Glycoside Hydrolase Family 1 (GH1) are $(\beta/\alpha)_8$ barrel folded enzymes, which consist of eight twisted, parallel β-strands, located in the internal part of the protein, surrounded by eight α-helixes in the external part. The C-terminal end (in the β strand 8) of all known $(\beta/\alpha)_8$ barrel proteins hosts the catalytic residues within the $\beta \rightarrow \alpha$ loop.[41,42] Different types of active site can be found in the GH1 family such as pocket, cleft, and tunnel as shown in Figure 1.4. Figure 1.4 shows examples of the $(\beta/\alpha)_8$ barrel structures of three different GH1, fungous (BGL1A), bacterium (TmBGL), and plant (ZmBGL) and it also shows that BGL1A and TmBGL have the same tunnel-shaped active site, while the active site of ZmBGL is a flat pocket (Figure 1.4B). The $(\beta/\alpha)_8$ barrel structures of these three GH1 (Figure 1.4C) consist of extended loops on the C terminal side of the protein, which form the substrate binding pocket.[43] The quaternary structure of all GH1s is monomeric ;however, hyperthermophilic archaea GH1s are tetrameric proteins[44], and the myrosinases are dimers.[45]

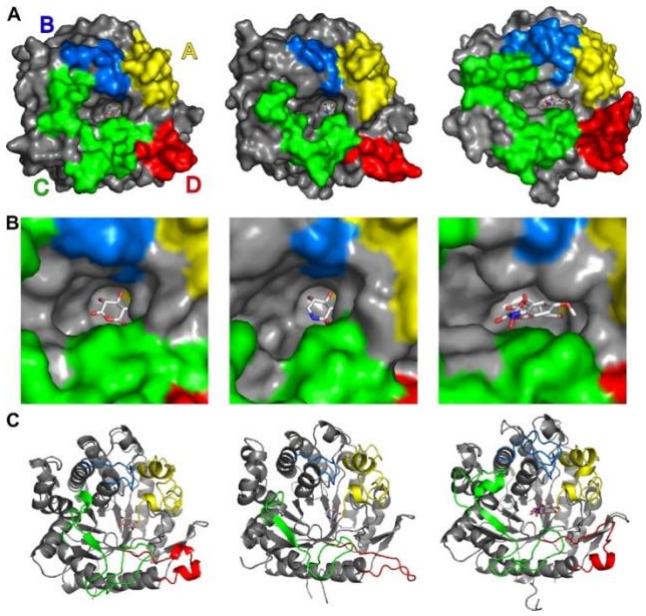


Figure 1.4. (A) Molecular surfaces of three glycoside hydrolases family 1 proteins (GH1s); BGL1A (GH1 from white-rot fungus *Phanerochaete chrysosprium*; left), TmBGL (GH1 from *Thermotoga maritima* bacterium; centre), ZmBGL E191D mutant (GH1 from zea mays plant; right). **(B)** Close-up views of active sites of the enzymes. **(C)** Ribbon representation of (A). Ligands are shown as stick models. Loop A (yellow) between β1 and α1; loop B (blue) between β4 and α4; loop C (green) between β6 and α6; and loop D (red) between β7 and α7. Loops A, B, and C are similar in the three GH1BGL structures. The aglycon binding site formed by loops B and C. Adapted from Nijikkena.[43]

The GH family 1 β-glucosidases are members of the 4/7 super family which includes other GHs such as GH2 β-galactosidase, GH5 cellulases, GH10 xylanase, and GH17 barley glucanases. It is a large family of eight-folded β/α barrel GHs that have the acid/base and the nucleophile residues at the end of β-strand 4 and 7, respectively.[46] However, many other GH1s activities are reported in the CAZY data base and some of them are β-galactosidase; β -mannosidase ; β-glucuronidase; β-xylosidase; β-D-fucosidase; phlorizin hydrolase; exo-β-1,4-glucanase; 6-phospho-β-galactosidase; 6-phospho-β-glucosidase; and thioglucosidase.[35]

The hydrolytic ability of GH1 is dependent on two critical glutamic acid residues:

E164 (*Thermus nonproteolyticus* glycoside hydrolase, *Tno*GH1 numbering), located in the $T^{161}LNEP^{165}$ motif (β-strands 4), is the acid catalyst.

E338 located in the $I^{336}TENG^{340}$ motif (β-strands 7) is the nucleophile.[47]

E164 plays an important role in the formation of the intermediate (enzyme-saccharide) of classical glycosidases as an activator of the glycosidic oxygen.

1.2.1 *Thermus nonproteolyticus* glycoside hydrolases (*Tno*GH1)

An example of GH1 is the thermophilic GH1 from the extremophilic bacterium *Thermus nonproteolyticus* (*Tno*GH1) which has been previously described in the literatures. *Thermus nonproteolyticus* is a thermophilic bacterium which was isolated from a hot spring in the Guangdong Province in Southern China. *Tno*GH1 is a monomeric enzyme and adopts the expected $(\beta/\alpha)_8$ barrel fold as other GH1 enzymes as shown in Figure 1.5, with a molecular weight around 50 kDa. It is a β-glucosidase determined to be active with β-D-glucoside, β-D-fucoside, β-D-galactoside and β-D-mannoside substrates.[47] This enzyme is part of this research project and was selected for its thermostability.

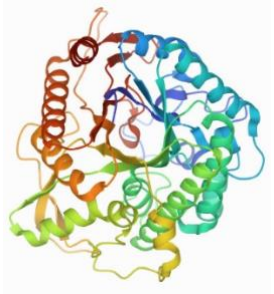


Figure 1.5. Protein database structure of the *Thermus nonproteolyticus* glycoside hydrolase (pdb:1NP2). Figure is adapted from the protein data bank (PDB).

One of the first efforts on gene identification, cloning, and expression of *Tno*GH1 was in 2001 by He Xiangyuan and co-workers. Their analysis of the amino acid sequence of *Tno*GH1, 437 amino acids (48,9 kDa), showed strong sequence similarities with other known GH1 amino acid sequences. In their study, they were also able to determine the optimal activity

conditions of *Tno*GH1 which is pH 5.6 and a temperature of 90 °C; the enzyme showed an impressive half-life of 2.5 h at 90 °C.[47]

Studies of the *Tno*GH1 structure have demonstrated that certain factors play an important role in the stability of this protein. The stability of the α-helix of *Tno*GH1, appears to be key and depends on the relative proportion of certain amino acid residues: amino acids with an α-helix-forming propensity, such as alanine (Ala) should be abundant while amino acids with helix structure destabilizing properties (β-branched), such as isoleucine (Ile), valine (Val), and threonine (Thr) should not. A previous study, which compared the stability of *Tno*GH1 and other mesophilic-equivalent GH enzymes, showed that, among all of the amino acid sequences of the mesophilic enzymes, the amino acid sequence of *Tno*GH1 had the highest concentration of Ala (17.9%) and the lowest concentration of β-branched residues (9.5%) which increases the stability of the α-helix structure, and consequently the stability of *Tno*GH1 structure.[47]

Another study, concerning the thermostability of *Tno*GH1, mentioned the presence of abundant critical amino acid residues, which enhance the stability of the enzyme structure, such as the high quantities of alanine (12.8%) and leucine (10.9%) found in *Tno*GH1. These residues play an important role in enhancing the core hydrophobic interactions of proteins. Another peculiar finding in *Tno*GH1 is the high abundance is arginine (9.6%.), which prevents hydrocarbon chain from unfavourable contact with water molecules that may increase enzymes stability. Moreover, high amounts of proline support the protein structure by providing a fixed tertiary structure, and this is also observed in *Tno*GH1 (8%). Another factor that enhances the stability of the protein structure is in the presence of electrostatic interactions (ion pairs). This type of interaction plays an important role in the stability of the

enzyme structure, particularly on the surface area, because it can cross-link the non-contiguous regions on the surface of the folded protein. Another study showed that *Tno*GH1 has the highest number of ion pairs per residue (0.1), compared to other GHs (0.08-0.05) that were included in the study.[48]

In 2003, the first three-dimensional structure of *Tno*GH1 was resolved by Xinquan Wang and his co-worker at a resolution of 2.4 Å. Their study showed that *Tno*GH1 has $(\beta/\alpha)_8$ barrel fold structure, which was highly conserved with the other eight GH1s analysed in their work, and presents the two signature glutamic acid residues in the active site, Glu164 in motif-TLNEP and Glu338 in motif-ITENG, which are critical for the mechanism of GH1.[47] Also, to further understand the possible basis of *Tno*GH1 thermostability, they compared *Tno*GH1 structure with other known TIM structures of mesophilic and hyperthermophilic glycosidases.

*Tno*GH1 was also evaluated in a previous work in the Paradisi lab for its activity and stability in a variety of organic co-solvents and at different pH values. Firstly, *Tno*GH1 was assayed for its enzymatic ability to catalyse the hydrolysis of β-glycoside linkages in four different carbohydrate moieties; p-nitrophenyl-β-D-glucoside (pNP-Glc), p-nitrophenyl-β-D galactoside (pNP-Gal), p-nitrophenyl-β-D-fucopyranoside (pNP-Fuc), and p-nitrophenyl-β-D-xylopyranoside (pNP-Xyl). The highest specific activity of the enzyme was towards pNP-Glc (1.7 U/mg) and pNP-Gal (1.67 U/mg). The kinetics profile elucidates that *Tno*GH1 showed low specificity (expressed in term of K_{cat}/K_M values) towards the aforementioned β-glycoside substrates at 25 °C when compared to another thermophile enzyme, *Thermobaculum terrenum* GH1 (*Tte*GH1), also included in this study. However, the highest K_{cat}/K_M value observed was with pNP-Glc; 0.033 $mM^{-1}\ s^{-1}$ (K_{cat} = 0.07 s^{-1}) while the lowest value was with pNP-Xyl; 0.004 $mM^{-1}\ s^{-1}$ (K_{cat} = 0.0051 s^{-1}).

*Tno*GH1 showed great resistance and 10-20% enhanced activity in the presence of DMSO and THF solvents. It was also stable over different pH values with a typical bell-shape profile at pH values ranging from 3 to 12 (unpublished results from Paradisi lab, University of Nottingham).

1.2.2 *Halothermothrix orenii* glycoside hydrolases (*Hor*GH1)

The second relevant enzyme in this project is the halotolerant GH1 isolated from *Halothermothrix orenii* (*Hor*GH1). *Halothermothrix orenii is* a salt adapted rod shaped bacterium, which was isolated from a hypersaline lake in Tunisia. Like *Tno*GH1, *Hor*GH1 is a monomeric β-glucosidase with an (β/α)8-barrel fold (Figure 1.6).

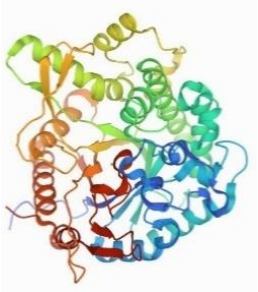


Figure 1.6. Protein database structure of the *Halothermothrix orenii* glycoside hydrolase (pdb:3TA9)**.** Figure is adapted from the protein data bank (PDB).

In 2009, the first genome sequence of an organism belonging to *Haloanaerobiales* was reported by Mavromatis and co-workers who sequenced the genome and predicted 2451 genes.[49]

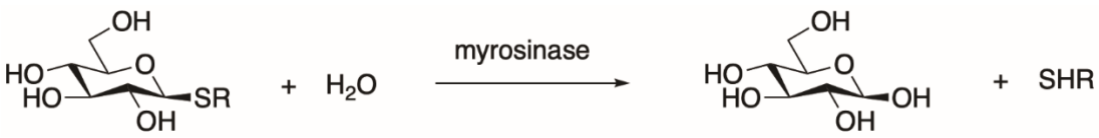
After that, in 2011 Kori and co-workers successfully cloned, expressed, purified, and crystallized *Hor*GH1. In their efforts; they were able to express *Hor*GH1 in *E. coli* DE3 cells competent cells, purify the enzyme by metal ion-affinity chromatography and resolve its structure at a resolution of 3.5 Å.[50]

In 2015, the optimum activity and thermostability conditions for *Hor*GH1 were determined by Hassan et al.[51] They observed that the optimum pH is 6.0 at a temperature range of 65-70 °C. The structural characterization and ligand binding evaluation of the enzyme led to the use of *Hor*GH1 in lactose conversion and galacto-oligosaccharide synthesis (GOS).[51]

Similarly, to *Tno*GH1, *Hor*GH1 was available in our laboratory as a recombinant protein and it was previously tested for its substrate scope with pNP-Glc, pNP-Gal, pNP-Fuc, and pNP-Xyl. The best performance of *Hor*GH1 activity was observed with pNP-Glc (1.86 U/mg) similar to the value observed with *Tno*GH1 (1.7 U/mg). However, the kinetics profile of the enzymes showed that *Hor*GH1 is more efficient towards pNP-Glc (K_{cat}/K_M = 0.109 mM^{-1} s^{-1} and K_{cat} = 0.106 s^{-1}) than *Tno*GH1 (K_{cat}/K_M = 0.033 mM^{-1} s^{-1} and K_{cat} =0.07 s^{-1}).

1.2.3 Myrosinase

Myrosinases (MYRs; EC 3.2.3.147) are unique members of the GH1 family able to hydrolyse glucosinolates (GSLs; also called thioglycosides) which are the most stable glycosidic molecules in nature (Scheme 1.5). Interestingly, they are the only enzymes capable of hydrolyzing GSLs as this activity is not found in any other GHs. MYRs are also known as thioglucoside glucohydrolases, S-glucosides, sinigrinases.[45] They are $(\beta/\alpha)_8$ barrel folded proteins, as are the other members of the GH1, but unlike β-glycosidases which are ubiquitous, myrosinases have been identified only in a handful of species such as *Sinapis alba*[45], *Brevicoryne brassicae*[52,53], *Verticillium longisporum*[54], *Arabidopsis thaliana,* and *Brassica napus.*[55]



Scheme 1.5. General equation of myrosinase.

GSLs are plant secondary metabolites, classically as 1-thio-β-D-glucosides.[56] In these structures, the glycoside is bridged to the aglycon moiety by a sulphur bond (Figure 1.7). One of the main natural sources of GSLs are plants belonging to the Brassicaceae species (Cruciferae). Examples of Cruciferae are kale, cabbage, brussels sprouts, and broccoli.[57,58] Different parts of the plant can have different amounts of GSLs, and consequently produce different hydrolytic activity of the plant MYR. The highest amounts are found in seeds and seedlings.[59] As shown in Figure 1.7, GSLs are rich in nitrogen and sulphur, and are classified into three main groups (aliphatic, aromatic, and indole). This classification depends on factors such as their aglycon, their amino acid predecessors, and their core structures.[62] Furthermore, genetic makeup of the host plant, geographic area, weather, growth stage, and time in the harvest cycle all play an important role in determining the type of GSLs group. A study from Spain, on Spinach and Turkish cabbage cultivars, showed the effects of geographic area on the type of GSL in the cultivars. While the Spinach cultivar was found to have high concentration of the aliphatic type, the Turkish one had higher indolic GSL.[57] The breakdown of S- glycosidic bonds in GSLs can release molecules with activity against pests and herbivores (part of the plant defence mechanism) and isothiocyanates (ITCs; chemoprotective agent linked to the prevention of cancers.[61,62] Interestingly, in plants, GSL and MYR are compartmentalised and kept separated, however when the tissues of the plant is damaged by chewing or chopping, the catalytic capacity of MYR is activated to interact with GSL, so that the hydrolytic reaction of GSL starts and the ITCs are released.[64]

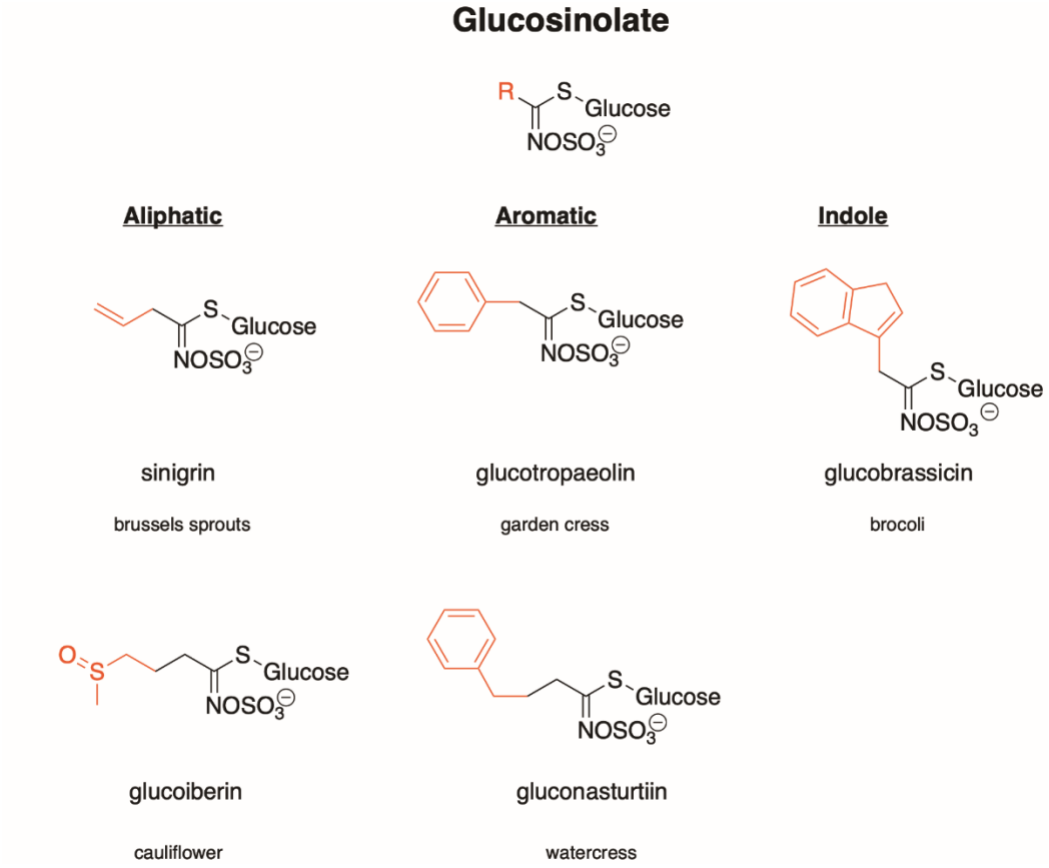


Figure 1.7. Chemical structures of plant glucosinolates (GSLs). Figure taken from Cruciferous Vegetables.[63]

1.2.3.1 *Sinapis alba* Myrosinase (*Sa*MYR)

Sinapis alba myrosinase (*Sa*MYR) is a plant species MYR from *Sinapis alba* plant (also known as white mustard). The quaternary structure of *Sa*MYR is a dimer as shown in Figure 1.8.

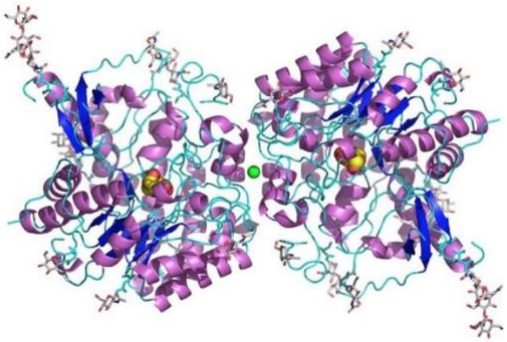


Figure 1.8. The overall structure of *Sinapis alba* myrosinase (*Sa*MYR) showing the dimer held together by a zinc (green) atom, fluor glucose (yellow). Published by Bones in 2006.[64]

*Sa*MYR is classified as a unique member of GH1 with an active site that differs from that of classical GH1, as it lacks the catalytic glutamic acid residue in the $T^{184}INQL^{188}$ motif (equivalent to $T^{161}LNEP^{165}$ in *Tno*GH1)[61], while it maintains the second one (E409 in motif $T^{408}ENG^{411}$) (Figure 1.9).[45]

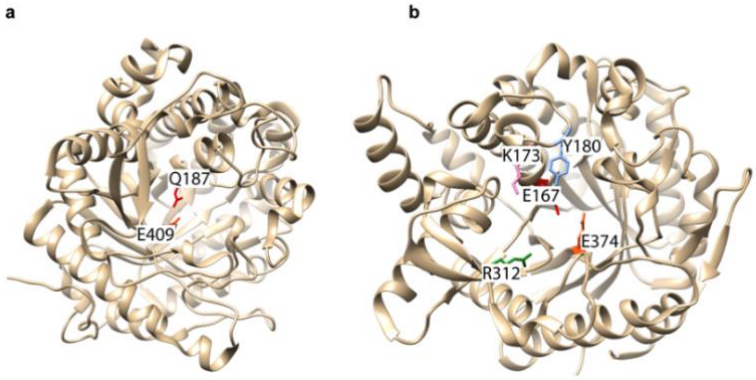


Figure 1.9. Modelling structures of the active sites of myrosinase enzymes. **a** *S. alba* myrosinases (pdb:1E4M)**;** Q187 and E409 are the active site residues. **b** *B. brassicae* myrosinases (pdb:1WCG); E167 and E374 are the active site residues.

As previously explained, the catalytic residue (Glu) plays an important role in the glycosylation step of classical glycosidases as activator of the glycosidic oxygen in this step (Schemes 1.4). E409 is capable of perfectly positioning the water molecule in the hydrolysis of the glycosyl enzyme intermediate, and subsequently releasing the products. In addition, Q187 plays another important role in the hydrolysis of β-thioglucosidases as it has the ability to form an important hydrogen bond between its amide nitrogen and the sulphate group of the substrate.[45,59] A study of the crystal structure of *Sa*MYR by Wim determined another reason for the ability of plant MYR to hydrolyse glucosinolates in the absence of the second catalytic acid residue. This is based on the presence of an excellent aglycon leaving group which is not present in the classical β-glucosidase. Wim and colleagues also mentioned another important amino acid residue in the active site of *Sa*MYR, which is Ser190 (Figure 1.10). It is a definer of

the position of E409 and possibly capable of hydrogen bonding to the sulphate group of the GSLs (featured prosperity of *Sa*MYR).[45,66]

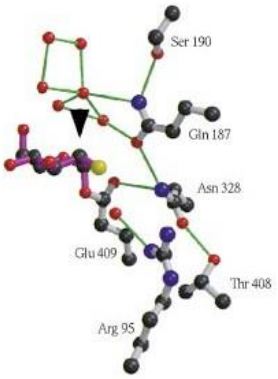
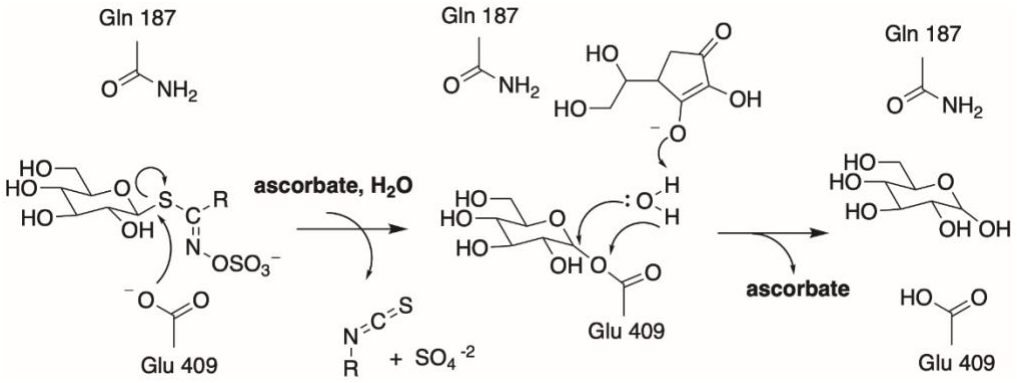


Figure 1.10. Side view of the active site of *Sinapis alba* myrosinase (*Sa*MYR) containing the sinigrin substrate.. The hydrogen-bond network (green) around Gln187 is shown. Water molecules are shown as red spheres. An arrow is drawn below the water molecule that is situated in a position suitable for the nucleophilic attack. Published by Burmeister in 1997.[45]

Another significant factor found in the mechanism of plant MYR is the ascorbic acid that exists in the vacuoles of plant cells. The importance of ascorbic acid was first described by Nagashima and Uchiyama, who noticed that ascorbate substituted for the function of the (missing) catalytic acid in the mechanism of GH1, and as an activator of the glycosidic oxygen in the glycosylation step (Scheme 1.6)[66,67]. In Scheme 1.6, the critical function of ascorbic acid as an activator for the catalysis of thioglycosidic substrates hydrolyses by *Sa*MYR is shown. This makes *Sa*MYR the first example of an enzyme that depends on an external cofactor in the glycosylation step, which differs from the normal function of ascorbic acid as a reducing agent of the reaction.[65]



Scheme 1.6. The retaining mechanism of *Sinapis alba* myrosinase (*Sa*MYR).[64]

The amino acid residues around the glucose ring of *Sa*MYR are Tyr330, Phe473, Trp457, Glu464, Phe465, Gln39, His141, and Asn186 are also important (Figure 1.11). Glu464, Gln39, His 41, and Asn186 provide six hydrogen bonds with the glucose. On the other hand, the hydrophobic environment around the glucose ring is formed by Tyr330, Phe473, Trp457, and Phe465.[67]

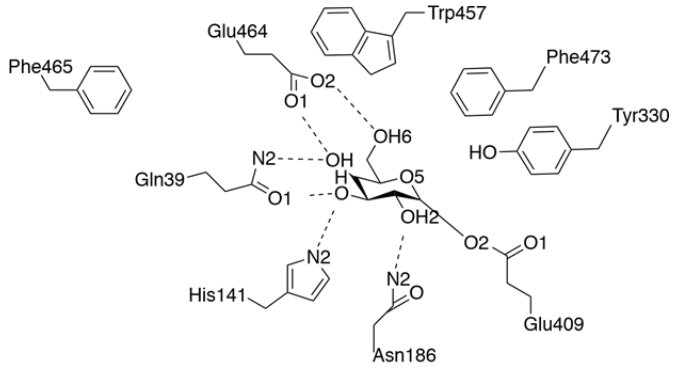


Figure 1.11. Amino acid residues around the glucose ring of *Sa*MYR.[45]

As with the discovery of *Sa*MYR, several studies were carried out to compare classical GH1 β -glycosidases and *Sa*MYR. One of the first study was by Wim which focused on determining the most similar plant MYR to cyanogenic β-glucosidase, a GH1 from *Trifolium repens* (CBG), in terms of amino acid sequences, protein structure, and active site residues. This study included three different plant MYRs which were from *Sinapis alba* (*Sa*MYR), *Brassica napus*

(MYRO_BRANA), and *Arabidopsis thaliana* (MYRO_ARATH). Among these plant MYRs, the *Sa*MYR was the most similar one to the CBG (46 %), but as expected CBG and *Sa*MYR differ in their catalytic residue (Glu183 in CBG and Gln187 in *Sa*MYR).[45]

Wim presented a second study based on a bacterial GH1 β-glucosidase from *Cellulomonas fimi* (Cex). In this study, a 3D structural comparison between the Cex and the glycosyl-MYR intermediate was explored. Based on this comparison, it was found that although Cex and *Sa*MYR were identical in the nucleophile residues (Glu) in their active sites, they differed in the shape of their active sites (cleft in Cex and pocket in *Sa*MYR), the size of their $(\beta/\alpha)_8$ barrel structure (Cex was smaller than *Sa*MYR), and as expected, they also differed in their catalytic residue (Glu127 in Cex and Gln187 in *Sa*MYR).[45]

1.2.3.2 Cabbage aphid *Brevicoryne brassicae* myrosinase *(Bb*MYR)

Cabbage aphid *Brevicoryne brassicae* myrosinase (*Bb*MYR) is an insect species MYR from the cruciferous specialist aphid *Brevicoryne brassicae*. This MYR is able to hydrolyze commonly known plant GSLs such as sinigrin and glucotropaeolin (Figure 1.7). *Bb*MYR is a dimeric $(\beta/\alpha)_8$ barrel structure GH1 as *Sa*MYR (Figure 1.12)[66]. It acts through retaining mechanism as other GH1 members (Scheme 1.7).

