UNIVERSITY of the
WESTERN CAPE

Biochemical characterisation of an alpha-amylase with pullulan hydrolase type III characteristics derived from a hot spring metagenomics library

Bianca Boersma

Student No.: 3985248

A thesis submitted in partial fulfilment of the requirements for the degree MAGISTER SCIENTIAE (MSc)

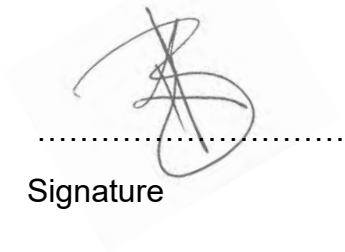
Institute for Microbial Biotechnology and Metagenomics Department of Biotechnology, University of the Western Cape, Bellville, Cape Town, South Africa

Supervisor: Prof. Marla Trindade

Co-supervisor: Dr Lonnie van Zyl

Declaration

I, Bianca Boersma, declare that “Biochemical characterisation of an alpha-amylase with pullulan hydrolase type III characteristics derived from a hot spring metagenomics library” is my own original work and that I have accurately reported and acknowledged all sources. Furthermore, this document has not previously been submitted, in part or its entirety, to any university for the purpose of obtaining an academic qualification.



…… ……..…22/02/2024……

Signature Date

Abstract

The field of metagenomics has provided biotechnological researchers with new enzymes and insight into previously unknown enzyme families with potential for novel modes of action and being able to function under extreme conditions, such as extremes of temperature, acidity or alkalinity. The enzymes can also often act on multiple substrates and show increased affinity for certain substrates. In this study, the aim was to biochemically characterise a putative type III pullulan hydrolase (PHTIII). A sequence-based metagenomic screen from previous performed by Xiao Ping Hu (2010) was used to identify novel glycosyl hydrolase enzymes in shotgun metagenomic sequence data from the Mphizi hot spring in Malawi. An identified ORF, named Pull3.1, displayed low sequence identity to a characterised α-amylase (62.71%) and 61.66% to a characterised pullulan hydrolase type III, and was, therefore, hypothesized to be an α-amylase (EC 3.2.1.1) with pullulan hydrolase type III (EC 3.2.1.x) characteristics. Further sequence analysis indicated that Pull3.1 had a glycogen binding site and that it belongs to the GH13_20 family as well as having a family 48 carbohydrate-binding module. Pull3.1 displayed the highest substrate conversion rates for starch, pullulan and glycogen. The products released were of the following degrees of polymerisation 3, 2 or 1, in accordance with the products released from the PHTIII from *Thermococcus aggregans*. Pull3.1 is a thermozyme that shows maximum activity at 75°C and an optimum pH of 7. It was thermostable up to 80°C and had a half-life of 11.6 min at 50°C. Further kinetic characterisation showed that Pull3.1 had a high affinity for starch. The K_M value of 0.048 mg/ml is within the range reported for other α-amylase enzymes, and Pull3.1 had a $V_{max}$of 0,729 (μmol mL^{-1}). All of this data indicates that Pull3.1 is a good candidate for a starch conversion process that requires thermozymes.

Acknowledgements

I thank the Lord for giving me the strength and courage to complete this thesis and for allowing me the opportunity to increase my knowledge and meet the most amazing people along the way.

I would also like to thank my parents, my sister, and new family, the Kooij family for their everyday support and encouragement. Thank you for your abundant love. Thank you for all the motivation and interest you always showed and for trying to understand even the most minor details.

To my husband, thank you for keeping me motivated at all hours of the day. Thank you for 'doing' long distance so that I could also follow my dreams. Your love and cheer kept me going.

Huge special thanks go to Hilda, my lab mate for life, and one of my very best friends, for her support with every aspect of my thesis and life during this degree course. I appreciate all your help and advice so much. I cannot thank you enough.

Thank you to my grandparents - Oupa Mark, Oupa Pieter, Ouma Bibi, Ouma Dina and Tannie Anriette - for always being excited about my project and for listening to all the developments.

The Chitin Knights: you are the best of friends and colleagues, and I already miss not seeing you all every day. Thank you Luq, Sibs, Michell and Steph, for all the coffee, discussions and fun we had.

Thank you, Prof. Patidar, for introducing me to Prof. Marla. It has been a fantastic journey and would not have been possible without the introduction your encouragement. Also, huge thanks to Prof. Marla for allowing me the opportunity to further my studies, for the support in difficult times and for your warm smile whenever I walked into your office.

Thank you, Dr Van Zyl, for all your guidance and especially for sharing exciting lab stories and discoveries.

Huge thanks go to Louwrens for pushing me to question everything I was working with and for helping me develop laboratory techniques. I would also like to thank all the post-docs and friends who shared their experiences and ideas on how to go about doing experiments. Thanks, Tslil for sharing all your thoughts and always sharing interesting facts about what you are or have worked on. Thank you most of all for your warm and kind heart.

Thank you, Anita, for allowing me to pop into your office after work for a chat and for the beautiful friendship and support you always give.

I also want to thank the bioinformatics team, Candace and Pumi, for making the boot camp even more fun and for the friendship that came from it.

Thank you to all my friends and colleagues at IMBM - Dug, Min-Ghah, Nasreen, Kanyisa, Emile, Lubabalo, Fazlin, Bongi, Phillip, Nombuso, Amy, Ross, Ellen, Anesu,

Mapula, Keith and Wesley - for the support and for creating a pleasant work environment.

Also, a big thank you to my friends Irene, Dawie, Chrisna, Angelique, Corne, Jennelee, Lesley and Mr Yogels.

I also want to thank Arina, Dr Kirby, Robin, Carmen and Leo for always helping with arrangements and for your friendly presence.

Thank you to my writing guide, Kassim, for your assistance.

Last, but not least, I want to extend my gratitude to the NRF, Prof. Michael Davies-Coleman and Prof. Marla for funding my MSc degree. This report would not have been possible without you.

Abbreviations and Acronyms

°C	Degrees Celsius
Amp	Ampicillin
APS	Ammonium persulfate
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BSA	Bovine Serum Albumin
CDase	Cyclomaltodextrinase
CD	Cyclodextrin
DNA	Deoxyribonucleic acid
DNS	3,5-dinitrosalicylic acid
dNTP	Deoxyribonucleotides
DP	Degree of polymerisation
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
Et al.	Et alia (and others)
g	Gram
IPTG	Isopropyl- β -D-thiogalactopyranoside
kDa	kilodalton
K_M	Michaelis constant
Lac	Lactose
LB	Lysogeny broth
M	Molar
MEGA X	Molecular Evolutionary Genetics Analysis X
min	Minutes
µl	Microlitre
ml	Millilitre
mM	Millimolar
MW	Molecular weight

NCBI	National Center for Biotechnology Information
NDP-sugars	Nucleotide diphospho-sugars
ng	Nanogram
OD	Optical density
ONT	Oxford Nanopore Technologies
ORF	Open reading frame
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PHTIII	Pullulan hydrolase type III
rpm	Revolutions per minute
s	Seconds
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
TAE	Tris Acetate EDTA
TEMED	Tetramethylethylenediamine
V	Volts
V_{max}	Maximum rate achieved by the system
UV	Ultraviolet
v/v	Volume per volume
w/v	Weight per volume
× *g*	Times Gravity

Table of Contents

List of Figures

List of Tables

Chapter 1: Literature review

1.1 Introduction

The field of metagenomics offers promising avenues for biotechnological research by unveiling novel enzymes with unique properties, including extreme temperature tolerance and substrate versatility. To date function-based screening has been the preferred approach for the discovery of enzymes with novel characteristics (Ngara & Zhang, 2018). This study aims to investigate the feasibility of employing sequence-based screening identify enzymes with improved characteristics, such as a higher optimal operating temperature.

The literature review will compare sequence-based screening with function-based screening, highlighting their respective strengths and weaknesses. It will demonstrate that despite the advantages of both methods, there remains a considerable pool of unexplored sequences that could be analysed bioinformatically for further research and development. Additionally, this review will discuss the potential role of Pull3.1 as a starch hydrolysing enzyme and emphasize the significance of these enzymes in various applications.

Glycoside hydrolases (GH) have the potential to be used in a wide range of commercial processes (Souza, 2010) and comprise a diverse group of enzymes with significant roles in carbohydrate metabolism, food processing, and various industrial applications. Their ability to cleave α-glycosidic bonds in starch and related substrates makes them essential components of enzymatic systems involved in carbohydrate degradation, modification, and utilisation. Pull3.1 may exhibit characteristics present in the families and subfamilies of GH, hence prompting an exploration of this enzyme family within the context of this literature review.

1.2 Starch

Starch can occur in various forms such as wheat, soybeans, rice corn and other raw starch products (Robyt, 2009). Therefore, the starch industry has mass-produced glucose syrup products to ensure sufficient production of cheap soft drinks and savoury sauces, as well as other products. They can also be added to preservatives and can enhance the appeal of foodstuff products (Marques *et al.*, 2017; Schwartz & Whistler, 2009). Starch is the second most abundant heterogeneous polyose

produced terrestrially (after cellulose). It is stored within the chloroplast, amyloplast and cytosol of plants (Falarz *et al.*, 2018; Hii *et al.*, 2012) (Figure 1.1).

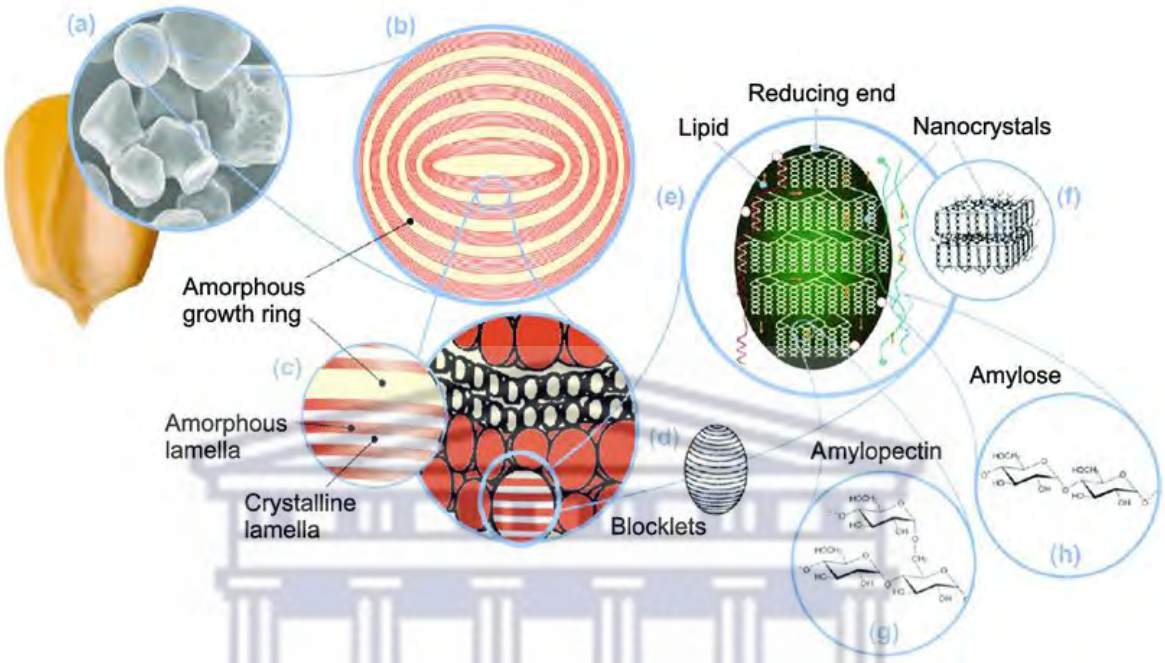


Figure 1. 1: The multi-scale structure of starch. (a) starch molecules in maize; (b) and (c) indicate the amorphous and semicrystalline growth rings and lamellae; (d) displays the blocklets; (e) shows the crystalline lamellae of the blocklet's lamellae that are formed by amylopectin double helixes; (f) starch nanocrystals illustrating the crystalline lamellae when separated by acid hydrolysis; (g) the molecular structure of amylopectin; (h) the molecular structure of amylose. Taken from Božić *et al.*, 2017.

Starch molecules are made up of glucose units linked *via* covalent chemical bonds to form glycosidic linkages. The glycosidic bond is formed between the carbon-1 on one sugar and the carbon-4 of the other sugar. These bonds found in starch are different from those found in cellulose. The location of the hydroxyl group in starch is on the carbon-1 below the glucose ring. This classifies the bond as an α-glycosidic bond (Figure 1.2) (Falarz *et al.*, 2018; Daintith & Martin, 2010). If the hydroxyl group was above the glucose ring, it would be regarded as a β-glycosidic bond, which is the type of bond found in cellulose. These glycosidic bonds are formed through a condensation reaction (Falarz *et al.*, 2018; Hehre, 1969).

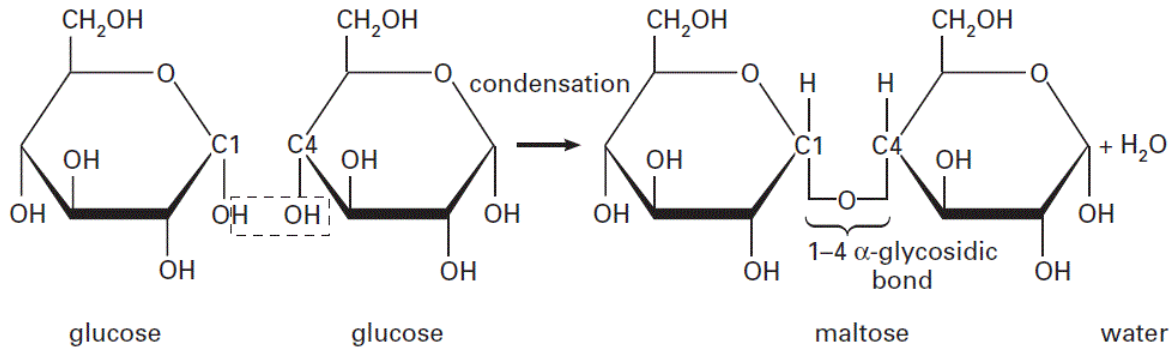


Figure 1. 2: Glycosidic bond formation. A condensation reaction between two glucose units indicating the α-glycosidic bond formation and the end products maltose and H_2O formed by this reaction. Taken from Daintith & Martin (2010).

Starch contains two polysaccharide molecules linked by α-1,6-glycosidic bonds that form amylopectin and α-amylose, as illustrated in Figure 1.3. The botanical origin of starch determines the ratio of amylose to amylopectin, which affects the physical properties of starch (Hii *et al.*, 2012; Ring *et al.*, 1987; Tavallaie *et al.*, 2019).

Amylopectin is considered one of the largest molecules in nature, as it has a polymerisation degree of 2 million (Ganguly *et al.*, 2018; Hii *et al.*, 2012). Amylopectin is highly branched, with short chains that contribute 72-82% of the starch weight. The majority of glucose syrups manufactured from starch hydrolysis contain 75-85% amylopectin. The rest of the weight percentage is attributed to α-amylose. This molecule has long chains and contains fewer branches (Buléon *et al.*, 1998; Lu *et al.*, 2018). The amylopectin branching can be ascribed to the occasional α-1,6-glycosidic linkages that occur at every 20-30 glucose units (Hii *et al.*, 2012; Yasmin *et al.*, 2016).

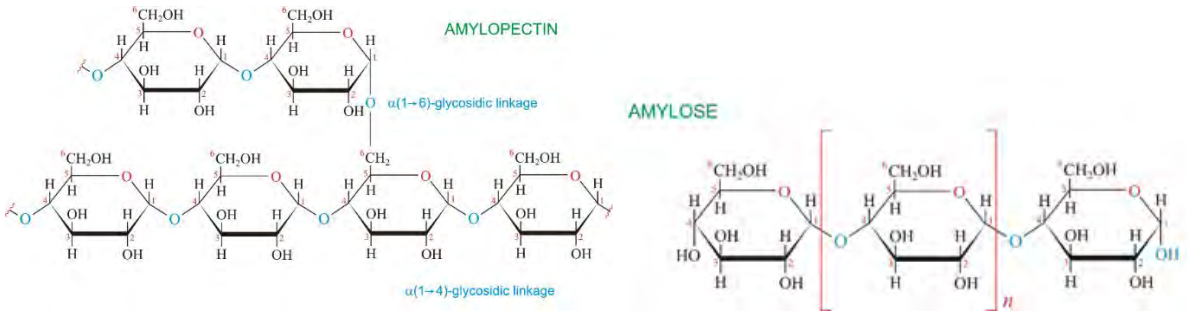


Figure 1. 3: Amylopectin and amylose bonds. The image displays exactly where the α-1,4 and α-1,6-glycosidics bonds are formed within the starch molecule; respectively on amylopectin and amylose. Amylose contains only linear α-1,4-glycosidic bonds, while amylopectin has both α-1,4-linear and α-1,6-branched glycosidic bonds. Taken from Generalic, E. (2022).

Starch is one of the most frequently-consumed carbohydrates in the human diet (Larbey *et al.*, 2019; Ludwig *et al.*, 2018). Because of this, humans have developed the ability to secrete α-amylase enzymes in their saliva and pancreas (Magallanes *et al.*, 2017). Since starch makes up such a considerable portion of consumed biomass, it is to be expected that a large amount of starch-degrading enzymes are readily available, but these enzymes are found in only a few glycoside hydrolase families (Khlestkin *et al.*, 2018; Stam *et al.*, 2006).

Starch hydrolysing enzymes help herbivores and omnivores to digest plant-based material. These starch-degrading enzymes are secreted by animals and plants, and are even produced by archaea in hydrothermal vents (Wang *et al.*, 2008). In comparison to chemical catalysts, microbial enzymes offer better selectivity and can be used under mild to more extreme conditions, making them very interesting biocatalysts. *Geobacillus* species isolated from Manikaran hot springs are the most common source of thermostable alpha-amylase (Far *et al.*, 2020). The alpha-amylase obtained from *B. amyloliquefaciens*, *B. licheniformis*, and *Bacillus stearothermophilus* strains are significant because of their ability to produce large amounts of alpha-amylase (Far *et al.*, 2020). Since being discovered in a wide variety of habitats, these degrading enzymes have made their way into the health, food and biotechnological industries (Garron & Henrissat, 2019).

1.3 Industrial importance of starch degrading enzymes

Degradation of starch can produce oligosaccharides, disaccharides or monosaccharides, depending on which enzymes or processes are used. These products are of value in various sectors of industry because of product demand and monetary value. This has led to a race to find the most suitable, sustainable and cost-effective method of degradation, be it physical, chemical or enzymatic (Khlestkin *et al.*, 2018; Magallanes *et al.*, 2017). These methods are indicated in Figure 1.4. Each method alters the rate at which starch is degraded after being consumed, which could have a significant effect on health (Magallanes *et al.*, 2017). Certain compounds present in foods, such as dietary fibers or polyphenols, can

inhibit or modulate the activity of digestive enzymes. For example, soluble fibers like β-glucans can delay starch digestion by forming viscous gels that impede enzyme access to starch molecules. This delayed digestion can promote satiety, improve insulin sensitivity, and support overall metabolic health (Giuntini, *et al*. 2022).

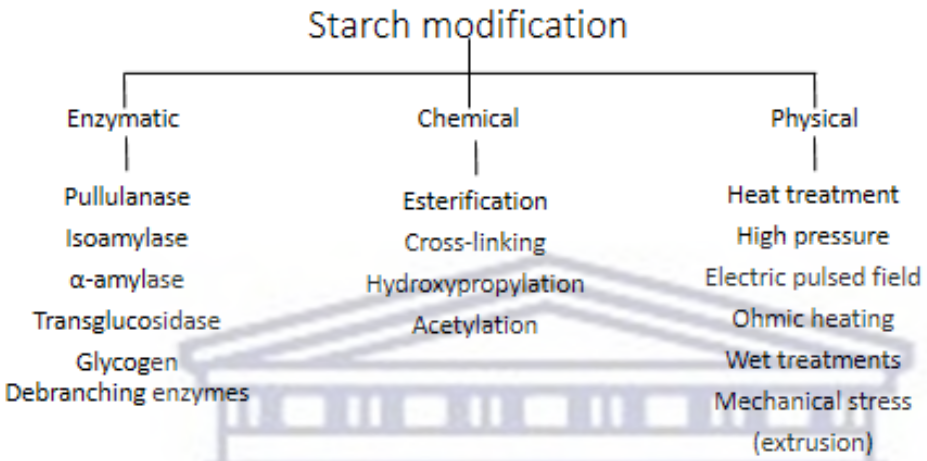


Figure 1. 4: Methods of starch modification. The three common methods used for stach modification and examples of each. Adapted from Magallanes *et al.* (2017).

These three methods of starch degradation have been of use in industries such as pulp and paper, textiles, detergents and pharmaceuticals. Until recently, acid treatment was the preferred method used to break these bonds (Ali & Ahmed, 2016; Hii *et al.*, 2012). Starch molecules are held together *via* glycosidic bonds, which remain stable under alkaline conditions. Now that industries are beginning to rely primarily on enzymatic methods (de Souza & de Oliveira Magalhães, 2010), starch-degrading enzymes make up a large portion of the global enzyme market (Abdelraheem *et al.*, 2019; Sarmiento *et al.*, 2015). Enzymatic hydrolysis processes have many undeniable advantages, as stated by Wilkerson *et al.* (2018), including being more environmentally friendly and having a “high stereo-, chemo-, and regiospecificity”. Enzymes used in hydrolysis reactions are typically derived from renewable sources such as microorganisms, plants, or animals. Additionally, enzymes are biodegradable, meaning they break down naturally into harmless compounds, reducing environmental pollution and toxicity risks associated with chemical catalysts (Choudhury, 2022). Another compelling factor is that proteins can be modified to suit the industrial purpose or requirements (Abdelraheem *et al.*, 2019; Park *et al.*, 2018; Wilkerson *et al.*, 2018).

Amylase, particularly that derived from the *Bacillus* genus, is one of the starch-degrading enzymes widely utilised across various sectors within the starch industry. (de Souza & de Oliveira Magalhães, 2010). The starch conversion process begins with starch being gelatinised. After liquefication, the saccharification step begins (Figure 1.5.). This step usually requires the addition of enzymes such as pullulanase, in order for oligosaccharides or dextrins at the α-1,6-linkages to debranch and form maltose and dextrose syrups. Maltose syrup is often used in semi-moist products, such as pet food or pastry fillings, while dextrose syrup helps prevent crystallisation of sucrose-containing products; such as, jam and table syrup.

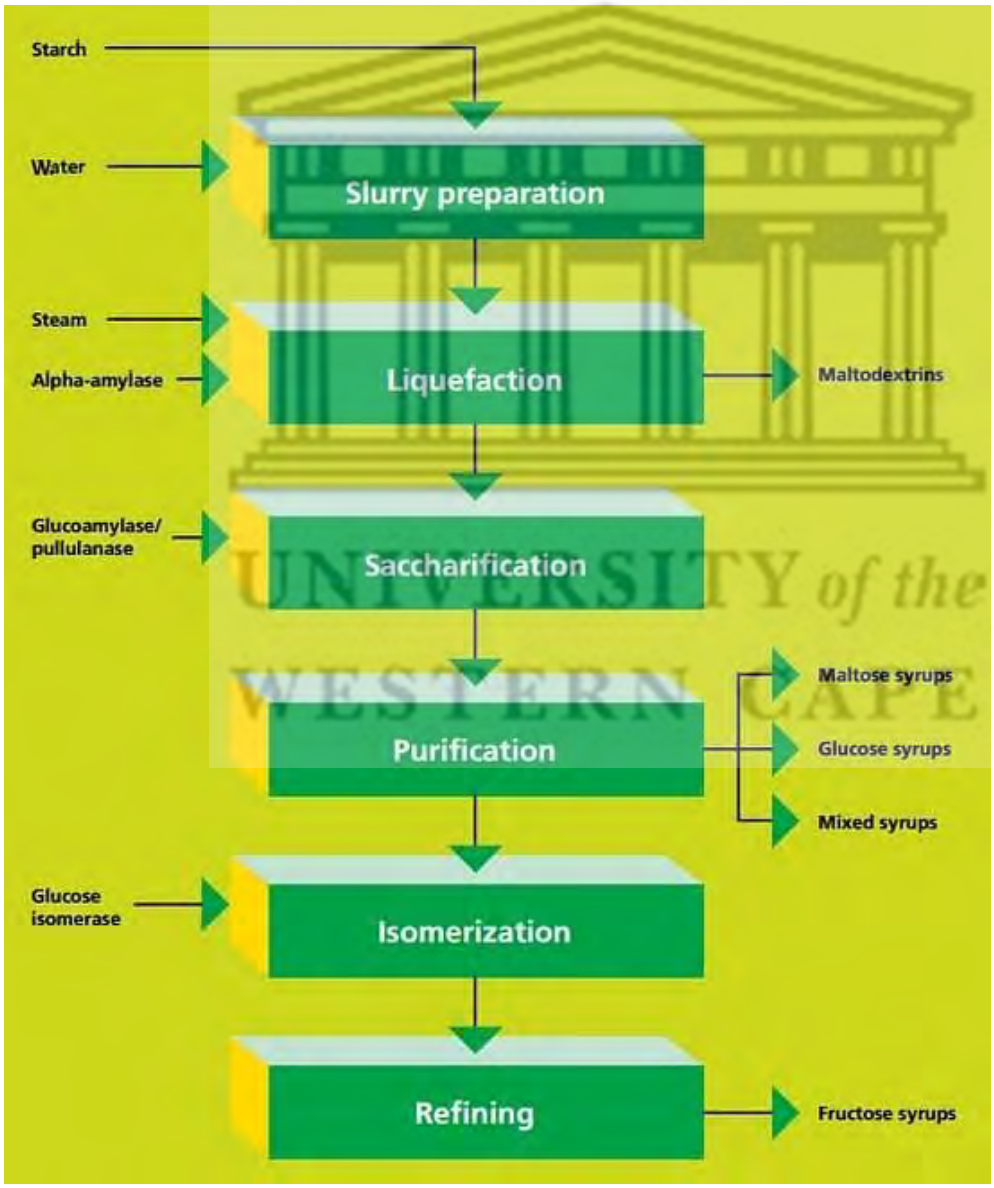


Figure 1. 5: Major steps in enzymatic starch conversion. The main steps involved in transforming raw starch into a viable product for human consumption. The steps mentioned in the figure also indicate the enzymes used during certain steps as well as the end product of each step. Taken from (Novozymes, 2014).

Glucoamylase is added during the saccharification process alongside pullulanase or other enzymes, to produce glucose syrup. The syrup can be further isomerised to form fructose syrup (Fernandes, 2014). The fructose syrup market is enormous, since this product is used as a sweetener in soft drinks, sauces, soups, ice cream and yoghurt (Jòzef, 2007; Parker *et al.*, 2010). It would be beneficial to simplify the steps involved in making these products. Pullulan hydrolase type III (PHTIII) would be a good addition to the starch conversion process, as it would facilitate: liquefication by hydrolysing α-1,4-linkages; hydrolysis of the α-1,6-linkages during saccharification. This would lead to a more efficient process.

Biotechnology and chemical companies, including Novozymes, Du Pont and BASF SE, have shifted their focus from chemical solutions to finding novel enzymes and becoming suppliers of valuable industrial enzymes (Cordeiro *et al.*, 2011; Dalmaso *et al.*, 2015; Roth *et al.*, 2019). These starch converting processes allow for the production of ethanol, maltose, glucose and fructose syrups (Van der Maarel *et al.*, 2002; Nisha & Satyanarayana, 2016).

To increase the efficiency of the starch conversion process, these companies utilise thermozymes (Figure 1.5.), which are able to function at higher temperatures than normal enzymes and are therefore, well suited to processes that entail heated steps (de Souza & de Oliveira Magalhães, 2010; Rosenberg *et al.*, 2006; Vieille *et al.*, 1996).

Additional mechanisms assist in stabilising enzymes to function under the harsh conditions typically encountered in industrial processes. Ion pairs, helix stabilisation and sub-unit assembly are a few of the mechanisms involved in stabilisation, which is a desirable feature in industrial processes (Bruins *et al.*, 2001). Again, this emphasises the need for continued screening to isolate improved enzyme variants either from new environmental samples using tools such as metagenomics, or by screening mutant libraries of characterised enzymes. Table 1.1 gives examples of starch-hydrolysing enzymes that function at very high temperatures.

Table 1. 1: Starch hydrolysing enzymes found in from hyperthermophiles able to function at a high (≥ 60°C) temperature (Niehaus *et al.* 1999)

Enzyme	Organism	Enzyme properties			Reference
		Optimal temperature (°C)	Optimal pH	MW (kDa)	
α-amylase	*Desulfurococcus mucosus*	100	5.5	-	(Canganella *et al.*, 1994)
	Pyrococcus furiosus	100	6.5 ± 7.5	129	(Laderman *et al.*, 1993)
	Pyrococcus woesei	100	5.5	68	(Koch *et al.*, 1991)
	Thermococcus profundus	80	5.5	42	(Young Chul Chung *et al.*, 1995)
	Thermococcus aggregans	100	5.5	-	(Canganella *et al.*, 1994)
Pullulanase type I	*Thermotoga maritima* MSB8	90	6.0	93	(Bibel *et al.*, 1998)
	Thermus caldophilus GK24	75	5.5	65	(Kim *et al.*, 1996)
Pullulanase type II	*Desulfurococcus mucosus*	100	5.0		(Canganella *et al.*, 1994)
	Pyrococcus woesei	100	6.0	90	(Rudiger *et al.*, 1995)
	Pyrococcus furiosus	105	6.0	90	(Dong *et al.*, 1997)
	Thermococcus aggregans	100	6.5	-	(Canganella *et al.*, 1994)
PHTIII	*Thermococcus aggregans*	95	6.5	80	(Niehaus *et al.*, 2000)
	Desulfurococcus mucosus	85	5.0	66	(Duffner *et al.*, 2000)
Cyclomalto-dextrinase	*Thermococcus kodakarensis*	65	7.5		(Sun *et al.*, 2015)
	Microbacterium terrae KNR 9	60	6.0	27.72	(Rajput *et al.*, 2016)

1.4 Hydrolases

Enzyme classes such as α-amylase (EC 3.2.1.1), glucoamylase (EC 3.2.1.3) and pullulanase (EC 3.2.1.41) perform well under saccharification or liquefication conditions. Together, these enzymes can hydrolyse the glycosidic bonds of a homopolymer down to single glucose units. These enzymes (α-amylases, glucoamylase and pullulanase) fall into the hydrolase class, and specifically glycoside hydrolase. The reaction they catalyse falls into the enzyme commission (EC) class 3 category (Toor *et al*. 2021).

Hydrolase enzymes are one of seven main EC classes (Dalkiran *et al.*, 2018; Dalmaso *et al.*, 2015; Kotera *et al.*, 2014). The other enzyme commission classes are EC 1 oxidoreductase, EC 2 transferase, EC 4 lyases, EC 5 isomerases, EC 6 ligases and EC 7 translocases. These classes are curated based on a functional classification system provided by the International Union of Biochemistry and Molecular Biology (IUBMB) Enzyme Nomenclature. The enzyme classes can also be found on the Explorenz database (McDonald *et al.*, 2009).

Enzymes such as α-amylase and pullulanase belong to the glycoside hydrolase EC 3.2.1.x family. Enzymes that belong to this family hydrolyse the glycosidic bonds found in carbohydrates (Dalmaso *et al.*, 2015; Kwan, 2017). Since many organisms have carbohydrates as a central portion of their diet, it is not unusual that the enzymes that aid digestion are found in high abundance throughout the hierarchy of life (Garron & Henrissat, 2019). The Carbohydrate-Active enZyme (CAZy) database was launched in 1999 to help researchers better understand how glycoside hydrolase and other enzymes can be classified to improve predictions of enzyme activity, based on structural similarities (Lombard *et al.*, 2014). There are currently five classes of enzymes to which a carbohydrate-active enzymes belong: glycoside hydrolases (GHs); glycosyltransferases (GTs); polysaccharide lyases (PLs); carbohydrate esterases (CEs); auxiliary activities (AAs) (Lombard *et al.*, 2014). Each of these classes will be discussed in the following sections.

GHs originate from various sources for instance, human saliva and organisms living in hydrothermal vents, and can differ significantly in substrate specificity, even though the enzymes belong to the same family (Bourne & Henrissat, 2001). This is because these families are designated based on amino acid sequence similarity, which translates to the tertiary structure. The three-dimensional structure is usually a good indication of the catalytic mechanism and the active site that the enzymes possess. From this structure, the folds relating to catalytic activity can be identified. The tertiary structure is far more conserved than the amino acid sequence (Bourne & Henrissat, 2001; Garron & Henrissat, 2019). Clans are groups of enzyme families with different primary structures that are placed together based on a highly similar tertiary structure and similarities in folds (Cantarel *et al.*, 2009; Vlasenko *et al.*, 2010).

1.4.1 Starch-converting enzymes

To date, 171 GH families have been listed in clans A to R on the CAZy database (Henrissat & Davies, 1997). Starch-converting enzymes can be found in glycoside hydrolase families 3 ,13, 14, 15, 31, 49, 57, 63, 77, 97, 119, 122, 126, AA13 and GT35 (Akeboshi *et al.*, 2004; Kelly *et al.*, 2016; Møller & Svensson, 2016; Volkov *et al.*, 2015). The enzymes described below were selected to provide a reference point as to the substrate specificities and end products that these enzymes can produce when acting on starch. These examples of enzymes show how a wide variety of enzymes can work in a mixture to produce products that are valuable to industry.

The catalytic mechanisms of these families of enzymes can either retain or invert the anomeric configuration of the saccharide (Vuong & Wilson, 2010) (Figure 1.6). The mechanism is usually well conserved (Angelov *et al.*, 2017; Vlasenko *et al.*, 2010). The retaining or inverting mechanism requires that the saccharides go through an oxocarbenium ion-like transition state during all the steps required. Both mechanisms require a base or acid and nucleophile when performing hydrolysis. The inverting mechanism only requires one step, while the retaining mechanism relies on a two-step double-displacement reaction involving glycosylation (step one) and deglycosylation (step two) (McCarter & Withers, 1994; McIntosh *et al.*, 1996) (Figure 1.6).

Four types of cleavage can occur when starch is hydrolysed into smaller sub-units. This is important, as the conversion of starch into oligosaccharides requires multiple enzymes due to the complexity of the polymer (Hii *et al.*, 2012). The four types are endo-acting, exo-acting, transferase and starch-debranching enzymes (Hii *et al.*, 2012; Van der Maarel *et al.*, 2002) (Figure 1.8).

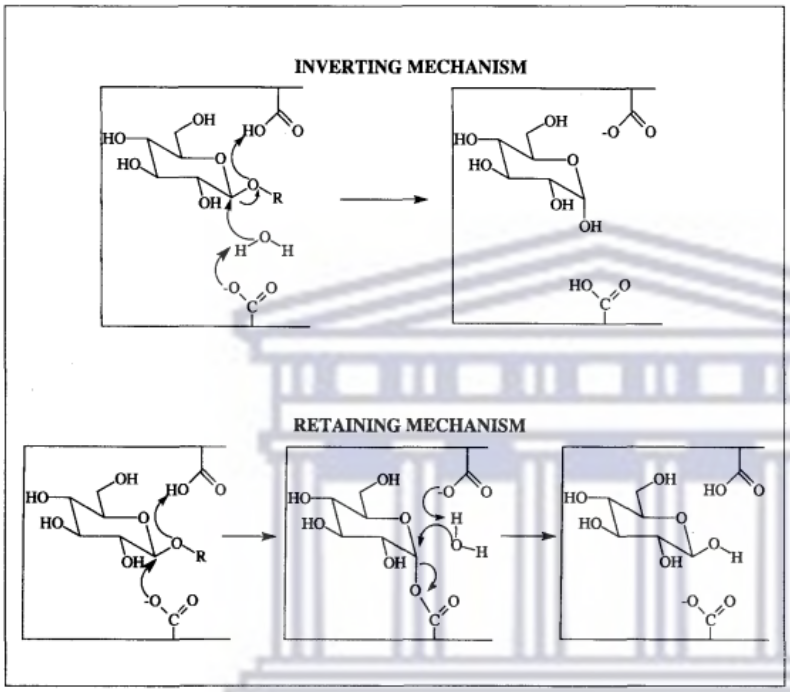


Figure 1. 6: Retaining and inverting mechanisms of glycosidases. The two catalytic mechanisms used by glycoside hydrolases to perform hydrolysis. Taken from (McCarter & Withers, 1994).

1.4.1.1 Amylases

Endo-acting and exo-acting amylase enzymes usually work in combination with debranching enzymes when converting starch to disaccharides or monosaccharides. Endo-acting enzymes such as α-amylase (EC 3.2.1.1) are adapted for cleaving α-1,4-glycosidic bonds at random sites found in the inner parts of the saccharide, as indicated in Figure 1.7 (Hamre & Sørlie, 2020; Van der Maarel *et al.*, 2002). Exo-acting amylase enzymes include glucoamylase (EC 3.2.1.3) and β-amylase (EC 3.2.1.2), which attack at the non-reducing end of the saccharide (Figure 1.9). These enzymes can hydrolyse both α-1,4-glycosidic bonds and α-1,6-glycosidic bonds from the non-reducing end of the saccharide (Esmaeili & Noorolahi, 2017; Frandsen & Svensson, 1998; Halima *et al.*, 2016; Van der Maarel *et al.*, 2002). The endo-acting, exo-acting and debranching enzymes are used to assist starch degradation

processes in industry (Figure 1.9). In terms of amylase enzymes, there are three types: α-amylase, β-amylase and γ-amylase (Yasmin *et al.*, 2016). Amylase enzymes differ from cellulase in terms of the mechanism used to hydrolyse saccharides.

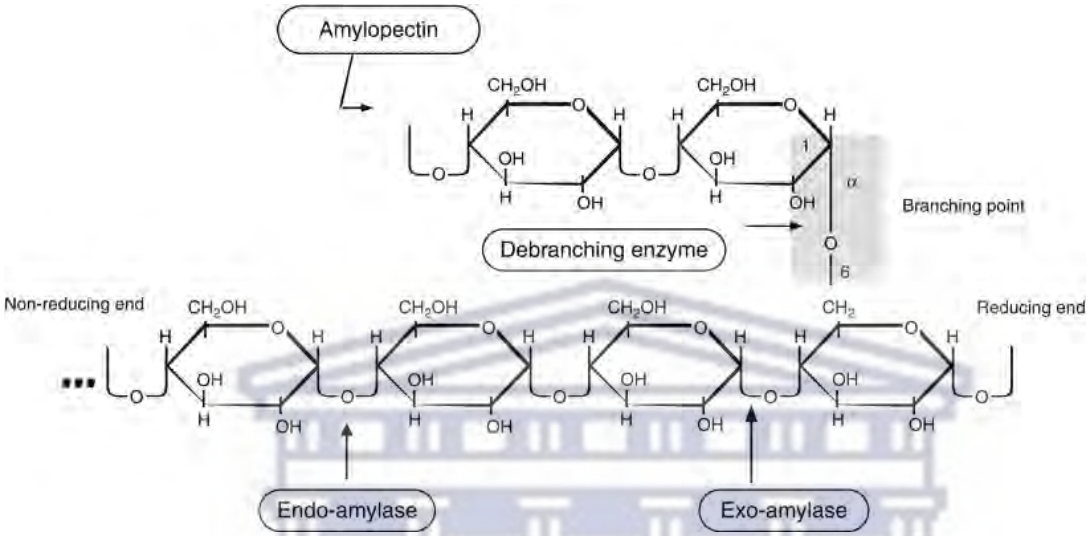
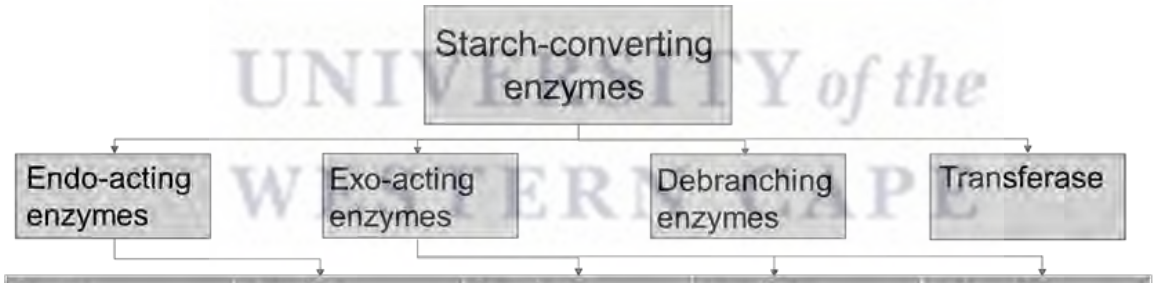


Figure 1. 7: Cleavage sites of the three main starch converting methods. The three main hydrolysis methods displayed in this image are performed by debranching enzymes, endo-amylase and exo-amylase. The image displays that the debranching enzyme will act on the α-1,6-glycosidic bonds. The endo-amylase and exo-amylase will act on the α-1,4-glycosidic bonds, the position where hydrolysis occurs determines the type of amylase. Taken from Rosenberg *et al.* (2006).



