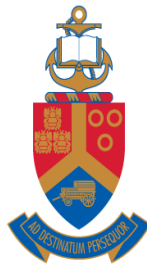


Structure-function analysis of a novel multi-functional glycoside hydrolase

by

Ntsoaki Leticia Mosina

Submitted in partial fulfilment of the requirements of the degree
Philosophiae Doctor (PhD) Biotechnology



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Department of Genetics
Faculty of Natural and Agricultural Science
University of Pretoria
June 2018

DECLARATION

I, Ntsoaki Leticia Mosina, declare that the thesis, which I hereby submit for the degree of *Philosophiae Doctor* (PhD) in Biotechnology at the University of Pretoria, is my own work and has not been submitted by me for a degree at this or any other tertiary institution.

.....

Signature of candidate

.....

Date

Acknowledgements

I would like to thank my supervisors, Prof. Don Cowan and Prof. Wolf-Dieter Schubert for their guidance and supervision.

I am thankful for financial support from the National Research Foundation (Innovation) and University of Pretoria.

I would also like to thank the DST-Women in Science and TATA for the 2015 Women in Science Doctoral Fellowship.

I would also like to thank the Structural Biology Group and Centre for Microbial Ecology and Genomics for their support.

To my adorable son Tshedza Mboneni, you truly brought the light that gave me the strength to pursue and complete this PhD.

To Tondani Mboneni and my family, I would like to thank you for your emotional support, sharing ideas, your contribution on my experiments and for always been there day in and day out.

Table of Contents

List of abbreviations	i
List of tables	iv
List of figures.....	v
List of publications and conference contributions.....	viii
Abstract.....	ix
Chapter 1	1
Literature review	1
1. Introduction	1
1.1. Structural composition of lignocellulose	1
1.1.1. Structural components of cellulose.....	1
1.1.2. Structural components of hemicellulose.....	2
1.1.2.1. Occurrence and function of xylan	2
1.1.2.2. Occurrence and function of mannan.....	3
1.2. Lignocellulose degradation.....	6
1.2.1. Cellulases.....	6
1.2.2. Hemicellulases.....	8
1.2.2.1. Xylanases.....	8
1.2.2.2. Mannanases	9
1.2.3. Industrial application of cellulases and hemicellulases	11
1.3. Microbial glycoside hydrolases	13
1.3.1. Classification and nomenclature of glycoside hydrolases	13
1.3.2. Catalytic mechanism of glycoside hydrolases.....	13
1.3.2.1. Retaining glycoside hydrolases	14
1.3.2.2. Inverting glycoside hydrolases	14
1.3.3. Glycoside hydrolase family 5	15
1.3.4. Glycoside hydrolase family 6	16
1.3.5. Carbohydrate-binding modules	17
1.4. Methods in protein structure determination	18
1.4.1. Data collection by X-ray diffraction.....	19
1.4.2. Phasing techniques	20
1.4.2.1. Molecular replacement	20
1.4.2.1.1. Factors affecting molecular replacement	21

1.4.3. Isomorphous replacement.....	23
1.4.4. Multi-wavelength anomalous dispersion.....	23
1.4.5. Crystal structure of glycoside hydrolases.....	24
1.5. Rationale.....	29
1.6. Aims and objectives.....	29
Chapter 2.....	30
Materials and Methods.....	30
2.1. Cultivation and cloning of the full length glycoside hydrolase and corresponding domains.....	32
2.1.1. Polymerase chain reaction.....	32
2.1.2. Restriction digestion and ligation reactions.....	34
2.1.3. Preparation of chemically competent <i>E.coli</i> cells.....	35
2.1.4. Transformation by heat shock using chemically competent cells.....	36
2.2. Optimisation of protein expression.....	37
2.2.1. Small scale protein expression and analysis of <i>PmGH</i>	37
2.2.2. Large scale protein expression.....	38
2.3. Protein purification.....	38
2.3.1. Immobilised metal affinity chromatography.....	39
2.3.2. Gel filtration chromatography of <i>PmGH</i>	40
2.4. Enzyme characterisation.....	40
2.4.1. Determination of the range of hydrolytic activities.....	41
2.4.2. pH and temperature optimum and stability.....	42
2.4.3. <i>PmGH</i> identification by mass spectrometry.....	43
2.5. Homology modelling.....	44
2.6. Protein crystallisation.....	45
2.6.1. Crystal soaking.....	45
2.7. Data collection, phasing and refinement.....	47
Chapter 3.....	48
Results.....	48
3.1. Cloning and expression.....	48
3.2. Protein purification.....	49
3.3. Functional characterisation.....	51
3.3.1. Substrate range of <i>PmGH</i> , GH5 and GH6.....	51
3.3.2. pH and temperature optima profiles of <i>PmGH</i> , GH5 and GH6.....	52

3.3.3. pH stability profiles of <i>PmGH</i> , GH5 and GH6	54
3.3.4. Thermostability profiles of <i>PmGH</i> , GH5 and GH6.....	58
3.4. Dinitrosalicylic acid assay standard curves	62
3.5. Homology modelling of <i>PmGH</i>	64
3.6. Screening and optimisation of protein crystallisation conditions.....	67
3.7. X-ray diffraction and data analysis	74
Chapter 4	76
Discussion	76
Chapter 5	85
Concluding remarks and future studies	85
References.....	87
Appendix A.....	99
Appendix B.....	100
Appendix C.....	105

List of abbreviations

°C	Degree Celsius
µg	Microgram
µl	Microlitre
3-D	Three dimension
Å	Angstrom
BLAST	Basic local alignment search tool
bp	Base pair
CAZy	Carbohydrate Active enZyme
CBM	Carbohydrate binding module
CMC	Carboxymethyl cellulose
COOT	Crystallographic Object Oriented Toolkit
Cryo-EM	Cryogenic electron microscopy
DNA	Deoxyribonucleic acid
DNS	Dinitrosalicylic acid
<i>E.coli</i>	<i>Escherichia coli</i>
ESRF	European Synchrotron Radiation Facility
FPLC	Fast Protein Chromatography
GH	Glycoside Hydrolase
GMQE	Global model quality estimation
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl- β-D-1-thiogalactopyranoside
kDa	Kilodalton
LB	Luria-bertani

LC-MS	Liquid chromatography-mass spectrometry
M	Molar
MAD	Multi-wavelength anomalous dispersion
Man	Mannanase
mg	Milligram
MgCl₂.6H₂O	Magnesium chloride hexahydrate
min	Minute
MIR	Multiple isomorphous replacement
ml	Millilitre
mM	Millimolar
MR	Molecular replacement
NCBI	National Center for Biotechnology Information
ng	Nanogram
nkat	Nanokatal
nmol	Nanomole
NMR	Nuclear magnetic resonance
OD	Optical density
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
pI	Isoelectric point
<i>PmGH</i>	<i>Paenibacillus mucilaginosus</i> glycoside hydrolase
R²	R-squared
s	Second
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SIR	Single isomorphous replacement
TbCl₂	Terbium chloride
USA	United States of America
V	Voltage
Xyl	Xylanase

List of tables

Table 1.1: A selection of industrial applications that utilise glycoside hydrolases	12
Table 1.2: Summary of the types of the prominent catalytic folds across the GH clans	25
Table 2.1: PCR primers and thermal cycling parameters for all amplicons	33
Table 2.2: List of plasmid-vector constructs generated and expected sizes	35
Table 2.3: Theoretical molecular weight and isoelectric points	39
Table 2.4: Summary of the glycoside hydrolase activities tested and the corresponding substrates	42
Table 2.5: Structural homologs of the composite domains of <i>PmGH</i>	44
Table 3.1: Polymeric substrates and hydrolytic activities detected in <i>PmGH</i> , GH5 and GH6 using the DNS assay	52
Table 3.2: Summary of homology modelling data using SwissModel and the homologs used for the native enzyme and its composite domains	66
Table 3.3: Screening conditions and incubation times for crystallisation of <i>PmGH</i>	67
Table 3.4: Data-collection statistics of the highest quality X-ray diffraction data obtained from <i>PmGH</i> protein crystals at the ESRF	75

List of figures

Figure 1.1: Structural composition of mannan and its derivatives: (a) mannan; (b) galactomannan; (c) glucomannan; (d) galactoglucomannan	5
Figure 1.2: (a) Structure of xylan and its xylanolytic target sites. (b) Hydrolysis of xylo-oligosaccharide by β -xylosidase	9
Figure 1.3: Schematic representation of galactoglucomannan hydrolysis by mannanases	10
Figure 1.4: General catalytic mechanism of (a) retaining and (b) inverting glycoside hydrolase	15
Figure 2.1: Schematic overview of the methodical approach employed in determining the functional and structural properties of the novel glycoside hydrolase, <i>PmGH</i>	30
Figure 2.2: Diagrammatic overview of the protein crystallisation strategy employed in the crystallisation of the multi-modular protein, <i>PmGH</i>	46
Figure 3.1: PCR products of (a) <i>PmGH</i> , (b) GH5 and (c) GH6 and subsequent identification by BLAST analysis and NCBI conserved domain search	48
Figure 3.2: Optimisation of the small scale production of <i>PmGH</i> by varying (a) induction time (hours) and (b) IPTG concentration	49
Figure 3.3: Purification of <i>PmGH</i> by gravity flow IMAC and gel filtration for crystallisation and enzyme characterisation	50
Figure 3.4: Partial purification of (a) GH5 and (b) GH6 catalytic domains by gravity flow IMAC using Talon CellThru resin charged with cobalt	51
Figure 3.5: pH optima of the native enzyme, <i>PmGH</i> , and its composite catalytic domains, GH5 and GH6	53

Figure 3.6: Temperature optima of the native enzyme, <i>PmGH</i> , and its composite catalytic domains, GH5 and GH6	53
Figure 3.7: pH stability profiles of the (a) cellulase, (b) mannanase and (c) xylanase activities detected in the native enzyme, <i>PmGH</i>	54
Figure 3.8: pH stability profiles of the (a) cellulase, (b) mannanase and (c) xylanase activities detected in the GH5 domain	56
Figure 3.9: pH stability profile of the cellulase activity detected in the GH6 domain at pH 4-9 over a 2 hour incubation period	57
Figure 3.10: Thermostability profiles of the (a) cellulase, (b) mannanase and (c) xylanase activities detected in the native enzyme, <i>PmGH</i>	58
Figure 3.11: Thermostability profiles of the (a) cellulase, (b) mannanase and (c) xylanase activities detected in the GH5 domain	60
Figure 3.12: Thermostability profile of GH6 over a 2 hour incubation period at temperatures ranging from 40-90°C under optimal pH conditions (pH 5)	61
Figure 3.13: Glucose standard curves for pH 4-9 using the DNS assay	62
Figure 3.14: Mannose standard curves for pH 4-9 using the DNS assay	63
Figure 3.15: Xylose standard curves for pH 4-9 using the DNS assay	64
Figure 3.16: Homology model of <i>PmGH</i> and its corresponding catalytic domains	65
Figure 3.17: Screening conditions showing potential conditions for <i>PmGH</i> crystallisation	68
Figure 3.18: Improvement of protein crystals grown in condition D by bead seeding with a 10 ¹ dilution	69

Figure 3.19: Optimisation of crystallisation condition E by varying the concentration of MgCl ₂ .6H ₂ O and PEG 10 000	70
Figure 3.20: Optimisation of <i>PmGH</i> protein crystals by bead seeding	71
Figure 3.21: Optimisation of <i>PmGH</i> crystals by a second round of bead seeding	72
Figure 3.22: Final <i>PmGH</i> crystals seeded with a 10 ⁻³ dilution and incubated at 12°C for 7 days	73
Figure 3.23: X-ray diffraction image of the highest quality data (2.53 Å) collected from <i>PmGH</i> protein crystals at the ESRF	74

List of publications and conference contributions

Publication

Mosina NL, Schubert –W-D, Cowan DA. (2019). Characterization and homology modelling of a novel multi-modular and multi-functional *Paenibacillus mucilaginosus* glycoside hydrolase. *Extremophiles* <https://doi.org/10.1007/s00792-019-01121-8>.

Conference contributions

Mosina NL, Schubert W-D, Cowan DA. Structure-function analysis of a novel, multi-functional glycoside hydrolase. Oral presentation, Thermophiles International Conference, Skukuza, South Africa, September 2017.

Mosina NL, Schubert W-D, Cowan DA. Functional characterisation of a multi-specificity functional exoglucanase from *Paenibacillus mucilaginosus*. Poster presentation, Genomics Research Institute Research Symposium, Pretoria, South Africa, October 2016.

Mosina NL, Schubert W-D, Cowan DA. Cloning, purification and crystallisation of a bi-functional *Paenibacillus mucilaginosus* exoglucanase. Poster presentation, South African Society for Microbiology, Durban, South Africa, January 2016.

Mosina NL, Schubert W-D, Cowan DA. Cloning, purification and crystallisation of a bi-functional *Paenibacillus mucilaginosus* exoglucanase. Poster presentation, Genomics Research Institute Research Symposium, Pretoria, South Africa, October 2015.

Abstract

The present study presents the research findings of the first ever multi-modular *Paenibacillus mucilaginosus* glycoside hydrolase (*PmGH*). Furthermore, we report the successful crystallisation of a multi-modular GH. The GH is composed of two catalytic modules (GH5 and GH6) and two carbohydrate binding modules (both CBM3). Functional analysis demonstrated that the cellulase, mannanase and xylanase activities of *PmGH* (130 kDa) were attributed to the GH5 catalytic domain. The presence of the GH6 catalytic domain resulted in slightly increased cellulase activity in *PmGH*. Optimal *PmGH* activity and functional stability was highest at pH 4-6 and at 40-60°C. The structural properties of *PmGH* that determine its robust nature were further investigated.

Homology modelling of *PmGH* showed the GH5 and GH6 domains to be independent but provided no structural information for the CBMs and linker regions. However, successful homology modelling of the individual domains indicated that the combination of the modules makes *PmGH* structurally and functionally novel. Glycoside hydrolases occur as independent modules or as part of a multi-modular protein with other catalytic and/or non-catalytic modules. Multiple combinations of these modules can occur in nature resulting in novel proteins such as *PmGH*.

In an attempt to determine the *PmGH* crystal structure, a range of crystallisation conditions were tested. After extensive screening and optimisation, multiple *PmGH* crystals were diffracted, using both local diffraction and Synchrotron radiation sources (ESRF, France). Overall ~90% of the *PmGH* protein crystals did not diffract and of the remaining ~10% yielded unsatisfactory data. Phasing by molecular replacement also yielded no structural solutions. Alternative phasing methods such as multi-wavelength anomalous dispersion were also unsuccessful due to the quality of the diffraction data collected. Given the severe lack of multi-

modular GH crystal structures in protein structure databases, the present study highlights the major limitations in structural studies of these important enzymes.

Chapter 1

Literature review

1. Introduction

1.1. Structural composition of lignocellulose

Lignocellulose is an organic polymer that forms the major structural component of all plant cell walls (Sunna and Antranikian, 1997). It is derived primarily from agricultural, industrial and forestry waste or residues and is composed primarily of cellulose, hemicellulose and lignin (Blum *et al.*, 2000; Järvinen *et al.*, 2012; Subramaniyan and Prema, 2002). As the most abundant polymer on Earth, cellulose forms the major component of the plant cell wall (42-50%) (Chandel *et al.*, 2012). Hemicellulose makes 25-30% of the plant cell wall and is composed of a heterogenous group of compounds such as xylan, α -1,5-L-arabinan, arabino-galactan, mannobiose and galacto-glucomannan (van den Brink and de Vries, 2011). Lignin, the third major component (20-25%) of the plant cell wall, is essential in providing mechanical support to the aerial parts of the plant (Järvinen *et al.*, 2012). This aromatic compound is composed of dimethoxylated, monomethoxylated and non-methoxylated phenyl propanoid units that form a three dimensional network. Three monolignols, namely *p*-coumaryl, coniferyl and synapyl alcohols are the monomers of lignin (Martínez *et al.*, 2005). Cellulose, hemicellulose and lignin are sandwiched together within the plant cell wall and each play a structural and functional role within the plant (Juturu and Wu, 2012).

1.1.1. Structural components of cellulose

Cellulose is a homo-polysaccharide composed of β -D-glucose monomers linked by β -1,4-glycosidic bonds. Linear polymers consisting of approximately 10 000 D-glucose monomers linked *via* β -1,4 glycosidic bonds (Schülein, 2000) adhere to each other in a parallel fashion to

form a crystalline cellulose matrix (Zhang and Lynd, 2004). The crystallinity of cellulose is a result of inter-polymer hydrogen bonds and van der Waals forces (van den Brink and de Vries, 2011). Cellulose crystallinity is not continuous as some regions may be crystalline or amorphous. The interaction of cellulose with hemicellulose and lignin further enhances its recalcitrance in nature. The degree of crystallinity and the overall amount of cellulose present varies across different plant species with higher plant species containing a greater proportion.

1.1.2. Structural components of hemicellulose

Hemicellulose is a heteropolysaccharide composed primarily of xylan, mannan, galactan and arabinan. Xylan is considered the major hemicellulose component of the plant cell wall (Ebringerová and Heinze, 1999). Mannan in plants play a structural role and act as a non-starch reserve in the wall of endosperms of plant seeds (Chauhan *et al.*, 2012). The distribution and chemical composition of hemicelluloses may vary depending on the type of plant.

1.1.2.1. Occurrence and function of xylan

Xylan is the most abundant hemicellulose and occurs as a homopolymer or heteropolymer in the plant cell wall. Xylan is a linear homopolymer composed of β -1,4 linked D-xylose residues and is prominent in lower plant species (Scheller and Ulvskov, 2010). However, some exceptions have been reported in green algae which have a xylan structure consisting of β -1,3 linked D-xylose residues. A mixture of β -1,4 and β -1,3 linked xylan has been reported in red algae. The xylan backbone can be substituted with acetyl groups, arabinose, glucuronic acids, ferulic acid and/or *p*-coumaric acid (Moreira and Filho, 2016; Scheller and Ulvskov, 2010). The occurrence and degree of substitution on xylan is dependent on the botanical source which in turn is related to its structural and functional role within in the plant cell wall (Ebringerová

and Heinze, 1999). For instance, acetylated xylan is essentially a xylan polymer substituted with acetyl groups at the *O*-3 position. The ratio of acetyl groups:D-xylose varies across different plant species with softwoods and hardwoods having a molar ratio of 1:9-12 and 1:5-6, respectively (Bastawde, 1992). Phenolic crosslinks with ferulic acid or 4-coumaric acid enhance the structural integrity of the plant cell wall. Ferulic acid is often esterified to xylan and several di-ferulic acid crosslinks connect the polymer to lignin (O'Neill *et al.*, 1996).

Xylans with a α -(1,2)-linked glucuronosyl and 4-*O*-methyl glucuronosyl modification are termed glucuronoxylans. Arabinoxylans refer to xylan polymers with an L-arabinose substitution on the main chain. Glucuronoarabinoxylan, an abundant hemicellulose in commelinid monocots, has several L-arabinose and glucuronic acid substitutions along the main chain. In hardwood and softwood plants, *O*-acetyl-4-D-methylglucuronoxylan and arabino-4-D-methylglucuronoxylan are the main xylan polymers, respectively (Kulkarni *et al.*, 1999; Sunna and Antranikian, 1997). Plants with a higher evolutionary level often have a higher distribution and frequency of substitutions on the xylan backbone. A number of variations can occur in the xylan structure and, as a result, xylans do not have a repeated structure (Moreira and Filho, 2016; Scheller and Ulvskov, 2010).

1.1.2.2. Occurrence and function of mannan

Based on its structural composition, mannan can be divided into four subfamilies: i) linear mannan; ii) glucomannan; iii) galactomannan; iv) galactoglucomannan (Figure 1.1). All mannan structures are composed of D-mannose backbone linked together *via* β -1,4 glycosidic bonds. However, the backbone of some mannan structures are composed of D-mannose and D-glucose residues (Chauhan *et al.*, 2012). Furthermore, the mannan backbone can have a side chain of α -1,6 linked galactose residues. Linear mannan is composed primarily of β -1,4 linked D-mannose residues and contains less than 5% galactose (Dhawan and Kaur, 2007). This

homopolysaccharide is common in *Aloe vera* and the endosperm of a range of plants such as ivory nuts and green coffee beans. Galactomannan is composed of a mannan backbone and 1,6-linked α -D-galactopyranosyl side chains along the main chains (Figure 1.1b) (Chauhan *et al.*, 2012). The degree of substitution of galactomannan with galactopyranosyl units varies across different plant species. ‘True’ galactomannan polysaccharides contain more than 5% D-galactose units along the mannan backbone. Galactomannan is abundant in the endospermic region of seeds and aids in water retention, particularly in regions with elevated atmospheric temperatures (Yamabhai *et al.*, 2016). The water solubility of galactomannan is increased due to the presence of the hydrophilic D-galactosyl side chains. Therefore, galactomannan polymers with a higher degree of substitution tend to be more water soluble than those with fewer D-galactosyl side chains.

Glucomannan is a heteropolysaccharide abundant in softwood and is composed of β -1,4 linked D-mannose residues and β -1,4 linked D-glucose residues (Figure 1.1c) (Moreira and Filho, 2008). In hardwood plants, glucomannan monomers are linked together either by a β -1,3 or β -1,6 glycosidic bond (Yamabhai *et al.*, 2016). The ratio of glucose to mannose in softwood and hardwood plants is 1:3 and 1:1.5-2, respectively. In coniferous woods, glucomannan and galactomannan can both occur as the main hemicellulose. Like all plant polymers, the degree of polymerisation of glucomannan may vary across different plant species. Furthermore, the nature of the glycosidic bond may vary and deviate from the typical β -1,4 linkage. In seeds and coniferous plants, glucomannan is often in close proximity to the cellulose and xylan component of the cell wall (Dhawan and Kaur, 2007). The paracrystalline arrangement formed by the glucomannan is located between the cellulose microfibrils. Glucomannan has a similar two-fold screw axis as cellulose, although glucomannan has a weaker and less tight crystal packing. Therefore, glucomannan is easily accessible compared to cellulose.

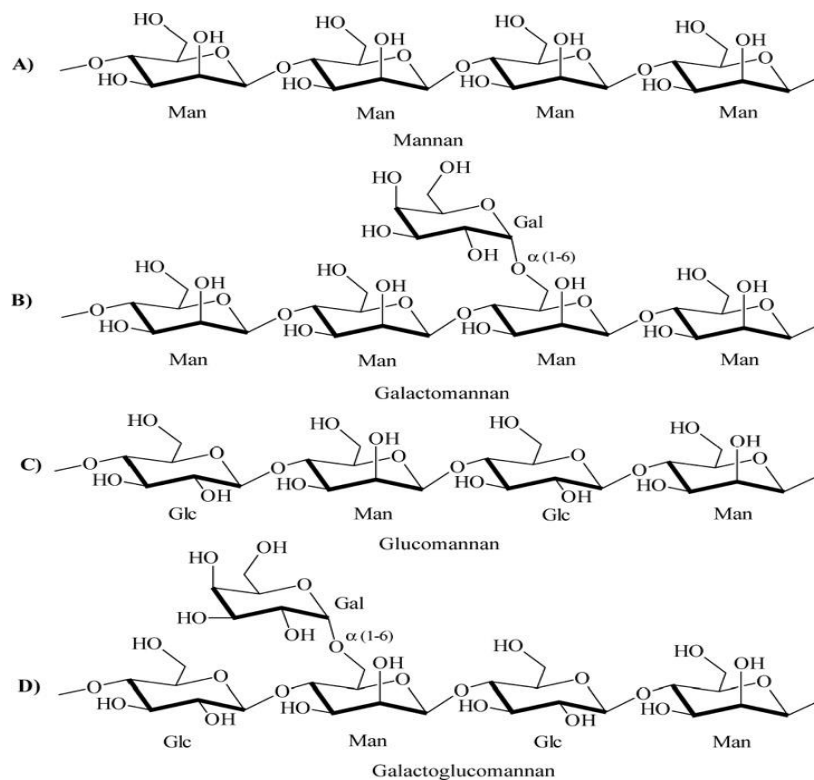


Figure 1.1: Structural composition of mannan and its derivatives: (a) mannan; (b) galactomannan; (c) glucomannan; (d) galactoglucomannan (Dhawan and Kaur, 2007).

The galactoglucomannan main chain is composed primarily of D-mannose residues interspersed with D-glucose residues and α -1,6 linked D-galactosyl sidechains (Figure 1.1d). As a prominent hemicellulose in coniferous plants, the molar ratio of mannose/glucose/galactose is 3:1:1 in galactoglucomannan. In softwood, galactoglucomannan is often acetylated at the C-2 and C-3 position of mannose residues (Lundqvist *et al.*, 2003). Softwoods with acetylated galactoglucomannan may be either galactose rich or galactose poor with a molar ratio of mannose/glucose/galactose of 3:1:1 and 3:1:0.1, respectively (Moreira and Filho, 2008). As for galactomannan, galactoglucomannan is highly soluble in water due to the presence of D-galactosyl side chains which are responsible for forming strong intramolecular hydrogen bonds.

1.2. Lignocellulose degradation

Glycoside hydrolases (GHs) (cellulases, endo- β -1,4-xylanases, xylosidases, α -L-arabinofuranosidases and α -D-glucuronidases, β -mannanase, β -mannosidase, *etc.*), carbohydrate esterases (ferulic acid esterases, acetyl xylan esterases, *p*-coumaric acid esterases, acetyl mannan esterases, *etc.*) and ligninases are the major enzyme groups involved in the enzymatic degradation of lignocelluloses (Biely, 2012; Chávez *et al.*, 2006).

1.2.1. Cellulases

Endo- β -1,4-glucanases, exo- β -1,4-glucanases and β -glucosidases, collectively known as cellulases, work in a synergistic manner to efficiently hydrolyse cellulose to D-glucose (Bayer *et al.*, 1998; Blumer-Schuetz *et al.*, 2008; van den Brink and de Vries, 2011). The amorphous region of cellulose is hydrolysed through the action of extracellular endoglucanases (carboxymethyl cellulase) to produce cello-oligosaccharides (Zhang and Lynd, 2004). Endoglucanases preferentially target the glycosidic bonds within the cellulose structure to increase the accessibility of exoglucanases (Haki and Rakshit, 2003). Numerous bacteria and fungi are known to produce these endo-acting enzymes (Obeng *et al.*, 2017). Optimal endoglucanase activity has been reported under acidic (pH 3.5-10) and thermophilic (30-109°C) conditions (Bischoff *et al.*, 2006; Dutta *et al.*, 2008; Garg *et al.*, 2016; Graham *et al.*, 2011; Pil Lee *et al.*, 2017). Exoglucanases, also referred to as cellobiohydrolases (CBH), are extracellular enzymes that preferentially target the free chain ends of cello-oligosaccharides to produce cellobiose (Chandel *et al.*, 2012). Optimal exoglucanase activity has been reported at pH conditions ranging from pH 3.5-10 (Lee *et al.*, 2012; Ma *et al.*, 2015; Tuohy *et al.*, 2002) and at temperatures ranging from 30-105°C (Kim *et al.*, 2008; Ruttersmith and Daniel, 1991; Yeoman *et al.*, 2010). Cellobiose can be effectively hydrolysed by β -glucosidases to D-glucose units, although glucose has been shown to act as a competitive inhibitor of cellulose hydrolysis.

Similarly, β -glucosidases are produced by a wide array of microorganisms, and can be found as intracellular, extracellular or cell wall-associated enzymes. β -Glucosidases are often active over a broad pH range and temperatures ranging from 45-75°C (Cantarel *et al.*, 2009). Aerobic microorganisms such as *Bacillus* sp., *Streptomyces* sp., *Trichoderma viride*, *Aspergillus niger*, *Phanaerochaete chrysosporium* and *Thermomyces lanuginosus* produce cellulases as individual enzymes which work in synergy (Kazemi *et al.*, 2014; Li *et al.*, 2012; Park *et al.*, 2005; Subramaniyan and Prema, 2002). In contrast to aerobic bacteria, anaerobic bacteria form supramolecular complexes in order to hydrolyse cellulose.

Anaerobic bacteria such as *Clostridium cellulovorans*, *Clostridium thermocellum* and *Ruminococcus flavefaciens* form high molecular weight extracellular enzyme complexes termed the cellulosome (Blumer-Schuette *et al.*, 2008). Several glycoside hydrolases (GHs) and some carbohydrate esterases arrange themselves around the non-protein component called scaffolding proteins (Xu *et al.*, 2009). Scaffolding proteins and the catalytic components of the cellulosome contain several highly conserved cohesin and dockerin modules, respectively (Béguin and Lemaire, 1996; Gilbert, 2007). Complex assembly is initiated by the binding of type I dockerins to type I cohesins. Attachment of the cellulosome to the bacterium is achieved through the binding of type II dockerins located on the C-terminus of the scaffolding proteins to type II cohesins that are secured to the peptidoglycan layer of the bacterial cell wall (Bayer *et al.*, 1998; Pinheiro *et al.*, 2009). Cohesins and dockerins have a high sequence homology, although cross-specificity does not occur. All cohesins and dockerins are species-specific; cohesins from one species cannot therefore recognise and bind to dockerins of another species (Xu *et al.*, 2009).

1.2.2. Hemicellulases

Hemicellulases refer to a diverse group of enzymes responsible for the degradation of hemicellulose. Considering the diverse nature of hemicellulose structures across different plant species, an array of enzymes is required for its hydrolysis (Ebringerová and Heinze, 1999). With xylan and mannan as the major hemicellulose components, a number of xylan- and mannan-degrading microbial enzymes have been identified. Xylan- and mannan-degrading enzymes have become the key enzymes for hemicellulose degradation.