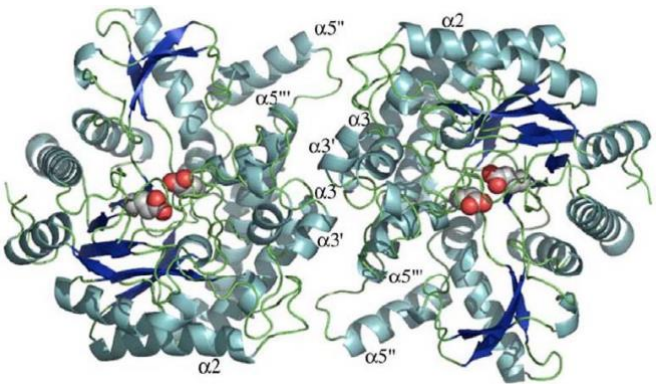
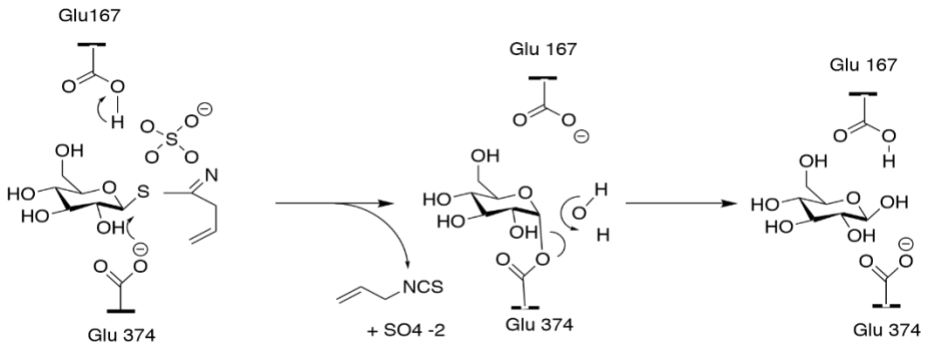


Figure 1.12. The structure of aphid myrosinase showing the dimer. The two catalytic glutamic acid residues are shown in red. Published by Bones in 2006.[64]



Scheme 1.7. The retaining mechanism of Cabbage aphid *Brevicoryne brassicae* myrosinase (*Bb*MYR).[64]

Unlike *Sa*MYR, *Bb*MYR relies on the typical catalytic acid/base residue(E167) utilized by β-glucosidase in the activation of the glycosidic sulphur, in GSLs, in the glycosylation step. *Bb*MYR is also lacks on ascorbic acid and Ser190 (*Sa*MYR numbering) which play important roles in the mechanism of *Sa*MYR. In *Bb*MYR, Ser190 is replaced by Ala170 which is unable to hydrogen bond to the sulphate group as Ser190 in *Sa*MYR.[67] Therefore, *Bb*MYR closely aligns with the classic β-glycosidases rather than *Sa*MYR in term of structure, amino acid residues present in the active site (E167 and E374; Figure 1.9b), and phylogenetic relatedness (Figure 1.13).[67] The glucose binding site of *Bb*MYR is also very similar to β-glucosidases which is another feature distinguished the *Bb*MYR from the *Sa*MYR. As discussed above, Figure 1.11 shows the glucose binding site of *Sa*MYR and the important amino acid residues around the glucose ring. In *Bb*MYR, there is a replacement of phenylalanine (Phe465) by tryptophan (Trp424), which leads to the formation of an additional hydrogen bond between the nitrogen atom of the indole ring (in Trp) and the hydroxyl group of the glucose which is a general feature of β-glucosidases.[66]

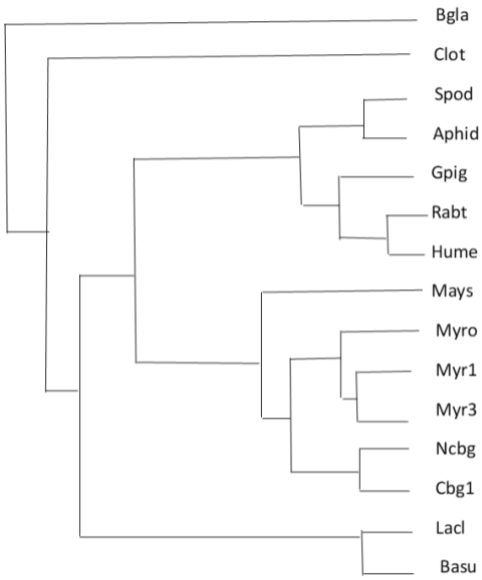


Figure 1.13. Cladogram of some members of glycosyl hydrolase family 1 (GH1). β-glucosidases are: Basu (from *Bacillus subtilis*), Bgla (from *Bacillus polymyxa*), Lacl (from *Lactococcus lactis*), Mays (from *Zea mays*), Clot (from *Clostridium thermocellum*), Spod (from *Spo-doptera frugiperda*), Gpig (from *Cavia porcellus*), Cbg1 (from *Trifolium repens*), Ncbg (*Trifolium repens*), Hume (from human) and Rabt (from rabbit). Myrosinases are: Myro (from *Arabidopsis thaliana*), Myr1 and Myr3 (from *Sinapis alba*), and Aphid (from *Brevicoryne brassicae*).Adapted from Jones et al.[67]

Even though the active site of *Bb*MYR lacks the Gln residue present in *Sa*MYR, which is essential in the mechanism of *Sa*MYR for hydrolysing GSL substrates, it is nonetheless capable of hydrolysing GSLs. In a previous study, Jones and colleagues mentioned that K173 and R312 in the active site of *Bb*MYR play a critical role in the hydrolysis of GSL, as they are directly involved in its recognition (Figure 1.9).[67] In *Sa*MYR, K173 and R312 are replaced by R194 and Q333 . Equally important, another study of the active site of *Bb*MYR by Harald Huseby and his group highlighted a possible catalytic role of Y180 due to its proximity to the thioglycosidic linkage in the substrate.[66] Interestingly, the presence of K173, R312, and Y180 is a feature of *Bb*MYR and cannot be seen in either β-glucosidases and other myrosinases.

1.3 Protein engineering

In chemical reactions, enzymes often prove superior to non-enzymatic catalysts as explained in Section 1.1.

The development of biocatalysis has found plenty of applications which have been extended outside the academic laboratories.

Enzyme technology has also been applied in medicine,[68–72] for biofuel generation such as in the production of an environmental friendly biodiesel[73], cleaning products such as detergents,[74] food manufacturing such as pectinases which have been essential for fruit juice processing industry [75] and cosmetics such as the use of a mixture of superoxide dismutase (SOD) and peroxidase in sunscreen creams and proteases in skin creams.[16] However, in industrial biotechnology these natural catalysts face several challenges such as an often limited substrate scope,(co-) solvent tolerance, short half-life, and sensitivity towards environmental changes, such as deviations from their optimal pH and/or temperatures, which pose a significant hurdle in maintaining the integrity of a sample for prolonged periods of time.[76] These issues limit the use of an enzyme and its practical applications. However, a significant evolution in biocatalysis was achieved with introducing the recombinant deoxyribonucleic acid (DNA) technology and the development of polymerase chain reaction (PCR; Figure 1.14) which allowed the replication of DNA and subsequently the modification of enzymes in protein engineering. Protein engineering is achieved through genetic modification (GM) of a gene coding for an enzyme of interest to enhance a variety of properties[77,78], including catalytic activity[79], substrate specificity[80], thermostability[81], and solvent tolerance.[82,83]

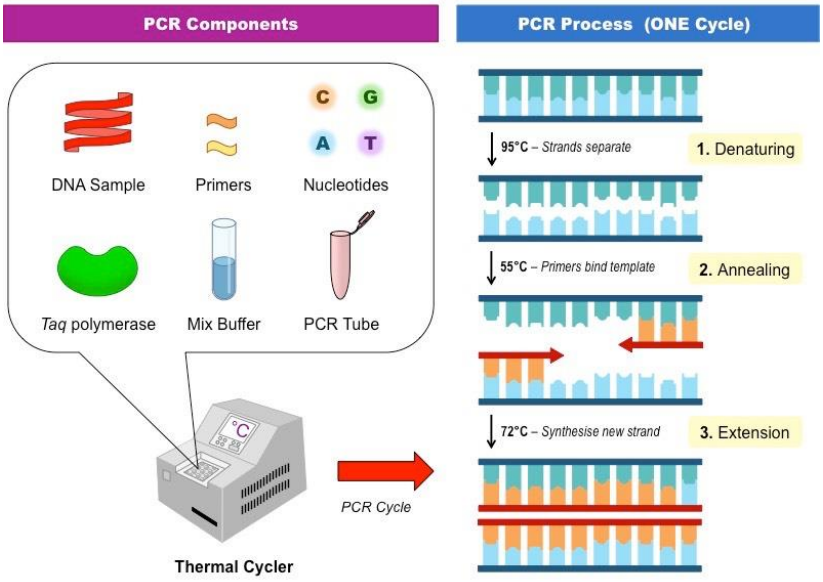


Figure 1.14. The principle of polymerase chain reaction (PCR). Figure is adapted from The Biology Notes.[84]

Therefore, numerous protein engineering studies have been carried out on enzyme modification and one of the first studies was in 1985 by David A. Estel and his group who were able to chemically modify a serine protease from *Bacillus* species which was the first example of engineering an industrial enzyme. It had been reported that a Met, Cys, and Tyr around the enzyme active site decreased the stability of the enzyme as a result of oxidation which subsequently inactivated the enzyme. An oxidation of M222 residue to methionine sulfoxide in the active site of serine protease was achieved in the presence of H_2O_2. In contrast, the mutant M222A showed great stability even in the presence of 1 M H_2O_2.[85]

Another example of using enzyme engineering to enhance the stability of enzymes can be seen in a mesophilic hydrolytic enzyme, an acetylcholinesterase, from *Drosophila*. Acetylcholinesterases are enzymes used as biosensors for their ability to detect organophosphate and carbamate pesticides in the environment. *Drosophila* acetylcholinesterase is considered to be one of the most sensitive yet less stable acetylcholinesterase. A common strategy to improve the sensitivity and stability of proteins

is to increase the peripheral polarity to maximise solvation. This can be achieved by mutating hydrophobic residues that are expose to the solvent on the surface of the enzyme into polar residues such as arginine. In this example, 14 hydrophobic residues at the surface of *Drosophila* acetylcholinesterase were mutated to arginine which resulted an increase in the *Drosophila* acetylcholinesterase stability.[86]

The most common two types of protein engineering are directed evolution and rational design are shown in Figure 1.15.[87]

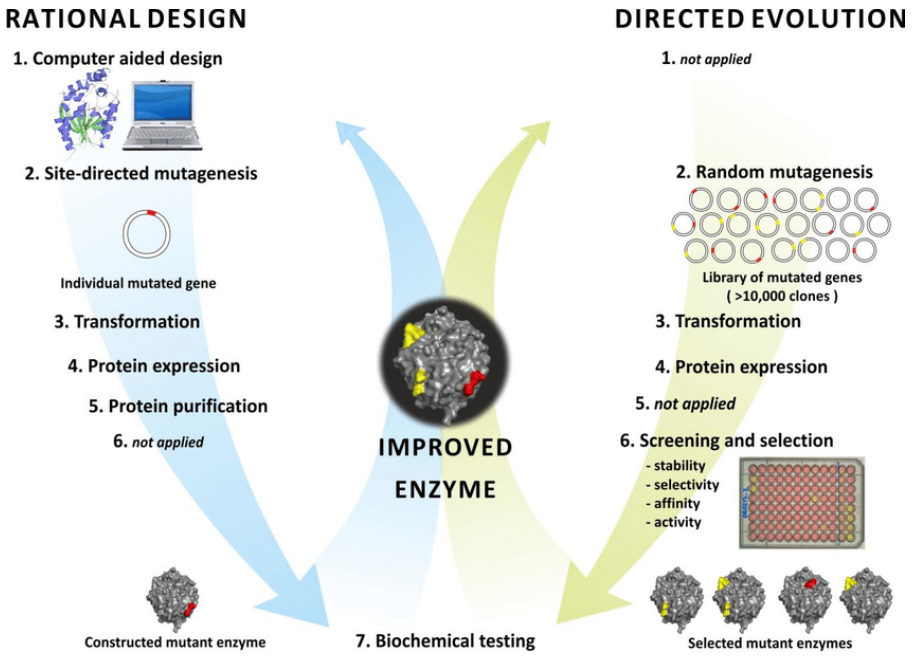


Figure 1.15. The two major types of protein engineering. Published by Bendl in 2016. [87]

Rational design, which is often implemented via site directed mutagenesis (SDM), is a protein engineering technique that involves an addition, deletion, or a change of one or more amino acid residues in a protein.[88,89] Computational studies regarding the protein of interest and prior knowledge of the enzyme mechanisms are essential in this technique to understand the protein structure then subsequently predict which amino acids need to be introduced,

deleted, or altered in the protein. In 1978, Michael Smith (Nobel Prize in Chemistry in 1993) investigated this method.[90]

SDM is particularly useful in determining the role of catalytic residues and to understand the correlation of a protein structure to its function.[91–96]

Directed evolution is another protein engineering technique based on creating a library of clones with multiple random mutations followed by an appropriate screening and selection of the mutants. Variants that have improved properties can be identified and possibly subjected to further cycles of optimization until the desired properties such as enhanced enzyme stability, affinity, specificity are achieved.[97–101]

Unlike the rational design approach, directed evolution mutagenesis does not require prior knowledge of the protein structure. Therefore, directed evolution is a beneficial technique that can be particularly useful when structural information is not available.[102]

Error-prone PCR (epPCR) is a commonly used method for the introducing of genetic diversity. Random mutations can be obtained by for example decreasing the DNA polymerases' fidelity and/or using agents such as Mg^{+} and Mn^{+2}.[103]

1.3.1 Protein engineering applied to glycoside hydrolases:

Recently, directed evolution has been a powerful tool in enzyme engineering field to improve the enzymes properties such as thermal stability[104], organic solvents resistance[105], specificity[106], and activity even for a thermophilic enzymes.[107]. Frances H. Arnold was awarded the 2018 Nobel Prize in Chemistry for her great achievement in the directed evolution of enzymes.[108] Protein engineering was applied for the improvement of *Hor*GH1 by investigating a series of mutants. Therefore, N222F, N294T, F417S, F417Y, and Y296F variants were produced to be assayed for lactose conversion and GOS. No improvement was achieved with

respect to the hydrolytic activity of these mutants, but F417S showed a significant improvement in the GOS production in comparison with the WT enzyme.[109]

Oh et al.[110] used protein engineering (SDM) to determine the importance of the catalytic residues in the activity of a GH1 from *Thermus caldophilus*. In this study, Glu164 and Glu338 were substituted by Gly and Gln, respectively. Extreme reduction in the enzymatic activity of β-glucosidase and β-galactosidase was noticed with the first mutation (E164G) and no activity remained in the second mutant (E338Q) which proves that Glu164 and Glu338 are the two key catalytic residues responsible for the hydrolysis of glycosidic bond in *Thermus caldophilus* GK24 β-glycosidase.[110]

At Lund University in Sweden, Pontus and his group were also able to improve the catalytic performance of a thermostable GH1 from *Thermotoga neapolitana* (TnBgl1A) by employing a rational design technique. TnBgl1A is one of the thermophilic GH1s that catalyse the synthesis of alkyl glycosides which are surfactants that have attractive properties such as biodegradability, high surface activity and high chemical stability.[111]

The activity, thermostability, and specificity of a metagenome-derived β-glucosidase (Bgl1D) were also enhanced by protein engineering (directed evolution). The first cycle of epPCR produced 50,000 colonies which were visually screening by using an agar plate-based technique that revealed a black hydrolyzing zones in 5,000 colonies. The crude enzyme activities of these interesting 5,000 colonies were assayed with the substrate ρ-nitrophenyl-β-d-glucopyranoside (ρNPG). This activity assay showed that Bgl1D58 (Q25L/K117N/M148K), Bgl1D94 (S28P/I57K/E154G), and Bgl1D47 (F68L/I70M) mutants have the highest activities towered ρNPG, so they were used as templates for the second round epPCR. From the second epPCR library which have 50,000 positive colonies, Bgl1D2 (Q25L/S28T/L115Q/K117N/M148K), Bgl1D6 (S28P/I57K/Y82S/W122G/E154G), and Bgl1D20

(Y37H/D44E/F68L/I70M/R91G) mutants showed the highest activities 2.3 to 3.6 times more than the activity of the WT Bgl1D.

The six variants of the first and second epPCRs, Bgl1D58, Bgl1D94, Bgl1D47, Bgl1D2, Bgl1D6, and Bgl1D20 were expressed in BL21 (DE3) pLysS and purified by His-tag chromatography.

The thermostability of Bgl1D2 increased almost seven times with an improvement of the half-life ($t_{1/2}$) from 45min (WT Bgl1D) to 298min (Bgl1D2), after incubating at 60 °C. The kinetics profile also showed great enhancements in the specificity and turnover rate of Bgl1D20 (295.7 k_{cat}/K_M $s^{-1}mM^{-1}$,139.1 s^{-1}) compared to the WT Bgl1D (24.8 k_{cat}/K_M $s^{-1}mM^{-1}$,13.4 s^{-1}).[112] SDM was also used to increase the efficiency of TnBgl1A for the synthesis of alkyl glycosides by a mutation of a major influence residue on substrate binding (Asn220) in TnBgl1A into a hydrophobic aromatic amino acid (Phe) which was found to be conserved in natural glycosyltransferases and determined to be an important residue at the aglycone (1) subsite of TnBgl1A. Thus, the N220F mutation improved the TnBgl1A specificity for transglycosylation (7-fold) and increased the potential yield of alkyl glycoside (17%-58%).[113] Ks5A7 (GenBank HV348683) is another β-glucosidase that has been engineered by directed evolution to improve its thermostability. It is a hyperactive glucose-tolerant β-glucosidase with limited industrial applications because of its low thermostability. The directed evolution of Ks5A7 included four rounds of epPCR random mutagenesis (Figure 1.16) and 26,000-50,000 colonies were produced on each epPCR library which were then screened by using a petri dish double-layer high-throughput screening strategy.

The 4R1 mutant (T167I/V181F/K186T/A187E/A298G) was stable at 50 °C with a great $t_{1/2}$ value (8640 min) compared to the $t_{1/2}$ of the WT Ks5A7 (1 min). Therefore, a 8640-fold increase in thermostability of Ks5A7 was obtained with the directed evolution technique

which also improved the specific activity of Ks5A7 (374.26 U/mg for 4R1 and 243.18 U/mg for the WT Ks5A7) with no loss in glucose tolerance (IC50 of 1.5 M).[114]

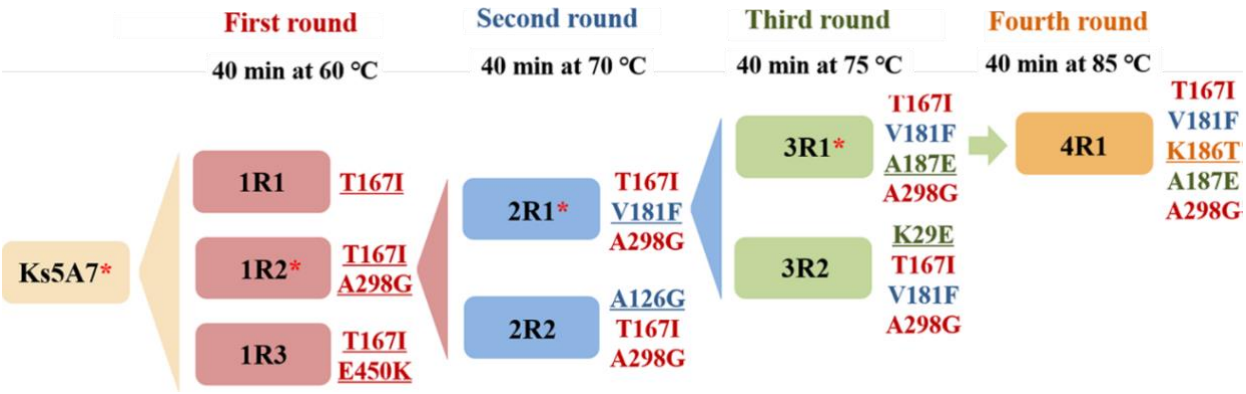


Figure 1.16. Amino acid changes in mutants of Ks5A7. Mutants selected as starting points for the next round of evolution are marked with asterisks. Newly introduced mutations in each generation are underlined. Figure reproduced from Cao et al.[114]

Beside the previously mentioned *in vitro* techniques, there are *in silico* tools which can be used to predict the engineered proteins as well as to evaluate them in their possible new positions which can then be experimentally evaluated.[115] Some of these tools are EASE-MM[116], I-Mutant[117], and mCSM.[118] In 2018, Milos and his group in Technical University of Ostrava were able to develop these classic tools which were limited to single-point mutations into an advanced web server called FireProt that is able to design multiple point mutants for thermostable proteins.[119]

In this study, protein engineering aided the introduction of amino acid mutations by mapping the *Bb*MYR active site onto the extremophilic *Tno*GH1 enzyme first, and then onto *Hor*GH1 to further confirm the key role played by selected residues in the recognition and hydrolysis of GSLs.

1.4 Aims and objectives

This study will focus on β-glucosidases from thermophilic and halophilic sources, which are particularly attractive because of their biotechnological advantages over many mesophilic biocatalysts. Specifically, the engineering of β-thioglucosidase activity into two glycosyl hydrolases isolated from the thermophilic bacterium *Thermus nonproteolyticus* (*Tno*GH1) and the halo-thermotolerant *Halothermothrix orenii* (*Hor*GH1) will be investigated. It is expected that a variant can be created which is capable of enhanced hydrolytic activity towards thioglycosides i.e., the hydrolysis of the sulphur bond contained in glucosinolate (GSLs). In nature, the only β-thioglucosidase capable of hydrolyzing sulphur-containing secondary metabolites (GSLs) is an enzyme called myrosinase.

Myrosinases are not particularly extremophilic and hence it is of interest to generate new biocatalysts from halo-thermophiles that are able to catalyse the hydrolysis of thioglycosides, which can be thermostable with considerable activity in organic solvents and have other unique features in biotechnological processes especially at high temperature which make them attractive in industrial biocatalysts.

Therefore, this research will focus on:

- GH1 from halophile-thermophiles range
- Engineering novel extremophilic GH1s towards thioglycoside substrates
- Exploring the activities of the WT enzymes and the mutants toward standard substrates (O-glucosides) and thioglucosides
- Investigation of the kinetics parameters of the WT enzyme and the mutants

1.5 Bibliography

1 R. M. Sharpe and D. S. Irvine, *Bmj*, 2004, **328**, 447–451.

2 P. Anastas and N. Eghbali, *Chem. Soc. Rev.*, 2010, **39**, 301–312.

3 I. T. Horváth and P. T. Anastas, *Chem. Rev.*, 2007, **107**, 2167–2168.

4 A. Brunning, The Twelve Principles of Green Chemistry: What it is, & Why it Matters, https://www.compoundchem.com/2015/09/24/green-chemistry/#disqus_thread.

5 Y. Ni, D. Holtmann and F. Hollmann, *ChemCatChem*, 2014, **6**, 930–943.

6 W. Shao, G. Yeretssian, K. Doiron, S. N. Hussain and M. Saleh, *J. Biol. Chem.*, 2007, **282**, 36321–36329.

7 J. M. Reiner, *Behavior of Enzyme Systems: An Analysis of Kinetics and Mechanism*, Burgess Publishing Company, 1959.

8 Y. F. Rego, M. P. Queiroz, T. O. Brito, P. G. Carvalho, V. T. de Queiroz, Â. de Fátima and F. Macedo, *J. Adv. Res.*, 2018, **13**, 69–100.

9 C. C. Worthington, V. Worthington and A. Worthington, Introduction to enzymes, http://www.worthington-biochem.com/introbiochem/Enzymes.pdf.

10 S. Wenda, S. Illner, A. Mell and U. Kragl, *Green Chem.*, 2011, **13**, 3007–3047.

11 M. F. Ali, B. M. El Ali and J. G. Speight, *Handbook of Industrial Chemistry: Organic Chemicals*, McGraw-Hill Companies, New York, 2005.

12 T. Hudlicky and J. W. Reed, *Chem. Soc. Rev.*, 2009, **38**, 3117–3132.

13 B. M. Nestl, S. C. Hammer, B. A. Nebel and B. Hauer, *Angew. Chemie - Int. Ed.*, 2014, **53**, 3070–3095.

14 J. Thiem, *FEMS Microbiol. Rev.*, 1995, **16**, 193–211.

15 G. M. Whitesides and C.-H. Wong, *Angew. Chemie Int. Ed. English*, 1985, **24**, 617–638.

16 S. Li, X. Yang, S. Yang, M. Zhu and X. Wang, *Comput. Struct. Biotechnol. J.*, 2012, **2**,

e201209017.

17 D. C. Demirjian, F. Morís-Varas and C. S. Cassidy, *Curr. Opin. Chem. Biol.*, 2001, **5**, 144–151.

18 S. Gupta, P. Sharma, K. Dev and A. Sourirajan, *Biochem. Res. Int.*

19 M. W. Bauer, L. E. Driskill and R. M. Kelly, *Curr. Opin. Biotechnol.*, 1998, **9**, 141–145.

20 I. V. O. Valchev and P. Tsekova, *Appita J.*, 2010, **63**, 45–52.

21 B. Henrissat and G. Davies, *Curr. Opin. Struct. Biol.*, 1997, **7**, 637–644.

22 G. Davies and B. Henrissat, *Structure*, 1995, **3**, 853–859.

23 Y. Bhatia, S. Mishra and V. S. Bisaria, *Crit. Rev. Biotechnol.*, 2002, **22**, 375–407.

24 D. G. Naumoff, *Biochem.*, 2011, **76**, 622–635.

25 C. Murphy, J. Powlowski, M. Wu, G. Butler and A. Tsang, *Database*, 2011, **2011**, 1–14.

26 S. Park, B. Lee and Kwanhwa Park, *J. Nutr. Heal. Food Eng.*

27 M. Okuyama, M. Kitamura, H. Hondoh, M. S. Kang, H. Mori, A. Kimura, I. Tanaka and M. Yao, *J. Mol. Biol.*, 2009, **392**, 1232–1241.

28 B. Henrissat, *Biochem. J.*, 1991, **280**, 309–316.

29 Glycoside Hydrolase family classification, http://www.cazy.org/Glycoside-Hydrolases.html.

30 V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucleic Acids Res.*, 2014, **42**, 490–495.

31 R. Opassiri, B. Pomthong, T. Akiyama, M. Nakphaichit, T. Onkoksoong, M. Ketudat Cairns and J. R. Ketudat Cairns, *Biochem. J.*, 2007, **408**, 241–249.

32 Glycoside Hydrolase Family 13, http://www.cazy.org/GH13.html.

33 M. R. Stam, E. G. J. Danchin, C. Rancurel, P. M. Coutinho and B. Henrissat, *Protein Eng. Des. Sel.*, 2006, **19**, 555–562.

34 M. Okuyama, *Biosci. Biotechnol. Biochem.*, 2011, **75**, 2269–2277.

35 X. Zang, M. Liu, Y. Fan, J. Xu, X. Xu and H. Li, *Biotechnol. Biofuels*, 2018, **11**, 1–13.

36 L. F. Mackenzie, Q. Wang, R. A. J. Warren and S. G. Withers, *J. Am. Chem. Soc.*, 1998, **120**, 5583–5584.

37 S. J. Williams and S. G. Withers, *Aust. J. Chem.*, 2002, **55**, 3–12.

38 M. Jahn, J. Marles, R. A. J. Warren and S. G. Withers, *Angew. Chemie - Int. Ed.*, 2003, **42**, 352–354.

39 M. Jahn, H. Chen, J. Müllegger, J. Marles, R. A. J. Warren and S. G. Withers, *Chem. Commun.*, 2004, **4**, 274–275.

40 R. R. Singhania, A. K. Patel, R. K. Sukumaran, C. Larroche and A. Pandey, *Bioresour. Technol.*, 2013, **127**, 500–507.

41 M. Henn-Sax, B. Höcker, M. Wilmanns and R. Sterner, *Biol. Chem.*, 2001, **382**, 1315–1320.

42 J. a Silverman, R. Balakrishnan and P. B. Harbury, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 3092–3097.

43 Y. Nijikken, T. Tsukada, K. Igarashi, M. Samejima, T. Wakagi, H. Shoun and S. Fushinobu, *FEBS Lett.*, 2007, **581**, 1514–1520.

44 L. M. Zanphorlin, P. O. De Giuseppe, R. V. Honorato, C. C. C. Tonoli, J. Fattori, E. Crespim, P. S. L. De Oliveira, R. Ruller and M. T. Murakami, *Sci. Rep.*, 2016, **6**, 1–14.

45 W. P. Burmeister, S. Cottaz, H. Driguez, R. Iori, S. Palmieri and B. Henrissat, *Structure*, 1997, **5**, 663–675.

46 J. Jenkins, L. Lo Leggio, G. Harris and R. Pickersgill, *FEBS Lett.*, 1995, **362**, 281–285.

47 X. Wang, X. He, S. Yang, X. An, W. Chang and D. Liang, *J. Bacteriol.*, 2003, **185**, 4248–4255.

48 X. Y. He, X. Q. Wang, S. J. Yang, W. R. Chang and D. C. Liang, *Acta Crystallogr D Biol Crystallogr*, 2001, **57**, 1650–1651.

49 K. Mavromatis, N. Ivanova, I. Anderson, A. Lykidis, S. D. Hooper, H. Sun, V. Kunin, A. Lapidus, P. Hugenholtz, B. Patel and N. C. Kyrpides, *PLoS One*, 2009, **4**, e4192.

50 L. D. Kori, A. Hofmann and B. K. C. Patel, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, 2011, **67**, 111–113.

51 N. Hassan, T. H. Nguyen, M. Intanon, L. D. Kori, B. K. C. Patel, D. Haltrich, C. Divne and T. C. Tan, *Appl. Microbiol. Biotechnol.*, 2015, **99**, 1731–1744.

52 A. M. E. Jones, M. Bridges, A. M. Bones, R. Cole and J. T. Rossiter, *Insect Biochem. Mol. Biol.*, 2001, **31**, 1–5.

53 E. Kazana, T. W. Pope, L. Tibbles, M. Bridges, J. A. Pickett, A. M. Bones, G. Powell and J. T. Rossiter, *Proc. R. Soc. B Biol. Sci.*, 2007, **274**, 2271–2277.

54 K. Witzel, F. S. Hanschen, R. Klopsch, S. Ruppel, M. Schreiner and R. Grosch, *Front. Plant Sci.*, 2015, **6**, 1–11.

55 H. Nong, J. M. Zhang, D. Q. Li, M. Wang, X. P. Sun, Y. J. Zhu, J. Meijer and Q. H. Wang, *J. Integr. Plant Biol.*, 2010, **52**, 879–890.

56 D. J. Kliebenstein, J. Gershenzon and T. Mitchell-Olds, *Genetics*, 2001, **159**, 359–370.

57 D. Šamec, I. Pavlović and B. Salopek-Sondi, *Phytochem. Rev.*, 2017, **16**, 117–135.

58 N. Rakariyatham, B. Butrindr, H. Niamsup and L. Shank, *Brazilian J. Microbiol.*, 2005, **36**, 242–245.

59 A. Bourderioux, M. Lefoix, D. Gueyrard, A. Tatibouët, S. Cottaz, S. Arzt, W. P. Burmeister and P. Rollin, *Org. Biomol. Chem.*, 2005, **3**, 1872–1879.

60 B. A. Halkier and J. Gershenzon, *Annu. Rev. Plant Biol.*, 2006, **57**, 303–333.

61 V. Dufour, M. Stahl and C. Baysse, *Microbiology*, 2015, **161**, 229–243.

62 I. Winde and U. Wittstock, *Phytochemistry*, 2011, **72**, 1566–1575.

63 J. Higdon, Cruciferous Vegetables, https://lpi.oregonstate.edu/book/export/html/388.

64 A. M. Bones and J. T. Rossiter, *Phytochemistry*, 2006, **67**, 1053–1067.

65 W. P. Burmeister, S. Cottaz, P. Rollin, A. Vasella and B. Henrissat, *J. Biol. Chem.*, 2000, **275**, 39385–39393.

66 H. Husebye, S. Arzt, W. P. Burmeister, F. V Ha and A. Brandt, *Insect Biochem. Mol. Biol.*, 2005, **35**, 1311–1320.

67 A. M. E. Jones, P. Winge, A. M. Bones, R. Cole and J. T. Rossiter, *Insect Biochem. Mol. Biol.*, 2002, **32**, 275–284.