Enzyme	α-amylase	β-amylase	γ-amylase	α-glucosidase
Enzyme commission	3.2.1.1	3.2.1.2	3.2.1.3	3.2.1.20
Bonds processes	α-(1,4)	α-(1,4)	Terminal non-reducing end α-(1,4) and α-(1,6)	Terminal non-reducing end α-(1,4) and α-(1,6)
Preferred substrate	Starch, glycogen, polysaccharides, and oligosaccharides	Starch, glycogen, polysaccharides, and oligosaccharides	Starch, amylopectin, and amylose	Starch, amylopectin, and amylose
End products	Glucose, α-limit dextrin, maltose, and linear oligosaccharides	Maltose and β-limit dextrin	Glucose	Glucose

Figure 1. 8: Starch converting enzymes with the focus on endo- and exo-acting enzymes. The enzymes focused on in this figure is the endo-acting α-amylase and the three exo-acting enzymes: β-amylase, γ-amylase and α-glucosidase, information is given about the EC, bonds processed, the preferred substrate and end products. Adapted from **(**Henrissat & Davies, 1997; Niehaus *et al*., 2000; Van der Maarel *et al*., 2002).

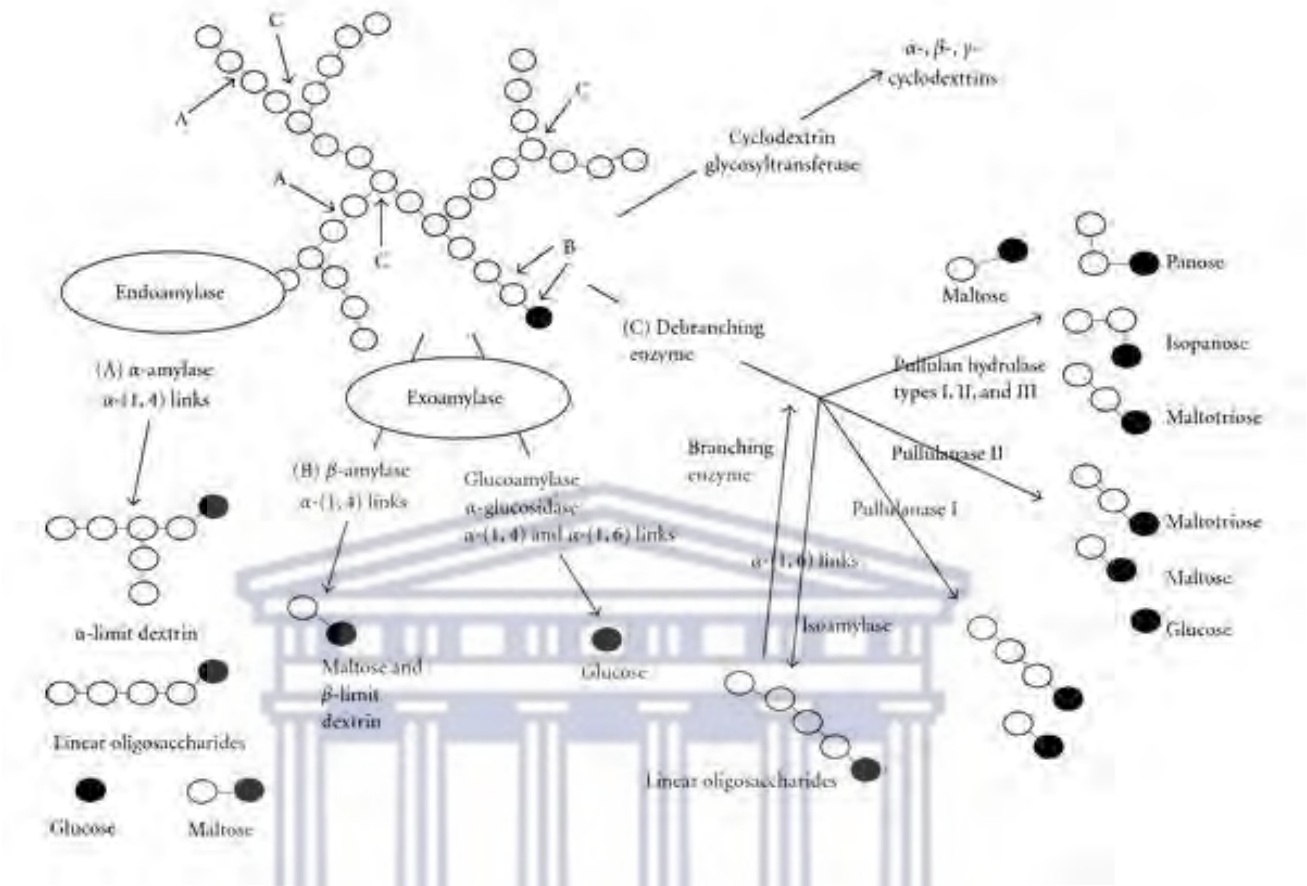


Figure 1. 9: The type of bonds processed by either endoamylase, exoamylase or debranching enzymes and their end products. This figure is and extension of figure 1.8 where the end product of the different endo- and exo-acting enzymes as well as debranching enzymes are displayed. Examples of the three different enzyme types are also given. The black circles indicate the non-reducing ends. Taken from Van der Maarel *et al*. (2002).

1.4.1.2 Endo-acting enzymes

Of all the enzymes involved in starch conversion, the most is known about α-amylase (EC 3.2.1.1), especially the α-amylase enzymes belonging to GH family 13. These enzymes were the first starch-converting enzymes to be discovered in 1811 (Singh *et al.*, 2018). Families in which α-amylase enzymes can also be found are GH57, GH119 and GH126 (Janeček *et al.*, 2014; Janeček & Zámocká, 2019; Zhang *et al.*, 2017; Zona *et al.*, 2004) (Table 1.2). However, there are two families (GH70 and GH77) that contain starch-converting enzymes that belong to clan H - the same clan to which the GH13 α-amylase enzymes belong. However, these two families do not contain any α-amylase enzymes (Miao *et al.*, 2018; de Souza & de Oliveira Magalhães, 2010; Zhang *et al.*, 2017).

Hydrolysis by α-amylase enzymes occurs randomly on α-1,4-glycosidic bonds (Zhang *et al.*, 2017). This makes them perfect for use in saccharification, liquefication, the baking industry and anti-stalling agents (Esmaeili & Noorolahi, 2017; Rosell *et al.*, 2001). α-Amylase hydrolyses starch into maltose, glucose, α-limit dextrin, and branched and unbranched oligosaccharides of various lengths (Khlestkin *et al.*, 2018; Van der Maarel *et al.*, 2002).

α-amylase enzymes used for industrial purposes are usually derived from either bacterial or fungal sources. However, there have been promising signs that archaeal α-amylase from *Chromohalobacter* species and *Halobacillus* species (and others) also satisfy industry needs, especially high-salinity wastewater that contains starch (Prakash *et al.*, 2009; Singh *et al.*, 2018). The α-amylase used for liquefication frequently originates from bacterial sources; for example, *B. licheniformis*, *B. amyloliquefaciens* and *B. stearothermophilus. B. amyloliquefaciens* has been especially valuable in the animal feed industry, where it is used in the liquefying process (Isaksen *et al.*, 2011). α-Amylase obtained from *B. licheniformis* has a half-life 100 times longer than *B. amyloliquefaciens* with which it has an 80% similarity in terms of amino acid sequencing. This α-amylase is also Ca^{2+} independent, acidophilic and extremely thermostable. (Van der Maarel *et al.*, 2002; Wu *et al.*, 2018).

Novel archaeal amylase enzymes have been found in species such as *Pyrococcus furiosus, P. woesei, Thermococcus kodakarensis* and a few other Thermococcale (Janeček & Zámocká, 2019; Lévêque *et al.*, 2000). α-Amylase derived from *P. furiosus* has all the requirements to function optimally during liquefication, but produces the enzyme in inclusion bodies and is consequently considered inappropriate for industry applications. The enzyme remained active at 100°C, is Ca^{2+} independent and had a half-life of 13 hours (Smith & Schindler, 2009). A *Thermococcus* species was developed by BASF - called Fuelzyme® - that could function in acidic environments and maintain functionality at a temperature above 110°C. This made the α-amylase especially suitable for liquefication during ethanol production (Dalmaso *et al.*, 2015). α-Amylase is the preferred source for liquefication; therefore, these endo-acting enzymes are referred to as liquefying enzymes (Robyt, 2009).

1.4.1.3 Exo-acting enzymes

Exo-acting enzymes perform hydrolysis from the nonreducing ends of either amylose or amylopectin (Figure 1.7). Exo-acting enzymes include β-amylase, γ-amylase and α-glucosidase.

1.4.1.3.1 β-amylase and glucoamylase (γ-amylase)

There are two classes of β-amylase enzymes (EC 3.2.1.2) that belong to GH13 and 14 (Table 1.2). One class produces β-maltose and β-limit dextrin, while the other class, glucoamylases (EC 3.2.1.3) produces β-D-glucose. These enzymes are referred to as γ-amylase enzymes (Van der Maarel *et al.*, 2002; Onuma *et al.*, 2018; Robyt, 2009). The β-amylase enzymes that produce maltose as a product are usually found in plants such as barley and wheat, but can also be found in bacteria, especially the *Bacillus* species (Vajravijayan *et al.*, 2018).

Fungi are the primary source of glucoamylase, particularly the *Aspergillus* species. One of the traits of glucoamylase is its ability to act on both α-1,4-glycosidic bonds and α-1,6-glycosidic bonds, which have made glucoamylase invaluable in the production of glucose syrup, since it can completely convert starch to β-D-glucose units (Onuma *et al.*, 2018; Robyt, 2009). Section 1.3 states that, α-amylase enzymes play a major role in many liquefication steps. Liquefication entails the breakdown of large or insoluble molecules into smaller molecules such as dextrins or maltose units, which can be utilised by micro-organisms and further metabolised into glucose units *via* saccharification (Hopkins, 2006). Glucoamylase and pullulanase are responsible for the saccharification step during the starch conversion process (Pang *et al.*, 2019).

β-amylase is only able to hydrolyse from the second α-1,4-glycosidic bond and release a β-maltose. Both γ-amylase and β-amylase are exo-acting amylase. Although most glucoamylase types belong to the GH15 family, some can be found in GH97 (Lee & Paetzel, 2011; Onuma *et al.*, 2018; Robyt, 2009; Volkov *et al.*, 2015). Both these types of amylase use an inverting mechanism during hydrolysis (Lee & Paetzel, 2011; Vajravijayan *et al.*, 2018) (Figure 1.6.).

Table 1. 2: Glycoside hydrolase families of endo-acting and exo-acting enzymes

Enzyme	Enzyme commission	Glycoside hydrolase family
α-amylase	3.2.1.1	13, 57, 119, 126
β-amylase	3.2.1.2	13 and 14
γ-amylase	3.2.1.3	15 and 97
α-glucosidase	3.2.1.20	4, 13, 31, 63, 76, 97 and 122

1.4.1.3.2 α-Glucosidase

Starch-converting enzymes have been incorporated into many products in the food industry, but these enzymes also play an invaluable role in human health in terms of the gut microbiome. The α-glucosidases found in the pioneer genus *Bifidobacterium*, which are the first organisms to colonise the human gut, help prevent lactose (Lac) intolerance, help control the immune response and fight pathogens (Okuyama, 2011).

The exo-acting α-glucosidase is responsible for the high yield of glucose as the main product of starch conversion using this enzyme (Taylor & Dewar, 1994). The α-glucosidases (EC 3.2.1.20) belong to GH4, 13, 31, 63, 97 and 122 (Okuyama, 2011). (Table 1.2). According to the CAZy database, α-glucosidase can also be found in GH76.

1.4.1.4 Debranching enzymes

Starch debranching enzymes are divided into two groups: indirect debranching enzymes and direct debranching enzymes (Fogarty & Kelly, 1990). Indirect debranching enzymes are enzymes such as amylo-1,6-glucosidase (EC 3.2.1.33), which are involved in glycogen hydrolysis. They can only perform hydrolysis after the substrate has been modified by another enzyme so that, for instance, the correct residue is available for hydrolysis. For example, hydrolysis by amylo-1,6-glucosidase can only occur once the transglucosylase has modified the substrate to contain only one glucose residue on the side chain (Bilal & Iqbal, 2019; Fogarty & Kelly, 1990; Hii

et al., 2012). The direct debranching enzymes can immediately perform hydrolysis on α-1,6-glycosidic bonds. These enzymes are listed in Figure 1.8.

There are two types of direct debranching enzymes based on the preferred substrates, namely pullulanase (EC 3.2.1.41) or R-enzymes and isoamylase (EC 3.2.1.68) (Hii *et al.*, 2012; Van der Maarel *et al.*, 2002; Nakamura, 1996). The main distinction between pullulanase and isoamylase is that pullulanase can hydrolyse the α-1,6-glycosidic bond of pullulan and amylopectin, while isoamylase can only hydrolyse the amylopectin α-1,6-glycosidic bond (Van der Maarel *et al.*, 2002). As the enzyme being investigated in this study shows similarity to pullulanase enzymes, these enzymes will be discussed further.

Pullulanase enzymes have a high degree of specificity for pullulan, β-limit dextrin and α-limit dextrin (Nakamura, 1996). There are five types of pullulan hydrolysing enzymes, based on substrate specificity. These are further arranged in two groups: pullulanase enzymes (type I and type II) and pullulan hydrolase type I, II and III (Ahmad *et al.*, 2014; Elleuche *et al.*, 2015; Møller *et al.*, 2016) (Figure 1.10). All pullulanase enzymes fall under the GH13 family, except for isopullulanase, which belongs to GH49. Amylopullulanase (pullulan hydrolases type II) belongs to the GH13 and GH57 families (Akeboshi *et al.*, 2004; Han *et al.*, 2013).

1.4.1.5 Pullulanase type I and II

Both pullulanase type I and II can hydrolyse α-1,6-glycosidic bonds in pullulan to produce maltotriose, but pullulan hydrolase type I (EC 3.2.1.41) can hydrolyse these bonds in amylopectin (Hii *et al.*, 2012; Pang *et al.*, 2019). Pullulanase type II (amylopullulanase) (EC 3.2.1.1/41) can hydrolyse the α-1,4-glycosidic bonds of starch and pullulan, while pullulanase type I cannot (Van der Maarel *et al.*, 2002). Pullulanase type II (amylopullulanase) belongs to the GH57 family. The GH57 family contains enzymes that are mainly derived from extremophiles.

One example of this is amylopullulanase, which is derived from *Geobacillus thermoleovorans* (Noorwez *et al.*, 2006). The thermostability of the enzyme amylopullulanase isolated from *G. thermoleovorans* and α-amylase described in section 1.3. ensures a high temperature optimum (70°C), which makes it useful in a

single step starch-conversion process with PHTIII, as explained in section 1.3 (Nisha & Satyanarayana, 2016; Noorwez *et al.*, 2006; Zona *et al.*, 2004).

Pullulanase is also used in conjunction with glucoamylase or β-amylase in the saccharification process to hydrolyse the α-1,6-glycosidic linkages of pullulan or amylopectin branches, in order to produce maltotriose (Bertoldo *et al.*, 1999; Rüdiger *et al.*, 1995). This product is beneficial as maltotriose is the preferred substrate for glucoamylase. Glucoamylase only has to hydrolyse the remaining α-1,4-glycosidic linkages to produce glucose. Type I pullulanase was successfully isolated from thermophiles by Rüdiger *et al*. (1995), who were the first to produce data on type I pullulanase isolated from the hyperthermophilic archaeon *Pyrococcus woesei* (Rüdiger *et al.*, 1995; Messaoud *et al.*, 2002).

1.4.1.6 Pullulan hydrolase type I, II and III

The pullulan hydrolase enzymes are direct-debranching enzymes. Pullulan hydrolase type I (neopullulanase) and pullulan hydrolase type II (isopullulanase) cannot act on starch as a substrate. Pullulan is the preferred substrate for these two enzymes. Hydrolysis occurs at the α-1,4-glycosidic bonds of pullulan. Type I produces panose, while type II produces isopanose. They can also act on cyclodextrins (CDs) (Ahmad *et al.*, 2014; Hii *et al.*, 2012). Isopullulanase (EC 3.2.1.57) belongs to the GH49 family and is the only pullulan hydrolase to be found in this family. Neopullulanase (EC 3.2.1.135) belongs to the GH13 family (Akeboshi *et al.*, 2004; Labes *et al.*, 2008; Okuyama, 2011) (Table 1.2).

As mentioned in section 1.4.1.5, pullulanase and glucoamylase are invaluable during the saccharification processes. However, there are drawbacks to these enzymes in terms of temperature and pH. Liquefication occurs at a temperature above 100°C, while the second step (saccharification) takes place at 60°C. This temperature change is very costly. For this reason, it is worth trying to find or create thermostable enzymes that can function at a higher temperature (Zhang *et al.*, 2013). Organisms such as *Geobacillus thermocatenulatus* and *Klebsiella variicola* have been screened for enzymes with a higher optimal temperature, but the screenings were not successful.

A breakthrough was made in 2000 when pullulan hydrolase type III (EC 3.2.1.x) was first identified by Niehaus *et al*. (2000). It was regarded as a hyperthermophilic pullulanase, and was seen to show activity at 95°C with an optimal pH of 6.5 (Zhang *et al.*, 2013). The gene was retrieved from *Thermococcus aggregans,* an archaeon, and expressed in *E. coli* (Demirjian *et al.*, 2001; Machovič & Janeček, 2008; Niehaus *et al.*, 2000).

A second type of III pullulan hydrolase was identified in another archaeon, *Thermococcus kodakarensis* KOD1 (Ahmad *et al.*, 2014). This pullulan hydrolase can remain active at temperatures above 100°C, which allows liquefication and saccharification to occur in a single step. This was a significant breakthrough in starch processing (Ahmad *et al.*, 2014; Niehaus *et al.*, 2000), because if PHTIII is used in combination with amylopullulanase, it reduces the number of enzymes required in the starch conversion process, since amylopullulanase can be substituted for α-amylase and glucoamylase (Nisha & Satyanarayana, 2016).

The products produced when PHTIII acts on pullulan are maltose, maltotriose, panose and glucose (Ahmad *et al.*, 2014; Hii *et al.*, 2012; Nisha & Satyanarayana, 2016). When pullulan hydrolyse type III hydrolyses the amylose and amylopectin found in starch, it produces maltotriose and maltose (Hii *et al.*, 2012; Niehaus *et al.*, 2000).

1.4.1.6.1 Isoamylase

Isoamylase (EC 3.2.1.68) belongs to the GH13 family. It acts exclusively on the α-1,6-glycosidic bonds. These debranching enzymes can only act on glycogen and amylopectin's branch points and not on those of pullulan. Since isoamylase prefers a higher molecular weight (MW), these enzymes produce long linear oligosaccharides. This is a disadvantage because other enzymes will also be required to hydrolyse the oligiosacchrides to glucose units. The amylase involved can only produce β-limit dextrins and smaller branched oligosaccharides. Isoamylase was first isolated from yeast and then found in bacteria such as *Pseudomonas* species (Harada *et al.*, 1968). Pullulanase enzymes are better adapted in terms of temperature and pH range for use in the saccharification process (Hii *et al.*, 2012). One issue with industry applications was the low yield of protein and end product, but this has since been investigated. In 2016, the highest yield of isoamylase isolated from

Thermobifida fusca was achieved using the *E. coli* strain MDS42. The study also reported the highest conversion yet of CDs, which led to an increase in the application of isoamylase for industrial starch processing (Ran *et al.*, 2016) The *E. coli* strain MDS42 was a suitable host and the reason for optimal expression of isoamylase.

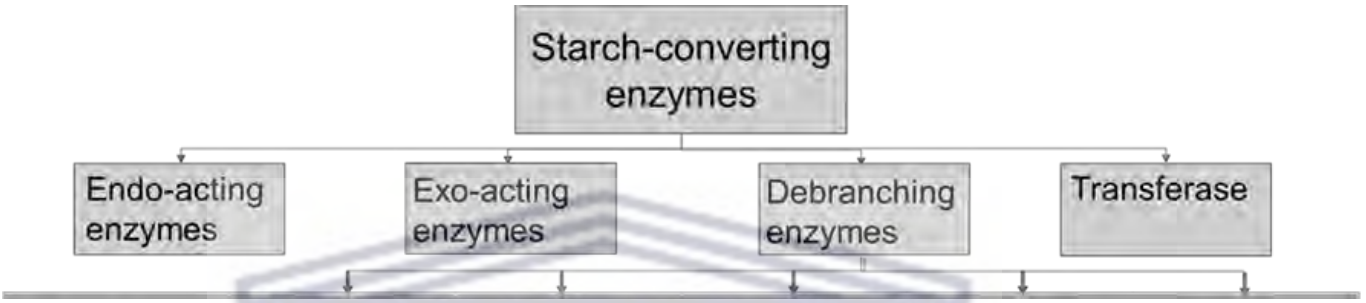
1.4.1.7 Cyclomaltodextrinase

Cyclomaltodextrinase (CDase) enzymes share the GH13_20 sub-family and GH57 with α-amylase and pullulanase enzymes (Figure 1.10). Three enzyme commissions are grouped under CDase; they are referred to as maltogenic amylase (MA; EC 3.2.1.133), CDase (EC 3.2.1.54) and neopullulanase (NPase; EC 3.2.1.135). These enzymes are grouped under CDases because of their ability to hydrolyse all three of the following substrates: cyclomaltodextrins (CDs), starch and pullulan (Park *et al.*, 2000). This distinguishes them from α-amylase, which cannot hydrolyse pullulan. However, on rare occasions, they can hydrolyse CDs very slowly (reference), but not to the extent that the CDases are able to. The only distinguishing factor between CDase and PHTIII is substrate preference: CDase prefers CDs over other substrates; PHTIII prefers pullulan over other substrates.

CDs are enzymatically derived from starch molecules. The first characterised enzyme that was known to act on potato starch came from *Bacillus amylobacter*. These experiments were performed by Antoine Villiers in the late 19th century. He noted the formation of crystals and their specific properties; Antoine had discovered CD molecules. This interest in CDs led to the discovery of the three forms of CDs: α-, β- and γ-CDs (Crini, 2014). CDs have a wide range of applications in the pharmaceutical, food, cosmetic and other industries. This molecule is classified as a molecular cage or carrier (Rasheed *et al.*, 2008). The three types of CDs differ in the number of glucopyranose units that the CD is made up of, i.e. 6, 7 or 8 units for α-, β- and γ-CDs (Crini, 2014). The CDase isolated from *P. furiosus* hydrolysed pullulan to panose, with CDs and starch being converted to maltose.

CDase also performs transglycosylation on oligosaccharides (Park *et al.*, 2000; Yang *et al.*, 2004). This activity occurs when excess glucose is present after a process such as saccharification. The glucose is then converted to branched

oligosaccharides (BOS). BOS have gained significant traction in various industries and have been used in sweeteners. There have even been reports that they support the growth of intestinal bacteria such as *Bifidobacterium* (Tungland, 2018). BOS also delays starch retrograding, which is beneficial in the food industry; therefore, there has been an increase in demand for BOS (Lee *et al.*, 1995).



Enzyme	Pullulanase type 1	Pullulanase type 2 (amylopullulanase)	Pullulan hydrolase type 1 (neopullulanase)	Pullulan hydrolase type 2 (isopullulanase)	Pullulan hydrolase type 3
Enzyme commission	3.2.1.41	3.2.1.41	3.2.1.135	3.2.1.57	3.2.1.41
Glycoside hydrolase family	13 and 57	13 and 57	13	49	13_20 and 57
Bonds processes	α-(1,6)	α-(1,6) of pullulan and α-(1,4) of starch	α-(1,4)	α-(1,4)	α-(1,6) and α-(1,4) of both pullulan and starch
Preferred substrate	Pullulan, poly- and oligiosaccharides	Pullulan and starch the poly-and oligiosaccharides	Pullulan	Pullulan	Pullulan, starch, amylopectin, and amylose
End products	Maltotriose	Pullulan end product = malotriose Starch end product = glucose, maltose and maltotriose	Panose	Isopanose	Pullulan end product = panose, maltose and maltotriose Starch end product = Maltotriose and maltose

Figure 1. 10: Starch converting enzymes with the focus on the direct debranching enzymes. The debranching enzymes listed in the above figure are the five types of pullulanases. The end products, preferred substrates and where they can be found in the glycoside hydrolase family is indicated. Adapted from Henrissat & Davies, 1997; Lombard *et al*. 2014.

1.4.2 Examples of well-characterised α-amylase and PHTIII enzymes

α-Amylases and PTH III enzymes represent two significant classes of glycoside hydrolases with diverse industrial applications. These enzymes play pivotal roles in the hydrolysis of α-glycosidic bonds within starch and related polysaccharides, contributing to processes such as carbohydrate metabolism, food production, and biofuel generation. Well-characterised examples of these enzymes serve as valuable tools in biotechnology, offering insights into their catalytic mechanisms, substrate specificity, and potential applications.

1.4.2.1 Thermostable α-amylase of *B. licheniformis*

α-amylase from the *Bacillus* species have received the most attention among the α-amylase types. The biggest focus has been on the enzyme from *B. licheniformis* because of its thermostability. The presence of calcium, the type of substrate and other stabilizing agents can affect thermostability (Gupta *et al.*, 2003). The high thermostability of *B. licheniformis* α-amylase (BLA) inspired further investigation into what could be achieved with protein engineering. The altering of amino acid composition has shown that particular amino acids are essential in thermostability, while other amino acids could be altered to maintain stability at even higher temperatures (Declerck *et al.*, 2003).

Declerck and colleagues (2003) conducted a thorough investigation on BLA, with all amino acids being tested at substituting side chains for His133 and Ala209. They discovered that Ile and Val were the best stabilising residues at these sites. Mutational analysis directed their attention to the protein region spanning from the variable B domain to a partial section of the central A domain (Declerck *et al*., 2003). BLA inactivation via deamination probed their focus to mutations on Asn/ Gln residues, seven sites were identified, and substitution occurred at positions 172 and 190. With these mutations, they found the single most effective thermostabilising substitution to be Asn190Phe (Declerck *et al.*, 2003).

Thoma *et al*. (1971) found that the catalytic proficiency of α-amylase on substrates is affected by substrate pre-treatment, the amount of substrate branching, swelling and granule size (Thoma *et al.*, 1971). The substrate preference is determined by the binding region. An in-depth study was performed by Kandra *et al*. (2002), who found BLA to have several glycone and aglycone-binding sites in its binding region. They also found a 'barrier' sub-site that allows for dual product specificity and prefers to act on longer substrates (Kandra *et al.,* 2022). The study found that BLA acted on longer maltooligosaccharides and maltohexaose, but struggled to digest maltopentaose and maltotetraose (Kandra *et al.*, 2002).

In general, the catalytic mechanism of the α-amylase family is the retaining double displacement mechanism mentioned in section 1.4.1. This is also the mechanism used by BLA to hydrolyse substrates. The active site uses two of the three conserved catalytic residues: a glutamic acid as an acid/base catalyst and an aspartate as the nucleophile. There are five steps involved in the double displacement mechanism. The first step takes place after the substrate binds to the active site. A proton is donated from the glutamic acid to the glycosidic bond oxygen found between two glucose molecules. Nucleophilic aspartate attacks the C1 of glucose. During step two, an oxocarbonium ion-like transition state is formed, as well as a covalent intermediate state. In step three, the protonated glucose formed during step one leaves the active site, allowing either a water molecule or another glucose molecule to enter the active site. The covalent bond of the glucose molecule and the aspartate from step one is then attacked and another oxocarbonium ion-like transition state is formed as part of step four. In step five, the base catalyst glutamate receives either a proton from water or a new glucose molecule, and the oxygen from the water or the new glucose molecule replaces the oxocarbonium bond between the glucose molecule and aspartate to form a new hydroxyl group at the C1 position of the glucose (Figure 1.11). (Koshland Jr., 1953; Van der Maarel *et al.*, 2002).

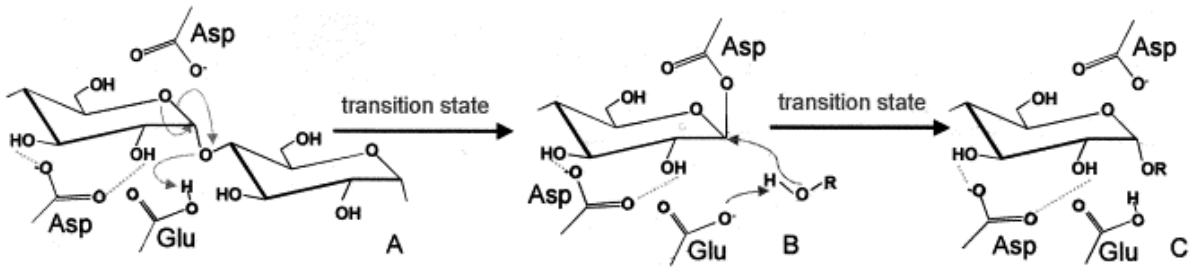


Figure 1. 11: Double displacement mechanism of alpha-amylase involving two sequential steps of substrate binding, nucleophilic attack, and product release. The catalytic mechanism used by α-amylase, the catalytic triad that will be discussed in section 3.4 is displayed and used to indicate how the reaction would occur. Taken from (Van der Maarel *et al*. 2002).

1.4.2.2 Type III pullulan hydrolase from *Thermococcus kodakarensis*

The type III pullulan hydrolase from *T. kodakarensis* was found in 2013. This is a highly thermostable enzyme. According to Guo *et al*. (2018), the thermostability could be due to multiple factors, such as: an increase in the level of the salt bridges; tighter packing; fewer cavities; helical segments; increased Pro, Arg and Tyr residues; and a decrease in the level of serine. *T. kodakarensis* is unique in its ability to convert maltotriose to maltose and glucose. This PHTIII is able to act on α-1,4-glycosidic bonds and branched α-1,6-glycosidic bonds. This substate specificity could be because of the Cys342 and Cys343 disulfide bridge near the calcium binding site. The high electron density at the loop of the α-helix at residues 310–320 and the proline-rich 325–330 could be responsible for the α-1,6-glycosidic bond substrate preference. This assumption is made because a similar region 310-330 is also found in type I pullulan hydrolase from *T. aggregans.* A peripheral interaction between the enzyme and its pullulan substrate is likely to occur as a result of this mechanism (Guo *et al.*, 2018).

Further investigation was performed on aromatic residues of PHTIII to establish what residues are responsible/necessary for catalysis. Treatment of the enzyme with N-bromosuccinimide, which is classically used to oxidise and inactivate tryptophan, tyrosine, histidine and cysteine residue, did affect catalytic efficiency, which indicates that aromatic residue does play a role in catalysis. Type III pullulan hydrolase has the same catalytic mechanism as that described in section 1.4.2.1. (Guo *et al.*, 2018; Koshland Jr., 1953; Van der Maarel *et al.*, 2002).

1.5 Additional classes of enzymes involved in starch hydrolysis

This section is an overview of the four remaining classes of enzymes that are either redox enzymes, perform hydrolysis on carbohydrate esters, form glycosidic bonds or perform non-hydrolytic cleavage. These enzymes differ from starch degrading enzymes since they are not involved in the hydrolysis and/or rearrangement of glycosidic bonds. The additional classes are: glycosyltransferases, polysaccharide lyases, carbohydrate esterases and auxiliary activities.

1.5.1 Glycosyltransferase

Glycosyltransferase (EC 2.4.x.y) and glycoside hydrolase are two of the enzyme classes included in the CAZy database (Campbell *et al.*, 1998). Glycosyltransferase is responsible for the formation of complex carbohydrates. These play a crucial role in energy storage and contribute to the structure of the plants and micro-organisms in which they are found (Campbell *et al.*, 1998). These enzymes are involved in the biosynthesis of saccharides and also act *via* the retaining or inverting mechanism to perform transglycosylation as described in section 1.4.1. The glycosidic bonds are formed by transferring a sugar moiety for an active donor molecule to a specific acceptor molecule (Campbell *et al.*, 1998). This enzyme class is also part of the starch-conversion family, as it contains an enzyme that aids in starch degradation (Rathore *et al.*, 2009).

1.5.2 α-Glucan phosphorylase

α-Glucan phosphorylase (EC 2.4.1.1) belongs to the glycosyltransferase class, but does not contribute to glycosidic bond formation. It does however contribute to the cleavage of glycosidic bonds. This trait places these enzymes in the hexosyltransferase group (EC 2.4.1.x). The enzymes in this group are different from other glycosyltransferase types, as they contain a wide diversity of sugar donors, which are helpful to organisms in extreme environments where nutrient content is low. It has been reported by Mizanur *et al*. (2008) that the α-D-glucose-1-phosphate

synthesised by α-glucan phosphorylases serves as a substrate donor for the synthesis of nucleotide diphospho-sugars (NDP-sugars). These nucleotide sugars are essential components in glycosyltransferase reactions (Mizanur *et al.*, 2008).

α-glucan phosphorylase belongs to the GT35 family and can be found in various organisms. This enzyme acts at the non-reducing end, hydrolysing α-1,4-glycosidic bonds and leaving a highly branched limit dextrin. EC 2.4.1.1 enzymes act on starch, amylopectin, maltodextrin and glucan. The substrate acted upon determines the names of these enzymes, e.g. if an enzyme acts on amylopectin or α-glucan, it is referred to as amylopectin phosphorylase or α-glucan phosphorylase, respectively (Humblot *et al.*, 2014).

The following is an example of starch degradation involving enzymes from the glycosyltransferase class and the glycoside hydrolase class. It shows the involvement of glycosyltransferases in starch degradation. One enzyme was isolated from an archaeon *Archaeoglobus fulgidus* (Labes & Schönheit, 2001). This CD glucanotransferase (EC 2.4.1.19) from GH13 acts on CDs to produce the linearised maltooligodextrins that are converted by maltodextrin phosphorylase (EC 2.4.1.1) to glucose-1-phosphate (Labes & Schönheit, 2001, 2007). A maltodextrin phosphorylase has also been isolated from *P. furiosus* DSM 3638. It was demonstrated that the activity of this enzyme led to the degradation of maltodextrin to glucose-1-phosphate (Lee *et al.*, 2006).

The degradation of starch to the glucose-1-phosphate intermediate is of industrial value; therefore, enzymes capable of achieving degradation are also of value. The and α-glucan phosphorylase has been reported to be functional at a temperature of 70°C. The enzyme referred to was derived from *Thermus caldophilus* GK24 (Bae *et al.*, 2005).

1.5.3 Auxiliary activity

Auxiliary activity is the newest addition to the five classes of enzymes in the CAZy database. This class includes lytic polysaccharide monooxygenase (LPMO) and redox enzymes. The redox enzymes found in this auxiliary activity are usually involved in the degradation of lignocellulosic biomass, while the LMPO perform vital

oxidative activities in the degradation of starch, chitin, xylan, hemicellulose and cellulose (Frommhagen *et al.*, 2015; Levasseur *et al.*, 2013).

1.5.4 Lytic polysaccharide monooxygenase

Lytic polysaccharide monooxygenase (LPMO) can be found in auxiliary activity (AA) 9, 10, 11 and 13 (Frommhagen *et al.*, 2015). The LPMO's AA9 and AA10 were initially classified as part of the GH61 family and CBM33, but these enzymes have since been moved to the auxiliary activity enzymes category as mentioned in the previous section (Lo Leggio *et al.*, 2015; Vu *et al.*, 2014). LPMO's use oxygen as the electron source to increase bond susceptibility to hydrolysis. Oxidation occurs at the carbon-1 and carbon-4 positions of polysaccharides. The enzymes increase the susceptibility of carbohydrates to enzymatic degradation. When LPMO (EC 1.14.99.55) from AA13 - which is the only AA that can perform the starch-degrading activity - is used in conjunction with β-amylase, the activity increases 100-fold (Lo Leggio *et al.*, 2015). Therefore, there has been an increase in demand for these starch-degrading enzymes. A study completed by Lo Leggio *et al.* (2015) found that all the AA13 enzyme entries came from micro-organisms that belong to the phylum Ascomycota, which indicates that the phylum Ascomycota is a leading source of these enzymes.

1.5.5 GH13_20 Subfamily

The GH13_20 subfamily is one of a range of subfamilies that fall under the GH13 family. GH13_20 is a polyspecific sub-family (Liu et al., 2021; Stam et al., 2006) and contains enzymes from a variety of different enzyme commissions (Table 1.3).

Table 1. 3: Subfamily GH13_20 and the enzymes within this subfamily (Stam *et al.*, 2006)

EC Number	Reported enzyme activity
3.2.1.x	Cyclic α-1,6-maltosyl-maltose hydrolase
3.2.1.x	Reducing-end specific α-glucosidase
3.2.1.1	α-Amylase
3.2.1.41	Pullulanase
3.2.1.54	Cyclomaltodextrinase
3.2.1.133	Maltogenic α-amylase
3.2.1.135	Neopullulanase

1.6 Carbohydrate-binding modules of each GH

Carbohydrate-binding modules (CBMs) are used in bioprocessing and remediation, have also been applied as analytical tools (Shoseyov *et al.*, 2006). With bioprocessing, CBMs are used as affinity tags that assist in purifying and recovering active molecules.

CBMs have also yielded benefits in terms of whole-cell immobilisation. This was first demonstrated by attaching a CBM to the cell surface of *E. coli*. The cell was then immobilised onto cellulose (Francisco *et al.*, 1993). The replacement of CBMs is a method used for enzyme engineering, as it aids in creating a higher affinity for the substrate, which results in higher binding and hydrolytic action. This technique was used on an amylolytic strain of *Saccharomyces cerevisiae* that lacked a CBM. The CBM from glucoamylase found in *Aspergillus niger* was incorporated *via* in-frame fusion. The glucoamylase from *S. cerevisiae* could then act on starch (Latorre-García *et al.*, 2005).

For diagnostic purposes, Shoseyov *et al*. (1999) developed a system that is able to detect foodborne pathogens. It employs a CBM bound to a bacterium binding protein, which is loaded onto a cellulose matrix. The example given was a cotton gauze that allowed water to pass through, but captured microbes for further investigation (Shoseyov *et al.*, 2006, 1999).

CBMs are generally associated with GHs that range from cellulose to starch and glycogen degrading enzymes. GH enzymes are usually modular and contain a catalytic site plus one or more non-catalytic sites (Boraston *et al.*, 2004; Oliveira *et al.*, 2015). These non-catalytic sites are usually CBMs. CBMs have been used to search for enzymes that act on starch using a method called “module walking”. The modules are usually indicators of substrate specificity. This is an excellent approach because the catalytic domain is usually attached to the CBM through a linker molecule (Lo Leggio *et al.*, 2015). CBMs increase activity because of their mode of action, i.e., they bind the substrate, which allows the active site and substrate to spend more time in close proximity, thereby improving reaction stoichiometry.

CBMs fulfil three essential functions, in conjunction with the enzyme that they work with. The first is to establish and maintain close proximity of the enzyme to the

substrate it is meant to act on. This enables a higher degree of degradation of the substrate (Boraston *et al.*, 2004). The second key function relies on substrate targeting, and it is evident that CBMs possess the ability to distinguish between key structural differences. This feature permits CBMs to be substrate-specific and so allow a targeted approach during degradation. This function has been applied in studies to identify different types of polysaccharides in the walls of plant cells. The third function is that CBMs disrupt crystalline structures. This increases the amount of underlying substrate available for degradation (Oliveira *et al.*, 2015). These key functions play an invaluable role in protein engineering, as they can increase and alter substrate specificity, which improves enzymatic reactions. Enzyme engineering mechanisms include creating chimeras by combining new related CBMs and active sites (Armenta *et al.*, 2017).

CBMs have signature folds that suggest carbohydrate-binding activity (Oliveira *et al.*, 2015). The tertiary structure and folds of the CBMs are also used as a classification system. This is the same system used to determine which clan each GH family belongs to. However, folds do not predict function (Oliveira *et al.*, 2015). Boraston created a classification system that uses folds, and this has been updated several times (Table 1.4). The β-sandwich is the most common fold seen in CBMs. This fold has two β-sheets that contain three to six antiparallel β-stands on each of the two strands. The fold always has at least one single bound metal ion (Boraston *et al.*, 2004; Oliveira *et al.*, 2015). There are currently 86 CBM families listed in the CAZy database; however, the tertiary structure of these families is only indicated up to CBM family 70.

Table 1. 4: Carbohydrate binding module classification based on the fold (Boraston *et al.*, 2004; Guillén *et al.*, 2010; Hashimoto, 2006; Oliveira *et al.*, 2015)

Fold family	Fold	CBM family
1	β-Sandwich	2, 3, 4, 6, 9, 11, 15, 16, 17, 20, 21, 22, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 40, 41, 42, 44, 48, 47, 51, 70
2	β-Trefoil	13, 42
3	Cysteine knot	1
4	Unique	5, 12
5	OB fold	10
6	Hevein fold	18
7	Unique; contains hevein-like fold	14

Another classification system used for CBMs is based on function and structure, with CBMs divided into three types. Type A are surface binding CBMs, type B contain glycan-chain-binding CBMs, while type C are small-sugar-binding CBMs (Boraston *et al.*, 2004; Oliveira *et al.*, 2015).

Type B CBMs contain grooves with sub-sites. These sub-sites allow for the binding of sugars found on the substrate. Binding efficiency is determined by the degree of polymerisation of the substrates. These types of CBMs prefer a higher degree of polymerisation and therefore bind to hexasaccharides. Type B also prefers binding to glucan chains rather than crystalline structures (Boraston *et al.*, 2004). Table 1.5 was adapted from Armenta *et al.* (2017), Janeček *et al.* (2019) and the CAZy database. It shows the CBM families, GH families and substrate preference. CBM families with references to the ligand-binding function was updated.