1.2.2.1. Xylanases

Xylan degrading enzymes refer to a group of enzyme responsible for the conversion of xylan to D-xylose (Figure 1.2). Depending on the nature of the xylan substrate, additional enzymes termed accessory enzymes may be required for the complete degradation of xylan to its monomeric sugar. Xylan-degrading enzymes include endo-1,4- β -xylanases, exo-1,4- β -xylosidases (Walia *et al.*, 2017). Accessory enzymes may include acetyl xylan esterases, ferulic acid esterases and *p*-coumaric acid esterases, depending on the side chain residues (Bastawde, 1992; Collins *et al.*, 2005). Xylanases target the internal β -(1,4)-glycosidic bonds of the xylan backbone to produce xylo-oligosaccharides (Kulkarni *et al.*, 1999; Sunna and Antranikian, 1997). The xylo-oligosaccharides are further hydrolysed to D-xylose monomers by xylosidases. For more complex, modified xylan substrates, various carbohydrate esterases such as acetyl xylan esterase can be employed to remove acetyl side chains or other side groups. Carbohydrate esterases indirectly increase xylanase activity by removing side chains from the xylan backbone which often sterically hinder xylanases. The presence of side chains often limits xylanase activity due to the inability to access the xylan substrate.

Xylanases are common in nature and have been reported in numerous bacteria and fungi such as *Bacillus* sp., *Clostridium* sp., *Aspergillus niger* and *Streptomyces* sp. (Kumar *et al.*, 2016; Subramaniyan and Prema, 2002; Sunna and Antranikian, 1997). Furthermore, microbial xylanases differ in their structural and functional properties. Xylanases have been reported to be active under a wide range of pH (pH 2-11) and temperature (5-105°C) conditions (Collins *et al.*, 2005).

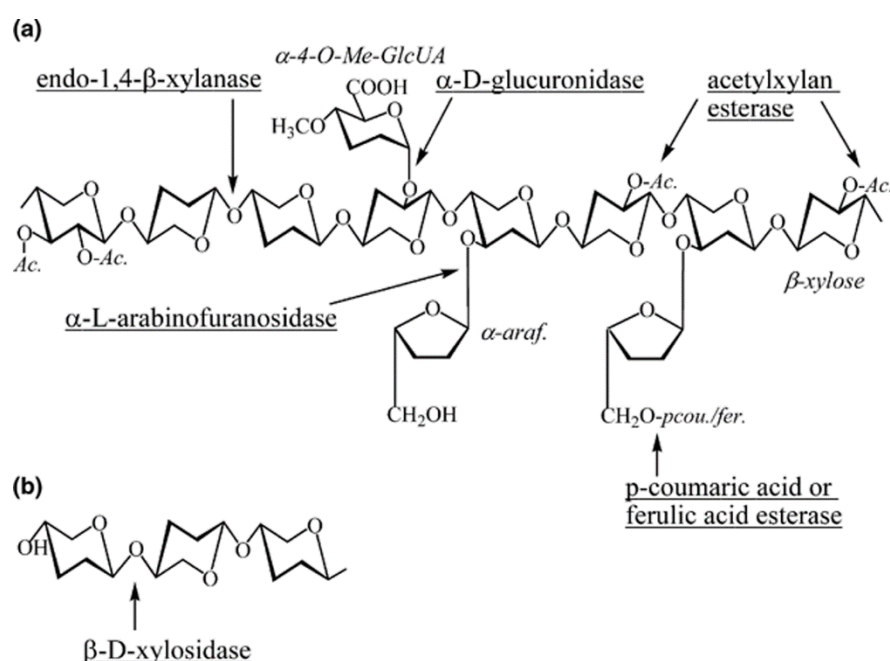


Figure 1.2: (a) Structure of xylan and its xylanolytic target sites. (b) Hydrolysis of xylo-oligosaccharide by β -xylosidase (Collins *et al.*, 2005).

Ac: acetyl group; α -araf: α -arabinofuranose; α -4-O-Me-GlcUA: α -4-O-methylglucuronic acid; *pcou*: *p*-coumaric acid; *fer*: ferulic acid.

1.2.2.2. Mannanases

The complete degradation of mannan is achieved by various mannan-degrading enzymes and accessory enzymes (Figure 1.3). Mannan-degrading enzymes include 1,4- β -D-mannan

mannohydrolase (β -mannanase), 1,4- β -D-mannopyranoside hydrolase (β -mannosidase) and 1,4- β -D-glucoside hydrolase (β -glucosidase). Accessory enzymes such as acetyl mannan esterases and 1,6- α -D-galactoside galactohydrolase (α -galactosidase) remove substituents from the side chain that sterically hinder β -mannanases. Therefore, accessory enzymes increase the accessibility of mannan main chain, rendering it susceptible to enzymatic attack.

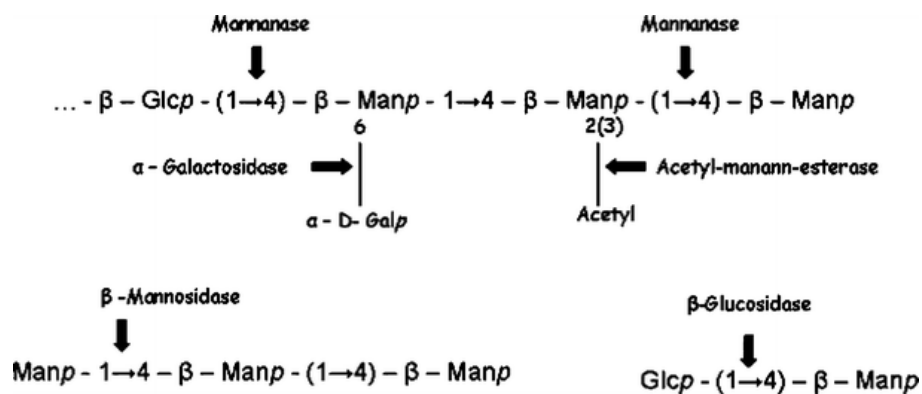


Figure 1.3: Schematic representation of galactoglucomannan hydrolysis by mannanases (Moreira and Filho, 2008).

β -Mannanases hydrolyse the internal β -1,4-glycosidic linkages of the main chain releasing mannosides which are susceptible to hydrolytic attack by β -mannosidases (Moreira and Filho, 2008; Yamabhai *et al.*, 2016). β -Mannosidases subsequently release monomeric D-mannose sugars from the non-reducing end of mannan and manno-oligosaccharides (Dias *et al.*, 2004). The hydrolysis of glucomannan and galactoglucomannan also requires the hydrolytic action of β -glucosidases which target 1,4- β -D-glucopyranose from the non-reducing end of the manno-oligosaccharides. α -Galactosidases target α -1,6 linked D-galactopyranosyl units released from galactomannan and galactoglucomannan by β -mannanases (Reddy *et al.*, 2016). Acetyl mannan esterases target the acetyl groups on galactoglucomannan, allowing the mannan-degrading enzyme to access the main chain (Kanelli and Topokas, 2017). The complete

degradation of mannan and its derivatives to its monomeric sugars is achieved through the synergistic action of mannan-degrading enzymes and accessory enzymes. Mannan-degrading enzymes have been identified in a range of bacteria and fungi such as *Bacillus* sp., *Aeromonas hydrophila*, *Aspergillus oryzae* and *Paenibacillus* sp. (Chauhan and Gupta, 2017; Chauhan *et al.*, 2012; Sakai *et al.*, 2017). Like most glycoside hydrolases, microbial mannan-degrading enzymes are robust biocatalysts that are active over a broad pH (pH 3-7.11) and temperature (45-92°C) range (Dhawan and Kaur, 2007; Moreira and Filho, 2008; Zhao *et al.*, 2011).

1.2.3. Industrial application of cellulases and hemicellulases

Glycoside hydrolase, particularly cellulases have numerous applications in the production of bioethanol, clarification of juices and wines, extraction of plant oils and production of coffee and laundry detergents (Table 1.1) (Aspeborg *et al.*, 2012; Bhat, 2000; Chandel *et al.*, 2012; Sunna and Antranikian, 1997). Cellulases and hemicellulases are necessary in the production of bio-ethanol in order to convert the lignocellulose to fermentable sugars under mild conditions (pH 4-6) (Dondelinger *et al.*, 2016; Harris *et al.*, 2014; Kuhad *et al.*, 2011). However, in the paper and pulp industry, the production and bleaching of pulp is conducted under alkali conditions and at elevated temperatures. As a result, thermostable, alkali-stable and cellulase-free hemicellulases are required for bio-pulping and bio-bleaching processes. (Kamali and Khodaparast, 2015; Walia *et al.*, 2017). Therefore, the type of glycoside hydrolase required is dependent on the type of industrial bio-process and its production parameters.

Table 1.1: A selection of industrial applications that utilise glycoside hydrolases

(adapted from (Collins *et al.*, 2005))

Industry	Application	Function	Glycoside hydrolase	Reference
Bioenergy	Bio-conversion of cellulose and hemicellulose	Enzymatic conversion of cellulose and hemicellulose to fermentable sugars for bioethanol production. Production of eco-friendly bio-fuels.	Cellulases Hemicellulases	(Ribeiro <i>et al.</i> , 2012)
Paper and pulp	Biobleaching Bio-pulping Bio-modification of fibres	Reduces the use of chlorine in the bleaching of pulp. Facilitates the production of eco-friendly pulp-reduces energy consumption. Improves production efficiency and strength of paper by improving fibrillation and drainage properties of pulp.	Hemicellulases	(Kumar <i>et al.</i> , 2009) (Suurnakki <i>et al.</i> , 1997)
Textiles	Retting of hemp	Enzymatic retting of textiles. Alternative to chemical retting	Cellulases Hemicellulases	(Zhao <i>et al.</i> , 2013)
Animal feed	Ruminant feeds Swine and poultry feed	Decreased non-starch polysaccharide content. Easily digested by animals.	Cellulase Xylanases Mannanases	(Costa <i>et al.</i> , 2014) (Lei <i>et al.</i> , 2016)
Food and beverage	Extraction of fruit and vegetable juices Beer brewing	Clarification of juices. Overall improvement in process by increasing extraction yield, process performance and product quality.	Cellulases Hemicellulases	(Sharma <i>et al.</i> , 2017)
Baking	Dough and bakery products	Enhances the elasticity and strength of baking dough. Allows for easy handling of dough, larger quantities of dough and improved texture in product.	Cellulase Hemicellulases	(Waters <i>et al.</i> , 2011)

1.3. Microbial glycoside hydrolases

1.3.1. Classification and nomenclature of glycoside hydrolases

The term Glycoside Hydrolase (GH) refers to a group of enzymes that catalyse the hydrolysis of glycosidic bonds (Cantarel *et al.*, 2009). The classification of glycoside hydrolases has evolved over the years as initial classification schemes were primarily based on substrate specificity (Henrissat and Davies, 2000). However, many glycoside hydrolases are poly-specific and display a range of activities with different lignocellulosic substrates. Divergent evolution has resulted in GHs of varying properties including thermophilic enzymes and enzymes with a broad range of hydrolytic activities (Henrissat, 1991). The current classification system is based on similarities in amino acid sequence, the catalytic mechanism and structural properties (Henrissat, 1991; Henrissat and Bairoch, 1993). Therefore, the classification of glycoside hydrolases is sequence related and not function related. As a result, a single GH family may consist of enzymes of high sequence homology but with different functions. Some GH families can be grouped into clans based on the fold of the protein (Henrissat and Bairoch, 1993). At present there are 14 clans identified on the Carbohydrate Active enZyme (CAZy) database (<http://www.cazy.org/>). Most clans consist of two or three GH families, with the exception of a clan GH-A which consists of 19 GH families.

1.3.2. Catalytic mechanism of glycoside hydrolases

Glycoside hydrolases are highly efficient enzymes with some having a catalytic enhancement greater than 10^{17} fold (Rye and Withers, 2000). The structural analysis of CAZyme enzyme-substrate complexes has contributed greatly to the understanding of the catalytic mechanism of glycoside hydrolases. Hydrolases belonging to the same GH family often utilize the same catalytic mechanism and catalytic residues. The hydrolysis of lignocellulosic substrates by GHs

occurs either by inversion or retention of the anomeric configuration (Collins *et al.*, 2005; Davies and Henrissat, 2002; Henrissat, 1991).

1.3.2.1. Retaining glycoside hydrolases

Retaining enzymes have two glutamate residues located approximately 5.5 Å apart which serve as the catalytic residues (Henrissat *et al.*, 1995). Using a double displacement mechanism, retaining enzymes are able to retain the anomeric configuration (Figure 1.4a) (Davies *et al.*, 1998). One carboxylate group acts as an acid catalyst and protonates the substrate. The carboxylate residue acts as a general base and deprotonates the water which then performs a nucleophilic attack on the anomeric carbon. This leads to the formation of an α -glycosyl-enzyme intermediate via an oxocarbenium ion-like transition state (Vasella *et al.*, 2002). Oxocarbenium ions are formed upon ionisation of glycosidic bond during enzymatic hydrolysis of the substrate. The positively charged carbocation is stabilised by resonance and performs a nucleophilic attack on the carboxylate residues.

1.3.2.2. Inverting glycoside hydrolases

Inverting glycoside hydrolases utilize a single glutamate and aspartate residue, located approximately 10 Å apart, as the catalytic residues (Behera *et al.*, 2017; Henrissat *et al.*, 1995). The distance between the catalytic residues of inverting hydrolases is greater than that of retaining hydrolases in order to accommodate a water molecule between the anomeric carbon and carboxylate residue (Figure 1.4b). Using a single displacement mechanism, inverting enzymes catalyse the hydrolysis of the substrate resulting in a product of inverted anomeric configuration. A carboxylate residue from the catalytic residue acts as a general acid and catalyses the departure of the leaving group. The second carboxylate group acts as a general

base and activates the water molecule which subsequently performs a nucleophilic attack on the anomeric carbon. This results in the cleavage of the glycosidic bond and inversion of the anomeric configuration.

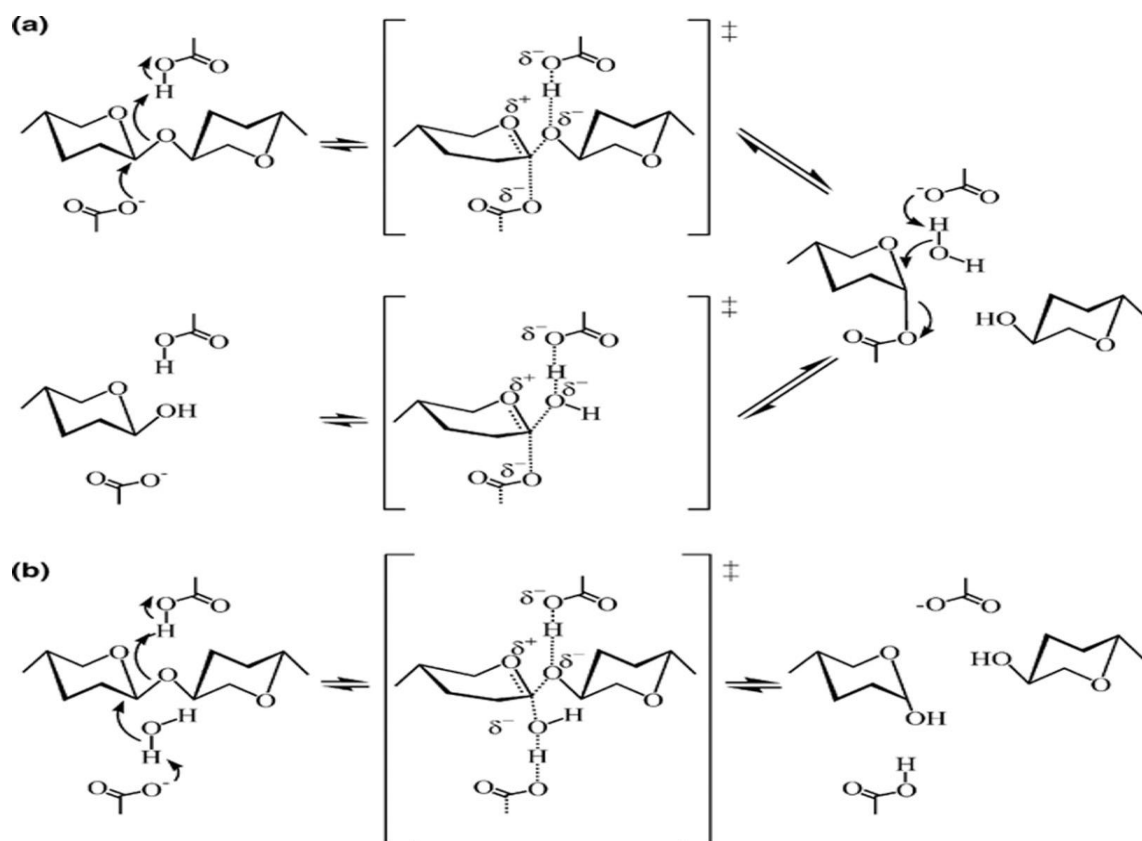


Figure 1.4: General catalytic mechanism of (a) retaining and (b) inverting glycoside hydrolases (Collins *et al.*, 2005).

1.3.3. Glycoside hydrolase family 5

Glycoside Hydrolase family 5 (GH5) is one of the largest GH families and includes of a wide range of hydrolytic activities such as xylanases, cellulases, mannanases, chitosanases and licheninases (Aspeborg *et al.*, 2012). Therefore, GH5 enzymes often have a broad range of specificities and can exist independently or as part of multi-modular proteins together with

other catalytic or non-catalytic domains. As one of the largest GH families, GH5 is, surprisingly, largely unexplored. Initial sequence-based classification resulted in several carbohydrate active enzymes being incorrectly classified due to the limited availability of sequence data (Naumoff, 2011). Analysis of three-dimensional structural data revealed inconsistencies in the organisation of secondary structural elements adjacent to the conserved, characteristic $(\beta/\alpha)_8$ fold of the catalytic domain (St John *et al.*, 2010). This discovery resulted in the re-classification of various GH5 enzymes to GH30. The GH5 family has since been further sub-divided into 56 sub-families (GH5-1 to GH5-56), although some sub-families have not been fully characterized and some enzyme activities remain unknown (Aspeborg *et al.*, 2012). Similarly, GH30 enzymes also exhibit broad substrate specificities and are further divided into sub-families (GH30-1 to GH30-9) with both families belonging to clan GH-A.

All GH5 enzymes are retaining GHs with two glutamates as the catalytic residues. Glycoside hydrolase family 5 enzymes have a novel and unique catalytic triad that is conserved across GH5 three-dimensional structures (Santos *et al.*, 2012; Shaw *et al.*, 2002). GH5 enzymes are the first group of $(\beta/\alpha)_8$ proteins that are known to have a catalytic triad. This catalytic triad is assumed to be of functional significance in the GH5 family. Although the exact function of this triad is unknown, it has been suggested that it plays a role in attenuation of the pKa of the proton donor glutamic acid.

1.3.4. Glycoside hydrolase family 6

Glycoside Hydrolase family 6 (GH6) enzymes are strictly cellulose degrading enzymes, often found in bacteria and fungi as single modules or as part of a multi-modular enzyme. The GH6 family is composed of endo- β -1,4-glucanases (EC 3.2.1.4) and cellobiohydrolases (EC 3.2.1.91) and, unlike GH5, the GH6 family is not divided into sub-families

(<http://www.cazy.org/>). Exoglucanases belonging to this family processively cleave cellulose from the non-reducing end to release cellobiose units (Poidevin *et al.*, 2013). At the catalytic site two aspartate residues serve as the catalytic residues involved in hydrolysis. At present, only 65 enzymes belonging to this family have been fully characterised and only 13 crystal structures have been determined. Although some GH6 enzymes may have a conserved catalytic fold, unlike the GH5 family, the GH6 family does not belong to any clan due to a lack of sufficient three-dimensional structural data. This lack of structural data for GH6 enzymes highlights the need for further studies in order to fully understand the structure and function of GH6 enzymes.

1.3.5. Carbohydrate-binding modules

Carbohydrate-binding modules (CBM) are non-catalytic modules found in many different glycoside hydrolases (Shoseyov *et al.*, 2006; Varnai *et al.*, 2013). The main function of the CBM is to co-ordinate polysaccharide recognition, facilitate substrate binding and to contribute to the structure and function of the catalytic site. These autonomously folding and functioning proteins enhance enzyme activity, although their mechanism of action has not been fully elucidated. However, two mechanisms have been proposed. The first mechanism involves the CBM in physical disruption of the plant cell wall resulting in the exposure of single chain glycans that can be easily degraded by enzyme hydrolysis (Gilbert *et al.*, 2013). However, limited evidence is available to support this hypothesis. The second mechanism suggests that the CBM enhances enzyme activity by increasing or prolonging interactions between the soluble enzyme and insoluble polysaccharide (Hashimoto, 2006; Subramaniyan and Prema, 2002). As non-catalytic modules, CBMs are not essential for enzyme hydrolysis and do not affect the synergism between cellulolytic enzymes.

Based on alignments of the primary structure, CBMs are classified into 83 families (<http://www.cazy.org/>). Selected CBMs of bacterial and fungal origin display dual specificity for cellulose and xylan. Folding of the CBM is conserved across different enzymes of varying specificity and taxonomic groups. In glycoside hydrolases, the CBMs are located either on the N- or C- termini and can be found between two catalytic modules, as a single unit or arranged in tandem (Guillén *et al.*, 2010). The catalytic domain must initially be positioned in close proximity to the substrate through the CBM to allow the catalytic domain to hydrolyse the substrate at the active site (Reyes-Ortiz *et al.*, 2013). The CBM can subsequently be relocated to new regions on the substrate to allow for continuous hydrolysis. Aromatic residues (tyrosine and tryptophan) that are essential in the binding of the CBM form stacking interactions with the sugar residues (Boraston *et al.*, 2004). The strong van der Waals forces formed stabilise the protein structure (Hashimoto, 2006). Studies suggest that the orientations of these aromatic residues are implicated in the varying substrate specificities of CBMs (Guillén *et al.*, 2010; Shoseyov *et al.*, 2006). Alternatively, the side chain of other polar amino acid residues may form hydrogen bonds with sugar residues and stabilise the structure. The diversity of CBMs mimics that of plant polysaccharides in nature. The heterogeneous structure of cell wall polysaccharides and the accessibility of ligands, which depends on the cell type, tissue and botanical source, has resulted in the multiplicity of CBMs of varying substrate specificities (Boraston *et al.*, 2004).

1.4. Methods in protein structure determination

The three-dimensional structure of proteins can be determined by X-ray crystallography, nuclear magnetic resonance (NMR) and electron microscopy (Stuart and Abrescia, 2013). X-ray crystallography has proven to be the most popular technique as it is suitable for a broad range of proteins and provides higher resolution protein structures when compared to NMR

(Krishnan and Rupp, 2012). X-ray diffraction is considered a form of extremely high resolution microscopy that is used to determine the three-dimensional structure of macromolecules including glycoside hydrolases (Rhodes, 2010). Since the elucidation of the first crystal structure of a *Trichoderma reesei* cellulase in 1990 (Rouvinen *et al.*, 1990), numerous other glycoside hydrolase structures have been determined (Maehara *et al.*, 2014; Qin *et al.*, 2015; Sato *et al.*, 2017; You *et al.*, 2018).

The success of any X-ray diffraction experiment is highly dependent of the quality of the protein crystals. Several criteria need to be met for the protein in order to enable successful protein crystallisation. A pure, active, stable, soluble and monodisperse protein (Asherie, 2004; Drenth, 2007) is considered a pre-requisite to any successful crystallisation experiment. However, there are exceptions to these rules as some proteins have proven to crystallise under atypical conditions. Protein crystallisation can occur under a range of conditions that vary in pH, type of buffer, precipitant, salt and temperature (Asherie, 2004). The most common methods for protein crystallisation are the sitting-drop or hanging-drop methods. The hanging drop vapour diffusion method is the most inexpensive method using for growing crystals. High quality macromolecular crystals may grow overnight or even after weeks.

1.4.1. Data collection by X-ray diffraction

During X-ray diffraction, a rotating anode (copper or molybdenum) generates an X-ray beam of a distinct wavelength that passes through the protein crystal mounted on the goniometer. The X-ray beams are scattered by the electrons present in the crystallised protein molecule resulting in constructive and destructive interference in the diffraction pattern. The diffraction pattern appears on the detector as discrete spots referred to as reflections. Each reflection contains information on the atoms in the protein and each atom contributes to the intensity of

each reflection (Davis *et al.*, 2003). The direction in which the X-rays are scattered is dependent on the geometry of the crystal unit cell and wavelength of the X-rays (Chatterjee, 2010). X-ray scattering is often favoured in selected directions and are minimal in other directions. Therefore, the constellation of the atoms determines the effectiveness of interference and the intensity of the diffracted X-rays in each direction (Chatterjee, 2010; Wlodawer *et al.*, 2013). The diffracted x-ray beams are collected with an area detector and produce a diffraction pattern. However, several diffractions patterns need to be collected in order to cover the required crystal orientations.

1.4.2. Phasing techniques

X-rays have both amplitudes and phases and during X-ray diffraction the phase information, which is a critical for protein structure determination, is lost (Lattman and Loll, 2008). The loss of this information is referred to as the crystallographic phase problem. Therefore, protein structure determination is essentially solving the phases. Several phasing techniques can be employed to solve the ‘phase problem’ and produce an electron density map: these include molecular replacement, isomorphous replacement and multi-wavelength anomalous dispersion (Wlodawer *et al.*, 2013).

1.4.2.1. Molecular replacement

Molecular replacement (MR) makes use of a known molecular model to solve an unknown crystal structure of a related molecule (Rossman and Blow, 1962; Wlodawer *et al.*, 2013). MR enables the solution of the crystallographic phase problem by providing initial estimates of phases of the new structure from a previously known structure. Approximately 70% of

deposited protein structures have been solved by molecular replacement (Abergel, 2013). Molecular replacement determines the highest correlation between experimental data and those calculated from the model. Therefore, the success of molecular replacement is dependent on the co-ordinates of a similar protein. Using these co-ordinates the crystal structure of a protein can be solved by rotating and translating these co-ordinates into the new crystal system until the best match to the experimental data is obtained. Molecular replacement places the model in the correct orientation and position within the unknown unit cell (Evans and McCoy, 2008). The orientation and positioning of a molecule requires three rotation angles and three translation angles, respectively. The intramolecular vectors depend on the orientation of the molecule. However, intermolecular vectors depend on both the orientation and position of the molecule. Therefore, the intramolecular and intermolecular vectors can be exploited in the rotation and translation function, respectively. Fourier transform can subsequently be used to calculate the amplitudes and phases from the molecular replacement solution and, in combination with the experimental data, an electron density map can be produced (Wlodawer *et al.*, 2008). Therefore, MR produces an electron density map and its corresponding atomic co-ordinates and together with the protein backbone a model can be built and refined.

1.4.2.1.1. Factors affecting molecular replacement

A number of factors need to be taken into account during MR to ensure accurate structure determination. These include the nature of the protein crystal, the quality of the X-ray data, the MR search probe, the rotation function integration radius and temperature factor (Evans and McCoy, 2008). Molecular replacement is most effective when a suitable structural homolog of the protein is available (Scapin, 2013). The sequence identity between a homolog and protein of interest determines whether two proteins are structurally similar. Therefore, MR is likely to be successful if the model is fairly complete and shares at least 25% sequence identity with the

unknown structure (Taylor, 2003). As a consequence, crystal structure determination becomes increasingly difficult as the model becomes less complete or shares less sequence identity. High quality X-ray data, with a completeness as close to 100% as possible, reduce any difficulties with molecular replacement. The resolution, the accuracy and multiplicity of the X-ray data also need to be taken into account. Therefore, the quality of the protein crystal and corresponding X-ray diffraction data are critical in any X-ray diffraction experiment (McPherson, 2004). In addition to this, the ideal search model should have a low r.m.s deviation ($< 2 \text{ \AA}$), a high completeness and a high amino acid sequence identity (Evans and McCoy, 2008; Taylor, 2010). The side chains of some residues in the model can be trimmed to alanine or glycine residues particularly in regions where large differences with the unknown structure can be expected.

Molecular replacement as a phasing technique often favours protein crystals with a single protein molecule per asymmetric unit. Protein crystals with greater than one protein molecule per asymmetric unit often have a reduced signal-to-noise ratio. However, several protein structures have been determined by MR where multiple copies of the protein molecule were present in the asymmetric unit. Furthermore, the rotation function is calculated from intra-molecular vectors and not on the intermolecular vectors (Rossman and Blow, 1962). Therefore, the choice of the integration radius is dependent on the conformation of the molecules. The signal-to-noise ratio corresponding to the correct molecular replacement solution can be increased by examining the deviations from these starting values in spherical and elongated molecules.

1.4.3. Isomorphous replacement

In the absence of a starting model, single isomorphous replacement (SIR) or multiple isomorphous replacement (MIR) can be used. This approach makes use of heavy atoms that can be introduced into specific sites within the unit cell without disrupting the crystal lattice (Perutz, 1956; Taylor, 2010). Considering that heavy atoms are more electron dense it allows for measurable differences in the intensities of the spots the diffraction pattern (Lu and Sun, 2014). As a result, the differences in each reflection are measured in order to estimate the phase angle using vector summation methods. Unlike MR, SIR/MIR produces an electron density map and the protein backbone needs to be fitted into the model. As a result, the structural detail will depend on the quality and resolution of the phases. During SIR/MIR, regions of high flexibility are typically not visible due to static or dynamic disorder (Egli, 2016). Static disorder is a consequence of the variation between the different molecules within the protein crystal. Dynamic disorder occurs when a region of the structure is mobile within the crystal but this can be overcome by cryogenic data collection. In contrast to MR, SIR/MIR data can be collected at a single wavelength using several protein crystals (Blessing, 2014). However, data collection requires a tuneable X-ray source which is typically available at a synchrotron.