68 S. J. Pilkis, I. T. Weber, R. W. Harrison and G. I. Bell, *J. Biol. Chem.*, 1994, **269**, 21925–21928.

69 M. Shepherd and A. Steele, What Is Glucokinase (GCK)?, https://www.diabetesgenes.org/what-is-mody/what-is-glucokinase-gck/.

70 Z. Wang, F. Ping, Q. Zhang, J. Zheng, H. Zhang, M. Yu, W. Li and X. Xiao, *J. Diabetes Investig.*, 2018, **9**, 199–203.

71 J. J. M. Cachumba, F. A. F. Antunes, G. F. D. Peres, L. P. Brumano, J. C. Dos Santos and S. S. Da Silva, *Brazilian J. Microbiol.*, 2016, **47**, 77–85.

72 R. Pieters, S. P. Hunger, J. Boos, C. Rizzari and C. Pui, *Cancer*, 2012, **117**, 238–249.

73 S. Al-Zuhair, K. B. Ramachandran, M. Farid, M. K. Aroua, P. Vadlani, S. Ramakrishnan and L. Gardossi, *Enzyme Res.*, 2011, **2011**, 2–4.

74 D. Basketter, N. Berg, C. Broekhuizen, M. Fieldsend, S. Kirkwood, C. Kluin, S. Mathieu and C. Rodriguez, *Regul. Toxicol. Pharmacol.*, 2012, **64**, 117–123.

75 A. R. Tapre and R. K. Jain, *Int. Food Res. J.*, 2014, **21**, 447–453.

76 H. Breithaupt, *EMBO Rep.*, 2001, **2**, 968–971.

77 N. Rueda, J. C. S. dos Santos, C. Ortiz, R. Torres, O. Barbosa, R. C. Rodrigues, Á. Berenguer-Murcia and R. Fernandez-Lafuente, *Chem. Rec.*, 2016, **16**, 1436–1455.

78 U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore and K. Robins, *Nature*, 2012, **485**, 185–194.

79 A. Nobili, F. Steffen-Munsberg, H. Kohls, I. Trentin, C. Schulzke, M. Höhne and U. T. Bornscheuer, *ChemCatChem*, 2015, **7**, 757–760.

80 N. Gupta and E. T. Farinas, *Protein Eng. Des. Sel.*, 2010, **23**, 679–682.

81 H. Zhang, J. Sang, Y. Zhang, T. Sun, H. Liu, R. Yue, J. Zhang, H. Wang, Y. Dai, F. Lu and F. Liu, *Int. J. Biol. Macromol.*, 2019, **137**, 1190–1198.

82 M. Ishibashi, R. Kawanabe, N. Amaba, S. Arai, F. A. Laksmi, K. Komori and M. Tokunaga, *Protein Expr. Purif.*, 2018, **145**, 39–44.

83 P. Martinez, M. E. Van Dam, A. C. Robinson, K. Chen and F. H. Arnold, *Biotechnol. Bioeng.*, 1992, **39**, 141–147.

84 S. Banerjee, Polymerase Chain Reaction (PCR), https://thebiologynotes.com/polymerase-chain-reaction-pcr/.

85 D. A. Estell, T. P. Graycar and J. A. Wells, *J. Biol. Chem.*, 1985, **260**, 6518–6521.

86 C. Strub, C. Alies, A. Lougarre, C. Ladurantie, J. Czaplicki and D. Fournier, *BMC Biochem.*, 2004, **5**, 1–6.

87 Jaroslav Bendl, Icahn School of Medicine at Mount Sinai, 2016.

88 H. Hussain and N. F. M. Chong, *Biomed Res. Int.*

89 O. Edelheit, A. Hanukoglu and I. Hanukoglu, *BMC Biotechnol.*, 2009, **9**, 1–8.

90 M. A. Shampo and R. A. Kyle, *Mayo Clin. Proc.*, 2003, **78**, 804.

91 L. Xu, X. Liu, Z. Yin, Q. Liu, L. Lu and M. Xiao, *Appl. Microbiol. Biotechnol.*, 2016, **100**, 10385–10394.

92 C. Wang, F. P. Douillard, W. Zhou and Y. Hao, *Curr. Microbiol.*, 2014, **69**, 423–428.

93 X. J. Fan, C. Yang, C. Zhang, H. Ren and J. D. Zhang, *Appl. Biochem. Biotechnol.*, 2018, **184**, 12–24.

94 J. Sauer, B. W. Sigurskjold, U. Christensen, T. P. Frandsen, E. Mirgorodskaya, M. Harrison, P. Roepstorff and B. Svensson, *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.*, 2000, **1543**, 275–293.

95 X. C. Wang, S. P. You, J. X. Zhang, Y. M. Dai, C. Y. Zhang, W. Qi, T. Y. Dou, R. X. Su and Z. M. He, *Enzyme Microb. Technol.*, 2018, **118**, 50–56.

96 X. Liu, L. Cao, J. Zeng, Y. Liu and W. Xie, *Int. J. Biol. Macromol.*, 2019, **136**, 1052–1059.

97 E. M. Brustad and F. H. Arnold, *Curr. Opin. Chem. Biol.*, 2011, **15**, 201–210.

98 M. S. Packer and D. R. Liu, *Nat. Rev. Genet.*, 2015, **16**, 379–394.

99 M. D. Lane and B. Seelig, *Curr. Opin. Chem. Biol.*, 2014, **22**, 129–136.

100 M. Goldsmith and D. S. Tawfik, *Curr. Opin. Struct. Biol.*, 2012, **22**, 406–412.

101 O. Kuchner and F. H. Arnold, *Trends Biotechnol.*, 1997, **15**, 523–530.

102 C. Zeymer and D. Hilvert, *Annu. Rev. Biochem.*, 2018, **87**, 131–157.

103 J. L. Porter, R. A. Rusli and D. L. Ollis, *ChemBioChem*, 2016, **17**, 197–203.

104 V. G. H. Eijsink, S. GÅseidnes, T. V. Borchert and B. Van Den Burg, *Biomol. Eng.*, 2005, **22**, 21–30.

105 F. Zhu, B. He, F. Gu, H. Deng, C. Chen, W. Wang and N. Chen, *J. Biotechnol.*, 2020, **309**, 68–74.

106 J. K. Lee, E. Jeong, J. Lee, M. Jung, E. Shin, Y. hoon Kim, K. Lee, I. Jung, D. Kim, S. Kim and J. S. Kim, *Nat. Commun.*

107 G. Li, M. A. Maria-Solano, A. Romero-Rivera, S. Osuna and M. T. Reetz, *Chem. Commun.*, 2017, **53**, 9454–9457.

108 R. Fasan, S. B. Jennifer Kan and H. Zhao, *ACS Catal.*, 2019, **9**, 9775–9788.

109 N. Hassan, B. Geiger, R. Gandini, B. K. C. Patel, R. Kittl, D. Haltrich, T. H. Nguyen, C. Divne and T. C. Tan, *Appl. Microbiol. Biotechnol.*, 2016, **100**, 3533–3543.

110 E. J. Oh, Y. J. Lee, J. J. Choi, M. S. Seo, M. S. Lee, G. A. Kim and S. T. Kwon, *J. Microbiol. Biotechnol.*, 2008, **18**, 287–294.

111 M. Y. Rather and S. Mishra, *Sustain. Chem. Process.*, 2013, **1**, 1–15.

112 B. Yin, Q. Hui, M. Kashif, R. Yu, S. Chen, Q. Ou, B. Wu and C. Jiang, *Int. J. Mol. Sci.*

113 P. Lundemo, P. Adlercreutz and E. N. Karlsson, *Appl. Environ. Microbiol.*, 2013, **79**, 3400–3405.

114 L. Cao, S. Li, X. Huang, Z. Qin, W. Kong, W. Xie and Y. Liu, *J. Agric. Food Chem.*, 2018, **66**, 13228–13235.

115 M. C. Ebert and J. N. Pelletier, *Curr. Opin. Chem. Biol.*, 2017, **37**, 89–96.

116 L. Folkman, B. Stantic, A. Sattar and Y. Zhou, *J. Mol. Biol.*, 2016, **428**, 1394–1405.

117 E. Capriotti, P. Fariselli and R. Casadio, *Nucleic Acids Res.*, 2005, **33**, 306–310.

118 D. E. V. Pires, D. B. Ascher and T. L. Blundell, *Bioinformatics*, 2014, **30**, 335–342.

119 M. Musil, J. Stourac, J. Bendl, J. Brezovsky, Z. Prokop, J. Zendulka, T. Martinek, D. Bednar and J. Damborsky, *Nucleic Acids Res.*, 2017, **45**, W393–W399.

Chapter 2: Material and Methods

2.1 Reagents

Substrates 4-nitrophenyl β-D-thioglucopyranoside (pNT-Glc) and 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc) were purchased from Carbosynth. *Escherichia coli* bacterial strains and the QuikChange II Site-Directed Mutagenesis Kit were from Agilent Technologies. The plasmid miniprep kit was from Macherey-Nagel. Growth media and assay components were procured from Fisher Scientific. All other chemicals were purchased from Sigma.

2.2 General instruments

- DNA amplification, site directed mutagenesis, and temperature stability assays were carried out using a PCR thermal Cycler SensoQuest® Labcycler 48 and the Applied Biosystems Verti® Thermal Cycler.
- DNA electrophoresis and Protein SDS-PAGE were performed in a power supply EPS 301 from Amersham Biosciences.
- Plasmid purification was carried out in the NucleoSpin Plasmid purification kit from Macherey-Nagel®, PCR product purification was performed with the QIAquick PCR purification kit from Qiagen® and agarose gel DNA extraction was performed using the QIAquick Gel Extraction kit from Qiagen®.
- Solid medium plates and liquid cultures were incubated in a Thermo Scientific MaxQ6000 static incubator with temperature control from 4 to 60 °C or the New Brunswick Scientific Model C24 for temperatures from 30 to 37 °C.
- General centrifuge cycles were performed using Thermo® Microcentrifuge Accuspin Micro 17R, with controlled temperature for volumes up to 2 mL.

- Cell debris were removed from cell lysate suspensions using a centrifuge Thermo® Haraeus Multifuge with Thermo Fiberlite F15 for volumes larger than 2 mL, at 4 °C.
- Cells were lysed by sonication using a Fisher brand Ultrasonic Liquid processor FB120 with timer and pulser from Fisher brand.
- Protein purifications were carried out in a chromatography system AKTAPrime ® Start from GE Healthcare using HisTrap FF 1mL, HisTrap FF 5mL (GE Healthchare) or Ni-NTA Superflow Cartridges (Qiagen).
- DNA, Protein quantification, absorbance reading and scanning, and conventional enzymatic tests and general well-reading were performed on an EPOCH2 microplate and cuvette reader from BioTek®.
- General pipetting was performed with Micropipettes P5000, P1000, P200, P20, P10 and P2.5 from StarLab® and Eppendorf® Multipipette E2.
- pH-meter Orion Star A111, calibrated regularly with standard solutions (pH 4, pH 7 and pH 10 from Sigma-Aldrich).
- Chemicals were weighted using balances model PS5602, model CSC2000 or an analytical series PAS214C, all from Thermo Scientific®.
- Culture media and glassware were generally sterilised using a Thermostatic autoclave ST-19T from Dixons Surgical Instruments Ltd®
- Further common instruments used during the experiments: Lab- Line® Multi-Block Heater, Benchtop Shaker Carl Stuart Limited IKA®, Vortex mixer Wizard from FisherBrand® and microwave cooker from Daewoo®.

2.3 Cell culture and manipulation

2.3.1 Bacterial strains

The following *Escherichia coli* strains were used:

- *E. coli* XL10-Gold as propagation strain

Genotype: endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCBhsdSMR-mrr)173 tetR F'[proAB lacIqZΔM15 Tn10(TetR Amy CmR)]

- *E. coli* BL21(DE3) as protein expression strain

Genotype: *E. coli* str. B F^- ompT gal dcm lon $hsdS_B$ ($r_B^- m_B^-$) λ (DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+] $_{K-12}$(λS).

2.3.2 Culture media, buffers, and antibiotics preparation

- LB medium: N-Z amine (10 g), yeast extract (5 g), NaCl (10 g) and dH_2O (1 L).
- LB agar plate: N-Z amine (10 g), yeast extract (5 g), NaCl (10 g), agar (15 g) and dH_2O (1 L).
- TB medium: N-Z amine (12 g), yeast extract (24 g), glycerol (5 g), K_2HPO_4 (2.2 g) and KH_2PO_4 (9.4 g) and dH_2O (1 L).
- NZY+ medium: N-Z amine casein enzymatic hydrolysate (0.5 g), yeast extract (0.25 g), and NaCl (0.25 g) were dissolved in 50 mL dH_2O and the pH adjusted to 7.5. After autoclaving, $MgCl_2$ (1 M, 0.625 mL), MgSO4 (1 M, 0.625 mL), and glucose (2 M, 0.5 mL) were filter-sterilized and added to the 50 mL NZY^+ broth buffer prior to use.
- Universal Buffer: 25 mM citric acid (0.96 g), 25 mM KH_2PO_4 (0.68 g), 25 mM Tris (0.61 g), 12.5 mM sodium tetraborate decahydrate (1.91 g), and 25 mM KCl (0.37 g) were dissolved in 200 ml of dH_2O, then the buffer was filtered and degassed.

- Ampicillin stock solution: (100 mg/mL) 1 g of sodium ampicillin was dissolved in sufficient dH_2O to make a final volume of 10 mL. An 0.22 µm sterile filter was first prewashed by drawing through 50-100 mL of dH_2O to be used to sterilize the solution. Then the ampicillin solution was passed through the washed filter. The ampicillin was stored in aliquots at -20 °C for 1 year (or at 4 °C for 3 months).
- IPTG solution (1 M) 2.38 g of isopropyl β-D-1-thiogalactopyranoside was dissolved in sufficient dH_2O to make a final volume of 10 mL. An 0.22 µm sterile filter was first prewashed by drawing through 50-100 mL of dH_2O to be used to sterilize the solution. Then the IPTG solution was passed through the washed filter. The IPTG stored in aliquots at -20 °C for 1 year (or at 4 °C for 3 months).

2.3.3 Preparation of chemically competent cells

A single colony of *E. coli* (either XL10-Gold® or BL21 (DE3)) was inoculated into LB medium (50 mL) and grown at 37 °C, 150 rpm to an OD_{600} of ca. 0.4. The cultures were then placed on ice for 20 min and subsequently harvested by centrifugation (3000 g, 15 min, 4 °C). Pellets were resuspended in 10 mL of an ice-cold solution of magnesium chloride (100 mM), combined, and harvested by centrifugation (3000 g, 15 min, 4 °C). The pellet was resuspended in 20 mL of an ice-cold solution of calcium chloride (100 mM) and incubated on ice for ca. 30 min. After harvesting by centrifugation (3000 g, 15 min, 4 °C), pellets were resuspended in 10 mL of an ice cold solution of calcium chloride (85 mM) and glycerol (15 % v/v). Cells were harvested (3,000 g, 10 min, 4 °C), resuspended in 0.5 mL of the previous solution, aliquoted (80 µL) to 1.5 mL microfuge tubes, and frozen at -80 °C.

2.3.4 Chemically competent cell transformation

The plasmid DNA (pET45b Amp resistant) harbouring the GH1 gene was used to transform chemically competent *E. coli* cells under sterile conditions.

The plasmid (1 µL) was placed in a micro-centrifuge tube containing BL21 cells (40 µL). The competent cells were left on ice for 30 min, then heat-shocked at 42 °C for 1 min and 30 s. It was then cooled on ice for a subsequent 5 min. Autoclaved LB media was then added (250 µL) to the sample and left shaking (180 rpm) at 37 °C for 1 h. Aliquots of 50 and 200 µL of the culture were plated to LB agar dishes containing ampicillin (150 µL in each 150 ml LB). They were grown O/N at 37 °C; the LB agar plates were then stored at 4 °C..

A single colony was isolated from each plate and inoculated in individual falcon tubes containing liquid LB medium (6 mL) and ampicillin (6 µL). The tubes were incubated at 37 °C shaking incubator for 16 h.

2.4 General procedures:

2.4.1 Site-directed mutagenesis (SDM)

2.4.1.1 Mutagenic primers design

- *Tno*GH1 mutations:

Protein sequence:

The below FASTA file represents the sequence of the *Tno*GH1 enzyme following N-terminal His tagging. The active site motifs are in bold.

in pET45b Amp resistant
>1NP2_1|Chains A,B|beta-glycosidase|*Thermus nonproteolyticus* / Mw: 49 kDa

MTENAEKFLWGVATSAYQIEGATQEDGRGPSIWDTFARRPGAIRDGSTGEPACDHYHRYEEDIALMQSLGVGVYRFSVAWPRI
LPEGRGRINPKGLAFYDRLVDRLLAAGITPFLTLYHWDLPQALEDRGGWRSRETAFAFAEYAEAVARALADRVPFFATL**NEP**W
CSAFLGHWTGEHAPGLRNLEAALRAAHHLLLGHGLAVEALRAAGARRVGIVLNFAPAYGEDPEAVDVADRYHNRYFLDPILGR
GYPESPFQDPPPAPILSRDLEAIARPLDFLGVNYYAPVRVAPGTGPLPVRYLPPEGPVTAMGWEVYPEGLYHLLKRLGREVPW
PL**YITENG**AAYPDLWTGEAVVEDPERVAYLEAHVEAALRAREEGVDLRGYFVWSLMDNFEWAFGYTRRFGLYYVDFPSQRRIP
KRSALWYRERIARAQTGGSAH

Mutations were introduced into *Tno*GH1[1] using the QuikChange SDM kit (Agilent Technologies) according to the manufacturers instructions.

The mutagenic primers were designed using QuikChange Primer Design Program available online at www.agilent.com/genomics/qcpd. Table 2.1 shows the three *Tno*GH1 mutagenic primers which were used individually to generate single and double mutants (generating 3 additional variants with all possible permutations).

Table 2.1. List of primers used in this study for *Thermus nonproteolyticus* glycoside hydrolase SDM.

Enzyme	Primer (5'-3')
*Tno*GH1-Lue171Lys	GTGCTCTGCGTTTAAAGGTCATTGGACG
*Tno*GH1-Val287Arg	TACTACGCTCCGCGCAGAGTAGCACCTG
*Tno*GH1-His178Tyr	TCATTGGACGGGTGAATATGCACCGGGTCTGCG

- *Hor*GH1 mutations:

Protein sequence:

The below FASTA file represents the sequence of the HorGH1 enzyme following N-terminal His tagging. The active site motifs are in bold.

in pET45b Amp resistant

>3TA9_1|Chains A,B|Glycoside hydrolase family 1|*Halothermothrix orenii* / Mw: 52 kDa

MHHHHHHMAKIIFPEDFIWGAATSSYQIEGAFNEDGKGESIWDRFSHTPGKIENGDTGDIACDHYHLYREDIELMKEIGI
RSYRFSTSWPRILPEGKGRVNQKGLDFYKRLVDNLLKANIRPMITLYHWDLPQALQDKGGWTNRDTAKYFAEYARLMFEE
FNGLVDLWVTH**NEP**WVVAFEGHAFGNHAPGTKDFKTALQVAHHLLLSHGMAVDIFREEDLPGEIGITLNLTPAYPAGDSE
KDVKAASLLDDYINAWFLSPVFKGSYPEELHHIYEQNLGAFTTQPGDMDIISRDIDFLGINYYSRMVVRHKPGDNLFNAE
VVKMEDRPSTEMGWEIYPQGLYDILVRVNKEYTDKPL**YITENG**AAFDDKLTEEGKIHDEKRINYLGDHFKQAYKALKDGV
PLRGYYVWSLMDNFEWAYGYSKRFGLIYVDYENGNRRFLKDSALWYREVIEKGQVEAN

*H*orGH1 was mutated using the QuikChange SDM kit (Agilent Technologies) according to the manufacturers instructions.. As with *Tno*GH1, the *Hor*GH1mutagenic primers were designed using QuikChange Primer Design Program available online at www.agilent.com/genomics/qcpd. The three different mutagenic primers of *H*orGH1 which were used individually to generate single and double mutants (generating 3 additional variants with all possible permutations) are shown in Table 2.2.

Table 2.2. List of primers used in this study for *Halothermothrix orenii* glycoside hydrolase SDM.

Enzyme	Primer (5'-3')
*Hor*GH1-Glu173Lys	GAACCGTGGGTTGTGGCTTTTAAAGGTCATG
*Hor*GH1-Met299Arg	AATTATTACAGTAGAAGGGTCGTACGCCATAAG
*Hor*GH1-His180Tyr	GCATTCGGCAATTATGCCCCTGGTACA

2.4.1.2 Thermocycling reaction

The mutant strand synthesis reactions for the thermal cycling were prepared using 2.5 µL of 10× QuikChange Lightning Multi reaction buffer, X µL to final volume of 25 µL of dH_2O, 0.3 µL of QuikSolution, 100 ng of ds-DNA template, 100 ng of each primer (for 1–3 primers), 1 µL of dNTP mix, and 1 µL of QuikChange Lightning Multi enzyme blend. These components were added in order, then mixed gently by pipetting or tapping the reaction tube.

The thermocycling conditions of the reaction were 30 cycles of denaturation, annealing, and elongation phases as following (20 s in 95 °C, 30 s in 55 °C, 30 s in 65 °C). Then the reactions were cooled to ≤37 °C by placing the reaction thin-walled tube on ice for 2 min.

2.4.1.3 Dpn I digestion of the original plasmid

In order to digest the original unmutated plasmid, the amplification product was digested by adding 1 µL of Dpn I restriction enzyme directly into the amplification reaction and the solution was gently mix by pipetting the reaction mixture up and down several times. After that the reaction mixture was centrifuge for 1 min in a microcentrifuge and then incubated for 5 - 60 min at 37 °C.

2.4.1.4 Transformation of XL-10-Gold ultracompetent cells

After generating the mutated plasmid via PCR, the plasmid was introduced into the XL-10-Gold Ultracompetent cells as follows:

The XL10-Gold ultracompetent cells were gently thawed on ice. 45 µL of the ultracompetent cells was add into a new autoclaved 1.5 mL Eppendorf tube and 2 µL of XL10-Gold β-mercaptoethanol (β-ME) mix was added to the 45 µL ultracompetent cells. The contents of the tube were gently swirled and then incubated on ice and swirled gently every 2 min for 10 min. 1.5 µL of the *Dpn* I-treated DNA from each mutagenesis reaction (see 2.5.2.3) was transferred to the ultracompetent cells and the transformation reaction was gently swirled to mix, then incubated on ice for 30 min.

50 mL of NZY$^+$ broth buffer was prepared as described above in the media preparation section (see 2.3.4). The NZY$^+$ broth was preheated at 42 °C to be used in a further step.

The transformation reactions tubes were heat-shocked at 42 °C for 30 s then they were incubated on ice for 2 min.

0.5 mL of the preheated (42°C) NZY$^+$ broth was added to each transformation reactions tube. The tubes were incubated at 37 °C shaking incubator (225-250 rpm) for 1 h. 50 µL of the transformation reaction was plated on agar plates containing the appropriate antibiotic for

selection of the plasmid vector (supplemented with amp, 150 µL in each 150 ml LB). The transformation plates were incubated at 37 °C for >16 h.

2.4.2 Plasmid DNA isolation

The plasmid was isolated using the NucleoSpin Plasmid purification kit according to the manufacturers instructions; isolation of high-copy plasmid DNA from *E. coli*.

In order to isolate the plasmid, mini prep was prepared as followed; a single colony from the SDM plate was incubated in 5 mL LB medium (supplemented with amp, 5 µL). After growing the cells O/N at 37 °C, 200 rpm, the cells were gently pelleted (11,000 x *g*, RT, 1 min) and the supernatant was discarded. Then the pellet was lysed by adding 250 µL Buffer A1 (resuspending buffer), 250 µL Buffer A2 (lysis buffer), and 300 µL Buffer A3 (neutralization buffer). The lysate was clarified by centrifugation for 10 min at 11,000 x *g* at RT and the DNA was bound by placing a NucleoSpin® Plasmid / Plasmid (NoLid) Column onto a 2 mL collection tube. 750 µL of the colourless supernatant was transferred onto the column. After centrifugation for 1 min at 11,000 x *g*, the flow through was discarded and the column placed back into the collection tube. Then the silica membrane was washed by 600 µL Buffer A4 (supplemented with ethanol). The collection tube was then centrifuged for 1 min at 11,000 x *g* at RT. To release the plasmid DNA, the spin column was placed into a new 1.5 mL Eppendorf tube and 30 µL of low salt buffer (elution buffer) was added to the centre of the membrane and incubated for 2 min at RT. Then it was centrifuged for 1 min at 11,000 x *g* at RT. The plasmid DNA concentration was finally quantified by using an EPOCH2 microplate and cuvette reader (BioTeck).

2.4.3 Plasmid DNA propagation

1 µL of plasmid DNA was gently mixed with 40 µL of XL10-Gold competent cells under sterile conditions. The mixture was placed on ice for 30 min, then heat shocked at 42 °C for 90 s and placed back on ice for 5 min. Subsequently, 250 µL of autoclaved LB media was added and then the culture incubated at 37 °C for 1 h. After that, 50 µL and 200 µL of the culture was plated on agar plates (supplemented with amp, 150 µL in each 150 ml LB) and the plates were incubated at 37 °C O/N then stored at 4 °C.

2.4.4 β-Glycoside hydrolases expression

2.4.4.1 Small scale expression

E. coli BL21 (DE3) was transformed with a plasmid harbouring one of the β-glycosylase genes and a single colony was inoculated into 10 mL of LB medium (supplemented with amp, 10 µL). After growing the cells at 37 °C, 200 rpm for ca. 7 h, the cells were gently pelleted (1,400 g, 4 °C, 10 min) and resuspended in 10 mL of fresh LB (supplemented with amp, 10 µL). IPTG (1mM) was added and cells grown at 37 °C, 200 rpm for 16 h. Cells were harvested (3500 g, 4 °C, 20 min) and resuspended in 0.5 mL buffer (HEPES 50 mM, sodium chloride 150 mM, pH 7.5). The cells were lysed (on ice) by sonication (10 s on, 10 s off, 8 cycles, Amplitude: 40%) and the debris pelleted (17,000 g, 4 °C, 15 min). The cell-free extracts were then stored at RT.

2.4.4.2 Medium scale expression

E. coli BL21 (DE3) was transformed with a plasmid harbouring one of the β-glycosylase genes and a single colony was inoculated into 5 mL of LB medium (supplemented with amp, 5 µL) and incubated at 37 °C, 200 rpm O/N. Then, 1 mL of the starter culture was inoculated into

50 mL of LB and TB medium (supplemented with amp, 50 µL) and grown at 37 °C, 200 rpm until an OD_{600} of (0.6-0.8) was reached at which point IPTG (1 mM) was added and cells grown at 37 °C, 200 rpm for ca. 16 h. Cells were harvested (3,500 g, 4 °C, 20 min) and the pellets stored at -20 °C until purification.

2.4.4.3 Large scale expression

E. *coli* BL21 (DE3) was transformed with a plasmid harbouring one of the β-glycosylase genes and a single colony was inoculated into 6 mL of LB medium (supplemented with amp, 6 µL) and incubated at 37 °C, 200 rpm O/N. Then, 3 mL of the starter culture were inoculated into 300 mL of LB medium (supplemented with amp, 300 µL) and grown at 37 °C, 200 rpm until an OD_{600} of (0.6-0.8) was reached at which point IPTG (1 mM) was added and cells grown at 37 °C, 200 rpm for ca. 16 h. Cells were harvested (3,500 g, 4 °C, 20 min) and the pellet was stored at -20 °C until purification.

2.4.5 β-Glycoside hydrolase purification

The pellet from the 300 mL culture was resuspended in loading buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 10 mM imidazole; 3 mL of loading buffer/1 g of pellet) and lysed by sonication for 6 min on ice (5 s on, 5 s off, 60 cycles, Amplitude: 40%). The soluble fraction was decanted following centrifugation at 22,800 × *g*, 4 °C for 1 h, and filtered with a 0.45 µm filter. Filtered supernatant was loaded onto a 1 mL HisTrap FF crude® column, using an AKTA™ Start. The column was washed with eight column volumes of loading buffer, followed by fifteen column volumes of loading buffer with 10 % elution buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 300 mM imidazole). The protein was then eluted with eight column volumes of 100% elution

buffer. Pure fractions were pooled and dialysed for 20 h at room temperature with dialysis buffer (50 mM HEPES pH 7.5, 150 mM NaCl), with one buffer exchange after 2 h.

2.4.6 β-Glycoside hydrolase quantification

The concentration of the purified enzymes was determined by absorbance at 280 nm. The extinction coefficient ε was estimated using the EXPasy ProtParam tool[2] as shown in Table 2.3.

Table 2.3. List of the extinction coefficient at 280 nm of the *Thermus nonproteolyticus* glycoside hydrolase and of the *Halothermothrix orenii* glycoside hydrolase WTs and mutant proteins. The values estimated by using the EXPasy ProtParam tool.

Enzyme	**Extinction coefficient /** M^{-1} cm^{-1}
WT *Tno*GH1	107260
*Tno*GH1-L171K	107260
*Tno*GH1-V287R	107260
*Tno*GH1-H178Y	110240
*Tno*GH1-L171K/ V287R	107260
*Tno*GH1-L171K/ H178Y	110240
*Tno*GH1-V287R/H178Y	110240
WT *Hor*GH1	106230
*Hor*GH1-E173K	106230
*Hor*GH1-M299R	106230
*Hor*GH1-H180Y	107720
*Hor*GH1-E173K/ M299R	106230
*Hor*GH1-E173K/ H180Y	107720
*Hor*GH1-M299R/H180Y	107720

2.4.7 β-Glycoside hydrolase characterization

2.4.7.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were analysed by 12% SDS-PAGE and staining the resulting gel with InstantBlue.