Table 1. 5: Carbohydrate binding module families and their ligand binding function

CBM family	GH, (GT, AA) family	Ligand binding function of CBM	References
20	13, 14, 15, 31, 57, 77, 119, AA13	Starch and glycogen	Armenta *et al*. (2017), Janeček *et al*. (2019)
21	13, 15	Starch	Armenta *et al*. (2017), Janeček *et al*. (2019)
25	13, 14, 119	Starch, amylose and amylopectin	Armenta *et al*. (2017), Janeček *et al*. (2019)
26	13, 31	Starch, amylose and amylopectin	Armenta *et al*. (2017), Janeček *et al*. (2019)
34	13	Starch	Armenta *et al*. (2017), Janeček *et al*. (2019)
41	13	Amylose, amylopectin, starch, pullulan and malto-oligosaccharides	Armenta *et al*. (2017), Janeček *et al*. (2019)
45	13	Starch	Armenta *et al*. (2017), Janeček *et al*. (2019)
48	13, GT2	Glycogen and pullulan	Armenta *et al*. (2017), Janeček *et al*. (2019)
53	57, GT5	Starch, amylose, amylopectin and glycogen	Armenta *et al*. (2017), Janeček *et al*. (2019)
58	13	Starch, amylopectin, pullulan and β-cyclodextrin	Armenta *et al*. (2017), Janeček *et al*. (2019)
68	13	Maltotrioses and pullulan	Zeng *et al.*, 2019
69	13	Starch	Peng *et al.*, 2014
74	13	Starch, amylose, amylopectin and Maltose	Valk *et al.*, 2016
82	13	Starch, amylose, amylopectin, pullulan, β-cyclodextrin, glycogen, maltoheptaose	Ramsay *et al.*, 2006; Sreekanth *et al.*, 2016
83	13	Starch, amylose, amylopectin, pullulan, β-cyclodextrin, glycogen, maltoheptaose	Ramsay *et al.*, 2006; Sreekanth *et al.*, 2016

1.7 Metagenomics for discovery of novel glycoside hydrolases

The great plate count anomaly refers to the realisation that many micro-organisms have yet to be identified or cultivated (Staley & Konopka, 1985). It is based on the inconsistency between 16S rRNA genes retrieved from environmental samples and the number of colonies cultured in that same environment.

The factors that limit the growth of micro-organisms include inadequate culture media, temperature limit and the pH range required for the survival of organisms (Handelsman, 2004). Most organisms live in a community of other organisms or in a host system. The conditions present in these systems are difficult to replicate in vitro. Another problem encountered is that the organisms might not produce the same compounds that it does when the organism is discovered in its natural habitat. Pace and colleagues realised this in the 1980s and determined that new means should be used to identify and study micro-organisms since all micro-organisms cannot be cultivated in the laboratory (Handelsman, 2004; Hugenholtz & Tyson, 2008). This led to the birth of metagenomics as a new field or research, which aims to access and gain knowledge on the wider diversity of organisms, and within the context of their communities, that were never able to be studied in depth previously.

Metagenomics has been used for the past few decades as a technique for discovering the concealed treasures that micro-organisms hold (Gupta & Gupta, 2014) and includes techniques for studying and expressing genes from environmental DNA without the need to culture the original host organism (Gilbert & Dupont, 2011; Hugenholtz & Tyson, 2008). We can now find industrially relevant enzymes, and even analyse their genetic as well as metabolic diversity without the need for culturing (Liebl *et al.*, 2014).

Lipase and esterase are starting to replace chemical alternatives in the detergents industry, and amylase and pectinase are becoming critical for clarification, saccharification and maceration in the food and beverage industry (Almeida *et al.*, 2019; Borchert *et al.*, 2017; Castilla *et al.*, 2018; Lewin *et al.*, 2016; Liu & Kokare, 2017; Niu *et al.*, 2018). Environmentally-derived enzymes allow companies to produce the same or better, sustainable products while contributing to a greener future (Castilla *et al.*, 2018). The untapped diversity of microbes can be accessed,

characterised, and quantified using metagenomics to uncover novel enzymes, metabolic pathways, and important products (Culligan *et al.*, 2014). When sampling from an environment that resembles that in which an enzyme would be functioning during industrial use, these techniques allow access and the expression of enzymes with desired properties more suited for industrial application.

Academics have collaborated with companies from various industries (e.g. Syngenta, Genencor International, Degussa and TerraGen) to detect and produce enzymes using metagenomic screening techniques (Lorenz & Eck, 2005). The use of sequence based-screening led to the discovery of the novel pullulanase, PersiPul1. Research conducted by Motahar *et al.* (2020) created a cocktail of enzymes including PersiPul1 that decreases the hardness and chewiness of the bread containing quinoa protein. This mixture of enzymes helped in creating a potential bio-additive for the antioxidant filled bread (Motahar *et al.*, 2022).

The discovery of a novel alkaline-adapted pullulanase PulSS4 was made possible through function-based screening. *Hermetia illucens'* gut microbiota was the subject of the screening; the requirement for these insects to endure -10°C temperatures as well as other unfavourable environmental factors sparked interest in studying the microbiome of these larvae. PulSS4 was able to function at 10°C (Lee *et al.*, 2016).

Handelsman and fellow researchers were the first to create metagenomic libraries and to detect cellulolytic activity from some of the clones (Handelsman, 2004). This provided an indication of the potential that metagenomic screening has to offer, i.e. the possibility of retrieving hidden genes that could potentially deliver new pharmaceutical and industry-altering products (Ekkers *et al.,* 2012). Genes encoding LMPO with improved efficiency and stability have been identified through metagenomics, enhancing the feasibility and sustainability of biofuel production (Frommhagen *et al.*, 2015). Using studies conducted on the genetic diversity of antibiotic resistance genes in environmental samples, metagenomics contributed to the development of novel antimicrobial agents and strategies for combating drug-resistant pathogens (de Castro *et al*, 2014).

Metagenomic libraries need to be screened to identify genes of interest, these genes expressed as enzymes are then tested to see that they meet the specific needs

within the industry. This screening can be performed in two primary ways - either by sequenced-based screening or by functional screening. The latter method was used in the study performed by Handelsman *et al.* (2004) to discover new cellulase activity. To create these libraries, the genomic DNA from environmental samples has to be fragmented and then cloned into a suitable vector. Sampling specifications have, to date, been more standardised to ensure that the total genomic DNA sample is representative of the community in a specific area (Culligan *et al.*, 2014; Gilbert & Dupont, 2011; Thomas *et al.*, 2014).

1.7.1 Function-based screening

Function-based screening provides an opportunity to discover novel genes, pathways and products. This method of screening uses a foreign expression host to produce the enzyme. The enzyme expressed then undergoes activity testing on the substrate of choice, which indicates the enzyme function (Ekkers *et al.*, 2012). Xylanases, cellulases, esterases, lipases and many other classes have been detected using this method of screening (Almeida *et al.*, 2019; Borchert *et al.*, 2017; Ellilä *et al.*, 2019).

Various factors contribute to the success of function-based screening. The correct choice of vector is necessary to successfully express clones at the desired level (Handelsman, 2004). The size and applicable features of the vector should also be taken into consideration for cloning purposes. If the products are toxic, a low number of copies would be desirable, to ensure the host organism survives long enough to express sufficient amounts of the enzyme (Kakirde *et al.*, 2010). Another vector attribute that should be considered is the insert size. The insert size ranges from <15 kb for plasmids and up to 200 kb for bacterial artificial chromosomes (BACs). Within this range there are fosmids (25-45 kb) and cosmids (15-40 kb) to choose from (Ekkers *et al.*, 2012).

The challenges presented by host-dependent systems include toxicity, protein folding and codon use bias. Additionally, only 40% of environmentally isolated genes are expressed in *E. coli,* since the promotors for transcription and the ribosomal binding sites in *E. coli* are too different from that of the organism from which the genes were derived (Culligan *et al.*, 2014; Gabor *et al.*, 2004; Nora *et al.*, 2019).

These challenges can be circumvented by cloning the gene fragments into a shuttle vector in order to also test expression in other hosts (Castilla *et al*., 2018; Dalmaso *et al.*, 2015; Ekkers *et al.*, 2012; Kakirde *et al.*, 2010; Liebl *et al.*, 2014; Nora *et al.*, 2019).

Of the various heterologous expression hosts to choose from *E. coli* strains BL21 and K12 are usually used (Zhang *et al*. 2019). Another frequently used genus is *Streptomyces* – more specifically *S. lividans* SBT5 and *S. avermitillis* SUKA5 – and other genera such as *Bacillus*, *Myxococcus, Rhodobacter* and *Saccharomyces* (Xu & Wright, 2019; Zhang *et al.*, 2019). Having a broader range of expression hosts has proved to be more successful in yielding a greater number of genes being expressed (Craig *et al.*, 2010; Liebl *et al.*, 2014; Mirete *et al.*, 2016). Another strategy used to ensure that genes are transcribed, regardless of orientation, is to create a vector with dual-orientation promoters (Uchiyama & Miyazaki, 2009).

Although there are drawbacks, functional screening allows for all of the gene regions needed for the activity to be retrieved. The advantage of this is that all the enzymes found are definitively active and there is gene expression. Therefore, it could be said that functional screening can lead to the discovery of entirely novel genes without previous knowledge of the gene sequences (Mirete *et al.*, 2016). This was the case for a study conducted by Hu (2010), where a fosmid library was created using the metagenomic DNA from a Malawian hot spring, the same metagenomic library from which Pull3.1 was selected. Using function-based screening a novel thermophilic endoglucanase, XPgene12 from this metagenomic library was discovered. Another advantage to using function-based metagenomics is that these novel sequences usually exhibit lower sequence similarity and enables easier commercialisation since patents should not pose an issue (Kahar *et al.*, 2022).

1.7.2 Sequence-based screening

Sequence-based screening can be seen as a high throughput mechanism for finding new variants of genes that are of specific interest, through comparative analysis against known sequences. If there are similarities covering conserved regions, it could indicate similar biochemical activity. This method has aided the discovery of a plethora of novel gene variants, especially from extreme environments (Kotik, 2009;

Maimanakos *et al.*, 2016). However, there are limitations associated with annotating genes based on conserved regions found on open databases and even the tools used for annotation. A study conducted by Randle-Boggis *et al*. (2016) explained the compromise between a taxonomic resolution and annotation accuracy when setting the parameters involved in gene annotation using various programs that analyse sequence data from metagenomes. The compromise can cause incorrect annotation. This happened with samples collected at a subway station in 2015, when reads detected were annotated to be *Yersinia pestis* (Afshinnekoo *et al.*, 2015). On further inspection, it was found that reads could also have mapped closely to any other bacterial species if the sequences were compared against a more extensive dataset. The information on the reads was also sourced from institutes that did not study *Y. pestis*. The researchers also mentioned that they believed the genus and species of the sequences would be determined to be something completely different if the sequences were to be compared to a better-annotated database that contains complete genomes (Afshinnekoo *et al.*, 2015). Another drawback is that sequence-based screening does not guarantee retrieval of a complete gene, as is the case with functional screening (Yun & Ryu, 2005).

However, there are many advantages to using sequence-based screening, including that there is no need for a suitable host to be able to detect novel genes (Escobar-Zepeda *et al.*, 2015). Table 1.6 summarises a few novel enzymes with interesting attributes that have been discovered in recent years.

Table 1. 6: Enzymes discovered using sequence-based screening

Enzyme	Source	Reference
Acetyl xylan esterase	Hot desert hypolith	Adesioye *et al.*, 2018
Amine transferase	Hot spring	Ferrandi *et al.*, 2017
β-Galactosidase	Hot spring water	Liu *et al.*, 2015
β-Galactosidase	Dairy industry stabilisation ponds	Eberhardt *et al.*, 2021
β-Xylosidase	Hot spring soil	Sato *et al.*, 2017
α-L arabinofuranosidase, β-glucosidase, β-xylosidase,	*Erethizon dorsatum* (porcupine) microbiome	Thornbury *et al.*, 2019

and endo-1,4-β-xylanase		
Bifunctional cellulose/ hemicellulose	Black goat rumen	Lee *et al.*, 2018
Bifunctional cellulase/xylanase	Cow rumen microbiota	Ariaeenejad *et al.*, 2021
Bilirubin-oxidizing enzyme	Wastewater treatment activated sludge	Kimura & Kamagata, 2016
Carbohydrate-active enzyme	Antarctic tundra soil	Oh *et al.*, 2019
Carbohydrate-active enzyme	Camel and goat intestinal tract	Al-Masaudi *et al.*, 2019
Carbonic anhydrase	Active hydrothermal vent chimney	Fredslund *et al.*, 2018
Cellulases	Oil reservoir	Lewin *et al.*, 2017
Cytochrome P450	Hot spring	Nguyen *et al.*, 2020
Endoglucanase	Hot spring sediment	Zhao *et al.*, 2017
Epoxide hydrolase	Hot spring	Ferrandi *et al.*, 2015
Esterase	Heated compost and hot spring	Leis *et al.*, 2015
Esterase	Hot spring mud	Zarafeta *et al.*, 2016
Glycoside hydrolase GH16 SCLam	Sugarcane soil metagenome	Liberato *et al.*, 2021
Hydrolase	Hot environments	Wohlgemuth *et al.*, 2018
Lignocellulose-degrading enzyme	Gut microbiome of the common black slug *Arion ater*	Joynson *et al.*, 2017
Pectinase	Apple pomace-adapted compost	Zhou *et al.*, 2017
Tauroursodeoxycholic acid biotransformation enzymes	Gut microbiome of black bears	Song *et al.*, 2017
Toluene monooxygenase	Hydrocarbon-polluted sediment	Bouhajja *et al.*, 2017

Many enzymes with novel or bifunctional properties have been discovered using the sequence-based screening method (Madhavan *et al.*, 2017). These findings may be intentional and produced via computational pipelines and multi-stage in-silico screening. Substrate specific searches on samples isolated from Antarctic tundra soil yielded enzymes capable of functioning at very low temperatures. This enzymes' functional adaption displays the advantages of doing a function-specific search using the sequence-based screening approach (Oh *et al.*, 2019). In a research study on the metabolic potential of organisms isolated from the sulphur chimneys, an amylase from the Mothra hydrothermal vent with an optimum temperature of 90°C was discovered (Wang *et al.*, 2011).

Bifunctional screening using computational pipelines has helped overcome obstacles encountered when researchers have tried enzyme fusion or creating enzyme cocktails. The purification, production and loading costs are reduced when applied to a bifunctional enzyme instead of a single enzyme. The bifunctional enzyme discovered by Ariaeenejad *et al*. (2021) was a novel cellulase/xylanase. The benefits of this enzyme are that lignocellulosic biomass is more accessible when the xylan layer on the surface of the lignocellulose has been degraded. The mode of action performed by the cellulase/xylanase bifunctional enzyme increases the effectiveness of hydrolysis. This increase in hydrolysis performance can be tailored by using a computational pipeline that explicitly searches for bifunctional enzymes. Therefore, sequence-based screening using computational pipelines is a potent tool when applied correctly (Ariaeenejad *et al.*, 2021).

1.7.3 Computational workflow for sequence-based screening for novel enzymes

Sequencing plays a pivotal role in the discovery of novel enzymes by unveiling the genetic blueprints of diverse organisms, providing insights into their metabolic potential. Sequencing enables researchers to identify genes encoding enzymes with unique catalytic activities, facilitating the exploration of new biocatalysts for various industrial applications. Sequencing can be conducted using either short-read or long-read technologies, and it can also be executed simultaneously, integrating both approaches.

Sequencing technologies from Pacific Biosciences and Oxford Nanopore Technologies (ONT) offer longer reads that provide better genome assemblies from environmental samples and facilitate the reconstruction of complete pathways, but have lower throughput than second-generation sequencers (Damme *et al*., in press; Weirather *et al*., 2017). Short read length technologies on the other hand, such as Illumina, have scaled their sequencers, which have lower throughput but are more cost-effective and so serves the needs of smaller sequencing laboratories (Haynes *et al*., 2019; Heather & Chain, 2016; Vollmers *et al*., 2017). Short reads are well-suited for detecting single genes due to their high accuracy and coverage and is suitable for sequencing metagenomic libraries (Buermans & Den Dunnen, 2014). Therefore, sequencers are selected based on read length, the number of reads, run time, cost per sequencer, throughput and the aim of the project (Buermans & Den Dunnen, 2014).

The workflow for shotgun sequencing and screening, from DNA extraction to the bioinformatic analysis steps is displayed in Figure 1.12. This workflow facilitates metagenome assembled genomes (MAGS) and can generate a picture of the type of organisms and the genes found in the sample environment. The various steps will be elaborated on below. The metagenomic library of relevance to this project was sequenced using the Illumina MiSeq, therefore more focus will be given to the workflow for this technology.

The quality of the assemblies relies on the correct genome fragmentation. There are many variations of fragmentation, including chemical fragmentation, physical fragmentation and enzymatic digestion (Solonenko & Sullivan, 2013). With the Miseq platform, transposon-based shearing (tagmentation) is the preferred method of DNA fragmentation. Tagmentation via transposon mutagenesis allows for dsDNA to be sheared at random sites, while undergoing incorporation of the adapters necessary for binding to adapter-specific primers on the flow cell. Once adapters are added, they are amplified via PCR (polymerase chain reaction) and purified (Divoll *et al.*, 2018; Volpicella *et al.*, 2012). These PCR fragments are diluted and then added to the flow cell. Multiple copies of each attached fragment are formed on the flow cell via bridge amplification, which creates clusters of the sequences (Bruijns *et al.*, 2018; Dickens & Graham, 2002; Heather & Chain, 2016). As each nucleotide is added, a fluorescent signal is detected that corresponds to one of the four ddNTPs

(Gupta & Gupta, 2014; Haynes *et al.*, 2019). The captured signal allows for the sequence to be determined and picked up by the sequencer for each fragment of DNA (Bruijns *et al.*, 2018).

The data generated from the sequencer can then be used to assemble the genomes sampled. As data increases, so does the need to process it, in order to prevent bottlenecks from occurring (Haynes *et al.*, 2019; Vollmers *et al.*, 2017). The assembly process requires specific computational steps, including the alignment of sequencing reads. The assembly process can either be a reference-based or a de novo assembly process (Dhar *et al.*, 2019). The choice of sequencer also plays a significant role at this stage, since longer reads give an added advantage when doing de novo assembly (Dhar *et al.*, 2019; Khan *et al.*, 2018), especially as the quality of base calling improves on platforms such as ONT.

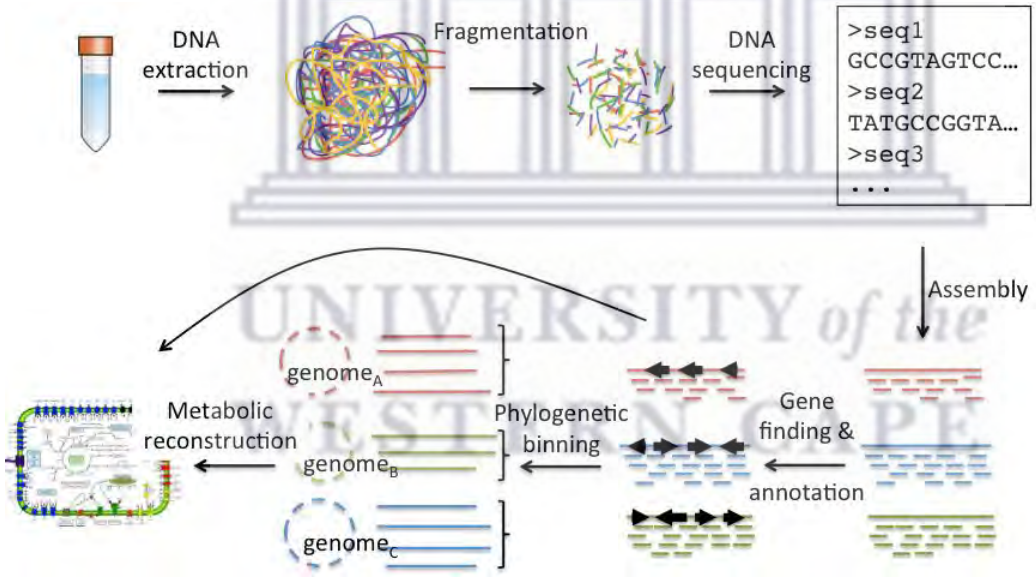


Figure 1. 12: The metagenomic sequence-based screening workflow. The steps involve DNA extraction followed by DNA sequencing. The gene sequences are assembled and annotated. Following this process, phylogenetic binning will occur, and assembled genomes can be determined. Taken from Envgen (https://envgen.github.io/metagenomics.html).

With assemblers such as Newbler, the software will recognise the overlapping regions and start forming contigs that do not contain any gaps (Ghurye & Pop, 2019; Qiu, 2019). The contigs and the gaps between them are paired together to create scaffolds, based on characteristics such as its likely position and orientation when found in a genome.

Binning is the next step in sorting through reads. Binning should allow differentiation between genomes and identification of which taxon the organisms probably belong to. So binning is a crucial step. The original software available is for either long reads or short reads. However, new tools such as MetaProb and MUFFIN combine the original tools, which enables short reads or long reads to be processed in combination. A disadvantage of sequence-based screening that arises is that binning relies on a reference database, but the databases cannot possibly contain all genomes. This means they could skew the data (Damme *et al.*, in press; Girotto *et al.*, 2016).

Gene calling for putative open reading frames (ORF) can be performed for known homologs by using BLAST. However, other algorithms for gene calling of novel ORFs are also available (Wooley *et al.*, 2010). Functional annotation can be performed to understand better the functional opportunity that the sampled microbial system offers. This is especially valuable when screening for a specific enzyme or family of enzymes. Conserved regions and sequence signatures are indicators of sequence function.

Binning and functional annotation usually correlate with the type of environment sampled, which could be a good reference point to compare results generated with metagenomes previously found in similar environments. Sampling in an extreme environment would generate frequent findings of extremophile genomes, which could open the door to the functional capacity of these genomes (Cowan *et al.*, 2004). Sequence-based screening offers the means to discover novel modes of action and discover enzymes with enhanced features that might be invaluable for industry use.

1.8 Aim and objectives of the study

Problem Statement and Aim of the Study

The discovery of novel variants of genes through sequence-based screening has led to the worthwhile discovery of enzymes functioning at the preferred pH, temperatures and on substrates that are used in industry. However, the potential of this approach for the discovery of novel biocatalysts has been not yet been fully explored. A fosmid library has previously been prepared from a hot spring and was next generation sequenced. Through sequence analysis several ORFs with moderate to low sequence similarity to glycosyl hydrolases were identified. Given the interest in discovering novel thermozymes with greater functional capacity and tolerances, these constituted the start of a study aiming to characterise the enzymes encoded by these ORFs and to assess whether they presented thermostable properties.

This study specifically aimed to characterise a putative pullulanase, Pull3.1. This enzyme was selected because of the possibility that it could form part of a novel group of enzymes (PHTIII) and function at a high temperature. This was prompted by the successful characterisation of other thermophilic glycoside hydrolases from this metagenomic library. PHTIII enzymes have potential for use in the food processing industry, specifically in terms of liquefication and saccharification, therefore Pull3.1 will be characterised.

Objectives

The objectives for this study were to:

Perform bioinformatic analysis of the Pull3.1 sequence.

Recombinantly express and purify Pull3.1.

Characterise Pull3.1 to determine temperature optimum, pH optimum, thermostability, substrate preference, effect of inhibitors, and kinetic behavior.

Chapter 2: Materials and methods

2.1 Bioinformatic analysis

Contig 492, assembled by Hu (2010) was selected for analysis in this thesis. Among the ORFs within this contig, ORF8 exhibited promising characteristics for functioning as a starch-converting enzyme. Pull3.1 (ORF8) was selected and annotated using KBase (Allen *et al*. 2017). The ORF nucleotide sequences used for further analysis were selected based on similarity to other sequences on the NCBInr database using the Standard Nucleotide Basic Local Alignment Search Tool (BLASTn) (Altschul *et al.*, 1990). The nucleotide sequence was translated to an amino acid sequence using ExPASY. Related proteins, including conserved domains, were identified by doing BLASTp, Pfam and InterProscan searches (Altschul *et al.*, 1990; Gasteiger, *et al*., 2005; Jones *et al.*, 2014; Mistry *et al.*, 2021). ClustalW was used to perform multiple sequence alignments (MSA). The secondary structure was predicted using PSIPRED (Jones *et al.*, 1999) and the 3D structure used for homology modelling by SWISS-MODEL (Swissmodel.expasy; Buchan, 2013; Waterhouse *et al.*, 2018).

The features on the 3D structure and the superimposed image were generated using PyMol (Schrodinger, 2015). The MEGA X program (Megasoftware) was used to construct the maximum-likelihood tree (Tamura *et al*. 2021), with the protein sequences retrieved from BLASTp using similarity searches against sequences on the IMG/M and CAZy databases (Altschul *et al.*, 1990; Chen *et al.*, 2021; Henrissat & Davies, 1997). The maximum-likelihood tree was constructed using the JTT matrix-based model with a bootstrap value of 1000. The Bioedit alignment program was used to produce the visualisation of the MSA (Hall TA.1999).

2.2 Chemicals and reagents

The chemicals used in this study were supplied by: Merck Chemicals and Laboratory Supplies (Darmstadt, Germany); Sigma Aldrich Chemical Company (Missouri, United States); Kimix Chemical and Laboratory Supplies (South Africa). Culture media was supplied by Oxoid Ltd (Hampshire, England) and New England Biolabs (Massachusetts, USA). The primers used for the PCR reactions were synthesised by Inqaba Biotech (Johannesburg, South Africa). The enzymes needed for restriction digestion of DNA samples, Phusion DNA polymerase and DNA, and the protein MW markers were from Fermentas Life Sciences Ltd, ThermoFisher (South Africa) and New England Biolabs (Massachusetts, USA) respectively.

2.3 Bacterial strains, polymerase chain reaction and culturing

The *E. coli* strains used are indicated in table 2.1 and were grown in LB (Lysogeny broth) medium Sambrook & Russell (2001). The LB medium contained (per liter) 10 g of tryptone, 5 g of yeast extract and 10 g of NaCl. If required, 100 mg/ml ampicillin (Amp) was added to reach a final concentration of 100 µg/ml. For solid media, agar was added at a concentration of 15 g per liter. All media was autoclaved at 121°C for

20 min, unless stated otherwise. Strains were inoculated into broth using aseptic technique. Unless otherwise stated, cultures were incubated at 37ºC, and for liquid cultures accompanied by agitation at 150-225 rpm. *E. coli* clones containing pET-21a (+) (Table 2.1) were routinely cultured on LB medium, or in LB broth at 37°C overnight, supplemented with Amp to perform selection of clones containing the Pull3.1 gene.

Table 2. 1: Bacterial strains and plasmids

Strains and plasmids	Standard features or genotype	Source or reference
Strains		
E. coli BL21	*F-ompThsdSB(rB-mB-) gal dcm gal λ(DE3)*	Invitrogen, USA
Plasmids		
pET21a (+)	Expression vector with a C- terminal His-tag, Amp^R, T7 promoter and terminator, MCS	Novagen, USA
pETPull3.1	pET21a containing the Pull3.1 gene cloned into the NdeI (CATATG) and NotI (GCGGCCGC) restriction sites (Appendix, Figure A). This construct was synthesised by Biomatik (Figure 2.1)	Biomatik

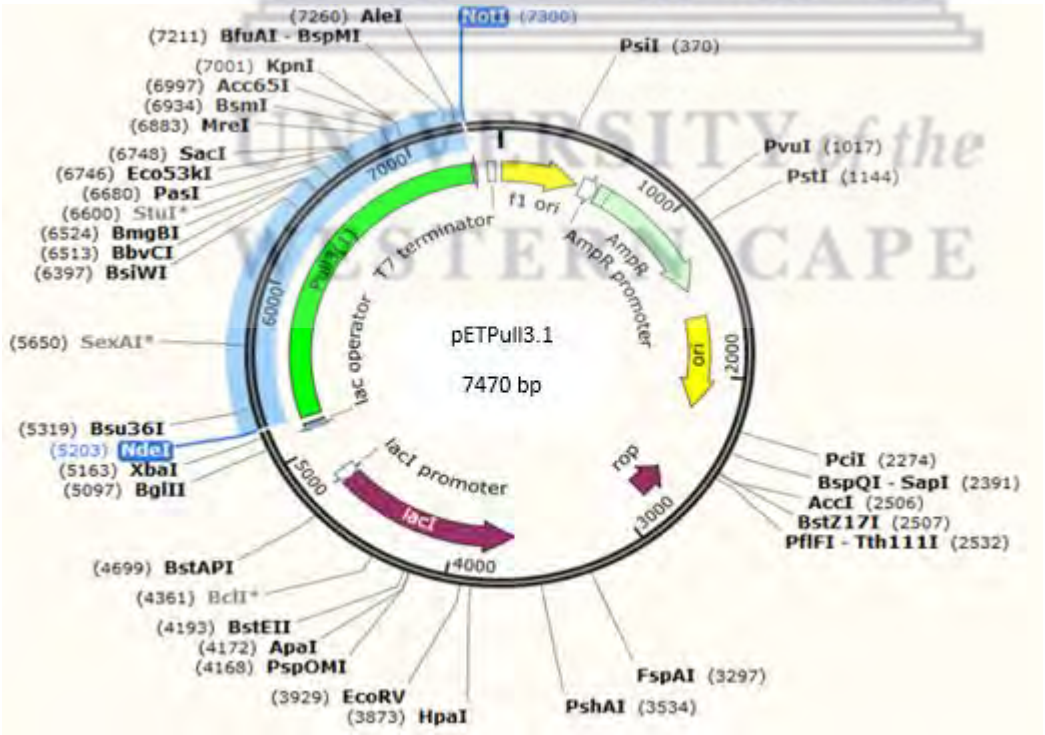


Figure 2. 1: A map of the Pull3.1 gene cloned in the NotI and NdeI sites in the pET21a+ expression vector indicating the C-terminal His-tag

2.3.1 PCR for purified DNA templates

Polymerase chain reaction (PCR) amplifications were set up in 0.2 ml thin-walled PCR tubes. Each PCR reaction contained 1 × Phusion buffer or GC buffer, 2 mM dNTPs (deoxyribonucleotides) (dATP, dCTP, dTTP, dGTP), 10 µM each of the forward and reverse primer, 1 U of Phusion DNA polymerase or *Taq* DNA polymerase and 1 ng of template DNA, pETPull3.1. The final volume was adjusted to 50 µl with nuclease-free water. The PCR amplification took place in a thermocycler (T100 thermal cycler Bio-Rad). The primers and thermocycling parameters are detailed in Table 2.2 (McPherson & Møller, 2000).

2.3.2 Colony PCR

Colony PCR was performed to ensure that the correct target sequence was within the synthesised gene construct (Figure 2.1, Appendix, Figure A). A colony was picked with a clean pipette tip and inoculated into a 0.2 ml thin-walled PCR tube containing 20 µl ultrapure water. The tube was vortexed and heated for 10 min at 95°C in a thermocycler (T100 thermal cycler Bio-Rad). After 10 min, the tube was centrifuged for 10 min at 10 000 × *g*. Four µl of the supernatant was used as a template, and PCR reaction continued as detailed in Table 2.2.

Table 2. 2: Primers and thermocycling conditions

Primer	Target sequence	Sequence	Size (bp)	Thermocycling parameters
Pull3.1-1392-F	PULL3-cont4921	5'-TATACATATGTTG GTGGTCCTCTGGGCC-3'	2105	Initial denaturation 98°C for 90 s. 30 cycles of 98°C for 30 s, 55°C for 30 s and 72°C for 1 min. The final extension took place for 5 min at 72°C.
Pull3.1-1392-R		5'-TGCTCGAGTGCGGCCGCTTA-3'		

2.3.3 Agarose gel electrophoresis

Electrophoresis was used to separate plasmid DNA and PCR products and to assess the quality thereof. A 1% (w/v) agarose gel was made using 1 × TAE buffer (40 mM Tris-base, 1 mM EDTA, 20 mM glacial acetic acid) with 0.5 µg/ml of ethidium bromide. The solution was heated until melted and poured into a tray kept together by casting clamps, a comb was also inserted (Sambrook & Russell, 2001). Once solidified the comb was removed and placed in an electrophoresis tank filled with 1 × TAE buffer. DNA samples (5 µl) were each mixed with 1 µl of 6 × loading dye (30% (v/v) glycerol, 5% (w/v) bromophenol blue). The samples were then added to a well alongside a 1kb DNA MW marker. The gels were electrophoresed at 90 V for 90 min and were then visualised using an AlphaImager™ HP 2000 (Alpha Innotech, USA).

2.3.4 Preparation of chemically competent *E. coli* cells

E. coli BL21(DE3) was streaked out on LB agar and incubated overnight at 37°C to allow it to form colonies. A preculture of LB broth was prepared. A single colony of *E. coli* BL21(DE3) was added to McCartney bottles containing 5 ml LB broth, then incubated at 37°C overnight, with shaking at 170 rpm. A volume of 0.5 ml of the overnight starter culture was inoculated into a 250 ml Erlenmeyer flask containing 50 ml prewarmed LB broth on a shaker set at 170 rpm at 37°C for approximately 3 hours. The flask was placed on ice. The cells were kept at 4°C from this point on. Cells were collected by centrifuging at 964 × *g* for 5 min when the OD_{600} reached 0.6. The supernatant was decanted, and 40 ml of ice-cold $MgCl_2$ was added. The tube was incubated on ice for an hour, thereafter the tube was centrifuged again for 5 min at 964 × *g.* The supernatant was then decanted, and the pellet was resuspended in 1 ml of 100 mM $CaCl_2$ /10% glycerol. Aliquots of the cells (50 µl) were transferred to 1.5 ml Eppendorf tubes and stored at -80°C.

2.3.5 Transformation of *E. coli* BL21(DE3) cells

Aliquots of *E. coli* BL21(DE3) cells stored at -80°C were thawed on ice. The DNA concentration and purity of plasmid DNA (pET21a+ or pETPull3.1) were measured using a NanoDrop ND-1000 (NanoDrop Technologies, USA), as per the manufacturer's specifications. Five µl of the prepared plasmid was pipetted into 50 µl of the chemically competent cell mixture. The cell mixture and plasmids were mixed and left on ice for 15 min. The cells were then heat shocked at 42°C for 3 min. After 3 min, 1 ml of LB broth was added to the 55 µl in the Eppendorf tubes and incubated in a water bath at 37°C for 1 hour. The transformation mixture (100 µl) was spread onto selective LB media containing 100 µg/ml Amp, and incubated overnight at 37°C (Pope & Kent,1996).

2.3.6 Plasmid extraction

For plasmid extraction from *E. coli,* the colonies which tested positive using colony PCR, were cultured in 5 ml LB broth with 100 µg/ml of Amp and incubated overnight at 37°C with shaking (170 rpm). The cells were harvested by centrifuging at 5687 × *g* for 3 min. The plasmid DNA extraction was performed using QIAprep Spin Miniprep Kit (Qiagen), following the manufacturer's instructions.

2.3.7 Restriction digestion of plasmid DNA and PCR products.

Restriction digestion reactions were set up in 0.2 ml thin-walled PCR tubes. 1 U of the desired restriction enzyme, NotI and NdeI was added to a 20 µl reaction, and 16 µl of nuclease-free water, 1 µg of plasmid DNA (pETPull3.1) and 2 µl of the appropriate buffer supplied by the manufacturer was added and mixed. The reaction was incubated at 37°C for 1 h to ensure complete digestion. Thereafter, the reaction was inactivated at 80°C for 10 min.

2.4. Protein expression and purification

2.4.1 Pull3.1 Protein expression

E. coli BL21(DE3) harbouring pETPull3.1 was streaked onto LB agar plates containing Amp (100 µg/ml) and incubated at 37ºC overnight. A starter culture was prepared by inoculating a colony into 10 ml LB broth containing Amp (100 µg/ml) and incubating overnight at 37ºC in an incubator with shaking at 170 rpm. Five ml of the starter culture was added to 50 ml LB media in an Erlenmeyer flask containing Amp (100 µg/ml) and isopropyl-β-d-thiogalactopyranoside (IPTG) (1 mM) and the culture was incubated at room temperature overnight. The culture was centrifuged at 20 784 × *g* for 10 min at 4ºC. The supernatant was discarded and the pellets were re-suspended in 100 mM Tris-HCl buffer pH 7. Each sample was sonicated on ice for 300 s, with a 30 s pause after each 30 s to prevent over-heating. The sonicated samples were centrifuged and the supernatants were collected and filtered through 0.45 µm syringe filters (Lasec, SA) to decrease the amount of cell debris. The filtered sample was processed for Pull3.1 purification as detailed in the next section.

2.4.2 Protein purification

The Pull3.1 protein was purified using metal affinity chromatography (Ni-charged) with the His•Bind® Resin Chromatography and Buffer kit (Novagen), according to the manufacturer's instructions. A 2 ml resin bed was prepared and equilibrated with three volumes of sterile deionised water followed by the addition of 5 volumes of charge buffer (50 mM $NiSO_4$) and three volumes of binding buffer (0.5 M NaCl, 20 mM tris-HCl, 5 mM imidazole, pH 7.9). After equilibration, an additional 1 ml of binding buffer was added, followed by the total volume of the supernatant. The column was washed with ten volumes of binding buffer. After the binding buffer had passed through six volumes of wash buffer (0.5 M NaCl, 60 mM imidazole, 20 mM tris-HCl, pH 7.9) was added.

The His-tag fused Pull3.1 was released from the column using the elution buffer (0.5 M imidazole, 0.25 M NaCl, 10 mM tris-HCl, pH 7.9). Concentration and buffer exchange of the eluted His-tag purified Pull3.1 was performed using the Amicon® Ultra-15 Centrifugal Filter Unit (Sigma) with 50 kDa MW cut off. The sample (8 ml) with the phosphate citrate buffer pH 6 (7 ml) was added to the Amicon® Ultra filter device and centrifuged at 3075 × *g* for 15 min at 4ºC. Another 7 ml of phosphate citrate buffer pH 6 (7 ml) was added, and centrifugation was repeated until the sample was concentrated down to 1 ml.

2.4.3 Determining protein concentrations

The purified enzyme concentration was estimated using the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific). The standard curve was determined using BSA (bovine serum albumin), as specified by the manufacturers. The absorbance of the standards was measured in a SPECTROstar® Nano microplate reader (BMG

Labtech) at 562 nm. The standards and unknown samples were determined in triplicate by adding 200 µl of the BCA working reagent and 25 µl of the protein sample into the 96-well flat-bottom microtiter plate. The samples were placed at 37ºC for 30 min and then allowed to cool to room temperature. Using the standard curve generated from the BSA absorbance readings, the protein concentrations of the unknown protein samples were determined.

2.4.4 Sodium Dodecyl-Sulphate Polyacrylamide Gel Electrophoresis

Sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyse protein expression and purification as described by (Laemmli, 1970). The SDS-PAGE gel was cast in a hand casting system (Bio-Rad) and consisted of a stacking gel and resolving gel, as described in Table 2.3. After the resolving gel (12%) was prepared, it was poured into the gel cassette. The stacking gel was added with a ten well comb and inserted on top of the stacking gel. Once set, the gels were placed in mini-PROTEAN tetra cells within the MINI-PROTEAN System. 1% running buffer (0.25 mM tris-HCl, 2 M glycine, 1% (w/v) SDS) was added prior to sample loading.

Protein sample preparation consisted of mixing protein (10 µl) and 5 µl of 2x sample buffer (0.18 M tris-HCl, pH 6.8; 9% (w/v) SDS; 0.075% bromophenol blue 2-mercaptoethanol; 15% (v/v) glycerol), and heating to 95˚C for 10 min prior to loading onto the polymerized polyacrylamide gel.

The colour pre-stained protein standard (broad range, 11-245 kDa) (New England Biolabs, USA) was used as a marker. The gel was run at 20 V for 10 min, then 200 V for 45 min. As soon as the dye front reached the end of the gel, the run was stopped, and the gel was removed and rinsed.

The rinsed gel was placed in a Coomassie brilliant blue staining solution (isopropanol 12.5% (v/v), acetic acid 10% (v/v) and Coomassie brilliant blue R250 0.25% (w/v) overnight. The stained gel was then placed in a de-staining solution (isopropanol 12.5% (v/v), acetic acid 10% (v/v), distilled water 77.5% (v/v) that was changed repeatedly until no background stain was left and only the bands could be seen.

Table 2. 3: SDS-PAGE gel composition

4% stacking gel		**12% resolving gel**	
Composition	Volume (10 ml)	Composition	Volume (10 ml)
Water (ml)	5.86	Water (ml)	3.2
30% acrylamide (ml)	1.34	30% acrylamide (ml)	4.0
0.5 M tris-HCl, pH 6.8 (ml)	2.6	1.5 M tris-HCl, pH 8.8 (ml)	2.6
10% SDS (µl)	100	10% SDS (µl)	100
10% (Ammonium Persulfate) (APS) (µl)	100	10% APS (µl)	100
TEMED (µl)	30	TEMED (µl)	10

2.5 Chromatographic methods

2.5.1 Fast performance liquid chromatography

Gel filtration chromatography was used to determine the quaternary structure of the Pull3.1 protein. A buffer exchange was conducted on His-tag purified Pull3.1 to 100 mM phosphate-citrate buffer with 100 mM NaCl at a pH of 6 using an Amicon® Ultra 50K device, with a 50 kDa MWCO as described in the protein purification section, the sample was concentrated down to 0.5 ml. The sample was filtered through a 0.22 µm syringe filter (GVS, USA).

The ÄKTA FPLC purifier system (Amersham Pharmacia Biotech Inc., USA) Superdex^{TM} 75 10/300 GL column was used to determine the size of the protein. The run conditions were as follows: the pressure limit was set to 1.8 MPa; the mobile phase was 100 mM phosphate-citrate buffer with 100 mM NaCl (pH 6) and a flow rate of 0.4 ml/min. The retention time of each peak was identified using a UV detector at 280 nm. Data analysis and peak identification was performed using Unicorn ver. 4.0 software (Amersham Pharmacia Biotech Inc., USA). β-amyalse from Barley (Sigma-Aldrich) and BSA (Thermo Fisher Scientific) were used as protein standards.