1.4.4. Multi-wavelength anomalous dispersion

Multi-wavelength anomalous dispersion (MAD) makes use of anomalous scattering data of certain atoms at or near the X-ray adsorption edges to obtain the necessary phase information (Taylor, 2003). The success of MAD is dependent on the anomalous diffraction of one or more anomalously diffracting atoms in the crystal (Su *et al.*, 2015). For a single protein crystal, more than three data sets are recorded at different wavelengths around the X-ray adsorption edge of the anomalous scattering atom. For instance, selenium, an anomalous scatterer, can be incorporated into proteins that have been over-expressed in *Escherichia coli* strains, grown in

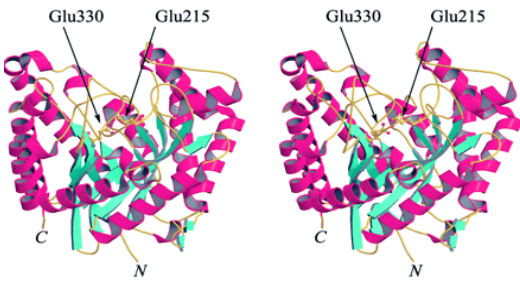
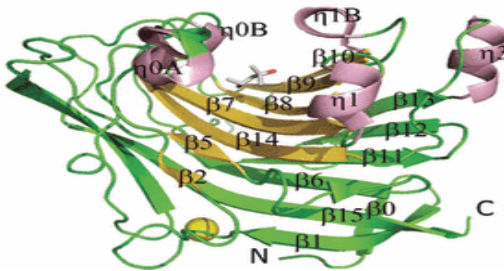
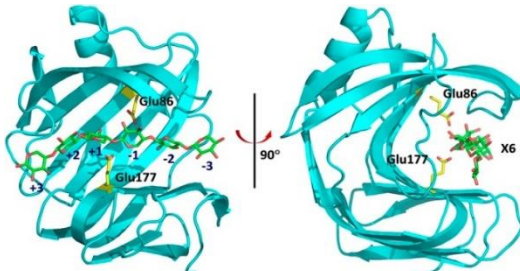
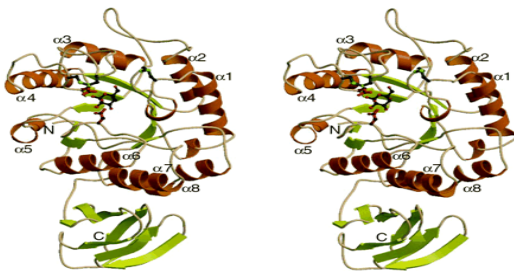
minimal media supplemented with seleno-L-methionine instead of methionine (Boggon and Shapiro, 2000). One major advantage of MAD is that there are no non-isomorphous atoms, which aids in data collection. The phase information obtained often produces a higher quality electron density map than MR.

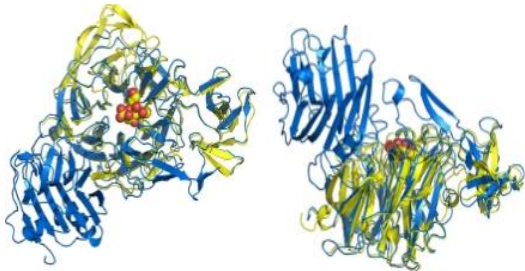

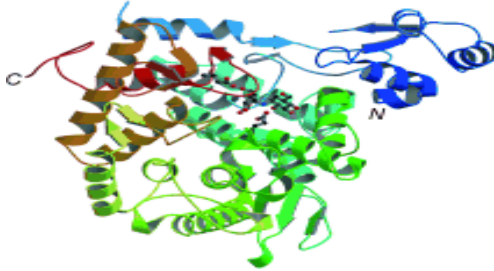
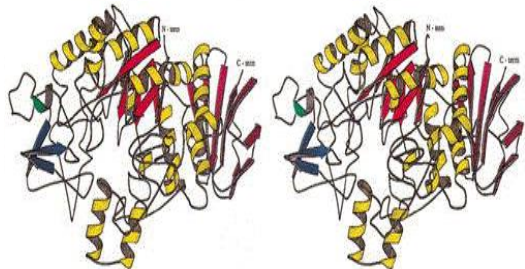
1.4.5. Crystal structure of glycoside hydrolases

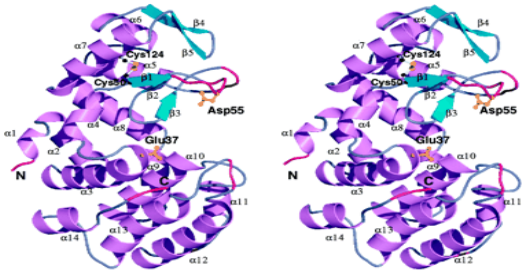
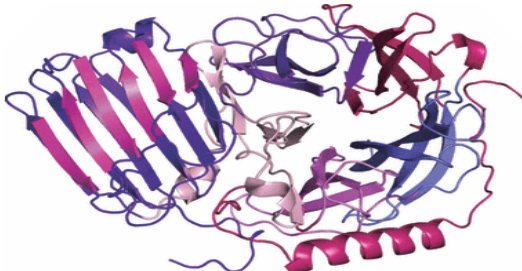
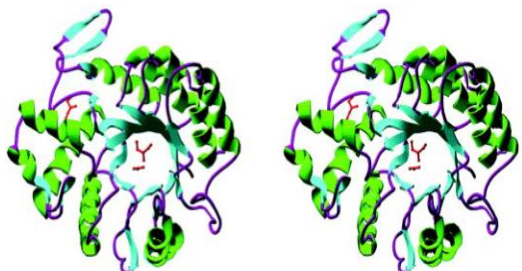
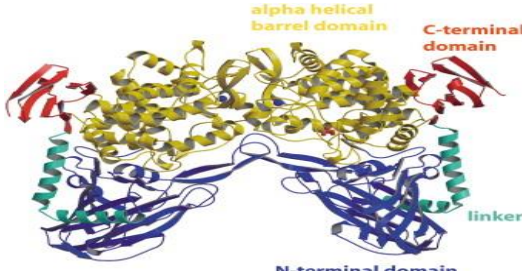
In nature, glycoside hydrolases occur in various structural conformations and molecular arrangements. There have been several reports of multi-modular microbial glycoside hydrolases which have one or more carbohydrate binding modules and/or catalytic modules belonging to different families (Davies *et al.*, 1998; Lombard *et al.*, 2014; Mewis *et al.*, 2016). The multiplicity of microbial glycoside hydrolases highlights the many variations in structural conformation across the different families. The catalytic fold of glycoside hydrolases is often highly conserved and many of these families have been grouped into different clans (GH-A to GH-N), with many families sharing a similar catalytic fold (Henrissat and Bairoch, 1996) as indicated in Table 2 below. The catalytic fold of proteins is dictated by the spatial arrangement of the secondary structural elements (Schaeffer and Daggett, 2011).

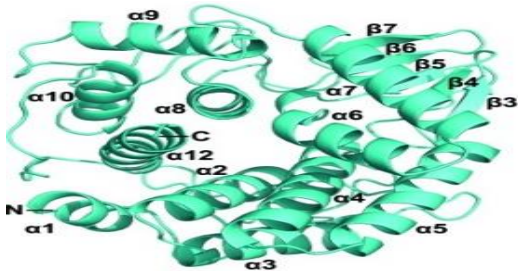
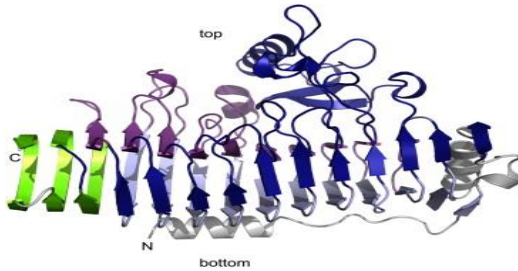
Several crystal structures of glycoside hydrolases across the various GH-clans, have been elucidated by X-ray diffraction. Dias *et al* (2004) elucidated the crystal structure of a GH 5 exo- β -mannosidase from *Cellvibrio mixtus* utilizing the MAD method with selenomethionine. The 1.5 Å crystal structure displayed the $(\beta/\alpha)_8$ fold typical of the GH-A clan (Table 1.2). Further analysis indicated that the β -mannosidase utilizes two glutamate residues in a retaining hydrolytic mechanism. Additional key amino acids residues were identified as highly conserved in GH5, including Arg⁸⁰, Asn²¹⁴, Glu²¹⁵, His²⁸³, Glu³³⁰ and Trp³⁷⁶.

Table 1.2: Summary of the types of the prominent catalytic folds across the GH clans.

GH-Clan	Protein fold	GH families	Representative structures
GH-A	$(\beta/\alpha)_8$	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113, 128	 <p>(Dias <i>et al.</i>, 2004)</p>
GH-B	β -jelly roll	7, 16	 <p>(Ilari <i>et al.</i>, 2009)</p>
GH-C	β -jelly roll	11, 12	 <p>(Wan <i>et al.</i>, 2015)</p>
GH-D	$(\beta/\alpha)_8$	27, 31, 36	 <p>(Fujimoto <i>et al.</i>, 2003)</p>

GH-E	6-fold β -propeller	33, 34, 83, 93	 <p>(Newstead <i>et al.</i>, 2008)</p>
GH-F	5-fold β -propeller	43, 62	 <p>(de Sanctis <i>et al.</i>, 2010)</p>
GH-G	$(\alpha/\alpha)_6$	37, 63, 100, 125	 <p>(Gibson <i>et al.</i>, 2007)</p>
GH-H	$(\beta/\alpha)_8$	13, 70, 77	 <p>(Watanabe <i>et al.</i>, 1997)</p>

GH-I	$\alpha+\beta$	24, 46, 80	 <p>(Saito <i>et al.</i>, 1999)</p>
GH-J	5-fold β -propeller	32, 68	 <p>(Bujacz <i>et al.</i>, 2011)</p>
GH-K	$(\beta/\alpha)_8$	18, 20, 85	 <p>(Ramasubbu <i>et al.</i>, 2005)</p>
GH-L	$(\alpha/\alpha)_6$	15, 65	 <p>(Egloff <i>et al.</i>, 2001)</p>

GH-M	$(\alpha/\alpha)_6$	8, 48	 <p>(Attigani <i>et al.</i>, 2016)</p>
GH-N	β -helix	28, 49	 <p>(Pijning <i>et al.</i>, 2009)</p>

Using the MAD method and selenomethionine derivatives, the crystal structure of a GH43 (clan GH-F) endo-1,5-L-arabinanase from *Bacillus subtilis* was determined at a resolution of 1.75 Å (de Sanctis *et al.*, 2010). The arabinanase displayed a 5-fold β -propeller catalytic fold composed of five β -sheets (blades), with each blade composed of four anti-parallel β -strands (Table 1.2). All five blades were arranged radially around a pseudo 5-fold axis. Analysis of the crystal structure also indicated that the catalytic domain was composed of 20 β -strands and three α -helices.

The crystal structure of a GH8 endoglucanase (EC 3.2.1.4) from *Klebsiella pneumoniae* was elucidated by MR (Attigani *et al.*, 2016) and displayed a $(\alpha/\alpha)_6$ barrel fold (Table 1.2) at a resolution of 1.76 Å. However, the crystal structure was composed of 11 α -helices instead of the typical 12 helices with the twelfth α -helix being replaced with a flexible loop. Comparison of this endoglucanase with a very well characterized GH8 (clan GH-M) endoglucanase from *Clostridium thermocellum* (Alzari *et al.*, 1996) revealed distinct structural differences between

the two catalytic sites. The endoglucanase from *K. pneumoniae* had a more open substrate binding cleft compared to that of *C. thermocellum*. As a result the endoglucanase from *K. pneumoniae* was unable to binding and cleave cellobiose and its enzyme activity was restricted to cello-oligosaccharides with greater than five sugar residues.

1.5. Rationale

Glycoside hydrolases are essential in the bio-conversion of plant biomass to various bio-products. A *Paenibacillus mucilaginosus* glycoside hydrolase (*PmGH*) presumed to be thermophilic in nature was previously isolated from a compost metagenome (Tshukudu, 2012). Therefore, this study cloned, expressed and characterised the recombinant glycoside hydrolase, *PmGH* in order to evaluate its functional properties. Furthermore, we investigated the structural properties of *PmGH* through extensive optimisation of protein crystallisation experiments providing the first ever protein crystal of the multi-modular and multi-functional, *PmGH*. X-ray data collection was optimised using local and synchrotron radiation sources.

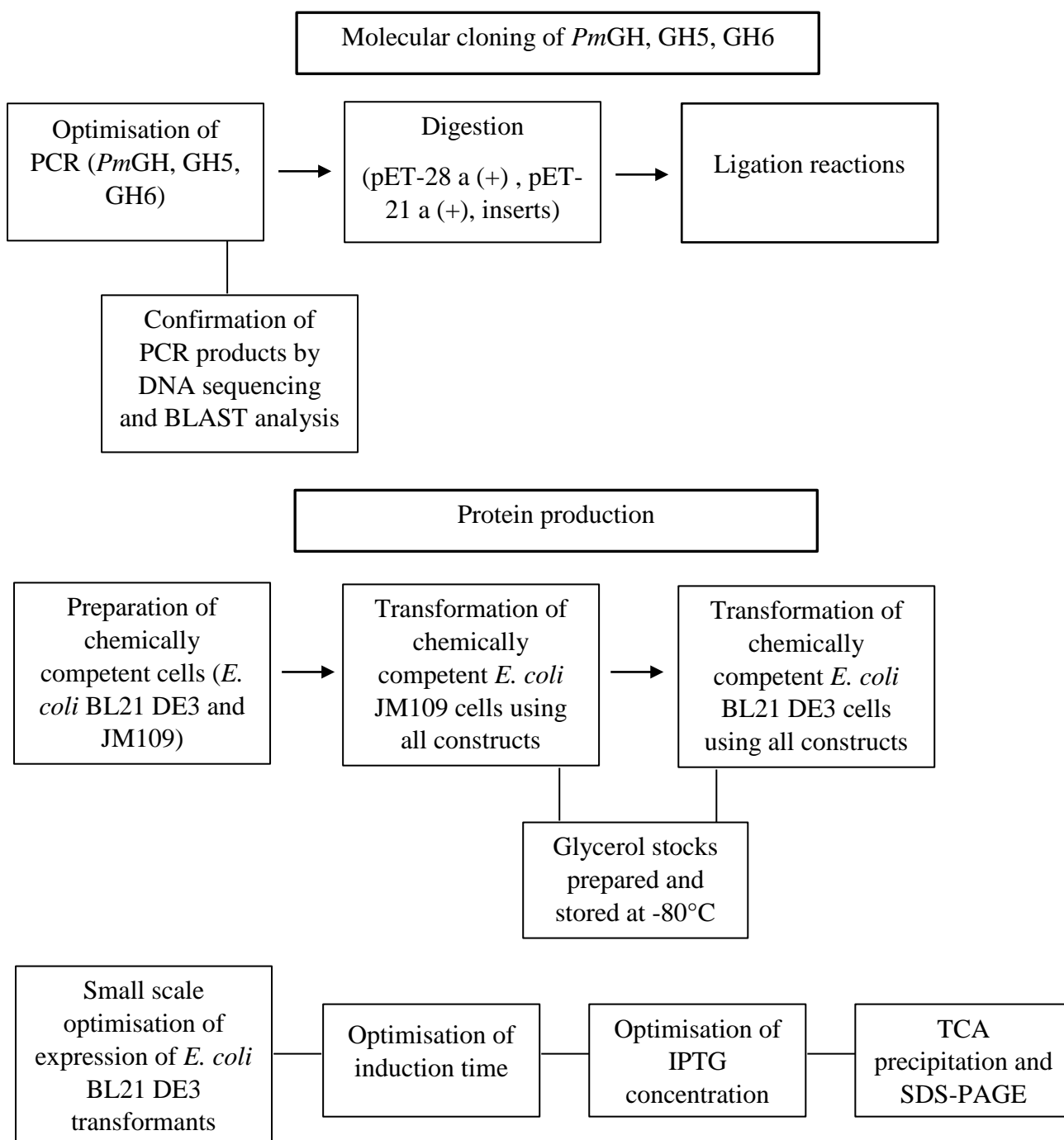
1.6. Aims and objectives

- To determine the functional properties of the novel glycoside hydrolase, *PmGH*, by determining the types of hydrolytic activities, the functional thermostability and pH-stability of this recombinant enzyme.
- To determine the structural properties of *PmGH* by establishing and optimising suitable protein crystallisation conditions. Furthermore, to optimise X-ray data collection in order to elucidate the *PmGH* crystal structure.

Chapter 2

Materials and Methods

In determining the functional and structural properties of the novel, multi-modular glycoside, *PmGH*, the strategy shown in Figure 2.1 was employed. The overall steps included molecular cloning, protein production, protein purification, functional and structural analysis. All protocols and experiments were optimised in order to investigate the properties of *PmGH*.



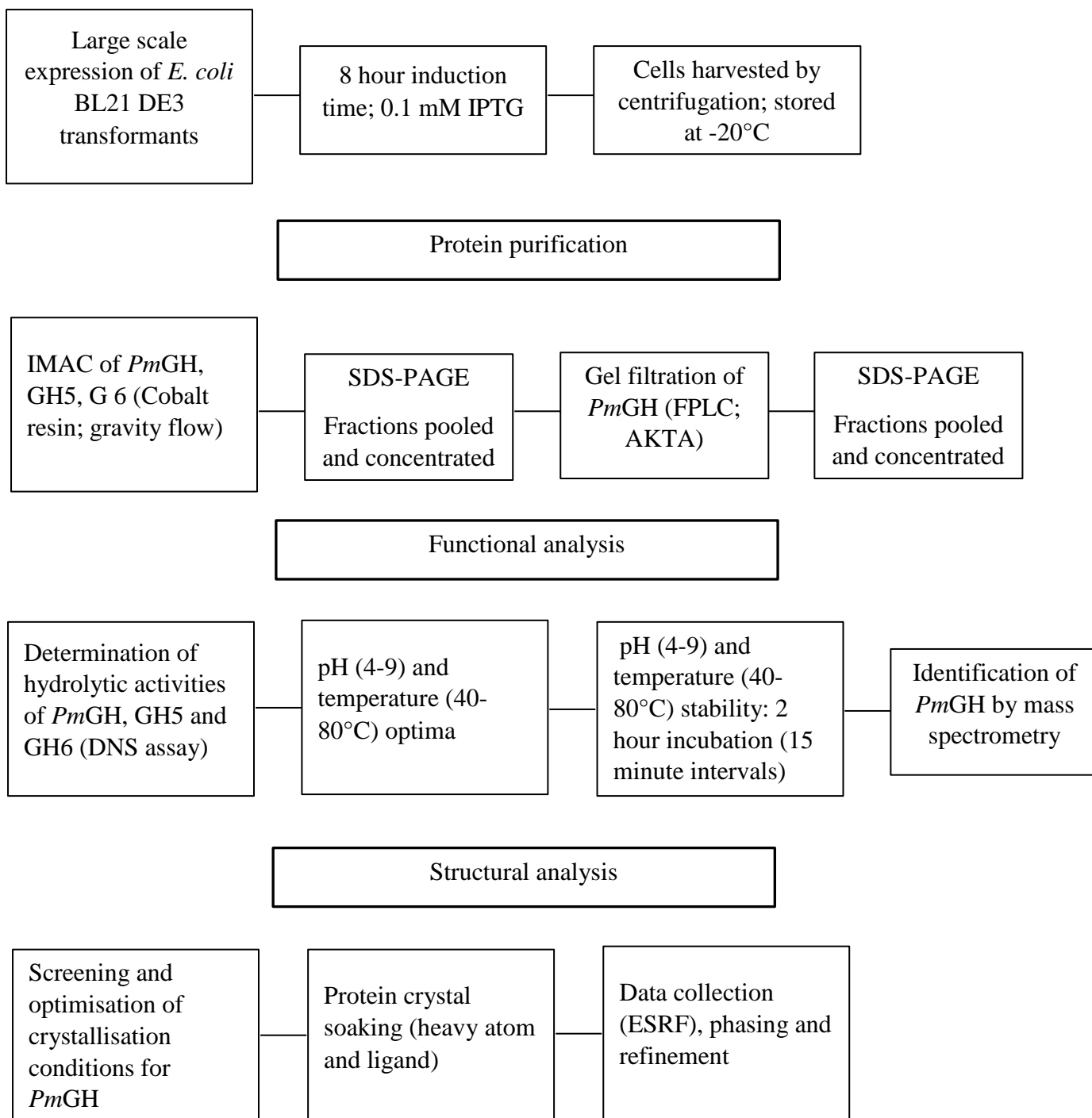


Figure 2.1: Schematic overview of the methodical approach employed in determining the functional and structural properties of the novel glycoside hydrolase, *PmGH*.

2.1. Cultivation and cloning of the full length glycoside hydrolase and corresponding domains

A cellulase positive fosmid clone was previously identified from a high temperature compost, metagenomic library (Tshukudu, 2012). Bioinformatics analysis suggested that one of the open reading frames within the 24 447 bp contig encoded a putative cellulase, *PmGH*. The *PmGH* gene was obtained from the Institute for Microbial Biotechnology and Metagenomics (University of Western Cape, South Africa) and *Escherichia coli* Rosetta (pLysS) harbouring the pET-21 a (+)-GH recombinant construct was transported to the Centre for Microbial Ecology and Genomics (University of Pretoria, South Africa) for subsequent analysis. The recombinant *E. coli* [pET-21 a (+)] strain was inoculated into five millilitre of Luria-Bertani (LB) and incubated at 37°C overnight. Long term glycerol stocks were prepared and stored at -70°C. Plasmid DNA was extracted from the bacterial culture using a Plasmid Mini-Prep kit (Qiagen, USA) and electrophoresed on a 1% (w/v) agarose gel. All plasmid DNA was purified using a NucleoSpin gel and PCR clean-up kit (Macherey-Nagel, Germany). The purified plasmid DNA was electrophoresed on a 1% (w/v) agarose gel and quantified using a NanoDrop 2000c spectrophotometer (ThermoScientific, USA).

2.1.1. Polymerase chain reaction

All polymerase chain reaction (PCR) reactions were carried using the iProof high fidelity PCR kit (Bio-Rad, USA). All PCR reactions were performed as per manufacturers' instructions using the recommended reaction mixtures and PCR run conditions (Table 2.1). As recommended, all PCR reaction components were used at a final concentration of 1x iProof HF buffer, 200 µM dNTP, 0.5 µM forward and reverse primers and 0.02 U/µl iProof DNA polymerase per 20 µl PCR reaction. The optimal Mg²⁺ concentration recommended was 0.5 to

1 mM per PCR reaction and was optimised by increase or decreasing the Mg²⁺ concentration in 0.2 mM increments. The Mg²⁺ concentrations were optimised to a final concentration of 1 mM, 0.4 mM and 0.8 mM for the *PmGH*, GH5 and GH6 PCR reactions, respectively.

Table 2.1: PCR primers and thermal cycling parameters for all amplicons.

Sample	*Forward primer	*Reverse primer	‡PCR parameters	Expected product size (base pairs)	Restriction sites
<i>PmGH</i>	AAAACCATGG AGCCGCATGTG GA	TTTTAAGCTTT TACTCGAGTG CGGCCGCAA	98°C (30s); 98°C (10s); 63°C (90s); 72°C (90s); 72°C (10 min); 4°C (∞)	~3000 bp	NcoI HindIII
GH5	GCAAAAGCTTA ATGGTCGACCA AAAT GG	TTTTGGATCC TACTTGTCTT CAACGAC	98°C (30s); 98°C (10s); 63°C (30s); 72°C (90s); 72°C (10 min); 4°C (∞)	~765 bp	HindIII BamHI
GH6	GCAAAAGCTTA ATGGTTTGTAG GAGC GA	TTTTTGTCTGA CTTAGTCAAA CCGCTTGC	98°C (30s); 98°C (10s); 63°C (30s); 72°C (60s); 72°C (10 min); 4°C (∞)	~1159 bp	HindIII SalI

*Restriction sites indicated in bold. ‡35 cycles of denaturation, annealing and extension.

The recommended PCR thermal cycling parameters were used and further optimised by varying the extension time (72°C) for each PCR reaction as recommended (15-30 s/kb). Polymerase chain reaction amplicon sizes were confirmed by agarose gel electrophoresis and purified as previously described in Section 2.1. The purified PCR amplicons were quantified spectrophotometrically using a NanoDrop 2000c spectrophotometer (ThermoScientific, USA) and sequenced. All DNA sequencing was performed at the Central Analytical Facilities DNA Sequencing Unit (Stellenbosch University, South Africa). A Basic Local Alignment Search Tool (BLAST) analysis of the amplified regions was conducted to confirm the identity of the full length glycoside hydrolase gene. The multi-modular nature of *PmGH* was confirmed by a National Centre for Biotechnology Information (NCBI) conserved domain search and the composite domains were identified.

2.1.2. Restriction digestion and ligation reactions

The purified PCR amplicons described in Section 2.1.1 above were double digested with the corresponding restriction enzymes shown in Table 2.1 above, and subsequently analysed by agarose gel (1%; w/v) electrophoresis. The plasmid vectors, pET-21 a (+) and pET-28 a (+), were also double digested with restriction enzymes that correspond to the respective PCR amplicon to create compatible sticky ends (Table 2.1). All restriction analysis was performed using FastDigest restriction enzymes (ThermoScientific, USA). As per manufacturer's instructions, 2 µl of Fast digest buffer (1x), 1 µg of DNA (plasmid or insert) and 0.5 µl of each restriction enzyme was added per restriction digestion reaction. Each reaction was brought to final reaction volume of 20 µl by the addition of nuclease free water. All digestion reactions were incubated at 37°C (5 minutes for PCR products and 10 minutes for plasmid vectors) and inactivated by incubation at 80°C for 20 minutes. The digested amplicons and plasmid vectors were subsequently purified using a NucleoSpin gel and PCR clean-up kit (Macherey-Nagel,

Germany). Sticky end ligation reactions were setup for all PCR inserts and the plasmid vectors with the corresponding sticky ends. All ligation reactions were performed using a rapid DNA ligation kit (ThermoScientific, USA) using T4 DNA ligase at 22°C. A plasmid vector: PCR insert ratio of 3:1 was used in all ligation reactions. All ligation reaction mixtures contained 4 µl of ligation buffer, 1 µl of T4 DNA ligase and brought to a final volume of 20 µl by the addition of PCR grade water. Ligation of the PCR inserts and plasmid vectors were confirmed by agarose gel electrophoresis. Positive constructs (Table 2.2) were purified using a NucleoSpin gel and PCR clean-up kit (Macherey-Nagel, Germany) and stored at -20°C for further use.

Table 2.2: List of plasmid-vector constructs generated and expected sizes.

Construct	Expected construct size	Construct	Expected construct size
<i>PmGH-pET21-a (+)</i>	~8443 bp	<i>PmGH-pET28-a (+)</i>	~8369 bp
<i>GH5-pET21-a (+)</i>	~6208 bp	<i>GH5-pET28-a (+)</i>	~6134 bp
<i>GH6-pET21-a (+)</i>	~6602 bp	<i>GH6-pET28-a (+)</i>	~6528 bp

2.1.3. Preparation of chemically competent *E.coli* cells

Chemically competent cells of *Escherichia coli* (*E. coli*) JM 109 and BL21 DE3 were prepared as described by Sambrook and Russel (2006). *E.coli* cells (JM 109 and BL21 DE3) were streaked on LB agar plates and incubated overnight at 37°C. A single colony was placed in 10 ml LB media and grown overnight at 37°C. Five millilitres of overnight culture was transferred into 500 ml LB media in a 2 L Erlenmeyer flask. The flasks were incubated at 37°C until OD₆₀₀ of 0.4. The cells were subsequently transferred to two 250 ml centrifuge bottles and placed on

ice for 20 minutes. The cells were centrifuged for 10 minutes at 4°C at 3 000 x g (JA 25.50 fixed angle rotor; Beckman Coulter; USA). The supernatant was discarded and the pellet was resuspended on ice in 30 ml cold 0.1 M CaCl₂. The cell suspension was subsequently transferred into 50 ml propylene falcon tubes (Merck, USA) and incubated on ice for 30 minutes. Upon incubation the cells were centrifuged for 10 minutes at 4°C at 3 000 x g, the supernatant was discarded and the cells were resuspended in 8 ml cold 0.1 M CaCl₂ containing 15% (v/v) glycerol. The cell suspensions were aliquoted (100 µl) into 1.5 ml Eppendorf tubes and subsequently snap frozen in liquid nitrogen. All chemically competent *E. coli* cells were stored at -80°C for further use.

2.1.4. Transformation by heat shock using chemically competent cells

All constructs listed in Table 2.2 above were transformed by heat shock into *E. coli* (JM109 and BL21 DE3). For all transformation experiments, frozen chemically competent cells were removed from -80°C and thawed on ice. To an Eppendorf tube containing 25 µl of chemically competent cells, 1-50 ng of DNA was added and mixed by flicking the tube several times. The tubes were placed on ice for 30 minutes and the cells were subsequently heat-shock for 45-50 seconds at 42°C and placed on ice for 2 minutes. To each tube, 900-950 µl of Super Optimal broth with Catabolite repression (SOC medium) was added and incubated for 30-60 minutes at 37°C. The cells were subsequently plated on LB with the appropriate antibiotic (ampicillin or kanamycin). The plates were allowed to dry and were incubated at 37°C overnight. Ampicillin (100 µg.ml⁻¹) and kanamycin (30 µg.ml⁻¹) were used to select positive transformants containing the pET21-a (+) gene constructs and pET28-a (+) gene constructs, respectively. Glycerol stocks of all positive transformants were stored at -80°C for further use in protein expression trials.