An SDS-PAGE gel (12% acrylamide) was prepared from the following: running gel [dH_2O (2.475 mL), Tris buffer 1.5 M pH 8.8 (1.95 mL), acrylamide 33% (3 mL), SDS 10% solution pH 7.4 (75 µL), ammonium persulphate 10% solution (75 µL), and Tetramethylethylenediamine (7.5 µL)] was prepared and allowed to set in an SDS-PAGE mould. The stacking gel [dH_2O (1.5 mL), Tris buffer 1 M pH 6.8 (250 µL), acrylamide 33% (330 µL), SDS 10% solution pH 7.4 (20 µL), ammonium persulphate 10% solution (20 µL) and TEMED (3 µL)] was prepared and allowed to set. Loading dye (7 µL) was added to each protein sample and then heated at approximately 100 °C for 5 min; finally it was pipetted into individual wells in the gel. Running buffer [glycine (14.5 g), Tris (3.1 g), SDS 10% solution (10 mL) in 1 L dH_2O] was used as gel buffer to run the assay. The gel was ran at 300 V, 30 mA, for 1 h and 10 min. The gel was then removed from the mould and either stained with Coomassie blue staining solution (2% Coomassie brilliant blue R-250 in aqueous solution 50% methanol and 10% acetic acid) for 15-30 minutes following distaining with the distaining solution (aqueous solution 7.5% methanol and 10% acetic acid) overnight or with Instant Blue (Expedeon®) solution overnight. The gel was then placed between two plastic sheets for clarity of visual analysis.

2.4.7.2 DNA electrophoresis

DNA assay gel electrophoresis was performed using an agarose concentration of 0.8% (w/v). The solution was made by dissolving agarose powder (0.32 g) in 40 mL of TAE buffer by heating it in a standard microwave oven until the solution was completely clear. When the temperature was low enough without leaving it solidify, 4 µL of SYBR safe DNA staining were added. The solution was then loaded into the mould and left to solidify. Samples were

prepared appropriately mixing them with the Gel Loading Dye Purple (6x) from New England Biolabs. The electrophoresis was conducted routinely at 75V and 150 mA for 45

minutes.

2.4.7.3 Calibration Curves of 4-Nitrophenol & 4-Nitrothiophenol

As a first step before starting enzyme characterization, the calibration curves of two glycosidic compounds; 4-Nitrophenol (pNP) and 4-Nitrothiophenol (pNT) were prepared to determine the absorbance of PNP and PNT products at 420 nm.

In order to investigate that, a range of different concentrations of the products (PNP and PNT) were tested (0 - 1 mM) in individual wells in duplicate. The samples were assayed at 25 °C in buffer pH 7.5 (100 mM HEPES, 500 mM NaCl).

2.4.7.4 Activity assay

The activity of the purified enzyme was tested against two substrates pNP-Glc and pNT-Glc.

10 µL of appropriately diluted samples of the enzyme were added to a 96-well plate included 290 µL of the reaction solution (500 mM sodium chloride in 100 mM in HEPES buffer, pH 7.5) containing either pNP-Glc or pNT-Glc at a final concentration of 10 mM. Assays were performed at 50 °C for *Tno*GH1 and at 25 °C for *Hor*GH1.

The formation of pNP and pNT was followed at 420 nm over 15 min. The specific activity (U/mg) was expressed as µmol of product formed per minute per milligram of protein.

2.4.7.5 Kinetics assays

Enzyme activity was measured spectrophotometrically in triplicate by monitoring the change in absorbance at 420 nm of the p-nitrophenol (pNP) or p-nitrothiophenol (pNT).

Assays were conducted in 200 µL at 50 °C for *Tno*GH1 and at 25 °C for *Hor*GH1. A typical reaction mixture contained 100 mM HEPES buffer pH 7.5, 500 mM sodium chloride.

pNT-Glc was dissolved in 30% DMSO, the concentration of DMSO was controlled at 9 % across pNT-Glc assays.

All assay components were filtered with a 0.45 µm filter prior to use. Assays were initiated with the addition of enzyme. The kinetics assays presented in primary non-linear regression plots.

2.4.7.6 pH effect on activity profile

The activities of the enzyme were measured at pH values ranging from 3 to 12 and they were presented as relative activities to standard buffer (100 mM HEPES, 500 Mm NaCl, pH 7.5).

2.4.7.7 pH stability profile

The pH stability profile was evaluated by incubating the enzyme for 24 h at pH values ranging from 3 to 12. The activities then were presented as relative activities to standard buffer (100 mM HEPES, 500 Mm NaCl, pH 7.5).

2.4.7.8 Temperature stability profile

Temperature stability profiles were investigated by incubating the enzyme at different temperatures and testing the activity after 0.5 h, 2 h, 4 h, 24 h, 48 h, 72 h, 120 h, and 240 h. the activities were presented as relative activities to the starting activity (Time 0) at 25 °C.

2.4.8 Data analysis

2.4.8.1 UCSF Chimera software

In silico modelling of crystal structures of β-Glycoside hydrolases and the sequences of glycoside hydrolases were performed using the UCSF Chimera software, production version 1.12.[3]

2.4.8.2 The Universal Protein Resource (UniProt)

Myrosinases amino acid sequences were obtained from UniProt. https://www.uniprot.org.[4]

2.4.8.3 The European Molecular Biology Open *Software* Suite (EMBOSS)

Similarities and identities, including homology modelling, between the glycoside hydrolases sequences were calculated using EMBOSS software. https://www.ebi.ac.uk/Tools/psa/emboss_needle/.[6]

2.4.8.4 Igor Pro software

Kinetic parameters were evaluated by nonlinear regression analysis in Igor Pro (Version 7.0.8.1; Wavemetrics Inc., Lake Oswego Oregon).[7]

2.4.8.5 GraphPad Prism

Bar graphs were produced in GraphPad Prism.[8] The primary plots were analysed using equation 1.

$$\frac{v_i}{[E]_T} = \frac{k_{cat}[S]}{K_M + [S]} \qquad \text{(Equation 1)}$$

Where: v_i is the initial velocity of the reaction.

k_{cat} is the turnover number.

K_M is the Michaelis- Menten constant.

[S] is the substrate concentration.

$[E]_t$ is the total concentration of the enzyme.

2.4.8.6 Pfam protein family database

For consensus analysis of amino acids, β-glycosidase sequences were taken from the Pfam protein family database (https://pfam.xfam.org - sequence search), ~4000 sequences of representative proteome were used for amino acid analysis.[9]

2.4.8.7 The Molecular Evolutionary Genetics Analysis software (MEGA X)

The representative proteome at 15 % comembership threshold, as defined by Pfam (see 3.4.8.7) which is an even sampling of the sequences of the glycosyl hydrolase family PF00232[10], was aligned with the WT sequences of the β-glycosidase used in this study with the ClustalW tool[11] (https://www.ebi.ac.uk/Tools/msa/clustalo/) with the MEGA X software (version 10.1.0).[12]

2.4.8.8 WebLogo

Sequence logo analysis was performed with WebLogo. http://weblogo.threeplusone.com/create.cgi.[13]

2.5 Bibliography:

1 C. M. Heckmann, L. J. Gourlay, B. Dominguez and F. Paradisi, *Front. Bioeng. Biotechnol.*, 2020, **8**, 1–13.

2 E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel and A. Bairoch, *Proteomics Protoc. Handb.*, 2005, 571–608.

3 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.

4 R. Apweiler, *Nucleic Acids Res.*, 2008, **36**, 190–195.

5 X. Robert and P. Gouet, *Nucleic Acids Res.*, 2014, **42**, 320–324.

6 P. Rice, L. Longden and A. Bleasby, *Trends Genet.*, 2000, **16**, 276–277.

7 D. Babonneau, *J. Appl. Crystallogr.*, 2010, **43**, 929–936.

8 M. L. Swift, *J. Chem. Inf. Comput. Sci.*, 1997, **37**, 411–412.

9 S. El-Gebali, J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart, E. L. L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S. C. E. Tosatto and R. D. Finn, *Nucleic Acids Res.*, 2019, **47**, D427–D432.

10 C. Chen, D. A. Natale, R. D. Finn, H. Huang, J. Zhang, C. H. Wu and R. Mazumder, *PLoS One*., 2011, **6**, e18910

11 F. Madeira, Y. M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. N. Tivey, S. C. Potter, R. D. Finn and R. Lopez, *Nucleic Acids Res.*, 2019, **47**, W636–W641.

12 S. Kumar, G. Stecher, M. Li, C. Knyaz and K. Tamura, *Mol. Biol. Evol.*, 2018, **35**, 1547–1549.

13 G. Crooks, G. Hon, J. Chandonia and S. Brenner, *Genome Res*, 2004, **14**, 1188–1190.

Chapter 3: Computational study of Glycoside hydrolases family 1 (GH1)

3.1 Overview of Myrosinase types

As already mentioned in the introduction, myrosinases (MYRs; EC 3.2.3.147) are unique members of the GH1 family able to hydrolyse thioglycosides.

Unlike β-glycosidases which are ubiquitous, MYRs have been identified only in a handful of species such as *Sinapis alba*[1], *Arabidopsis thaliana* and *Brassica napus*[2], *Brevicoryne brassicae*[3], and *Verticillium dahliae*.[4] *Sinapis alba*, *Arabidopsis thaliana* and *Brassica napus* species belong to the Plantae kingdom and Brassicaceae family. *Brevicoryne brassicae* species is from Animalia kingdom and Aphididae family. *Verticillium dahliae* is a fungus which is placed Incertae sedis within the order Hypocreales (i.e., its exact taxonomic position is undefined). These species are shown in Figure 3.1.

Figure 3.1. Species producing myrosinases*: A. thaliana* (top left), *B. napus* (top right), *B. brassicae* (bottom left), *V. dahliae* (centre), *B. napus* (bottom right).

As discussed in the introduction, studies of *Sa*MYR and *Bb*MYR have shown some interesting similarities and differences between them. Even though the plant and aphid MYRs have the

expected $(\beta/\alpha)_8$ barrel folded structure, like the structure of β-glucosidases, they differ in the active site residues, the cofactors, and the enzymatic activity. Moreover, the catalytic activity of *Bb*MYR is not affected by the presence or absence of ascorbate, as is *Sa*MYR..

The fungal MYR, on the other hand, shows higher K_M value than *Sa*MYR towards sinigrin, but lower hydrolytic activity as well as lower stability at high temperatures and more acidic conditions.[5]

Therefore, in this chapter a computational study of five MYRs and β-glycosidases will carry out first to find out the similarities and differences between them. Following by a general study of the most similar MYR and the extremo-adapted β-glycosidases. After that, for better understanding of the catalytic performance of the most similar MYR, a bioinformatic analysis of its active site including the essential residues for activity and for the hydrolysis of GSL will perform. Then, comparative studies between the active site motifs of the most similar MYR and the extremo-adapted β-glycosidases will carry out.

3.2 A comparative study between types of MYRs and the extremo-adapted β-glycosidases

To investigate which MYR presented closer affinity to the extremo-adapted β-glycosidases, a comparative study between the structures (amino acid sequences) of the previously mentioned MYR and GH1s was carried out.

Figure 3.2 shows the amino acid sequence alignment of *Brevicoryne* brassicae, *Verticillium dahliae*, *Arabidopsis thaliana*, *Sinapis alba* and *Brassica napus* myrosinases (here referred to as MYR_Bbrassica, MYR_Verticill, MYR_Arabidops, MYR_Salba, MYR_Nbrassica, respectively). The active site of *Sinapis alba*, *Arabidopsis thaliana*, and *Brassica napus* myrosinases differ from that of classical β-glycosidases, as they lack the catalytic glutamic acid residue in the

$T^{184}INQL^{188}$ motif in MYR_Salba, $T^{202}INQL^{206}$ in MYR_Arabidops, and $T^{204}INQL^{208}$ in MYR_Nbrassica (equivalent to $T^{161}LNEP^{165}$ in *Tno*GH1)[6], while it maintains the second one (E409 in motif $T^{408}ENG^{411}$ in MYR_Salba,E420 in motif $T^{419}ENG^{422}$ in MYR_Arabidops, E429 in motif $T^{428}ENG^{431}$ in MYR_Nbrassica.[1]

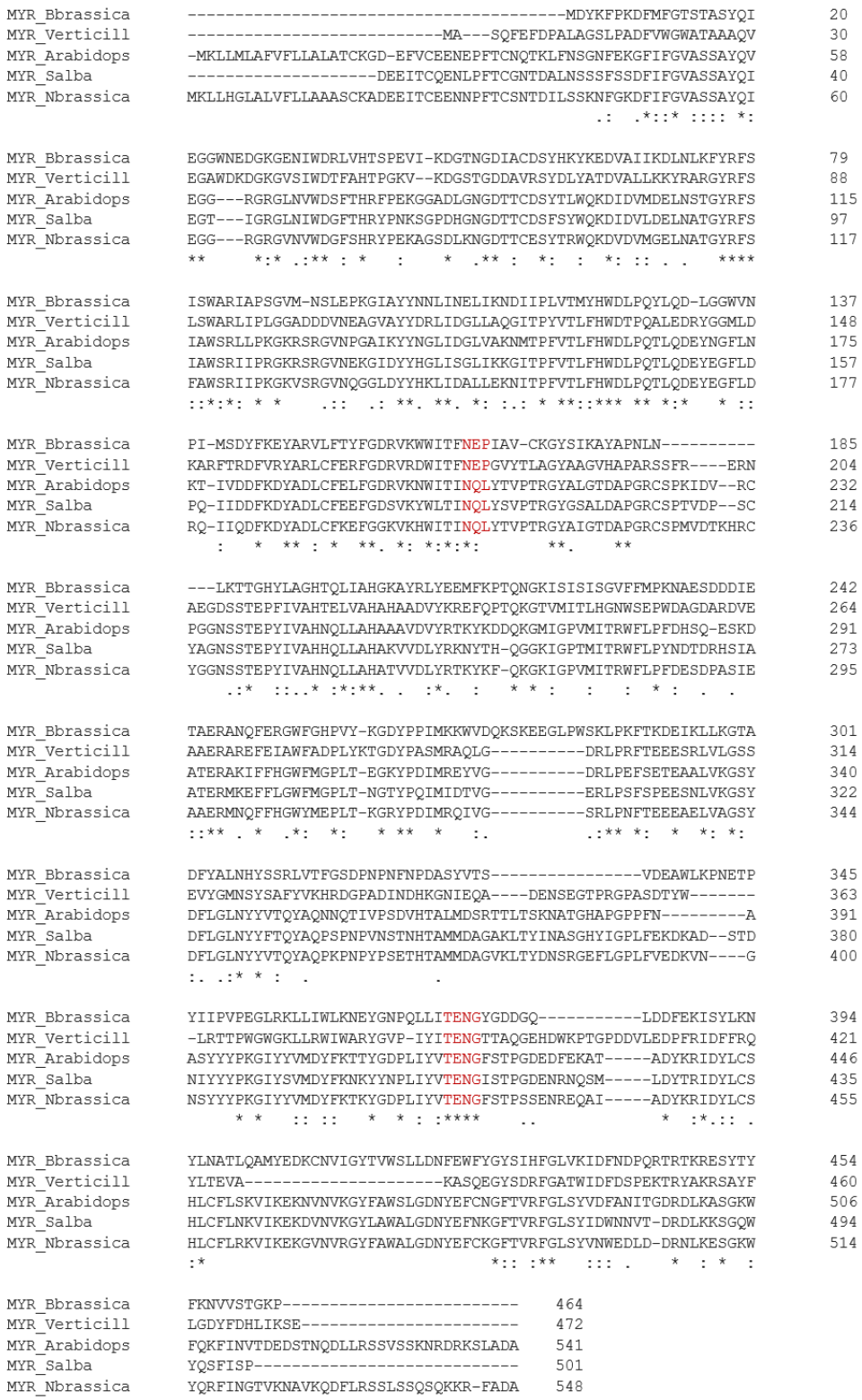


Figure 3.2. Amino acid sequence alignment of *Brevicoryne brassicae* (MYR_Bbrassica), *Verticillium dahliae* (MYR_Verticill), *Arabidopsis thaliana (*MYR_Arabidops*), Sinapis alba* (MYR_Salba) *and Brassica napus (*MYR_Nbrassica). The active site residues are coloured in red. The alignment performed in EMBOSS software (https://www.ebi.ac.uk/Tools/psa/emboss_needle/). Myrosinases sequences were obtained from UniProt. MYR_Bbrassica from https://www.uniprot.org/uniprot/Q95X01, MYR_Verticill from https://www.uniprot.org/uniprot/G2XC15, MYR_Arabidops from https://www.uniprot.org/uniprot/P37702, MYR_Salba from https://www.uniprot.org/uniprot/P29736, and MYR_Nbrassica from https://www.uniprot.org/uniprot/Q00326.

MYRs from the cabbage aphid *Brevicoryne brassicae* (*Bb*MYR) and from *Verticillium dahliae*, on the other hand, rely on the typical catalytic acid/base system found in β-glycosidases (E167 and E374 in *Bb*MYR and E179 and E390 in MYR_Verticill; Figure 3.2) and these MYRs are more closely aligned with classic β-glycosidases than the plant MYRs in term of amino acid residues present in the active site.

Among all the MYRs mentioned in this study, the cabbage aphid *Brevicoryne brassicae* and the *Verticillium dahliae* MYRs are those with higher similarity to the extremo-adapted β-glycosidase, in term of the active site residues, and were therefore selected as models to engineer their thioglucoside hydrolase activity into *Tno*GH1 and *Hor*GH1.

3.2.1 Sequence comparison of MYR and *T. nonproteolyticus* GH1 (*Tno*GH1)

*Tno*GH1, as expected, presents the greatest sequence similarity to the *Bb*MYR and MYR_Verticill. *Tno*GH1 and *Bb*MYR show a 52.3% similarity (33.5% identity), and *Tno*GH1 and MYR_Verticill show 44.5 % similarity (31.7 % identity) as shown in Table 3.1.

Table 3.1. The identity, similarity and gap between *Brevicoryne brassicae* (MYR_Bbrassica), *Verticillium dahliae* (MYR_Verticill), *Arabidopsis thaliana* (MYR_Arabidops), *Sinapis alba* (MYR_Salba) and *Brassica napus* (MYR_Nbrassica) and *Thermus nonproteolyticus* glycosyl hydrolase.

Type of MYR	Identity	Similarity	Gaps in the alignment
MYR_Bbrassica	33.50 %	52.30 %	11.70 %
MYR_Verticill	31.70 %	44.50 %	18 %
MYR_Arabidops	28 %	42.40 %	26.80 %
MYR_Salba	27.90 %	43.50 %	19.80 %
MYR_Nbrassica	28.50 %	43.80 %	25.50 %

Although high similarities were found with the *Verticillium dahliae* MYR and those from *Brassica napus* (43.8 %; Table 3.1), *Sinapis alba* (43.5 %), and *Arabidopsis thaliana* (42.4 %; Table 3.1), *Bb*MYR is the most similar one to the *Tno*GH1 (Table 3.1) presenting the highest degree of homology.

3.2.2 Sequence comparison of MYR and *H. orenii* GH1 (*Hor*GH1)

The sequence alignments of each MYR and the amino acid sequence of *Hor*GH1 shows similar results to those obtained with *Tno*GH1 as shown in Table 3.2.

Table 3.2. The identity, similarity and gap between Brevicoryne brassicae (MYR_Bbrassica), Verticillium dahliae (MYR_Verticill), Arabidopsis thaliana (MYR_Arabidops), Sinapis alba (MYR_Salba) and Brassica napus (MYR_Nbrassica) and *Halothermothrix orenii* glycosyl hydrolase.

Type of MYR	Identity	Similarity	Gaps in the alignment
MYR_Bbrassica	36.02 %	56.20 %	10.30 %
MYR_Verticill	33.50 %	48.80 %	19.80 %
MYR_Arabidops	28.90 %	43.90 %	24.70 %
MYR_Salba	30.70 %	48.40 %	17.30 %
MYR_Nbrassica	31.80 %	49.50 %	20.40 %

Both *Tno*GH1 and *Hor*GH1 shows the highest degree of similarity with *Bb*MYR (52.3% and 56.2% respectively) and the lowest gaps values (Table 3.1 and Table 3.2). *Hor*GH1 has a 36% identity with *Bb*MYR and 30.7 % with *Sa*MYR. *Verticillium dahliae, Arabidopsis thaliana*, and *Brassica napus* MYRs show almost ≥ 30% identity and > 40% similarity with *Hor*GH1 (Table 3.2).

3.3 Bioinformatic analysis of the crystal structures of cabbage aphid *Brevicoryne brassicae* (*Bb*MYR)

The amino acid FASTA sequence and the protein database structure of *Bb*MYR are shown in Figure 3.3.

>1WCG:A|PDBID|CHAIN|SEQUENCE

MDYKFPKDFMFGTSTASYQIEGGWNEDGKGENIWDRLVHTSPEVIKDGTNGDIACDSYHKYKED
VAIIKDLNLKFYRFSISWARIAPSGVMNSLEPKGIAYYNNLINELIKNDIIPLVTMYHWDLPQY
LQDLGGWVNPIMSDYFKEYARVLFTYFGDRVKWWITF**NEP**IAVCKGYSIKAYAPNLNLKTTGHY
LAGHTQLIAHGKAYRLYEEMFKPTQNGKISISISGVFFMPKNAESDDDIETAERANQFERGWFG
HPVYKGDYPPIMKKWVDQKSKEEGLPWSKLPKFTKDEIKLLKGTADFYALNHYSSRLVTFGSDP
NPNFNPDASYVTSVDEAWLKPNETPYIIPVPEGLRKLLIWLKNEYGNPQLLI**TEN**GYGDDGQLD
DFEKISYLKNYLNATLQAMYEDKCNVIGYTVWSLLDNFEWFYGYSIHFGLVKIDFNDPQRTRTK
RESYTYFKNVVSTGKP

Figure 3.3. Amino acid FASTA sequence and the protein database structure (pdb:1WCG) of *Bb*MYR. The active site motifs NEP and TEN are highlighted in bold.

*Bb*MYR has unique structural features related to its ability to breakdown the S-glycosidic bonds in GSL, not observed in either β-glycosidases or other MYRs.

Two amino acid residues in the active site of *Bb*MYR have been highlighted to play critical roles in the hydrolysis of GSL, as they are directly involved in its recognition.[3] These residues are K173 and R312. Figure 3.4 shows the active site structure of *Bb*MYR. K173, R312, as well as Y180 (see later) are the key residues which are not conserved in the extremophilic GH1 and their introduction by mutagenesis as part of this study will be discussed in section 3.3.1 and 3.3.2 while E167 and E374 are the conserved catalytic residues.

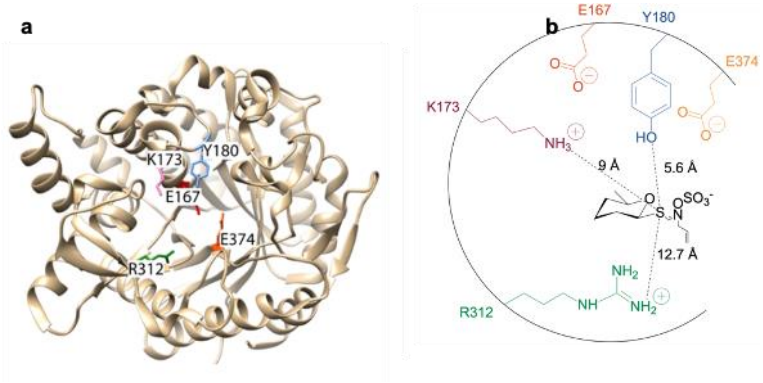


Figure 3.4. Protein database structures of the active sites of cabbage aphid *Brevicoryne brassicae* myrosinase enzyme. **a** *B. brassicae* myrosinase (pdb:1WCG); target residues of mutagenesis indicated, K173, R312, and Y180; residues essential for activity also indicated, E167 and E374. **b** diagram indicating the relationship between the thioglycosidic substrate and the side chains of the residues of WT *B. brassicae* myrosinase selected for targeted mutagenesis (section 4.3.1 and 4.3.2), essential residues also displayed. Distances between substrate and side chains predicted with UCSF Chimera.

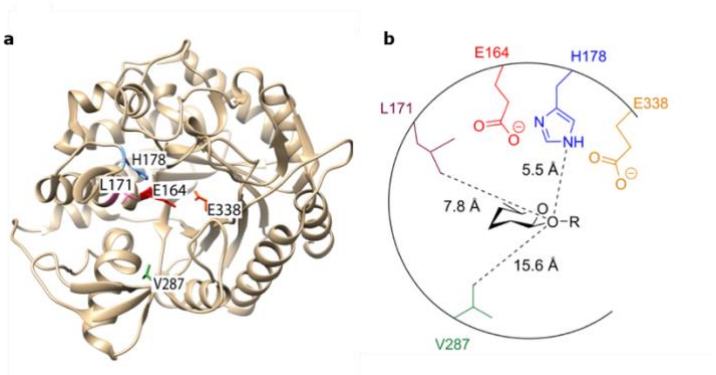


Figure 3.5. Protein database structures of the active sites of *Thermus nonproteolyticus* glycoside hydrolase enzyme (pdb:1NP2). **a** residues to be mutated in this study are indicated, L171, V287, and H178. Residues essential for activity also indicated, E164 and E338. **b** diagram indicating the relationship between the substrate and the side chains of the residues of WT *Thermus nonproteolyticus* to be mutated in this study, essential residues also displayed. Distances between the glycosidic substrate and side chains predicted with UCSF Chimera.

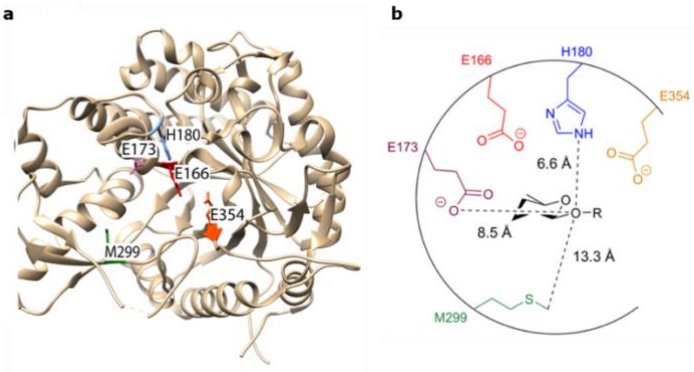


Figure 3.6. Protein database structures of the active sites of *Halothermothrix orenii* glycoside hydrolase (pdb:3TA9). **a** residue to be mutated in this study are indicated, E173, M299, and H180. Residues essential for activity also indicated, E166 and E354. **b** diagram indicating the relationship between the glycosidic substrate and the side chains of the residues of WT *Halothermothrix orenii* to be mutated in this study, essential residues also displayed. Distances between substrate and side chains predicted with UCSF Chimera.

Arginine R312 appears to be also conserved in both *Sa*MYR and *Bb*MYR and it clearly plays an important role in the recognition of the sulphate group of sinigrin. A study by Jones and his co-worker mentioned that R194 and R259 in the active site of *Sa*MYR might be critical for its activity. K173 and V228 are the equivalently positioned amino acid residues in *Bb*MYR.

Another interesting amino acid residue for this study, present in the active site of *Bb*MYR is Y180. This tyrosine may also play a role in substrate recognition due to its proximity to the thioglycosidic linkage in the substrate.[7]

The distance between the sulphur of thioglycosidic substrate and the side chains of K173 residue is 9 Å, to the side chain of Y180 was predicted to be 5.6 Å, and to the side chain of R312 is 12.7 Å (Figure 3.4 above). Therefore, Y180 appears to be in close proximity to the substrate, in fact closer than K173 and R312 and might interact with the aglycon (-R). The proximity of Y180 to the sulphur of the thioglycosidic linkage suggests also a possible catalytic role of this residue.

Previous studies of the crystal structure of *Bb*MYR mention all of these residues (K173, R312, and Y180) but do not offer any experimental proof and this was the starting point of this research.

3.3.1 Cabbage aphid *Brevicoryne brassicae (Bb*MYR) and *T. nonproteolyticus* GH1 (*Tno*GH1)

The sequence alignment between *Bb*MYR and *Tno*GH1 (Figure 3.7) reveals that *Tno*GH1 presents a non-polar amino acid leucine (L171) at the equivalent position of the polar amino acid lysine (K173) in *Bb*MYR; lysine is basic, positively charged, and is a bulkier amino acid than leucine. Therefore, even though K173 appears to point away from the active site of *Bb*MYR, if the charge and/or the size are important in this position of the active site of *Bb*MYR,

a positive result in term of the specificity of *Tno*GH1 towards β-thioglycoside, may be achieved by introducing this mutation.