2.5.2 High performance liquid chromatography

HPLC was performed to determine the products released when Pull3.1 reacted on different substrates. Reaction mixtures of 1 ml were prepared, consisting of 100 µl Pull3.1 enzyme (20 µg), 100 µl of the substrate (starch, glycogen or pullulan) (0.2% (w/v) stock) and 800 µl buffer (100 mM phosphate-citrate buffer pH 6). The reaction was incubated overnight at 70ºC and placed on ice for 5 min, then centrifuged at 20 784 × *g* for 5 min at 22ºC. Zero-point five ml of the supernatant was added to an autosampler vial. The reactions were analysed using a Dionex Ultimate 3000 HPLC system. Hundred µl was injected on a Rezex RSO-Oligosaccharide Ag^{+} (4) % (12 µm) column at a flow rate of 0.3 ml/min, with the column oven set to 48ºC. The mobile phase was water. The following standards were selected: D- (+)-glucose (Merck); maltose; maltotriose (Sigma-Aldrich); maltotetraose (Megazyme); maltoheptaose (Megazyme). These standards were prepared in concentrations of 2, 4, 8 and 16 mM. Detection was performed via RI (refractive index). Chromeleon Dionex v6.80 software was used to analyse the peaks.

2.6 Enzyme assays

The DNS assay was used to determine the biochemical parameters for Pull3.1 catalysis. This included determining the temperature and pH optima, thermostability, substrate preference, the effect of cations detergents and inhibitors, and the enzyme kinetics.

2.6.1 DNS assay

The DNS (3,5-dinitrosalicylic acid) assays were performed using 0.2% (w/v) potato starch as the substrate. Every reaction contained 20 µg of Pull3.1 enzyme. Analysis of hydrolysis of the glycosidic bonds and the reducing sugars released was performed at pH 6 and a temperature of 70ºC for 20 min, unless stated otherwise. The reducing sugars released were measured according to Bernfield's protocol for performing the DNS assay (Bernfield, 1955).

Reactions were centrifuged at 20 784 × *g* for 5 min at 22ºC to reduce the size of the standard error in the OD readings. The cloudiness that occurs when working with polymerizing starch, especially at increasing concentrations, influences the OD readings. Once the samples were centrifuged, 100 µl of the sample was pipetted into 96-well plates with a clear flat bottom well (Greiner CELLSTAR®). A SPECTROstar® Nano microplate reader (BMG Labtech, Germany) was used to measure the OD at a 540 nm pathlength and volume set to 100 µl. Unless otherwise stated, all reactions were performed in phosphate-citrate buffer pH 6 for 20 min at 70°C with 2% (w/v) potato starch as the substrate. All experiments were performed with three biological repeats and three technical repeats.

The maltose standard was used to determine the concentration of sugars released, and a standard curve was prepared with each assay to estimate the values. The maltose standard curve is constructed using of a series of dilutions of a 0.2% (w/v) maltose reaction. DNS was added to the eight different concentrations of maltose and boiled for 15 min then immediately placed on ice for 5 min. The OD was then measured at 540 nm using a spectrophotometer. The SPECTROstar® Nano microplate reader (BMG Labtech, Germany) was used to measure the optical density (OD) reading from 96 well plate samples of enzyme assays, as specified by the manufacturer.

2.6.2 Temperature optimum

The optimum temperature was determined using phosphate-citrate buffer pH 6 and with 2% (w/v) potato starch as the substrate. Samples were incubated in 0.2 ml thin-walled PCR tubes in a thermocycler for 20 min. The temperature levels selected were 34.9, 40.7, 44.6, 50, 55, 60.8, 64, 70, 75, 80.9, 84.4, 89.4, 95.3 and 99. These were the temperatures that could fit the widest range within a thermocycler, where these assays were conducted. DNS assay was performed (section 2.5.1) and samples were centrifuged at 20 784 × *g* for 5 min at 22ºC. They were then transferred to the 96 flat-bottom well microtiter plates and the absorbance levels were read at 540 nm.

2.6.3 pH optimum

Buffers were used to determine the pH optimum and extended over a pH range from 3-11: 100 mM phosphate-citrate buffer pH 3-8; 100 mM tris-HCL Buffer pH 7-9;

100mM glycine-NaOH buffer pH 9-11; Na_2CO_3-$NaHCO_3$ Buffer pH 9-11 (Cusabio technology LLC, 2020). The samples were treated the same as described in section 2.6.2 and a DNS assay was performed. The samples were then centrifuged and the OD was measured.

2.6.4 Cations, detergents and inhibitors

Various cations, detergents and inhibitors were tested, i.e.: $CaCl_2$ (Sigma-Aldrich); $MgCl_2 \cdot 6H_2O$ (Merck); $NiCl_2 \cdot 6H_2O$ (Riedel-de Haën); $CuCl_2 \cdot 2H_2O$ (Sigma-Aldrich); KCl (Merck); NaCl (Labchem); $FeCl_3$ (Sigma); $ZnCl_2$ (Merck); β-mercaptoethanol (Sigma-Aldrich); EDTA (Merck); DTT (Sigma-Aldrich); Triton X (Sigma-Aldrich); Tween 20 (Sigma-Aldrich); Tween 80 (Sigma-Aldrich); SDS (Sigma-Aldrich). The compounds were tested at a concentration of 1 mM and 5 mM to be able to make the best comparison to other enzymes (Wu *et al*., 2018; Cruz *et al.*, 2018).

2.6.5 Substrate preference

The following substrates were tested using the DNS assay at 0.2% (w/v): potato starch (Sigma-Aldrich), amylum starch (Millipore), amylose (Sigma-Aldrich), pullulan (Sigma-Aldrich), glycogen from oyster (Sigma-Aldrich), glycogen from bovine liver (Sigma-Aldrich), beta-glucan from barley (Megazyme), lichenan (Megazyme), carboxymethyl cellulase (CMC) (Sigma-Aldrich) and 1,4-beta-D-mannan (Megazyme).

2.6.6 Kinetic characterisation

The enzyme reaction rate was determined using potato starch (Sigma-Aldrich) as the substrate at increasing concentrations. The maximum concentration that could be tested was 0.10 mg/ml because of the increase in viscosity due to the high starch concentration. The experiment was performed at 55ºC for 20 min at pH 6, in triplicate and using three biological repeats. The kinetic parameters (K_M and V_{max}) were determined by generating Michaelis-Menten plots using GraphPad Prism 4 (GraphPad Software, Inc).

2.6.7 Thermostability

For the thermostability assay the Pull3.1 enzyme was first incubated on ice, 50, 60, 70 and 80ºC for 0, 10, 20, 30, 40, 50 and 60 min. Afterwards the DNS assay method described in section 2.6.1 was followed.

Chapter 3: Results and discussion

3.1 Introduction

Metagenomic screening has become a widely-used method for discovering novel proteins, since the organisms harbouring these proteins are not always culturable. There are two main methods for discovering proteins: function-based screening and sequence-based screening (Nazir, 2016). Enzymes have been identified from environments such as sheep rumen, soil and hot springs, using sequence-based screening (Hua *et al.*, 2015; Motahar *et al.*, 2020; Zhao *et al.*, 2017). The Pull3.1 ORF identified from the metagenomic screen on the Mphizi hot spring sequence data was considered an interesting candidate for enzymatic studies. The preliminary bioinformatic analysis indicated that the enzyme could possibly act on glycosidic linkages in starch, pullulan and cyclomaltodextrin (section 3.2).

The enzymes that break down the above-mentioned substrates are glycoside hydrolase enzymes that act on either the α-1,4- or the α-1,6-glycosidic bonds in these substrates. The enzymes usually involved are amylase, pullulanase and cyclomaltodextrinase (CDase), respectively. The biggest differences between amylases and pullulanases are their substrate preferences and the products released after hydrolysis. Pullulanase enzymes are debranching enzymes that prefer to act on pullulan on the α-1,6-glycosidic bonds. These enzymes mainly produce maltotriose, panaose and isopanose based on the exact type of pullulanase at work (Figure 1.9). Amylase usually prefers starch, amylopectin, and amylose, and mainly produces glucose or maltose after hydrolysis. α-Amylases can be found in four varieties: the endo-acting amylase is α-amylase; the exo-amylase enzymes include β-amylase, γ-amylase, and α-glucosidase.

Glycoside hydrolase enzymes use a retaining double-displacement mechanism to break the bonds between glucose units (Figure 1.11). Enzymes that are able to act on these bonds are used in a multitude of industries, but what makes them worth investigating is the possibility that there might be dual modes of action, which is desirable as two steps of a process (e.g., liquefication and saccharification) can then be completed in a single step. α-Amylase is probably one of the most comprehensively documented enzymes and retrieved from a vast number of environmental niches, including desert soil (Nithya *et al.*, 2017). However, if an α-amylase were to be discovered with dual modes of action it would allow users of

these enzymes to save both time and money, therefore the screening for α-amylases with altered properties/functions is of value.

Antibiotics and antimicrobials cannot penetrate biofilms easily. α-Amylase is a proposed solution to this problem and has been used in the medical industry, where it helps to destroy harmful biofilm (Lahiri *et al.*, 2021). The increase in renewable energy sources also favours α-amylase, since it is a crucial ingredient in bioethanol production, as α-amylase can break down the starch in grains (Wang *et al.*, 2014). The same can be said for pullulanase, which is being studied for use in biorefineries (Toor *et al.*, 2021). Thermostable pullulanases can perform both liquefication and saccharification without the need to adjust pH or temperature (Amin *et al.*, 2021; Nisha & Satyanarayana, 2016). Pull3.1 could possibly function at these temperatures, which could be valuable in industry. It was therefore selected as the characterisation target for this study.

3.2 Analysis of primary amino acid sequence

For this study, the sequenced fosmid library clones that contained metagenomic DNA created by Hu (2010) were screened. The library was constructed using mDNA from a sediment sample from the Mphizi hot spring in Malawi and contained roughly 10,000 clones, with an average insert size of ±30 kb. From this library, 4416 clones were sequenced using the Illumina MiSeq platform. Following de novo assembly, annotation of contig 492 identified 24 ORFs (Table 3.1), with ORF8 predicted to encode a possible pullulanase, designated Pull3.1, which was selected for further investigation and characterisation in this study.

Pull3.1 (ORF8) was annotated in KBase as a 1,4-alpha-glucan branching enzyme; ORF9 as a hypothetical protein and ORFs 10-12 that follow are maltose binding and maltose transporter proteins, respectively. These proteins suggest that during earlier processes involved in glycolysis the end products produced are maltose, for this reason there are maltose transporters downstream. ORF13, predicted to be an ATP-dependent 6-phosphofructokinase 1 (Table 3.1) may be important in glycolysis converting fructose-6-phosphate into fructose 1,6-bisphosphate and ADP.

Table 3. 1: Putative assignment of the 24 ORFs predicted on contig 492

ORF number	ORF name	Gene position	Gene length (bp)
1	Putative 2-aminoethylphosphonate transport system permease protein PhnU	3..662	659
2	Thiamine-binding periplasmic protein	638..1654	1016
3	Hypothetical protein	1667..3133	1466
4	Pyridoxine/pyridoxal/pyridoxamine kinase	3151..4482	1331
5	Thiamine thiazole synthase	4461..5258	797
6	Hypothetical protein	5471..5761	290
7	Thiamine-monophosphate kinase	5765..6697	932
8	1,4-alpha-glucan branching enzyme GlgB	6987..9080	2093
9	Hypothetical protein	9198..11273	2075
10	Maltose/maltodextrin-binding periplasmic protein	11357..12544	1187
11	Maltose/maltodextrin transport system permease protein MalF	12638..14161	1523
12	Maltose/maltodextrin transport system permease protein MalG	14151..15008	857
13	ATP-dependent 6-phosphofructokinase 1	15011..15979	968
14	Ferredoxin-NADP reductase	15976..17307	1331
15	Ribosomal RNA small sub-unit methyltransferase D	17381..17929	548
16	5'-nucleotidase SurE	17926..18744	818
17	Hypothetical protein	18766..19491	725
18	5,6-dimethylbenzimidazole synthase	19515..20057	542
19	Threonine synthase	20293..21165	872
20	Hypothetical protein	21167..21655	488
21	Thiosulfate sulfurtransferase GlpE	21715..22761	1046
22	Hypothetical protein	22758..23651	893
23	Cyclic 2,3-diphosphoglycerate synthetase	23755..25065	1310
24	Hypothetical protein	25148..25687	539

Further investigation was conducted specifically on ORF8. ORF8 encoding Pull3.1 is 2093 nucleotides long that translates to 686 amino acids for the mature peptide (excluding signal sequence). The MW of the putative mature protein is 77,85 KDa. The signal sequence has a typical tripartite architecture: the N-region starts at M^1,

the H-region spans L^2-L^8, and the C-region spans G^9-A^{12} (Figure 3.2). BLASTp analysis was used to find similar proteins either characterised or uncharacterised. The protein sequences with the highest percentage similarity to Pull3.1 were from the following organisms: *Candidatus* Bipolaricaulota bacterium (MBC7092855.1; 81.7%) from an oil production facility metagenome; *Candidatus* Atribacteria bacterium (HCL89767.1; 75.5%) from a marine sediment metagenome; *Thermococcus* sp. EP1 (CAB94218.1; 73.7%); and *Thermococcus aggregans* (WP_082391529.1, 74.2%) from a yellow meal worm (Liu *et al.*, 2020; Parks *et al.*, 2018; Strobl *et al.*, 1998) (Table 3.2). These are all annotated as pullulanase or α-amylase domain-containing proteins, thus are thought to belong to the GH13 family.

Table 3. 2: Similarity/identity matrix analysis of Pull3.1 and the closest relatives identified through BLASTp analysis

Organism or protein	Accession number	1	2	3	4	5	6	7	8	9
1.Pull3.1	N/A		70.9	63.1	61.6	60.7	59.1	61.3	57.8	57.1
2.Candidatus_Bipolaricaulota_bacterium	MCB7092855.1	81.7		63.8	63.7	62.2	62.8	63.3	59	59.4
3.Candidatus_Atribacteria_bacterium	HCL89767.1	75.5	79.4		89.2	82.3	55	83.9	52.5	53.5
4.Candidatus_Atriarchaeales	PKP54169.1	74.3	80.1	94.4		82	55.4	83.3	52.2	52.9
5.Candidatus_Atribacteria_bacterium	MCB8499447.1	73.5	78.5	90.4	91		53.4	78.1	51.3	52.3
6.Thermococcus_sp._EP1	WP_0823981529.1	73.7	76.8	71.8	72	71.5		54.9	73.9	73.1
7.Atribacteria_sp.	MBA3062470.1	75.4	79.2	90.6	90.6	86.7	70.1		50.9	51.9
8.Thermococcus_aggregans	CAB94218.1	74.2	75.2	71.3	71.3	70.7	86.8	69.1		81
9.Thermococcus_sp._101_C5	WP_152879761.1	73.6	74.1	71.7	72.1	71.5	85.9	69.9	91.7	

The similarity is indicated in green; the identity is indicated in blue.

Further analyses through various databases (Pfam, InterProscan, the superfamily HMM library, the genome assignments server, and BLASTp nucleotide architecture) were performed, and identified that Pull3.1 belongs to the glycoside hydrolase family 13 (Figure 3.1) and might have CDase properties.

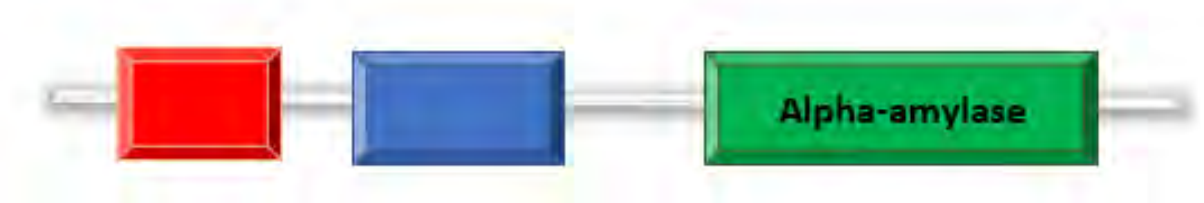
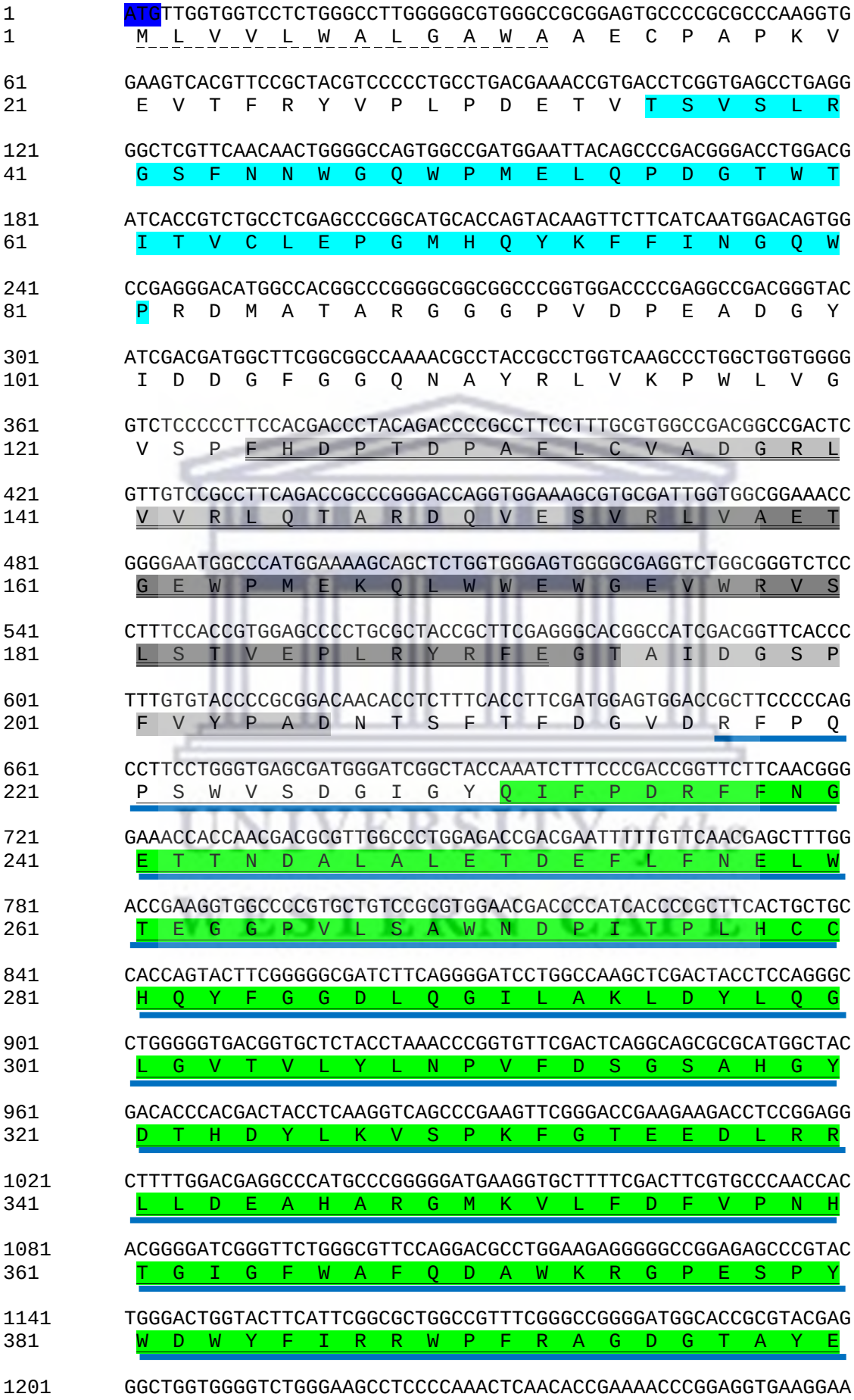


Figure 3. 1: Pfam analysis: α-amylase domain identification. The green box indicates an alpha-amylase catalytic domain. The blue box indicates an alpha-amylase, N-terminal Ig-like domain. The red box indicates a glycogen recognition site of AMP-activated protein kinase.

Region P^{221}-R^{621} was identified to have a transglycosidase domain (Figure 3.2). This is a well-known attribute that is often associated with CDase (Park *et al.*, 2000). The standard fold found in more than half of the GH families' catalytic domains is known as the TIM alpha / beta-barrel, which contains the active site is found in region R^{217}-A^{631} (Figure 3.2) (Park *et al.*, 2000; Rigden *et al.*, 2003).

The glycogen recognition site at region T^{35}-W^{80} shares homology with the sites found in AMP-activated protein kinases. These sites help with glucose uptake as well as in the binding to glycogen and it is also responsible for creating a tighter grip on β-cyclodextrin via the two tryptophan molecules that cradle the ring. At the same time, the leucine pierces the CD ring (Polekhina *et al.*, 2005).

The Ig-like N-terminal domain is present throughout the α-amylase family. This indicates that the putative domains are linked to the catalytic domain of the CDase, as well as to the N-terminal of α-amylase, which may act as a multimer interface domain. The Pfam E-value for the putative carbohydrate binding module family 48 (CBM48) domain showed very low confidence. However, considering that the enzyme showed greatest similarity to α-amylase, pullulan hydrolase type III (PHTIII), and CDases, a CBM48 would be expected to be present. CBM48 is closely related to CBM20 and they do sometimes co-occur, as seen with α-amylase from *Bacillus* sp. AAH-31. CBM48 is the only CBM for PHTIII, while there are 4 CBMs associated with CDase; CBM48 being one of them. Thus, the low E-value indicates that this maybe a novel CBM48 domain.



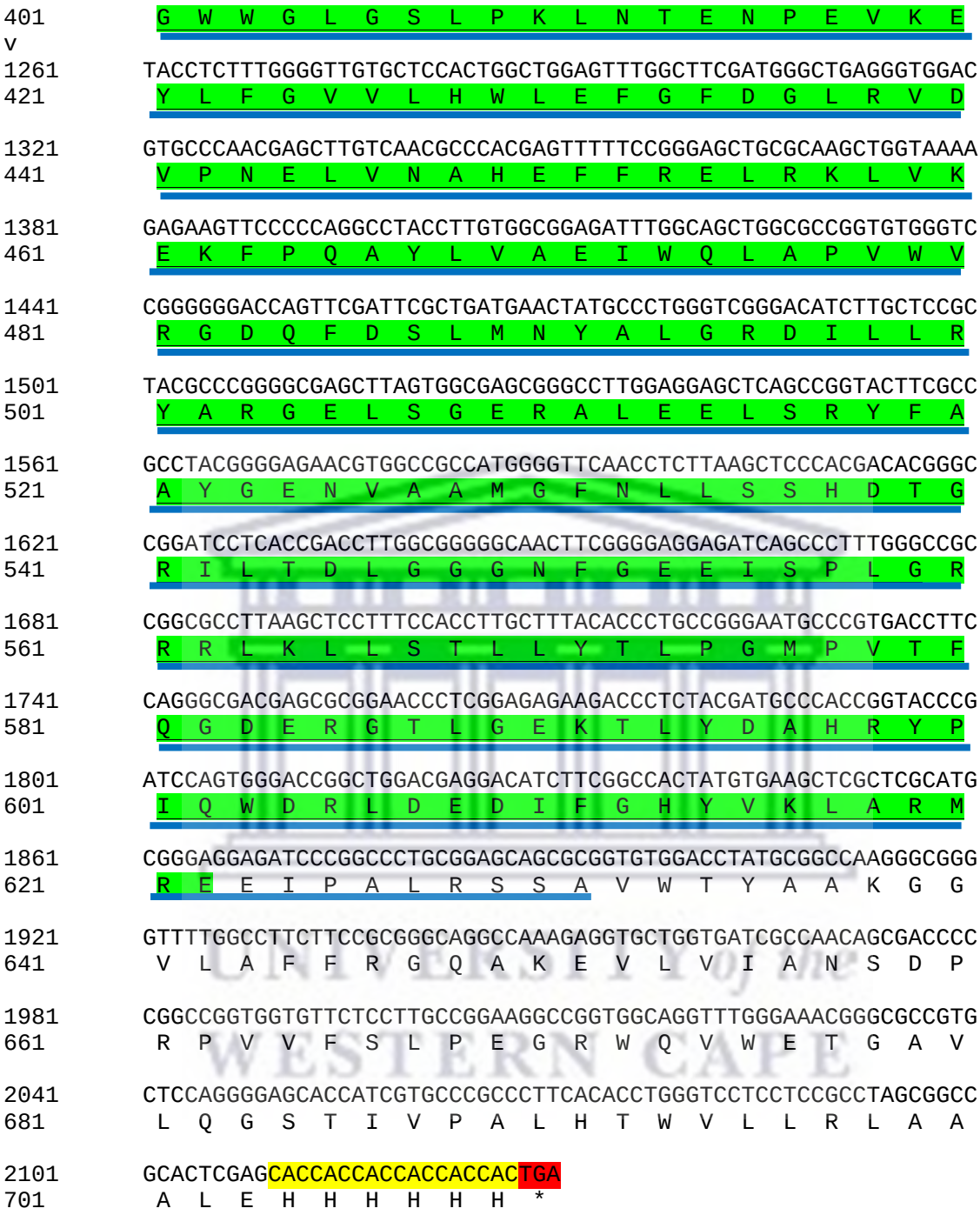


Figure 3. 2: Translated nucleotide sequence and identified features of Pull3.1 in pETPull3.1 vector. The start codon is highlighted in blue; the signal peptide is indicated with a dashed line. A glycoside hydrolase family 13 domain was detected (bright green), while the transglycosidase superfamily domain that overlaps with the GH13 domain is underlined. The putative glycogen binding site is shaded in turquoise. The light grey area shows the α-amylase, N-terminal Ig-like domain. This domain overlaps with the catalytic domain of CDase (the double-underlined section) and CBM48 (the dark grey section). The TIM barrel structure which is composed of residues 217–631 is underlined with a blue line. The his-tag is highlighted in yellow. The stop codon is highlighted in red.

3.3 Phylogenetic analysis

The phylogenetic tree was composed using the characterised and uncharacterised proteins identified from BLASTp analysis (Table 3.2) and the IMG/M database containing metagenomes from different environments. Various hot springs, marine soils and gut microbiomes were used to find similar proteins and generate the phylogenetic tree. Inspection of the phylogenetic tree indicated that Pull3.1 has ancestors in clade A.2 (Figure 3.3). Clade A.2 generally contains proteins from thermophilic Archaea, *Thermococcus*. It has previously been noted that pullulanase and CDase are widespread in the *Thermococcus* genus (Lee *et al.*, 2013a) but in this clade there is a nearly equal amount of α-amylase and pullulanase. Pull3.1 clustered with the putative bacterial proteins isolated from hot springs in North America found in clade A.1 (Canada: Ga0209381 100091947 and USA: Ga0308401 10055651). These proteins in Clade A.1 were all pullulanases. This is suggesting that Pull3.1 will perform pullulanase activity and because of the ancestral mix of α-amylase and pullulanse it could be expected that Pull3.1 may also have α-amylase properties. The CDases clustered on a separate branch it was assumed when the bioinformatics was initially performed that this will not form part of Pull3.1's range of activities. It was subsequently tested experimentally and discussed in section 3.6.1.

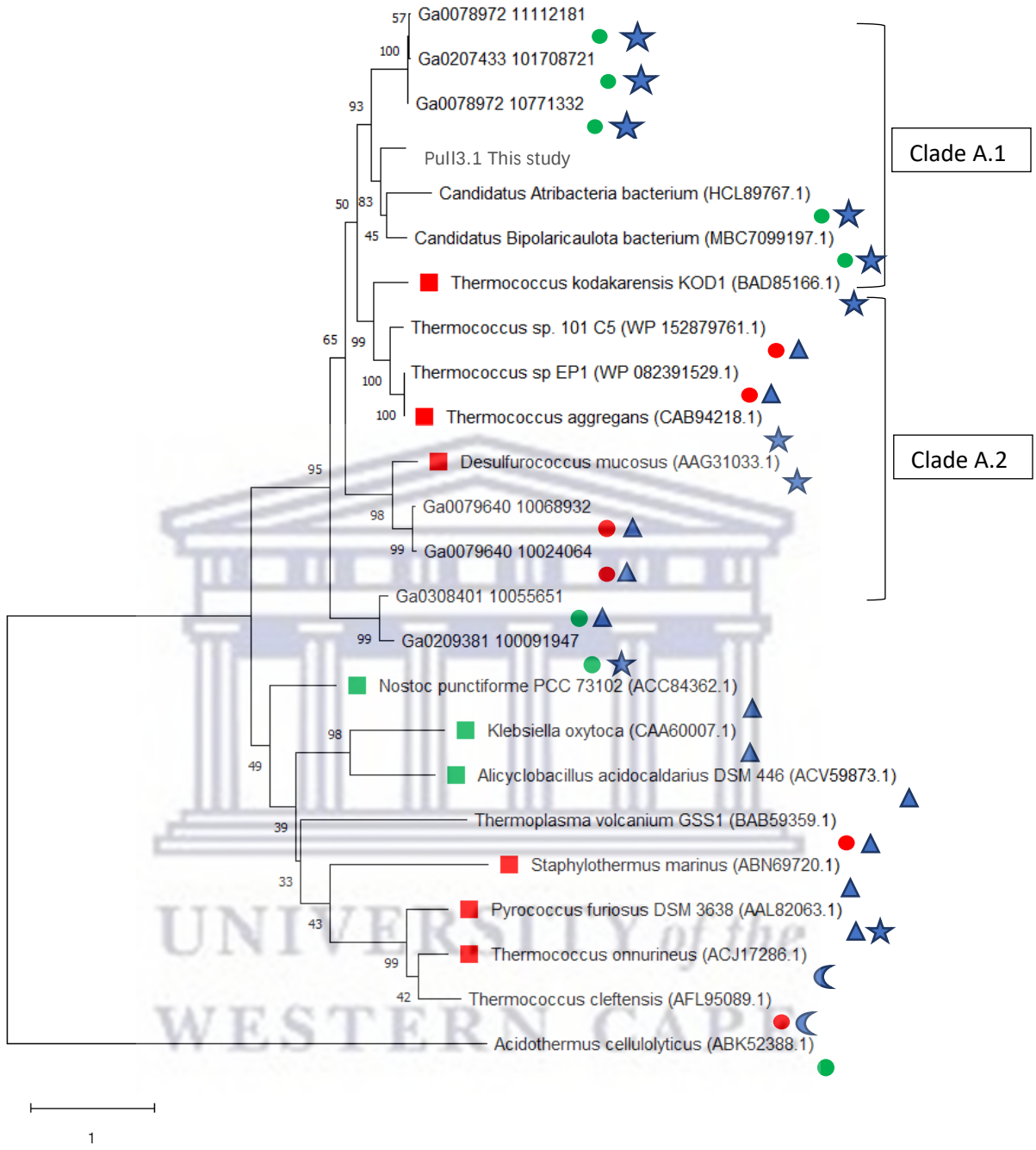


Figure 3. 3: Unrooted phylogenetic tree indicating the evolutionary relationship of Pull 3.1 relative to proteins to which it shows the highest sequence identity. The tree was constructed using the JTT matrix-based model and the Maximum-likelihood statistical method. A 1000 bootstrap replicates were used as a test of phylogeny. The scale indicates the number of substitutions per site. The sequences were labelled based on the genus and species names of the organisms in which these proteins were discovered. The accession numbers are provided in brackets. The sequences that have labels that start with Ga were retrieved from Integrated Microbial Genomes and Microbiomes (IMG/M) and are uncharacterised. The label is the gene ID. The sequences with a square are characterised and belong to the GH family 13 sub-family GH13_20. The red coloured shapes are proteins from the Archaeal species, while the green coloured shapes are proteins from bacterial species, the squares indicate characterised proteins while the circles indicate uncharacterised proteins. The blue stars indicate pullulanse proteins, the blue triangle indicates the alpha-amylases and the crescents represent the cyclomaltodextrinases. *A. cellulolyticus* 11B was chosen as an

outgroup, as it is a characterised chitinase from glycoside hydrolase family 6. The protein sequences that do not have a square next to the label are the best BLASTp hits obtained when comparing the amino acid sequence to the NCBInr database that are uncharacterised enzymes.

3.4 Multiple sequence alignment

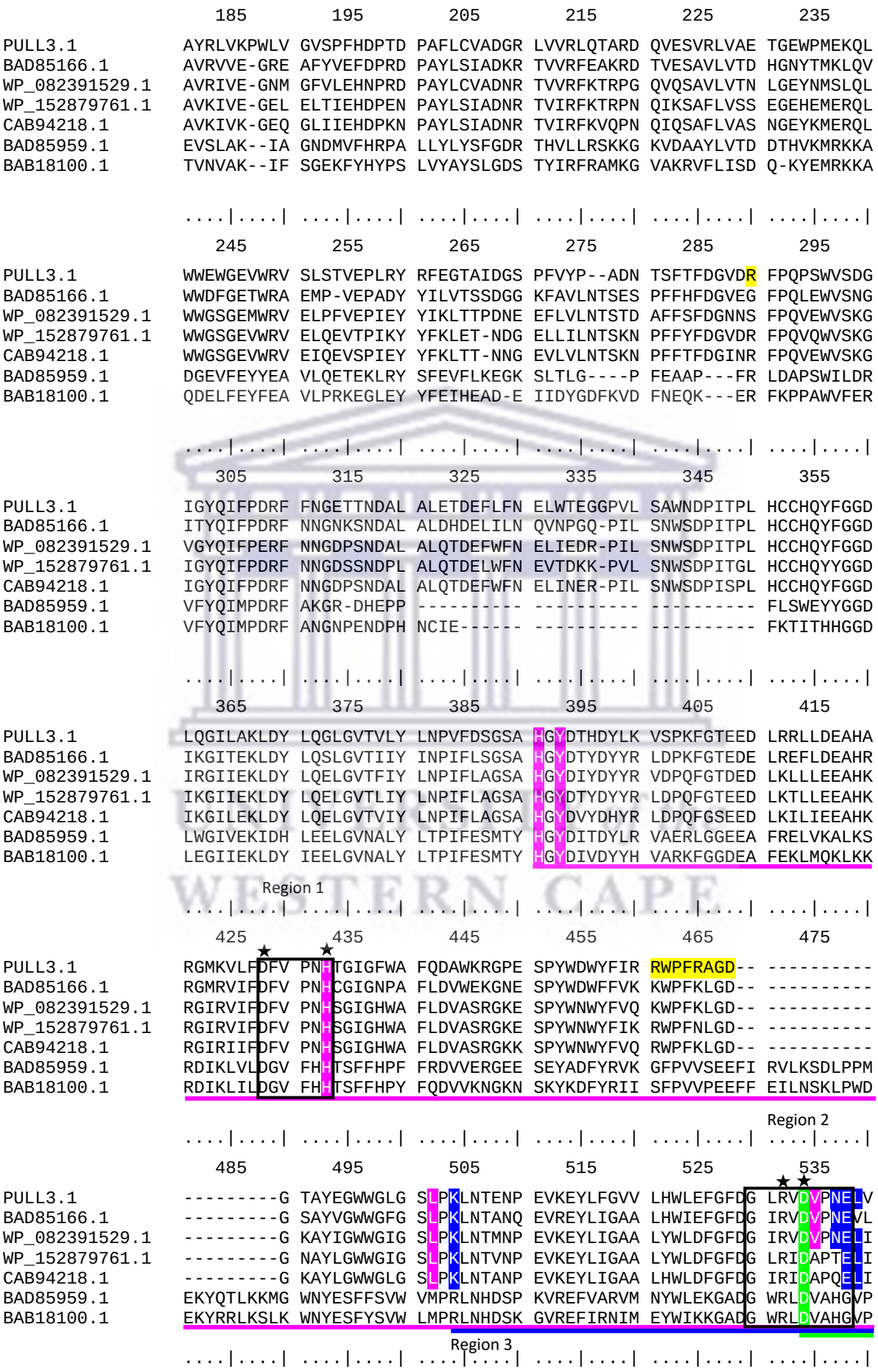
The sequences chosen for the MSA were chosen because they were the most closely related characterised enzymes and because they all belong to GH13_20. They were placed in the order of highest percentage identity to Pull3.1. These enzymes can convert starch, pullulan and CDs, but are classified under different EC numbers. The two most related enzymes to Pull3.1, i.e., *T. kodakarensis* KOD1 (BAD85166.1) and *T. aggregans* (WP 082391529.1), are both pullulan hydrolase type III (PHTIII) enzymes. The two at the bottom of the list are CDases from *T. kodakarensis* KOD1 (BAD85959.1) and *Thermococcus* sp. B1001 (BAB18100.1). Both displayed much poorer alignment with Pull3.1. The alignment suggests that Pull3.1 is more likely to be a PHTIII than a CDase.

```
                  ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    5         15         25         35         45         55
PULL3.1           -------MLV VLWALGAWAA ---------- ---------- ---------- ----------
BAD85166.1        MKK--GGLLL ILLILVSIAS GCISESNENQ TATASTVPPT SVTPSQSSTP TTSTSTYGPS
WP_082391529.1    MEKRYLSLIF ILLILSVTFA GCITQTN--- ---------- ---------- ---------T
WP_152879761.1    MKRWTSVGIV LVFIIGIIAS GCLQQIKSQ- ---------- ---------- ---------G
CAB94218.1        MRRRFGALLL LMFMLNIIAS GCLQSP---- ---------- ---------- ----------
BAD85959.1        -------MYK VFGFEENFIH G--------- ---------- ---------- ----------
BAB18100.1        -------MYK IFGFKDNDYL G--------- ---------- ---------- ----------

                  ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    65         75         85         95        105        115
PULL3.1           ---------- ---------- ---ECPAPKV EVTFRYVPLP DETVTSVSLR GSFNNWGQWP
BAD85166.1        ERTELKLPSV NYTPIYVG-- IEKGCPSGRV PVKFTYNPGN -KTVKSVSLR GSFNNWGEWP
WP_082391529.1    KTETLVIPEG NYTTIYLNEK GTSQCPASKV PVTFTYTPPK NESIDSVSLR GSFNDWAEWP
WP_152879761.1    TEQTLTLPED NYKTLYLNEK VKGTCPTGKV PVTFTYNPQG -KNVTSVSLR GSFNNWKELP
CAB94218.1        TTQELKLPSG NYPPIYINEK SQNMCPPGKV PVTFRYQPE- -ENVTSVSLR GSFNDWGELP
BAD85959.1        ---------- ---------- -----RVARV EFSLPDAGRW ----DYAYLL GNFNAFNEGS
BAB18100.1        ---------- ---------- -----KVGIT EFSIPKSG-- ----SYAYLL GNFNAFNEGS

                  ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    125        135        145        155        165        175
PULL3.1           MELQP-DGTW TITVCLEPGM HQYKFFINGQ WPRDMATARG GGPVDPEADG YIDDGFGGQN
BAD85166.1        MELK--NGTW ETTVCLRPGR YEYKYFINGQ WVKDMSDDGT GRPYDPDADA YAPDGYGGKN
WP_082391529.1    MQKE--NGAW SITVCLEPGR YEYKFFINGE WVKDMSSDKY GNPVDPEADE YVDDGFGGKN
WP_152879761.1    MENE--NGVW KKTVCLEPGQ YEYKFFINGE WVKDMS---- --TVDPTADS YIDDGFGGKN
CAB94218.1        MKNE--NGTW VRTVCLNPGR YEYKFFVDGE WIKDMS---- --AVDPTADA YVDDGFGGKN
BAD85959.1        FRMKHEDKRW IIEIKLPEGL WRYAFSAGGE FLLDPE---- ----NPEKEL YRRPSYKFER
BAB18100.1        FRMREKGDRW YIKVELPEGI WYYTFSVDGN LILDFE---- ----NNEKTV YRRLSYKFEK

                  ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
```



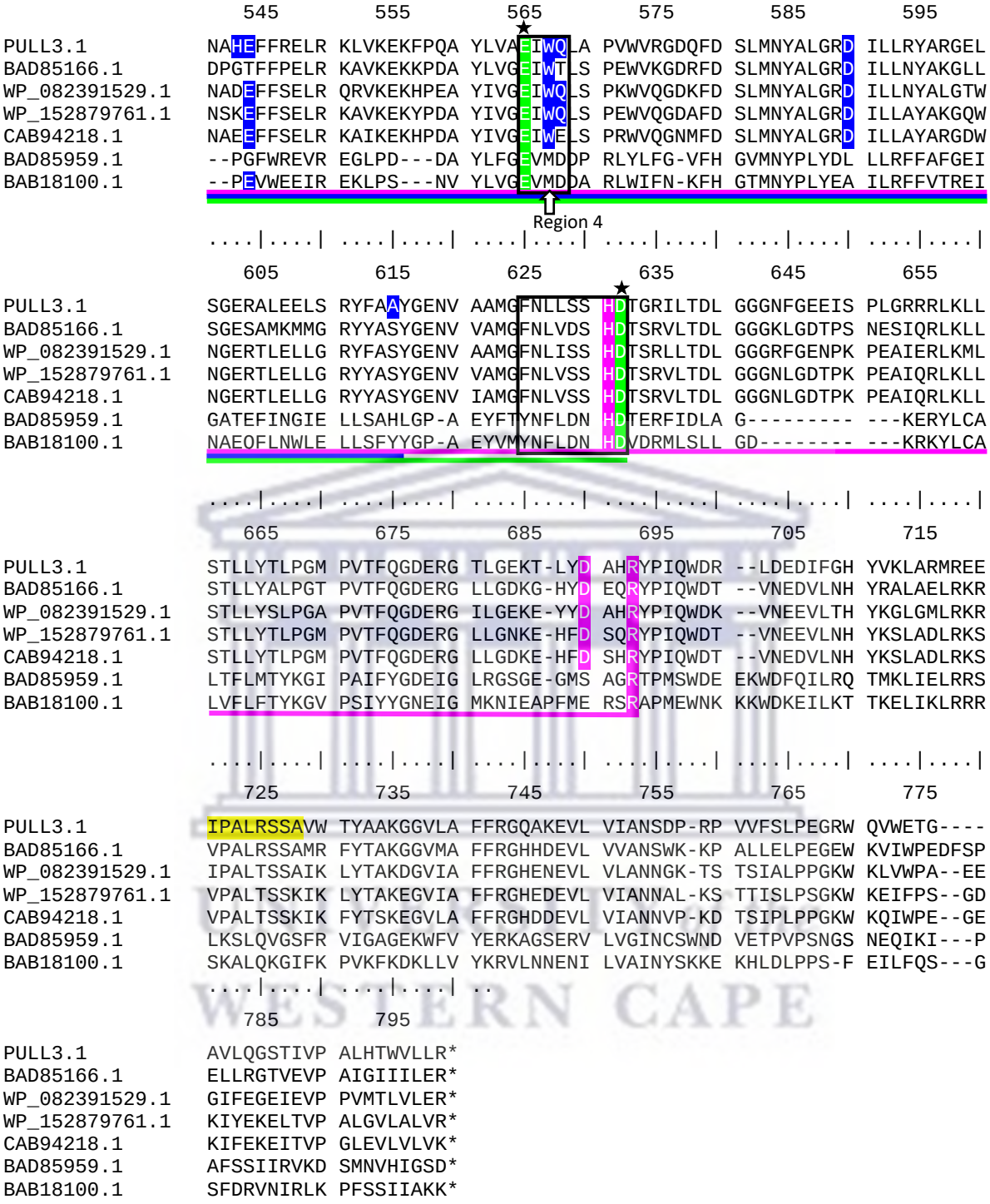


Figure 3. 4: Multiple sequence alignment of Pull3.1 with closely related, characterised enzymes within the GH13 subfamily 20. The following protein sequences are represented: pullulan hydrolase type III from *T. kodakarensis* KOD1 (BAD85166.1); pullulan hydrolase type III from *T. aggregans* (WP_082391529.1); α-amylase C-terminal domain-containing protein from *Thermococcus* sp. 101 C5 (WP_152879761.1); α-amylase C-terminal domain-containing protein from *Thermococcus* sp. EP1 (CAB94218.1); CDase from *T. kodakarensis* KOD1 (BAD85959.1); CDase from *Thermococcus* sp. B1001 (BAB18100.1). The green line indicates the glycogen binding site. The green shaded letters on the green line are the conserved feature residues. The pink line indicates the active site and the highlighted residues are the conserved regions. The bright green line indicates the catalytic site, which completely overlaps the substrate binding site. The three bright green highlighted letters are

conserved regions for the catalytic sites. The blue line represents the homodimer interface, and the highlighted letters indicate the conserved regions. The four regions conserved in all GH13 family members are indicated with the rectangular outlined boxes numbered 1-4, and the conserved residues within the regions are indicated with black stars. The black arrow at conserved region 4 indicates a highly conserved tryptophan related to transglycosylation.