2.2. Optimisation of protein expression

2.2.1. Small scale protein expression and analysis of *PmGH*

To further optimise protein expression, pre-cultures of *E.coli* strains containing the prepared constructs were prepared by inoculating three millilitres LB containing 100 $\mu\text{g}\cdot\text{ml}^{-1}$ ampicillin for *E.coli* (pET21-a (+) gene constructs) or 30 $\mu\text{g}\cdot\text{ml}^{-1}$ kanamycin for *E.coli* (pET28-a (+) gene constructs). All tubes were incubated at 37°C overnight. One millilitre of the starter culture was inoculated into 10 ml of LB containing the suitable antibiotic and incubated at 37°C until an OD₆₀₀ of 1 was reached. Optimal induction temperature was tested at 20°C and 24°C. Optimal induction time was determined by withdrawing a one millilitre volume of sample at different time intervals ($T_0 = 0$ hours; $T_1 = 2$ hours; $T_2 = 4$ hours; $T_3 = 6$ hours; $T_4 = 24$ hours) at 20°C and 24°C. Protein expression was induced by the addition of 0.1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) (Melford, United Kingdom). Once the optimum induction time was established, a range of IPTG concentrations (0-0.1 mM) were tested to establish the ideal conditions for *PmGH* expression.

All samples were precipitated by the trichloroacetic acid (TCA) precipitation method (Koontz, 2014). One hundred microliters of ice cold TCA was added to one millilitre of sample and vortexed for 30 seconds. The protein was allowed to precipitate on ice for 30 minutes and subsequently centrifuged at $10\,000 \times g$ for 15 minutes at 4°C. The supernatant (insoluble fraction) was removed and stored for further analysis. The pellet (soluble fraction) was washed twice by re-suspending in 500 μl of ice cold acetone and centrifuged at $10\,000 \times g$ for 5 minutes at 4°C. The pellet was air dried and re-suspended in 50 μl SDS-PAGE gel loading buffer. The soluble and insoluble fractions were analysed by SDS-PAGE (10%; w/v) at 90 V (Laemmli, 1970).

2.2.2. Large scale protein expression

Large scale expression of *PmGH* and corresponding catalytic domains (GH5 and GH6) were performed under optimal conditions in two litre Erlenmeyer flasks. Pre-cultures (10 ml) in LB containing the suitable antibiotic were prepared and incubated overnight at 37°C. To one litre of LB containing antibiotic, 10 ml of pre-culture was added and incubated at 37°C to an OD₆₀₀ of 1. Protein expression was induced by the addition of 0.1 mM IPTG (Melford, United Kingdom). After an 8 hour induction time, the cells were harvested by centrifugation at 5 000 × g (JA 25.50 fixed angle rotor; Beckman Coulter; USA) for 25 minutes at 4°C. The supernatant was discarded and the cell pellet was re-suspended in lysis buffer (50 mM phosphate buffer; pH 7, 300 mM NaCl) and stored at -20°C for further analysis.

2.3. Protein purification

The cell pellets were thawed and lysed by sonication for 2 minutes (30 seconds on and 30 seconds off) at 40% amplitude on ice (Qsonica, USA). The cell lysate was centrifuged at 37 000 × g for 60 minutes at 4°C Fiberlite fixed angle rotor (Thermo Scientific, USA) to separate the soluble and insoluble fraction. The soluble fractions (supernatant) was used for further analysis by immobilised metal affinity chromatography and gel filtration chromatography. The insoluble fraction (pellet) was resuspended in SDS-PAGE loading buffer for further analysis. The soluble and insoluble fractions were loaded onto an SDS-PAGE gel and the gel was stained and subsequently destained (Laemmli, 1970). The theoretical molecular weight and isoelectric point (pI) were computed for all proteins using an online prediction tool (http://web.expasy.org/compute_pi/) (Table 2.3).

Table 2.3: Theoretical molecular weight and isoelectric points.

Protein sample	Molecular weight (kDa)	pI
<i>PmGH</i> (native enzyme)	~ 130	5.9
GH5	~ 30	5
GH6	~ 40	5.1

2.3.1. Immobilised metal affinity chromatography

The soluble fraction was purified by Immobilised Metal Affinity Chromatography (IMAC) using Talon CellThru resin (Clontech, USA) charged with cobalt (2 ml bed volume) to bind his-tagged proteins. The cobalt beads were washed in 10-20 times the resin bed volume and equilibrated using lysis buffer (50 mM potassium phosphate buffer (pH 7); 300 mM NaCl) as per manufacturer's instructions. The soluble fraction was then added directly to the charged resin and the his-tagged protein was allowed to bind overnight on a tube roller at 4°C. Upon binding, the resin was separated from the unbound protein in a gravity flow column. The resin was then washed by adding 10-20 bed volumes of wash buffer (50 mM potassium phosphate buffer; 300 mM NaCl) (pH 7) and the proteins were eluted by the addition of 5 bed volumes of elution buffer (50 mM potassium phosphate buffer; 300 mM NaCl; 150 mM imidazole) (pH 7) to the resin. The collected fractions were run on a 10% (w/v) SDS-PAGE gel at 90 V, stained with Coomassie brilliant blue and destained (Laemmli, 1970). For all protein samples, the pure fractions were pooled and concentrated using Amicon ultra-15 centrifugal filters (Merck Millipore, USA) with the corresponding molecular weight cut-off (Table 2.3) (100 kDa cut-off for *PmGH*; 10 kDa cut-off for GH5 and GH6). Protein samples were further analysed by

functional analysis. The native enzyme, *PmGH*, was further purified by gel filtration as a high degree of purity was required for protein crystallisation trials.

2.3.2. Gel filtration chromatography of *PmGH*

The concentrated protein samples were loaded onto a HiPrep 16/60 Sephacryl S 200 HR (GE Healthcare, United Kingdom) pre-packed gel filtration column (60 cm x 16 mm) and eluted in 20 mM phosphate buffer (pH 6) by Fast Protein Liquid Chromatography (FPLC; AKTA). The proteins were eluted in 2 column volumes and 2 ml fractions were collected at a constant flow rate of 0.5 ml.min⁻¹. The fractions (2 ml) were analysed by SDS-PAGE (Laemmli, 1970) and subsequently pooled and concentrated using Amicon ultra-15 centrifugal filters (Merck Millipore, USA). Protein purity and molecular weights were subsequently analysed by SDS-PAGE and the structural and functional properties of all proteins were further investigated.

2.4. Enzyme characterisation

All enzyme assays were performed using the dinitrosalicylic (DNS) acid assay (Bailey *et al.*, 1992). As described by Bailey *et al.* (1992), blanks were prepared by incubating 900 µl of 1% (w/v) substrate (sodium citrate; pH 5) at 50°C for 5 minutes followed by the addition of 1.5 ml DNS and 100µl buffer and subsequently boiled for 5 minutes. To 900 µl of 1% (w/v) substrate incubated at 50°C, 100µl of enzyme was added and incubated at 50°C for 5 minutes. The enzyme reaction was stopped by the addition of 1.5 ml dinitrosalicylic acid (DNS) and boiled for 5 minutes. Reaction mixtures were cooled and end product release was measured spectrophotometrically at 540 nm against the blanks. For all control reactions 900 µl of 1% (w/v) substrate was incubated at 50°C (5 minutes) followed by the addition of 1.5 ml DNS and

100 μ l enzyme solution. The controls were subsequently boiled (5 minutes), cooled and absorbance was measured 540 nm.

Glucose, mannose and xylose standard curves were prepared for all pH and temperature intervals used in determining cellulase, mannanase and xylanase activities, respectively. All enzyme reactions were carried out in triplicate and the average absorbance readings were plotted. Enzyme activity was expressed in nanokatals per millilitre ($\text{nkat}\cdot\text{ml}^{-1}$) where one nkat is defined as the amount of enzyme required to catalyse the conversion of one nmol of substrate in one second under the specified reaction conditions. Specific activity was expressed in nkat per milligram of protein ($\text{nkat}\cdot\text{mg}^{-1}$).

2.4.1. Determination of the range of hydrolytic activities

Enzymes were tested on a range of substrates to determine the type of glycoside hydrolase activity. The hydrolytic activities of *PmGH*, GH5 and GH6 were determined using the DNS assay as previously described (Section 2.4) (Bailey *et al.*, 1992). The following substrates solutions were prepared (1%; w/v) in 20 mM sodium citrate buffer (pH 5): carboxymethyl cellulose (CMC), cellobiose, cellotriose, cellotetraose and cellopentaose, oat spelt xylan, beechwood xylan, locust bean gum, acacia gum (Table 2.4).

Table 2.4: Summary of the glycoside hydrolase activities tested and the corresponding substrates.

Hydrolytic activity	Substrate	Structural composition	Supplier
Exo-/endo-glucanase	Carboxymethyl cellulose	Soluble cellulose	Sigma, USA
Exo-/endo-glucanase	Avicel	Crystalline cellulose	
β -glucosidase	Cellobiose	Glucose di-saccharide	Megazyme, Ireland
	Cellotriose	Glucose tri-saccharide	
	Cellotetraose	Glucose tetra-saccharide	
	Cellopentaose	Glucose penta-saccharide	
Endo-xylanase	Beechwood xylan	Xylose polysaccharide ~10% arabinose	Sigma, USA
	Oat spelt xylan	Xylose polysaccharide ~15% glucose ~10% arabinose	Sigma, USA
Endo-mannanase	Gum Arabic	Arabinogalactan	Sigma, USA
	Locust bean gum	Galactomannan	Sigma, USA

2.4.2. pH and temperature optimum and stability

The optimum pH values for *PmGH*, GH5 and GH6 were determined for the corresponding hydrolytic activities detected in Section 2.1 above, using the DNS assay (Bailey *et al.*, 1992).

Enzyme assays were conducted under standard assay conditions at 50°C. A pH range of pH 4-9 was tested using the following buffers: sodium citrate buffer (pH 4-5), potassium phosphate buffer (pH 6-8) and Tris-HCl (pH 9). The temperature optima were assayed within the temperature range of 40-90°C at the pre-determined pH optimum as described above for each variant (*PmGH*, GH5 and GH6). All substrate solutions were prepared at a concentration of 1% (w/v) CMC, beechwood xylan and acacia gum.

The functional pH stability and thermostability profiles of *PmGH*, GH5 and GH6 were determined at pH 4-9 and 40-90°C, respectively. The enzymes were incubated at each pH and samples were withdrawn every 15 minutes for two hours. The residual activities of the samples were determined using the DNS assay. Similarly, thermostability was assayed within the temperature range of 40-90°C over a two hour incubation period and samples were withdrawn every 15 minutes. Enzyme activity was determined using the DNS assay and expressed as percentage residual activity.

2.4.3. *PmGH* identification by mass spectrometry

The pure *PmGH* protein was loaded onto a 10% SDS-PAGE gel and electrophoresed at 100V for one hour. The protein gel was Coomassie stained and subsequently destained. The molecular weight of the protein was estimated against a PAGERuler prestained protein ladder (ThermoScientific, USA). The excised band was destained until light blue and fixed in glutaraldehyde. The sample was placed in an Eppendorf tube and sent to the Central Analytical Facilities Unit (Stellenbosch University, South Africa) for peptide analysis by liquid chromatography-mass spectrometry (LC-MS). Proteolytic digestion and data analysis was performed at the Central Analytical Facilities Unit (Stellenbosch University, South Africa).

2.5. Homology modelling

Three-dimensional structure prediction of *PmGH* was carried out by alignment of target sequences with template structures using SwissModel online tool (Guex *et al.*, 2009; Waterhouse *et al.*, 2018). Due to the multi-modular nature of *PmGH*, no suitable structural homolog was identified. As a result the composite domains of *PmGH* were modelled and all figures were prepared in PyMOL (De Lano, 2002). Template sequences used in homology modelling and corresponding sequence identities and percentage sequence coverage are indicated in Table 2.5.

Table 2.5: Structural homologs of the composite domains of *PmGH*.

Protein	Homolog PDB Id.	Sequence identity	Sequence coverage	Source	Reference
GH5	1LF1	56%	100%	Alkaline <i>Bacillus</i> sp.	(Shaw <i>et al.</i> , 2002)
GH6	4AVN	55%	93%	<i>Thermobifida fusca</i>	(Sandgren <i>et al.</i> , 2013)
CBM3a	2WO4	33%	100%	<i>Clostridium thermocellum</i>	(Yaniv <i>et al.</i> , 2013)
CBM3b	1NBC	43%	87%	<i>Clostridium thermocellum</i>	(Yaniv <i>et al.</i> , 2012)

2.6. Protein crystallisation

Crystallisation screening trials were conducted using a range of commercially available screens (Hamptons Research, USA) for all pure protein samples. Initial crystallisation screens were set up using the sitting-drop vapour diffusion method (Mosquito crystal, TTP LabTech, United Kingdom) and further optimised by hanging-drop (Figure 2.2). Hanging drops composed of 2 μ l protein solution and 2 μ l reservoir solution were prepared. Protein crystallisation conditions were further optimised by varying the type of buffer, concentration of buffer, pH of buffer, concentration of precipitant, salt concentration and temperature. The quality of the all protein crystals were assessed and further enhanced by bead seeding (10^{-1} to 10^{-10} serial dilutions per round) and streak seeding.

2.6.1. Crystal soaking

Protein crystals of *PmGH* and GH5 were soaked in saturated solutions of glucose, xylose, mannose and cellobiose prepared in the corresponding reservoir solution. Glycoside hydrolase 6 domain protein crystals were soaked in a saturated solution of glucose prepared in reservoir solution. Competitive inhibitors of *PmGH*, GH5 and GH6 were selected based on the enzyme activities determined by the DNS assay (Section 2.4). The soaked protein crystals of *PmGH*, GH5 and GH6 were incubated for 2 hours and washed in reservoir solution to remove unbound inhibitor (sugar). The crystals were subsequently used in diffraction experiments to screen for the best diffraction conditions.

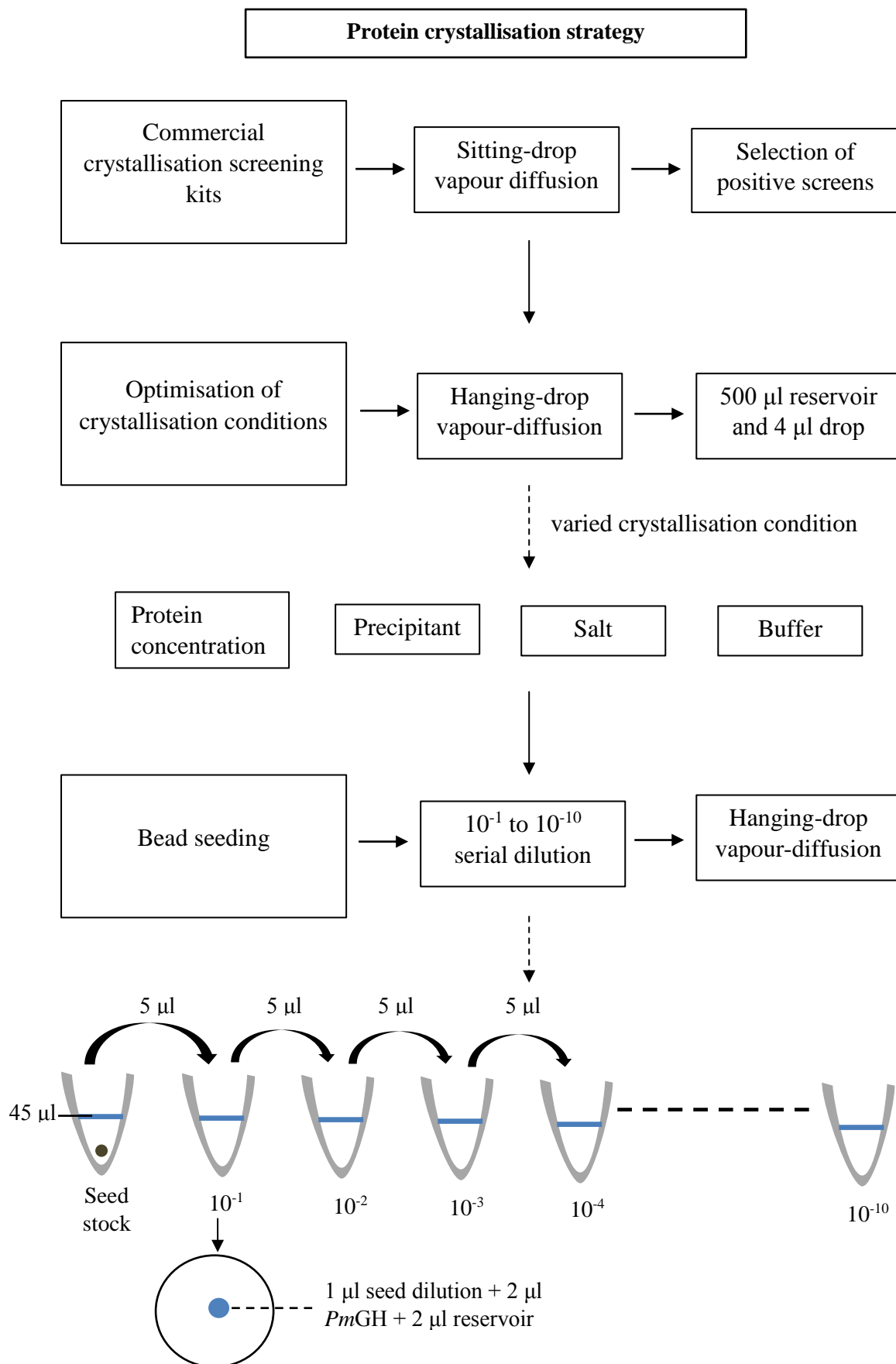


Figure 2.2: Diagrammatic overview of the protein crystallisation strategy employed in the crystallisation of the multi-modular protein, *PmGH*.

Native protein crystals of *PmGH* were soaked in various heavy atom derivatives to aid in the elucidation of the crystal structure of the novel protein *PmGH*. Commercially available heavy atom screens of mercury, platinum and gold (Hamptons Research, USA) were tested on the native *PmGH* protein to increase the intensity of the diffraction data at the various heavy atom binding sites. Protein crystals were harvested, transferred to the corresponding cryo-protectant (25% (v/v) PEG 400 or 30% (v/v) glycerol) and flash cooled in liquid nitrogen. Protein crystals were stored in a cooled cryogenic storage dewar and shipped to the European Synchrotron Radiation Facility (ESRF) (Grenoble, France).

2.7. Data collection, phasing and refinement

X-ray diffraction data for *PmGH* was collected using synchrotron X-ray radiation on beamlines MX1845 and ID30A at the ESRF (European Synchrotron Radiation Facility) (Grenoble, France). Phasing by molecular replacement was carried out using Phaser (McCoy *et al.*, 2007) within the PHENIX package (Adams *et al.*, 2010) using previously determined structures as search models. Extensive analysis of the diffraction data by MR involved: i) creating ensembles using search models for the individual domains (Table 2.5), ii) the use of more than one ensemble or search model for each domain, iii) the removal of loops from the search models to account for structural flexibility. Several attempts were also made to solve the *PmGH* phases by MAD. However, data collection proved to be a major bottleneck.

Chapter 3

Results

3.1. Cloning and expression

Polymerase chain reaction amplicons of *PmGH* (~3000 bp), GH6 (~1159 bp) and GH5 (~765 bp) were sequenced and subjected to BLAST analysis. The full length gene (*PmGH*) and its catalytic domains were further identified by an NCBI conserved domain search as shown in Figure 3.1 below.

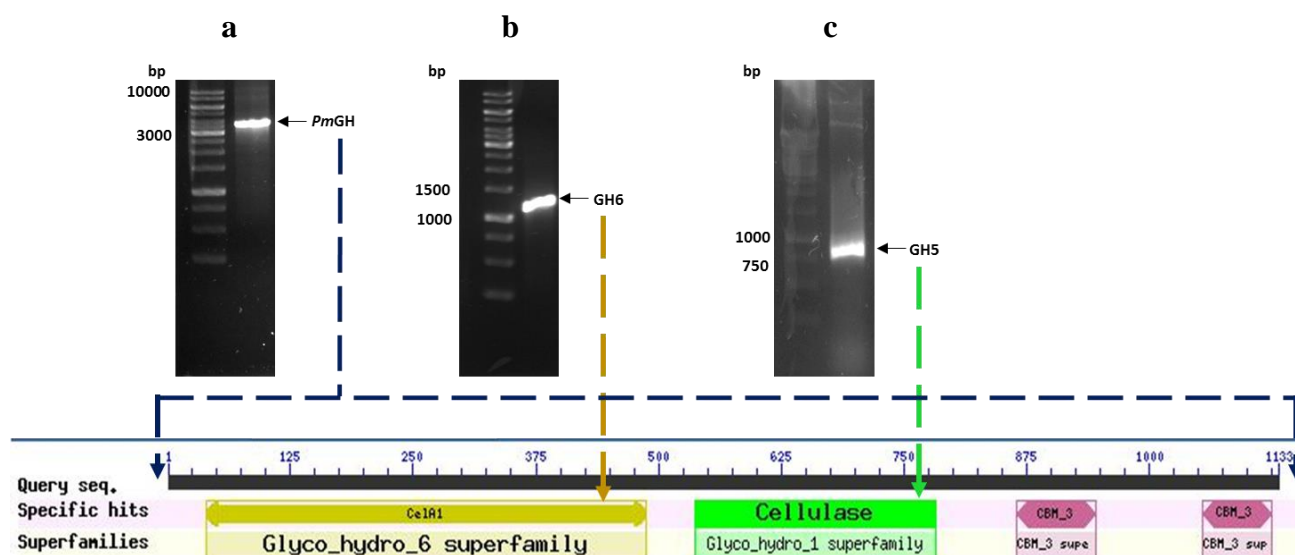


Figure 3.1: PCR products of (a) *PmGH*, (b) GH5 and (c) GH6 and subsequent identification by BLAST analysis and NCBI conserved domain search.

Expression of the *PmGH* gene in *E.coli* [pET-21 a (+)] was optimised by varying induction time and IPTG concentration (Figure 3.2). Small scale *PmGH* expression trials showed that the addition of 0.1 mM IPTG and an 8 hour induction time allowed for optimal protein production.

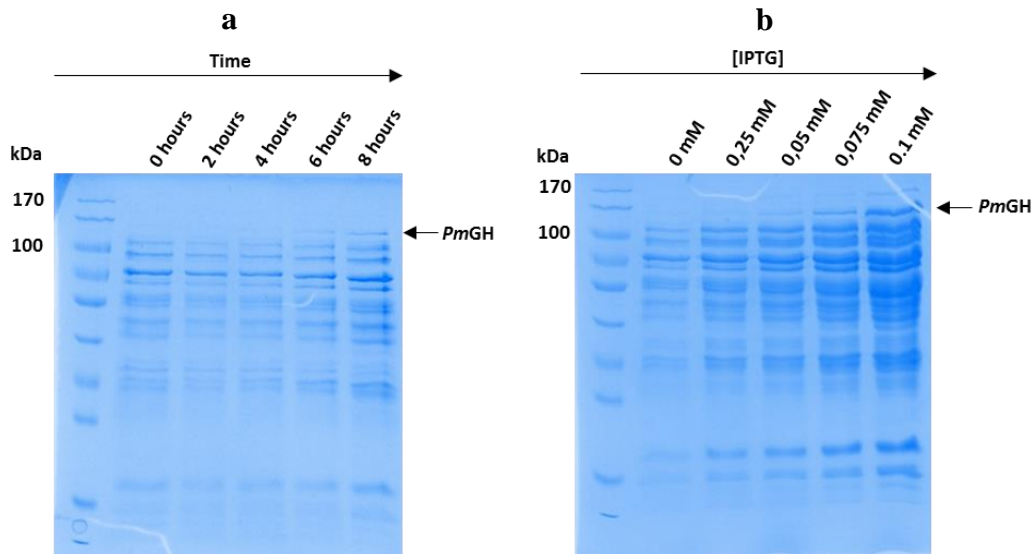


Figure 3.2: Optimisation of the small scale production of *PmGH* by varying (a) induction time (hours) and (b) IPTG concentration.* A PageRuler Prestained protein ladder (10-170 kDa) was used as a molecular weight marker.

*Induction times ranged from 0 to 8 hours (2 hour intervals). IPTG concentration varied from 0 to 0.1 mM (0.025 mM increments).

3.2. Protein purification

Purification of *PmGH* (~130 kDa) by gravity flow IMAC resulted in the partial purification of the enzyme as evidenced by SDS-PAGE (Figure 3.3). Further purification by gel filtration resulted in a highly pure sample suitable for protein crystallisation. Protein fractions A6 and A7 were pooled, concentrated, analysed by SDS-PAGE and used in subsequent analysis.

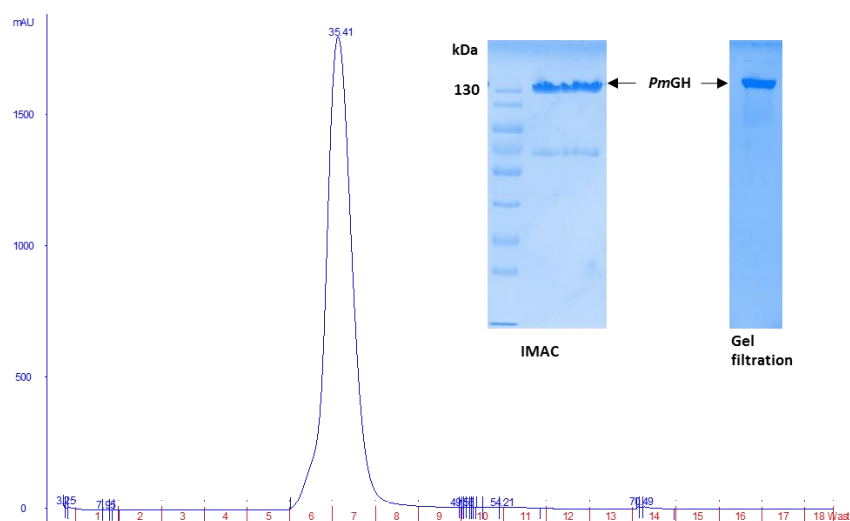


Figure 3.3: Purification of *PmGH* by gravity flow IMAC and gel filtration for crystallisation and enzyme characterisation. SDS-PAGE gels show the pooled and concentrated *PmGH* fractions after IMAC and gel filtration chromatography.

The catalytic domains GH5 (~30 kDa) and GH6 (~40 kDa) were produced (pET-28 a (+); 8 hour induction time; 0.1 mM IPTG) as described for *PmGH* and subsequently partially purified by gravity flow IMAC (Figure 3.4). Unlike *PmGH*, a high degree of purity was not required as both enzymes were subsequently used for functional analysis. Functional analysis was dependent on the ability of the enzyme to bind to the substrate. This is typically a highly specific reaction which is unlikely to be affected by the presence of minor protein contaminants. Furthermore, DNS is specific for reducing sugars, the end product of the hydrolysis of polysaccharides by GHs.

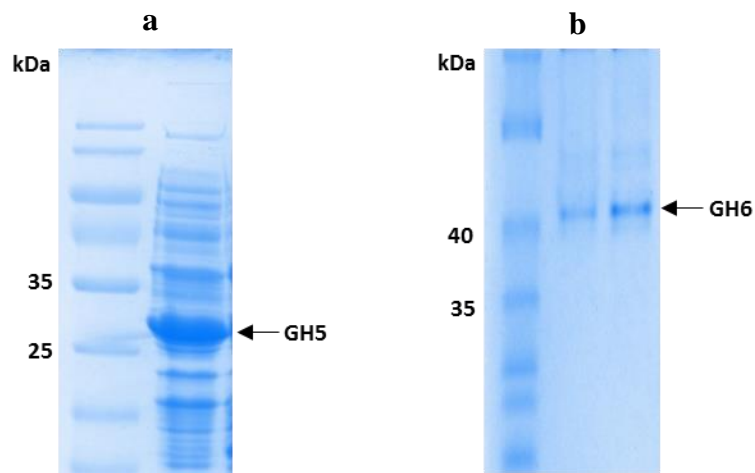


Figure 3.4: Partial purification of (a) GH5 and (b) GH6 catalytic domains by gravity flow IMAC using Talon CellThru resin charged with cobalt.

3.3. Functional characterisation

3.3.1. Substrate range of *PmGH*, GH5 and GH6

The range of hydrolytic activities of *PmGH*, GH5 and GH6 were determined using the DNS assay (Table 3.1). A range of substrates were tested as described in Table 2.4 (Section 2.4.1) above. Cellulase, xylanase and mannanase activities were detected in *PmGH* and GH5. Strictly cellulase activity was detected in GH6.

Table 3.1: Polymeric substrates and hydrolytic activities detected in *PmGH*, GH5 and GH6 using the DNS assay

Activity (nkat.ml⁻¹)			
Substrates	<i>PmGH</i>	GH5	GH6
CMC	18	12	17
Beechwood xylan	22	23	0
Locust bean gum	17	13	0

3.3.2. pH and temperature optima profiles of *PmGH*, GH5 and GH6

The pH optima of the native enzyme *PmGH* and its composite catalytic domains were determined for each of the hydrolytic activities in Table 3.1 above. All hydrolytic activities, *i.e.* cellulase, xylanase and mannanase, detected in *PmGH* displayed optimal activity at pH 6 (Figure 3.5). The sub-cloned GH5 catalytic domain showed optimal cellulose and xylanase activity at pH 5. Similarly, optimal cellulase activity was detected at pH 5 for the GH6 catalytic domain.

The temperature optima profiles of *PmGH* indicated optimal cellulase, mannanase and xylanase activities at 60°C (Figure 3.6). Optimal cellulase and mannanase activity was demonstrated at 70°C for the GH5 catalytic domain. Similarly, optimum cellulase activity was displayed at 70°C for the GH6 catalytic domain.