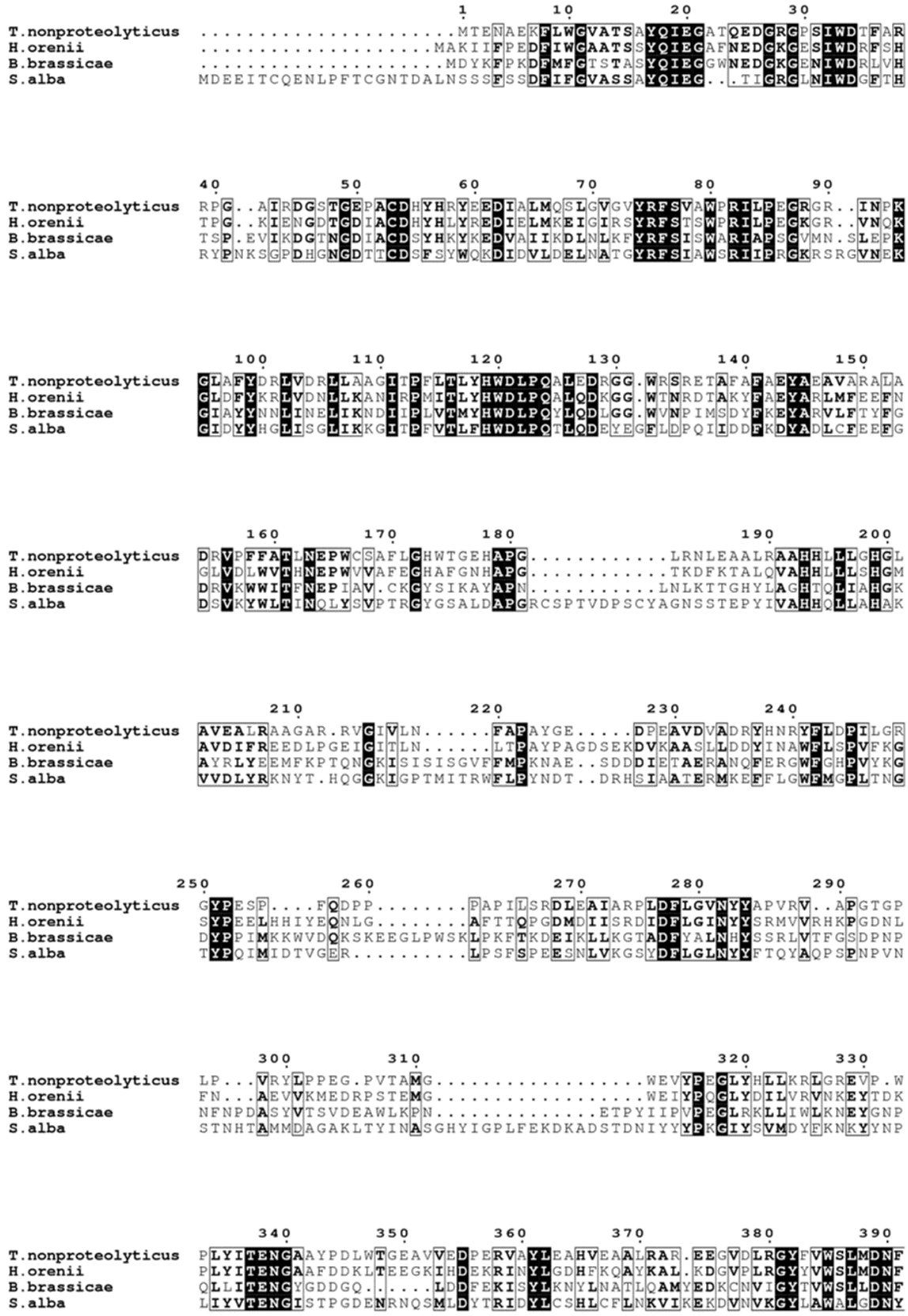


Figure 3.7. Full sequence of *Thermus nonproteolyticus* glycoside hydrolase (*Tno*GH1), *Halothermothrix orenii* (*Hor*GH1), cabbage aphid *Brevicoryne brassicae* (*Bb*MYR), and *Sinapis alba* myrosinase (*Sa*MYR). The alignment generated from ENDscript server software and the amino acid numbers refer to *Thermus nonproteolyticus* glycoside hydrolase (*Tno*GH1).

From the sequence alignment between *Bb*MYR and *Tno*GH1 it is also clear that in position R312 (*Bb*MYR) *Tno*GH1 presents the non-polar amino acid valine (V287); arginine is also basic, positively charged, and is again a bulkier amino acid than valine. Therefore, as with *Tno*GH1-L171K mutation case, an increase in the specificity of *Tno*GH1 towards β-thioglycoside might be achieved with *Tno*GH1-V287R mutation if the thiogluco-hydrolase activity of *Bb*MYR depends on the charge and/or the size of the amino acid residues at this position.

This sequence alignment also shows that *Tno*GH1 has a histidine (H178), which is an aromatic amino acid with an imidazole side chain at the equivalent position of the aromatic benzene ring side chain tyrosine (Y180) in *Bb*MYR. Tyrosine prefers to substitute with other amino acid of the same type such as phenylalanine, which differs from tyrosine only in the hydroxyl group in the ortho position on the benzene ring. On the other hand, histidine is commonly replaced by cysteine and it is quite rare to see histidine exchanges for any amino acid at all.[8] Therefore, even though Y180 may be in the proximity of the sulphur of the thioglycosidic linkage, introducing this mutation in *Tno*GH1 may negatively affect the specificity of *Tno*GH1 towards β-thioglycoside and/or the turnover of the enzymatic reaction, and this should be taken into consideration in the evaluation of the experimental results. Figure 3.8 shows the aforementioned amino acid residues in *Thermus nonproteolyticus* glycoside hydrolase (*Tno*GH1) and the amino acid residues at the equivalent positions in *Bb*MYR.

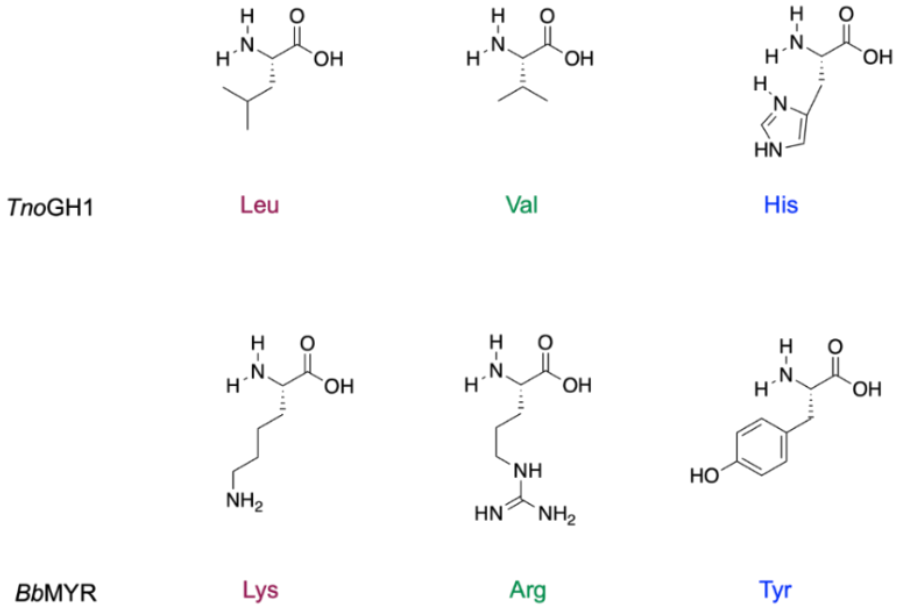


Figure 3.8. The amino acid residues in positions 171, 287, and 178 in *Thermus nonproteolyticus* glycoside hydrolase (*Tno*GH1) and the amino acid residues at the equivalent positions in cabbage aphid *Brevicoryne brassicae* (*Bb*MYR).

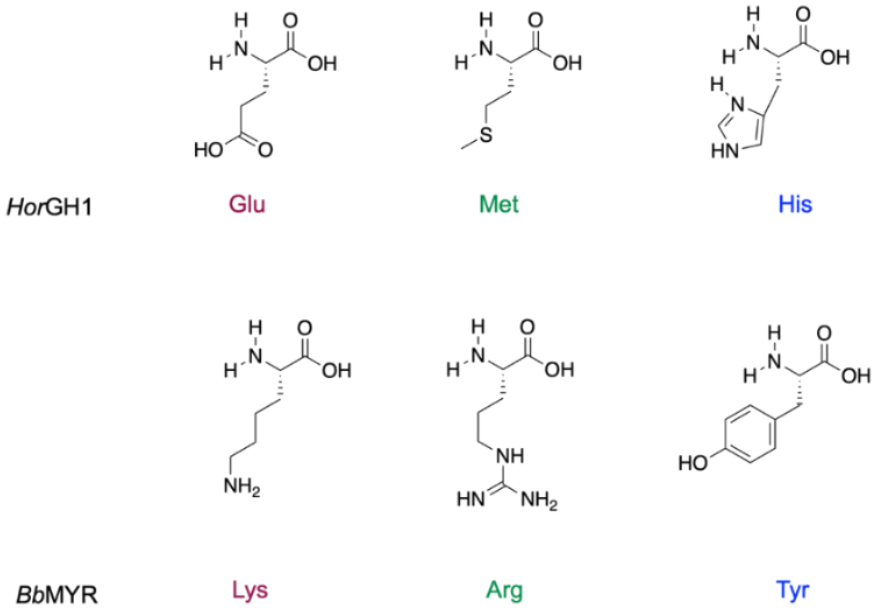


Figure 3.9. The amino acid residues in positions 173, 299, and 180 in *Halothermothrix orenii* glycoside hydrolase (*Hor*GH1) and the amino acid residues at the equivalent positions in cabbage aphid *Brevicoryne brassicae* (*Bb*MYR).

The protein database structures of the active sites of *Tno*GH1 and the diagram indicating the relationship between the substrate and the side chains of the residues of WT *Tno*GH1 mutated in this study is shown in Figure 3.5.

Therefore, the bioinformatic analysis of the crystal structures and the sequence alignment of both *Bb*MYR and *Tno*GH1 indicate three single mutations which may be able to shift the activity in the *Tno*GH1 enzyme towards thioglycosides: *Tno*GH1-L171K, *Tno*GH1-V287R, and *Tno*GH1-H178Y. Of course, a combinatorial approach of the mutations may also lead to better variants: *Tno*GH1-L171K/V287R, *Tno*GH1-L171K/H178Y, and *Tno*GH1-V287R/H178Y will also be experimentally tested to assess whether they play a critical role in the hydrolysis of thioglucosides.

3.3.2 Cabbage aphid *Brevicoryne brassicae (Bb*MYR) and *H. orenii* (*Hor*GH1)

The sequence alignment between *Bb*MYR and *Hor*GH1 (Figure 3.7 above) shows that a polar amino acid glutamic acid (E173) in *Hor*GH1 amino acid sequence is present at the equivalent position of another polar amino acid lysine (K173) in *Bb*MYR; glutamic acid is an acidic negatively charged amino acid and lysine is a basic positively charged amino acid and both amino acid residues have similar molecular weight (ca. 147 g/mol; Figure 3.9). Therefore, if the positive charge is not important in this position (K173) of the active site of *Bb*MYR a decrease in the turnover rate and a loss in the specificity of a *Hor*GH1-E173K mutant compared with the WT *Hor*GH1, might be observed.

The second interesting position for this study in the sequence alignment between *Bb*MYR and *Hor*GH1 is R312. In the amino acid sequence of *Hor*GH1, a methionine (M299) found at the same position of arginine (R312) in the amino acid sequence of *Bb*MYR (Figure 3.7 above). Methionine is a non-polar sulphur containing amino acid (Figure 3.9 above). Moreover, the

M299R mutation is a mutation from hydrophobic amino acid (Met) to a longer positively charged amino acid (Arg). Therefore, if the positive charge is important in this position (R312) of the active site of *Bb*MYR an increase in the turnover rate and an improvement in the specificity of a *Hor*GH1-M299R mutant compared with the WT *Hor*GH1, will be noticed.

Y180 in the amino acid sequence of *Bb*MYR is the third interesting amino acid residue for this study. The sequence alignment between *Bb*MYR and *Hor*GH1 (Figure 3.7 above) reveals that *Hor*GH1 presents again histidine (H180) at the equivalent position of tyrosine (Y180) in *Bb*MYR (Figure 3.7 and Figure 3.9) aligning perfectly with *Tno*GH1. Therefore, similar predicted results might be achieved with *Hor*GH1-H180Y mutant as with *Tno*GH1-H178Y in term of the specificity of *Hor*GH1 towards β-thioglycoside and/or the turnover of the enzymatic reaction. Also in the case of *Hor*GH1, three amino acid residues Lys, Arg, and Tyr will introduced into *Hor*GH1 at 173, 299, and 180 positions, as single and double mutants, procuring six mutant *Hor*GH1 enzymes; *Hor*GH1-E173K, *Hor*GH1-M299R, *Tno*GH1-H180Y, *Hor*GH1-E173K/M299R, *Hor*GH1-E173K/H180Y, and *Hor*GH1-M299R/H180Y in order to experimentally assay if they are important in the hydrolysis of thioglucosides. Figure 3.6 shows the crystal structure of the active site of *Hor*GH1 (a) and the diagram indicating the interaction between the substrate and the amino acids side chains of the targeted WT *Hor*GH1.

3.4 Bibliography:

1 W. P. Burmeister, S. Cottaz, H. Driguez, R. Iori, S. Palmieri and B. Henrissat, *Structure*, 1997, **5**, 663–675.

2 H. Nong, J. M. Zhang, D. Q. Li, M. Wang, X. P. Sun, Y. J. Zhu, J. Meijer and Q. H. Wang, *J. Integr. Plant Biol.*, 2010, **52**, 879–890.

3 A. M. E. Jones, P. Winge, A. M. Bones, R. Cole and J. T. Rossiter, *Insect Biochem. Mol. Biol.*, 2002, **32**, 275–284.

4 K. Witzel, F. S. Hanschen, R. Klopsch, S. Ruppel, M. Schreiner and R. Grosch, *Front. Plant Sci.*, 2015, **6**, 1–11.

5 M. Ohtsuru, I. Tsuruo and T. Hata, *Agric. Biol. Chem.*, 1969, **33**, 1309–1314.

6 A. Bourderioux, M. Lefoix, D. Gueyrard, A. Tatibouët, S. Cottaz, S. Arzt, W. P. Burmeister and P. Rollin, *Org. Biomol. Chem.*, 2005, **3**, 1872–1879.

7 H. Husebye, S. Arzt, W. P. Burmeister, F. V Ha and A. Brandt, *Insect Biochem. Mol. Biol.*, 2005, **35**, 1311–1320.

8 M. J. Betts and R. B. Russell, *Consequences of Substitutions*, 2006, vol. 9.

Chapter 4: *Thermus nonproteolyticus* glycoside hydrolase

4.1 Introduction

As mentioned in the introduction, *Tno*GH1 is a thermophilic GH1 from the extremophilic bacterium *T. nonproteolyticus,*[1–3] so it is attractive for industrial applications, particularly those requiring high temperature conditions. Thermophilic enzymes are usually thermostable enzymes with long shelf-life, high tolerance to organic solvents, and low risk of microbial contamination.[4,5] Therefore, engineering S-glycosidase activity into *Tno*GH1 will broaden its substrate scope towards β-thioglycosidase activity and could significantly increase its potential applications in an industrial setting.[6]

In this chapter, the introduction of mutations into the extremophilic *Tno*GH1 enzyme will be discussed as the first example of engineering of GH1 in which myrosinase activity has been introduced. Three mutations L171K, V287R, and H178Y will be introduced into the protein, on their own or in combination, thereby creating single and double mutant forms of the active site of *Tno*GH1, respectively. Then, the kinetic properties of all variants and WT enzymes with test substrates β-D-thioglucopyranoside (pNT-Glc) and 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc) will be carried out.

4.2 Expression and purification

Mutated variants of *Tno*GH1 were generated by SDM as described in the Material and Methods section 2.4.1 and cloned into *E. coli* vectors pET45b to allow overexpression and subsequent purification. In order to optimize expression and purification conditions for *Tno*GH1, various trials and different techniques were examined. Initially, a small-scale experiment (10 mL LB; see Material and Methods section 2.4.4.1) was set up, and even though

the enzyme expressed very well as shown in Figure 4.1, the pellet produced was minute (ca. 0.07 g).

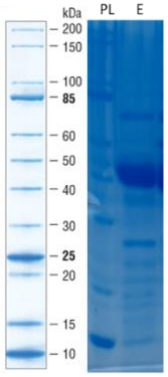


Figure 4.1. 12% acrylamide SDS-PAGE of the WT *Thermus nonproteolyticus* glycoside hydrolase (WT *Tno*GH1). PL is protein ladder (5 µL) and E (enzyme) is pure fraction of the enzyme (10 µL).

Therefore, the volume of the expression medium was increased to 50 mL and 300 mL with trying two types of medium; LB and TB. TB medium allowed faster bacterial cells growth (6 g pellet and 2 g pellet for TB and LB, respectively), but LB was better for the overall expression of *Tno*GH1 and the final yield (6.5 mg/L and 25 mg/L for TB and LB, respectively) of the purified enzyme.

Although *Tno*GH1 expressed very well in both cases, 300 mL medium gave larger pellets and obviously higher concentration of *Tno*GH1 to enable activity assays with both the natural substrate (pNP-Glc) and the substrate mimic (pNT-Glc). Figure 4.2 below shows the SDS-PAGE of *Tno*GH1 with a well-defined and clear band at 49 kDa, with 25 mg/L protein yield.

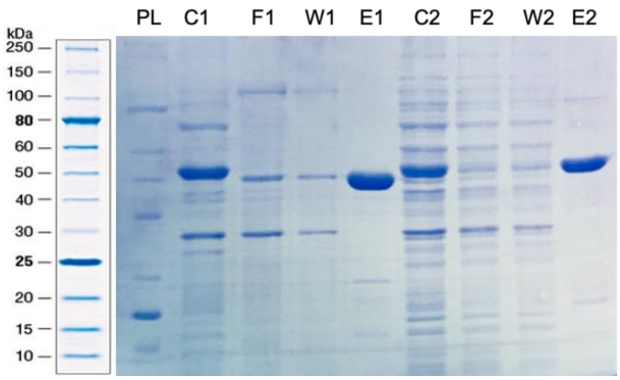


Figure 4.2. 12% acrylamide SDS-PAGE of the WT *Thermus nonproteolyticus* glycoside hydrolase (WT *Tno*GH1). PL is the protein ladder (5 µL), C1 is crude extract of the enzyme expressed in 50 ml LB (10 µL), F1 is flow through of the enzyme expressed in 50 ml LB (10 µL), W1 is washing sample of the enzyme expressed in 50 ml LB (10 µL) eluted with 100% loading buffer, and E1 is the pure enzyme expressed in 50 ml LB (10 µL). C2 is crude extract of the enzyme expressed in 300 ml LB (10 µL), F2 is flow through of the enzyme expressed in 300 ml LB (10 µL), W2 is washing sample of the enzyme expressed in 300 ml LB (10 µL) eluted with 100% loading buffer, and E2 is the pure enzyme expressed in 300 ml LB (10 µL).

4.3 Substrate specificity

The characterization of the WT *Tno*GH1 enzyme with two substrates (Figure 4.3) was carried out; assays were performed with a glucosidic (pNP-Glc) and a thioglycoside (pNT-Glc) substrate to understand both β-glucosidase and the (potential) β-thioglucosidase activities of the *Tno*GH1 enzyme under standard conditions.

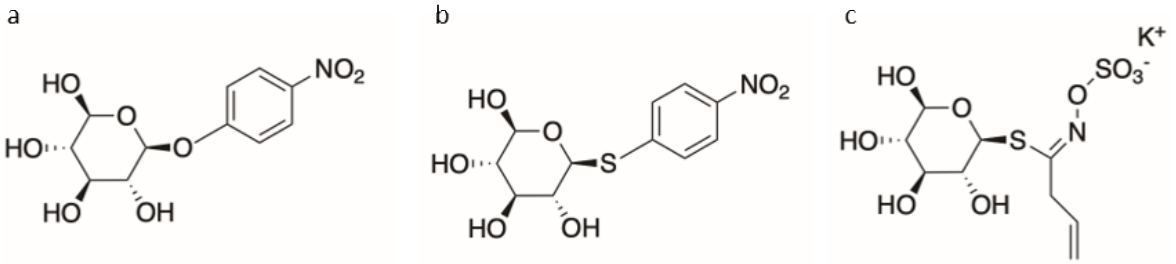


Figure 4.3. Structures of the glycoside and thioglycoside substrates used in this study**.** The native substrate of myrosinase is added to show the similarity in the structure of the molecules. **a** 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc). **b** 4-nitrophenyl-β-D-thioglucopyranoside (pNT-Glc). **c** sinigrin, the native substrate of myrosinase.

Commercial substrates **a** and **b** release upon cleavage 4-nitrophenol (pNP) and 4-nitrothiophenol (pNT) (Figure 4.4) which absorb at 420 nm, therefore the activity of the

enzyme can be monitored directly by UV-vis spectrophotometry and the specific activity (U/mg) of the pure enzyme can be calculated.

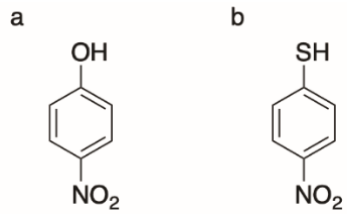


Figure 4.4. Structures of the products. **a** 4-nitrophenol (pNP). **b** 4-nitrothiophenol (pNT).

Calibration curves (Figures 4.5) of the two glycosidic products were prepared to determine experimentally the molar absorptivity (ε). All experiments were performed in triplicate. pNP had a molar absorptivity of 7783.8 $M^{-1}cm^{-1}$ while pNT had 3547.6 $M^{-1}cm^{-1}$.

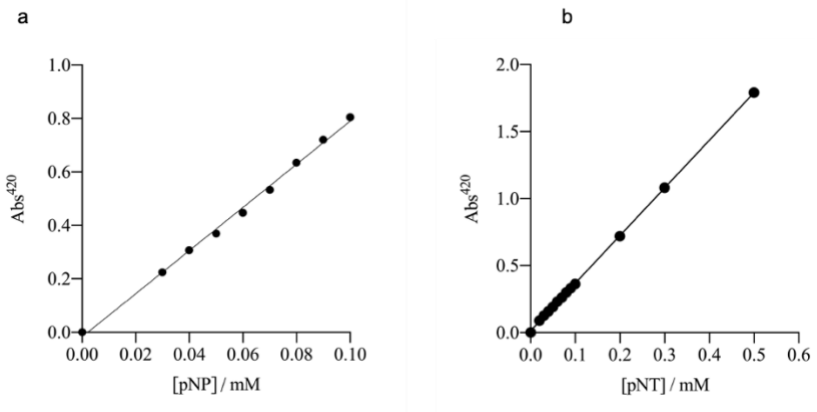


Figure 4.5. Calibration curves of **a** 4-nitrophenol (pNP) and **b** 4-nitrothiophenol (pNT). Absorbance at 420 nm was measured at the indicated concentrations.

4.4 Activity assays with 4-nitrophenyl-β-D-glucopyranoside and 4-nitrophenyl-β-D-thioglucopyranoside substrates

For the activity assay with pNP-Glc substrate, 10 mM of pNP-Glc was freshly prepared in reaction buffer (2.4.7.4), then 0.0003 mg/mL of the purified enzyme was added to the pNP-Glc substrate solution.

The specific activity of WT *Tno*GH1 with pNP-Glc substrate at 25 °C was measured to be 1.7 U/mg, in line with previous results obtained in our lab. However, when the temperature of the activity assay was increased to 50 °C, the specific activity of WT *Tno*GH1 doubled (3.5 U/mg).

The activity assay tests with pNT-Glc substrate (10 mM pNT-Glc, 10% DMSO), was carried out under the same conditions as for the pNP-Glc substrate.

The enzyme was considerably less active with pNT-Glc, requiring significantly higher enzyme concetrations to obtan reliable readings. With a final concetration of 0.50 mg/mL of the enzyme, the activity with pNT-Glc was measured to be 0.014 U/mg at 25 °C. Again, when the assay was carried out at 50 °C, the activity increased to 0.027 U/mg.

The activity of *Tno*GH1 was previously tested in Paradisi lab in the present of organic co-solvents such as MeOH, EtOH, iPrOH, THF, ACN as well as DMSO as mentioned in the introduction (see section 1.2.1). In this study, *Tno*GH1 showed in fact an enhancement in the specific activity (20%) in the presence of 10% DMSO with no loss of activity after incubating at the solvent for 24 h which is a common feature of organic solvent-tolerant enzyme to be stable in organic solvent.[7] DMSO was therefore used to dissolve pNT-Glc which is poorly water soluble.

The ability of DMSO to both activate or inhibit enzymes activity has been proved in literature.[8–12] Wiggers and his co-worker reported that the enzyme activity of *Trypanosoma*

cruzi glyceraldehyde-3-phosphate dehydrogenase increased by 2-fold when the enzyme was assayed in the presence of DMSO (5%, v/v) similar to what observed here with *Tno*GH1.[13] The specific activity of the WT *Tno*GH1 with pNP-Glc substrate was also tested in the presence of 10% DMSO which enhanced 2-fold (3.5 U/mg) at 25°C and 3-fold (5.5 U/mg) at 50°C. Interestingly, these results show that the addition of DMSO improves the activity of the enzyme to a value similar to that seen at at 50°C (without DMSO).

4.5 Kinetics assays with 4-nitrophenyl-β-D-glucopyranoside and 4-nitrophenyl-β-D-thioglucopyranoside substrates

As with the activity assay in previous section,, the kinetics for pNP-Glc were measured both at 25 °C (0.06 µM) and at 50 °C (0.006 µM). The kinetics assay showed that the performance of *Tno*GH1 is significantly better at 50°C. The turnover rate increased from 25.14 s^{-1} to 132.5 s^{-1} and the specificity increased from 17340 M^{-1} s^{-1} to 246000 M^{-1} s^{-1} when comparing values at 25 °C and 50 °C (Figure 4.6).

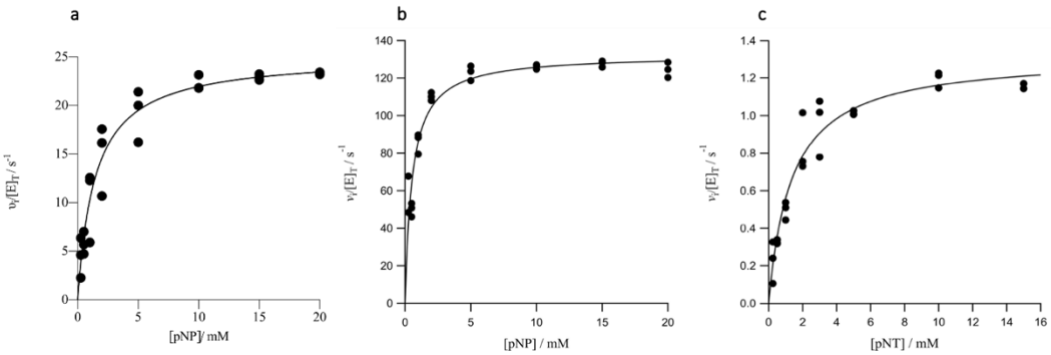


Figure 4.6. Michaelis-Menten primary plot of the activity of WT *Thermus nonproteolyticu* glycoside hydrolase (WT *Tno*GH1)**. a** with varying concentrations of 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc) substrate at 25°C (v_i vs substrate concentration) Michaelis-Menten kinetics behaviour described by equation 1, closed circles represent individual data points. The line is characterised by the parameters k_{cat} = 25.14 s^{-1}, K_M = 1.45 mM, k_{cat}/K_M = 17.34 × 10^3 M^{-1} s^{-1}. **b** with 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc) substrate concentration at 50°C (v_i vs substrate concentration) Michaelis-Menten kinetics behaviour described by equation 1, markers represent individual data points. The line is characterised by the parameters k_{cat} = 132.5 s^{-1}, K_M = 0.54 mM, k_{cat}/K_M = 246 × 10^3 M^{-1} s^{-1}. **c** with 4-nitrophenyl-β-D-thioglucopyranoside (pNT-Glc) substrate concentration at 50°C (v_i vs substrate concentration). Michaelis-Menten kinetics behaviour described by equation 1, closed circles represent individual data points. The line is characterised by the parameters k_{cat} = 1.34 s^{-1}, K_M = 1.4 mM, k_{cat}/K_M= 957 M^{-1} s^{-1}.

The kinetics assays with pNT-Glc substrate could be performed only at 50 °C (Figure 4.6 above), given the low activity previously measured.

The reaction set up with pNT-Glc (0.3 µM) differs from that with pNP-Glc as 10% DMSO needs to be added to solubilise the substrate as mentioned above.

As expected, Figure 4.6 above shows that, the kinetics parameters of WT *Tno*GH1 with pNT-Glc were lower than that one measured with pNP-Glc ($k$cat/K_M= 957 M^{-1} s^{-1} and k_{cat}= 1.34 s^{-1} for pNT-Glc and k_{cat}/K_M= 246000 M^{-1} s^{-1} and k_{cat}= 132.5 s^{-1} for pNP-Glc).

4.6 *T. nonproteolyticus* glycoside hydrolase mutagenesis

As previously mentioned, three residues in the amino acid sequence of *Bb*MYR, K173, R312, and Y180, were identified to be important in the hydrolysis of thioglucosides.[14,15] The sequence alignment between *Bb*MYR and *Tno*GH1 revealed that *Tno*GH1 presents a leucine, a valine, and a histidine at the equivalent positions (L171, V287, and H178-*Tno*GH1 numbering). Thus, three single mutants were generated in *Tno*GH1 (L171K, V287R, and H178Y) and double mutant permutations were then created (L171K/V287R, L171K/H178Y, and V287R/H178Y) to elucidate any synergistic effect among these amino acids.

As with the WT *Tno*GH1 previous studies, the activity assays and the kinetic properties of all variants and WT enzyme with test substrates, pNP-Glc and pNT-Glc were then carried out as described in the below sections.

4.6.1 Single mutants

The mutants were generated by SDM (see Material and Methods section 2.4.1) and the variants were confirmed by sequencing. The three single mutants were expressed and purified following the same protocol optimised for the WT enzyme as shown in Figure 4.7.

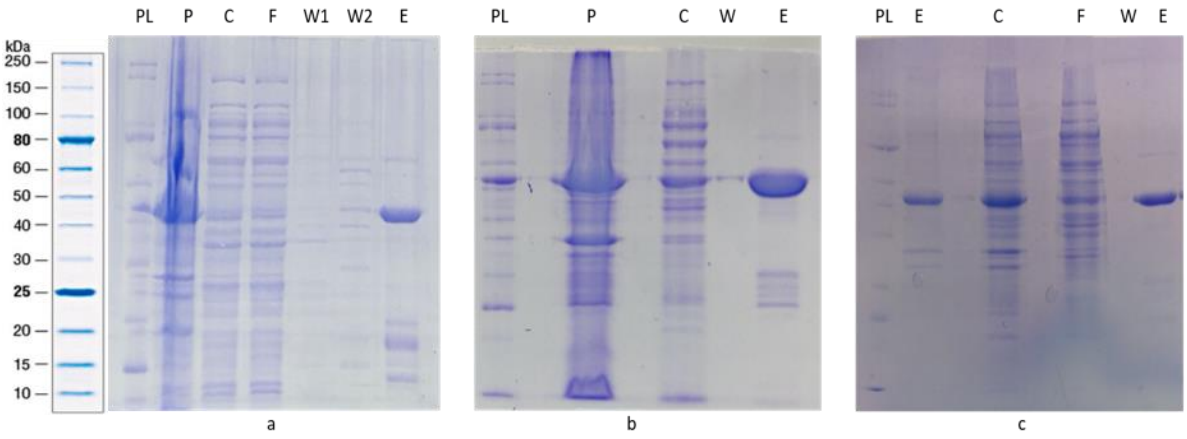


Figure 4.7. Purification of *Thermus nonproteolyticus* glycoside hydrolase double single as analysed by 12% acrylamide SDS-PAGE. **a** *Tno*GH1-L171K. **b** *Tno*GH1-V287R. **c** *Tno*GH1-H178Y. PL is the protein ladder fraction (5 µL), P is the pellet fraction after sonication(10 µL), C is crude extract fraction of the mutants (10 µL), F is flow through fraction of the mutants (10 µL), W1 is first wash fraction of the mutants (10 µL) eluted with 100% loading buffer, W2 is second wash fraction of the mutants (10 µL) eluted with 90% loading buffer and 10% elution buffer, and E is the pure fraction of the mutant (10 µL).