To gain a better understanding of the protein structure of Pull3.1 the SWISS-MODEL sever was used. The SWISS-MODEL server identified *T. kodakarensis* KOD1 (BAD85166.1) (TK-PUL), the PHTIII, as the model with highest percentage identity (62.07%) (Figure 3.5 A). This agrees with what is seen in the phylogenetic tree (Figure 3.3) since Pull3.1 clustered with the pullulanases. The crystal structure of TK-PUL (Figure 3.5 A) was used as a template to build the putative Pull3.1 model. The loop identified in Figure 3.5 A is a unique feature of PUL-TK, which is believed to possibly have a strong influence on substrate binding, and especially on the cleavage of α-1,6-glycosidic linkages (Guo *et al.*, 2018). The superimposed structure suggests that Pull3.1 also has this feature and therefore should be able to also cleave α-1,6-glycosidic linkages. The Ca^{2+} binding site is also present in the sequence of Pull3.1 (Figure 3.5 C). According to Guo *et al*. (2018), the Ca^{2+} ion is buried and entirely dehydrated. The Ca^{2+} binding site could also contribute to stabilising the protein. The salting-out effect of Ca^{2+} binding might cause the formation of a more compressed structure, which then results in higher thermostability of the enzyme (Santos & Barbosa-Tessmann, 2019).

The four conserved regions found in the TIM barrel structure among amylolytic sequences are present as well as the catalytic triad composed of Asp^{538}, Glu^{569} and Asp^{636} (Figure 3.5 B) that are highly conserved in the GH13 family (Figure 3.5). As mentioned in section 1.3, during starch hydrolysis the CBM identifies the substrate and once the active site binds to the substate, two amino acids- one Asp^{538} and the Glu^{569}- are involved in the retaining double displacement mechanism (Figure 1.6). Pull3.1 also contains the conserved Trp^{585} responsible for transglycosylation and for stabilising the catalytic triad intermediate (Casa-Villegas *et al.*, 2018).

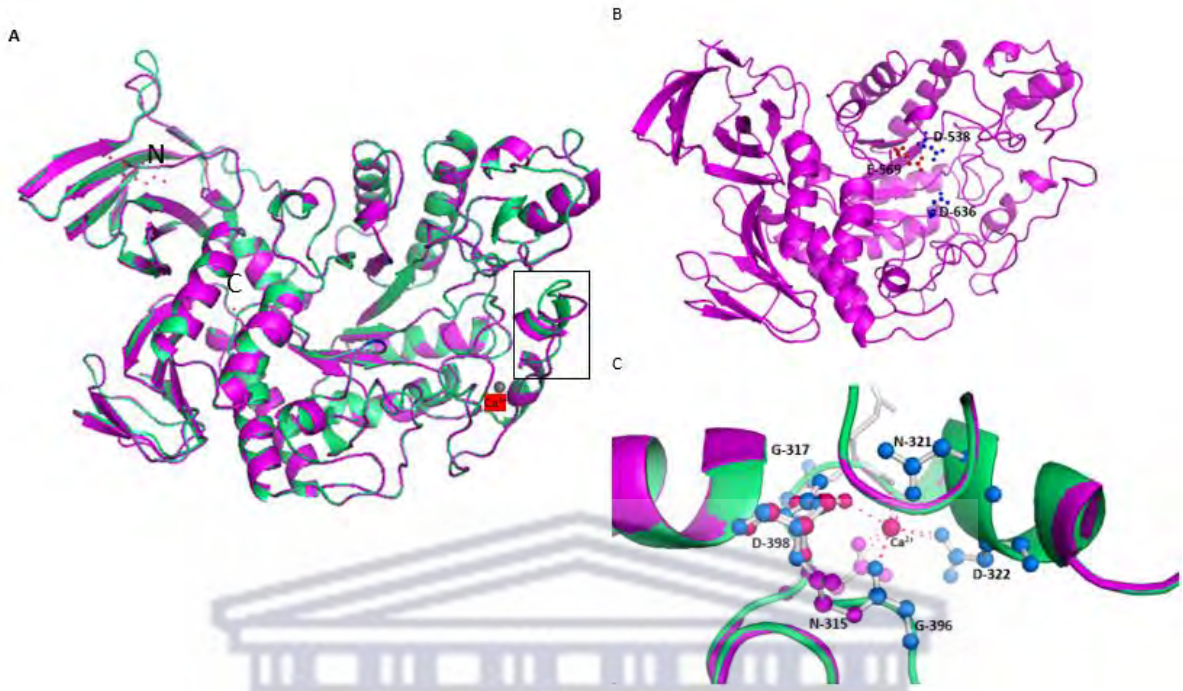


Figure 3. 5: Structural comparison and functional insights of the predicted model for Pull3.1 and *T. kodakarensis* KOD1 TK-PUL. A) Superimposition of the predicted model for Pull3.1 (pink) and the *T. kodakarensis* KOD1 (BAD85166.1) TK-PUL used as the template (green). The N-terminal and C-terminal positions are indicated. The dots indicate the atom selection of the terminal site residues. The black box indicates the loop responsible for influencing strong substrate binding. The grey dot represents the Ca^{2+} ion. B) The catalytic triad is composed of two aspartic acid residues in blue and glutamic acid in red displayed in the ball-and-stick representation. C) The Ca^{2+}binding site is surrounded by main chains and side chains. The pink highlighted ball-and-stick icons were selected from the TK-PUL sequence to indicate the Ca^{2+} site and how it would bind at the Asp-398 site of TK-PUL. The blue ball-and-stick figures are from Pull3.1 and represent the residue involved in Ca^{2+}binding Phe-314, Gly-317, Asn-321, Asp-309, Gly348 and Asp-398.

Overall, the sequence analysis indicated that Pull3.1 could be a novel amylase/PHTIII or less likely a CDase, and that no close relatives have been characterised. Therefore, characterisation of this enzyme may shed light on the relationship between sequence divergence from characterised enzymes and associated biochemical behavior, such as substrate preference, temperature stability, optima, and kinetic performance. For these reasons, biochemical characterisation of the enzyme was performed. This characterisation is explained in the following sections.

3.5 Confirmation of the clone pET-Pull3.1

The results presented in this section were used to confirm that Pull3.1 was correctly cloned and to assess if enzyme over-expression would be at an adequate level to perform enzyme assays.

3.5.1 Restriction enzyme digest analysis

The ORF encoding Pull3.1 was cloned and expressed in the *E. coli* BL21(DE3)-pET-21a system. A construct consisting of Pull3.1 cloned into pET-21a at the NdeI and NotI sites (pETPull3.1) was synthesized by Biomatik (Figure 2.1) and was checked on arrival through restriction endonuclease digestion (Figure 3.6). The vector backbone can be seen at 5.4 kb and the Pull3.1 insert at 2.1 kb. This confirmed that the construct was correct, and it was transformed into *E. coli* BL21(DE3) for protein expression.

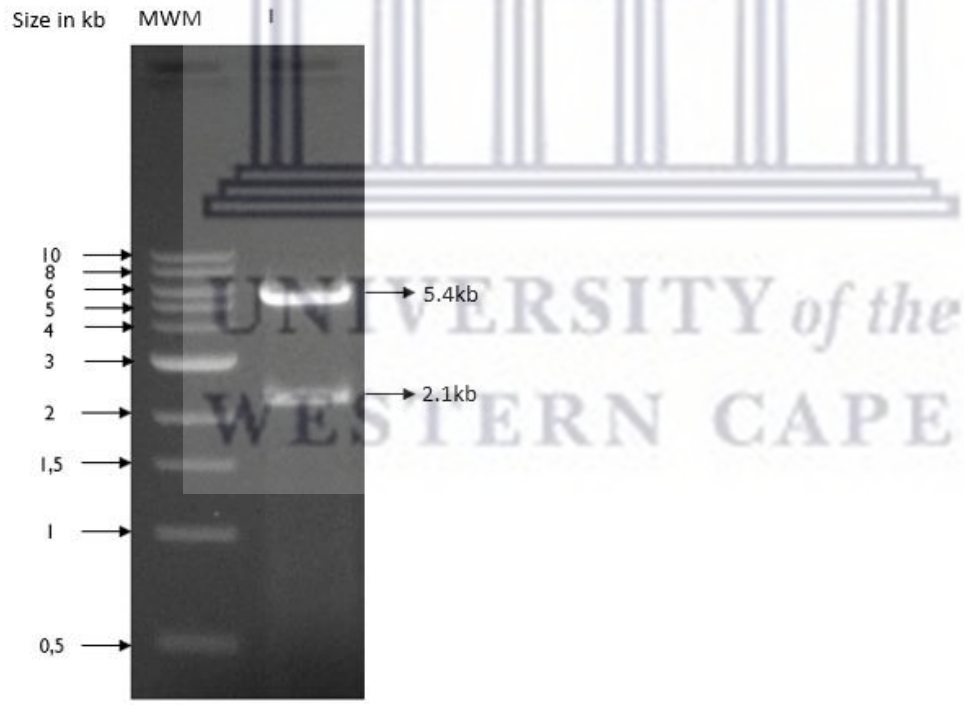


Figure 3. 1: Restriction digestion of pETPull3.1 with NdeI and NotI enzymes analysed on a 1% agarose gel. Lane MWM, 1 kb MWM marker; Lane 1, digested pETPull3.1 with a vector backbone size of ~5.4 kb and band size of ~2.1 kb corresponding to the size of Pull3.1.

3.5.2 Over-expression and purification of Pull3.1

Following expression and purification of the protein (Figure 3.7), a pure protein band was seen at between the 58 kDa to 80 kDa range which matches the estimated protein size. The α-amylases from *Geobacillus stearothermophilus* and *Laceyella sacchari* have a MW of 83 kDa and 79 kDa, respectively (Kuriki *et al.*, 1988; Odibo & Obi, 1988). The Pull3.1 protein was purified well enough to proceed with activity assays.

Size exclusion chromatography of Pull3.1 showed a peak at 28.39 min in Figure 3.10 compared to Figure 3.9, i.e., the BSA, (66 kDa) with a peak at 28.30 min (Thermo Fisher Scientific) and β-amylase (Figure 3.8), (112 kDa) with a peak at 18.55 min. Using these values to generate a slope intercept equation, Pull3.1's protein size is estimated at 66.19 kDa. The BLASTp analysis and Uniprot results indicated that homodimer formation is possible since the sequence of Pull3.1 contained a putative polypeptide binding site for the homodimer interface (Figure 3.5 B). The binding site overlaps with the catalytic triad as seen in Figure 3.5. This interface is also present in the neopullulanase of *G. stearothermophilus*, and the homodimer interface is essential for the formation of active enzymes and possibly affects substrate preference, since the active site cleft is narrower because of dimerisation (Hondoh *et al.*, 2003). However, this is not supported by the size exclusion analysis and would need to be further investigated through X-ray crystallography to reveal binding site positions (Wüthrich. 1990).

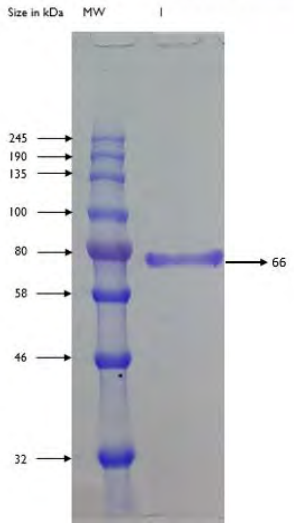


Figure 3. 2: 12% SDS-PAGE analysis of purified pETPull3.1. Lane MWM, protein standard (11-245 kDa); Lane 1, Pull3.1 ~ 66 kDa.

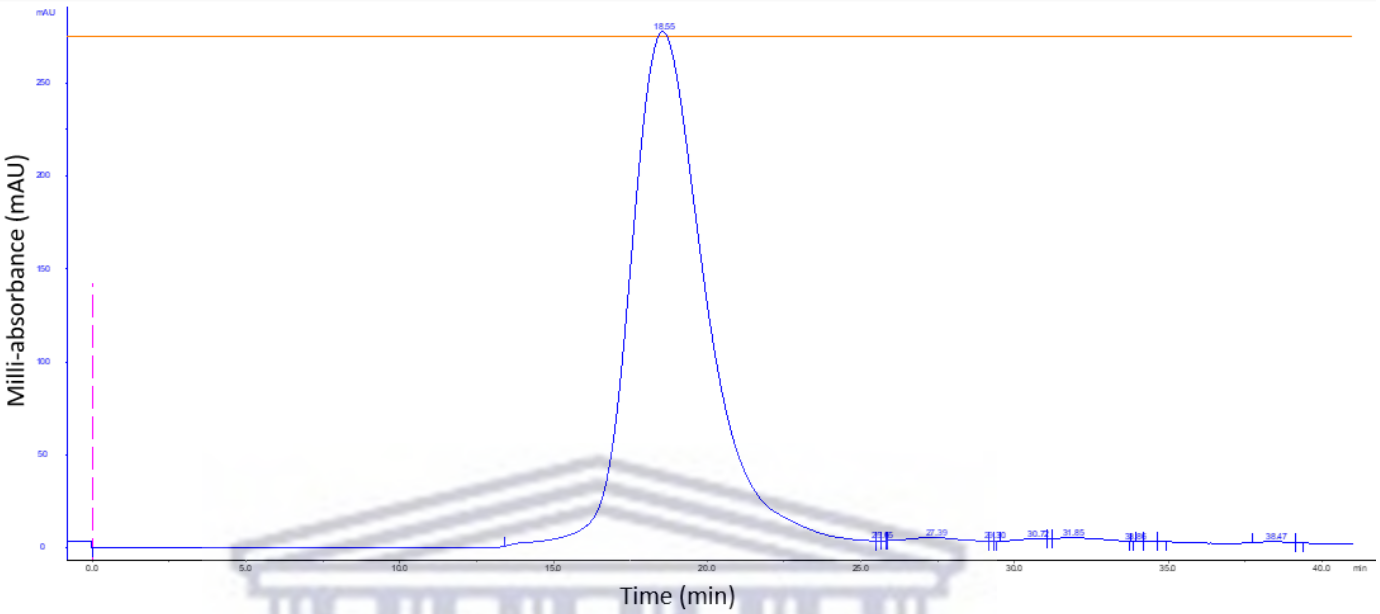


Figure 3. 8 FPLC analysis of the β-amylase standard used to estimate the molecular weight of Pull3.1. The BSA standard has a molecular weight of 112 kDa in this figure the β-amylase was detected at 18.55 min.

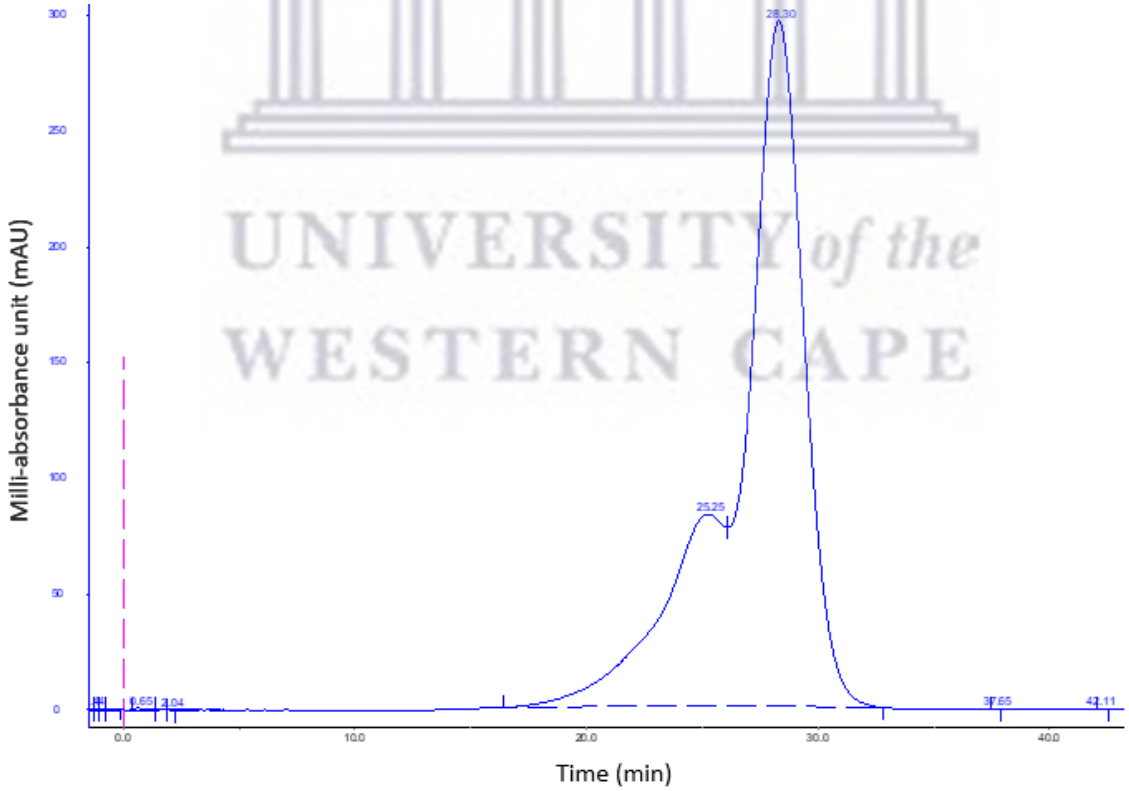


Figure 3. 9: FPLC analysis of the BSA standard used to estimate the molecular weight of Pull3.1. The BSA standard has a molecular weight of 66.5 kDa in this figure the BSA was detected at 28.30 min.

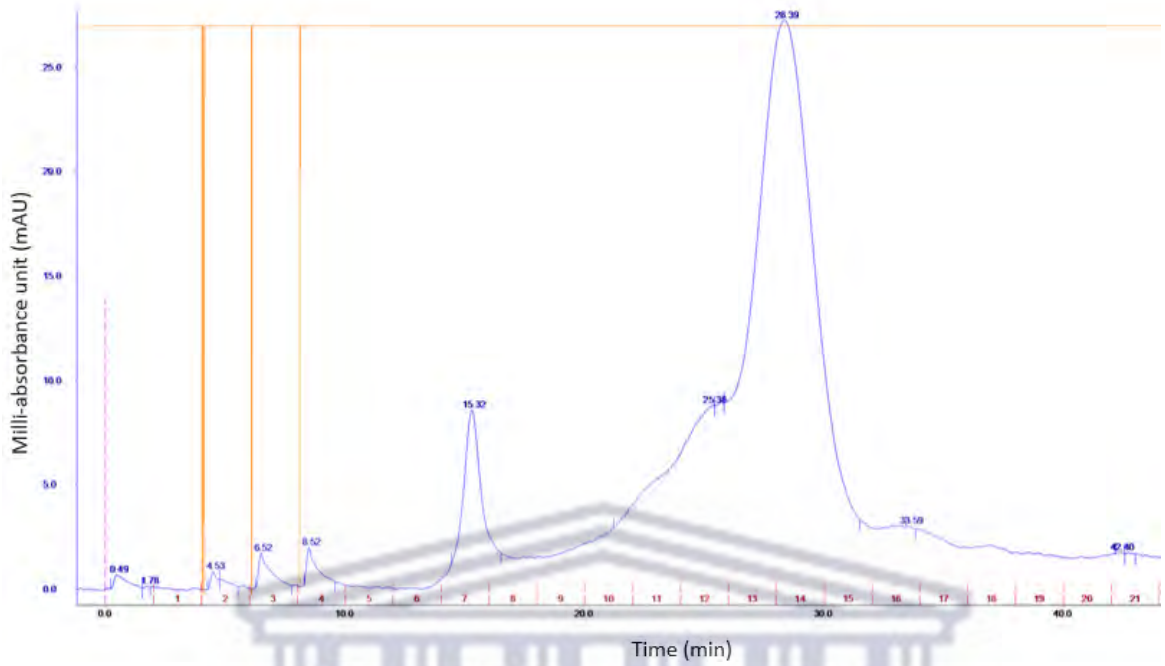


Figure 3. 10: FPLC analysis of Pull3.1 protein peaks for molecular weight estimation. Pull3.1 showed a peak at 28.39 min and a small buffer peak at 15.32 min.

3.6 Biochemical characterisation

Different tests were performed including substrate utilisation, temperature and pH optimum tests, enzyme kinetics and thermostability. The optimal conditions for Pull3.1 activity were determined.

3.6.1 Substrate utilisation

Pull3.1 displayed the highest specific activity on starch; for non-starch substrates, Pull3.1 performed best on pullulan and glycogen from oyster (Table 3.3). Furthermore, there was no activity detected on substrates with β-glycosidic bonds. This is a clear indication of Pull3.1's specificity for the hydrolysis of the α-glycosidic bond. Only pullulanase and CDase can hydrolyse bonds on starch as well as pullulan, therefore both these substrates were used for further testing. The PHTIII from *T. aggregans* and *D. mucosus* could not hydrolase glycogen, whereas it was possible for PHTIII from *T. kodakarensis* to do so (Ahmad *et al.*, 2014). As indicated in Figure 3.5, a glycogen binding site was identified in Pull3.1, which suggests it has the ability to use glycogen as a substrate. As seen in Table 3.3 the relative activity on glycogen substrates was 58.17% and 53.19% indicating that there is definite activity on these substrates and the glycogen binding site, confirming the in-silico

prediction. This might also reflect the higher sequence similarity between Pull3.1 and *T. kodakerenis* rather than to *T. aggregans* or *D. mucosus,* as seen in the phylogenetic tree (Figure 3.3). The test on cyclomaltodextrin also rules out the possibility of Pull3.1 act as a CDase as expected during the primary amino acid analysis (section 3.2).

The hydrolysis products of Pull3.1 activity on both starch and glycogen were assessed through HPLC analysis. Peaks at 24 min and 27 min, respectively, were observed (Figure 3.12). When comparing the chromatograms from the digested samples to the standards, the hydrolysis product at 27 min was maltose (G2). The product released from starch and glycogen at 24 min is possibly isomaltooligio saccharide (IMO) with a degree of polymerisation (DP) of 3 (Bai *et al.*, 2015). Maltose was the primary product released when starch was hydrolysed by the PHTIII from *T. aggregans*, although it was accompanied by glucose and maltotriose (Niehaus *et al.*, 2000). The pullulan hydrolysis products were converted to the peak seen at 24 min. This peak is also possibly an IMO, panose or isopanose, but it is not possible to differentiate between these products with ion exclusion chromatography (RSO oligosaccharide Ag^+ 4% column) (Corradini *et al.*, 2012).

Table 3. 3: Substrate preference of Pull3.1

Substrate	Type of glycosidic bond	Specific activity (µmol/min/mg of protein)	Relative activity (%)
Starch (potato)	α	0.82 ± 0.1	100.0
Starch (amylum)	α	0.43 ± 0.06	53.19
Amylose	α	0.44 ± 0.06	54.26
Pullulan	α	0.48 ± 0.06	58.47
Glycogen (oyster)	α	0.48 ± 0.07	58.17
Glycogen (bovine)	α	0.43 ± 0.06	53.19
Glucan (barley)	α	0.43 ± 0.08	52.17
Lichenan	α	0.43 ± 0.06	52.24
Mannan	α	0.43 ± 0.06	52.74
CMC	β	0	0
Cellulose	β	0	0
α-cyclodextrin	β	0	0

The specific activity value is based on three biological repeats, followed by the standard deviation (±). For the relative activity, the activity of Pull3.1 on soluble starch from potato, Sigma was used as the 100% measure. The reactions were incubated at 70ºC for 20 min.

High-performance anion-exchange chromatography (HPAEC) analysis could be used to determine the pullulan hydrolysis products produced based on its structure. As an alternative to HPAEC, digesting the pullulan hydrolysis product with glucoamylase would also indicate which one of the two isomers was released (Kuriki *et al.*, 1988). There is also a slight peak at 27 min for the pullulan hydrolysis product. This peak could be another maltose (G2) peak. The tiny peak at 32 min probably indicates glucose (G1), as this has also been observed with a CDase from *Lactobacillus plantarum* and the PHTIII from *T. aggregans* (Jang *et al.*, 2018; Niehaus *et al.*, 2000).

The smallest polymer that Pull3.1 was able to digest was maltotetraose, with maltose, maltotriose and some undigested maltotertaose as the end products (Figure 3.11-13). All three substrates were selected for further testing, the original polymer was converted entirely to hydrolysis products. From the hydrolysis products it appeared that Pull3.1 may be a multi-specific alpha-amylase with PHTIII-like (EC 3.2.1.41) properties. This assumption was made on the premise that α-amylase does not act on pullulan and rarely acts on CDs. All PHTIII and CDase enzymes can act on CDs, starch, and pullulan. When these two enzymes act on CDs and starch the by-products formed are maltose and when they act on pullulan, panose is formed (Park *et al.*, 2000). It is also evident that the smallest oligosaccharide hydrolysed is maltotetraose, since all the end products are either DP3 or smaller, the remaining DP4 was undigested maltotetraose which is similar to the PHTIII from *T. aggregans*; whereas, the smallest oligiosaccharide that the *T. kodakarensis* enzyme can act on is maltotriose (Ahmad *et al.*, 2014).

Another possibility is that Pull3.1 is a glucoamylase, since one of the byproducts formed by this enzyme is glucose, one of the byproducts formed by Pull3.1. It has been recognised, especially in the food industry, that if starch is hydrolysed using glucoamylase, it produces glucose, but once it exceeds the 30-35% threshold, the glucose is converted to maltose and isomaltose (Jòzef, 2007). This theory was tested by subjecting Pull3.1 to a 30% glucose reaction; however, no higher DP end products formed during this reaction. This suggests that Pull3.1 is unable to perform transglycosylation.

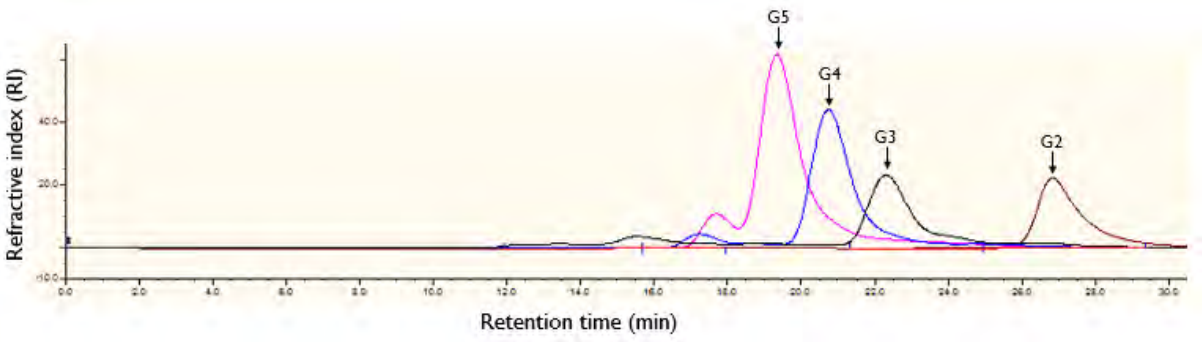


Figure 3. 11: HPLC analysis of 2 mM oligosaccharides used as standards. The labels indicate the number of glucose units and the retention time at which these units appear. The peaks indicate the following: G2, maltose; G3, maltotriose; G4, maltotetraose; G5, maltopentose.

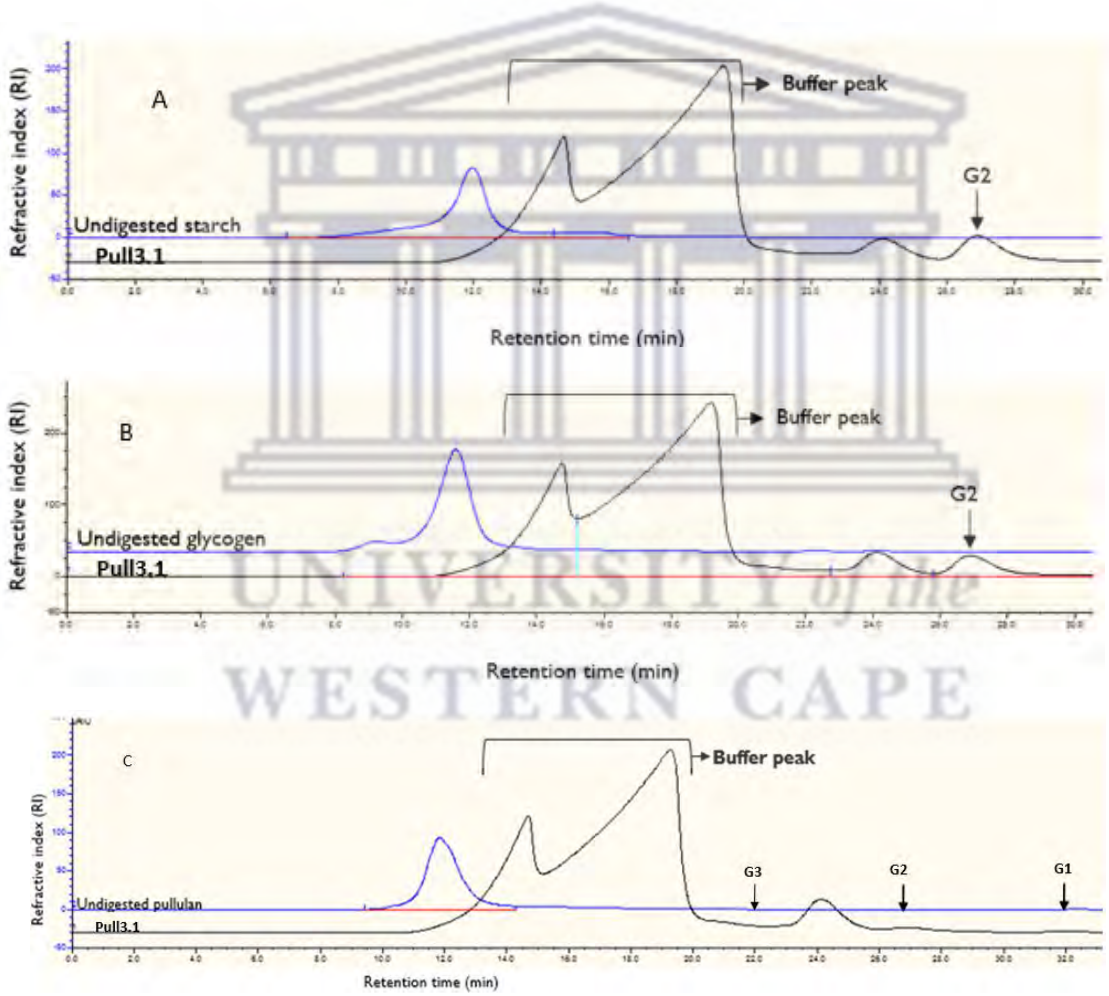


Figure 3. 3: HPLC chromatograms that indicate the hydrolysis products of starch (A), glycogen (B) and pullulan (C) after digestion with purified Pull3.1. The chromatograms also include the undigested sample. The peaks indicate the following: G1, glucose and G2, maltose. For starch the end products were G2 and a product between G2 and G3, similar end products were seen for glycogen in image B. Pullulan, image C also displayed these two peaks but with an additional peak at G1.

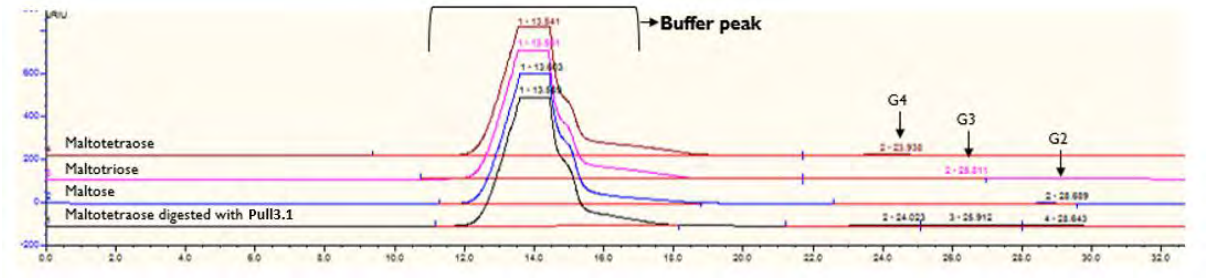


Figure 3. 4: Maltotetraose digested, the smallest sub-unit that Pull3.1 was able to digest. The peaks indicate the following: G2, maltose; G3, maltotriose; G4, maltotetraose. Pull3.1 did not digest all the maltotetraose and therefore a peak was seen at G4, but the rest of the maltotetraose was converted to G3 and G2.

3.6.2 Interaction, stability, and activity of Pull3.1 in the presence of various metals, detergents, and inhibitors

Metals are known to occasionally act as a stabilising factor that influences thermostability (Mrudula *et al.*, 2011; Rudiger *et al.*, 1995; Vieille & Zeikus, 2001). It is well-known that Ca^{2+} increases activity among pullulanase enzymes, while the activity for Pull3.1 remains almost unchanged with the addition of $CaCl_2$ (Table 3.4). The same results were previously observed by Cruz *et al.* (2018) when they tested the effect of $CaCl_2$ and NaCl on enzyme activity for α-amylase at values below 1 mM and above 5 mM (Cruz *et al.*, 2018).

Slight inhibition was noticed for $MgCl_2{\cdot}6H_2O$, $NiCl_2{\cdot}6H_2O$, KCl and NaCl for Pull3.1, while other studies have found Ni^{2+} and Zi^{2+} to be much more inhibitory (Deutch, 2002; Endo *et al.*, 2001). Pull3.1 only had 61,73% relative activity in $FeCl_3$, it appears to have an inhibiting effect on Pull3.1. Copper caused complete inhibition of Pull3.1, which is the case for all characterised pullulanases and CDases. It has been reported that cupric ions compete with biologically important metal ions. It is also the most competitive ion in the Irving Williams series and can cause reduced enzyme activity (Ndata *et al.*, 2021; Rubino & Franz, 2012).

Complete inhibition was also documented for enzymes from *Thermococcus* sp. CL1, i.e. *Geobacillus thermoleovorans*, *Thermotoga neapolitana* and *Thermus thermophilus* (Kang *et al.*, 2011; Lee *et al.*, 2013b; Wu *et al.*, 2014; Zouari Ayadi *et al.*, 2008). $ZnCl_2$ usually has an inhibiting effect on pullulanase enzymes and CDases, but this was not the case for Pull3.1, in this study a minor increase in activity has been observed (Duffner *et al.*, 2000; Kaulpiboon & Pongsawasdi, 2004).

Since pullulanases are ideal for use in the detergent industry, it is crucial that they are able to handle high surfactant concentrations (Ece *et al.*, 2014). The enzymes should be able to function at temperatures of 40-60°C and should be able to function in very alkaline conditions (Bertoldo *et al.*, 2004). Of the detergents tested, β-mercaptoethanol (β-ME) and dithiothreitol (DTT) caused a substantial increase in activity. This has been documented in other studies on pullulanase enzymes that showcased the use of thiol groups during enzyme activity (Kahar *et al.*, 2016; Rudiger *et al.*, 1995).

DTT and β-ME can cause a reduction of disulphide bonds in thiol groups of proteins in the cysteine, glutathione, or thiamine residues. This alteration in the tertiary structure may lead to a more favourable position for substrate binding to occur. The reconfiguration could help the substrate reach the active site cleft more easily (Cecarini *et al.*, 2007; Wermuth, 2003). EDTA, SDS, Triton X, Tween 20 and Tween 80 inhibited Pull3.1.

Table 3. 4: Effect of various cations, inhibitors, and detergents on the activity of Pull3.1

Cations, inhibitors and detergents	Concentration	Specific activity (µmol/min/mg of protein)	Relative activity (%)
None	-	0.7 ± 0.1	100.0
$CaCl_2$	1 mM	0.74 ± 0.1	102.01
	5 mM	0.73 ± 0.2	101.53
$MgC_{l2}·6H_2O$	1 mM	0.64 ± 0.2	90.16
	5 mM	0.57 ± 0.2	84.53
$NiCl_2 ·6H_2O$	1 mM	0.7 ± 0.2	95.02
	5 mM	0.68 ± 0.2	95.54
$CuCl_2·2H_2O$	1 mM	0	0
	5 mM	0	0
KCl	1 mM	0.74 ± 0.2	102.51
	5 mM	0.7 ± 0.2	98.02
NaCl	1 mM	0.74 ± 0.2	103.02
	5 mM	0.67 ± 0.2	96.18
$FeCl_3$	1 mM	0.4 ± 0.1	61.73
	5 mM	0.53 ± 0.2	75.42
$ZnCl_2$	1 mM	0.87 ± 0.1	114
	5 mM	0.89 ± 0.2	116.07
β-mercaptoethanol	1 mM	0.9 ± 0.2	119.12
	5 mM	1.26 ± 0.4	154.38
EDTA	1 mM	0.55 ± 0.1	79.97
	5 mM	0.47 ± 0.1	69.62
DTT	1 mM	1.23 ± 0.3	155.54
	5 mM	1.42 ± 0.1	174.48
Triton X	0.1% (v/v)	0.48 ± 0.1	79.63
	0.5% (v/v)	0.32 ± 0.2	91.88
Tween 20	0.1% (v/v)	0.49 ± 0.1	78.5
	0.5% (v/v)	0.25 ± 0.1	76.58
Tween 80	0.1% (v/v)	0.45 ± 0.1	73.28
	0.5% (v/v)	0.29 ± 0.2	85.01
SDS	0.1% (w/v)	0.19 ± 0.1	42.82
	0.5% (w/v)	0.12 ± 0.0	51.42

The specific activity value is followed by the standard deviation (±). For the relative activity, the activity of Pull3.1 on soluble starch from potato was used as the 100% measure.

3.6.3 Temperature and pH optimum

Industrial practicality can be estimated using the pH and temperature optimum. By way of example, certain industrial process steps can be conducted at elevated

temperatures, which eliminates the risk of bacterial contamination and enhances process kinetics (Seibert *et al.*, 2016). Another added advantage is the reduced cost, because less energy is required during steps such as filtration and pumping because of the lower viscosity achieved when performing activities at a higher temperature. Thermozymes are usually more stable than mesozymes. This stability might permit higher activity in the presence of organic solvents, detergents and a wider pH range, an ideal advantage for an enzyme used in household and industrial cleaning products (Bruins *et al.*, 2001; Józef Synowiecki, 2010).

Pull3.1 had an optimal pH of 7 in the 100 mM phosphate-citrate buffer and a wide pH range of 5-9 (Figure 3.14). An interesting increase in activity was observed at pH 10 when Pull3.1 was assayed in 100 mM glycine-NaOH buffer. The effect of the glycine buffer was also tested at different temperatures, but did not show an increase in activity. This suggests that glycine does not help to enhance activity at different temperatures and that this phenomenon is purely an occurrence related to glycine buffer pH 10. The ability of Pull3.1 to operate at >80% of its optimum activity at pH 10 might make this an attractive feature of the enzyme for industry.