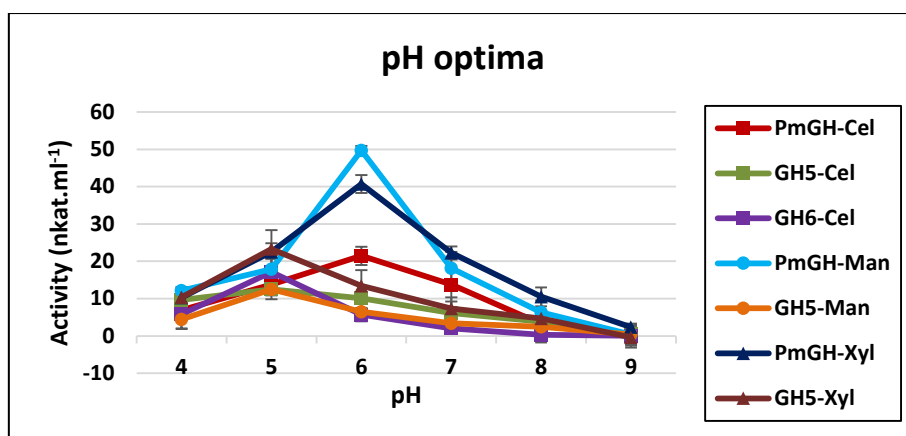


Figure 3.5: pH optima of the native enzyme, *PmGH*, and its composite catalytic domains, GH5 and GH6. The pH optima was determined for each enzyme and its corresponding hydrolytic activities (cellulase, mannanase, xylanase).*

*Red: *PmGH* cellulase pH profile; olive green: GH5 cellulase pH profile; purple: GH6 cellulase pH profile; light blue: *PmGH* mannanase pH profile; orange: GH5 mannanase pH profile; dark blue: *PmGH* xylanase pH profile; brown: GH5 xylanase pH profile.

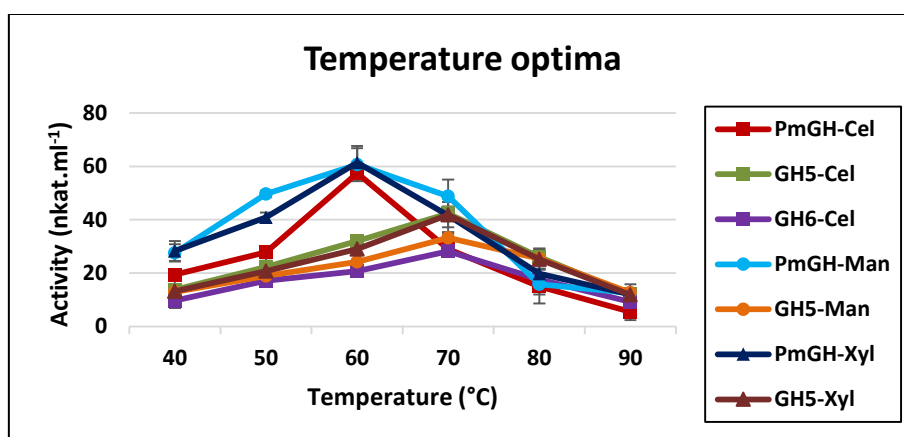


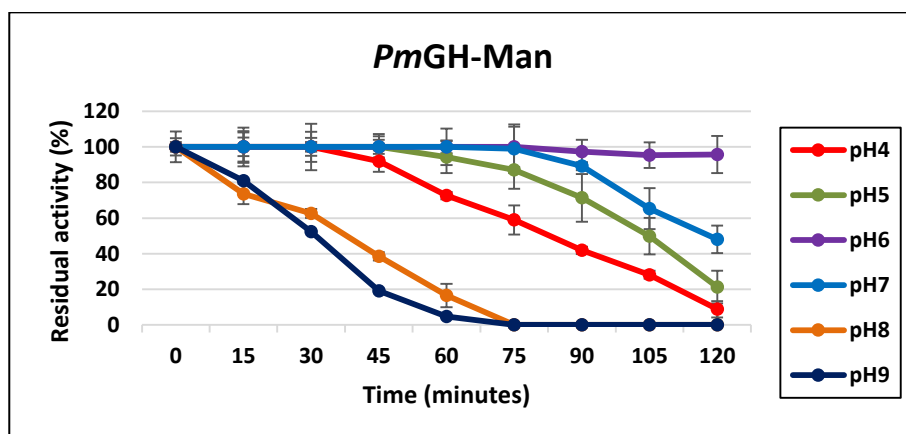
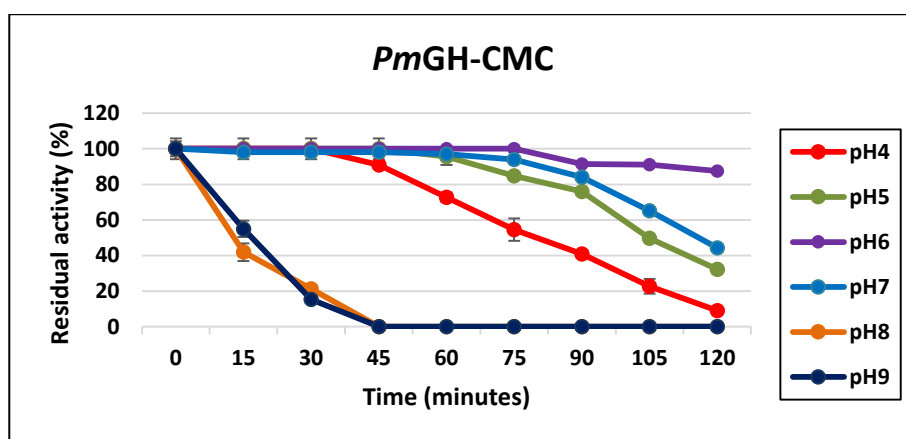
Figure 3.6: Temperature optima of the native enzyme, *PmGH*, and its composite catalytic domains, GH5 and GH6. Temperature conditions were determined for each enzyme and their corresponding hydrolytic activities under the pre-determined optimum pH conditions.*

*Red: *PmGH* cellulase temperature profile; olive green: GH5 cellulase temperature profile; purple: GH6 cellulase temperature profile; light blue: *PmGH* mannanase temperature profile; orange: GH5 mannanase temperature profile; dark blue: *PmGH* xylanase temperature profile; brown: GH5 xylanase temperature profile.

3.3.3. pH stability profiles of *PmGH*, GH5 and GH6

The functional pH stability of the native glycoside hydrolase, *PmGH*, and its catalytic domains, GH5 and GH6, were tested over a 2 hour incubation period. Functional stability was determined for each enzyme and its corresponding hydrolytic activities from pH 4-9.

The pH stability profiles of the cellulase, mannanase and xylanase activities of *PmGH* indicated that highest functional stability was attainable at pH 6 (Figure 3.7). Overall, greater than 85% of enzyme activities were retained at optimum pH (pH 6) after the 2 hour incubation. A minimum of 94% residual activity was retained at pH 7 for all hydrolytic activities after a 75 minute incubation period followed by a gradual decrease in enzyme activities. Although moderately pH stable at pH 4, *PmGH* retained 100% of hydrolytic activities for 30 minutes followed by a sharp decrease in residual activities. Reduced pH stability was observed under alkaline conditions (pH 8-9) resulting in a rapid decrease in enzyme activities over time.



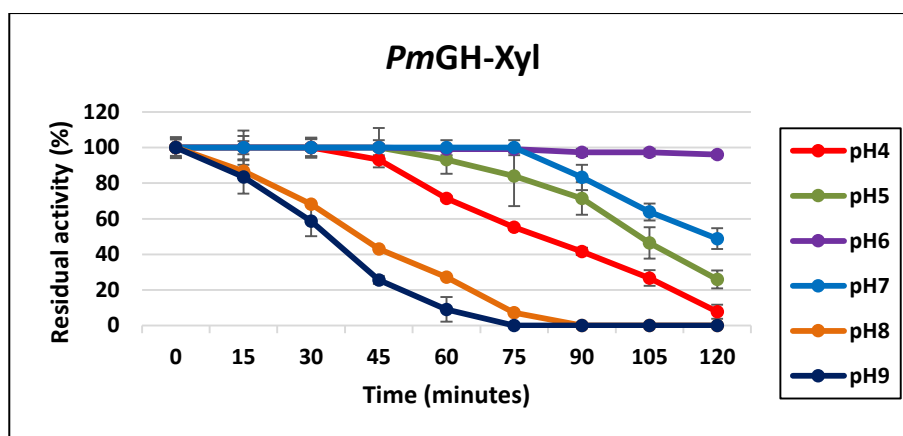


Figure 3.7: pH stability profiles of the (a) cellulase, (b) mannanase and (c) xylanase activities detected in the native enzyme, *PmGH*. Native enzyme was prepared and incubated at each pH condition (pH 4-9). Residual enzyme activity was determined on the substrates, carboxymethyl cellulose (CMC), beechwood xylan and locust bean gum using the DNS assay*.

*Red: *PmGH* pH stability profile at pH 4; olive green: *PmGH* pH stability profile at pH 5; purple: *PmGH* pH stability profile at pH 6; light blue: *PmGH* pH stability profile at pH 7; orange: *PmGH* pH stability profile at pH 8; dark blue: *PmGH* pH stability profile at pH 9.

Overall, maximum and minimum pH stability were detected from pH 4-6 and pH 7-9, respectively, in the GH5 catalytic domain (Figure 3.8). At optimum pH (pH5) approximately 81% of cellulase, mannanase and xylanase activities were retained for 2 hours. Approximately 70% of enzyme activities were retained at pH 4 after 2 hours. Close to 100% of enzyme activities were retained at pH 6 after 60 minutes followed by a gradual decrease in residual enzyme activities. However, a rapid decrease in residual enzyme activities was observed at pH 7 after 15 minutes. Minimal pH stability was observed under alkaline conditions (pH 8-9) resulting in a sudden reduction in enzyme activities.

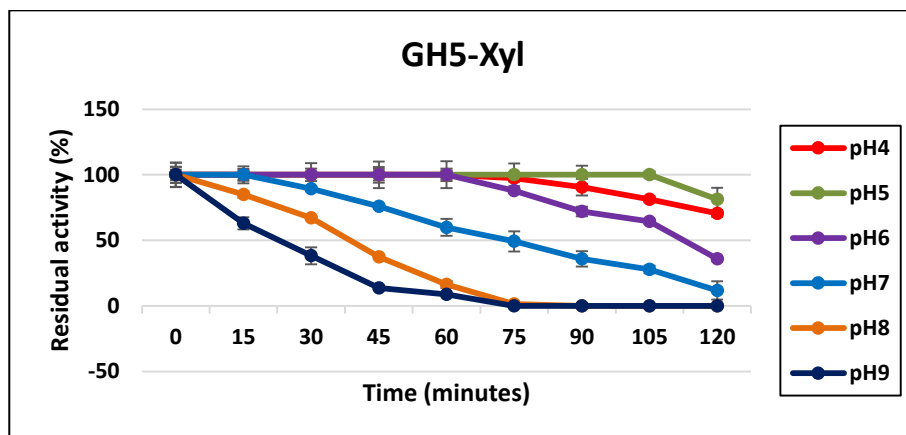
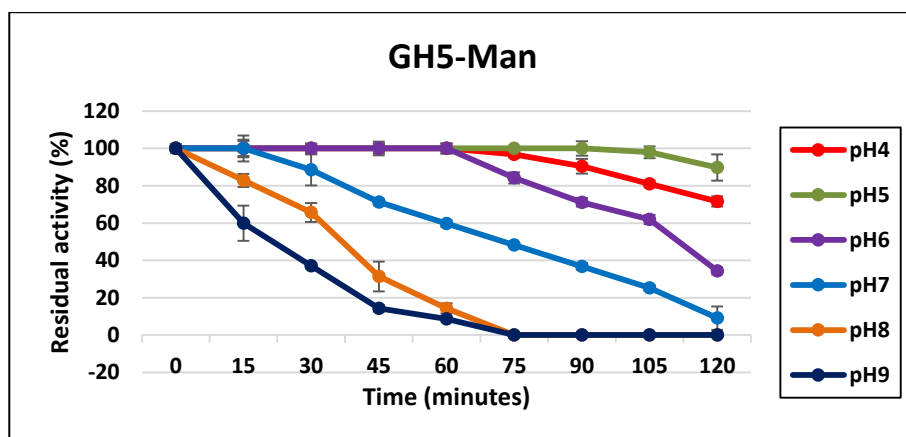
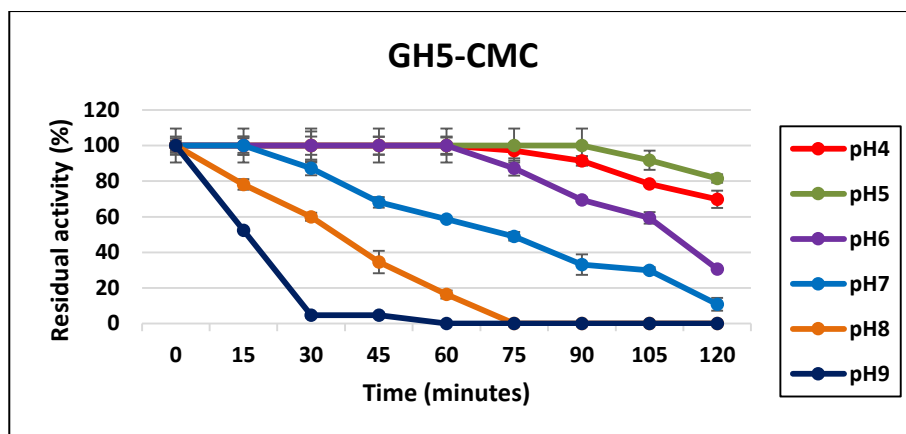


Figure 3.8: pH stability profiles of the (a) cellulase, (b) mannanase and (c) xylanase activities detected in the GH5 domain. GH5 was prepared and incubated at each pH condition (pH 4-9). Residual enzyme activity was determined on the substrates, carboxymethyl cellulose (CMC), beechwood xylan and locust bean gum using the DNS assay*.

*Red: GH5 pH stability profile at pH 4; olive green: GH5 pH stability profile at pH 5; purple: GH5 pH stability profile at pH 6; light blue: GH5 pH stability profile at pH 7; orange: GH5 pH stability profile at pH 8; dark blue: GH5 pH stability profile at pH 9.

The GH6 cellulase remained relatively stable at pH values ranging from 4-7 with an extreme reduction in residual activity at pH 8-9 (Figure 3.9). At optimum pH (pH 5), 100% of cellulase activity was retained after for 90 minutes followed by a steady decrease resulting in 62% residual cellulase activity after 2 hours. Approximately 100% and 85% cellulase activity was retained at pH 4 and pH 6, respectively after 75 minutes. Under neutral pH conditions, 100% cellulase activity was retained for 30 minutes, thereafter a continuous reduction in activity was observed.

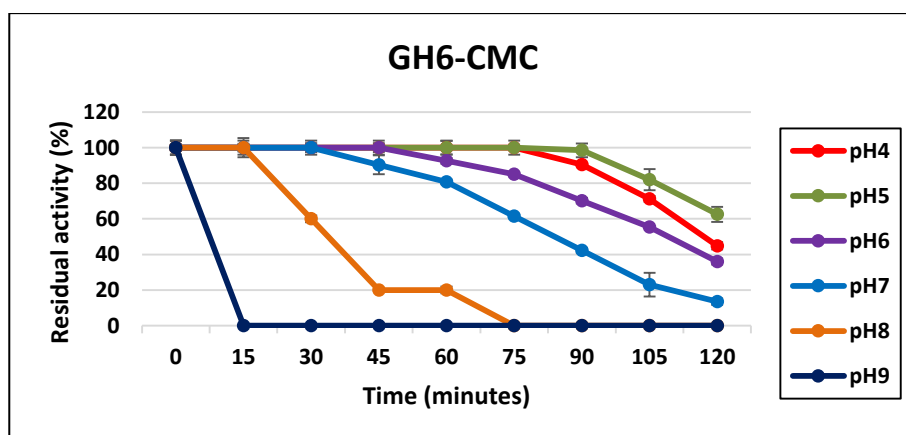


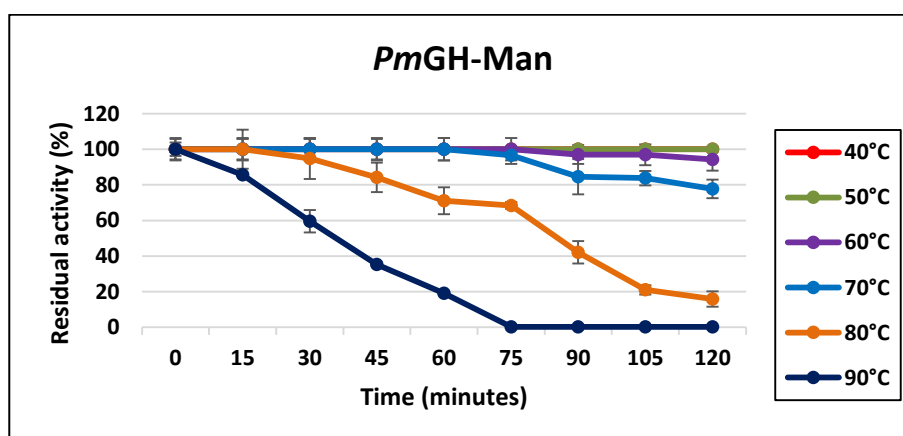
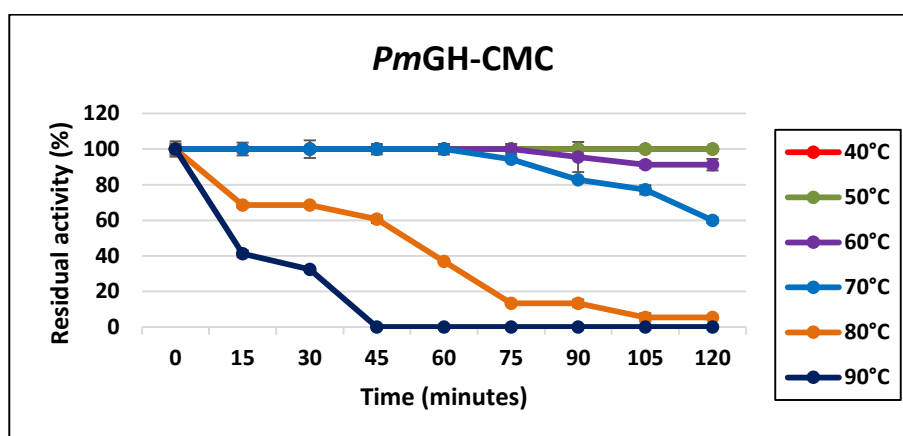
Figure 3.9: pH stability profile of the cellulase activity detected in the GH6 domain at pH 4-9 over a 2 hour incubation period. GH6 was prepared and incubated at each pH condition (pH4-9) and tested for residual activity using the DNS assay. Residual enzyme activity was determined on the substrate, carboxymethyl cellulose (CMC) using the DNS assay*.

*Red: GH6 pH stability profile at pH 4; olive green: GH6 pH stability profile at pH 5; purple: GH6 pH stability profile at pH 6; light blue: GH6 pH stability profile at pH 7; orange: GH6 pH stability profile at pH 8; dark blue: GH6 pH stability profile at pH 9.

3.3.4. Thermostability profiles of *PmGH*, GH5 and GH6

Functional thermostability of *PmGH*, and GH5 and GH6 catalytic domains was tested over a 2 hour incubation period. Temperatures ranging from 40 to 90°C were used in determining the thermostability of the hydrolytic activities of the different enzyme and domain preparations.

The general trend observed in the thermostability profiles of *PmGH* showed high thermostability at 40-70°C and reduced thermostability at elevated temperatures (80-90°C) (Figure 3.10). The enzyme retained 100% of cellulase, mannanase and xylanase activities at 40°C and 50°C over 2 hours. At 60°C, 91-94% of hydrolytic activities were retained after 2 hours. Approximately, 78% mannanase and 72% xylanase activity were retained at 70°C for 2 hours. In contrast, 60% of *PmGH* cellulase activity was retained under the same conditions. A rapid decline in all *PmGH* hydrolytic activities was observed at 80°C and 90°C over 2 hours.



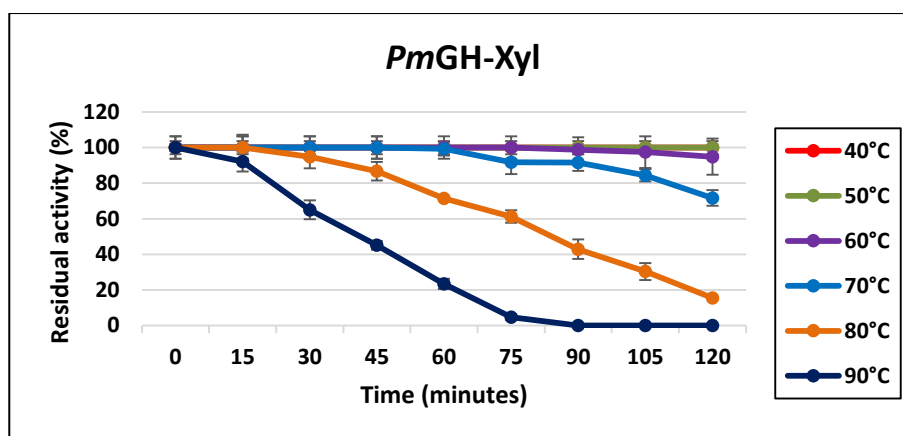


Figure 3.10: Thermostability profiles of the (a) cellulase, (b) mannanase and (c) xylanase activities for the native enzyme, *PmGH*. The native enzyme was incubated for 2 hours at 40-90°C under optimal pH conditions (pH 6). Residual enzyme activity was determined on the substrates, carboxymethyl cellulose (CMC), beechwood xylan and locust bean gum using the DNS assay*.

*Red: *PmGH* thermostability profile at 40°C; olive green: *PmGH* thermostability profile at 50°C; purple: *PmGH* thermostability profile at 60°C; light blue: *PmGH* thermostability profile at 70°C; orange: *PmGH* thermostability profile at 80°C; dark blue: *PmGH* thermostability profile at 90°C.

Overall, the thermostability profiles of GH5 highlight its highly thermostable nature at 40-70°C and reduced thermostability at 80-90°C for all hydrolytic activities (cellulase, xylanase and mannanase) (Figure 3.11). At 40-50°C, 100% of hydrolytic activities were retained after the 2 hour incubation period. Approximately 90-92% of hydrolytic activities were retained at 60°C after 2 hours. At 70°C, 100% of all hydrolytic activities were retained for 60 minutes, followed a gradual decrease in residual activities. Thermostability of the mannanase and xylanase activities were significantly reduced at 80°C and 90°C followed by a steady decline after 15 minutes.

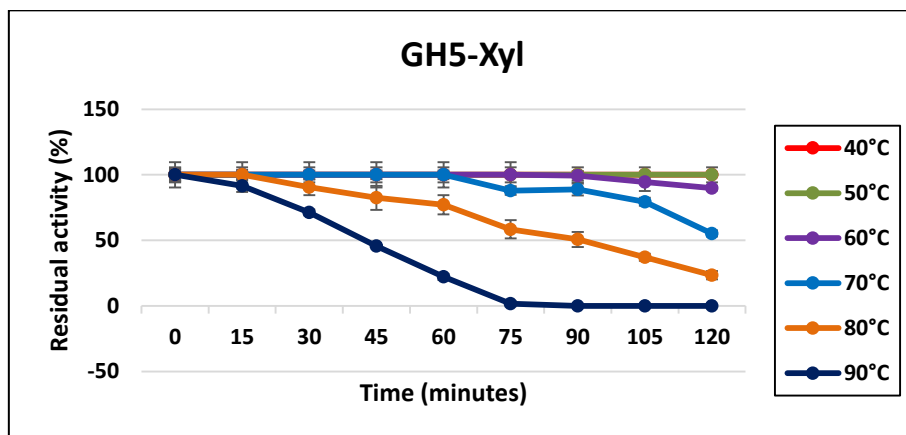
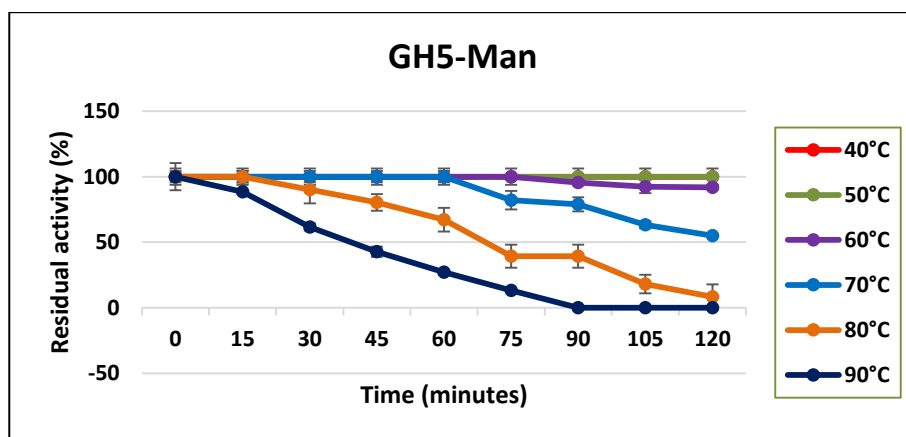
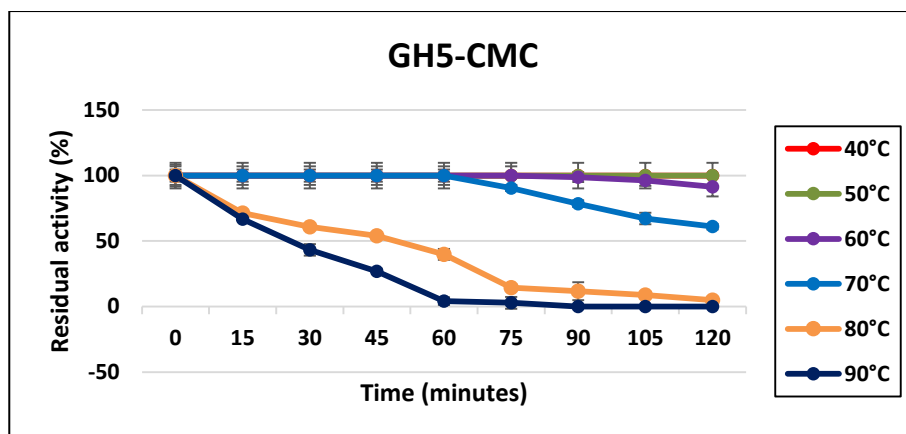


Figure 3.11: Thermostability profiles of the (a) cellulase, (b) mannanase and (c) xylanase activities for the GH 5 domain. GH5 was incubated for 2 hours at 40-90°C under optimal pH conditions (pH 5). Residual enzyme activity was determined on the substrates, carboxymethyl cellulose (CMC), beechwood xylan and locust bean gum using the DNS assay*.

*Red: GH5 thermostability profile at 40°C; olive green: GH5 thermostability profile at 50°C; purple: GH5 thermostability profile at 60°C; light blue: GH5 thermostability profile at 70°C; orange: GH5 thermostability profile at 80°C; dark blue: GH5 thermostability profile at 90°C.

Compared to *PmGH* and GH5, the GH6 catalytic domain displayed increased functional thermostability, especially under extreme conditions (80-90°C) (Figure 3.12). Prolonged thermostability was observed between 40-50°C with 100% of cellulase activity being retained after 2 hours. At 60°C and 70°C, 78% and 48% cellulase activity were retained for 2 hours, respectively. Unlike *PmGH* and GH5, 100% of cellulase activity was retained for 30 minutes at 80°C, followed by a steady decrease in residual activity. Although minimal, cellulase activity was still detectable after 2 hours at 80°C (20%) and 90°C (8%).

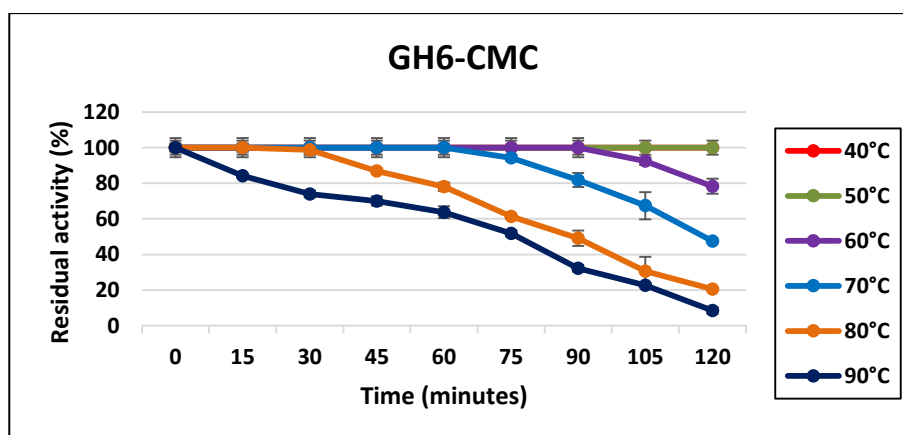


Figure 3.12: Thermostability profile of GH6 (cellulase) over a 2 hour incubation period at temperatures ranging from 40-90°C under optimal pH conditions (pH 5). Residual enzyme activity was determined on the substrate, carboxymethyl cellulose (CMC) using the DNS assay.*

*Red: GH6 thermostability profile at 40°C; olive green: GH6 thermostability profile at 50°C; purple: GH6 thermostability profile at 60°C; light blue: GH6 thermostability profile at 70°C; orange: GH6 thermostability profile at 80°C; dark blue: GH6 thermostability profile at 90°C.

3.4. Dinitrosalicylic acid assay standard curves

Glucose, mannose and xylose standard curves were prepared for all DNS assays. All standard curves had an R-squared (R^2) of approximately 0.9. An R^2 value closer to 1 indicated a high degree of accuracy. Cellulase (Figure 3.13), xylanase (Figure 3.14) and mannanase (Figure 3.15) activities were determined using the equations shown in the standard curve and expressed in nkat.ml^{-1} .