Analysis by 12% polyacrylamide gel (Figure 4.7 above) showed a decrease in the expression level of the mutants (Table 4.1) when compared with the expression level of the WT enzyme.

Table 4.1. Protein yields of WT and single mutants *Thermus nonproteolyticu* glycoside hydrolase (*Tno*GH1).

	Protein yield mg/L
WT	25
*Tno*GH1-L171K	20
*Tno*GH1-V287R	15
*Tno*GH1-H178Y	10

4.6.1.1 Activity assays of the single mutants with pNP-Glc and pNT-Glc substrates

*Tno*GH1 single mutants were tested under the same WT conditions in term of substrate concentration (10 mM) and temperatures (25 °C and 50 °C). The single mutants showed activities toward pNP-Glc and pNT-Glc particularly at 50 °C, but the WT was clearly more active in all conditions. The specific activities of the mutants were slightly increased when 10% DMSO added to the assays as shown in Table 4.2.

Table 4.2. Specific activities of the WT enzyme and single mutants of *Thermus nonproteolyticus* glycoside hydrolase with 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc) and with 4-nitrophenyl-β-D-thioglucopyranoside (pNT-Glc) at 25 °C and at 50 °C. The values in brackets under the pNP-Glc columns are the specific activities of the enzymes with 10% DMSO. Experiments were conducted in triplicate.

	Specific activity/ U/mg (pNP-Glc) 25 °C	Specific activity/ U/mg (pNP-Glc) 50 °C	Specific activity/ U/mg (pNT-Glc) 25 °C	Specific activity/ U/mg (pNT-Glc) 50 °C
WT	1.7 ± 0.3 (3.5 ± 0.5)	3.5 ± 0.2 (5.5 ± 0.5)	0.014 ± 0.002	0.028 ± 0.004
*Tno*GH1-L171K	0.5 ± 0.1 (0.6 ± 0.1)	1.3 ± 0.1 (1.4 ± 0.1)	0.006 ± 0.001	0.012 ± 0.001
*Tno*GH1-V287R	0.7 ± 0.1 (0.9 ± 0.1)	2.3 ± 0.2 (2.6 ± 0.1)	0.004 ± 0.001	0.010 ± 0.001
*Tno*GH1-H178Y	0.7 ± 0.1 (0.9 + 0.1)	1.1 ± 0.1 (1.6 ± 0.1)	0.001 ± 0.001	0.002 ± 0.001

The relative activities of *Tno*GH1 single mutants to the WT *Tno*GH1 are shown as columns in Figure 4.8 **a**. Although the three mutants show similar specific activities values towards pNP-Glc substrate at 25 °C, *Tno*GH1-V287R shows the best mutant performance at 50 °C with three times fold increasing comparing with its specific activity at 25 °C.

Increasing the temperature from 25 °C to 50 °C naturally improved the specific activity values of the single mutants, by almost 2-fold, with pNT-Glc (Table 4.2). Figure 4.8 **b** shows the relative activities of the *Tno*GH1 single mutants to the WT *Tno*GH1 with pNT-Glc substrate.

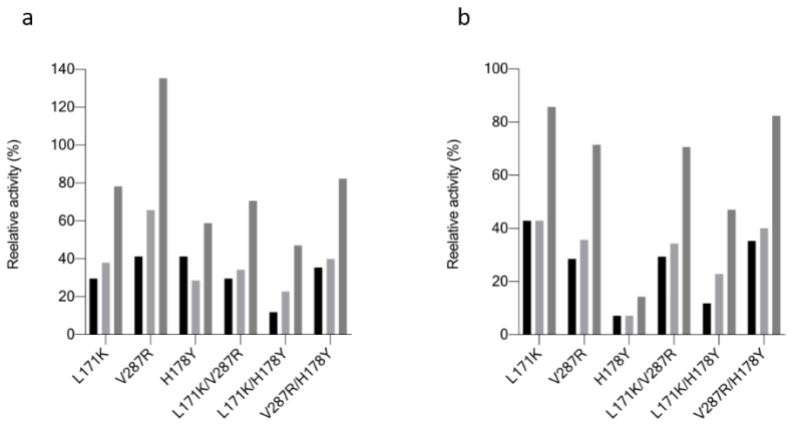


Figure 4.8. Relative activities of *Thermus nonproteolyticu* glycoside hydrolase mutants to the WT *Thermus nonproteolyticu* glycoside hydrolase. **a** with 4-nitrophenyl-β D-glucopyranoside (pNP-Glc) and **b** with the 4-nitrophenyl-β D-thioglucopyranoside (pNT-Glc) substrates. Black columns present the relative activities of *Tno*GH1 mutants at 25 °C to the specific activity of the WT *Tno*GH1 at 25 °C. Light grey columns present the relative activities of *Tno*GH1 mutants at 50 °C to the specific activity of the WT *Tno*GH1 at 50 °C. Dark grey columns present the relative activities of *Tno*GH1 single mutants at 50 °C to the specific activity of the WT *Tno*GH1 at 25 °C.

4.6.1.2 Kinetics assays of the single mutants with pNP-Glc and pNT-Glc substrates

The concentrations of the freshly purified enzymes were measure by UV using the EXPasy ProtParam tool to estimate the extinction coefficients of the *Tno*GH1 single mutants (see Material and Methods section 2.4.6, Table 2.3). This information was then used for the kinetic assays by ensuring that equal molar concentrations for each enzyme variant were used.

From the kinetics parameters, Table 4.3, it is apparent that the specificity of the WT *Tno*GH1 with its natural substrate (pNP-Glc) is in all cases better than all single mutants.

Table 4.3. Table summarizing the kinetic parameters of the WT enzyme and single mutants of *Thermus nonproteolyticus* glycoside hydrolase (*Tno*GH1) with the 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc) and 4-nitrophenyl β-D-thioglucopyranoside (pNT-Glc). Experiments were conducted in triplicate. Standard errors are given, based on fitted theoretical curves.

	k_{cat}/K_M (pNP-Glc)/ $M^{-1}s^{-1}$	k_{cat}(pNP-Glc)/ s^{-1}	K_M (pNP-Glc)/ mM	k_{cat}/K_M (pNT-Glc) / $M^{-1}s^{-1}$	k_{cat} (pNT-Glc)/ s^{-1}	K_M (pNT-Glc)/ mM
WT *Tno*GH1	$246 \times 10^3 \pm 3 \times 10^3$	133 ± 3	0.54 ± 0.06	957 ± 70	1.34 ± 0.04	1.40 ± 0.10
*Tno*GH1-L171K	$91 \times 10^3 \pm 9 \times 10^3$	136 ± 4	1.50 ± 0.20	1200 ± 100	1.10 ± 0.04	0.90 ± 0.10
*Tno*GH1-V287R	$192 \times 10^3 \pm 20 \times 10^3$	117 ± 3	0.61 ± 0.07	2500 ± 400	1.40 ± 0.01	0.56 ± 0.01
*Tno*GH1-H178Y	$230 \times 10^3 \pm 36 \times 10^3$	77 ± 2	0.34 ± 0.06	560 ± 50	0.15 ± 0.01	0.27 ± 0.07

Moreover, having lysine instead of leucine in the active site decreases the specificity of *Tno*GH1-L171K mutant by 2.7-fold, with a retention of turnover rate (133 s^{-1} for WT *Tno*GH1 and 136 s^{-1} for *Tno*GH1-L171K mutant) and reduced of substrate affinity with 3x increased in K_m value as shown in Table 4.3 and Figure 4.9. Loss of specificity was also noticed with *Tno*GH1-V287R mutant (almost 30% loss comparing with the WT *Tno*GH1 under the same experimental conditions). This mutation had only murgured impact on the turnover with about 10% loss and almost a retention of substrate affinity. Even though the *Tno*GH1-H178Y mutant showed a near retention of specificity and an increase in substrate affinity compared to the WT *Tno*GH1, this mutation represented the lowest single mutant turnover rate (by 2-fold) with pNP-Glc substrate comparing with the turnover values of *Tno*GH1-L171K and *Tno*GH1-V287R single mutants.

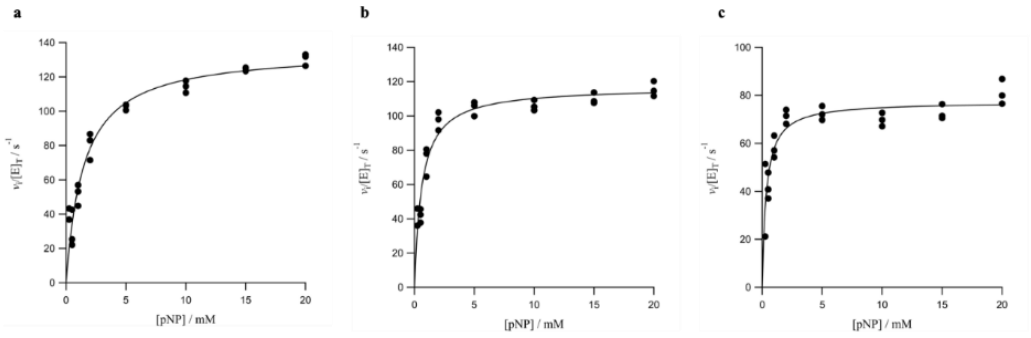


Figure 4.9. Michaelis-Menten curves for the *Thermus nonproteolyticus* glycoside hydrolase single mutants with 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc) substrate at 50 °C; **a** L171K, **b** V287R, **c** H178Y.

With pNT-Glc substrate (Table 4.3 and Figure 4.10) the mutant L171K showed a near retention kinetic parameters in term of specificity (125 %), turnover rate (82 % of WT), and substrate affinity (65 %) to the WT *Tno*GH1 (957 M^{-1} s^{-1}, 1.34 s^{-1}, and 1.40 mM). This mutation resulted in little perturbation in turnover rate with both substrates.

The kinetic assay for *Tno*GH1-V287R with pNT-Glc substrate showed an increase in the specificity, with a K_M value falling to 0.56 mM (3-fold improvement) with no loss in turnover rate, compared with the WT enzyme. The enhancement in k_{cat}/K_M together with the retention of k_{cat} suggest that V287R would be the best mutant for practical application. Interestingly, this mutation has only a small effect on the K_M for pNP-Glc (113% with respect to the WT), indicating that the binding is just as favourable.

A decrease in the performance of *Tno*GH1-H187Y mutant with pNT-Glc, was noticeable with a 3-fold loss in specificity (560 M^{-1} s^{-1}) and a 10-fold loss in turnover rate (0.15 s^{-1}). This mutant showed the lowest single mutant turnover rate, but the highest single mutant substrate affinity with both substrates.

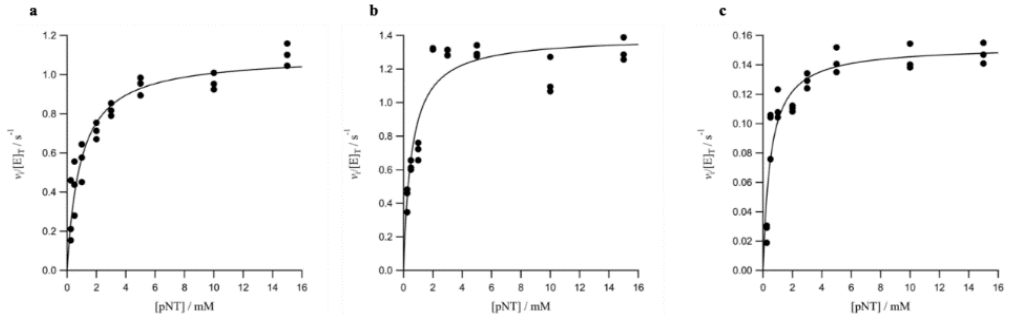


Figure 4.10. Michaelis-Menten curves for the *Thermus nonproteolyticus* glycoside hydrolase single mutants with 4-nitrophenyl-β-D-thioglucopyranoside (pNT-Glc) substrate at 50 °C; **a** L171K, **b** V287R, **c** H178Y.

The kinetics study of *Tno*GH1 single mutants with the thioglycoside substrate, showed a novel significantly enhanced activity and an improvement in the kinetic parameters compared with the WT *Tno*GH1 and the other two variants of the enzyme, supporting the validity of the *in-silico* rational approach.

This experimental work showed that the mutation V287R in *Tno*GH1 active site yielded the greatest improvement in specificity while retaining similar turnover number to the WT. It has

been shown previously R312 plays a critical role in aphid myrosinase as a sulphate recognition.[14] Its introduction in *Tno*GH1 to replace the hydrophobic side chain of valine in the same position is highly beneficial. A mutation to arginine introduces a positively charged guanidinium side chain into the active site of *Tno*GH1, which plays an important role in protein stabilization by hydrogen and/or ionic bonding to a negatively charged anion.[16] The prediction of the distance between the V287 in *Tno*GH1 and the oxygen of glycosidic substrate in chapter 3 (see section 3.3) showed that this bond can not be hydrogen bond. Therefore, the recorded improvement in k_{cat}/K_M may correspond to electrostatic interaction between the sulphur atom in pNT-Glc substrate and the positively charged side chain of arginine in the mutant which potentially stabilize the transition state resulting in an increase in k_{cat}/K_M.

The mutation that introduces lysine into the active site (L171K) is correlated with K173 in *Bb*MYR. This mutation also adds a positive side chain into the active site however a benefit in k_{cat}/K_M was not observed. In the crystal structure of *Bb*MYR, K173 points away from the active site, possibly reducing its direct involvement in substrate binding.

The mutation to tyrosine at position 180 (*Bb*MYR number), to replace histidine, resulted in a dramatic decrease in k_{cat}/K_M towards pNT-Glc substrate compared to all other variants. Y180 has been suggested to have a possible catalytic role in *Bb*MYR due to the proximity of the side chain to the thiosidic linkage in GSL. From a chemical perspective, this mutation increases the hydrophobicity at that position. In addition, the histidine appears to be highly conserved among β-glycosidases, suggesting that a mutation at this position is poorly tolerated. Specifically, tyrosine, while not forbidden, has an incidence of less than 3% in the data set. The other two targeted positions (L171 and V287) are not as highly conserved (Figure 4.11 below).

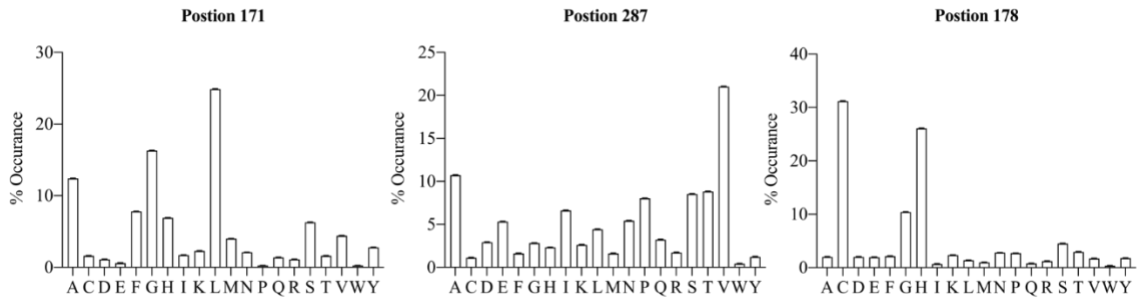


Figure 4.11. The targeted mutation positions in the WT *Thermus nonproteolyticus* glycoside hydrolase. Position 171 presents L171K mutant (leucine is 24 % conserved in this postion), position 287 presents V287R mutant (valine is 21 % conserved in this postion), and position 178 presents H178Y mutant (histidine is 26 % conserved in this postion). The data was obtained from the Pfam protein family database (https://pfam.xfam.org - sequence search), ~4000 sequences of representative proteome were used for amino acid analysis.

4.6.2 Double mutants

The *Tno*GH1 double mutants were engineered by starting from either one of the single mutant plasmids or by double mutation using two primers at the same time (see Material and Methods section 2.4.1). Successful mutants were confirmed by sequencing and expressed as previously described (see section 4.2).

All *Tno*GH1 double mutant showed a sufficient level of expression, were purified, and analysed with a 12% SDS PAGE (Figure 4.12).

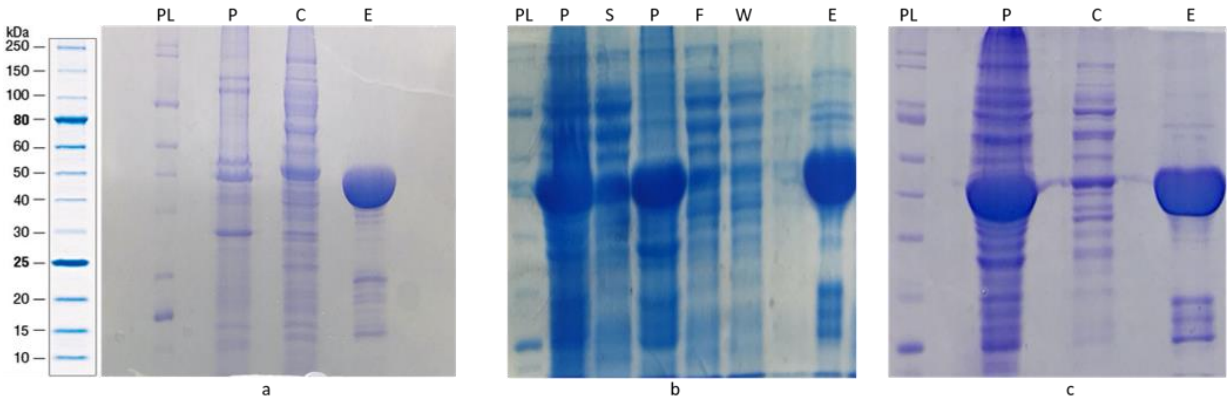


Figure 4.12. Purification of *Thermus nonproteolyticus* glycoside hydrolase double mutants as analysed by 12% acrylamide SDS-PAGE. **a** *Tno*GH1-L171K/V287R. **b** *Tno*GH1-L171K/H178Y. **c** *Tno*GH1-V287R/H178Y. PL is the protein ladder fraction (5 µL), P is the pellet fraction after sonication (10 µL), C is crude extract fraction of the mutants (10 µL), F is flow through fraction of the mutants (10 µL), W is wash fraction of the mutants (10 µL) eluted with 90% loading buffer and 10% elution buffer, and E is the pure fraction of the mutant (10 µL).

The amounts of purified WT and mutant enzyme are compared in Tables 4.1 and 4.4. Both single and double mutants showed lower expression level than the expression level obtained for WT.

Table 4.4. Protein yields of WT and mutants *Thermus nonproteolyticu* glycoside hydrolase.

	Protein yield/ mg/L
WT	25
*Tno*GH1-L171K/V287R	13
*Tno*GH1-L171K/H178Y	8
*Tno*GH1-V287R/H178Y	20

4.6.2.1 Activity assays of the double mutants with pNP-Glc and pNT-Glc substrates

In order to investigate the effect of combining two mutations, the variants were subjected to the same experiments carried out for the single mutants.

Overall, with the pNP-Glc substrate better performance, in terms of specific activity, was recorded for the WT *Tno*GH1 than the double mutants at both 25 °C and 50 °C (Table 4.5).

Table 4.5. Specific activities of the WT enzyme and double mutants of *Thermus nonproteolyticus* glycoside hydrolase with 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc) and with 4-nitrophenyl-β-D-thioglucopyranoside (pNT-Glc) at 25 °C and at 50 °C. The values in brackets under the pNP-Glc columns are the specific activities of the enzymes in the presence of 10% DMSO.

	Specific activity/ U/mg (pNP-Glc) 25 °C	Specific activity/ U/mg (pNP-Glc) 50 °C	Specific activity/ U/mg (pNT-Glc) 25 °C	Specific activity/ U/mg (pNT-Glc) 50 °C
WT	1.7 ± 0.3 (3.5 ± 0.5)	3.5 ± 0.2 (5.5 ± 0.5)	0.014 ± 0.002	0.028 ± 0.004
*Tno*GH1-L171K/V287R	0.5 ± 0.1 (0.6 ± 0.1)	1.2 ± 0.1 (1.4 ± 0.1)	0.002 ± 0.001	0.005 ± 0.001
*Tno*GH1-L171K/H178Y	0.3 ± 0.1 (0.5 ± 0.1)	0.8 ± 0.1 (0.9 ± 0.1)	0.0003 ± 0.0001	0.001 ± 0.001
*Tno*GH1-V287R/H178Y	0.7 ± 0.1 (0.9 ± 0.1)	1.4 ± 0.1 (1.6 ± 0.1)	0.005 ± 0.001	0.011 ± 0.001

The double mutants again performed better at 50 °C (2-3 folds increase), however, with the exception of H178Y, all single mutants had a higher specific activity than any of the double mutants. The H178Y mutation, on the other hand, provide significant benefits when combined with the V287R mutation at 50 °C; the specific activity of the enzyme doubling from 0.7 U/mg to 1.4 U/mg. As can be seen from Table 4.5 above, slight increases in the activities of the WT and double mutant enzyme with pNP-Glc were noticed when the assays were carried out in the presence of 10% DMSO. The relative activities of *Tno*GH1 double mutants to the WT *Tno*GH1 with pNP-Glc substrate is shown in Figure 4.8 above. Moreover, *Tno*GH1 double mutants showed similar performance compared single mutants with the β-glucosidase natural substrate (pNP-Glc), with the exception of *Tno*GH1-V287R.

The activity of the double mutants with pNT-Glc was also assessed. Again, it must be noted that these experiments were carried out with the addition of a 10% DMSO.

In general, it was confirmed that introducing additional mutations lowered the overall efficiently of the enzymes and the WT *Tno*GH1 again showed better specific activity values than the double mutants (Table 4.5. Figure 4.8 above). However, the positive effect of the addition of 10% DMSO and/or increasing the temperature to 50 °C was evident.

The decrease in the catalytic activity of an enzyme following mutations in the active site has also been reported in the literature.[17-20]

4.6.2.2 Kinetics assays of the double mutants with pNP-Glc and pNT-Glc

Kinetics parameters were determined for the double mutants (Table 4.6). Not surprisingly, the WT *Tno*GH1 was still the best with pNP-Glc, however, among the double mutants, some differences were observed.

Table 4.6. Table summarizing the kinetic parameters of the WT enzyme and double mutants of *Thermus nonproteolyticus* glycoside hydrolase (*Tno*GH1) with 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc) and 4-nitrophenyl β-D-thioglucopyranoside (pNT-Glc). Experiments were conducted in triplicate. Standard errors are given, based on fitted theoretical curves.

	k_{cat}/K_M(pNP-Glc)/ $M^{-1}s^{-1}$	k_{cat}(pNP-Glc)/ s^{-1}	K_M(pNP-Glc)/ mM	k_{cat}/K_M (pNT-Glc) / $M^{-1}s^{-1}$	k_{cat} (pNT-Glc)/ s^{-1}	K_M(pNT-Glc)/ mM
WT *Tno*GH1	$246 \times 10^3 \pm 3 \times 10^3$	133 ± 3	0.54 ± 0.06	957 ± 70	1.34 ± 0.04	1.40 ± 0.10
*Tno*GH1-L171K/V287R	$58 \times 10^3 \pm 5 \times 10^3$	111 ± 3	1.90 ± 0.20	1810 ± 320	0.38 ± 0.01	0.21 ± 0.04
*Tno*GH1-L171K/H178Y	$99 \times 10^3 \pm 10 \times 10^3$	69 ± 2	0.70 ± 0.10	364 ± 80	0.08 ± 0.04	0.22 ± 0.05
*Tno*GH1-V287R/H178Y	$260 \times 10^3 \pm 40 \times 10^3$	106 ± 3	0.41 ± 0.07	2400 ± 500	0.84 ± 0.04	0.35 ± 0.08

The *Tno*GH1-L171K/V287R mutant showed a 4-fold loss in specificity, a 1/3-loss in turnover rate, and 3-fold loss in substrate affinity compared with the WT. *Tno*GH1-L171K/H178Y mutant has a 2-fold loss is observed in both specificity and turnover rate, while *Tno*GH1-V287R/H178Y mutant shows similar specificity and substrate affinity towards pNP-Glc as the WT, but still a 30% decrease in turnover rate (Figure 4.13).

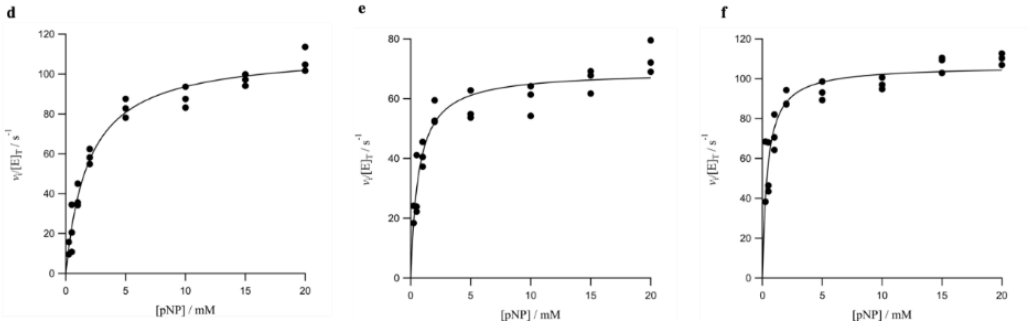


Figure 4.13. Michaelis-Menten curves for the *Thermus nonproteolyticus* glycoside hydrolase double mutants with 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc) substrate at 50 °C; **d** L171K/V287R, **e** L171K/H178Y, **f** V287R/H178Y.

More interesting of course is the performance of these variants with pNT-Glc (Table 4.6 above and Figure 4.14 below). The double mutant *Tno*GH1 L171K/V287R shows a 2-fold increase in specificity, a 3.5-fold decrease in turnover rate and 6.6-fold increase in substrate affinity, whereas with pNP-Glc substrate, this mutant shows a decrease in specificity, turnover rate, and substrate affinity compared with the WT.

The combination of a lysine with tyrosine in *Tno*GH1-L171K/H178Y mutant decreases both the specificity (3-fold) and the turnover rate (17-fold), consistent with what was observed with pNP-Glc, but an increase in substrate affinity (6.6-fold) similar to what was observed above with *Tno*GH1 L171K/V287R.

The V287R/H178Y mutants showed a 2.5-fold increase in specificity, a 1/3-fold decrease in turnover rate, and 4-fold increase in substrate affinity compared with the WT enzyme. This is comparable to what was observed in the single mutant V287R. *Tno*GH1-V287R/H178Y had previously shown a similar specificity to the WT enzyme for pNP-Glc, while here with pNT-Glc a gain in specificity was observed.

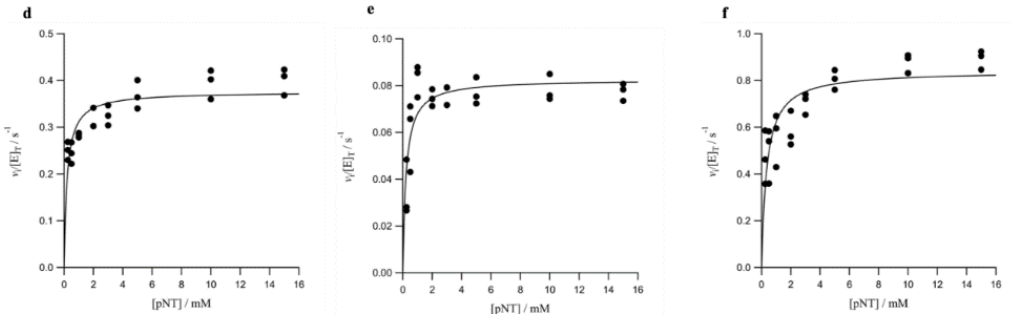


Figure 4.14. Michaelis-Menten curves for the *Thermus nonproteolyticus* glycoside hydrolase double mutants with 4-nitrophenyl-β-D-thioglucopyranoside (pNT-Glc) substrate at 50 °C; **d** L171K/V287R, **e** L171K/H178Y, **f** V287R/H178Y.

All the *Tno*GH1 double mutants showed lower turnover rates (0.08 - 0.84 s^{-1}) than the WT (1.34 s^{-1}) and any single mutant (1.10 – 1.40 s^{-1}) with pNT-Glc substrate except *Tno*GH1-H178Y (0.15 s^{-1}).

4.7 *T. nonproteolyticus* glycoside hydrolase conclusion

The single mutant *Tno*GH1-V287R showed a 3-fold increase in specificity (2500 M^{-1} S^{-1}) with no loss in turnover rate (1.40 s^{-1}) when compared with the WT enzyme (K_{cat}/K_M is 957 M^{-1} s^{-1} and K_{cat} is 1.34 s^{-1}) which makes this mutant the best of this selection.

Moreover, from the kinetics characterization of WT *Tno*GH1, single, and double mutants we can clearly observe a combined effect of mutations on kinetic parameters of the double mutants with pNT-Glc substrate, for instance the combination of the best single mutant, *Tno*GH1 V287R with L171K (K_{cat}/K_M is 1200 M^{-1} s^{-1}) and with H178Y (K_{cat}/K_M is 560 M^{-1} s^{-1}) mutants improves the specificity of these mutations to 1810 M^{-1} s^{-1} for *Tno*GH1-L171K/V287R and to 2400 M^{-1} s^{-1} for *Tno*GH1-V287R/H178Y.

Thus, the role of arginine (*Tno*GH1-V287R) was investigated in this chapter as a key to induce β-thioglycosidase activity into *Tno*GH1 with a remarkable enhancement of the specificity (k_{cat}/K_M) of *Tno*GH1 towards pNT-Glc substrate.