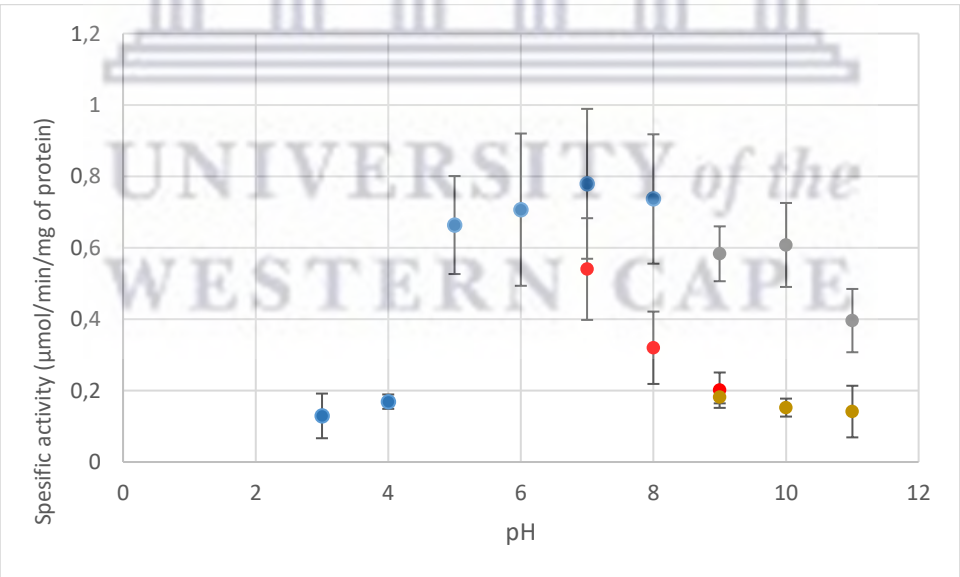


Figure 3. 5: Determination of the pH optimum for His-tag purified Pull3.1. The figure represents the mean of the three technical repeats (n=3) and the error bars indicate the standard deviation between samples, optimal pH for Pull3.1 to perform at is pH 7. The following buffers were tested: 100 mM phosphate-citrate buffer (blue dots pH 3-8); 100 mM tris-HCL buffer (red dots pH 7-9); 100mM glycine-NaOH buffer (grey dots pH 9-11); Na_2CO_3-$NaHCO_3$ Buffer (yellow dots pH 9-11).

The temperature optimum for Pull3.1 was between 60°C to 80°C, which is desirable for certain applications. Pull3.1 retained over 40% of its maximum activity over the 50°C to 85°C temperature range (Figure 3.15). Based on data from the BRENDA database the characterised pullulanase enzymes that showed the most similarity to Pull3.1 were pullulan 6-glucanohydrolase (EC 3.2.1.41) isolated from *G. stearothermophilus* and *L. sacchari.* Both enzymes have a pH optimum of 7, and a temperature optimum of 65°C and 70°C, respectively (Kuriki *et al.*, 1988; Odibo & Obi, 1988). A characterised CDase from *T. kodakarensis* had an optimum pH of 7.5 and a temperature optimum of 65°C (Sun *et al.*, 2015). However, Pull3.1 is not unique in its ability to function optimally at high temperatures, as other PHTIII enzymes characterised from *T. aggregans* and *D. mucosus* have been able to maintain an optimum temperature of 95°C and 85°C (Duffner *et al.*, 2000; Niehaus *et al.*, 2000).

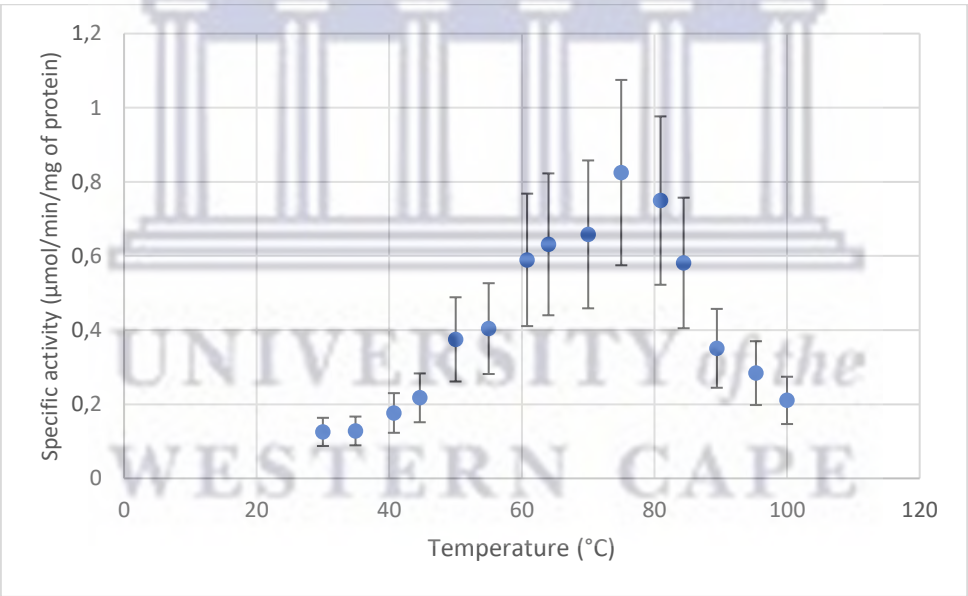


Figure 3. 6: Effect of temperature on the activity of His-tag purified Pull3.1 enzyme on starch. The data points represent the mean (n=3) and the error bars indicate the standard deviation between samples. The reaction was carried out in 100 mM phosphate-citrate buffer at pH 6 for 20 min.

3.6.4 Thermostability

It was hypothesised that Pull3.1 would be thermostable, since the sampling environment (Mphizi hot spring, Malawi) has a temperature range of 72-78°C (Hu, 2010). Pull3.1 was therefore assessed for its temperature optimum (Figure 3.15) as well as thermostability (Figure 3.16). Pull3.1 is thermostable and retained enzyme activity very well at 60°C and 70°C. At 80°C, the enzyme activity diminished drastically (Figure 3.16).

Protein structure and composition influences the thermostability of proteins. It has been reported that a high percentage of proline, arginine and tyrosine, and a high helical content can increase the thermostability of a protein (Guo *et al.*, 2018). In Pull3.1, there are 23 helices (Figure 3.18) that make up 26.9% of the protein content, alongside a lower serine content found in thermophiles, which is consistent with the results obtained from other thermophiles, such as TK-PUL (Table 3.5).

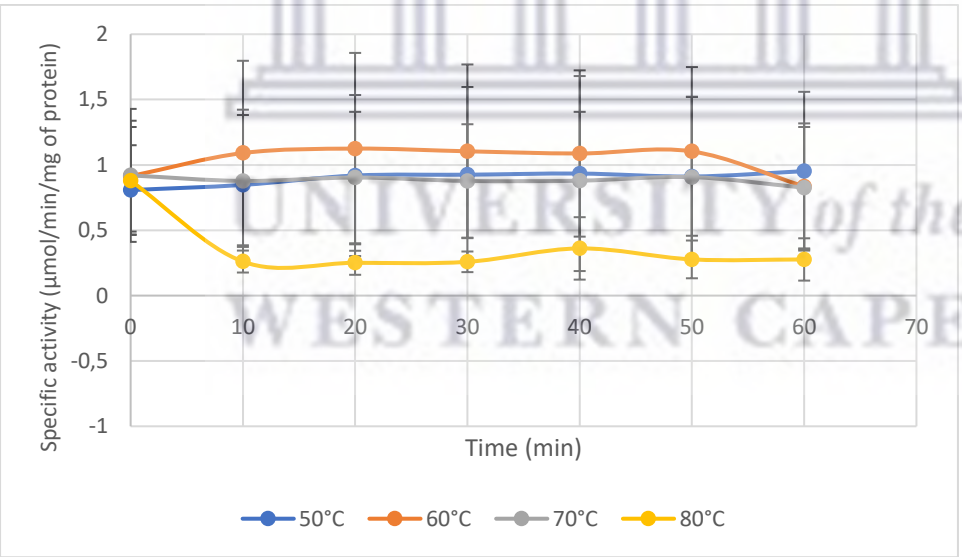


Figure 3. 16: Thermostability profile of Pull3.1. The reaction occurred in 100 mM phosphate-citrate buffer pH 6. The effect of temperature was analysed for over 60 min by taking measurements at 10 min intervals. The data points represent the mean (n=4); the error bars indicate the standard deviation between samples.

Proline helps with tighter packing of the hydrophobic core, strengthening loops and controlling the folding of the enzyme, which helps to prevent the backbone unfolding (Guo *et al.*, 2018; Kumar *et al.*, 2019; Suzuki *et al.*, 1991). There is also a hypothesis

that proline residues near the N-terminal contribute to thermostability (Takano *et al.*, 2009). The suggestion is that the residues found at the N-terminal cap, the second residue found at β-turns or the residues of the first turns in the α-helices help to achieve thermostability (Chen *et al.*, 2015; Guo *et al.*, 2018; Suzuki *et al.*, 1991; Takano *et al.*, 2009). Proline occurs at a higher percentage in thermophiles than in mesophiles (Table 3.5). This is also true for arginine and tyrosine. Arginine is regularly involved in the formation of salt bridges by employing the guanidinium side chain.

The amino acid composition and its effect on the thermal stability of the thermostable α-amylase from *B. licheniformis* has received more attention than other α-amylase enzymes (Declerck *et al.*, 2000; Hiteshi & Gupta, 2014). In terms of mutations performed on this enzyme, the most important domain involved in thermostability is domain B, where the Asn and Gln residues were replaced with Asn^{192}Phe (Hiteshi & Gupta, 2014). At the time that it was performed, it was the most thermostabilising mutation achieved with *B. licheniformis* α-amylase (BLA). When aligned to position 192 of *B. licheniformis* α-amylase, Pull3.1 is at position 339 with an Arg residue (Figure 3.17). These residues fall within a conserved region where glycogen recognition also occurs, and they do not agree with the BLA residues. However, further investigation using the other Asn/Gln residues in the Pull3.1 domain could also be performed. Substituting these residues could provide information regarding factors that influence the thermostability of Pull3.1.

It is possible that the thermostability of Pull3.1 could improve if the residues involved in deamination were replaced. The following residues found in α-amylase Termamyl LC^{TM} have been identified by Novozymes (Bisgaard-Frantzen *et al.*, 1999) as being able to increase thermostability: His156Tyr (H156Y) and Ala181Thr (A181T). Part of the investigation conducted by Bisgaard-Frantzan *et al*. (1999) was to substitute a histidine in domain B to reduce tension. This substitution with a tyrosine relieved the tension created because of protonation at position 156 at a low pH (Bisgaard-Frantzen *et al.*, 1999). The thermostability of all α-amylase enzymes could possibly be altered within domain B or the four conserved regions and result in significant changes to thermostability. It is also possible that the conserved tyrosine, tryptophan and phenylalanine found in regions 1 and 2 in Pull3.1 are responsible for the thermostability of Pull3.1. This speculation is based on the results of enzyme

treatment performed on *T. kodakarensis* by Guo *et al.* (2018) using N-bromosuccinimide, as described in section 1.4.2.2.

Figure 3. 17: Multiple sequence alignment of Pull3.1 to the *B. licheniformis* α-amylase (BLA). Indicated in yellow the serine overlaps and in green the arginine overlapping position the mutation sites responsible for the increased thermal stability in BLA.

The leading cause of increased stability during these mutations in thermostable α-amylase of *B. licheniformis* is salt bridge formation. According to Ban *et al*. (2021), a higher number of salt bridges of a suitable configuration at a suitable location has a significant effect on the thermostability of α-amylase enzymes because they are also important stabilisers of the protein's tertiary and quaternary structure (Borders *et al.*, 1994). According to the ESBRI (Evaluating the Salt BRIdges in Proteins), there were a total of 63 salt bridges in Pull3.1, all within a 4 Å distance. The distance of the salt bridges is important, as destabilisation can occur if the salt bridge exceeds a distance of 5 Å (Kumar & Nussinov, 1999). Protein stability *via* salt bridges is reliant upon where in the protein these salt bridges occur, the geometry, and the electrostatic interaction between residues. The salt bridges and hydrogen bonds create a more rigid structure that causes the protein to maintain its shape at lower temperatures but changes the conformation during higher temperatures, so that it becomes more active (Kumar & Nussinov, 1999). Many salt bridges in Pull3.1 exceeded the 5 Å distance, and none were in the positions in relation to the *B. licheniformis* α-amylase. Therefore it's not clear whether the salt bridges identified for Pull3.1 contribute to its thermostability (Ban *et al.*, 2021). Further investigation and mutations could be performed on the unique salt bridges found in Pull3.1, to determine if they have a stabilising effect.

Table 3. 5: Comparison of amino acid composition that contributes to thermostability between mesophilic and thermostable enzymes (Guo *et al.* 2018)

	Enzyme-content %	Helix	Proline	Arginine	Tyrosine	Serine
Thermophilic	Pull3.1	26.9	6.3	6.3	3.7	4.1
	TK-PUL	25	6.3	4.8	4.8	6.3
	1sma - Maltogenic amylases from *Thermus* sp.	23	6.0	6.0	5.1	2.9
	1j0h-Type I pullulan hydrolases *B. stearothermophilus*	23	6.0	5.8	5.4	3.6
Mesophilic	2yoc- Pullulanase from *Klebsiella oxytoca*	22	3.9	3.9	3.6	9.2
	2fh6- Pullulanase from *K. aerogenes*	25	4.8	4.1	3.6	9.3
	2wan- Pullulanase from *B. acidopullulyticus*	15	4.8	2.1	4.3	6.8

A)

B)

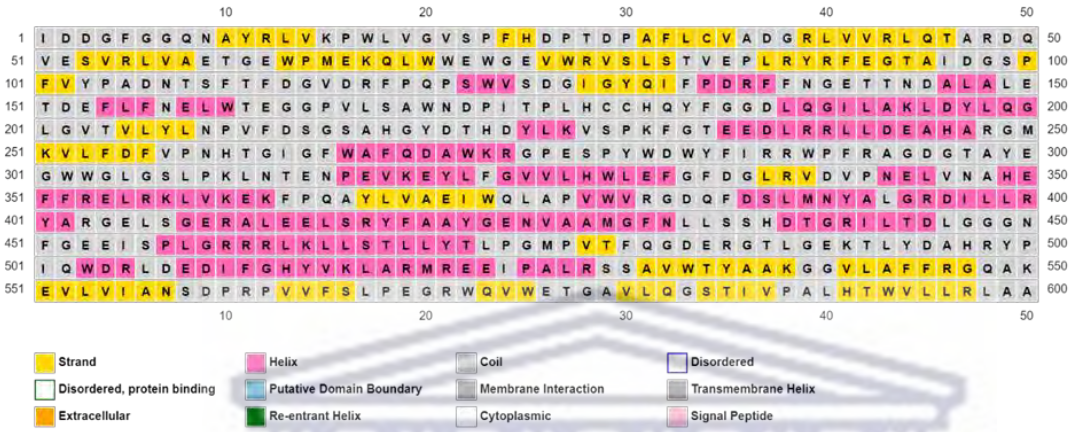


Figure 3. 18: Sequence plot of Pull3.1 generated by Psipred view. Psipred can only compute a limited number of amino acids; therefore, the sequence was run in sections A and B. This figure indicates the different helixes, strands, and coils within Pull3.1.

3.6.5 Enzyme kinetic characterisation

To date, only two PHTIII enzymes have been characterised (*T. aggregans* and *T. kodakarensis*) that allow for comparison with Pull3.1. Unfortunately, these PHTIII enzymes were only kinetically characterised on pullulan, but it was confirmed that PHTIII is suitable for starch conversion (Niehaus *et al.*, 2000; Toor *et al.*, 2021). Based on the Michaelis-Menton plot (Figure 3.18) it appears that Pull3.1 has a low K_M on starch (Table 3.6), which indicates good affinity for this substrate; and is substantially lower than that of other α-amylase enzymes and comparable to that of *C. flavus* (Table 3.6). This possible strong affinity for starch is also an indicator that Pull3.1 is most likely an α-amylase with low PHTIII activity. It is possible that if more α-amylase enzymes were assayed on pullulan, they too would show slight activity on pullulan and *vice versa*.

Table 3. 6: Kinetic parameters for Pull3.1 and other debranching enzymes acting on soluble starch as the substrate

Enzyme	Organism	Reference	K_M (mg/ml)	V_{max} (µmol ml^{-1})	K_{cat} (s^{-1})	K_{cat}/K_M (s^{-1} mg^{-1} ml)
α-amylase	*Streptomyces fragilis* DA7-7	(Nithya *et al.*, 2017)	0.000624	0.836	-	-
Pull3.1	Unknown	This study	0.048	0.729	0.007	0.141
α-amylase	*Cryptococcus flavus*	(Wanderley *et al.*, 2004)	0.057	-	-	-
α-amylase	*Thermotoga neapolitana*	(Park *et al.*, 2010)	0.07		0.02	0.286
α-amylase	*Bacillus subtilis* DM-03	(Das *et al.*, 2004)	1.2	1.96	-	-
PHTIII	*Thermococcus aggregans*	(Niehaus *et al.*, 2000)	3.72	22.7	0.283	0.076
CDase	*Thermococcus kodakarensis*	(Sun *et al.*, 2015)	3.1	34.6	34.41	11.1
CDase	*Microbacterium terrae* KNR 9	(Rajput *et al.*, 2016)	10	146	5.303	0.53
β-CDase	*Bacillus* sp.	(Upadhyay *et al.*, 2018)	2.613	0.309	0.003	0.001

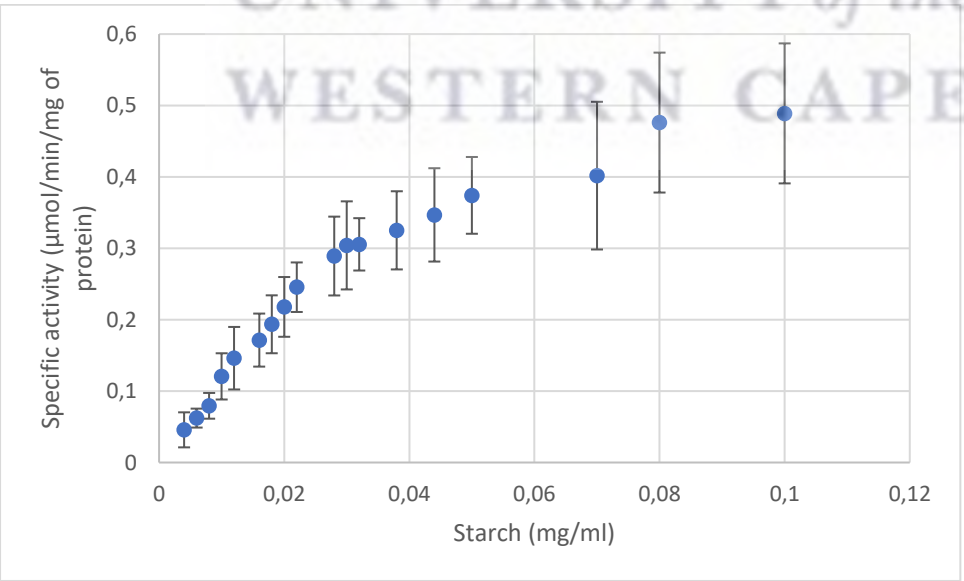


Figure 3. 19: Michaelis-Menten plot of Pull3.1 assayed on potato starch

As indicated in section 1.6, the presence of CBMs can indicate substrate specificity. Studies have shown that CBM48 evolved based on substrate specificities, rather than species (Machovič & Janeček, 2008). Because of their close structural and sequence similarities, Janeček *et al*. (2011) assumed that CBM20 and CBM48 evolved from a common ancestor. The hypothesis is that CBM48 was acquired by a protein without the need for CBM48 to be located in a specific region within the protein, i.e., it could be either before or after the catalytic region. It is then assumed that once the CBM has been acquired, evolution happens concurrently for both the enzyme and CBM. The result is that closely related proteins will also possibly contain closely related CBM modules (Janeček *et al.*, 2011; Machovič & Janeček, 2008). It is possible that the CBM for Pull3.1 substrate is specific for starch, pullulan, or glycogen, and therefore directs binding and conversion. However, Pull3.1 converts starch in preference to pullulan or glycogen. It may be that the active site could evolve to convert these substrates optimally due to these specific substrates being in the environment that the organism was sampled from.

The delay in reaching maximum activity has also been recorded in other studies that used sucrose as the substrate to determine α-amylase activity, and it has been hypothesised that high viscosity and substrate inhibition could be influencing factors (Chan *et al.*, 2021; Transtrum *et al.*, 2015). In this study, high viscosity was an issue when measuring Pull3.1 activity at increased concentrations of soluble starch. It should be noted that Pull3.1 was outside of the linear range while assayed, hence the values provided are estimations based on the available data.

Chapter 4: Conclusion

The food, cosmetic, detergent, and confectionary industries are constantly growing to meet ever-expanding demand. There will always be a search for the most affordable, least time-consuming, and most environmentally friendly method to satisfy the demand. Starch degrading methods are required by each of these industries, and enzymes, particularly those from thermophilic sources, have been relied upon to date (Lévêque *et al*., 2000).

Enzymes that function at a high temperature are usually retrieved from organisms that reside in hot environments; however, culturing these organisms under laboratory conditions has proven a challenge with most micro-organisms used (Staley & Konopka, 1985). Metagenomics offers a way to overcome the unculturability of some organisms by providing a means to study the genetic material directly isolated from the environment without the need to keep the host organism alive. Function-based screening is the preferred method of screening, since it allows the discovery of entirely novel proteins, but this method has certain drawbacks, as detailed in Chapter 1.

Sequence-based screening might not always lead to the discovery of completely novel proteins, but it does permit the discovery of novel functions from proteins in known families that may be able to tolerate extreme conditions. Sequence-based screening also allows for a search to be performed of the gene's products' mode of action by using conserved sequences as a guide for selection. Sequence-based screening relies on bioinformatic tools and has been used in various studies to estimate the type of protein, the protein family, and the three-dimensional structure, including the catalytic and binding sites. Once enough data has been retrieved to provide confidence in the type of enzyme, it is characterised further based on the initial bioinformatic background characterisation performed. This process also helps in planning the experimental design, since the background information hints at the substrate preference, as well as the temperature and pH ranges in which the organism would function well. Comparisons with other similar enzymes can also be accomplished easily. In this study, a sequence-based screen was used to select an enzyme from the Mphizi hot spring for characterisation.

The enzyme Pull3.1 has an optimum temperature range between 60°C to 80°C and an optimum pH of 7, which correlated with the sampling environment. At the time of

sampling the Mphizi hot spring recorded a temperature range of 72-78°C and a pH of 6.2 (Hu, 2010). As Handelsman (2004) stated, the sampling environment, and especially extreme environments, play a crucial role in discovering thermozymes, which was the case with this Pull3.1. The PHTIII from *T. aggregans* remains active at 100ºC. The same temperature optimum 95–100°C was found for PHTIII from *T. kodakarensis* (Guo *et al.*, 2018; Niehaus *et al.*, 2000). The inclusion of a His-tag can have an impact on the stability and function of some recombinant proteins (Booth *et al*. 2018). Since the temperature and pH parameters measured correlated with those of the Mphizi hot spring it is likely that the His-tag did not affect the functionality of the enzyme, but it remains to be tested and confirmed.

All three PHTIII enzymes characterised to date are thermozymes, but the two previously characterised enzymes are hyperthermozymes, since their optimal activities are above 80°C. As more thermozymes are added to the available databases, it becomes easier to study and understand how stability is achieved. Crystallization, site-directed mutation, biochemical assays and analysis of the structure and mechanisms of Pull3.1 may explain why Pull3.1 showed a sudden spike of activity when tested in a glycine pH 10 buffer. New tools such as Alphafold may help in predicting more novel protein structures such as Pull3.1.

The low K_M value of 0.048 mg/ml indicated that Pull3.1 had very high affinity for starch compared to some α-amylase enzymes, pullulanase and CDase, all of which had values above 2.6 mg/ml. The K_M values were compared to α-amylase from the BRENDA database, and enzymes with similar K_M values were found. Pull3.1 was the second highest of the nine α-amylase enzymes that were kinetically characterised on the database. For industrial purpose this will ensure that the enzyme will function at a constant rate regardless of the concentration of substrate and will be very advantageous when working with large volumes or concentrations (Robinson. 2015).

Pull3.1 showed the highest activity on starch, pullulan, and glycogen. As indicated in Chapter 3, these results agree with the bioinformatic analysis that Pull3.1 is probably an α-amylase with PHTIII characteristics. The sequence alignment also indicated a glycogen binding site that is evident in the substrate specificity results. The MSA also showed that the four conserved regions found in all GH13 enzymes are present. The sequence architecture from BLASTp and the Pfam results indicated that Pull3.1 had

domain hits for α-amylase, pullulanase and CDase, which all belong to GH13_20. α-Amylase was among the highest occurring domain hits.

The results from BLASTp indicate that Pull3.1 is relatively novel, since the protein that scored the highest sequence identity percentage (71.55%) was an uncharacterised pullulan hydrolase from an uncultured organism, possibly from the phylum *Bipolaricaulota*. The closest scoring characterised enzyme was an α-amylase from *Thermococcus* sp. EP1, which had a sequence identity score of 62.71%. Of the characterised enzymes included in the phylogenetic tree, Pull3.1 also grouped closest to the PHTIII's from the *Thermococcus* species. This type of enzyme is the only enzyme along with CDase that can act on both starch and pullulan, which is a known attribute of GH13_20. Only two of these types of enzymes have been characterised in the literature, which makes Pull3.1 the third enzyme with some PHTIII activity. The characterisation of Pull3.1 performed in this study suggests that this enzyme is probably an α-amylase; however, the bioinformatic analysis indicated that it is a pullulanase. This demonstrates the limitations of bioinformatic-based gene discovery, as nuanced amino acid sequence changes can result in altered substrate specificity.

The close relation of Pull3.1 to pullulanase enzymes offers the opportunity to learn which features of amylase and pullulanase determine substrate specificity. Niehaus *et al*. (2000) suggested that a slight alteration of sequence in one of the four conserved regions is what determines a substrate preference shift. The shift was identified as a change from AAQY to VANE (Niehaus *et al.*, 2000). The VANE sequence represents a pullulan hydrolase type II, while AAQY represents pullulan hydrolase type I. which indicates that the sequence from PHTIII, APQE is the result of a mutation. This was not the case with PHTIII from *T. kodakarensis*, since this PHTIII has the sequence VPNE, which is the same as Pull3.1. The enzyme from *T. kodakarensis* could hydrolyse both α-1,4- and α-1,6 glycosidic linkages (Guo *et al.*, 2018).

Further investigation could be performed on conserved region II to understand substrate specificity better. Another aspect worth investigating is the CBM, as mentioned in section 3.6.5. The CBM of this module is an indicator of substrate specificity, and it is thought that the module evolves with the active site. CBM48 is

closely related to CBM20, which favours starch or glycogen as opposed to CBM48 that favours glycogen and pullulan (Christiansen *et al.*, 2009). However, since the confidence level for CBM48 is very low, it could be that the actual CBM, that is yet to be identified, has a higher affinity for starch but is transitioning to ligand binding that favours pullulans and glycogen. Mutations that have been performed on CBM have also shown an increase in binding capacity (Armenta et al., 2017). Another aspect that CBM could provide information about is substrate specificity or tolerance of structural differences (Armenta *et al.*, 2017; Linder *et al.*, 1995).

An alternative interpretation of the discrepancy between bioinformatic classification and the biochemical assay results is a phenomenon called "annotation creep" (Randle-Boggis *et al.*, 2016). This results in the database containing proteins annotated as pullulanase enzymes that were annotated in relation to other hypothetical proteins which in turn are also annotated as such due to their relation to either a characterised enzyme or yet more hypothetical proteins, which were similarly annotated.

Characterisation experiments such as the thermal stability assay indicated that Pull3.1 maintained activity for at least an hour at temperatures between 40ºC and 70ºC. Together with the wide pH range and high temperature optimum, this makes Pull3.1 a good candidate for use in saccharification, but it will not perform well during liquefication because this starch degradation step takes place at 100ºC. Future testing could include upscaling to validate the use of Pull3.1 in the starch processing industry.

As described in Chapter 1, starch has multiple linkages and enzymes specific to each linkage. Some enzymes can have a dual mode of action, and these enzymes usually perform better than single substate enzymes. Pull3.1 offers a dual mode since it acts on starch and pullulan. An example of a multi-functional enzyme such as Pull3.1 was found by Lui *et al*. (2021), and this *B. subtilis* BS-5 strain showed α-amylase, endoglucanase and xylanase activity. The researchers noted that most of the GH13 sub-families to which Pull3.1 and the multi-functional enzyme from *B. subtilis* belong are usually monofunctional, but this is not the case with the GH13_20 sub-family (Liu *et al.*, 2021).

Another factor that influences the successful breakdown of plant matter is community mode action, where there is a synergistic flow of multiple enzymes working together, which has been reported with lignocellulosic enzymes (Conway *et al.*, 2018; Suleiman *et al.*, 2020). Further testing of Pull3.1 in combination with other starch degrading enzymes is, therefore, also suggested to explore the potential synergies and optimise the enzymatic breakdown of starch, enhancing overall efficiency in bioconversion processes.

It is well known that thermophilic organisms metabolise organic and inorganic carbon sources effectively; therefore, they could make use of a wide range of substrates. Accordingly, it was hypothesised that thermozymes from the glycoside hydrolase family can be found in hot springs (Baysal *et al.*, 2003; Debnath *et al.*, 2019; Kumar *et al.*, 2014). Some hyperthermophilic archaea produce α-amylase, which suggests that amylases play a significant role in their metabolism. Studies conducted on *Sulfolobus solfataricus* suggest the presence of a sensory system that the organism uses to detect exogenous starch (Haseltine *et al.*, 1996). Starch is an abundant, readily available energy source that can be found in plants that are able to survive in extreme environments or by photosynthesising algae (Brock, 1985). Many metagenomic studies have been performed on hot springs and new glycoside hydrolase types have been found as a result (Busch *et al.*, 2021; Chuzel *et al.*, 2018; Reichart *et al.*, 2021). Xiao Ping Hu (2010) stated that organic matter such as grass, plants, human waste, and animal waste surrounded the Mphizi hot spring. These rich organic sources provide hints as to which substrates are available for breakdown and guided the sequence-based screening performed in this study.

This study has again confirmed that metagenomic sequence-based screening of hot environments is an efficient method for discovering thermozymes that show potential for use in industrial applications. It has also broadened the library of enzymes found in a unique environmental niche.

Future work could include testing the synergistic effect of Pull3.1 in combination with α-amylase and pullulanase. Pull3.1 represents an appealing candidate for crystallisation research, because of its high affinity for starch. Research by Divne et al. (1994) employed X-ray crystallography to determine the three-dimensional structure of Cel7A at different temperatures. They found that at higher temperatures,

Cel7A undergoes structural changes in its quaternary structure, transitioning from a closed form to a more open, extended form. This conformational change allows for better accessibility of the cellulose substrate to the enzyme's active site, thus enhancing its catalytic efficiency at elevated temperatures. Overall, Cel7A serves as a prominent example of a glycoside hydrolase where the temperature profile is reflected in its quaternary structure. Such investigations may uncover small changes between similar proteins which have not been structurally solved, as well as provide insight into binding site properties., highlighting the importance of understanding the structural dynamics of enzymes for optimising their performance in industrial applications.

References

Abdelraheem, E.M.., Busch, H., Hanefeld, U. & Tonin, F. 2019. Biocatalysis explained: from pharmaceutical to bulk chemical production. *Reaction Chemistry and Engineering*. 4(11):1878–1894.

Adesioye, F.A., Makhalanyane, T.P., Vikram, S., Sewell, B.T., Schubert, W.D. & Cowana, D.A. 2018. Structural characterization and directed evolution of a novel acetyl xylan esterase reveals thermostability determinants of the carbohydrate esterase 7 family. *Applied and Environmental Microbiology*. 84(8):1–16.

Afshinnekoo, E., Meydan, C., Chowdhury, S., Jaroudi, D., Boyer, C., Bernstein, N., Maritz, J.M., Reevs, D., Gandara, J., Chhangawala, S. & Ahsanuddin, S., 2015. Geospatial resolution of human and bacterial diversity with city-scale metagenomics. *Cell Systems*. 1(1):72–87.

Ahmad, N., Rashid, N., Haider, M.S., Akram, M. & Akhtar, M. 2014. Novel maltotriose-hydrolyzing thermoacidophilic type III pullulan hydrolase from *Thermococcus kodakarensis*. *Applied and Environmental Microbiology*. 80(3):1108–1115.

Akeboshi, H., Tonozuka, T., Furukawa, T., Ichikawa, K., Aoki, H., H., Shimonishi, A., Nishikawa, A. & Sakano, Y., 2004. Insights into the reaction mechanism of glycosyl hydrolase family 49: Site-directed mutagenesis and substrate preference of isopullulanase. *European Journal of Biochemistry*. 271(22):4420–4427.

Al-Masaudi, S., El Kaoutari, A., Drula, E., Redwan, E.M., Lombard, V. & Henrissat, B. 2019. A metagenomics investigation of carbohydrate-active enzymes along the goat and camel intestinal tract. *International Microbiology*. 22(4):429–435.

Ali, A. & Ahmed, S. 2016. Natural polymers: an overview. *Natural Polymers: Derivatives, Blends and Composites*. *1(1)1-22.*

Allen B, Drake M, Harris N. & Sullivan T. 2017. Using KBase to assemble and annotate prokaryotic genomes. *Current Protocols in Microbiology.* 46(1):1E.13.

Almeida, J.M., Martini, V.P., Iulek, J., Alnoch, R.C., Moure, V.R., Mueller-Santos, M., Souza, E.M., Mitchell, D.A. & Krieger, N., 2019. Biochemical characterization and application of a new lipase and its cognate foldase obtained from a metagenomic library derived from fat-contaminated soil. *International Journal of Biological Macromolecules*. 137:442–454.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology*. 215(3):403–410.

Amin, K., Tranchimand, S., Benvegnu, T. & Abdel-razzak, Z. 2021. Glycoside hydrolases and glycosyltransferases from hyperthermophilic archaea : Insights on their characteristics and applications in biotechnology. *Biomolecules*. 11(11): 1557.

Angelov, A., Pham, V.T.T., Übelacker, M., Brady, S., Leis, B., Pill, N., Brolle, J., Mechelke, M., Moerch, M., Henrissat B. & Liebl, W. 2017. A metagenome-derived thermostable β-glucanase with an unusual module architecture which

defines the new glycoside hydrolase family GH148. *Nature scientific reports*. 7(1): 17306.

Ariaeenejad, S., Kavousi, K., Mamaghani, A.S.A., Motahar, S.F.S., Nedaei, H. & Salekdeh, G.H. 2021. In-silico discovery of bifunctional enzymes with enhanced lignocellulose hydrolysis from microbiota big data. *International Journal of Biological Macromolecules*. 177:211–220.

Armenta, S., Moreno-Mendieta, S., Sánchez-Cuapio, Z., Sánchez, S. & Rodríguez-Sanoja, R. 2017. Advances in molecular engineering of carbohydrate-binding modules. *Proteins: Structure, Function and Bioinformatics*. 85(9):1602–1617.

Bae, J., Lee, D., Kim, D., Cho, S.J., Park, J.E., Koh, S., Kim, J., Park, B.H., Choi, Y., Shin, H.J. & Hong, S.I., 2005. Facile synthesis of glucose-1-phosphate from starch by *Thermus caldophilus* GK24 α-glucan phosphorylase. *Process Biochemistry*. 40(12):3707–3713.

Bai, W., Fang, X., Zhao, W., Huang, S., Zhang, H. & Qian, M. 2015. Determination of oligosaccharides and monosaccharides in Hakka rice wine by precolumn derivation high-performance liquid chromatography. *Journal of Food and Drug Analysis*. 23(4):645–651.

Ban, X., Xie, X., Li, C., Gu, Z., Hong, Y., Cheng, L., Kaustubh, B. & Li, Z., 2021. The desirable salt bridges in amylases: Distribution, configuration and location. *Food chemistry*. 354(3):129475.

Baysal, Z., Uyar, F. & Aytekin, Ç. 2003. Solid state fermentation for production of α-amylase by a thermotolerant *Bacillus subtilis* from hot-spring water. *Process Biochemistry*. 38(12):1665–1668.

Bernfeld, P. 1955. Amylases, alpha and beta. *Methods in enzymology I*. I(540):149–158.

Bertoldo, C., Duffner, F., Jorgensen, P.L. & Antranikian, G. 1999. Pullulanase type I from *Fervidobacterium pennavorans* Ven5: Cloning, sequencing, and expression of the gene and biochemical characterization of the recombinant enzyme. *Applied and Environmental Microbiology*. 65(5):2084–2091.

Bertoldo, C., Armbrecht, M., Becker, F., Schäfer, T., Antranikian, G. & Liebl, W. 2004. Erratum: Cloning, sequencing, and characterization of a heat- and alkali-stable type I pullulanase from *Anaerobranca gottschalkii*. *Applied and Environmental Microbiology*. 70(8):5054.

Bibel, M., Brettl, C., Gosslar, U., Kriegshäuser, G. & Liebl, W. 1998. Isolation and analysis of genes for amylolytic enzymes of the hyperthermophilic bacterium *Thermotoga maritima*. *FEMS Microbiology Letters*. 158(1):9–15.

Bilal, M. & Iqbal, H.M. 2019. State-of-the-art strategies and applied perspectives of enzyme biocatalysis in food sector — current status and future trends. *Critical Reviews in Food Science and Nutrition*. 60(12):1–15.

Bisgaard-Frantzen, H., Svendsen, A., Norman, B., Pedersen, S., Kjaerulff, S., Outtrup, H. & Borchert, T.V., 1999. Development of Industrially Important alpha-amylases. *Journal of Applied Glycoscience*. 46(2):199–206.

Booth, W.T., Schlachter, C.R., Pote, S., Ussin, N., Mank, N.J., Klapper, V.,

Offermann, L.R., Tang, C., Hurlburt, B.K. & Chruszcz, M., 2018. Impact of an N-terminal polyhistidine tag on protein thermal stability. ACS omega, 3(1):760-768.

Boraston, A.B., Bolam, D.N., Gilbert, H.J. & Davies, G.J. 2004. Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochemical Journal*. 382(3):769–781.

Borchert, E., Selvin, J., Kiran, S.G., Jackson, S.A., O'Gara, F. & Dobson, A.D.W. 2017. A novel cold active esterase from a deep sea sponge *Stelletta normani* metagenomic library. *Frontiers in Marine Science*. 4(287):1–13.

Borders Jr, C.L., Broadwater, J.A., Bekeny, P.A., Salmon, J.E., Lee, A.S., Eldridge, A.M. & Pett, V.B., 1994. A structural role for arginine in proteins: Multiple hydrogen bonds to backbone carbonyl oxygens. *Protein Science*. 3(4):541–548.

Bouhajja, E., McGuire, M., Liles, M.R., Bataille, G., Agathos, S.N. & George, I.F. 2017. Identification of novel toluene monooxygenase genes in a hydrocarbon-polluted sediment using sequence- and function-based screening of metagenomic libraries. *Applied Microbiology and Biotechnology*. 101(2):797–808.

Bourne, Y. & Henrissat, B. 2001. Glycoside hydrolases and glycosyltransferases: families and functional modules. *Current Opinion in Structural Biology*. 11(5):593–600.

Božić, N., Lončar, N., Slavić, M. Š., & Vujčić, Z. 2017. Raw starch degrading α-amylases: an unsolved riddle. *Amylase*. *1*(1): 12–25.

Brock, T.D. 1985. Life at high temperatures. *Science*. 230(4722):132–138.

Bruijns, B., Tiggelaar, R. & Gardeniers, H. 2018. Massively parallel sequencing techniques for forensics: A review. *Electrophoresis*. 39(21):2642–2654.

Bruins, M.E., Janssen, A.E.M. & Boom, R.M. 2001. Thermozymes and their applications: A review of recent literature and patents. *Applied Biochemistry and Biotechnology - Part A Enzyme Engineering and Biotechnology*. 90(2):155–186.

Buchan, D.W., Minneci, F., Nugent, T.C., Bryson, K. & Jones, D.T., 2013. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic acids research, 41(1):349-357.

Buermans, H.P.J. & den Dunnen, J.T. 2014. Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta - Molecular Basis of Disease*. 1842(10):1932–1941.

Buléon, A., Colonna, P., Planchot, V. & Ball, S. 1998. Starch granules: Structure and biosynthesis. *International Journal of Biological Macromolecules*. 23(2):85–112.

Busch, P., Suleiman, M., Schäfers, C. & Antranikian, G. 2021. A multi-omic screening approach for the discovery of thermoactive glycoside hydrolases. *Extremophiles*. 25(2):101–114.

Campbell, J.A., Davies, G.J., Bulone, V. & Henrissat, B. 1998. Correction: A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochemical Journal*. 329(3):719.

Canganella, F., Andrade, C.M. & Antranikian, G. 1994. Characterization of amylolytic and pullulytic enzymes from thermophilic archaea and from a new *Fervidobacterium* species. *Applied Microbiology and Biotechnology*. 42(3):239–245.

Cantarel, B.I., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. 2009. The carbohydrate-active enzymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Research*. 37(1):233–238.

Casa-Villegas, M., Marín-Navarro, J. & Polaina, J. 2018. Amylases and related glycoside hydrolases with transglycosylation activity used for the production of isomaltooligosaccharides. *Amylase*. 2(1):17–29.