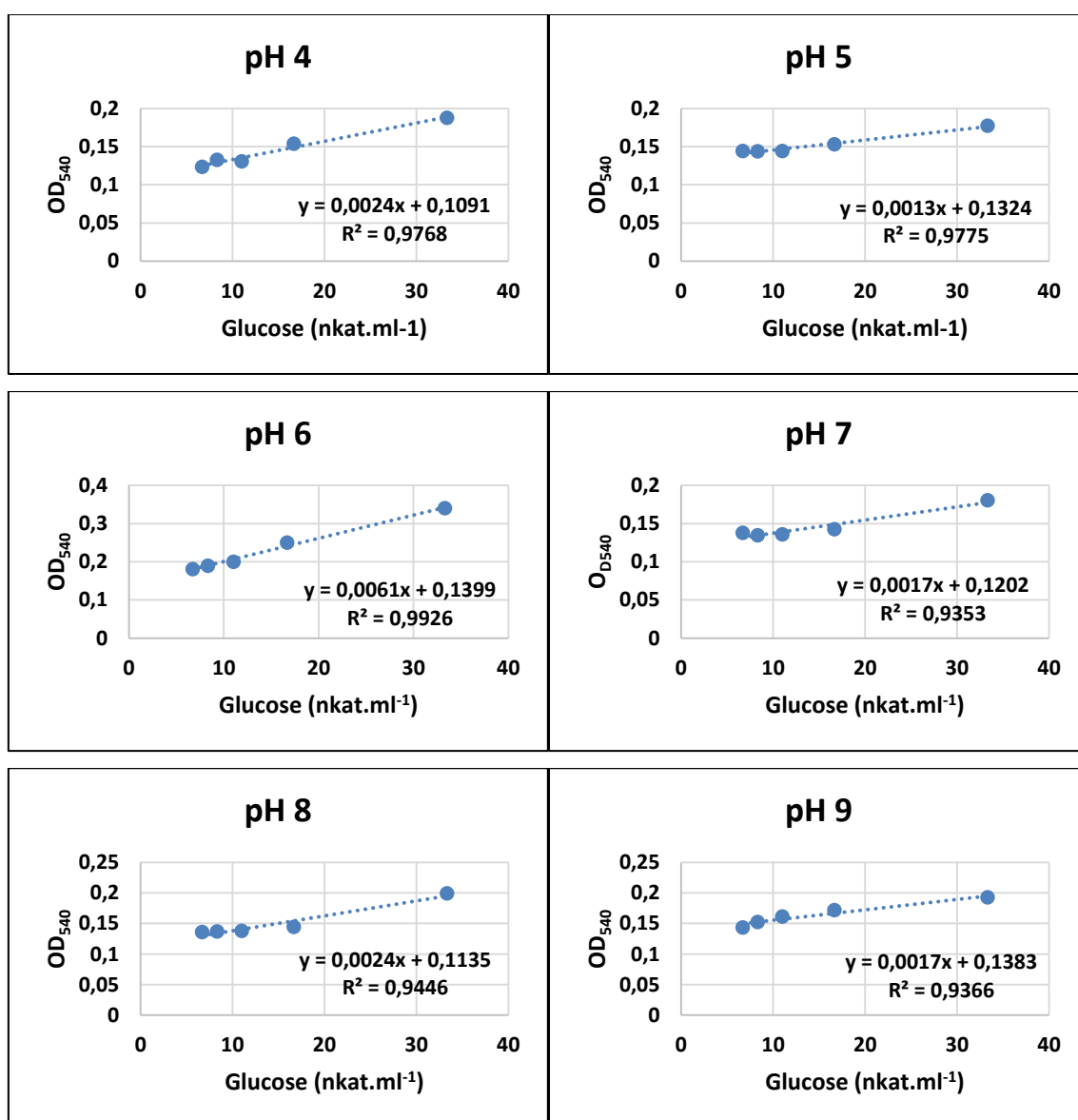


Figure 3.13: Glucose standard curves for pH 4-9 using the DNS assay. Cellulase activity was determined using the standard curve equation and expressed in nkat.ml^{-1} .

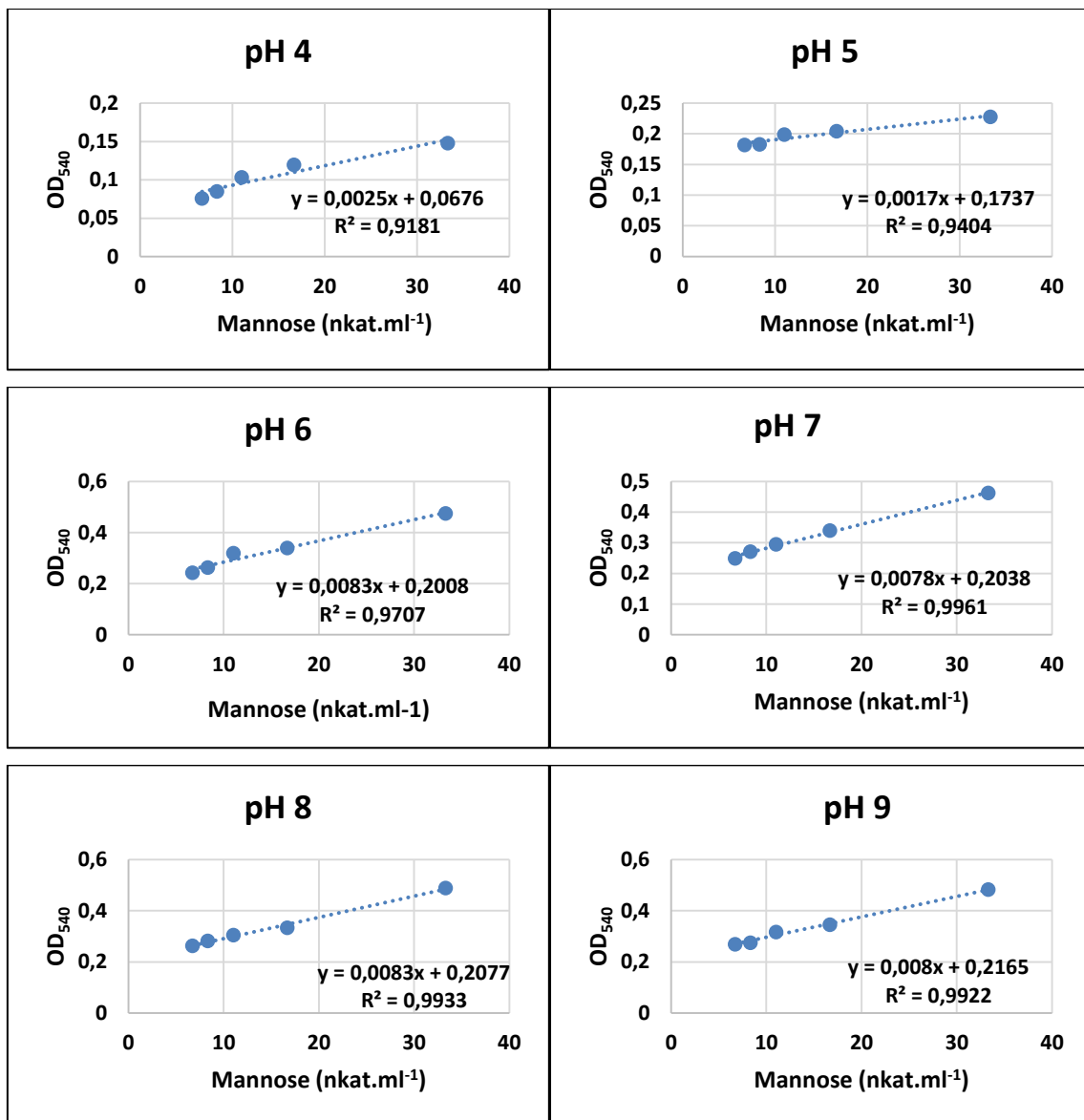


Figure 3.14: Mannose standard curves for pH 4-9 using the DNS assay. Mannanase activity was determined using the standard curve equation and expressed in nkat.ml⁻¹.

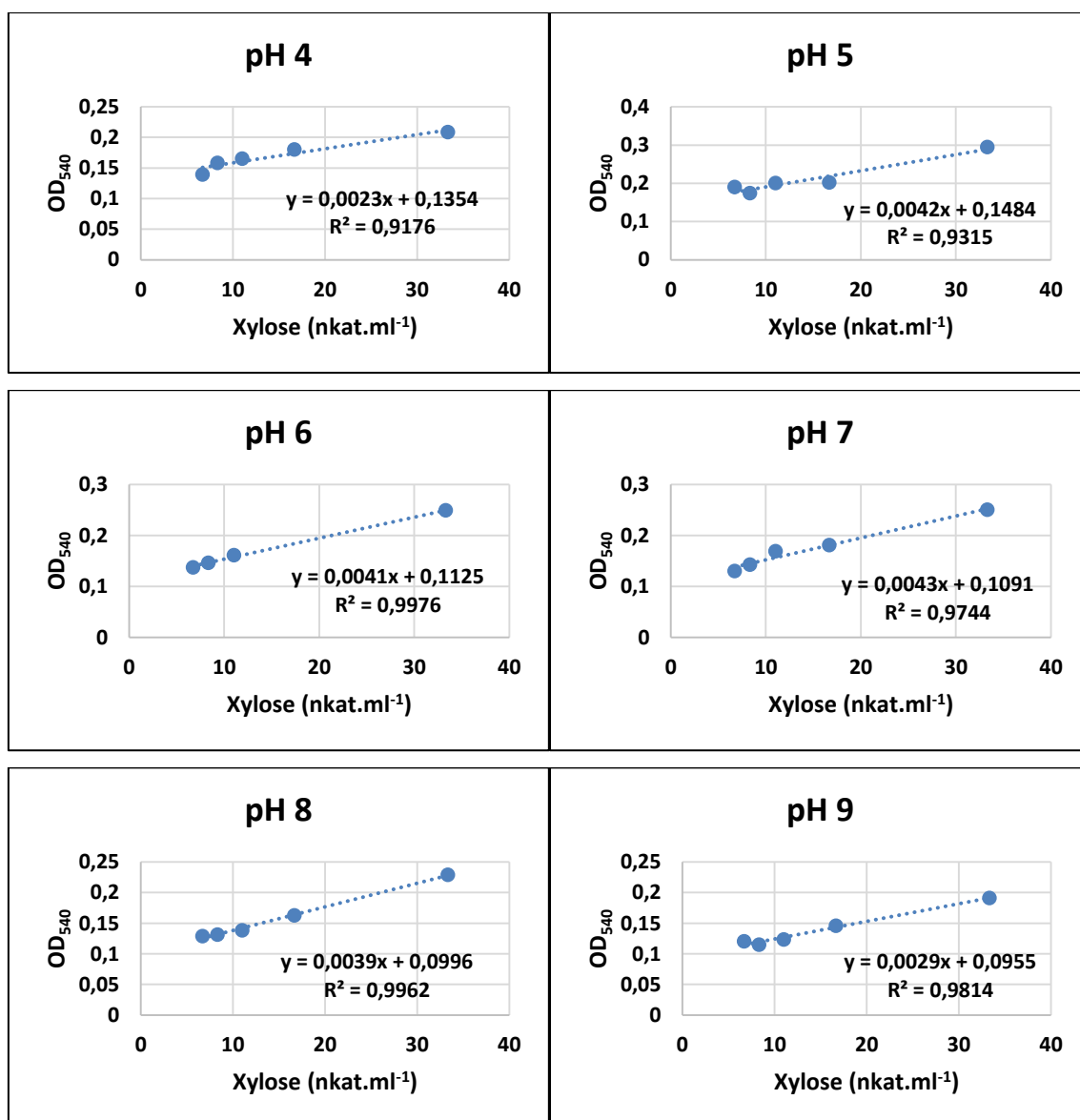


Figure 3.15: Xylose standard curves for pH 4-9 using the DNS assay. Xylanase activity was determined using the standard curve equation and expressed in nkat.ml⁻¹.

3.5. Homology modelling of *PmGH*

Homology modelling using SwissModel resulted in an incomplete model of *PmGH* (Figure 3.16a). The incomplete homology model represented the larger domains (GH5 and GH6) and excluded the CBMs and adjoining linker regions of the native enzyme (*PmGH*). The low global model quality estimation (GMQE) and QMEAN values indicated that the model was not an accurate representation of *PmGH* (Table 3.2).

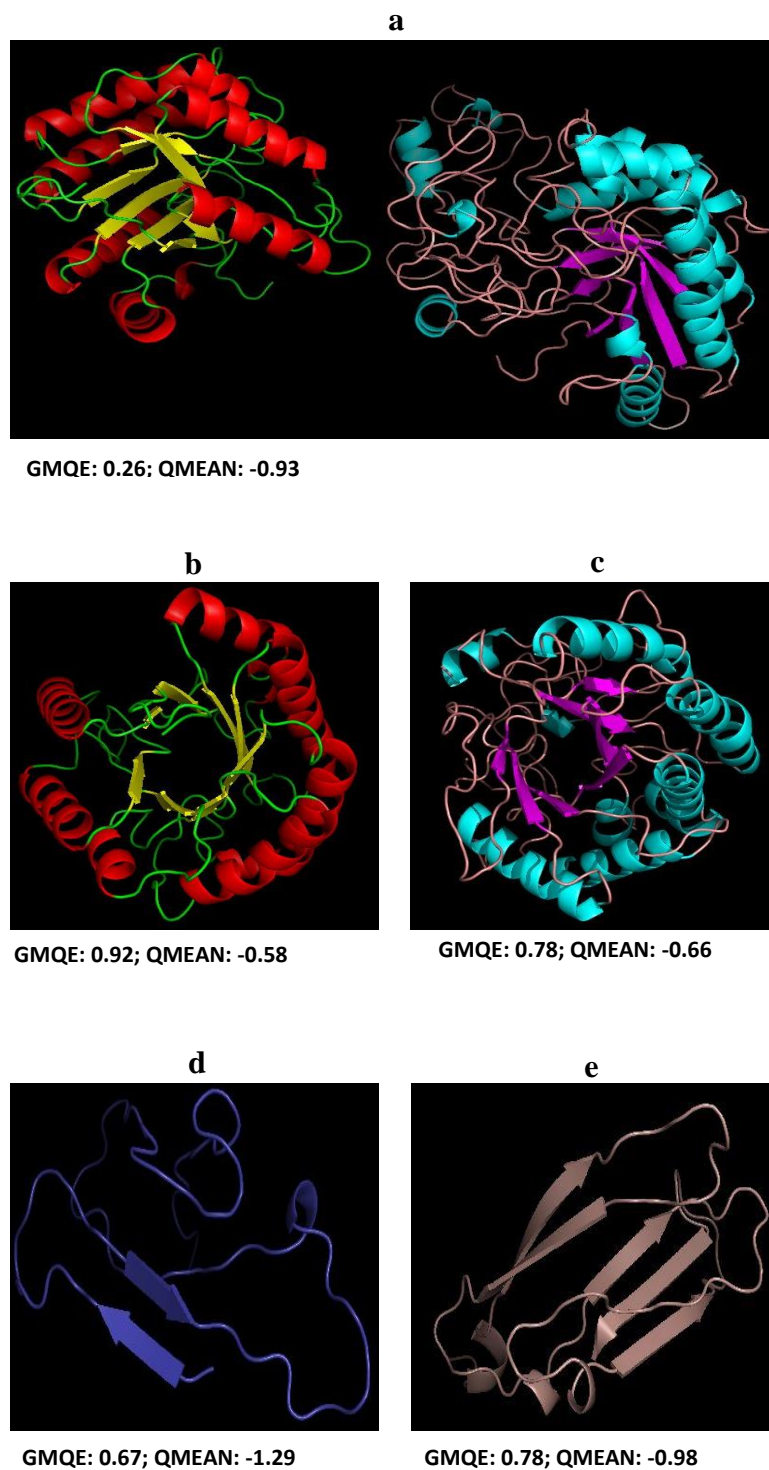


Figure 3.16: Homology model of *PmGH* and its corresponding catalytic domains. (a) *PmGH* homology model; (b) GH5 catalytic domain; (c) GH6 catalytic domain; (d) CBM3a; (e) CBM 3b. Global model quality estimation (GMQE) and QMEAN are indicated below each figure (a-e).*

*GMQE scores closer to 1 indicates a high degree of accuracy and similarity.

When modelled independently, homology models of the catalytic domains, GH5 (Figure 3.16b) and GH6 (Figure 3.16c), showed a higher level of accuracy as indicated by the higher GMQE and QMEAN values. Figure 3.16b shows the conserved TIM barrel (β/α)₈ catalytic fold typical of GH5 catalytic domains. A distorted seven stranded β/α conserved catalytic fold was observed in the GH6 catalytic domain (Figure 3.16c). A β -sandwich characteristic of CBM3 was observed in both models of the CBMs found in *PmGH* (Figure 3.16c-d).

Table 3.2 Summary of homology modelling data using SwissModel and the homologs used for the native enzyme and its composite domains.

Protein	Sequence identity	Sequence coverage	PDB Id.	GMQE*	QMEAN	Fold
<i>PmGH</i>	—	—	1LF1; 4AVN	0.26	-0.93	—
GH5	56%	100%	1LF1	0.92	-0.58	(β/α) ₈
GH6	55%	93%	4AVN	0.78	-0.66	Distorted 7 stranded β/α
CBM3a	33%	100%	2WO4	0.67	-1.29	β -sandwich
CBM3b	43%	87%	1NBC	0.78	-0.98	β -sandwich

*GMQE (Global Quality Model Estimation) scores closer to 1 indicates a high degree of accuracy and similarity.

3.6. Screening and optimisation of protein crystallisation conditions

Table 3.3 shows the crystallisation screening conditions that yielded protein crystals of *PmGH* over different time periods at 18°C. Magnesium chloride hexahydrate, buffers ranging from pH 6.5 to 6.8 and polyethylene glycol (PEG) provided general working conditions for *PmGH* crystallisation. The protein crystallisation conditions above (A-E) resulted in the protein crystals shown in Figure 3.17 below.

Table 3.3: Screening conditions and incubation times for crystallisation of *PmGH*.

Crystallisation condition	Salt	Buffer	Precipitant	Incubation time
A	0.01 M Zinc sulphate	0.1 M MES (pH 6.5)	25% PEG MME	10 days
B	0.2 M Magnesium acetate tetrahydrate	Sodium cacodylate trihydrate (pH 6.5)	20% (w/v) PEG 8000	7 days
C	0.2 M Sodium citrate tribasic dehydrate	0.2 M Tris hydrochloride (pH 8.5)	30% (w/v) PEG 400	7 days
D	0.2 M Magnesium chloride hexahydrate	0.1 M MES (pH 6.8)	10% (w/v) PEG 10 000	12 months
E	0.2 M Magnesium	–	15% PEG 10 000	24 days

	chloride hexahydrate			
--	-------------------------	--	--	--

PEG: Polyethylene glycol

Figure 3.17a shows the hollow, fragmented *PmGH* crystals grown over a 10 day period under crystallisation condition A. Minute protein crystals and a brown precipitate were observed under crystallisation condition B as shown in Figure 3.17b. Crystallisation condition C resulted in the growth of rare quasi-crystals as shown in Figure 3.17c.

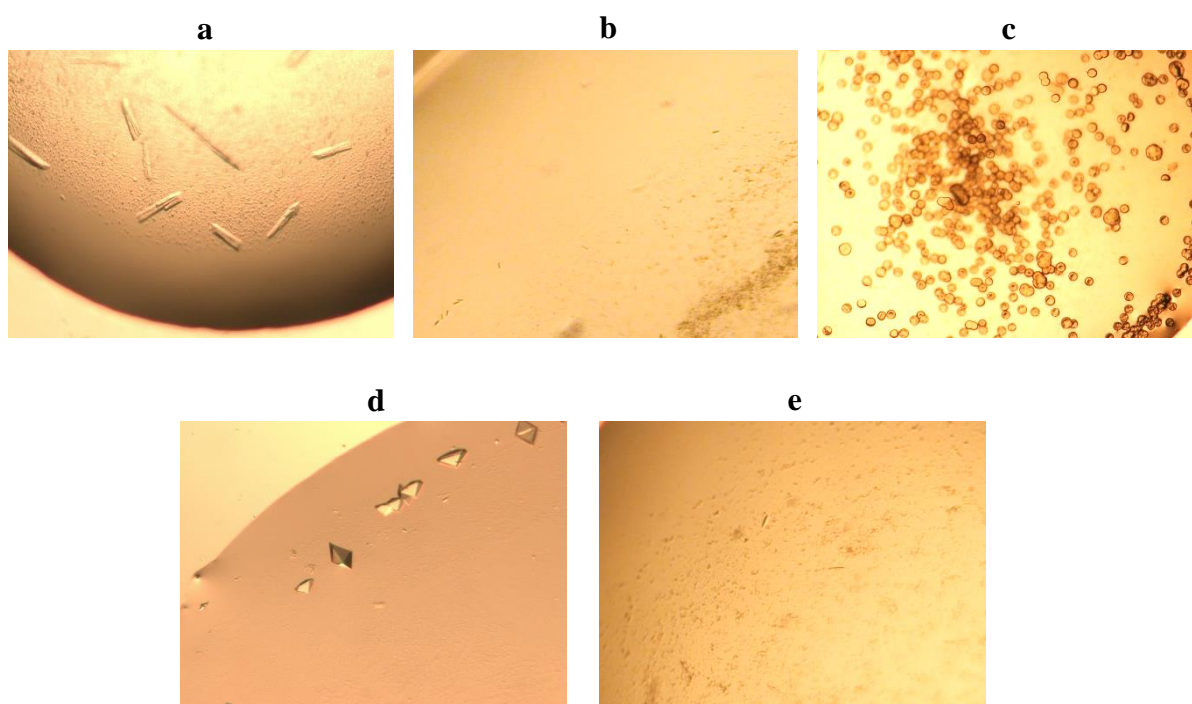


Figure 3.17: Screening conditions showing potential conditions for *PmGH* crystallisation. * (a) Hollow, fragmented *PmGH* protein crystals grown under condition A; (b) Brown precipitate and minute protein crystals grown under condition B; (c) Rare quasi-crystals grown in condition C; (d) Protein crystals grown in condition D over 12 months; (e) Minute protein crystals in the presence of a brown precipitate (condition E).

*Crystallisation conditions A-E represented in Table 3.3.

Over a 12 month period crystallisation condition D resulted in three-dimensional (3-D) protein crystals (Figure 3.17d). Figure e shows the minute protein crystals grown over a 24 day period under crystallisation E.

Protein crystallisation conditions (A-E) were extensively further optimised in order to identify suitable conditions for *PmGH* crystallisation. Crystallisation conditions D and E proved to be successful upon further optimisation, resulting in protein crystals suitable for X-ray diffraction. A hanging drop seeded with a 10^{-1} dilution of protein crystals harvested from the crystallization shown in Figure 3.14d resulted in significantly larger, better defined 3-D protein crystals after 6 months at 18°C (Figure 3.18).

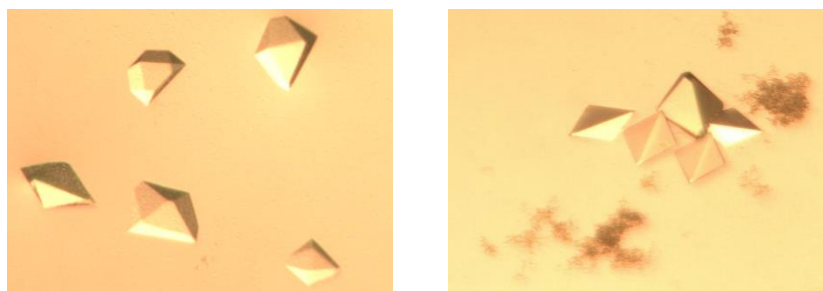


Figure 3.18: Improvement of protein crystals grown in condition D by bead seeding with a 10^{-1} dilution.*#

* All protein crystals grown at 18°C ; hanging drops equilibrated for one hour prior to seeding.

$3\text{ mg}\cdot\text{ml}^{-1}$ protein used in crystallisation suspended in 20 mM sodium citrate buffer (pH 6.5).

The protein crystallisation condition E was further optimised by varying magnesium chloride hexahydrate ($\text{MgCl}_2\cdot 6\text{H}_2\text{O}$) and PEG 10 000 concentration (Figure 3.19). Magnesium chloride hexahydrate concentrations ranging from 0.025 M to 0.2 M were tested. Polyethylene glycol concentrations ranging from 6-30% (w/v) were tested in combination with the optimal $\text{MgCl}_2\cdot 6\text{H}_2\text{O}$ concentration. At 0.15 M $\text{MgCl}_2\cdot 6\text{H}_2\text{O}$, an increase in the number of protein

crystals was observed. A lower concentration of PEG 10 000 (8%; w/v) and optimal $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ concentration (0.15 M), resulted in the growth of protein clusters. The protein clusters were subsequently harvested and used in further optimisation experiments by bead seeding.

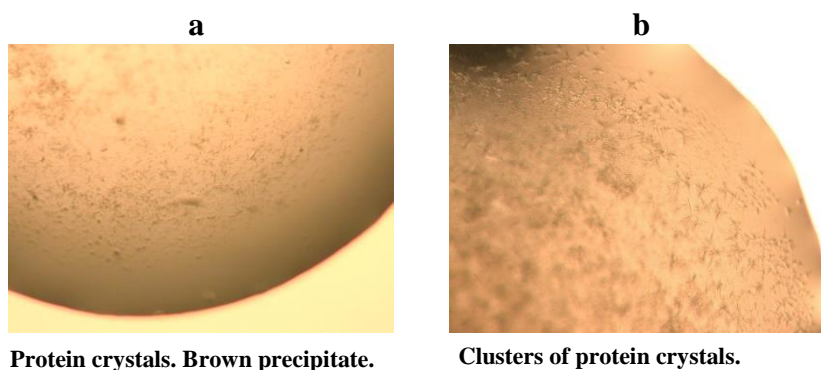


Figure 3.19: Optimisation of crystallisation condition E by varying the concentration of $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ and PEG 10 000. *#⁺ (a) Optimal salt concentration of 0.15 M $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$; (b) Optimal precipitant concentration of 8% PEG 10 000 in combination with 0.15 M $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$.

* A concentration range of 0.025-0.2 M $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ was tested.

A concentration range of 6-30% (w/v) of PEG 10 000 was tested.

⁺8 $\text{mg} \cdot \text{ml}^{-1}$ protein used in crystallisation suspended in 20 mM phosphate buffer (pH 6).

Protein crystallisation conditions were further optimised by bead seeding. Serial dilutions ranging from 10^{-1} to 10^{-10} were prepared in each round of bead seeding. All hanging drops were equilibrated for one hour prior to seeding and incubated at 18°C . Figure 3.20a-d below provides a photographic summary of the first round of bead seeding. Initial crystallisation conditions yielded a dense brown precipitate with micro-clusters of protein crystals after 4 days as shown in Figure 3.19b above. When used as a seed (Figure 3.19b) for

the trial shown in Figure 3.20a, the brown precipitate was reduced, resulting in a dense cluster of protein crystals after 4 days. Seeding with a 10^{-2} dilution increased protein crystal size and reduced clustered growth of the protein crystals after 4 days (Figure 3.20b). However, an increase in the brown precipitate was observed. Further seeding with 10^{-3} reduced clustering and increased dispersion of the protein crystals after a 4 day incubation period (Figure 3.20c). Significantly larger and more dispersed protein crystal were observed (Figure 3.20d), Figure 3.20a. Protein crystals shown in Figure 3.20e were seeded with a 10^{-4} dilution and were observed after 4 days.

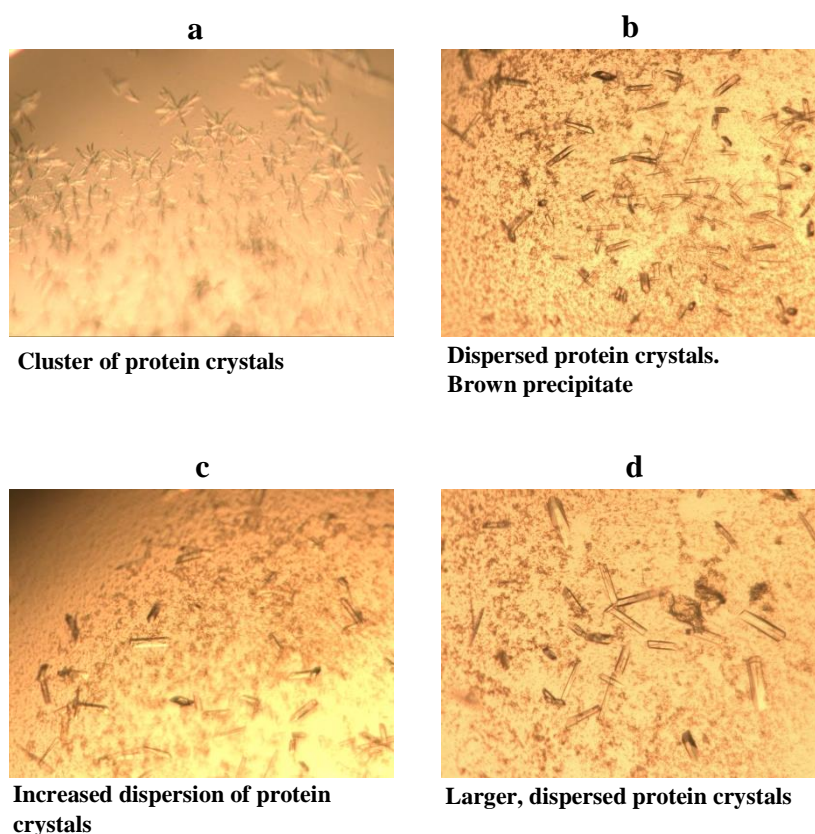


Figure 3.20: Optimisation of *PmGH* protein crystals by bead seeding.^{*#} (a) Hanging drop seeded with 10^{-1} dilution of *PmGH* protein crystals; (b) Hanging drop of *PmGH* seeded with 10^{-2} dilution ; (c) Hanging drop seeded with 10^{-3} dilution; (d) *PmGH* hanging drop seeded with 10^{-4} dilution.

*All protein crystals grown at 18°C; hanging drops equilibrated for one hour prior to seeding.