The results of this chapter raised few questions for future developments such as:

- are these generalized conclusion transferable to another GH1, i.e. will analogous mutations reproduce the observed effects?
- or they are specific to *Tno*GH1?
- is R312 (*Bb*MYR numbering) playing similar or different roles in another GH1?

Therefore, and to be able to answer these questions the selected residues (K173, R312, and Y180; *Bb*MYR numbering) were introduced into another GH1 which generated new GH1 mutants to be characterized in the hydrolysis of both glycosidic (pNP-Glc) and thioglycosidic (pNT-Glc) substrates and this is will be the focus of the next chapter.

4.8 Bibliography

1 X.-Y. He, X.-Q. Wang, S.-J. Yang, W.-R. Chang and D.-C. Liang, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2001, **57**, 1650–1651.

2 X. Wang, X. He, S. Yang, X. An, W. Chang and D. Liang, *J. Bacteriol.*, 2003, **185**, 4248–4255.

3 J. C. Yang XP, Yang SJ, Han BZ, *Chin. J. Biotechnol.*, 2005, **21**, 84–91.

4 P. Turner, G. Mamo and E. N. Karlsson, *Microb. Cell Fact.*, 2007, **6**, 1–23.

5 G. M. Mathew, A. Madhavan, K. B. Arun, R. Sindhu, P. Binod, R. R. Singhania, R. K. Sukumaran and A. Pandey, *Appl. Biochem. Biotechnol.*, , DOI:10.1007/s12010-020-03416-5.

6 J. Chapman, A. E. Ismail and C. Z. Dinu, *Catalysts*, 2018, **8**, 20–29.

7 V. Dachuri, J. Boyineni, S. Choi, H. S. Chung, S. H. Jang and C. W. Lee, *J. Mol. Catal. B Enzym.*, 2016, **131**, 73–78.

8 N. Doukyu and H. Ogino, *Biochem. Eng. J.*, 2010, **48**, 270–282.

9 D. H. Rammler, *Ann. N. Y. Acad. Sci.*, 1967, **141**, 291–299.

10 M. Patil, *Int. J. Mol. Vet. Res.*, 2013, **3**, 23–33.

11 N. A. David, *Annu. Rev. Pharmacol.*, 1972, **12**, 353–374.

12 L. Misuri, M. Cappiello, F. Balestri, R. Moschini, V. Barracco, U. Mura and A. Del-Corso, *J. Enzyme Inhib. Med. Chem.*, 2017, **32**, 1152–1158.

13 H. J. Wiggers, J. Cheleski, A. Zottis, G. Oliva, A. D. Andricopulo and C. A. Montanari, *Anal. Biochem.*, 2007, **370**, 107–114.

14 A. M. E. Jones, P. Winge, A. M. Bones, R. Cole and J. T. Rossiter, *Insect Biochem. Mol. Biol.*, 2002, **32**, 275–284.

15 H. Husebye, S. Arzt, W. P. Burmeister, F. V Ha and A. Brandt, *Insect Biochem. Mol. Biol.*, 2005, **35**, 1311–1320.

16 C. L. Hannon1 and E. V. Anslyn1, *The Guanidinium Group: Its Biological Role and Synthetic Analogs*, Bioorganic Chemistry Frontiers. Springer, Berlin, Heidelberg Springer-Verlag Berlin Heidelberg, 1993.

17 C. Diggle, P. G. Quirk, T. Bizouarn, R. L. Grimley, N. P. J. Cotton, C. M. Thomas and J. B. Jackson, *J. Biol. Chem.*, 1996, **271**, 10109–10115.

18 F. J. GIORDANO, P. PING, M. D. McKIRNAN, S. NOZAKI, A. N. DEMARIA, W. H. DILLMANN, O. MATHIEU-COSTELLO and H. K. HAMMOND, *Nat. Med.*, 1996, **2**, 534–539.

19 I. Weisberg, P. Tran, B. Christensen, S. Sibani and R. Rozen, *Mol. Genet. Metab.*, 1998, **64**, 169–172.

20 J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St.Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael and D. Baker, *Science.*, 2010, **329**, 309–313.

Chapter 5*: Halothermothrix orenii* glycoside hydrolase

5.1 Introduction

As mentioned in Chapter 4, this chapter will focus on another extremophilic GH1, to evaluate the transferability of the mutations studied in *Tn*oGH1 and their effect on enzyme kinetics.

It is possible for specific mutations to be transferable within the same group of enzymes, leading to similar effects. For instance, previous work in our group (2019) proved the role of the amino acid Asn120 in pyridoxal-5-phosphate (PLP) binding in three homologous transaminases of *Halomonas elongata* (HeTA), *Chromobacterium violaceum* (CvTA), and *Pseudomonas fluorescens* (PfTA). In the *Halomonas elongata* transaminase (HeTA), the position of Asn120 plays a critical role in the interaction with Asp that protonate the PLP pyridinium nitrogen; however, Val was the equivalent residue in other TAs. The thermostability of both CvTA-V124N and PfTA-V129N was enhanced when compared with the thermal stability of the WT CvTA and WT PfTA. Therefore, the introduction of Asn120 by mutation had the same effect in these three transaminases.[1]

In another example reported recently in 2020 by Sun and his group, an excellent improvement in the glucose-tolerance of a novel GH1 from a marine bacterium was achieved by a mutation of an amino acid residue which had been previously identified to be responsible for the glucose tolerance in another GH1 (*Humicola insolens*, HiBG). W168 and L173 in HiBG were shown to provide a hydrophobic environment which faciliate glucose ingress into the deep and narrow GH1 active site. In the marine bacterium GH1, the F171 and L176 are the two equivalent residues to W168 and L173 of HiBG. Comparing with the WT enzyme, the F171W mutant enhanced the glucose-tolerance of the marine bacterium GH1 and subsequently lead to increase the enzymatic applications of the enzyme in biofuels production and in the fermentation of lignocellulosic sugars.[2–4]

In Chapter 4, rational mutations to the active site of *Tno*GH1 were carried out to enhance the capacity of the enzyme to hydrolyse thioglycosides. In this chapter, a second glycosyl hydrolase, *Hor*GH1 is used as a template to further confirm the key role played by the selected residues (K173, R312, and Y180; *Bb*MYR numbering) in the recognition and hydrolysis of thioglycosides. Hence, it was of interest to introduce these mutations into *Hor*GH1and to determine the resulting effects on the enzymes kinetic properties.

As mentioned in the introduction (see section 1.2.2), *Hor*GH1 is a halotolerant GH1 isolated from *Halothermothrix orenii* which, as GH1, has two critical glutamic acid residues in the active site; the acid catalyst is E166 located in the T163LNEP167motif (β-strands 4), and the nucleophile is E354, in the I352TENG356 motif (β-strands 7).

A small collection of extremophilic GH1s (six enzymes) including *Hor*GH1 were available in the group and *Hor*GH1 was selected as it is a well characterized enzyme[5,6] with high performance at 25 °C in terms of activity and kinetic parameters as mentioned in the introduction (see section 1.2.2). It also shows high level of expression in BL21 with an excellent protein yield compared to the other five GH1s, even better than *Tno*GH1. In addition to that, there is a higher similarity between the amino acid sequence of *Hor*GH1 and that of the target *Bb*MYR (*Hor*GH1 shows 56.2% similarity and 36 % identity while *Ton*GH1 presents 52.3% similarity and 33.5% identity, see section 3.2). The greater versatility of *Hor*GH1 (active in a range of temperatures, not just at higher temperatures as is the case for *Tno*GH1) makes it possibly a better biocatalyst for larger scale applications.

The aim of the study described in this chapter was to t probe whether the selected mutations designed initially for *Tno*GH1 were transferable mutations to *Hor*GHG1 and what effects they had with respect to engineering S-glycosidase activity into an easier to handle GH1 enzyme.

5.2 Expression and purification

Single and double mutants were created for *Hor*GH1, carrying equivalent mutations to those created for *Tno*GH1. These *Hor*GH1 variants were expressed following the protocol reported in section 2.4.4.3. In terms of molecular weight (Mw), *Hor*GH1 is heavier than *Tno*GH1 (52 KDa and 49 KDa, respectively Figure 5.1). As shown in the SDS-PAGE, *Hor*GH1 expressed well in LB medium and showed a protein yield (35 mg/L) better than the protein yield of *Tno*GH1 (25 mg/L).

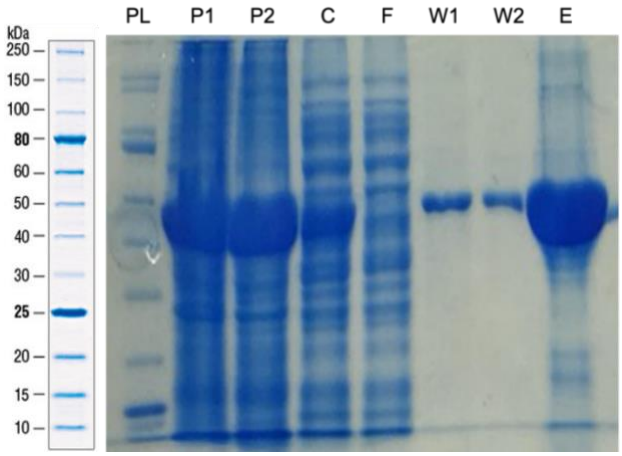


Figure 5.1. Purification of WT *Halothermothrix orenii glycoside* hydrolase (WT *Hor*GH1) as analysed by 12% SDS-PAGE. PL is the protein ladder (5 µL), P1 is the pellet fraction after sonication (10 µL), P2 is the pellet fraction after centrifuge the enzyme (10 µL), C is crude extract fraction of the enzyme (10 µL). F is flow through fraction of the enzyme (10 µL), W1 is first wash fraction of the enzyme eluted with 100% loading buffer, W2 is second wash fraction of the enzyme eluted with 90% loading buffer and 10% elution buffer, and E is the pure fraction of the enzyme (10 µL).

5.3 Activity assays with pNP-Glc and pNT-Glc substrates

As with *Tno*GH1, the activity of *Hor*GH1 was assayed with both glucosidic substrates; pNP-Glc and pNT-Glc. In this case however, the activity of the salt adapted *Hor*GH1 with both substrates was assayed only at the standard 25 °C.

Thus, a stock solution of 10 mg/mL of freshly purified enzyme was diluted several-fold and tested with 10 mM pNP-Glc or pNT-Glc (10% DMSO).

Unlike *Tno*GH1, *Hor*GH1 was less active in the presence of DMSO and it showed 30%-40% loss of activity with pNP-Glc substrate in the presence of 10%-20% DMSO, respectively; this is not uncommon.[7,8] Many examples have been reported in literature about the effect of DMSO on enzymatic reactions[9–11], which can be positive or negative. For instance, the activity of *Trypanosoma cruzi* glyceraldehyde-3-phosphate dehydrogenase was enhanced in the presence of DMSO (as for*Tno*GH1), whereas that of human aldose reductase (AR) was reduced similar to that was observed here for *Hor*GH1. [12,13]

Therefore, the activity of WT *Hor*GH1 and the mutants with pNP-Glc was not tested with 10% DMSO (as *Tno*GH1) as it negatively impacts the activity of the enzyme. However, DMSO is added to solubilise pNT-Glc substrate to enable the same substrate concentrations and reaction conditions established for the *Tno*GH1 pNT-Glc assays.

With pNP-Glc, at 25 °C *Hor*GH1 yielded the same specific activity (1.9 U/mg) which was obtained before in the lab. Together with the results obtained with *Tno*GH1, under the same standard conditions (25 °C), no real difference can be appreciated between the enzymes as both GH1s show equal level of activity towards the natural substrate. The active site is highly conserved across these two GH1s with a high level of similarity (62%). Their catalytic residues are also identical in term of position and amino acids residues as mentioned in Chapter 4 and as shown in Figure 5.2 below.

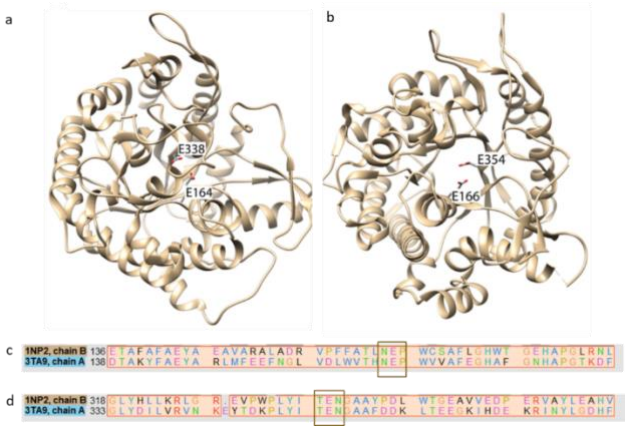


Figure 5.2. Comparison of *Hor*GH1 and *Tno*GH1 structures and catalytic site sequences **a.** Protein database structures of *Thermus nonproteolyticus* glycoside hydrolase enzyme (pdb:1NP2) showing the active site residues (E164 and E338), **b.** structures of *Halothermothrix orenii* glycoside hydrolase enzyme (pdb:3TA9) showing the active site residues (E166 and E354), **c.** amino acid sequence alignment of *Thermus nonproteolyticus* glycoside hydrolase (1NP2, chain B) and *Halothermothrix orenii* glycoside hydrolase (3TA9, chain A) showing T^{161}LNEP^{165}motif in the active site of *Tno*GH1 (β-strands 4) and T^{163}HNEP^{167}motif in the active site of *Hor*GH1 (β-strands 4), **d.** amino acid sequence alignment of *Thermus nonproteolyticus* glycoside hydrolase (1NP2, chain B) and *Halothermothrix orenii* glycoside hydrolase (3TA9, chain A) showing I^{336}TENG^{340} motif in the active site of *Tno*GH1(β-strands 7) and I^{352}TENG^{356} motif in the active site of *Hor*GH1 (β-strands 7).

It is important to note that *Tno*GH1 is a thermophilic enzyme and while its activity at 25 °C towards pNP-Glc was similar to *Hor*GH1, its optimal activity was recorded at 50 °C where it displayed a specific activity of 3.8 U/mg.

Unexpectedly, *Hor*GH1 was quite active also with pNT-Glc, and a similar specific activity value was observed (1.85 U/mg). For comparison, *Tno*GH1 had a specific activity of 0.014 U/mg at 25 °C which increased only to 0.028 U/mg at 50 °C. Therefore, unlike *Tno*GH1, *Hor*GH1 has a good natural activity towards the thioglycoside substrate, and for the purpose of this research this poses an interesting question as to whether this can be further enhanced through the identified mutations.

5.4 Kinetics assays with pNP-Glc and pNT-Glc substrates

The experimental set up for *Hor*GH1 mimicked that of *Tno*GH1 to be able to compare performance under the same conditions. Importantly, the temperature in all *Hor*GH1

experiments was kept at 25 °C but it is compared with the optimal performance of *Tno*GH1 at 50 °C. The kinetics assay summarised in the primary plot of *Hor*GH1 activity with pNP-Glc (Figure 5.3 **a**) shows that *Hor*GH1 is indeed a better catalyst with better kinetics parameters (k_{cat}/K_M = 102000 M^{-1} s^{-1}, k_{cat} = 80 s^{-1}) than *Tno*GH1 (k_{cat}/K_M = 17340 M^{-1} s^{-1}, k_{cat} = 25.14 s^{-1}). In addition, the lower working temperature further increases the potential of this enzyme for industrial applications.

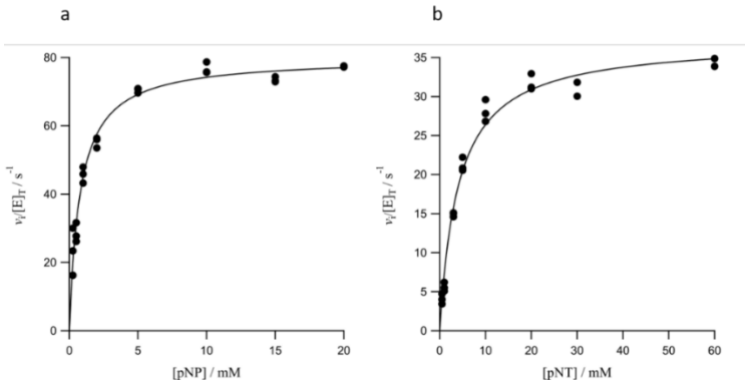


Figure 5.3. Michaelis-Menten primary plot of the activity of WT *Halothermothrix orenii* glycoside hydrolase (WT *Hor*GH1). **a** with 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc) substrate concentration (v_i vs substrate concentration) Michaelis-Menten kinetics behaviour described by equation 1, markers represent individual data points. The line is characterised by the parameters k_{cat} = 80 s^{-1}, K_M = 0.5 mM, k_{cat}/K_M = 102 × 10^3 M^{-1} s^{-1}. **b** with 4-nitrophenyl-β-D-thioglucopyranoside (pNT-Glc) substrate concentration (v_i vs substrate concentration) Michaelis-Menten kinetics behaviour described by equation 1, markers represent individual data points. The line is characterised by the parameters k_{cat} = 37.2 s^{-1}, K_M = 4.1 mM, k_{cat}/K_M = 9083 M^{-1} s^{-1}.

With pNT-Glc, Figure 5.3 **a** and **b** show that *Hor*GH1 required a higher substrate concentration to reach saturation (60 mM) with respect to *Tno*GH1 (15 mM). The measured K_M for *Hor*GH1 is in fact 4.10 mM, almost 4-fold the value of *Tno*GH1 (1.43 mM) but the k_{cat} as well as k_{cat}/K_M were significantly better at 37.20 s^{-1} and 9083 M^{-1} s^{-1}, respectively. Clearly, *Hor*GH1 is as a better catalyst with both pNT-Glc (k_{cat} of ~ 28 - fold higher than *Tno*GH1) and pNP-Glc (k_{cat} of ~ 3 - fold higher than *Tno*GH1).

5.5 *H. orenii* glycoside hydrolase mutagenesis

In order to Investigate the role of R312 (*Bb*MYR numbering) in the active site of *Hor*GH1 and to confirm if the effects bestowed by specific *Tno*GH1 mutations are transferable to another GH1, six variants of *Hor*GH1 mutants were engineered by rational design as a result of the sequence alignment between *Bb*MYR and *Hor*GH1 (Figure 2.7).

E173K, M299R, H180Y, E173K/M299R, E173K/H180Y, and M299R/H180Y were the *Hor*GH1 mutants to be investigated as single and double in the recognition and hydrolysis of both glycosidic (pNP-Glc) and thioglycosidic (pNT-Glc) substrates.

Details of the activity and kinetics assays of each mutant are outlined below.

5.5.1 Single mutants

The *Hor*GH1-E173K, *Hor*GH1-M299R, and *Hor*GH1-H180Y mutants were generated by SDM as described in section 2.4.1. Successful variants were confirmed by sequencing.

The three single mutants were expressed, purified, and analysed by SDS PAGE, as shown in Figure 5.4.

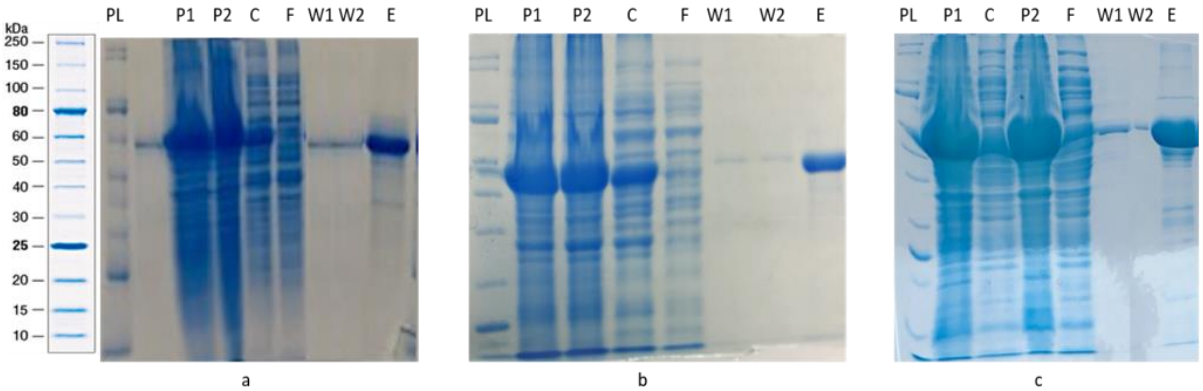


Figure 5.4. Purification of *Halothermothrix orenii* glycoside hydrolase single mutants as analysed by 12% acrylamide SDS-PAGE. **a** *Hor*GH1-E173K. **b** *Hor*GH1-M299R. **c** *Hor*GH1-H180Y. PL is the protein ladder fraction (5 µL), P1 is the pellet fraction after sonication (10 µL), P2 is the pellet fraction after centrifuge (10 µL), C is crude extract fraction of the mutants (10 µL), F is flow through fraction of the mutants (10 µL), W1 is first wash fraction of the mutants (10 µL) eluted with 100% loading buffer, W2 is second wash fraction of the mutants (10 µL) eluted with 90% loading buffer and 10% elution buffer, and E is the pure fraction of the mutant (10 µL).

Similar to what previously observed with *Tno*GH1, all mutants gave a good level of expression though lower than WT *Hor*GH1 as shown in Table 5.1.

Table 5.1. Protein yields of WT and single mutants *Halothermothrix orenii* glycoside hydrolase.

	Protein yield mg/L
WT	35
*Hor*GH1-E173K	15
*Hor*GH1-M299R	25
*Hor*GH1-H180Y	10

5.5.1.1 Activity assays of the single mutants with pNP-Glc and pNT-Glc substrates

As observed for *Tno*GH1, the mutants show a different affinity for the natural ligand. The data summarised in Table 5.2 shows that the replacement of glutamic acid by lysine in the active site of *Hor*GH1, position 173, decreased the specific activity of *Hor*GH1 by 70% towards pNP-Glc, while both M299R and H180Y mutations induced a loss of almost 50%. The single mutations effect on the activity of *Hor*GH1 with pNP-Glc is similar to what observed with *Tno*GH1 single mutants (30 - 70 % loss).

Table 5.2. Specific activities of *Halothermothrix orenii* glycoside hydrolase WT and single-mutant variants at 25 °C with 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc) and with 4-nitrophenyl-β-D-thioglucopyranoside (pNT-Glc).

	Specific activity/ U/mg (pNP-Glc)	Specific activity/ U/mg (pNT-Glc)
WT	1.95 ± 0.05	1.85 ± 0.15
*Hor*GH1-E173K	0.56 ± 0.02	0.97 ± 0.02
*Hor*GH1-M299R	0.97± 0.03	1.75 ± 0.15
*Hor*GH1-H180Y	0.85 ± 0.05	0.52 ± 0.11

The experimental work carried out with pNT-Glc however, showed that the performance of both *Hor*GH1-E173K and *Hor*GH1-M299R is better with pNT-Glc than the standard substrate. *Hor*GH1-E173K, in fact, retains 50% of its activity while *Hor*GH1-M299R is almost as active as the WT with a specific activity of 1.75 U/mg (WT activity 1.85 U/mg) towards pNT-Glc. *Hor*GH1-M299R appears to have a higher preference for the S-glycoside than the O-glycoside, considerably better than what was achieved with the best performing mutant of *Tno*GH1 (*Tno*GH1-V287R). The WT *Tno*GH1 enzyme had been found to display a very low specific activity with the pNT-Glc substrate (0.028 U/mg) and that the *Tno*GH1 V287R mutant was even lower (0.01 U/mg) (Table 5.2). On the other hand, the *Hor*GH1-H180Y mutant lost almost 72% of its activity with pNT-Glc confirming the negative impact of this mutation also in this glycosyl hydrolase. The relative activities of *Hor*GH1 single mutants to the WT *Hor*GH1 are shown as columns in Figure 5.5.

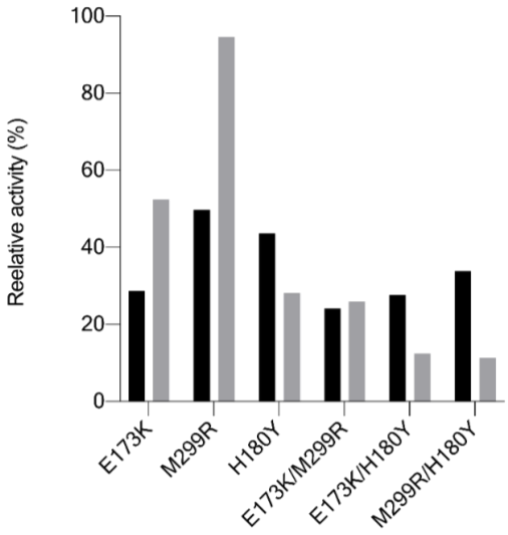


Figure 5.5. Relative activities of *Halothermothrix orenii* glycoside hydrolase mutants to the WT *Halothermothrix orenii* glycoside hydrolase. Black columns present the relative activities with 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc). Grey columns present the relative activities with the 4-nitrophenyl-β-D-thioglucopyranoside (pNT-Glc). This referred to standard conditions, 25C, saturating substrate concentrations, presence of 10% DMSO to pNT-Glc but not pNP-Glc assays.

Unlike what was observed with *Tno*GH1 single mutants data, *Hor*GH1 single mutants gave more encouraging results; despite a remarkable starting activity of the WT which virtually did not discriminate between O- or S- linkage, two of the mutants (E173K and M299R) performed better with the thio substrate than with the O-glycoside.

5.5.1.2 Kinetics assays of the single mutants with pNP-Glc and pNT-Glc substrates

The kinetics parameters of *Hor*GH1 with pNP-Glc (Table 5.3) shows that the lysine instead of glutamic acid in the active site of *Hor*GH1 (E173K), not only strongly diminished the activity towards the substrate with a 70% loss, but also negatively affected the kinetic parameters with 10-fold decrease in the specificity constant , 3-fold decrease in turnover rate, and clearly reduced substrate affinity with K_m values 5x increased which is another major contributing factor. In *Tno*GH1, the negative effect of this mutation in the kinetics parameters was less pronounced with 2.5-fold loss in the specificity constant but a retention of turnover rate with pNP-Glc (Table 4.3).

Table 5.3. Table summarizing the kinetic parameters of the WT enzyme and single mutants of *Halothermothrix orenii* glycoside hydrolase (*Hor*GH1) with the 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc) and 4-nitrophenyl β-D-thioglucopyranoside (pNT-Glc). Experiments were conducted in triplicate. Standard errors are given, based on fitted theoretical curves.

	k_{cat}/K_m(pNP-Glc)/ $M^{-1}s^{-1}$	k_{cat}(pNP-Glc)/ s^{-1}	K_m(pNP-Glc)/ mM	k_{cat}/K_m (pNT-Glc) / $M^{-1}s^{-1}$	k_{cat} (pNT-Glc)/ s^{-1}	K_m(pNT-Glc)/ mM
WT HorGH1	160000 ± 7 × 10^3	80 ± 1	0.5 ± 0.1	9073 ± 529	37.2 ± 0.7	4.1 ± 0.3
HorGH1-E173K	11000 ± 1 × 10^3	26 ± 2	2.5 ± 0.5	3351 ± 153	19.1 ± 0.3	5.7 ± 0.3
HorGH1-M299R	36000 ± 2 × 10^3	34 ± 1	0.9 ± 0.1	13480 ± 170	33.7 ± 1.2	2.5 ± 0.4
HorGH1-H180Y	55000 ± 5 × 10^3	38 ± 1	0.7 ± 0.1	2501 ± 131	9.4 ± 0.2	3.8± 0.2

Pronounced decreases in the kinetics parameters were also noticed with the *Hor*GH1-M299R mutant. This mutation showed a 4.4-fold loss in specificity towards the O-glycosidic substrate, 2.4-fold reduction in turnover, and the K_m value was almost doubled with this mutant. Again, the analogous *Tno*GH1 mutant performed better with this substrate than *Hor*GH1 mutant. *Hor*GH1-H180Y shows 3-fold loss in specificity constant, 2-fold loss in turnover number, and a decrease of the substrate affinity with the pNP-Glc substrate, similar to what was observed with the equivalent mutant in *Tno*GH1 (*Tno*GH1-H178Y).

The Michaelis-Menten curves for the *Hor*GH1 single mutants with pNP-Glc are shown in Figure 5.6.

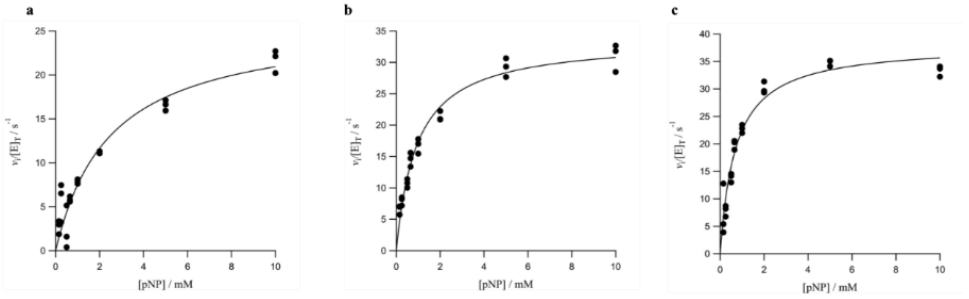


Figure 5.6. Michaelis-Menten curves for the *Halothermothrix* orenii glycoside hydrolase (*Hor*GH1) single mutants with 4-nitrophenyl β-D-glucopyranoside (pNP-Glc) substrate; **a**. E173K, **b**. M299R, **c**. H180Y.

However, when pNT-Glc was tested, the behaviour of the mutant variants shifted. *Hor*GH1-E173K still showed a decrease in the substrate affinity and other kinetic parameters (3-fold loss in specificity; 2-fold loss in turnover rate) but less marked than with pNP-Glc. The equivalent to *Hor*GH1-E173K in *Tno*GH1 is *Tno*GH1-L171K which showed a retention of kinetic parameters with pNT-Glc (Table 4.3). This mutation resulted in little perturbation in turnover rate for both substrates in *Tno*GH1, however with *Hor*GH1, while not a very active catalyst, the discrimination between the two substrates is evident.

The *Hor*GH1-M299R single mutant showed over 30% increase in specificity, a retention of turnover number, as well as an increase in the substrate affinity with the pNT-Glc substrate, was also obtained with this mutant. A similar pattern is observed in *Tno*GH1-V287R mutant which showed a significant shift in specificity towards pNT-Glc with no loss in turnover rate compared to the WT enzyme (Table 4.3).