Castilla, I.A., Woods, D.F., Reen, F.J. & O'Gara, F. 2018. Harnessing marine biocatalytic reservoirs for green chemistry applications through metagenomic technologies. *Marine drugs*. 16(7):1-227.

Cecarini, V., Gee, J., Fioretti, E., Amici, M., Angeletti, M., Eleuteri, A.M. & Keller, J.N., 2007. Protein oxidation and cellular homeostasis: Emphasis on metabolism. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research.* 1773(2):93–104.

Chan, W.K.D., Hicks, B., Kaliappan, A., Garbett, N.C., Hansen, L.D. & Kenealey, J.D. 2021. Employing calorimetric methods to determine the mechanism of the invertase maximal activity delay. *ACS Food Science & Technology*. 1(1):60–65.

Chen, A., Li, Y., Nie, J., McNeil, B., Jeffrey, L., Yang, Y. and Bai, Z., 2015. Protein engineering of *Bacillus acidopullulyticus* pullulanase for enhanced thermostability using in silico data driven rational design methods. *Enzyme and Microbial Technology*. 78:74–83.

Chen, I.M.A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., Hajek, P., Ritter, S., Varghese, N., Seshadri, R. & Roux, S., 2021. The IMG/M data management and analysis system v.6.0: New tools and advanced capabilities. *Nucleic Acids Research*. 49(D1):D751–D763.

Choudhury, A.K.R., 2020. Introduction to enzymes. *Sustainable Technologies for Fashion and Textiles.* 4:75-90.

Christiansen, C., Abou Hachem, M., Janeček, Š., Viksø-Nielsen, A., Blennow, A. & Svensson, B. 2009. The carbohydrate-binding module family 20 - Diversity, structure, and function. *FEBS Journal*. 276(18):5006–5029.

Chuzel, L., Ganatra, M.B., Rapp, E., Henrissat, B. & Taron, C.H. 2018. Functional metagenomics identifies an exosialidase with an inverting catalytic mechanism that defines a new glycoside hydrolase family (GH156). *Journal of Biological Chemistry*. 293(47):18138–18150.

Conway, J.M., Crosby, J.R., Hren, A.P., Southerland, R.T., Lee, L.L., Lunin, V.V., Alahuhta, P., Himmel, M.E., Bomble, Y.J., Adams, M.W. & Kelly, R.M., 2018 Novel multidomain, multifunctional glycoside hydrolases from highly lignocellulolytic *Caldicellulosiruptor* species. *AIChE Journal*. 64(12):4218–4228.

Cordeiro, A.L., Lenk, T. & Werner, C. 2011. Immobilization of *Bacillus licheniformis* α-amylase onto reactive polymer films. *Journal of Biotechnology*. 154(4):216–

221.

Corradini, C., Cavazza, A. & Bignardi, C. 2012. High-performance anion-exchange chromatography coupled with pulsed electrochemical detection as a powerful tool to evaluate carbohydrates of food interest: principles and applications. *International Journal of Carbohydrate Chemistry*. 2012:1–13.

Cowan, D.A., Arslanoglu, A., Burton, S.G., Baker, G.C., Cameron, R.A., Smith, J.J. & Meyer, Q., 2004. Metagenomics, gene discovery and the ideal biocatalyst. *Biochemical Society Transactions*. 32(2):298–302.

Craig, J.W., Chang, F.Y., Kim, J.H., Obiajulu, S.C. & Brady, S.F., 2010. Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. *Applied and Environmental Microbiology*. 76(5):1633–1641.

Crini, G., 2014. Review: A history of cyclodextrins. *Chemical Reviews*. 114(21):10940–10975.

Santos, F.C.D. & Barbosa-Tessmann, I.P., 2019. Recombinant expression , purification , and characterization of a cyclodextrinase from *Massilia timonae*. *Protein Expression and Purification*. 154:74–84.

Cruz, W.O., Sinhori, G.G.C., de Lima, C.A.R. & Pontes, E.G., 2018. Biochemical properties of α-amylase from midgut of *Alphitobius diaperinus* (Panzer)(Coleoptera: Tenebrionidae) larvae. *Neotropical Entomology*. 47(5):698–708.

Culligan, E.P., Sleator, R.D., Marchesi, J.R. & Hill, C.. 2014. Metagenomics and novel gene discovery: Promise and potential for novel therapeutics. *Virulence*. 5(3):399-412.

Cusabio technology LLC. 2020. *Buffer calculator*. https://www.cusabio.com/m-296.html#a15. Date of access:10 June 2022.

Dalkiran, A., Rifaioglu, A.S., Martin, M.J., Cetin-Atalay, R., Atalay, V. & Doğan, T., 2018. ECPred: A tool for the prediction of the enzymatic functions of protein sequences based on the EC nomenclature. *BMC Bioinformatics*. 19(1):1–13.

Dalmaso, G.Z.L., Ferreira, D. & Vermelho, A.B., 2015. Marine extremophiles a source of hydrolases for biotechnological applications. *Marine Drugs*. 13(4):1925–1965.

Damme, R. Van, Hölzer, M., Viehweger, A., Müller, B., Bongcam-Rudloff, E. & Brandt, C., 2021. Metagenomics workflow for hybrid assembly, differential coverage binning, transcriptomics and pathway analysis (MUFFIN). *PLoS Computational Biology*. 17(2):e1008716.

Das, K., Doley, R. & Mukherjee, A.K., 2004. Purification and biochemical characterization of a thermostable, alkaliphilic, extracellular α-amylase from *Bacillus subtilis* DM-03, a strain isolated from the traditional fermented food of India. *Biotechnology Applied Biochemistry*. 40:291–298.

de Castro, A., Fernandes, G. & Franco, O., 2014. Insights into novel antimicrobial compounds and antibiotic resistance genes from soil metagenomes. *Frontiers in Microbiology*.https://www.frontiersin.org/journals/microbiology/articles/10.3389/f

micb.2014.00489/full. Date of access: 26 January 2024.

de Souza, P.M., & de Oliveira Magalhães, P., 2010. Application of microbial - amylase in industry – A review. *Brazilian Journal of Microbiology*. 41:850–861.

Debnath, T., Kujur, R.R.A., Mitra, R. & Das, S.K., 2019. Diversity of microbes in hot springs and their sustainable use. *Microbial Diversity in Ecosystem Sustainability and Biotechnological Applications*. 1:159–186.

Declerck, N., Machius, M., Wiegand, G., Huber, R. & Gaillardin, C., 2000. Probing structural determinants specifying high thermostability in *Bacillus licheniformis* α-amylase. *Journal of Molecular Biology*. 301(4):1041–1057.

Declerck, N., Machius, M., Joyet, P., Wiegand, G., Huber, R. & Gaillardin, C., 2003. Hyperthermostabilization of *Bacillus licheniformis* α-amylase and modulation of its stability over a 50°C temperature range. *Protein Engineering*. 16(4):287–293.

Demirjian, D.C., Morís-Varas, F. & Cassidy, C.S. 2001. Enzymes from extremophiles. *Current Opinion in Chemical Biology*. 5(2):144–151.

Deutch, C.E., 2002. Characterization of a salt-tolerant extracellular a-amylase from Bacillus. *Letters in applied Microbiology.* 35(1):78-84.

Daintith, J. & Martin, E.A. eds., 2010. *A dictionary of science.* USA: Oxford University Press.

Dhar, R., Seethy, A., Pethusamy, K., Singh, S., Rohil, V., Purkayastha, K., Mukherjee, I., Goswami, S., Singh, R., Raj, A. & Srivastava, T., 2019. De novo assembly of the Indian blue peacock (*Pavo cristatus*) genome using Oxford Nanopore technology and Illumina sequencing. *GigaScience*. 8(5):1–13.

Dickens, C.W.S. & Graham, P.M., 2002. The south african scoring system (SASS) version 5 rapid bioassessment method for rivers. *African Journal of Aquatic Science*. 27(1):1–10.

Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J.K., Teeri, T.T. & Jones, T.A., 1994. The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from Trichoderma reesei. *Science*. *265*(5171):524-528.

Divoll, T.J., Brown, V.A., Kinne, J., McCracken, G.F. & O'Keefe, J.M. 2018. Disparities in second-generation DNA metabarcoding results exposed with accessible and repeatable workflows. *Molecular Ecology Resources*. 18(3):590–601.

Dong, G., Vieille, C. & Zeikus, J.G. 1997. Cloning, sequencing, and expression of the gene encoding amylopullulanase from *Pyrococcus furiosus* and biochemical characterization of the recombinant enzyme. *Applied and Environmental Microbiology*. 63(9):3577–3584.

Duffner, F., Bertoldo, C., Andersen, J.T., Wagner, K. & Antranikian, G. 2000. A new thermoactive pullulanase from *Desulfurococcus mucosus*: Cloning, sequencing, purification, and characterization of the recombinant enzyme after expression in Bacillus subtilis. *Journal of Bacteriology*. 182(22):6331–6338.

Eberhardt, M.F., Irazoqui, J.M. & Amadio, A.F. 2021. β-Galactosidases from a

sequence-based metagenome: cloning, expression, purification and characterization. *Microorganisms*. 9(1):1–14.

Ece, S., Evran, S., Janda, J.O., Merkl, R. & Sterner, R. 2014. Improving thermal and detergent stability of *Bacillus stearothermophilus* neopullulanase by rational enzyme design. *Protein Engineering, Design and Selection*. 28(6):147–151.

Ekkers, D.M., Cretoiu, M.S., Kielak, A.M. & Van Elsas, J.D. 2012. The great screen anomaly-a new frontier in product discovery through functional metagenomics. *Applied Microbiology and Biotechnology*. 93(3):1005–1020.

Elleuche, S., Qoura, F.M., Lorenz, U., Rehn, T., Brück, T. & Antranikian, G. 2015. Cloning, expression and characterization of the recombinant cold-active type-I pullulanase from *Shewanella arctica*. *Journal of Molecular Catalysis B: Enzymatic*. 116:70–77.

Ellilä, S., Bromann, P., Nyyssönen, M., Itävaara, M., Koivula, A., Paulin, L. & Kruus, K., 2019. Cloning of novel bacterial xylanases from lignocellulose-enriched compost metagenomic libraries. *AMB Express*. 9(1)1-12.

Endo, K., Hakamada, Y., Takizawa, S., Kubota, H., Sumitomo, N., Kobayashi, T. & Ito, S., 2001. A novel alkaline endoglucanase from an alkaliphilic *Bacillus* isolate: Enzymatic properties, and nucleotide and deduced amino acid sequences. *Applied Microbiology and Biotechnology*. 57(1):109–116.

ENVGEN. 2017. *Metagenomics 101*. http://envgen.github.io/metagenomics.html Date of access: 18 Feb. 2020.

Escobar-Zepeda, A., De León, A.V.P. & Sanchez-Flores, A., 2015. The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. *Frontiers in Genetics*. 6(348):1–15.

Esmaeili, S. & Noorolahi, Z., 2017. Application of starch processing enzymes in food technology-a review. *Carpathian Journal of Food Science and Technology*. 9(3):114–127.

Falarz, L.J., Deyholos, M.K. & Chen, G., 2018. Plant carbohydrates and production of renewable biofuel from starch, sugar, and cellulose. *Plant Bioproducts*. 1–268.

Far, B.E., Ahmadi, Y., Khosroushahi, A.Y. & Dilmaghani, A., 2020. Microbial alpha-amylase production: Progress, challenges and perspectives. *Advanced Pharmaceutical Bulletin*. 10(3):350–358.

Fernandes, P., 2014. Marine enzymes and food industry: Insight on existing and potential interactions. *Frontiers in Marine Science*. 1(46):1–18.

Ferrandi, E.E., Sayer, C., Isupov, M.N., Annovazzi, C., Marchesi, C., Iacobone, G., Peng, X., Bonch-Osmolovskaya, E., Wohlgemuth, R., Littlechild, J.A. & Monti, D., 2015. Discovery and characterization of thermophilic limonene-1,2-epoxide hydrolases from hot spring metagenomic libraries. *FEBS Journal*. 282(15):2879–2894.

Ferrandi, E.E., Previdi, A., Bassanini, I., Riva, S., Peng, X. & Monti, D., 2017. Novel thermostable amine transferases from hot spring metagenomes. *Applied Microbiology and Biotechnology*. 101(12):4963–4979.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M. & McKenney, K., 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*. 269(5223):496–512.

Fogarty, W.M. & Kelly, C.T., 1990. Recent advances in microbial amylases. *Microbial Enzymes and Biotechnology*. 1(1): 71–132

Francisco, J.A., Stathopoulos, C., J., W.A., Kilburn, D.G. & Georgiou, G., 1993. Specific adhesion and hydrolysis of cellulose by intact *Escherichia coli* expressing surface anchored cellulase or cellulose binding domains. *Nature Biotechnology*. (11):491–495.

Frandsen, T.P. & Svensson, B., 1998. Plant α-glucosidases of the glycoside hydrolase family 31. Molecular properties, substrate specificity, reaction mechanism, and comparison with family members of different origin. *Plant Molecular Biology*. 37(1): 1-13.

Fredslund, F., Borchert, M.S., Poulsen, J.C.N., Mortensen, S.B., Perner, M., Streit, W.R. & Leggio, L.L., 2018. Structure of a hyperthermostable carbonic anhydrase identified from an active hydrothermal vent chimney. *Enzyme and Microbial Technology*. 114:48–54.

Frommhagen, M., Sforza, S., Westphal, A.H., Visser, J., Hinz, S.W., Koetsier, M.J., van Berkel, W.J., Gruppen, H. & Kabel, M.A., 2015. Discovery of the combined oxidative cleavage of plant xylan and cellulose by a new fungal polysaccharide monooxygenase. *Biotechnology for Biofuels*. 8(1):4–15.

Gabor, E.M., Alkema, W.B.L. & Janssen, D.B., 2004. Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environmental Microbiology*. 6(9):879–886.

Ganguly, A., Bala, E., Banerjee, P. & Blanchard, R., 2017. Production of ethanol from waste potato using locally available biocatalyst. In *Utilization and Management of Bioresources: Proceedings of 6th IconSWM 2016*. 143-151.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S.E., Wilkins, M.R., Appel, R.D. & Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. 1:571-607.

Garron, M.L. & Henrissat, B., 2019. The continuing expansion of CAZymes and their families. *Current Opinion in Chemical Biology*. 53:82–87.

Generalic, E. *Starch*. *Croatian-English Chemistry Dictionary & Glossary.* https://glossary.periodni.com Date of access: 07 May 2020.

Ghurye, J. & Pop, M., 2019. Modern technologies and algorithms for scaffolding assembled genomes. *PLoS Computational Biology*. 15(6):1–20.

Gilbert, J.A. & Dupont, C.L., 2011. Microbial Metagenomics: Beyond the Genome. *Annual Review of Marine Science*. 3(1):347–371.

Girotto, S., Pizzi, C. & Comin, M., 2016. MetaProb: Accurate metagenomic reads binning based on probabilistic sequence signatures. *Bioinformatics*. 32(17):i567–i575.

Giuntini, E.B., Sardá, F.A.H. & de Menezes E.W., 2022 The effects of soluble dietary fibers on glycemic response: an overview and futures perspectives. *Foods.* 11(23): 3934.

Guillén, D., Sánchez, S. & Rodríguez-Sanoja, R., 2010. Carbohydrate-binding domains: Multiplicity of biological roles. *Applied Microbiology and Biotechnology*. 85(5):1241–1249.

Guo, J., Coker, A.R., Wood, S.P., Cooper, J.B., Keegan, R.M., Ahmad, N., Muhammad, M.A., Rashid, N. & Akhtar, M., 2018. Structure and function of the type III pullulan hydrolase from *Thermococcus kodakarensis. Acta Crystallographica Section D: Structural Biology*. 74(4):305-314.

Gupta, A.K. & Gupta, U.D., 2014. Next Generation Sequencing and Its Applications. *Animal Biotechnology. 345-367.*

Gupta, R., Gigras, P., Mohapatra, H., Goswami, V.K. & Chauhan, B., 2003. Microbial α-amylases: A biotechnological perspective. *Process Biochemistry*. 38(11):1599–1616.

Halima, N.B., Khemakhem, B., Fendri, I., Ogata, H., Baril, P., Pichon, C. & Abdelkafi, S., 2016. Identification of a new oat β-amylase by functional proteomics. *Biochimica et Biophysica Acta - Proteins and Proteomics*. 1864(1):52–61.

Hall TA., 1999. "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT." In *Nucleic acids symposium series* 41(41): 95-98.

Hamre, A.G. & Sørlie, M., 2020. Kinetic relationships with processivity in *Serratia marcescens* family 18 glycoside hydrolases. *Biochemical and Biophysical Research Communications*. 521(1):120–124.

Han, T., Zeng, F., Li, Z., Liu, L., Wei, M., Guan, Q., Liang, X., Peng, Z., Liu, M., Qin, J. & Zhang, S., 2013. Biochemical characterization of a recombinant pullulanase from *Thermococcus kodakarensis* KOD1. *Letters in Applied Microbiology*. 57(4):336–343.

Handelsman, J., 2004. Metagenomics: Application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews*. 68(4):669–685.

Harada, T., Yokobayashi, K. & Misaki, A., 1968. Formation of isoamylase by *Pseudomonas*. *Applied Microbiology*. 16(10):1439–1444.

Haseltine, C., Rolfsmeier, M. & Blum, P., 1996. The glucose effect and regulation of α-amylase synthesis in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Journal of Bacteriology*. 178(4):945–950.

Hashimoto, H., 2006. Recent structural studies of carbohydrate-binding modules. *Cellular and Molecular Life Sciences*. 63(24):2954–2967.

Haynes, E., Jimenez, E., Pardo, M.A. & Helyar, S.J., 2019. The future of NGS (Next Generation Sequencing) analysis in testing food authenticity. *Food Control*. 101(2):134–143.

Heather, J.M. & Chain, B., 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics*. 107(1):1–8.

Hehre, E.J., Okada, G. & Genghof, D.S., 1969. Configurational specificity: unappreciated key to understanding enzymic reversions and de novo glycosidic bond synthesis: I. reversal of hydrolysis by α-, β-and glucoamylases with donors of correct anomeric form. *Archives of Biochemistry and Biophysics* 135(1):74-89.

Henrissat, B. & Davies, G., 1997. Structural and sequence-based classification of glycoside hydrolases. *Current Opinion in Structural Biology*. 7:637–644.

Hii, S.L., Tan, J.S., Ling, T.C. & Ariff, A. Bin., 2012. Pullulanase: role in starch hydrolysis and potential industrial applications. *Enzyme Research*. 2012:1–14.

Hiteshi, K. & Gupta, R., 2014. Thermal adaptation of α-amylases: a review. *Extremophiles*. 18(6):937–944.

Hondoh, H., Kuriki, T. & Matsuura, Y., 2003. Three-dimensional structure and substrate binding of *Bacillus stearothermophilus* neopullulanase. *Journal of Molecular Biology*. 326(1):177–188.

Hopkins, R.H., 2006. The actions of the amylases. *Biochemical Journal*. 6:389–414.

Hu, X.P., 2010. Identification and characterisation of novel cellulolytic genes using metagenomics. Masters thesis. University of the Western Cape.

Hua, M., Zhao, S., Zhang, L., Liu, D., Xia, H., Li, F. & Chen, S., 2015. Direct detection, cloning and characterization of a glucoside hydrolase from forest soil. *Biotechnology Letters*. 37(6):1227–1232.

Humblot, C., Turpin, W., Chevalier, F., Picq, C., Rochette, I. & Guyot, J.P., 2014. Determination of expression and activity of genes involved in starch metabolism in *Lactobacillus plantarum* A6 during fermentation of a cereal-based gruel. *International Journal of Food Microbiology*. 185:103–111.

Isaksen , M.F., Cowieson, A.J. & Kragh, K.M., 2011. Starch-and protein-degrading enzymes: biochemistry, enzymology and characteristics relevant to animal feed use*. Enzymes in farm animal nutrition (2011): 85-95.*

Janeček, Š. & Zámocká, B., 2019. A new GH13 subfamily represented by the α-amylase from the halophilic archaeon *Haloarcula hispanica*. *Springer Nature*. 24:207–217.

Janeček, Š., Svensson, B. & MacGregor, E.A., 2011. Structural and evolutionary aspects of two families of non-catalytic domains present in starch and glycogen binding proteins from microbes, plants and animals. *Enzyme and Microbial Technology*. 49(5):429–440.

Janeček, Š., Svensson, B. & MacGregor, E.A., 2014. α-Amylase: An enzyme specificity found in various families of glycoside hydrolases. *Cellular and Molecular Life Sciences*. 71(7):1149–1170.

Janeček, Š., Mareček, F., MacGregor, E.A. & Svensson, B., 2019. Starch-binding domains as CBM families–history, occurrence, structure, function and evolution. *Biotechnology Advances*. 37(8):107451.

Jang, M.U., Kang, H.J., Jeong, C.K., Kang, Y., Park, J.E. & Kim, T.J., 2018. Functional expression and enzymatic characterization of *Lactobacillus*

plantarum cyclomaltodextrinase catalyzing novel acarbose hydrolysis. *Journal of Microbiology*. 56(2):113–118.

Jones, D.T., 1999. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*. 292(2):195–202.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G. & Pesseat, S., 2014. InterProScan 5: Genome-scale protein function classification. *Bioinformatics*. 30(9):1236–1240.

Joynson, R., Pritchard, L., Osemwekha, E. & Ferry, N., 2017. Metagenomic analysis of the gut microbiome of the common black slug arion ater in search of novel lignocellulose degrading enzymes. *Frontiers in Microbiology*. 8(1):2181.

Jòzef, Synowiecki. 2007. The use of starch processing enzymes in the food industry. *Industrial Enzymes: Structure, Function and Applications*. 16:19–34.

Józef Synowiecki. 2010. Some applications of thermophiles and their enzymes for protein processing. *African Journal of Biotechnology*. 9(42):7186–7191.

Kahar, U.M., Ng, C.L., Chan, K.G. & Goh, K.M. 2016. Characterization of a type I pullulanase from *Anoxybacillus* sp. SK3-4 reveals an unusual substrate hydrolysis. *Applied Microbiology and Biotechnology*. 100(14):6291–6307.

Kahar, U.M., Ab Latif, N., Amran, S.I., Liew, K.J. & Goh, K.M. 2022. A bibliometric analysis and review of pullulan-degrading enzymes—past and current trends. *Catalysts*. 12(2):143.

Kakirde, K.S., Parsley, L.C. & Liles, M.R., 2010. Size does matter: application-driven approaches for soil metagenomics. *Soil Biology and Biochemistry*. 42(11):1911–1923.

Kandra, L., Gyémánt, G., Remenyik, J., Hovánszki, G. & Lipták, A., 2002. Action pattern and subsite mapping of *Bacillus licheniformis* α-amylase (BLA) with modified maltooligosaccharide substrates. *FEBS Letters*. 518(1–3):79–82.

Kang, J., Park, K.M., Choi, K.H., Park, C.S., Kim, G.E., Kim, D. & Cha, J., 2011. Molecular cloning and biochemical characterization of a heat-stable type I pullulanase from *Thermotoga neapolitana*. *Enzyme and Microbial Technology*. 48(3):260–266.

Kaulpiboon, J. & Pongsawasdi, P. 2004. Expression of cyclodextrinase gene from *Paenibacillus* sp. A11 in *Escherichia coli* and characterization of the purified cyclodextrinase. *Journal of Biochemistry and Molecular Biology*. 37(4):408–415.

Kelly, E.D., Bottacini, F., O'Callaghan, J., Motherway, M.O.C., O'Connell, K.J., Stanton, C. & van Sinderen, D., 2016. Glycoside hydrolase family 13 α-glucosidases encoded by *Bifidobacterium breve* UCC2003; A comparative analysis of function, structure and phylogeny. *International Journal of Food Microbiology*. 224:55–65.

Khan, A.R., Pervez, M.T., Babar, M.E., Naveed, N. & Shoaib, M., 2018. A comprehensive study of de novo genome assemblers: Current challenges and future prospective. *Evolutionary Bioinformatics*. 14:1176934318758650.

Khlestkin, V.K., Peltek, S.E. & Kolchanov, N.A. 2018. Review of direct chemical and

biochemical transformations of starch. *Carbohydrate Polymers*. 181:460–476.

Kim, C.H., Nashiru, O. & Ko, J.H. 1996. Purification and biochemical characterization of pullulanase type I from *Thermus caldophilus* GK-24. *FEMS Microbiology Letters*. 138(3):147–152.

Kimura, N. & Kamagata, Y., 2016. A thermostable bilirubin-oxidizing enzyme from activated sludge isolated by a metagenomic approach. *Microbes and Environments*. 31(4):435–441.

Koch, R., Spreinat, A., Lemke, K. & Antranikian, G. 1991. Purification and properties of a hyperthermoactive α-amylase from the archaeobacterium *Pyrococcus woesei. Archives of Microbiology*. 155(6):572–578.

Koshland Jr., D.E. 1953. Stereochemistry and the mechanism of enzymatic reactions *Biol. Rev.* 28(4):416–436.

Kotera, M., Nishimura, Y., Nakagawa, Z.I., Muto, A., Moriya, Y., Okamoto, S., Kawashima, S., Katayama, T., Tokimatsu, T., Kanehisa, M. & Goto, S., 2014 PIERO ontology for analysis of biochemical transformations: Effective implementation of reaction information in the IUBMB enzyme list. *Journal of Bioinformatics and Computational Biology*. 12(6):1–17.

Kotik, M. 2009. Novel genes retrieved from environmental DNA by polymerase chain reaction: Current genome-walking techniques for future metagenome applications. *Journal of Biotechnology*. 144(2):75–82.

Kumar, S. & Nussinov, R. 1999. Salt bridge stability in monomeric proteins. *Journal of Molecular Biology*. 293(5):1241–1255.

Kumar, S., Dangi, A.K., Shukla, P., Baishya, D. & Khare, S.K. 2019. Thermozymes: Adaptive strategies and tools for their biotechnological applications. *Bioresource Technology*. 278:372–382.

Kumar, S.S., Sangeeta, R., Soumya, S., Ranjan, R.P., Bidyut, B. & Kumar, D.M.P. 2014. Characterizing novel thermophilic amylase producing bacteria from Taptapani hot spring, Odisha, India. *Jundishapur Journal of Microbiology*. 7(12):1–7.

Kuriki, T., Okada, S. & Imanaka, T. 1988. New type of pullulanase from *Bacillus stearothermophilus* and molecular cloning and expression of the gene in *Bacillus subtilis*. *Journal of Bacteriology*. 170(4):1554–1559.

Kwan, D.H. 2017. Structure-Guided Directed Evolution of Glycosidases: A case study in engineering a blood group antigen-cleaving enzyme. *Methods in Enzymology*. 1(597):25-53.

Labes, A. & Schönheit, P. 2001. Sugar utilization in the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus* strain 7324: Starch degradation to acetate and CO2 via a modified Embden-Meyerhof pathway and acetyl-CoA synthetase (ADP-forming). *Archives of Microbiology*. 176(5):329–338.

Labes, A. & Schönheit, P. 2007. Unusual starch degradation pathway via cyclodextrins in the hyperthermophilic sulfate-reducing archaeon *Archaeoglobus fulgidus* strain 7324. *Journal of Bacteriology*. 189(24):8901–8913.

Labes, A., Karlsson, E.N., Fridjonsson, O.H., Turner, P., Hreggvidson, G.O., Kristjansson, J.K., Holst, O. & Schönheit, P., 2008. Novel members of glycoside hydrolase family 13 derived from environmental DNA. *Applied and Environmental Microbiology*. 74(6):1914–1921.

Laderman, K.A., Davis, B.R., Krutzsch, H.C., Lewis, M.S., Griko, Y.V., Privalov, P.L. & Anfinsen, C.B., 1993. The purification and characterization of an extremely thermostable α- amylase from the hyperthermophilic archaebacterium *Pyrococcus furiosus*. *Journal of Biological Chemistry*. 268(32):24394–24401.

Laemmli, U.K. 1970. Cleavage of Structural proteins during the Assembly of the Head of Bacteriophage T4. *Nature*. 227(5259):680–685.

Lahiri, D., Nag, M., Banerjee, R., Mukherjee, D., Garai, S., Sarkar, T., Dey, A., Sheikh, H.I., Pathak, S.K., Edinur, H.A. & Pati, S., 2021. Amylases: Biofilm inducer or biofilm inhibitor? *Frontiers in Cellular and Infection Microbiology*. 11(11):1–13.

Larbey, C., Mentzer, S.M., Ligouis, B., Wurz, S. & Jones, M.K. 2019. Cooked starchy food in hearths ca. 120 kya and 65 kya (MIS 5e and MIS 4) from Klasies River Cave, South Africa. *Journal of Human Evolution*. 131:210–227.

Latorre-García, L., Adam, A.C., Manzanares, P. & Polaina, J. 2005. Improving the amylolytic activity of *Saccharomyces cerevisiae* glucoamylase by the addition of a starch binding domain. *Journal of Biotechnology*. 118(2):167–176.

Le Corre, D., Bras, J., & Dufresne, A. 2010. Starch Nanoparticles: A Review. *Biomacromolecules*. *11*(5):1139–1153.

Lee, J. & Paetzel, M. 2011. Structure of the catalytic domain of glucoamylase from *Aspergillus niger*. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*. 67(2):188–192.

Lee, H.S., Shockley, K.R., Schut, G.J., Conners, S.B., Montero, C.I., Johnson, M.R., Chou, C.J., Bridger, S.L., Wigner, N., Brehm, S.D. & Jenney Jr, F.E., 2006. Transcriptional and biochemical analysis of starch metabolism in the hyperthermophilic archaeon *Pyrococcus furiosus*. *Journal of Bacteriology*. 188(6):2115–2125.

Lee, J.E., Kim, I.H., Jung, J.H., Seo, D.H., Kang, S.G., Holden, J.F., Cha, J. & Park, C.S., 2013. Molecular cloning and enzymatic characterization of cyclomaltodextrinase from hyperthermophilic archaeon *Thermococcus* sp. CL1. *Journal of microbiology and biotechnology*. 23(8):1060–1069.

Lee, K.T., Toushik, S.H., Baek, J.Y., Kim, J.E., Lee, J.S. & Kim, K.S. 2018. Metagenomic mining and functional characterization of a novel KG51 bifunctional cellulase/hemicellulase from black goat rumen. *Journal of Agricultural and Food Chemistry*. 66(34):9034–9041.

Lee, S.J., Yoo, S.H., Kim, M.J., Kim, J.W., Seok, H.M. & Park, K.H., 1995. Production and characterization of branched oligosaccharides from liquefied starch by the action of *B. Licheniformis* Amylase. *Starch - Stärke*. 47(4):127–134.

Lee, Y.S., Seo, S.H., Yoon, S.H., Kim, S.Y., Hahn, B.S., Sim, J.S., Koo, B.S. & Lee,

C.M., 2016.. Identification of a novel alkaline amylopullulanase from a gut metagenome of *Hermetia illucens*. *International Journal of Biological Macromolecules*. 82:514–521.

Lo Leggio, L., Simmons, T.J., Poulsen, J.C.N., Frandsen, K.E., Hemsworth, G.R., Stringer, M.A., Von Freiesleben, P., Tovborg, M., Johansen, K.S., De Maria, L. & Harris, P.V., 2015. Structure and boosting activity of a starch-degrading lytic polysaccharide monooxygenase. *Nature Communications*. 6(1):5961.

Leis, B., Angelov, A., Mientus, M., Li, H., Pham, V.T., Lauinger, B., Bongen, P., Pietruszka, J., Goncalves, L.G., Santos, H. & Liebl, W., 2015. Identification of novel esterase-active enzymes from hot environments by use of the host bacterium *Thermus thermophilus*. *Frontiers in Microbiology*. 6(275):1–12.

Levasseur, A., Drula, E., Lombard, V., Coutinho, P.M. & Henrissat, B., 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels*. 6(1):1.

Lévêque, E., Janeček, Š., Haye, B. & Belarbi, A., 2000. Thermophilic archaeal amylolytic enzymes. *Enzyme and Microbial Technology*. 26(1):3–14.

Lewin, A., Zhou, J., Pham, V.T.T., Haugen, T., Zeiny, M.E., Aarstad, O., Liebl, W., Wentzel, A. & Liles, M.R., 2017. Novel archaeal thermostable cellulases from an oil reservoir metagenome. *AMB Express*. 7(1):1-17.

Lewin, A., Strand, T.A., Haugen, T., Klinkenberg, G., Kotlar, H.K., Valla, S., Drabløs, F. & Wentzel, A., 2016. Discovery and characterization of a thermostable esterase from an oil reservoir metagenome. *Advances in Enzyme Research*. 4(2):68–86.

Liberato, M.V., Prates, E.T., Gonçalves, T.A., Bernardes, A., Vilela, N., Fattori, J., Ematsu, G.C., Chinaglia, M., Gomes, E.R.M., Figueira, A.C.M. & Damasio, A., 2021. Insights into the dual cleavage activity of the GH16 laminarinase enzyme class on β-1,3 and β-1,4 glycosidic bonds. *Journal of Biological Chemistry*. 296(9):100385.

Liebl, W., Angelov, A., Juergensen, J., Chow, J., Loeschcke, A., Drepper, T., Classen, T., Pietruzska, J., Ehrenreich, A., Streit, W.R. & Jaeger, K.E., 2014. Alternative hosts for functional (meta)genome analysis. *Applied Microbiology and Biotechnology*. 98(19):8099–8109.

Linder, M., Lindeberg, G., Reinikainen, T., Teeri, T.T. & Pettersson, G. 1995. The difference in affinity between two fungal cellulose-binding domains is dominated by a single amino acid substitution. *FEBS Letters*. 372(1):96–98.

Liu, X. & Kokare, C. 2017. *Microbial Enzymes of Use in Industry*. *Biotechnology of Microbial Enzymes: Production, Biocatalysis and Industrial Applications*. 267-298.

Liu, Y.F., Chen, J., Liu, Z.L., Shou, L.B., Lin, D.D., Zhou, L., Yang, S.Z., Liu, J.F., Li, W., Gu, J.D. & Mu, B.Z., 2020. Anaerobic degradation of paraffins by thermophilic Actinobacteria under methanogenic conditions. *Environmental Science and Technology*. 54(17):10610–10620.

Liu, Z., Li, J., Jie, C., Wu, B. & Hao, N., 2021. A multifunctional α-amylase BSGH13

from *Bacillus subtilis* BS-5 possessing endoglucanase and xylanase activities. *International Journal of Biological Macromolecules*. 171:166–176.

Liu, Z., Zhao, C., Deng, Y., Huang, Y. & Liu, B., 2015. Characterization of a thermostable recombinant β-galactosidase from a thermophilic anaerobic bacterial consortium YTY-70. *Biotechnology and Biotechnological Equipment*. 29(3):547–554.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. & Henrissat, B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*. 42(D1):D490–D495.

Lorenz, P. & Eck, J. 2005. Metagenomics and industrial applications. *Nature Reviews Microbiology*. 3(6):510–516.

Lu, X., Wang, Y., Li, Y. & Huang, Q. 2018. Assembly of Pickering emulsions using milled starch particles with different amylose/amylopectin ratios. *Food Hydrocolloids*. 84:47–57.

Ludwig, D.S., Hu, F.B., Tappy, L. & Brand-Miller, J. 2018. Dietary carbohydrates: Role of quality and quantity in chronic disease. *Science and Politics of nutrition*. 361.

Van Der Maarel, M.J.E.C., Van Der Veen, B., Uitdehaag, J.C.M., Leemhuis, H. & Dijkhuizen, L. 2002. Properties and applications of starch-converting enzymes of the α-amylase family. *Journal of Biotechnology*. 94(2):137–155.

Machovič, M. & Janeček, Š., 2008. Domain evolution in the GH13 pullulanase subfamily with focus on the carbohydrate-binding module family 48. *Biologia*. 63(6):1057–1068.

Madhavan, A., Sindhu, R., Parameswaran, B., Sukumaran, R.K. & Pandey, A., 2017. Metagenome analysis: a Powerful tool for enzyme bioprospecting. *Applied Biochemistry and Biotechnology*. 183(2):636–651.

Magallanes, P.A., Flores, P.C. & Bello, L.A., 2017. Starch structure influences Its digestibility: A Review. *Journal of Food Science*. 82(9):2016–2023.

Maimanakos, J., Chow, J., Gaßmeyer, S.K., Güllert, S., Busch, F., Kourist, R. & Streit, W.R., 2016. Sequence-based screening for rare enzymes: New insights into the world of AMDases reveal a conserved motif and 58 novel enzymes clustering in eight distinct families. *Frontiers in Microbiology*. 7(1):1332).

Marques, S., Moreno, A.D., Ballesteros, M. & Gírio, F., 2017. Starch biomass for biofuels, biomaterials, and chemicals. *Biomass and Green Chemistry: Building a Renewable Pathway*. 69–94.

McCarter, J.D. & Withers, S.G., 1994. Mechanisms of enzymatic glycoside hydrolysis. *Current Opinion in Structural Biology*. 4:885–892.

McDonald, A. G., Boyce, S., & Tipton, K. F., 2009. ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic acids research.* 37(1):D593-D597.

McIntosh, L.P., Hand, G., Johnson, P.E., Joshi, M.D., Korner, M. & Withers, S.G., 1996. The pK(a) of the general acid/base carboxyl group of a glycosidase cycles during catalysis: A 13C-NMR study of *Bacillus circulans xylanase*. *Biochemistry*.

35(31):9958–9966.

McPherson, M., & Møller, S., 2000. Pcr. Taylor & Francis.

Messaoud, E., Ben Ammar, Y., Mellouli, L. & Bejar, S., 2002. Thermostable pullulanase type I from new isolated *Bacillus thermoleovorans* US105: Cloning, sequencing and expression of the gene in *E. coli*. *Enzyme and Microbial Technology*. 31(6):827–832.

Miao, M., Jiang, B., Jin, Z. & BeMiller, J.N., 2018. Microbial starch-converting enzymes: Recent insights and perspectives. *Comprehensive Reviews in Food Science and Food Safety*. 17(5):1238–1260.

Mirete, S., Morgante, V. & González-Pastor, J.E., 2016. Functional metagenomics of extreme environments. *Current Opinion in Biotechnology*. 38:143–149.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L., Tosatto, S.C., Paladin, L., Raj, S., Richardson, L.J. & Finn, R.D., 2021. Pfam: The protein families database in 2021. *Nucleic Acids Research*. 49(D1):D412–D419.

Mizanur, R.M., Griffin, A.K.K. & Pohl, N.L., 2008. Recombinant production and biochemical characterization of a hyperthermostable α-glucan/maltodextrin phosphorylase from *Pyrococcus furiosus*. *Archaea*. 2(3):169–176.

Møller, M.S. & Svensson, B., 2016. Structural biology of starch-degrading enzymes and their regulation. *Current Opinion in Structural Biology*. 40:33–42.

Møller, S., Henriksen, A. & Svensson, B., 2016. Structure and function of a-glucan debranching enzymes. *Cellular and Molecular Life Sciences*. 73:2619–2641.

Motahar, S.F.S., Khatibi, A., Salami, M., Ariaeenejad, S., Emam-Djomeh, Z., Nedaei, H., Kavousi, K., Mamaghani, A.S. & Salekdeh, G.H., 2021. A novel metagenome-derived thermostable and poultry feed compatible α-amylase with enhanced biodegradation properties. *International Journal of Biological Macromolecules*. 164:2124–2133.

Mrudula, S., Gopal, R. & Seenayya, G., 2011. Purification and characterization of highly thermostable amylopullulanase from a thermophilic, anaerobic bacterium *Clostridium thermosulfurogenes* SVM17. *Malaysian Journal of Microbiology*. 7(2):97–106.

Nakamura, Y., 1996. Some properties of starch debranching enzymes and their possible role in amylopectin biosynthesis. *Plant Science*. 121(1), 1-18.

Nazir, A., 2016. Review on Metagenomics and its Applications. *Imperial Journal of Interdisciplinary Research*. 2(3):277–286.

Ndata, K., Nevondo, W., Cekuse, B., van Zyl, L.J. & Trindade, M., 2021. Characterization of a highly xylose tolerant β-xylosidase isolated from high temperature horse manure compost. *BMC Biotechnology*. 21(1):1–10.

Ngara, T.R. & Zhang, H., 2018. Recent advances in function-based metagenomic screening. *Genomics Proteomics Bioinformatics.* 16(2018):405-415.

Nguyen, K.T., Nguyen, N.L., Tung, N.V., Nguyen, H.H., Milhim, M., Le, T.T.X., Lai,

T.H.N., Phan, T.T.M. & Bernhardt, R., 2020. A novel thermostable cytochrome p450 from sequence-based metagenomics of binh chau hot spring as a promising catalyst for testosterone conversion. *Catalysts*. 10(9):1–17.

Niehaus, F., Bertoldo, C., Kähler, M. & Antranikian, G., 1999. Extremophiles as a source of novel enzymes for industrial application. *Applied Microbiology and Biotechnology*. 51(6):711–729.

Niehaus, F., Peters, A., Groudieva, T. & Antranikian, G., 2000. Cloning, expression and biochemical characterisation of a unique thermostable pullulan-hydrolysing enzyme from the hyperthermophilic archaeon *Thermococcus aggregans*. *FEMS Microbiology Letters*. 190(2):223–229.

Nisha, M. & Satyanarayana, T., 2016. Characteristics, protein engineering and applications of microbial thermostable pullulanases and pullulan hydrolases. *Applied Microbiology and Biotechnology*. 100(13):5661–5679.

Nithya, K., Muthukumar, C., Kadaikunnan, S., Alharbi, N.S., Khaled, J.M. & Dhanasekaran, D. 2017. Purification, characterization, and statistical optimization of a thermostable α-amylase from desert actinobacterium *Streptomyces fragilis* DA7-7. *Biotechnology*. 7(5):1–13.