8 mg.ml⁻¹ protein used in crystallisation suspended in 20 mM phosphate buffer (pH 6).

The protein crystals shown in Figure 3.17e above were further used as a seed for the crystallization trial shown in Figure 3.21a-b. Seeding with a 10⁻¹ dilution resulted in the dispersed growth of the protein crystals.

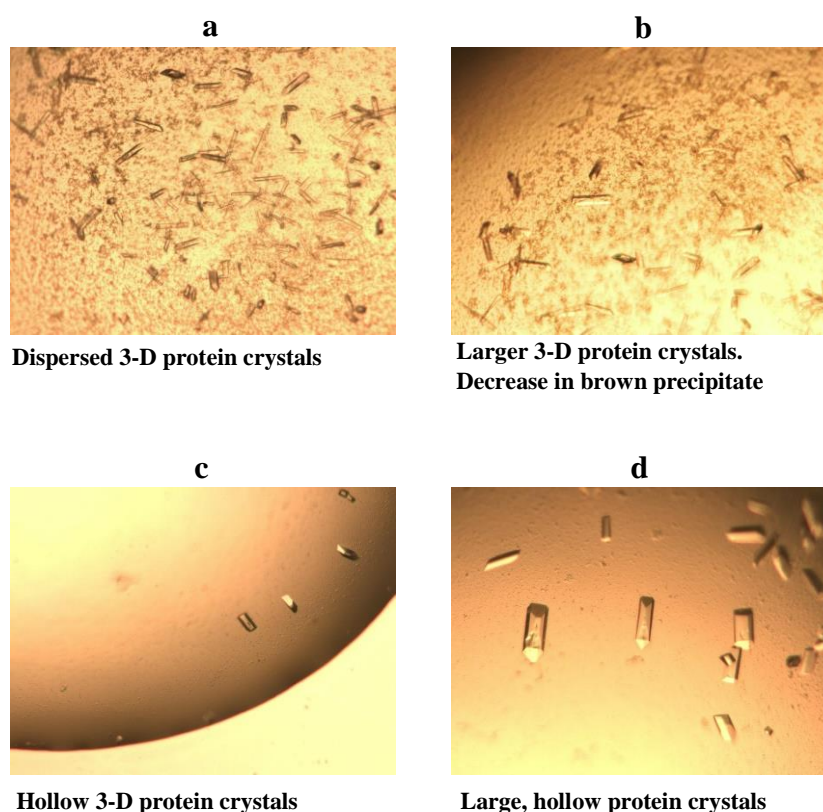


Figure 3.21: Optimisation of *PmGH* crystals by a second round of bead seeding.^{##} (a) Protein crystals grown by seeding with 10⁻¹ dilution; (b) Hanging drop seeded with 10⁻² dilution; (c) Protein crystals seeded with 10⁻³ dilution grown after 20 days; (d) Hollow *PmGH* crystals seeded with 10⁻³ dilution grown after 23 days.

*All protein crystals grown at 18°C; hanging drops equilibrated for one hour prior to seeding.

#8 mg.ml⁻¹ protein used in crystallisation suspended in 20 mM phosphate buffer (pH 6).

The appearance of three-dimensional protein crystals was clearly observed even in the presence of the brown precipitate (Figure 3.21a). An increase in the size of the protein crystals was observed (Figure 3.21b) with a decrease in the appearance of the brown precipitate. Protein crystals shown in Figure 3.21a-b appeared after 5 days. Figure 3.21c shows a hanging drop seeded with a 10^{-3} dilution. After a 20 day incubation period protein crystals and possible nucleation sites were observed. Further incubation for an additional five days resulted in several larger, hollow, three-dimensional protein crystals (Figure 3.21e).

A third and final round of bead seeding, together with a lower incubation temperature (12°C), resulted in protein crystals suitable for diffraction (Figure 3.22). Lower incubation temperatures resulted in large, fully formed three-dimensional protein crystals. However, differences in the shapes of the protein crystals were observed.

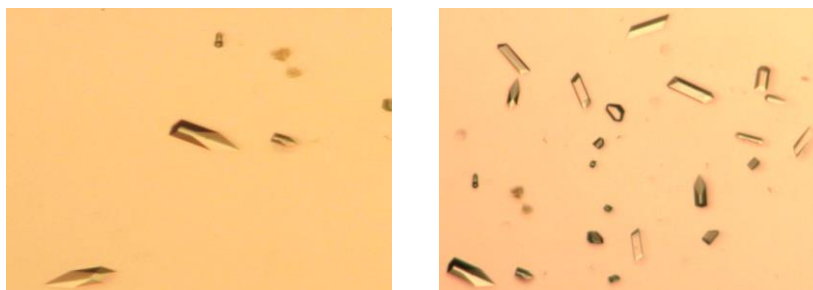


Figure 3.22: Final *PmGH* crystals seeded with a 10^{-3} dilution and incubated at 12°C for 7 days. * #

* Hanging drops equilibrated for one hour prior to seeding.

8 mg.ml^{-1} protein used in crystallisation suspended in 20 mM phosphate buffer (pH 6).

3.7. X-ray diffraction and data analysis

Figure 3.23 below provides details of the highest quality X-ray diffraction data that could be collected from *PmGH* protein crystals. Table 3.4 provides the corresponding data-collection statistics for the highest quality X-ray diffraction data set (see also Appendix D). The orthorhombic ($P2_12_12$) *PmGH* protein crystals diffracted to a resolution of 2.53-2.57 Å and a data set of 99.96% completeness was collected. Molecular replacement using Phenix proved to be unsuccessful after extensive analysis of all X-ray diffraction data collected at the ESRF (Grenoble, France).

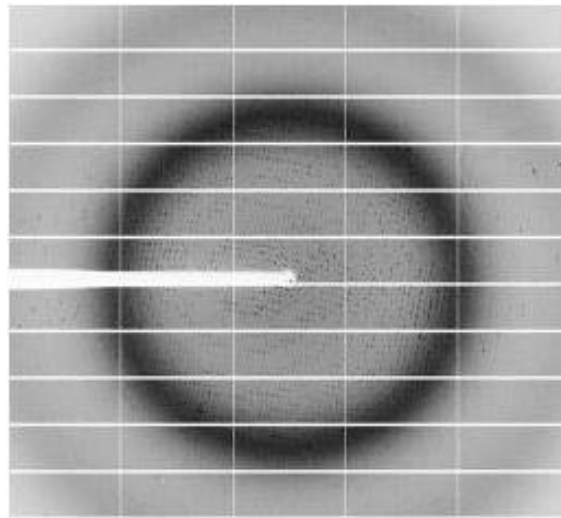


Figure 3.23: X-ray diffraction image of the highest quality data collected (2.53 Å) from *PmGH* protein crystals at the ESRF.

Table 3.4: Data-collection statistics of the highest quality X-ray diffraction data obtained from *PmGH* protein crystals at the ESRF.

Parameters	Values
Wavelength (Å)	0.976251
Space group	P2 ₁ 2 ₁ 2
Resolution (Å)	2.53 (2.57-2.53)
Unit cell dimensions	
a (Å)	110.309
b (Å)	116.359
c (Å)	92.821
α (°)	90.000
β (°)	90.000
γ (°)	90.000
Total number of observations	15 099 (14 475)
Total number of reflections	2 218 (1 950)
Completeness (%)	99.96
Multiplicity	7.27
I/sigma	3.87
Rmerge (I)	0.6009

Chapter 4

Discussion

The evolution of plants has resulted in the co-evolution of glycoside hydrolase (Scheller and Ulvskov, 2010). An increase in the complexity of plant polysaccharides over evolutionary time have resulted in an array of glycoside hydrolases and carbohydrate binding modules as shown in the CAZy database (Ebringerová and Heinze, 1999; Minic, 2008). In order to target a broader range of polysaccharides, microbial GHs are frequently organised into multi-modular enzymes. These multi-modular enzymes often combine one or more GH, CBM and/or accessory enzymes (e.g., carbohydrate esterases) (Varnai *et al.*, 2014). Therefore, the arrangement and type of modules are often indirectly dependent on the type of plant polysaccharides available in the microbial environment (Henrissat, 1991).

Multi-modular GHs are typically high molecular weight enzymes as shown in the present study. The current study demonstrated that a ~3000 bp gene of *Paenibacillus mucilaginosus* encoded for *PmGH*, a multi-modular, multi-functional and high molecular weight (130 kDa) glycoside hydrolase. The native enzyme (*PmGH*), which is composed of two catalytic domains, GH5 and GH6, displayed a broad range of hydrolytic activities. Cellulase, mannanase and xylanase activities were detected in *PmGH* and were attributed to the presence of the GH5 catalytic domain. Enzymes belonging to the GH5 family are known to possess more than one hydrolytic activity (Aspeborg *et al.*, 2012). In contrast, GH6 enzymes are thought to be strict cellulose hydrolases (Mertz *et al.*, 2005), consistent with the functional analysis in the present study. Functional differences across the three enzymes (*PmGH*, GH5 and GH6) were also observed across different pH and temperature conditions. The native enzyme (*PmGH*) and its composite catalytic domains (GH5 and GH6) displayed maximal activity at pH 6 and pH 5, respectively. A higher temperature optima (70°C) was observed for both catalytic

domains in comparison to *PmGH* (60°C) further highlighting differences in optimum conditions for substrate hydrolysis. Several reports support these findings as numerous thermophilic GH5 and GH6 enzymes with maximal activity at acidic pH have been reported (Ding *et al.*, 2018; Kim *et al.*, 2011; van Dyk *et al.*, 2010; Yin *et al.*, 2012). Fusco *et al.* reported a bifunctional *Dictyoglomus turgidum* endoglucanase/endomannanase belonging to the GH5 family with maximal activity at pH 5.4 and 70°C. Similarly, Zhou *et al.* (2017) reported a *Chaetomium thermophilum* cellulase (GH6) with maximal activity at pH 5 and 70°C. Overall, functional stability profiles indicated that all enzymes were functionally stable at pH 4-6 and 40-70°C. All enzymes retained 100% of activities at 40°C and 50°C after 2 hours incubation. The high thermostability at 60°C and 70°C for all enzymes confirming the thermophilic nature of the enzymes as thermophilic GHs typically function at temperatures above 50°C. At optimum pH, *PmGH* (pH 6), GH5 (pH 5) and GH6 (pH 6) retained ~85%, ~81% and ~62% of activities after 2 hours, but were all functionally unstable under alkaline conditions (pH 8-9). This is in agreement with numerous studies that have identified and characterised thermostable and pH stable microbial GHs (Bischoff *et al.*, 2006; Cha *et al.*, 2018; Jia *et al.*, 2016; Peng *et al.*, 2016). Overall functional analysis showed changes in pH and temperature for all enzymes. It is important to note that the net charge of a protein; *i.e.*, isoelectric point (pI), influences the optimum functional pH of any protein (Nehete *et al.*, 2013; P Bryan, 1978). A decrease in the theoretical pI of the catalytic domains (~5) when compared to the native enzyme (~5.9) resulted in a decrease in pH optima. As a result, a change in the net charge affects the intra-molecular and inter-domain interactions of the enzyme which inherently affects enzyme function *i.e.* pH and temperature optima and stability as seen in the present study. Furthermore, a number of structural elements are known to confer functional stability, particularly thermostability. Increased intramolecular interactions such as hydrogen bonding, salt bridges and the number of charged surface residues are known to contribute to the thermostability of GHs (Li *et al.*,

2005; Vieille and Zeikus, 2001). These inter- and intra-domain interactions differ across various GHs and may be increased or decreased in the individual catalytic modules and/or multi-modular GHs. As a consequence, the functional stability of the individual catalytic domains (GH5 and GH6) and the multi-modular enzyme (*PmGH*) may differ as evidenced by functional analysis data. Furthermore, functional analysis also indicated that the catalytic domains (GH5 and GH6) can function both independently and as part of the multi-modular enzyme, *PmGH*. These differences in enzyme function in the native enzyme (*PmGH*) and its composite catalytic domains (GH5 and GH6) warranted further investigation into the structural elements that dictate these functional properties. Functional analysis data further highlighted the importance of fully understanding the intermolecular and intramolecular interactions that are critical to enzyme function and stability. X-ray crystallography is considered the gold standard of protein structure determination and in the present study we made use of this technology in order to determine the crystal structure of the novel multi-modular glycoside hydrolase, *PmGH*. In an attempt to understand the structural properties of *PmGH* that contribute to its stability and catalytic specificity, a search for structurally similar proteins in the PDB was carried out. Homology modelling using the SWISS-MODEL pipeline, which searches for structurally similar GHs in the PDB and provides a model that best represents the protein of interest (*PmGH*), was used. Based on the sequence identity, coverage, and GMQE and QMEAN values, it was possible to determine the accuracy and novelty of *PmGH*. Homology search for structural homologs of the full length native enzyme, *PmGH*, confirmed that *PmGH* is novel since only an incomplete homology model composed only of the GH5 and GH6 catalytic domains was obtained. Although the inter-molecular interactions between GH5 and GH6 could not be identified as both modules were modelled independently within the native enzyme, this served as the first step in elucidating the structure of *PmGH*. The initial phases of homology modelling also revealed that the composite CBMs were not covered in the

model together with the adjoining linker regions. Homology modelling provided further evidence that glycoside hydrolases occur as independent modules or as part of a multi-modular protein with other catalytic and/or non-catalytic modules (Naumoff, 2011) as the modules were modelled independently. Consequently, multiple combinations of these modules can occur in nature, resulting in structurally and functionally novel proteins such as *PmGH*. In addition, homology modelling data further highlighted the inaccuracy of the homology model which was confirmed by the low GMQE score (0.26), as an accurate homology model should have a GMQE score as close to 1 as possible (Benkert *et al.*, 2011). Functional properties confirmed the complexity of *PmGH*. Consequently, a search for structural homologs of the individual composite modules (GH5, GH6, CBM 3a and CBM 3b) was carried out in order to establish whether these domains were structurally novel. The GH5 catalytic domain displayed a $(\beta/\alpha)_8$ catalytic fold, commonly referred to as a TIM barrel, which is characterised by the 8 parallel (β/α) subunits linked by hydrogen bonds to form a cylindrical core. Several GH5 crystal structures have been determined to date all of which correspond with a conserved TIM barrel (Davies *et al.*, 1998; Ducros *et al.*, 1995; Gloster *et al.*, 2007) as reported in the present study. A distorted seven stranded (β/α) catalytic fold exclusive to GH6 enzymes, was observed for the GH6 catalytic domain was also reported in the present study. Typically, the active site of GH6 exoglucanases and endoglucanase are often located in the tunnel or on the cleft, respectively (Sandgren *et al.*, 2013). As opposed to the catalytic domains, a β -sandwich fold characteristic of family 3 CBMs was reported for both CBM 3 domains. Overall, the high GMQE scores for all domain homology models indicated that the domains were not structurally novel and are highly similar to protein crystal structures that have already been deposited in the Protein Data Bank. Protein crystal structure determination by molecular replacement is most effective when a single suitable structural homolog of the protein of interest is available (Scapin, 2013). Furthermore, phasing by molecular replacement typically requires a sequence

identity of 25% or more with the unknown protein (Taylor, 2003). All homology models of the composite domains (GH5, GH6, CBM3a, CBM3b) of the full length protein, *PmGH*, had a sequence identity greater than 30%. As a result, the homology models of GH5, GH6, CBM3a and CBM3b were used as structural homologs in the molecular replacement of *PmGH*. However, the lack of a full length structural homolog further highlighted the need to determine the crystal structure of this novel enzyme (*PmGH*) in an attempt to further understand the molecular interactions that contribute to its functionality.

Crystallisation of high molecular weight proteins is typically challenging and often requires extensive optimisation. However, in the present study we successfully crystallised the native enzyme, *PmGH*, through a process of sequential optimisation. Several crystallisation conditions such as the amount and type of salt, precipitant and buffer; incubation temperature; protein concentration and bead seeding conditions were varied in order to obtain suitable crystallisation conditions for *PmGH*. As evidenced by the crystallisation data, changes in the crystallisation conditions significantly affected the nature and quality of the protein crystals. As a result, hollow protein crystals were obtained in initial optimisation experiments, although further optimisation trials resulted in the generation of complete three-dimensional protein crystals suitable for X-ray diffraction. This also indicated that numerous variables could be altered in any protein crystallisation experiment and as a result, the success rate may vary for different proteins. In an attempt to elucidate the crystal structure of *PmGH* and to provide structural data on multi-modular GHs, *PmGH* protein crystals were diffracted. Several crystallisation screening conditions were tested to determine the ideal conditions for *PmGH* crystallisation. Following extensive optimisation of protein crystallisation conditions, and the diffraction data was collected at the ESRF (beamline MASSIF/ID30A-1 and FIP/ID30A-3) on two separate occasions. Overall 90% of the *PmGH* protein crystals tested did not diffract. However, of the remaining 10% of protein crystals that diffracted medium to poor quality

diffraction data with resolutions ranging from 2.53-6 Å were collected. Figure 3.20 shows the diffraction pattern and dataset parameters of the highest quality diffraction data set (2.53-2.57Å) that could be collected. In the absence of a full length structural homolog, the structural homologs of GH5, GH6, CBM3a and CBM3b in phasing experiments by MR. However, MR proved to be unsuccessful after extensive analysis of all the diffraction data collected. Phasing experiments by MR are ultimately dependent on the quality of the diffraction data. A number of parameters such as resolution, completeness, multiplicity, signal-to-noise ratio can be used to assess the quality of the diffraction data (Evans and McCoy, 2008; Taylor, 2010). Ultimately, the abovementioned parameters are dependent on the nature of the protein crystal; *i.e.*, unit cell size and protein crystal order (Dobrianov *et al.*, 1999). The size of the unit cell dictates the number of unit cells in a crystal lattice which ultimately determines the scattering (intensity spots in diffraction pattern) and the volume of the protein. Conversely, the number of molecules in the asymmetric unit, protein crystal symmetry and crystal packing dictate the size of the unit cell (McPherson, 2004). Protein crystal order also affects scattering which in turn is dependent on how identical each unit cell is. Highly ordered protein crystals diffract to a higher resolution compared to disorder protein crystals. Linker regions play an important role in protein flexibility and contribute to the dynamic nature of enzymes which ultimately affects the tightness of the crystal packing (Reddy Chichili *et al.*, 2013; Ruiz *et al.*, 2016). The full length enzyme (*PmGH*) reported in this study has several linker regions adjoining the catalytic modules and CBMs. This multi-modular glycoside hydrolase targets high molecular weight polysaccharides; *i.e.*, cellulose, mannan and xylan. In order to hydrolyse these polysaccharides, *PmGH* must spatially accommodate the substrate in the active site. The dynamic nature of such proteins often results in structural translations and rotations, which could potentially affect protein structure determination (Ma *et al.*, 2011). As a result, the linkers may need to be stabilised or the protein structure of the composite domains may need to be determined

independently. However, it is highly likely that the full-length *PmGH* protein is not a tightly folded structure (due to the presence of the unordered linker regions), and that the partially disordered protein structure is responsible for the failure to yield coherent x-ray diffraction data.

Furthermore, the nature of the crystal packing indirectly determines the size of the unit cell and the number unit cells in the crystal lattice. The crystal packing of our protein could potentially have contributed to the lack of suitable diffraction data from the *PmGH* protein crystals. Furthermore, the unit cell properties of *PmGH* which could not be determined, could potentially have affected diffraction data collection and phasing experiments by MR. These factors have contributed to the lack of structural data on our protein and the overall lack of structural information on multi-modular proteins. However, given the various limiting factors of protein structure determination, a few well-resourced researchers have been able to overcome these limitations and were able to determine the protein structure of multi-modular enzymes.

Previous studies have reported that the crystal structure of a 140kDa, multi-modular *Paenibacillus* sp. 598K α -1,6-glucosyltransferase was determined by MAD at a resolution of 1.95-2.4Å (Fujimoto *et al.*, 2017). The enzyme was derivatised with terbium chloride (TbCl₂) and was composed of an N-terminal (β/α)₈ barrel catalytic domain and two C terminal catalytic domains that closely resemble GH31 enzymes. Three CBMs (two CBM35; CBM61) with a β -jelly roll fold were compactly located around the catalytic domains. This study also indicated that alternative phasing methods such as MAD and the use of rare heavy atoms can prove to be ideal for structure determination of a novel multi-modular enzyme.

Determination of the crystal structures of the GH5, GH6, CBM3a and CBM3b modules which make up the full length enzyme (*PmGH*) could provide an alternative strategy in the

determining the protein crystal structure of the full-length enzyme. Once determined, these crystal structures could be used as structural homologs in MR of *PmGH*. If successful, this approach could provide a complete high resolution, protein structure of the full-length enzyme. This approach could provide insights into the inter- and intra-molecular interactions of *PmGH* and could be used to visualise the net interactions that contribute to the functional properties of *PmGH*. However, it is important to note that this approach could yield a low resolution or incomplete protein structure. The incomplete protein structure could be a direct result of the absence of the linker regions in the structural homologs, which connect the modules to form the full length protein.

Future studies could also involve identification of the disordered regions in the enzyme. Once identified, the disordered regions could be deleted by genetic engineering, potentially resulting in an enzyme with a lower degree of flexibility. The crystal structure of the engineered enzyme could subsequently be determined by X-ray diffraction. Alternatively, cryogenic electron microscopy (cryo-EM) might be employed to determine the crystal structure of *PmGH*. Cryo-EM does not typically provide protein structures at atomic resolution; however, advances in the instrumentation and software used in cryo-EM technology have led to this method being considered as a viable alternative strategy to X-ray diffraction (Nogales, 2016; Shoemaker and Ando, 2018). The use of direct electron-detection cameras (direct detectors) as opposed to the use of film traditionally used in cryo-EM data collection now allow for the high-resolution signal to be preserved (Brilot *et al.*, 2012). Software packages such as RELION (Scheres, 2012) have since been designed to address conformational and compositional heterogeneity of cryo-EM samples which previously served as a bottleneck in data collection. Bartesaghi *et al.* (2014) determined the structure of a recombinant β -galactosidase complex (~465 kDa) at an average resolution of 3.2 Å by cryo-EM, while the authors were able to further improve the resolution of the β -galactosidase (~465 kDa) to 2.2 Å (Bartesaghi *et al.*,

2015) using cryo-EM. Improvements in cryo-EM imaging and imaging processing technologies together with specimens of adequate quality allowed Bartesaghi *et al.* (2014; 2015) to achieve the first high resolution cryo-EM protein structure of a glycoside hydrolase thus providing evidence that cryo-EM may be used as an alternative strategy for protein structure determination.

Overall we have presented the research findings of a novel glycoside hydrolase and its composite domains. The native enzyme and the GH5 catalytic domain are capable of hydrolysing the major plant polymers *i.e.* cellulose and hemicellulose (xylan and mannan). Due to the thermophilic nature of all enzymes and functionality at pH 4-6, these enzymes have potential for application in bio-ethanol production and the extraction of fruit juices. In an attempt to understand the molecular interactions which govern *PmGH* functionality we successfully crystallised this multi-modular GH. We have provided a detailed account of the intensive and numerous optimisation steps required to successfully crystallise and diffract protein crystals of a multi-modular glycoside hydrolase. Furthermore, we have provided the first evidence of successful crystallisation of a multi-modular *Paenibacillus* GH. Although challenging, this study has provided a strong starting point for the determination of the crystal structure of a multi-functional and multi-modular enzyme, *PmGH*. Inherently, the present study serves as a guide for future crystallographers working on enzymes of a similar nature.

Chapter 5

Concluding remarks and future studies

The *Paenibacillus mucilaginosus* GH presented in this study is a high molecular weight protein due to its multi-modular nature. The presence of two functional catalytic modules in *PmGH* contributed to its broad hydrolytic activities. The GH5 domain is responsible for the cellulase, xylanase and mannanase activities of *PmGH* with the GH6 domain providing additional cellulolytic activity. Functional analysis showed that *PmGH* is highly pH stable and thermostable. When sub-cloned and expressed independent of the native enzyme, both GH5 and GH6 are functional, pH stable and thermostable.

The present study also provided the first ever preliminary functional and crystallisation data for a *Paenibacillus mucilaginosus* GH. It is also one of the few studies that have provided evidence of the crystallisation of a multi-modular and multi-functional GH. Upon extensive screening and optimisation, *PmGH* crystals were obtained. Reports on the crystallisation and phasing of multi-modular GHs are often limited. As a result, a limited number of multi-modular GH crystal structures have been elucidated to date. Most multi-modular GHs reported have been elucidated by multi-wavelength anomalous dispersion and not by molecular replacement. The present study has highlighted the difficulties in crystallising and phasing multi-modular GHs. Furthermore, the difficulties experienced during data collection confirmed that not all protein crystals yield high quality data and not all protein crystals diffract. This has proven to be a major bottleneck and has contributed to the lack of existing crystal structures of multi-modular GHs.

In order to improve data collection in future studies, the composite catalytic domains of *PmGH* could be crystallised and diffraction data collected in order to solve the phases by molecular replacement. The crystal structures of the catalytic domains would subsequently be used as structural homologs in determining the crystal structure of *PmGH*. Alternatively, genetic engineering could be employed to delete disordered regions in the protein which affect protein folding. The crystal structure of the re-engineered enzyme could subsequently be elucidated using X-ray diffraction technology however, these structural changes may impact enzyme function. Though demanding, protein structure determination provides invaluable insights into macromolecular dynamics at each optimisation step thus expanding on existing functional, crystallographic and structural information.

References

- Abergel, C. (2013). Molecular replacement: tricks and treats. *Acta Crystallographica Section D Biological Crystallography* 69, 2167-2173.
- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography* 66, 213-221.
- Alzari, P.M., Souchon, H., and Dominguez, R. (1996). The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from *Clostridium thermocellum*. *Structure* 4, 265-275.
- Asherie, N. (2004). Protein crystallization and phase diagrams. *Methods* 34, 266-272.
- Aspeborg, H., Coutinho, P.M., Wang, Y., Brumer, H., and Henrissat, B. (2012). Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evolutionary Biology* 12, 186.
- Attigani, A., Sun, L., Wang, Q., Liu, Y., Bai, D., Li, S., and Huang, X. (2016). The crystal structure of the endoglucanase Cel10, a family 8 glycosyl hydrolase from *Klebsiella pneumoniae*. *Acta Crystallographica Section F Structural Biology Communications* 72, 870-876.
- Bailey, M.J., Biely, P., and Poutanen, K. (1992). Interlaboratory testing of methods for assay of xylanase activity. *Journal of Biotechnology* 23, 257-270.
- Bartesaghi, A., Matthies, D., Banerjee, S., Merk, A., and Subramaniam, S. (2014). Structure of β -galactosidase at 3.2 Å resolution obtained by cryo-electron microscopy. *PNAS*. 111: 11709-11714.
- Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X., Milne, J.L.S., and Subramaniam, S. (2015). 2.2 Å resolution cryo-EM structure of β -galactosidase in complex with a cell-permeant inhibitor. *Science*. 348: 1147-1151.
- Bastawde, K.B. (1992). Xylan structure, microbial xylanases, and their mode of action. *World Journal of Microbiology and Biotechnology* 8, 353-368.
- Bayer, E.A., Chanzy, H., Lamed, R., and Shoham, Y. (1998). Cellulose, cellulases and cellulosomes. *Current Opinion in Structural Biology* 8, 548-557.
- Béguin, P., and Lemaire, M. (1996). The cellulosome: an exocellular, multiprotein complex specialized in cellulose degradation. *Critical Reviews in Biochemistry and Molecular Biology* 31, 201-236.
- Behera, B.C., Behera, B.C., Sethi, B.K., Sethi, B.K., Mishra, R.R., Mishra, R.R., Dutta, S.K., Dutta, S.K., Thatoi, H.N., and Thatoi, H.N. (2017). Microbial cellulases - Diversity & biotechnology with reference to mangrove environment: a review. *Journal Genetic Engineering and Biotechnology* 15, 197-210.
- Benkert, P., Biasini, M., and Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27, 343-350.
- Bhat, M.K. (2000). Cellulases and related enzymes in biotechnology. *Biotechnology Advances* 18, 355-383.
- Biely, P. (2012). Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnology Advances* 30, 1575-1588.

- Bischoff, K.M., Rooney, A.P., Li, X.-L., Liu, S., and Hughes, S.R. (2006). Purification and characterization of a family 5 endoglucanase from a moderately thermophilic strain of *Bacillus licheniformis*. *Biotechnology Letters* 28, 1761-1765.
- Blessing, R. (2014). Phasing in crystallography: a modern perspective. By Carmelo Giacovazzo. IUCr Texts on Crystallography, No. 20. International Union of Crystallography/Oxford University Press, 2013. Pp. 432. Price GBP 65.00 (hardback). ISBN 978-0-19-968699-5. *Acta Crystallographica Section A* 70, 518-519.
- Blum, D.L., Kataeva, I.A., Li, X.L., and Ljungdahl, L.G. (2000). Feruloyl esterase activity of the *Clostridium thermocellum* cellulosome can be attributed to previously unknown domains of XynY and XynZ. *Journal of Bacteriology* 182, 1346-1351.
- Blumer-Schuette, S.E., Kataeva, I., Westpheling, J., Adams, M.W., and Kelly, R.M. (2008). Extremely thermophilic microorganisms for biomass conversion: status and prospects. *Current Opinion in Biotechnology* 19, 210-217.
- Boggon, T.J., and Shapiro, L. (2000). Screening for phasing atoms in protein crystallography. *Structure* 8, R143-149.
- Boraston, A.B., Bolam, D.N., Gilbert, H.J., and Davies, G.J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochemical Journal* 382, 769-781.
- Brilot, A.F., Chen, J.Z., Cheng, A., Pan, J., Harrison, S.C., Potter, C.S., Carragher, B., Henderson, R., and Grigorieff, N. (2012) Beam-induced motion of vitrified specimen on holey carbon film. *Journal of Structural Biology* 177, 630-637.
- Bujacz, A., Jedrzejczak-Krzepkowska, M., Bielecki, S., Redzynia, I., and Bujacz, G. (2011). Crystal structures of the apo form of β -fructofuranosidase from *Bifidobacterium longum* and its complex with fructose. *The FEBS Journal* 278, 1728-1744.
- Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* 37, D233-238.
- Cha, J.H., Yoon, J.J., and Cha, C.J. (2018). Functional characterization of a thermostable endoglucanase belonging to glycoside hydrolase family 45 from *Fomitopsis palustris*. *Applied Microbiology and Biotechnology* doi: 10.1007/s00253-018-9075-5.
- Chandel, A.K., Chandrasekhar, G., Silva, M.B., and Silvério da Silva, S. (2012). The realm of cellulases in biorefinery development. *Critical Reviews Biotechnology* 32, 187-202.
- Chatterjee, S.K. (2010). X-ray diffraction: Its theory and applications., 2nd edn (New Dehli: PHI Learning Pvt. Ltd.).
- Chauhan, P.S., and Gupta, N. (2017). Insight into microbial mannosidases: a review. *Critical Reviews Biotechnology* 37, 190-201.
- Chauhan, P.S., Puri, N., Sharma, P., and Gupta, N. (2012). Mannanases: microbial sources, production, properties and potential biotechnological applications. *Applied Microbiology and Biotechnology* 93, 1817-1830.
- Chávez, R., Bull, P., and Eyzaguirre, J. (2006). The xylanolytic enzyme system from the genus *Penicillium*. *Journal of Biotechnology* 123, 413-433.
- Collins, T., Gerday, C., and Feller, G. (2005). Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiology Reviews* 29, 3-23.