The *Hor*GH1-H180Y mutant showed a 3-fold decrease in specificity, a 4-fold decrease in turnover, and almost a retention of the k_m value compared to the WT, similar to the change observed in the respective mutant in *Tno*GH1 (except the substrate affinity). Therefore, mutation from histidine into tyrosine showed similar effects for both glycoside hydrolases. The Michaelis-Menten curves for the *Hor*GH1 single mutants with pNT-Glc are shown in Figure 5.7.

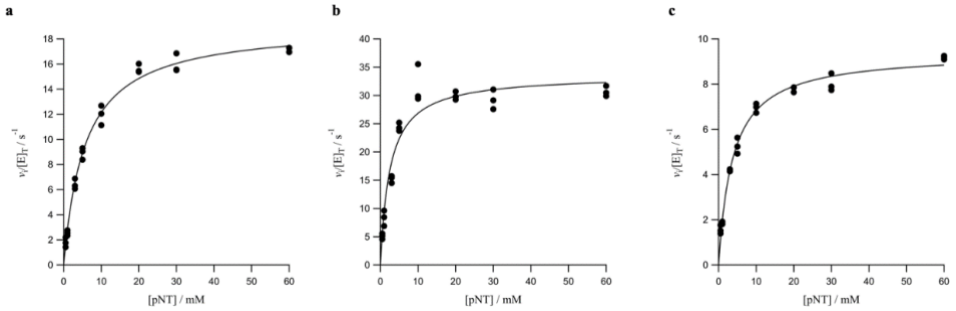


Figure 5.7. Michaelis-Menten curves for the *Halothermothrix orenii* glycoside hydrolase (*Hor*GH1) single mutants with 4-nitrophenyl β-D-thioglucopyranoside (pNT-Glc) substrate; **a**. E173K, **b**. M299R, **c**. H180Y.

The increased specificity constant seen for the *Hor*GH1-M299R single mutant adds weight to the hypothesis that R312 (*Bb*MYR numbering) is indeed playing an important role in the stabilization of the transition state as previously explained in chapter 4. On the other hand, the positive charge added by lysine in the active site of *Tno*GH1 and *Hor*GH1 did not enhance the k_{cat}/K_m values of both GH1s and conforms with the previous suggestions that K173

(*Bb*MYR numbering) might point away from the active site of *Bb*MYR. The dramatic decreases in k_{cat}/K_m and k_{cat} values towards both pNP-Glc and pNT-Glc substrates which was recorded for both extremophilic enzymes confirms that a histidine at that position is key for β-glycosidases activity.

5.5.2 Double mutants

As with *Tno*GH1, *Hor*GH1 double mutants were engineered starting from one of the single mutants plasmids or by using two primers at the same time. Following their confirmation by sequencing expression plasmids were introduced into *E. coli* strain vectors pET45b and the encoded protein variants produced and purified as described in section 2.4.1. Figure 5.8 shows that the *Hor*GH1 double mutants were well expressed and purified.

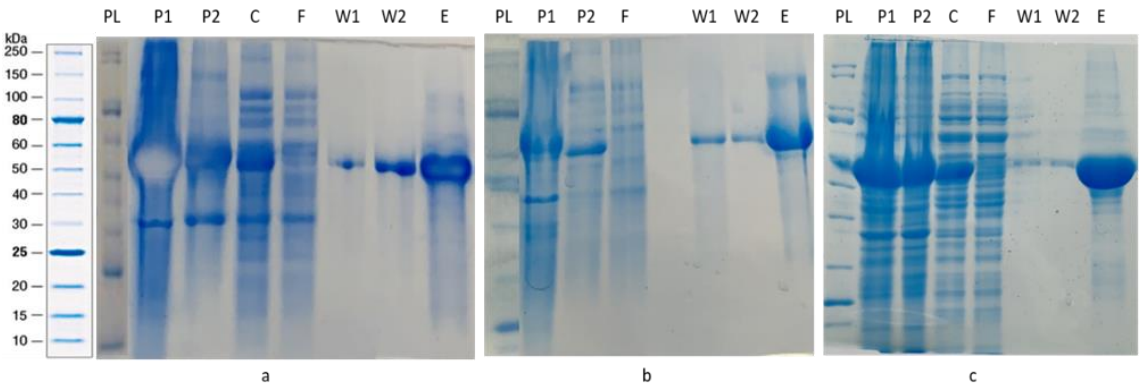


Figure 5.8. Purification of *Halothermothrix orenii* glycoside hydrolase double mutants as analysed by 12% acrylamide SDS-PAGE. **a** *Hor*GH1-E173K/M299R. **b** *Hor*GH1-E173K/H180Y. **c** *Hor*GH1- M299R/H180Y. PL is the protein ladder fraction (5 µL), P1 is the pellet fraction after sonication (10 µL), P2 is the pellet fraction after centrifuge (10 µL), C is crude extract fraction of the mutants (10 µL), F is flow through fraction of the mutants (10 µL), W1 is first wash fraction of the mutants (10 µL) eluted with 100% loading buffer, W2 is second wash fraction of the mutants (10 µL) eluted with 90% loading buffer and 10% elution buffer, and E is the pure fraction of the mutant (10 µL).

Similar to what observed above with the *Hor*GH1 single mutant proteins, the overall protein yield of the *Hor*GH1 double mutant proteins was lower than for the WT enzyme (Table 5.4).

Table 5.4. Protein yields of WT and double mutant *Halothermothrix orenii* glycoside hydrolases.

	Protein yield mg/L
WT	35
*Hor*GH1- E173K / M299R	16
*Hor*GH1- E173K /H180Y	20
*Hor*GH1- M299R /H180Y	30

5.5.2.1 Activity assays of the double mutants with pNP-Glc and pNT-Glc substrates

All double mutants showed lower specific activities with pNP-Glc than the WT *Hor*GH1 (1.9 U/mg) with losses higher than 70% for both *Hor*GH1-E173K/M299R (0.47 U/mg) and *Hor*GH1-E173K/H180Y (0.54 U/mg). The combination of both M299R and H180Y in *Hor*GH1-M299R/H180Y resulted in activities of around 60% of what was observed for the WT enzyme as shown in Table 5.5.

Table 5.5. Specific activities of WT and double mutant *Halothermothrix orenii* glycoside hydrolases at 25 °C with 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc) and with 4-nitrophenyl-β-D-thioglucopyranoside (pNT-Glc).

	Specific activity/ U/mg (pNP-Glc)	Specific activity/ U/mg (pNT-Glc)
WT	1.95 ± 0.05	1.85 ± 0.15
*Hor*GH1- E173K / M299R	0.47 ± 0.02	0.48 ± 0.01
*Hor*GH1- E173K /H180Y	0.54 ± 0.01	0.23 ± 0.01
*Hor*GH1- M299R /H180Y	0.66 ± 0.03	0.21 ± 0.02

These results are remarkably similar to *Tno*GH1 double mutants (60 - 80 % loss with pNP-Glc).

Figure 5.5 above shows the relative activities of *Hor*GH1 double mutant proteins to the WT *Hor*GH1. With pNT-Glc, the effect is even more dramatic (75 - 90% loss). As shown in Table 5.5 and Figure 5.5 above, the enzyme activity decreased to 0.48 U/mg, 0.23 U/mg, and 0.21

U/mg with *Hor*GH1-E173K/M299R, *Hor*GH1-E173K/H180Y, and *Hor*GH1-M299R/H180Y, respectively. The equivalent mutations in the active site of *Tno*GH1 showed similar results when comparing the performance of both GH1s with pNT-Glc.

5.5.2.2 Kinetics assays of the double mutants with pNP-Glc and pNT-Glc substrates

The kinetics parameters of the *Hor*GH1 double mutants are summarised in Table 5.6. These results showed similar trends to those obtained for *Tno*GH1 in chapter 4.

*Hor*GH1-E173K/ M299R showed a 21-fold loss in specificity, 7-fold loss in turnover rate, and 3x increased in K_m value (Table 5.6 and Figure 5.9). When comparing these results with those obtained for *Tno*GH1 in Table 4.6, we can see better performance of the combined of K173 and R312 mutations (*Bb*MYR numbering) towards pNP-Glc. The decrease in the kinetics parameters of the *Hor*GH1-E173K/H180Y mutant, with pNP-Glc is a 12-fold loss in specificity, 3-fold loss in turnover rate, and the K_m value increased almost 4x as a result of this mutation. However, in *Tno*GH1-L171K/H178Y mutant only a 2-fold loss was observed in the specificity and turnover rate of the enzyme with increased of K_m value from 0.54 to 0.70 for *Tno*GH1-L171K/H178Y and the WT enzyme respectively.

Table 5.6. Table summarizing the kinetic parameters of the WT enzyme and double mutants of *Halothermothrix orenii* glycoside hydrolase (*Hor*GH1) with the 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc) and 4-nitrophenyl β-D-thioglucopyranoside (pNT-Glc). Experiments were conducted in triplicate. Standard errors are given, based on fitted theoretical curves.

	k_{cat}/K_m(pNP-Glc)/ $M^{-1}s^{-1}$	k_{cat}(pNP-Glc)/ s^{-1}	K_m(pNP-Glc)/ mM	k_{cat}/K_m (pNT-Glc) / $M^{-1}s^{-1}$	k_{cat} (pNT-Glc)/ s^{-1}	K_m(pNT-Glc)/ mM
WT HorGH1	160000 ± 7 × 10^3	80 ± 1	0.5 ± 0.1	9073 ± 529	37.2 ± 0.7	4.1 ± 0.3
HorGH1-E173K/M299R	7500 ± 1 × 10^3	12 ± 1	1.6 ± 0.1	2660 ± 122	10.1 ± 0.2	3.8 ± 0.2
HorGH1-E173K/H180Y	13334 ± 1 × 10^3	24 ± 1	1.8 ± 0.2	1080 ± 61	5.2 ± 0.1	4.8 ± 0.3
HorGH1-M299R/H180Y	23000 ± 3 × 10^3	25 ± 2	1.1 ± 0.2	3840 ± 198	7.3 ± 0.1	1.9 ± 0.1

The *Hor*GH1-M299R/H180Y mutant showed a 7-fold loss in specificity, a 3-fold loss in turnover rate, and 2x increased in K_m value (Table 5.6 and Figure 5.9). This double mutant in *Tno*GH1 (*Tno*GH1-V287R/H178Y) was less affected showing similar specificity towards pNP-Glc as the WT enzyme with only 1/3-fold loss in turnover rate and a retention of the substrate affinity.

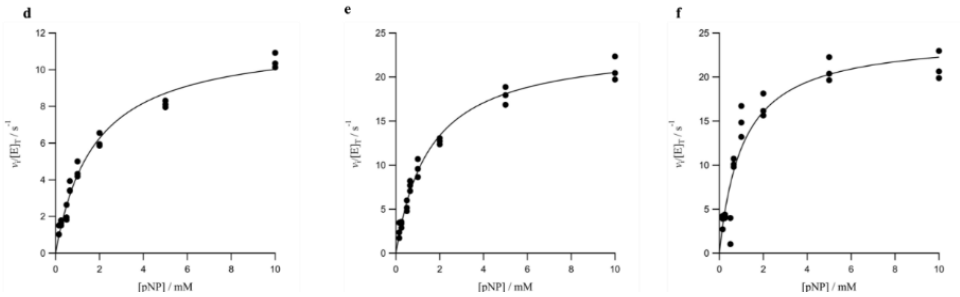


Figure 5.9. Michaelis-Menten curves for the *Halothermothrix* orenii glycoside hydrolase (*Hor*GH1) double mutants with 4-nitrophenyl β-D-glucopyranoside (pNP-Glc) substrate; **d**. E173K/M299R, **e**. E173K/H180Y, **f**. M299R/H180Y.

With pNT-Glc (Table 5.6 and Figure 5.10), *Hor*GH1-E173K/M299R showed an almost 3.4-fold decrease in specificity, 4-fold loss in turnover rate, and a retention of the substrate affinity compared to the WT enzyme. The equivalent mutant enzyme in *Tno*GH1 showed a 2-fold increase in specificity, a 3.5-fold decrease in turnover, and an increase in the substrate affinity. The *Hor*GH1-E173K/H180Y mutant showed an 8.4-fold decrease in specificity, an 7-fold in turnover resulting the lowest turnover rate (5 s^{-1}) among all other *Hor*GH1 single and double mutants. As well as a decrease in the substrate affinity was also observed with this mutant. The analogous *Tno*GH1 mutant showed decreases in both the specificity (3-fold) and the turnover rate (10-fold).

An enhancement of the substrate affinity value was achieved with the *Hor*GH1-M299R/H180Y mutant, but a near 2.4-fold decrease in specificity and a 5-fold decrease in turnover rate, similar to what was observed for *Hor*GH1-E173K/M299R. On the other hand, an improvement in the specificity of *Tno*GH1 analogous mutant (V287R/H178Y) was observed.

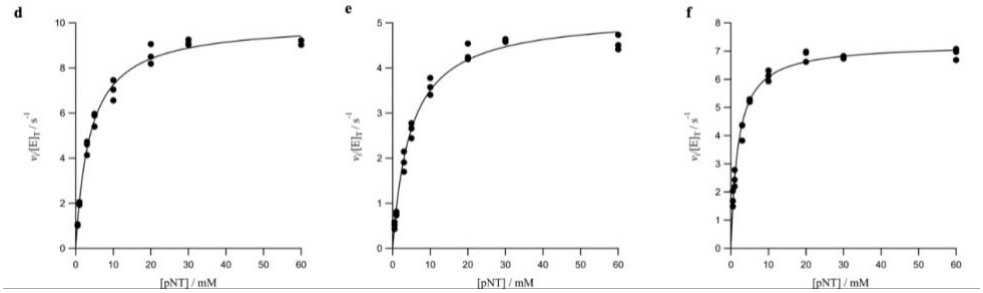


Figure 5.10. Michaelis-Menten curves for the *Halothermothrix orenii* glycoside hydrolase (*Hor*GH1) double mutants with 4-nitrophenyl β-D-thioglucopyranoside (pNT-Glc) substrate; **d**. E173K/M299R, **e**. E173K/H180Y, **f**. M299R/H180Y.

Unlike *Tno*GH1, the *Hor*GH1 double mutants showed the same shifts in kinetic parameters with both substrates as all *Hor*GH1 double mutants showed a loss in specificity and turnover rate with pNP-Glc and pNT-Glc substrates when compared to the WT enzyme.

5.6 Further studies of *Hor*GH1-M229R

The activity and stability of *Hor*GH1-M229R mutant was determined at pHs ranging from 3 to 12 (Figure 5.11).

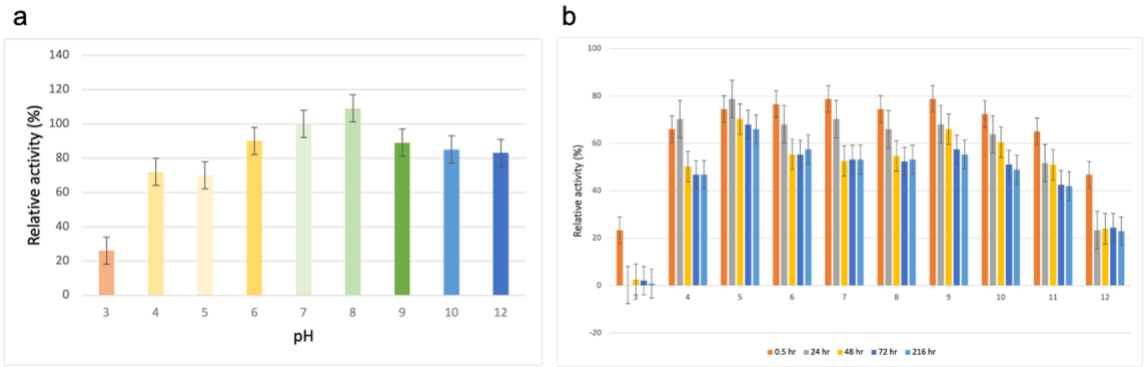


Figure 5.11. pH effect on **a** activity and **b** stability of *Hor*GH1-M299R. Experiments were conducted in triplicate.

*Hor*GH1-M229R presents the typical bell-shape profile and the optimum pH value was in the range of 7-8 as shown in Figure 5.11. However, the enzyme stored at pH 5 for more than 3

days was highly stable with virtually no loss of activity. Figure 5.12 shows the temperature stability profile of *Hor*GH1-M229R which was stable at 65 °C for 24 h confirming the enzyme displays a slight thermos adaption.[5]

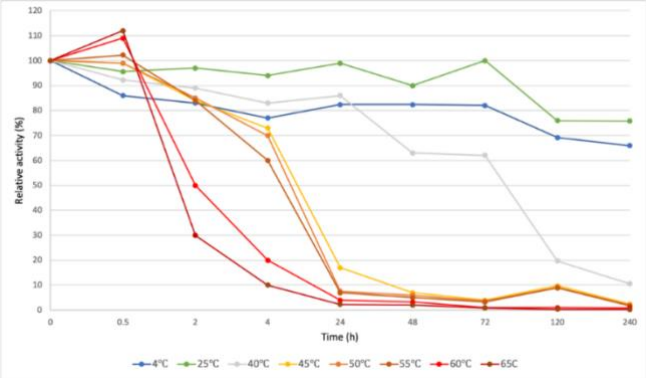


Figure 5.12. Thermostability of *Hor*GH1-M299R.

5.7 *H. orenii* glycoside hydrolase conclusion

In this chapter six variants of *Hor*GH1 were assayed for activity. Remarkable enhancement of the specificity (k_{cat}/K_m) towards β-thioglycoside was observed in the single mutant *Hor*GH1-M229R, (13480 $M^{-1}s^{-1}$) which showed a 3-fold increase with no loss in turnover rate when compared to the WT enzyme. These results are mimicking what was discussed in Chapter 4.

*Tno*GH1-V287R and *Hor*GH1-M229R are the best performing novel variants with β-thioglycosidase activity generated in this study. Among these, *Hor*GH1-M299R is the most promising mutant for further applications due to the larger turnover number (k_{cat} of *Tno*GH1-V287R is 1.40 s^{-1} and k_{cat} of *Hor*GH1-M229R is 33.7 s^{-1}). A higher turnover number of a mutant with respect to the WT enzyme has been reported in literature such as the increase in the k_{cat} as a result of mutation in the mitochondrial aldehyde dehydrogenase 2 (ALDH2; in the

liver) and in the CotA-laccases (1.4-fold) which improved the catalytic efficiency of the enzyme that was limiting its applications.[14-16]

The kinetics assays for the *Hor*GH1-M229R mutant showed that R312 (*Bb*MYR numbering) plays similar role in *Hor*GH1 as in *Tno*GH1. It also proves that, arginine 312 is key to improve β-thioglycosidase activity and its effect is not specific to *Tno*GH1, but it is a transferable to at least one other GH1. This improvement in k_{cat}/K_M and the retention of k_{cat} in both *Tno*GH1 and *Hor*GH1 suggest that V287R and M299R, respectively are the best mutants for practical application. However, among these two best mutants, *Hor*GH1- M299R is the best one.

5.8 Bibliography

1 D. Roura Padrosa, R. Alaux, P. Smith, I. Dreveny, F. López-Gallego and F. Paradisi, *Front. Bioeng. Biotechnol.*, 2019, **7**, 1–13.

2 J. Sun, W. Wang, Y. Ying and J. Hao, *Appl. Biochem. Biotechnol.*, 2020, **192**, 999–1015.

3 P. O. De Giuseppe, T. D. A. C. B. Souza, F. H. M. Souza, L. M. Zanphorlin, C. B. Machado, R. J. Ward, J. A. Jorge, R. D. P. M. Furriel and M. T. Murakami, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2014, **70**, 1631–1639.

4 Y. Yang, X. Zhang, Q. Yin, W. Fang, Z. Fang, X. Wang, X. Zhang and Y. Xiao, *Sci. Rep.*, 2015, **5**, 1–12.

5 L. D. Kori, A. Hofmann and B. K. C. Patel, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, 2011, **67**, 111–113.

6 K. Mavromatis, N. Ivanova, I. Anderson, A. Lykidis, S. D. Hooper, H. Sun, V. Kunin, A. Lapidus, P. Hugenholtz, B. Patel and N. C. Kyrpides, *PLoS One*, 2009, **4**, e4192.

7 N. Doukyu and H. Ogino, *Biochem. Eng. J.*, 2010, **48**, 270–282.

8 A. Gupta and S. K. Khare, *Crit. Rev. Biotechnol.*, 2009, **29**, 44–54.

9 D. H. Rammler, *Ann. N. Y. Acad. Sci.*, 1967, **141**, 291–299.

10 M. Patil, *Int. J. Mol. Vet. Res.*, 2013, **3**, 23–33.

11 N. A. David, *Annu. Rev. Pharmacol.*, 1972, **12**, 353–374.

12 H. J. Wiggers, J. Cheleski, A. Zottis, G. Oliva, A. D. Andricopulo and C. A. Montanari, *Anal. Biochem.*, 2007, **370**, 107–114.

13 L. Misuri, M. Cappiello, F. Balestri, R. Moschini, V. Barracco, U. Mura and A. Del-Corso,

J. Enzyme Inhib. Med. Chem., 2017, **32**, 1152–1158.

14 S. Jin, J. Chen, L. Chen, G. Histen, Z. Lin, S. Gross, J. Hixon, Y. Chen, C. Kung, Y. Chen, Y. Fu, Y. Lu, H. Lin, X. Cai, H. Yang, R. A. Cairns, M. Dorsch, S. M. Su, S. Biller, T. W. Mak and Y. Cang, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 9088–9093.

15 A. De la Vieja, C. S. Ginter and N. Carrasco, *J. Cell Sci.*, 2004, **117**, 677–687.

16 K. Z. Xu, H. R. Wang, Y. J. Wang, J. Xia, H. Ma, Y. J. Cai, X. R. Liao and Z. B. Guan, *J. Biosci. Bioeng.*, 2020, **129**, 405–411.

Chapter 6*:* Conclusion and future work

A study of myrosinase enzymes, which are unique members within the glycoside hydrolases family and are capable of hydrolysing the thioglycosidic bond in GSL, was carried out to determine the residues which play important roles in sulphate group recognition and the specific activity which distinguish MYRs from other GH1 members. The enzyme from *Brevicoryne brassicae* (*Bb*MYR) was deemed to be the best template for this work and the amino acids K173, R312, and Y180 (*Bb*MYR numbering) were highlighted in the literature as residues with key function in the hydrolysis of GSL by *Bb*MYR. However, there is no real experimental validation for this hypothesis reported in the literature.

Rational design was the technique used in this study to verify whether the selected residues did in fact have such function and could be integrated, to transfer myrosinase activity, into "standard" glycosyl hydrolases. Therefore, two extremo-adapted β-glucosidases were engineered by mapping the *Bb*MYR active site onto the *Tno*GH1 enzyme first, and then onto *Hor*GH for further confirmation. In *Tno*GH1, L171, V287, and H178 are the equivalent residues, in term of positions, to K173, R312, and Y180. Therefore, the mutations L171K, V287R, and H178Y were introduced to *Tno*GH1 as single and double mutants (generating 3 additional variants with all possible permutations). The equivalent single and double mutants in *Hor*GH1 were at positions E173K, M299R, and H180Y. The kinetic properties of all variants and WT enzymes with test substrates β-D-thioglucopyranoside (pNT-Glc) and 4-nitrophenyl-β-D-glucopyranoside (pNP-Glc) have been carried out.

Among the three residues, arginine (R312; *Bb*MYR numbering) was found to be pivotal as a marked improvement of activity towards the β-thioglucoside substrate has been observed in both mutant proteins. The *Tno*GH1-V287R and *Hor*GH1-M299R variants yielded the greatest

increase in specificity towards pNT-Glc while retaining similar turnover number to the WT. It has been shown previously that R312 plays a critical role in aphid myrosinase for sulphate group recognition. However, in this study it seems that R312 plays another important role in the enhancement of the hydrolysis of the sulphur bond contained in glucosinolate (GSLs). The pNT-Glc substrate does not have a sulphate group as the myrosinase natural substrate (sinigrin) which had been ordered several times from Carbosynth since March 2018 but was not received until today due to various issues related to the purity of the product and Covid-19 pandemic associated global lockdown. This is a clear limitation of the current study which needs to be taken into account as a starting point for future research testing the performance of *Tno*GH1-V287R and *Hor*GH1-M299R with the natural sinigrin substrate. This might either confirm the findings made here or lead to different interpretations. The introduction of a lysine in *Tno*GH1-L171K and *Hor*GH1 E173K which was highly recommended by Jones in his characterization and evolution study of a myrosinase from the cabbage aphid *Bb*MYR as mentioned in the introduction chapter (see section 1.2.3) also added a positively charged side chain into the active site, however an increase in the k_{cat}/ K_M was not observed. [1] This might also relate to the use of the tested substrate (pNT-Glc) rather than the natural one (sinigrin) since there is no sulphate group to be recognized in pNT-Glc substrate. The introduction of a tyrosine in both *Tno*GH1 and *Hor*GH1 to replace a histidine resulted in a dramatic decrease in k_{cat}/ K_M towards both pNP-Glc and pNT-Glc substrates compared to the WT. Y180 (*Bb*MYR numbering) had been suggested to have a possible catalytic role due to the proximity of the sidechain to the thioglycosidic linkage in glucosinolates.[2] However, in both extremophilic enzymes, the histidine displayed at that position appears to be highly conserved among β-glycosidases, suggesting that a mutation at this position is poorly tolerated. Specifically, tyrosine while not forbidden, had an incidence of less than 3% in the data set.

A more complex relationship was observed when the double mutants were compared to the WT and single mutants. From the single mutant results, we observed a correlation between the introduction of polar residues in the active site of *Tno*GH1 and the specificity for the pNT-Glc substrate. This is supported by the change in specificity observed between the *Tno*GH1-H178Y mutant and the *Tno*GH1-V287R/H178Y double mutant, where the latter has a much-improved specificity when compared to the former. Likewise, when *Tno*GH1-V287R was combined with *Tno*GH1-L171K, the double mutant enzyme had a higher specificity than the *Tno*GH1-L171K but lower specificity than the *Tno*GH1-V287R variant.

A similar effect was observed in *H*orGH1 specificity of pNT-Glc substrate with the *Hor*GH1-M299R/H180Y double mutant. Unlike *Tno*GH1, double mutant *Hor*GH1-E173K/M299R showed similar specificity to the *Hor*GH1-E173K single mutant. We saw a larger K_M (pNT-Glc) in the double mutant when compared to the *Hor*GH1-M299R single mutant, and a large reduction in k_{cat} compared to both single mutants. This would suggest that the same effect in the *Tno*GH1 mutants may be playing a role in the *Hor*GH1 mutants.

All *Hor*GH1 and *Tno*GH1 mutants exhibited a decrease in turnover rate and specificity with the native substrate pNP-Glc which indicates that the increase in specificity for pNT-Glc is at the expense of the native substrate. The increase observed in k_{cat}/ K_M (pNT-Glc) induced by the introduction of the arginine residues, is not observed with pNP-Glc. Considering the catalytic efficiency of the *Hor*GH1-M299R, this mutant had a ratio of k_{cat}/ K_M of glycoside to thioglycoside of 3:1, compared to the *Hor*GH1 WT which had a ratio of 11:1.

Therefore, the role of arginine proved to be key to improve β-thioglycosidase activity for the extremophilic β-glycosidases *Tno*GH1 and *Hor*GH1 by in-silico modelling of the *Bb*MYR. A 3-fold increase in specificity for the thioglycosidic substrate with no loss in turnover number

was observed by replacing of hydrophobic residues of both enzymes by arginine. These mutants were seen to have the greatest increase in specificity of all assayed mutants, including double mutants. Among the novel β-thioglycosidases addressed in this study, *Hor*GH1-M299R is the most promising mutant for the industrial application due to the larger turnover number. These results also included a thorough kinetic investigation of the different mutants which helped to shed light on the mechanism of β-glycosidases when acting on the native substrate.

Further work in this field will focus on the potential application in the synthetic direction, rather than the hydrolytic one, of *Hor*GH1 mutant variants. The aim being that of synthesising thioglycosides against the dominant hydrolytic reaction they enzyme was created to do. There are a number of strategies reported in literatures for the employment of mutant glycosidases in the formation of glycosidic bond in the synthesis of oligosaccharides. Briefly, GH1s have been successfully mutated to eliminate the catalytic nucleophile position, eliminating completely the hydrolytic activity but still offering the perfect pocket for the binding. The careful selection of glycosyl donors and acceptors yields the desired product under very mild conditions. The demonstration of the efficiency of these variants in synthetic applications would add yet a new tool to the toolbox of biocatalysts that can be used.

Bibliography

1. A. M. E. Jones, P. Winge, A. M. Bones, R. Cole and J. T. Rossiter, *Insect Biochem. Mol. Biol.*, 2002, **32**, 275–284.
2. H. Husebye, S. Arzt, W. P. Burmeister, F. V Ha and A. Brandt, *Insect Biochem. Mol. Biol.*, 2005, **35**, 1311–1320.

Publications

N. Almulhim, N. R. Moody and F. Paradisi, *Appl. Microbiol. Biotechnol.*, 2020, **104**, 4407–4415.

Acknowledgments

I would like to express my thanks and appreciation to my supervisor, Prof. Francesca Paradisi for the wonderful time I have had doing my PhD under her supervision. I am grateful for her being always available to support me academically and personally.

I would also like to especially thank Dr. Nicholas Moody for his help and support in the kinetics assay part. Thank you, Nick, for your time and patient.

Thanks, are also due to the FP (B11 lab) group. Studying under Prof. Francesca Paradisi guidance and being a member of the Paradisi's research group has been great opportunity to learn and develop in this fertile environment. I thus thank: Lidia, Dr. Eimear, Dr. David, Dr. Martina, Dr. Maria, Chris, Dr. Ana, Stephania, Valentina, and Irene for the fertile environment to learn and develop.

I would also like to show my gratitude to my home university at Saudi Arabia, King Faisal University and to the Saudi Cultural Bureau in London for all the support and advice during my study time in the UK.

My warm appreciation and infinite thanks go to my family for their love and ongoing encouragement. Special thanks to my husband Khalid for his help and support from the first moment I decided to do my postgraduate education and to my kids Mohammed and Danyah from who I was getting my everyday energy to go to the lab full of happiness and hope.