Niu, C., Zheng, F., Li, Y., Liu, C. & Li, Q., 2018. Process optimization of the extraction condition of β-amylase from brewer's malt and its application in the maltose syrup production. *Biotechnology and Applied Biochemistry*. 65(4):639–647.

Noorwez, S.M., Ezhilvannan, M. & Satyanarayana, T., 2006. Production of a high maltose-forming, hyperthermostable and Ca 2+-independent amylopullulanase by an extreme thermophile *Geobacillus thermoleovorans* in submerged fermentation. *Indian Journal of Biotechnology*. 5(3):337–345.

Nora, L.C., Westmann, C.A., Martins-Santana, L., Alves, L.D.F., Monteiro, L.M.O., Guazzaroni, M.E. & Silva-Rocha, R., 2019. The art of vector engineering: towards the construction of next-generation genetic tools. *Microbial Biotechnology*. 12(1):125–147.

Novozymes. 2014. *Major steps in enzymatic starch conversion*. https://www.potatopro.com/news/2014/novozymes-offers-starch-processors-new-enzyme Date of access: 17 Aug. 2020.

Odibo, F.J.C. & Obi, S.K.C., 1988. Purification and characterization of a thermostable pullulanase from *Thermoactinomyces thalpophilus*. *Journal of Industrial Microbiology*. 3(6):343–350.

Oh, H.N., Park, D., Seong, H.J., Kim, D. & Sul, W.J., 2019. Antarctic tundra soil metagenome as useful natural resources of cold-active lignocelluolytic enzymes. *Journal of Microbiology*. 57(10):865–873.

Okuyama, M. 2011. Function and structure studies of GH family 31 and 97 α-glycosidases. *Bioscience, Biotechnology and Biochemistry*. 75(12):2269–2277.

Oliveira, C., Carvalho, V., Domingues, L. & Gama, F.M., 2015. Recombinant CBM-fusion technology - Applications overview. *Biotechnology Advances*. 33(3):358–369.

Onuma, H., Uchiyama, H., Hara, K., Fukuta, Y. & Shirasaka, N. 2018. Enzymatic characterization of an extracellular glucoamylase from *Tricholoma matsutake* and its cloning and secretory expression in *Pichia pastoris*. *Bioscience, Biotechnology and Biochemistry*. 82(12):2180–2190.

Pacbio., 2016. *Genomes vs. Genes: The difference between contigs and scaffolds in genome assemblies*. https://www.pacb.com/blog/genomes-vs-gennnnes-difference-contigs-scaffolds-genome-assemblies/ Date of access: 18 Mar. 2020.

Pace, N. R., Stahl, D. A., Lane, D. J., & Olsen, G. J., 1986. The analysis of natural microbial populations by ribosomal RNA sequences. *Advances in microbial ecology.* 1-55.

Pang, B., Zhou, L., Cui, W., Liu, Z., Zhou, S., Xu, J. & Zhou, Z., 2019. A hyperthermostable Type II Pullulanase from a deep-sea microorganism *Pyrococcus yayanosii* CH1. *Journal of Agricultural and Food Chemistry*. 67(34):9611–9617.

Park, K.H., Kim, T.J., Cheong, T.K., Kim, J.W., Oh, B.H. & Svensson, B., 2000. Structure, specificity and function of cyclomaltodextrinase, a multispecific enzyme of the α-amylase family. *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*. 1478(2):165–185.

Park, K.M., Jun, S.Y., Choi, K.H., Park, K.H., Park, C.S. & Cha, J., 2010. Characterization of an exo-acting intracellular α-amylase from the hyperthermophilic bacterium *Thermotoga neapolitana*. *Applied Microbiology and Biotechnology*. 86(2):555–566.

Park, S.H., Na, Y., Kim, J., Kang, S.D. & Park, K.H., 2018. Properties and applications of starch modifying enzymes for use in the baking industry. *Food Science and Biotechnology*. 27(2):299–312.

Parker, K., Salas, M. & Nwosu, V.C., 2010. High fructose corn syrup : Production , uses and public health concerns. *Biotechnology and Molecular Biology Review*. 5(5):71–78.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.A. & Hugenholtz, P., 2018 A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*. 36(10):996.

Peng, H., Zheng, Y., Chen, M., Wang, Y., Xiao, Y. & Gao, Y., 2014. A starch-binding domain identified in α-amylase (AmyP) represents a new family of carbohydrate-binding modules that contribute to enzymatic hydrolysis of soluble starch. *FEBS Letters*. 588(7):1161–1167.

Philip Hugenholtz & Tyson, G.W., 2008. Metagenomics. *Nature*. 55(7212):481–483.

Polekhina, G., Gupta, A., van Denderen, B.J., Feil, S.C., Kemp, B.E., Stapleton, D. & Parker, M.W., 2005. Structural basis for glycogen recognition by AMP-activated protein kinase. *Structure*. 13(10):1453–1462.

Pope, B., & Kent, H. M., 1996. High efficiency 5 min transformation of *Escherichia coli*. *Nucleic acids research*. 24(3):536-537.

Prakash, B., Vidyasagar, M., Madhukumar, M.S., Muralikrishna, G. & Sreeramulu,

K., 2009. Production, purification, and characterization of two extremely halotolerant, thermostable, and alkali-stable α-amylases from *Chromohalobacter* sp. TVSP 101. *Process Biochemistry*. 44(2):210–215.

Qiu, S., 2019. Parallelizing de vovo Assembly with Heterogeneous Processors. Hong Kong University of Science and Technology.

Rajput, K.N., Patel, K.C. & Trivedi, U.B., 2016. A novel cyclodextrin glucanotransferase from an alkaliphile *Microbacterium terrae* KNR 9 : purification and properties. *3 Biotech*. 6(2):1–11.

Ramsay, A.G., Scott, K.P., Martin, J.C., Rincon, M.T. & Flint, H.J. 2006. Cell-associated α-amylases of butyrate-producing Firmicute bacteria from the human colon. *Microbiology*. 152(11):3281–3290.

Ran, H., Wu, J., Wu, D. & Duan, X. 2016. Enhanced production of recombinant *Thermobifida fusca* isoamylase in *Escherichia coli* MDS42. *Applied Biochemistry and Biotechnology*. 180(3):464–476.

Randle-Boggis, R., Helgason, T., Sapp, M & Ashton, P. 2016. Evaluating techniques for metagenome annotation using simulated sequence data. *FEMS Microbiology Ecology*.92(7): 1-15.

Rasheed, A., Kumar C.K., A. & Sravanthi, V.V.N.S.S., 2008. Cyclodextrins as drug carrier molecule: A review. *Scientia Pharmaceutica*. 76(4):567–598.

Rathore, R.S., Garg, N., Garg, S. & Kumar, A., 2009. Starch phosphorylase: Role in starch metabolism and biotechnological applications. *Critical Reviews in Biotechnology*. 29(3):214–224.

Reichart, N.J., Bowers, R.M., Woyke, T. & Hatzenpichler, R., 2021. High potential for biomass-degrading enzymes revealed by hot spring metagenomics. *Frontiers in Microbiology*. 12(4):1–13.

Rigden, D.J., Jedrzejas, M.J. & De Mello, L. V., 2003. Identification and analysis of catalytic TIM barrel domains in seven further glycoside hydrolase families. *FEBS Letters*. 544(3):103–111.

Ring, S.G., Colonna, P., I'Anson, K.J., Kalichevsky, M.T., Miles, M.J., Morris, V.J. & Orford, P.D., 1987. The gelation and crystallisation of amylopectin. *Carbohydrate Research*. 162(2):277–293.

Robinson, P. K. 2015. Enzymes: principles and biotechnological applications. *Essays in biochemistry*. 59:1.

Robyt, J.F. 2009. *Enzymes and Their Action on Starch*. 237-292.

Rosell, C.M., Haros, M., Escrivá, C. & Benedito de Barber, C., 2001. Experimental approach to optimize the use of α-amylases in breadmaking. *Journal of Agricultural and Food Chemistry*. 49(6):2973–2977.

Rosenberg, E., & Ron, E. Z. 2006. Biosurfactants. *The Prokaryotes; Rosenberg, E., DeLong, EF, Lory, S., Stackebrandt, E., Thompson, F., Eds*. 281-294.

Roth, C., Moroz, O.V., Turkenburg, J.P., Blagova, E., Waterman, J., Ariza, A., Ming, L., Tianqi, S., Andersen, C., Davies, G.J. & Wilson, K.S., 2019. Structural and

functional characterization of three novel fungal amylases with enhanced stability and pH tolerance. *International Journal of Molecular Sciences*. 20(19):1–15.

Rubino, J.T. & Franz, K.J., 2012. Coordination chemistry of copper proteins: How nature handles a toxic cargo for essential function. *Journal of Inorganic Biochemistry*. 107(1):129–143.

Rudiger, A., Jorgensen, P.L. & Antranikian, G., 1995. Isolation and characterization of a heat-stable pullulanase from the hyperthermophilic archaeon *Pyrococcus woesei* after cloning and expression of its gene in *Escherichia coli*. *Applied and Environmental Microbiology*. 61(2):567–575.

Shadegian Motahar, S.F., Salami, M., Ariaeenejad, S., Emam-Djomeh, Z. A. Mamaghani, A., Kavousi, K. & Moghadam, M. Salekdeh, G., 2022. Synergistic Effect of Metagenome-Derived Starch-Degrading Enzymes on Quality of Functional Bread with Antioxidant Activity. *Starch/Staerke*. 74(1):1–14.

Sambrook, J. & Russell, D.W., 2001. *Molecular Cloning: Ch. 8. In Vitro amplification of DNA by the polymerase chain reaction (Vol. 2).*Cold Spring Harbor Laboratory Press.

Sarmiento, F., Peralta, R. & Blamey, J.M., 2015. Cold and hot extremozymes: Industrial relevance and current trends. *Frontiers in Bioengineering and Biotechnology*. 3(148):1–12.

Sato, M., Suda, M., Okuma, J., Kato, T., Hirose, Y., Nishimura, A., Kondo, Y. & Shibata, D., 2017. Isolation of highly thermostable β-xylosidases from a hot spring soil microbial community using a metagenomic approach. *DNA Research*. 24(6):649–656.

Schrodinger, L. L. C., 2015. The PyMOL molecular graphics system. Version, 1:8.

Schwartz, D. & Whistler, R.L., 2009. History and future of starch. *Starch*. 1-10

Shoseyov, O., Shpigel, I., Goldstein, M.M. & Doi, R.H., 1999. Research development co of Hebrew university of Jerusalem. *Methods of detection using a cellulose binding domain fusion product*. U.S. Patent 5,856,201.

Shoseyov, O., Shani, Z. & Levy, I. 2006. Carbohydrate binding modules: Biochemical properties and novel applications. *Microbiology and Molecular Biology Reviews*. 70(2):283–295.

Singh, R., Singh, A. & Sachan, S,. 2018. *Enzymes used in the food industry: Friends or foes? Enzymes in Food Biotechnology: Production, Applications, and Future Prospects*. Elsevier Inc.

Smith, V.H. & Schindler, D.W., 2009. Eutrophication science: where do we go from here? *Trends in Ecology and Evolution*. 24(4):201–207.

Solonenko, S.A. & Sullivan, M.B., 2013. Preparation of metagenomic libraries from naturally occurring marine viruses. *Methods in Enzymology*. 531:143-165.

Song, C., Wang, B., Tan, J., Zhu, L. & Lou, D., 2017. Discovery of tauroursodeoxycholic acid biotransformation enzymes from the gut microbiome of black bears using metagenomics. *Scientific Reports*. 7(45495):1–8.

Souza, P. M. D., 2010. Application of microbial α-amylase in industry-A review. *Brazilian journal of microbiology*. 41(1):850-861.

Sreekanth, K.V., Alapan, Y., Elkabbash, M. & Ilker, E., 2016. Health Research Alliance. *Nature Materials*. 15(6):621–627.

Staley, J. & Konopka, A., 1985. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiology*. 39(1):321–346.

Stam, M.R., Danchin, E.G.J., Rancurel, C., Coutinho, P.M. & Henrissat, B., 2006. Dividing the large glycoside hydrolase family 13 into subfamilies: Towards improved functional annotations of α-amylase-related proteins. *Protein Engineering, Design and Selection*. 19(12):555–562.

Strobl, S., Maskos, K., Wiegand, G., Huber, R., Gomis-Rüth, F.X. & Glockshuber, R., 1998. Crystal structure of yellow meal worm alpha-amylase at 1.64 angstrom resolution. *Journal of Molecular Biology*. 278(3):617–628.

Suleiman, M., Krüger, A. & Antranikian, G., 2020. Biomass-degrading glycoside hydrolases of archaeal origin. *Biotechnology for Biofuels*. 13(1):1–14.

Sun, Y., Lv, X., Li, Z., Wang, J., Jia, B. & Liu, J., 2015. Recombinant cyclodextrinase from *Thermococcus kodakarensis* KOD1: Expression, purification, and enzymatic characterization. *Archaea*.

Suzuki, Y., Hatagaki, K. & Oda, H. 1991. A hyperthermostable pullulanase produced by an extreme thermophile, *Bacillus flavocaldarius* KP 1228, and evidence for the proline theory of increasing protein thermostability. *Applied Microbiology and Biotechnology*. 34(6):707–714.

Takano, K., Higashi, R., Okada, J., Mukaiyama, A., Tadokoro, T., Koga, Y. & Kanaya, S., 2009. Proline effect on the thermostability and slow unfolding of a hyperthermophilic protein. *Journal of Biochemistry*. 145(1):79–85.

Tamura, K., Stecher, G., & Kumar, S., 2021. MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Molecular Biology and Evolution*. 38(7):3022-3027.

Tavallaie, S., Khomeiri, M., Mousivand, M., Maghsoudlou, Y. & Hashemi, M., 2019. Starches from different sources hydrolysis using a new thermo-tolerant amylase complex produced by *Bacillus subtilis* T41a: Characterization and efficiency evaluation. *Food Science and Technology*. 112:108218.

Taylor, J.R.N. & Dewar, J. 1994. Role of alpha-glucosidase in the fermentable Sugar composition of sorghum malt mashes. *Journal of the Institute of Brewing*. 100(6):417–419.

Thermes, C. 2014. Ten years of next-generation sequencing technology. *Trends in genetics : TIG*. 30(9):418–426.

Thoma, J.A., Spradlin, J.E. & Dygert, S. 1971. Plant and animal amylases. *The enzymes*. 5(1):115–189.

Thomas, T., Gilbert, J. & Meyer, F. 2014. Metagenomics: A guide from sampling to data analysis. *Microbial Informatics and Experimentation*. 2(3):357–383.

Thornbury, M., Sicheri, J., Slaine, P., Getz, L.J., Finlayson-Trick, E., Cook, J., Guinard, C., Boudreau, N., Jakeman, D., Rohde, J. & McCormick, C., 2019. Characterization of novel lignocellulose-degrading enzymes from the porcupine microbiome using synthetic metagenomics. *PLoS ONE*. 14(1):1–18.

Toor, K.J., Ahmad, N., Muhammad, M.A. & Rashid, N., 2021. TK-PUL, a pullulan hydrolase type III from *Thermococcus kodakarensis*, a potential candidate for simultaneous liquefaction and saccharification of starch. *Amylase*. 4(1):45–55.

Transtrum, M.K., Hansen, L.D. & Quinn, C., 2015. Enzyme kinetics determined by single-injection isothermal titration calorimetry. *Methods*. 76:194–200.

Tungland, B. 2018. Overview of prebiotics: membership, physiological effects and their health attributes. *Human Microbiota in Health and Disease*.

Uchiyama, T. & Miyazaki, K., 2009. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Current Opinion in Biotechnology*. 20(6):616–622.

Upadhyay, D., Sharma, S., Shrivastava, D., & Kulshreshtha, N. M., 2019. Production and characterization of β-cyclodextrin glucanotransferase from *Bacillus* sp. ND1. *Journal of basic microbiology*, 59(2):192-205.

Vajravijayan, S., Pletnev, S., Mani, N., Pletneva, N., Nandhagopal, N. & K. Gunasekaran., 2018. Structural insights on starch hydrolysis by plant β-amylase and its evolutionary relationship with bacterial enzymes. *International Journal of Biological Macromolecules*. 113:329–337.

Valk, V., Lammerts van Bueren, A., van der Kaaij, R.M. & Dijkhuizen, L. 2016. Carbohydrate-binding module 74 is a novel starch-binding domain associated with large and multidomain α-amylase enzymes. *FEBS Journal*. 283:2354–2368.

Vieille, C. & Zeikus, G.J., 2001. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiology and Molecular Biology Reviews*. 65(1):1–43.

Vieille, C., Burdette, D.S. & Zeikus, J.G., 1996. Thermozymes. *Biotechnology Annual Review*. 2:1–83.

Vlasenko, E., Schülein, M., Cherry, J. & Xu, F., 2010. Substrate specificity of family 5, 6, 7, 9, 12, and 45 endoglucanases. *Bioresource Technology*. 101(7):2405–2411.

Volkov, P. V., Rozhkova, A.M., Gusakov, A. V., Zorov, I.N. & Sinitsyn, A.P., 2015. Glucoamylases from *Penicillium verruculosum* and *Myceliophthora thermophila*: Analysis of differences in activity against polymeric substrates based on 3D model structures of the intact enzymes. *Biochimie*. 110:45–51.

Vollmers, J., Wiegand, S., & Kaster, A. K., 2017. Comparing and evaluating metagenome assembly tools from a microbiologist's perspective-not only size matters!. PloS one, 12(1):e0169662.

Volpicella, M., Leoni, C., Costanza, A., Fanizza, I., Placido, A. & Ceci, L.R. 2012. Genome walking by next generation sequencing approaches. *Biology*. 1(3):495–507.

Vu, V. V., Beeson, W.T., Span, E.A., Farquhar, E.R. & Marletta, M.A. 2014. A family of starch-active polysaccharide monooxygenases. *Proceedings of the National Academy of Sciences of the United States of America*. 111(38):13822–13827.

Vuong, T. V. & Wilson, D.B., 2010. Glycoside hydrolases: Catalytic base/nucleophile diversity. *Biotechnology and Bioengineering*. 107(2):195–205.

Wanderley, K.J., Torres, F.A.G., Moraes, L.M.P. & Ulhoa, C.J., 2004. Biochemical characterization of α-amylase from the yeast *Cryptococcus flavus*. *FEMS Microbiology Letters*. 231(2):165–169.

Wang, H., Gong, Y., Xie, W., Xiao, W., Wang, J., Zheng, Y., Hu, J. & Liu, Z., 2011. Identification and characterization of a novel thermostable gh-57 gene from metagenomic fosmid library of the juan de fuca ridge hydrothemal vent. *Applied Biochemistry and Biotechnology*. 164(8):1323–1338.

Wang, R., Wang, D., Gao, X. & Hong, J., 2014. Direct fermentation of raw starch using a *Kluyveromyces marxianus* strain that expresses glucoamylase and Alpha-amylase to produce ethanol. *Biotechnology Progress*. 30(2):338–347.

Wang, S., Lu, Z., Lu, M., Qin, S., Liu, H., Deng, X., Lin, Q. & Chen, J., 2008. Identification of archaeon-producing hyperthermophilic α-amylase and characterization of the α-amylase. *Applied Microbiology and Biotechnology*. 80(4):605–614.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L. & Lepore, R., 2018. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*. 46(W1):W296–W303.

Weirather, J.L., de Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X.J., Buck, D. & Au, K.F., 2017.Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research*. 6(1):100.

Wermuth, C.G. 2003. *Specific Substituent Groups*. Second Edi ed. *The Practice of Medicinal Chemistry: Second Edition*. Elsevier Inc.

Wilkerson, J.W., Yang, S.O., Funk, P.J., Stanley, S.K. & Bundy, B.C. 2018. Nanoreactors: Strategies to encapsulate enzyme biocatalysts in virus-like particles. *New Biotechnology*. 44:59–63.

Wohlgemuth, R., Littlechild, J., Monti, D., Schnorr, K., van Rossum, T., Siebers, B., Menzel, P., Kublanov, I.V., Rike, A.G., Skretas, G. & Szabo, Z., 2018. Discovering novel hydrolases from hot environments. *Biotechnology Advances*. 36(8):2077–2100.

Wooley, J.C., Godzik, A. & Friedberg, I., 2010. A primer on metagenomics. *PLoS Computational Biology*. 6(2): e1000667.

Wu, H., Yu, X., Chen, L. & Wu, G. 2014. Cloning, overexpression and characterization of a thermostable pullulanase from *Thermus thermophilus* HB27. *Protein Expression and Purification*. 95:22–27.

Wu, X., Wang, Y., Tong, B., Chen, X. & Chen, J. 2018. Purification and biochemical characterization of a thermostable and acid-stable alpha-amylase from *Bacillus*

licheniformis B4-423. *International Journal of Biological Macromolecules*. 109:329–337.

Wüthrich, K. 1990. Protein structure determination in solution by NMR spectroscopy. Journal of Biological Chemistry, 265(36):22059-22062.

Xu, M. & Wright, G.D., 2019. Heterologous expression-facilitated natural products' discovery in actinomycetes. *Journal of Industrial Microbiology and Biotechnology*. 46(3):415–431.

Yang, S.J., Lee, H.S., Park, C.S., Kim, Y.R., Moon, T.W. & Park, K.H., 2004. Enzymatic analysis of an amylolytic enzyme from the hyperthermophilic archaeon *Pyrococcus furiosus* reveals its novel catalytic properties as both an α-amylase and a cyclodextrin-hydrolyzing enzyme. *Applied and Environmental Microbiology*. 70(10):5988–5995.

Yasmin, F., Abdullah, M., Sethi, A.A., Saleem, H., Narmeen, A., Ansari, A. & Qader, S.K.S., 2016. Solid state fermentation: a Cost effective approach for production of starch liquefying fungal amylase using agro industrial wastes. *Science International*. 28(3):2703–2706.

Young Chul Chung, Kobayashi, T., Kanai, H., Akiba, T. & Kudo, T., 1995. Purification and properties of extracellular amylase from the hyperthermophilic archaeon *Thermococcus profundus* DT5432. *Applied and Environmental Microbiology*. 61(4):1502–1506.

Yun, J. & Ryu, S., 2005. Screening for novel enzymes from metagenome and SIGEX, as a way to improve it. *Microbial Cell Factories*. 4(1):1–5.

Zarafeta, D., Moschidi, D., Ladoukakis, E., Gavrilov, S., Chrysina, E.D., Chatziioannou, A., Kublanov, I., Skretas, G. & Kolisis, F.N., 2016. Metagenomic mining for thermostable esterolytic enzymes uncovers a new family of bacterial esterases. *Scientific Reports*. 6(1):1–16.

Zeng, Y., Zheng, H., Shen, Y., Xu, J., Tan, M., Liu, F. & Song, H., 2019. Identification and analysis of binding residues in the CBM68 of pullulanase PulA from *Anoxybacillus* sp. LM18-11. *Journal of Bioscience and Bioengineering*. 127(1):8–15.

Zhang, J.J., Tang, X. & Moore, B.S., 2019. Genetic platforms for heterologous expression of microbial natural products. *Natural Product Reports*. 36(9):1313–1332.

Zhang, Q., Han, Y. & Xiao, H., 2017. Microbial alpha-amylase: A biomolecular overview. *Elesivier*. 53(1):88–101.

Zhang, Y., Liu, Y.H., Li, Y., Liu, X.G. & Lu, F.P., 2013. Extracellular expression of pullulanase from *Bacillus naganoensis* in *Escherichia coli*. *Annals of Microbiology*. 63(1):289–294.

Zhao, C., Chu, Y., Li, Y., Yang, C., Chen, Y., Wang, X. & Liu, B., 2017. High-throughput pyrosequencing used for the discovery of a novel cellulase from a thermophilic cellulose-degrading microbial consortium. *Biotechnology Letters*. 39(1):123–131.

Zhou, M., Guo, P., Wang, T., Gao, L., Yin, H., Cai, C., Gu, J. & Lü, X., 2017.

Metagenomic mining pectinolytic microbes and enzymes from an apple pomace-adapted compost microbial community. *Biotechnology for Biofuels*. 10(1):1–15.

Zona, R., Chang-Pi-Hin, F., O'Donohue, M.J. & Janeček, Š. 2004. Bioinformatics of the glycoside hydrolase family 57 and identification of catalytic residues in amylopullulanase from *Thermococcus hydrothermalis*. *European Journal of Biochemistry*. 271(14):2863–2872.

Zouari Ayadi, D., Ben Ali, M., Jemli, S., Ben Mabrouk, S., Mezghani, M., Ben Messaoud, E. & Bejar, S., 2008. Heterologous expression, secretion and characterization of the *Geobacillus thermoleovorans* US105 type I pullulanase. *Applied Microbiology and Biotechnology*. 78(3):473–481.

Appendix

A) Detailed Sequence of PULL3-cont492 1 in pET-21a (+)

1 TGGCGAATGG GACGCGCCCT GTAGCGGCGC ATTAAGCGCG GCGGGTGTGG TGGTTACGCG

61 CAGCGTGACC GCTACACTTG CCAGCGCCCT AGCGCCCGCT CCTTTCGCTT TCTTCCCTTC

121 CTTTCTCGCC ACGTTCGCCG GCTTTCCCCG TCAAGCTCTA AATCGGGGGC TCCCTTTAGG

181 GTTCCGATTT AGTGCTTTAC GGCACCTCGA CCCCAAAAAA CTTGATTAGG GTGATGGTTC

241 ACGTAGTGGG CCATCGCCCT GATAGACGGT TTTTCGCCCT TTGACGTTGG AGTCCACGTT

301 CTTTAATAGT GGACTCTTGT TCCAAACTGG AACAACACTC AACCCTATCT CGGTCTATTC

361 TTTTGATTTA TAAGGGATTT TGCCGATTTC GGCCTATTGG TTAAAAAATG AGCTGATTTA

421 ACAAAAATTT AACGCGAATT TTAACAAAAT ATTAACGTTT ACAATTTCAG GTGGCACTTT

481 TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTC TAAATACATT CAAATATGTA

541 TCCGCTCATG AGACAATAAC CCTGATAAAT GCTTCAATAA TATTGAAAAA GGAAGAGTAT

601 GAGTATTCAA CATTTCCGTG TCGCCCTTAT TCCCTTTTTT GCGGCATTTT GCCTTCCTGT

661 TTTTGCTCAC CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT TGGGTGCACG

721 AGTGGGTTAC ATCGAACTGG ATCTCAACAG CGGTAAGATC CTTGAGAGTT TTCGCCCCGA

781 AGAACGTTTT CCAATGATGA GCACTTTTAA AGTTCTGCTA TGTGGCGCGG TATTATCCCG

841 TATTGACGCC GGGCAAGAGC AACTCGGTCG CCGCATACAC TATTCTCAGA ATGACTTGGT

901 TGAGTACTCA CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA GAGAATTATG

961 CAGTGCTGCC ATAACCATGA GTGATAACAC TGCGGCCAAC TTACTTCTGA CAACGATCGG

1021 AGGACCGAAG GAGCTAACCG CTTTTTTGCA CAACATGGGG GATCATGTAA CTCGCCTTGA

1081 TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT ACCAAACGAC GAGCGTGACA CCACGATGCC

1141 TGCAGCAATG GCAACAACGT TGCGCAAACT ATTAACTGGC GAACTACTTA CTCTAGCTTC

1201 CCGGCAACAA TTAATAGACT GGATGGAGGC GGATAAAGTT GCAGGACCAC TTCTGCGCTC

1261 GGCCCTTCCG GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC GTGGGTCTCG

1321 CGGTATCATT GCAGCACTGG GGCCAGATGG TAAGCCCTCC CGTATCGTAG TTATCTACAC

1381 GACGGGGAGT CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA TAGGTGCCTC

1441 ACTGATTAAG CATTGGTAAC TGTCAGACCA AGTTTACTCA TATATACTTT AGATTGATTT

1501 AAAACTTCAT TTTTAATTTA AAAGGATCTA GGTGAAGATC CTTTTTGATA ATCTCATGAC

1561 CAAAATCCCT TAACGTGAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG AAAAGATCAA

1621 AGGATCTTCT TGAGATCCTT TTTTTCTGCG CGTAATCTGC TGCTTGCAAA CAAAAAAACC

1681 ACCGCTACCA GCGGTGGTTT GTTTGCCGGA TCAAGAGCTA CCAACTCTTT TTCCGAAGGT

1741 AACTGGCTTC AGCAGAGCGC AGATACCAAA TACTGTCCTT CTAGTGTAGC CGTAGTTAGG

1801 CCACCACTTC AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC

1861 AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA GACGATAGTT

1921 ACCGGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGGTTCG TGCACACAGC CCAGCTTGGA

1981 GCGAACGACC TACACCGAAC TGAGATACCT ACAGCGTGAG CTATGAGAAA GCGCCACGCT

2041 TCCCGAAGGG AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG

2101 CACGAGGGAG CTTCCAGGGG GAAACGCCTG GTATCTTTAT AGTCCTGTCG GGTTTCGCCA

2161 CCTCTGACTT GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA

2221 CGCCAGCAAC GCGGCCTTTT TACGGTTCCT GGCCTTTTGC TGGCCTTTTG CTCACATGTT

2281 CTTTCCTGCG TTATCCCCTG ATTCTGTGGA TAACCGTATT ACCGCCTTTG AGTGAGCTGA

2341 TACCGCTCGC CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA

2401 GCGCCTGATG CGGTATTTTC TCCTTACGCA TCTGTGCGGT ATTTCACACC GCATATATGG

2461 TGCACTCTCA GTACAATCTG CTCTGATGCC GCATAGTTAA GCCAGTATAC ACTCCGCTAT

2521 CGCTACGTGA CTGGGTCATG GCTGCGCCCC GACACCCGCC AACACCCGCT GACGCGCCCT

2581 GACGGGCTTG TCTGCTCCCG GCATCCGCTT ACAGACAAGC TGTGACCGTC TCCGGGAGCT

2641 GCATGTGTCA GAGGTTTTCA CCGTCATCAC CGAAACGCGC GAGGCAGCTG CGGTAAAGCT

2701 CATCAGCGTG GTCGTGAAGC GATTCACAGA TGTCTGCCTG TTCATCCGCG TCCAGCTCGT

2761 TGAGTTTCTC CAGAAGCGTT AATGTCTGGC TTCTGATAAA GCGGGCCATG TTAAGGGCGG

2821 TTTTTTCCTG TTTGGTCACT GATGCCTCCG TGTAAGGGGG ATTTCTGTTC ATGGGGGTAA

2881 TGATACCGAT GAAACGAGAG AGGATGCTCA CGATACGGGT TACTGATGAT GAACATGCCC

2941 GGTTACTGGA ACGTTGTGAG GGTAAACAAC TGGCGGTATG GATGCGGCGG GACCAGAGAA

3001 AAATCACTCA GGGTCAATGC CAGCGCTTCG TTAATACAGA TGTAGGTGTT CCACAGGGTA

3061 GCCAGCAGCA TCCTGCGATG CAGATCCGGA ACATAATGGT GCAGGGCGCT GACTTCCGCG

3121 TTTCCAGACT TTACGAAACA CGGAAACCGA AGACCATTCA TGTTGTTGCT CAGGTCGCAG

3181 ACGTTTTGCA GCAGCAGTCG CTTCACGTTC GCTCGCGTAT CGGTGATTCA TTCTGCTAAC

3241 CAGTAAGGCA ACCCCGCCAG CCTAGCCGGG TCCTCAACGA CAGGAGCACG ATCATGCGCA

3301 CCCGTGGGGC CGCCATGCCG GCGATAATGG CCTGCTTCTC GCCGAAACGT TTGGTGGCGG

3361 GACCAGTGAC GAAGGCTTGA GCGAGGGCGT GCAAGATTCC GAATACCGCA AGCGACAGGC

3421 CGATCATCGT CGCGCTCCAG CGAAAGCGGT CCTCGCCGAA AATGACCCAG AGCGCTGCCG

3481 GCACCTGTCC TACGAGTTGC ATGATAAAGA AGACAGTCAT AAGTGCGGCG ACGATAGTCA

3541 TGCCCCGCGC CCACCGGAAG GAGCTGACTG GGTTGAAGGC TCTCAAGGGC ATCGGTCGAG

3601 ATCCCGGTGC CTAATGAGTG AGCTAACTTA CATTAATTGC GTTGCGCTCA CTGCCCGCTT

3661 TCCAGTCGGG AAACCTGTCG TGCCAGCTGC ATTAATGAAT CGGCCAACGC GCGGGGAGAG

3721 GCGGTTTGCG TATTGGGCGC CAGGGTGGTT TTTCTTTTCA CCAGTGAGAC GGGCAACAGC

3781 TGATTGCCCT TCACCGCCTG GCCCTGAGAG AGTTGCAGCA AGCGGTCCAC GCTGGTTTGC

3841 CCCAGCAGGC GAAAATCCTG TTTGATGGTG GTTAACGGCG GGATATAACA TGAGCTGTCT

3901 TCGGTATCGT CGTATCCCAC TACCGAGATA TCCGCACCAA CGCGCAGCCC GGACTCGGTA

3961 ATGGCGCGCA TTGCGCCCAG CGCCATCTGA TCGTTGGCAA CCAGCATCGC AGTGGGAACG

4021 ATGCCCTCAT TCAGCATTTG CATGGTTTGT TGAAAACCGG ACATGGCACT CCAGTCGCCT

4081 TCCCGTTCCG CTATCGGCTG AATTTGATTG CGAGTGAGAT ATTTATGCCA GCCAGCCAGA

4141 CGCAGACGCG CCGAGACAGA ACTTAATGGG CCCGCTAACA GCGCGATTTG CTGGTGACCC

4201 AATGCGACCA GATGCTCCAC GCCCAGTCGC GTACCGTCTT CATGGGAGAA AATAATACTG

4261 TTGATGGGTG TCTGGTCAGA GACATCAAGA AATAACGCCG GAACATTAGT GCAGGCAGCT

4321 TCCACAGCAA TGGCATCCTG GTCATCCAGC GGATAGTTAA TGATCAGCCC ACTGACGCGT

4381 TGCGCGAGAA GATTGTGCAC CGCCGCTTTA CAGGCTTCGA CGCCGCTTCG TTCTACCATC

4441 GACACCACCA CGCTGGCACC CAGTTGATCG GCGCGAGATT TAATCGCCGC GACAATTTGC

4501 GACGGCGCGT GCAGGGCCAG ACTGGAGGTG GCAACGCCAA TCAGCAACGA CTGTTTGCCC

4561 GCCAGTTGTT GTGCCACGCG GTTGGGAATG TAATTCAGCT CCGCCATCGC CGCTTCCACT

4621 TTTTCCCGCG TTTTCGCAGA AACGTGGCTG GCCTGGTTCA CCACGCGGGA AACGGTCTGA

4681 TAAGAGACAC CGGCATACTC TGCGACATCG TATAACGTTA CTGGTTTCAC ATTCACCACC

4741 CTGAATTGAC TCTCTTCCGG GCGCTATCAT GCCATACCGC GAAAGGTTTT GCGCCATTCG

4801 ATGGTGTCCG GGATCTCGAC GCTCTCCCTT ATGCGACTCC TGCATTAGGA AGCAGCCCAG

4861 TAGTAGGTTG AGGCCGTTGA GCACCGCCGC CGCAAGGAAT GGTGCATGCA AGGAGATGGC

4921 GCCCAACAGT CCCCCGGCCA CGGGGCCTGC CACCATACCC ACGCCGAAAC AAGCGCTCAT

4981 GAGCCCGAAG TGGCGAGCCC GATCTTCCCC ATCGGTGATG TCGGCGATAT AGGCGCCAGC

5041 AACCGCACCT GTGGCGCCGG TGATGCCGGC CACGATGCGT CCGGCGTAGA GGATCGAGAT

5101 CTCGATCCCG CGAAATTAAT ACGACTCACT ATAGGGGAAT TGTGAGCGGA TAACAATTCC

5161 CCTCTAGAAA TAATTTTGTT TAACTTTAAG AAGGAGATAT ACATATGTTG GTGGTCCTCT

5221 GGGCCTTGGG GGCGTGGGCC GCGGAGTGCC CCGCGCCCAA GGTGGAAGTC ACGTTCCGCT

5281 ACGTCCCCCT GCCTGACGAA ACCGTGACCT CGGTGAGCCT GAGGGGCTCG TTCAACAACT

5341 GGGGCCAGTG GCCGATGGAA TTACAGCCCG ACGGGACCTG GACGATCACC GTCTGCCTCG

5401 AGCCCGGCAT GCACCAGTAC AAGTTCTTCA TCAATGGACA GTGGCCGAGG GACATGGCCA

5461 CGGCCCGGGG CGGCGGCCCG GTGGACCCCG AGGCCGACGG GTACATCGAC GATGGCTTCG

5521 GCGGCCAAAA CGCCTACCGC CTGGTCAAGC CCTGGCTGGT GGGGGTCTCC CCCTTCCACG

5581 ACCCTACAGA CCCCGCCTTC CTTTGCGTGG CCGACGGCCG ACTCGTTGTC CGCCTTCAGA

5641 CCGCCCGGGA CCAGGTGGAA AGCGTGCGAT TGGTGGCGGA AACCGGGGAA TGGCCCATGG

5701 AAAAGCAGCT CTGGTGGGAG TGGGGCGAGG TCTGGCGGGT CTCCCTTTCC ACCGTGGAGC

5761 CCCTGCGCTA CCGCTTCGAG GGCACGGCCA TCGACGGTTC ACCCTTTGTG TACCCCGCGG

5821 ACAACACCTC TTTCACCTTC GATGGAGTGG ACCGCTTCCC CCAGCCTTCC TGGGTGAGCG

5881 ATGGGATCGG CTACCAAATC TTTCCCGACC GGTTCTTCAA CGGGGAAACC ACCAACGACG

5941 CGTTGGCCCT GGAGACCGAC GAATTTTTGT TCAACGAGCT TTGGACCGAA GGTGGCCCCG

6001 TGCTGTCCGC GTGGAACGAC CCCATCACCC CGCTTCACTG CTGCCACCAG TACTTCGGGG

6061 GCGATCTTCA GGGGATCCTG GCCAAGCTCG ACTACCTCCA GGGCCTGGGG GTGACGGTGC

6121 TCTACCTAAA CCCGGTGTTC GACTCAGGCA GCGCGCATGG CTACGACACC CACGACTACC

6181 TCAAGGTCAG CCCGAAGTTC GGGACCGAAG AAGACCTCCG GAGGCTTTTG GACGAGGCCC

6241 ATGCCCGGGG GATGAAGGTG CTTTTCGACT TCGTGCCCAA CCACACGGGG ATCGGGTTCT

6301 GGGCGTTCCA GGACGCCTGG AAGAGGGGGC CGGAGAGCCC GTACTGGGAC TGGTACTTCA

6361 TTCGGCGCTG GCCGTTTCGG GCCGGGGATG GCACCGCGTA CGAGGGCTGG TGGGGTCTGG

6421 GAAGCCTCCC CAAACTCAAC ACCGAAAACC CGGAGGTGAA GGAATACCTC TTTGGGGTTG

6481 TGCTCCACTG GCTGGAGTTT GGCTTCGATG GGCTGAGGGT GGACGTGCCC AACGAGCTTG

6541 TCAACGCCCA CGAGTTTTTC CGGGAGCTGC GCAAGCTGGT AAAAGAGAAG TTCCCCCAGG

6601 CCTACCTTGT GGCGGAGATT TGGCAGCTGG CGCCGGTGTG GGTCCGGGGG GACCAGTTCG

6661 ATTCGCTGAT GAACTATGCC CTGGGTCGGG ACATCTTGCT CCGCTACGCC CGGGGCGAGC

6721 TTAGTGGCGA GCGGGCCTTG GAGGAGCTCA GCCGGTACTT CGCCGCCTAC GGGGAGAACG

6781 TGGCCGCCAT GGGGTTCAAC CTCTTAAGCT CCCACGACAC GGGCCGGATC CTCACCGACC

6841 TTGGCGGGGG CAACTTCGGG GAGGAGATCA GCCCTTTGGG CCGCCGGCGC CTTAAGCTCC

6901 TTTCCACCTT GCTTTACACC CTGCCGGGAA TGCCCGTGAC CTTCCAGGGC GACGAGCGCG

6961 GAACCCTCGG AGAGAAGACC CTCTACGATG CCCACCGGTA CCCGATCCAG TGGGACCGGC

7021 TGGACGAGGA CATCTTCGGC CACTATGTGA AGCTCGCTCG CATGCGGGAG GAGATCCCGG

7081 CCCTGCGGAG CAGCGCGGTG TGGACCTATG CGGCCAAGGG CGGGGTTTTG GCCTTCTTCC

7141 GCGGGCAGGC CAAAGAGGTG CTGGTGATCG CCAACAGCGA CCCCCGGCCG GTGGTGTTCT

7201 CCTTGCCGGA AGGCCGGTGG CAGGTTTGGG AAACGGGCGC CGTGCTCCAG GGGAGCACCA

7261 TCGTGCCCGC CCTTCACACC TGGGTCCTCC TCCGCCTAGC GGCCGCACTC GAGCACCACC

7321 ACCACCACCA CTGAGATCCG GCTGCTAACA AAGCCCGAAA GGAAGCTGAG TTGGCTGCTG

7381 CCACCGCTGA GCAATAACTA GCATAACCCC TTGGGGCCTC TAAACGGGTC TTGAGGGGTT

7441 TTTTGCTGAA AGGAGGAACT ATATCCGGAT

Pull3.1 is indicated in the highlighted grey area. The stop codon in red, the start codon in green and the his-tag in is highlighted in yellow.