- Costa, M., Fernandes, V.O., Ribeiro, T., Serrano, L., Cardoso, V., Santos, H., Lordelo, M., Ferreira, L.M., and Fontes, C.M. (2014). Construction of GH16 β -glucanase mini-cellulosomes to improve the nutritive value of barley-based diets for broilers. *Journal of Agricultural and Food Chemistry* *62*, 7496-7506.
- Davies, G.J., Dauter, M., Brzozowski, A.M., Bjørnvad, M.E., Andersen, K.V., and Schülein, M. (1998). Structure of the *Bacillus agaradherans* Family 5 Endoglucanase at 1.6 Å and its cellobiose complex at 2.0 Å resolution. *Biochemistry* *37*, 1926-1932.
- Davies, G.J., and Henrissat, B. (2002). Structural enzymology of carbohydrate-active enzymes: implications for the post-genomic era. *Biochemical Society Transactions* *30*, 291-297.
- Davis, A.M., Teague, S.J., and Kleywegt, G.J. (2003). Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angewandte Chemie International Edition in English* *42*, 2718-2736.
- De Lano, W.L. (2002). <http://www.pymol.org>.
- de Sanctis, D., Inácio, J.M., Lindley, P.F., de Sá-Nogueira, I., and Bento, I. (2010). New evidence for the role of calcium in the glycosidase reaction of GH43 arabinanases. *The FEBS Journal* *277*, 4562-4574.
- Dhawan, S., and Kaur, J. (2007). Microbial mannanases: an overview of production and applications. *Critical Reviews in Biotechnology* *27*, 197-216.
- Dias, F.M., Vincent, F., Pell, G., Prates, J.A., Centeno, M.S., Tailford, L.E., Ferreira, L.M., Fontes, C.M., Davies, G.J., and Gilbert, H.J. (2004). Insights into the molecular determinants of substrate specificity in glycoside hydrolase family 5 revealed by the crystal structure and kinetics of *Cellvibrio mixtus* mannosidase 5A. *The Journal of Biological Chemistry* *279*, 25517-25526.
- Ding, C., Li, M., and Hu, Y. (2018). High-activity production of xylanase by *Pichia stipitis*: purification, characterization, kinetic evaluation and xylooligosaccharides production. *International Journal of Biological Macromolecules* *117*, 72-77.
- Dobrianov, I., Caylor, C., Lemay, S.G., Finkelstein, K.D., and Thorne, R.E. (1999). X-ray diffraction studies of protein crystal disorder. *Journal of Crystal Growth* *196*, 511-523.
- Dondelinger, E., Aubry, N., Ben Chaabane, F., Cohen, C., Tayeb, J., and Rémond, C. (2016). Contrasted enzymatic cocktails reveal the importance of cellulases and hemicellulases activity ratios for the hydrolysis of cellulose in presence of xylans. *AMB Express* *6*, 24.
- Drenth, J. (2007). Principles of protein X-ray crystallography, 3rd edn (Netherlands: Springer).
- Ducros, V., Czjzek, M., Belaich, A., Gaudin, C., Fierobe, H., Belaich, J., Davies, G.J. and Haser, R. (1995). Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5. *Structure* *3*, 939-949.
- Dutta, T., Sahoo, R., Sengupta, R., Ray, S.S., Bhattacharjee, A., and Ghosh, S. (2008). Novel cellulases from an extremophilic filamentous fungi *Penicillium citrinum*: production and characterization. *Journal of Industrial Microbiology and Biotechnology* *35*, 275-282.
- Ebringerová, A., and Heinze, T. (1999). Xylan and xylan-derivatives biopolymers with natural properties. *Macromolecular Rapid Communications* *21*, 542-556.
- Egli, M. (2016). Diffraction Techniques in Structural Biology. *Current protocols in nucleic acid chemistry / edited by Serge L Beaucage [et al]* *65*, 7.13.11-17.13.41.

- Egloff, M.P., Uppenberg, J., Haalck, L., and van Tilbeurgh, H. (2001). Crystal structure of maltose phosphorylase from *Lactobacillus brevis*: unexpected evolutionary relationship with glucoamylases. *Structure* 9, 689-697.
- Evans, P., and McCoy, A. (2008). An introduction to molecular replacement. *Acta Crystallographica Section D Biological Crystallography* 64, 1-10.
- Fujimoto, Z., Kaneko, S., Momma, M., Kobayashi, H., and Mizuno, H. (2003). Crystal structure of rice alpha-galactosidase complexed with D-galactose. *The Journal of Biological Chemistry* 278, 20313-20318.
- Fujimoto, Z., Suzuki, N., Kishine, N., Ichinose, H., Momma, M., Kimura, A., and Funane, K. (2017). Carbohydrate-binding architecture of the multi-modular α -1,6-glucosyltransferase from *Paenibacillus* sp. 598K, which produces α -1,6-glucosyl- α -glucosaccharides from starch. *Biochemical Journal* 474, 2763-2778.
- Fusco, A.F., Ronca, R., Fiorentino, G., Pedone, E., Contursi, P., Bartolucci, S., and Limauro, D. (2018). Biochemical characterization of a thermostable endomannanase/endoglucanase from *Dictyoglomus turgidum*. *Extremophiles*. 22, 131-140.
- Garg, R., Srivastava, R., Brahma, V., Verma, L., Karthikeyan, S., and Sahni, G. (2016). Biochemical and structural characterization of a novel halotolerant cellulase from soil metagenome. *Scientific Reports* 6, 39634.
- Gibson, R.P., Gloster, T.M., Roberts, S., Warren, R.A., Storch de Gracia, I., García, A., Chiara, J.L., and Davies, G.J. (2007). Molecular basis for trehalase inhibition revealed by the structure of trehalase in complex with potent inhibitors. *Angewandte Chemie International Edition in English* 46, 4115-4119.
- Gilbert, H.J. (2007). Cellulosomes: microbial nanomachines that display plasticity in quaternary structure. *Molecular Microbiology* 63, 1568-1576.
- Gilbert, H.J., Knox, J.P., and Boraston, A.B. (2013). Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Current Opinion in Structural Biology* 23, 669-677.
- Gloster, T.M., Ibatullin, F.M., Macauley, K., Eklöf, J.M., Roberts, S., Turkenburg, J.P., Bjørnvad, M.E., Jørgensen, P.L., Danielsen, S., Johansen, K.S., Borchert, T.V., Wilson, K.S., Brumer, H and Davies, G.J. (2007). Characterization and three-dimensional structures of two distinct bacterial xyloglucanases from families GH5 and GH12. *The Journal of Biological Chemistry* 282, 19177-19189.
- Graham, J.E., Clark, M.E., Nadler, D.C., Huffer, S., Chokhawala, H.A., Rowland, S.E., Blanch, H.W., Clark, D.S., and Robb, F.T. (2011). Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment. *Nature communications* 2, 375.
- Guex, N., Peitsch, M.C., and Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* 30 Suppl 1, S162-173.
- Guillén, D., Sánchez, S., and Rodríguez-Sanoja, R. (2010). Carbohydrate-binding domains: multiplicity of biological roles. *Applied Microbiology and Biotechnology* 85, 1241-1249.
- Haki, G.D., and Rakshit, S.K. (2003). Developments in industrially important thermostable enzymes: a review. *Bioresource Technology* 89, 17-34.

- Harris, P.V., Xu, F., Kreel, N.E., Kang, C., and Fukuyama, S. (2014). New enzyme insights drive advances in commercial ethanol production. *Current Opinion in Chemical Biology* 19, 162-170.
- Hashimoto, H. (2006). Recent structural studies of carbohydrate-binding modules. *Cellular and Molecular Life Sciences* 63, 2954-2967.
- Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal* 280, 309-316.
- Henrissat, B., and Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal* 293, 781-788.
- Henrissat, B., and Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochemical Journal* 316, 695-696.
- Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.P., and Davies, G. (1995). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *PNAS* 92, 7090-7094.
- Henrissat, B., and Davies, G.J. (2000). Glycoside hydrolases and glycosyl transferases. Families, modules, and implications for genomics. *Plant Physiology* 124, 1515-1519.
- Ilari, A., Fiorillo, A., Angelaccio, S., Florio, R., Chiaraluce, R., van der Oost, J., and Consalvi, V. (2009). Crystal structure of a family 16 endoglucanase from the hyperthermophile *Pyrococcus furiosus*-structural basis of substrate recognition. *The FEBS Journal* 276, 1048-1058.
- Järvinen, J., Taskila, S., Isomäki, R., and Ojamo, H. (2012). Screening of white-rot fungi manganese peroxidases: a comparison between the specific activities of the enzyme from different native producers. *AMB Express* 2, 62.
- Jia, X., Qiao, W., Tian, W., Peng, X., Mi, S., Su, H., and Han, Y. (2016). Biochemical characterization of extra- and intra-cellular endoxylanase from thermophilic bacterium *Caldicellulosiruptor kronotskyensis*. *Scientific Reports* 6, 21672.
- Juturu, V., and Wu, J.C. (2012). Microbial xylanases: engineering, production and industrial applications. *Biotechnology Advances* 30, 1219-1227.
- Kamali, M., and Khodaparast, Z. (2015). Review on recent developments on pulp and paper mill wastewater treatment. *Ecotoxicology and Environmental Safety* 114, 326-342.
- Kanelli, M., and Topokas, E. (2017). Acylation of soluble polysaccharides in a biphasic system catalyzed by a CE2 acetyl esterase. *Carbohydrate Polymers* 163, 208-215.
- Kazemi, A., Rasoul-Amini, S., Shahbazi, M., Safari, A., and Ghasemi, Y. (2014). Isolation, identification, and media optimization of high-level cellulase production by *Bacillus* sp. BCCS A3, in a fermentation system using response surface methodology. *Preparative Biochemistry and Biotechnology* 44, 107-118.
- Kim, D.Y., Ham, S., Lee, H.J., Cho, H., Kim, J., Kim, Y., Shin, D., Rhee, Y.H. and Park, H. (2011). Cloning and characterization of a modular GH5 β -1,4-mannanase with high specific activity from the fibrolytic bacterium *Cellulosimicrobium* sp. strain HY-13. *Bioresource Technology* 102, 9185-9192.
- Kim, S.J., Lee, C.M., Han, B.R., Kim, M.Y., Yeo, Y.S., Yoon, S.H., Koo, B.S., and Jun, H.K. (2008). Characterization of a gene encoding cellulase from uncultured soil bacteria. *FEMS Microbiology Letters* 282, 44-51.

- Koontz, L. (2014). Chapter one - TCA precipitation. *In Methods in Enzymology*, J. Lorsch, ed. (Academic Press), pp. 3-10.
- Krishnan, V., and Rupp, B. (2012). Macromolecular structure determination: comparison of X-ray crystallography and NMR Spectroscopy. *In eLS*.
- Kuhad, R.C., Gupta, R., and Singh, A. (2011). Microbial cellulases and their industrial applications. *Enzyme Research 2011*, 280696.
- Kulkarni, N., Shendye, A., and Rao, M. (1999). Molecular and biotechnological aspects of xylanases. *FEMS Microbiology Reviews 23*, 411-456.
- Kumar, K.S., Manimaran, A., Permaul, K., and Singh, S. (2009). Production of beta-xylanase by a *Thermomyces lanuginosus* MC 134 mutant on corn cobs and its application in biobleaching of bagasse pulp. *Journal of Bioscience and Bioengineering 107*, 494-498.
- Kumar, V., Marín-Navarro, J., and Shukla, P. (2016). Thermostable microbial xylanases for pulp and paper industries: trends, applications and further perspectives. *World Journal of Microbiology and Biotechnology 32*, 34.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature 227*, 680-685.
- Lattman, E.E., and Loll, P.J. (2008). *Protein crystallography: A concise guide* (USA: The John Hopkins University Press).
- Lee, S.J., Lee, S.J., Lee, Y.J., Kim, S.B., Kim, S.K., and Lee, D.W. (2012). Homologous alkalophilic and acidophilic L-arabinose isomerases reveal region-specific contributions to the pH dependence of activity and stability. *Applied and Environmental Microbiology 78*, 8813-8816.
- Lei, Z., Shao, Y., Yin, X., Yin, D., Guo, Y., and Yuan, J. (2016). Combination of xylanase and debranching enzymes specific to wheat arabinoxylan improve the growth performance and gut health of broilers. *Journal of Agricultural and Food Chemistry 64*, 4932-4942.
- Li, C.H., Wang, H.R., and Yan, T.R. (2012). Cloning, purification, and characterization of a heat- and alkaline-stable endoglucanase B from *Aspergillus niger* BCRC31494. *Molecules 17*, 9774-9789.
- Li, W.F., Zhou, X.X., and Lu, P. (2005). Structural features of thermozymes. *Biotechnology Advances 23*, 271-281.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research 42*, D490-495.
- Lu, J., and Sun, P.D. (2014). A rapid and rational approach to generating isomorphous heavy-atom phasing derivatives. *The FEBS Journal 281*, 4021-4028.
- Lundqvist, J., Jacobs, A., Palm, M., Zacchi, G., Dahlman, O., and Stalbrand, H. (2003). Characterization of galactoglucomannan extracted from spruce (*Picea abies*) by heat fractionation at different conditions. *Carbohydrate Polymers 51*, 203-211.
- Ma, B., Tsai, C., Halilogu, T and Nussina, R. (2011). Dynamic allostery: linkers are not merely flexible. *Structure. 19*, 907-917.
- Ma, L., Yang, W., Meng, F., Ji, S., Xin, H., and Cao, B. (2015). Characterization of an acidic cellulase produced by *Bacillus subtilis* BY-4 isolated from gastrointestinal tract of Tibetan pig. *Journal of the Taiwan Institute Chemical Engineers 56*, 67-72.

- Maehara, T., Fujimoto, Z., Ichinose, H., Michikawa, M., Harazono, K., and Kaneko, S. (2014). Crystal structure and characterization of the glycoside hydrolase family 62 alpha-L-arabinofuranosidase from *Streptomyces coelicolor*. *Journal of Biological Chemistry* 289, 7962-7972.
- Martínez, A.T., Speranza, M., Ruiz-Dueñas, F.J., Ferreira, P., Camarero, S., Guillén, F., Martínez, M.J., Gutiérrez, A., and del Río, J.C. (2005). Biodegradation of lignocellulosics: microbial, chemical, and enzymatic aspects of the fungal attack of lignin. *International Microbiology* 8, 195-204.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography* 40, 658-674.
- McPherson, A. (2004). Introduction to protein crystallization. *Methods* 34, 254-265.
- Mertz, B., Kuczynski, R.S., Larsen, R.T., Hill, A.D., and Reilly, P.J. (2005). Phylogenetic analysis of family 6 glycoside hydrolases. *Biopolymer* 79, 197-206.
- Mewis, K., Lenfant, N., Lombard, V., and Henrissat, B. (2016). Dividing the large glycoside hydrolase family 43 into subfamilies: a motivation for detailed enzyme characterization. *Applied and Environmental Microbiology* 82, 1686-1692.
- Minic, Z. (2008). Physiological roles of plant glycoside hydrolases. *Planta* 227, 723-740.
- Moreira, L.R., and Filho, E.X. (2008). An overview of mannan structure and mannan-degrading enzyme systems. *Applied Microbiology and Biotechnology* 79, 165-178.
- Moreira, L.R., and Filho, E.X. (2016). Insights into the mechanism of enzymatic hydrolysis of xylan. *Applied Microbiology and Biotechnology* 100, 5205-5214.
- Naumoff, D.G. (2011). Hierarchical classification of glycoside hydrolases. *Biochemistry* 76, 622-635.
- Nehete, J.Y., Bhambar, R.S., Narkhede, M.R., and Gawali, S.R. (2013). Natural proteins: sources, isolation, characterization and applications. *Pharmacognosy Reviews* 7, 107-116.
- Newstead, S.L., Potter, J.A., Wilson, J.C., Xu, G., Chien, C.H., Watts, A.G., Withers, S.G., and Taylor, G.L. (2008). The structure of *Clostridium perfringens* NanI sialidase and its catalytic intermediates. *The Journal of Biological Chemistry* 283, 9080-9088.
- Nogales, E. (2016). The development of cryo-EM into a mainstream structural biology technique. *Nature Methods* 13, 24-27.
- O'Neill, F.H., Christov, L.P., Botes, P.J., and Prior, B.A. (1996). Rapid and simple assay for feruloyl and p-coumaroyl esterases. *World Journal of Microbiology and Biotechnology* 12, 239-242.
- Obeng, E.M., Adam, S.N.N., Budiman, C., Ongkudon, C.M., Maas, R., and Jose, J. (2017). Lignocellulases: a review of emerging and developing enzymes, systems, and practices. *Bioresources and Bioprocessing* 4, 16.
- P Bryan, W. (1978). The isoionic point of amino acids and proteins, Vol 6.
- Park, C.S., Kawaguchi, T., Sumitani, J., Takada, G., Izumori, K., and Arai, M. (2005). Cloning and sequencing of an exoglucanase gene from *Streptomyces* sp. M 23, and its expression in *Streptomyces lividans* TK-24. *Journal of Bioscience and Bioengineering* 99, 434-436.
- Peng, X., Su, H., Mi, S., and Han, Y. (2016). A multifunctional thermophilic glycoside hydrolase from *Caldicellulosiruptor owensensis* with potential applications in production of biofuels and biochemicals. *Biotechnology for Biofuels* 9, 98.

- Perutz, M.F. (1956). Isomorphous replacement and phase determination in non-centrosymmetric space groups. *Acta Crystallographica* 9, 867.
- Pijning, T., van Pouderooyen, G., Kluskens, L., van der Oost, J., and Dijkstra, B.W. (2009). The crystal structure of a hyperthermoactive exopolysaccharidase from *Thermotoga maritima* reveals a unique tetramer. *FEBS Letters* 583, 3665-3670.
- Pil Lee, J., Seo, G.-W., An, S.-D., and Kim, H. (2017). A cold-active acidophilic endoglucanase of *Paenibacillus* sp. Y2 isolated from soil in an alpine region. *Journal of Applied Biological Chemistry* 60, 257-263.
- Pinheiro, B.A., Gilbert, H.J., Sakka, K., Fernandes, V.O., Prates, J.A., Alves, V.D., Bolam, D.N., Ferreira, L.M., and Fontes, C.M. (2009). Functional insights into the role of novel type I cohesin and dockerin domains from *Clostridium thermocellum*. *Biochemical Journal* 424, 375-384.
- Poidevin, L., Feliu, J., Doan, A., Berrin, J.G., Bey, M., Coutinho, P.M., Henrissat, B., Record, E., and Heiss-Blanquet, S. (2013). Insights into exo- and endoglucanase activities of family 6 glycoside hydrolases from *Podospira anserina*. *Applied and Environmental Microbiology* 79, 4220-4229.
- Qin, Z., Yan, Q., Lei, J., Yang, S., Jiang, Z., and Wu, S. (2015). The first crystal structure of a glycoside hydrolase family 17 beta-1,3-glucanotransferase displays a unique catalytic cleft. *Acta Crystallographica Section D Biological Crystallography* 71, 1714-1724.
- Ramasubbu, N., Thomas, L.M., Raganath, C., and Kaplan, J.B. (2005). Structural analysis of dispersin B, a biofilm-releasing glycoside hydrolase from the periodontopathogen *Actinobacillus actinomycetemcomitans*. *Journal of Molecular Biology* 349, 475-486.
- Reddy Chichili, V.P., Kumar, V., and Sivaraman, J. (2013). Linkers in the structural biology of protein-protein interactions. *Protein Science* 22, 153-167.
- Reddy, S.K., Bagenholm, V., Pudlo, N.A., Bouraoui, H., Koropatkin, N.M., Martens, E.C., and Stalbrand, H. (2016). A β -mannan utilization locus in *Bacteroides ovatus* involves a GH36 α -galactosidase active on galactomannans. *FEBS Letter* 590, 2106-2118.
- Reyes-Ortiz, V., Heins, R.A., Cheng, G., Kim, E.Y., Vernon, B.C., Elandt, R.B., Adams, P.D., Sale, K.L., Hadi, M.Z., Simmons, B.A., *et al.* (2013). Addition of a carbohydrate-binding module enhances cellulase penetration into cellulose substrates. *Biotechnology for Biofuels* 6, 93.
- Rhodes, G. (2010). *Crystallography made crystal clear: a guide for users of macromolecular models*, 3rd edn (London: Academic Press).
- Ribeiro, D.A., Cota, J., Alvarez, T.M., Bröchli, F., Bragato, J., Pereira, B.M., Pauletti, B.A., Jackson, G., Pimenta, M.T., Murakami, M.T., *et al.* (2012). The *Penicillium echinulatum* secretome on sugar cane bagasse. *PLOS One* 7, e50571.
- Rossmann, M.G., and Blow, D.M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica* 15, 24-31.
- Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K., and Jones, T.A. (1990). Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science* 249, 380-386.
- Ruiz, D., R. Turowski, V., and Murakami, M. (2016). Effects of the linker region on the structure and function of modular GH5 cellulases. *Scientific Reports* 6, 28504.

- Ruttersmith, L.D., and Daniel, R.M. (1991). Thermostable cellobiohydrolase from the thermophilic eubacterium *Thermotoga* sp. strain FjSS3-B.1 purification and properties. *Biochemical Journal* 277, 887-890.
- Rye, C.S., and Withers, S.G. (2000). Glycosidase mechanisms. *Current Opinion in Chemical Biology* 4, 573-580.
- Saito, J., Kita, A., Higuchi, Y., Nagata, Y., Ando, A., and Miki, K. (1999). Crystal structure of chitosanase from *Bacillus circulans* MH-K1 at 1.6-Å resolution and its substrate recognition mechanism. *The Journal of Biological Chemistry* 274, 30818-30825.
- Sakai, K., Mochizuki, M., Yamada, M., Shinzawa, Y., Minezawa, M., Kimoto, S., Murata, S., Kaneko, Y., Ishihara, S., Jindou, S., *et al.* (2017). Biochemical characterization of thermostable β -1,4-mannanase belonging to the glycoside hydrolase family 134 from *Aspergillus oryzae*. *Applied Microbiology and Biotechnology* 101, 3237-3245.
- Sandgren, M., Wu, M., Karkehabadi, S., Mitchinson, C., Kelemen, B.R., Larenas, E.A., Ståhlberg, J., and Hansson, H. (2013). The structure of a bacterial cellobiohydrolase: the catalytic core of the *Thermobifida fusca* family GH6 cellobiohydrolase Cel6B. *Journal of Molecular Biology* 425, 622-635.
- Santos, C.R., Paiva, J.H., Sforca, M.L., Neves, J.L., Navarro, R.Z., Cota, J., Akao, P.K., Hoffmam, Z.B., Meza, A.N., Smetana, J.H., *et al.* (2012). Dissecting structure-function-stability relationships of a thermostable GH5-CBM3 cellulase from *Bacillus subtilis* 168. *Biochemical Journal* 441, 95-104.
- Sato, M., Liebschner, D., Yamada, Y., Matsugaki, N., Arakawa, T., Wills, S.S., Hattie, M., Stubbs, K.A., Ito, T., Senda, T., *et al.* (2017). The first crystal structure of a family 129 glycoside hydrolase from a probiotic bacterium reveals critical residues and metal cofactors. *Journal of Biological Chemistry* 292, 12126-12138.
- Scapin, G. (2013). Molecular replacement then and now. *Acta Crystallographica Section D Biological Crystallography* 69, 2266-2275.
- Schaeffer, R.D., and Daggett, V. (2011). Protein folds and protein folding. *Protein Engineering Design and Selection* 24, 11-19.
- Scheller, H.V., and Ulvskov, P. (2010). Hemicelluloses. *Annual Review of Plant Biology* 61, 263-289.
- Scheres, S.H.W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology* 180, 519-530.
- Schüle, M. (2000). Protein engineering of cellulases. *Biochimica et Biophysica Acta* 1543, 239-252.
- Sharma, H.P., Patel, H., and Sugandha (2017). Enzymatic added extraction and clarification of fruit juices: a review. *Critical Reviews in Food Science and Nutrition* 57, 1215-1227.
- Shaw, A., Bott, R., Vonnrhein, C., Bricogne, G., Power, S., and Day, A.G. (2002). A novel combination of two classic catalytic schemes. *Journal of Molecular Biology* 320, 303-309.
- Shoemaker, S.C., and Ando, N. (2018). X-rays in the cryo-EM era: structural biology's dynamic future. *Biochemistry* 57, 277-285.
- Shoseyov, O., Shani, Z., and Levy, I. (2006). Carbohydrate binding modules: biochemical properties and novel applications. *Microbiology and Molecular Biology Reviews* 70, 283-295.

- St John, F.J., Gonzalez, J.M., and Pozharski, E. (2010). Consolidation of glycosyl hydrolase family 30: a dual domain 4/7 hydrolase family consisting of two structurally distinct groups. *FEBS Letters* 584, 4435-4441.
- Stuart, D.I., and Abrescia, N.G. (2013). From lows to highs: using low-resolution models to phase X-ray data. *Acta Crystallographica Section D Biological Crystallography* 69, 2257-2265.
- Su, X.D., Zhang, H., Terwilliger, T.C., Liljas, A., Xiao, J., and Dong, Y. (2015). Protein crystallography from the perspective of technology developments. *Crystallography Reviews* 21, 122-153.
- Subramanian, S., and Prema, P. (2002). Biotechnology of microbial xylanases: enzymology, molecular biology, and application. *Critical Reviews in Biotechnology* 22, 33-64.
- Sunna, A., and Antranikian, G. (1997). Xylanolytic enzymes from fungi and bacteria. *Critical Reviews in Biotechnology* 17, 39-67.
- Suurnakki, A., Tenkanen, M., Buchert, J., and Viikari, L. (1997). Hemicellulases in the bleaching of chemical pulps. *Advances in Biochemical Engineering/Biotechnology* 57, 261-287.
- Taylor, G. (2003). The phase problem. *Acta Crystallographica Section D Biological Crystallography* 59, 1881-1890.
- Taylor, G.L. (2010). Introduction to phasing. *Acta Crystallographica Section D Biological Crystallography* 66, 325-338.
- Tshukudu, O. (2012). Cloning and characterisation of a cellulase from a metagenomic library. *In* Institute for Microbial Biotechnology and Metagenomics (University of Western Cape), unpublished.
- Tuohy, M.G., Walsh, D.J., Murray, P.G., Claeysens, M., Cuffe, M.M., Savage, A.V., and Coughlan, M.P. (2002). Kinetic parameters and mode of action of the cellobiohydrolases produced by *Talaromyces emersonii*. *Biochimica et Biophysica Acta* 1596, 366-380.
- van den Brink, J., and de Vries, R.P. (2011). Fungal enzyme sets for plant polysaccharide degradation. *Applied Microbiology and Biotechnology* 91, 1477-1492.
- van Dyk, J.S., Sakka, M., Sakka, K., and Pletschke, B.I. (2010). Identification of endoglucanases, xylanases, pectinases and mannanases in the multi-enzyme complex of *Bacillus licheniformis* SVD1. *Enzyme and Microbial Technology* 47, 112-118.
- Varnai, A., Makela, M.R., Djajadi, D.T., Rahikainen, J., Hatakka, A., and Viikari, L. (2014). Carbohydrate-binding modules of fungal cellulases: occurrence in nature, function, and relevance in industrial biomass conversion. *Advances in Applied Microbiology* 88, 103-165.
- Varnai, A., Siika-Aho, M., and Viikari, L. (2013). Carbohydrate-binding modules (CBMs) revisited: reduced amount of water counterbalances the need for CBMs. *Biotechnology for Biofuels* 6, 30.
- Vasella, A., Davies, G.J., and Bohm, M. (2002). Glycosidase mechanisms. *Current Opinions in Chemical Biology* 6, 619-629.
- Vieille, C., and Zeikus, G.J. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiology and Molecular Biology Reviews* 65, 1-43.
- Walia, A., Guleria, S., Mehta, P., Chauhan, A., and Parkash, J. (2017). Microbial xylanases and their industrial application in pulp and paper biobleaching: a review. *3 Biotech* 7, 11.
- Wan, Q., Parks, J.M., Hanson, B.L., Fisher, S.Z., Ostermann, A., Schrader, T.E., Graham, D.E., Coates, L., Langan, P., and Kovalevsky, A. (2015). Direct determination of protonation states

- and visualization of hydrogen bonding in a glycoside hydrolase with neutron crystallography. *PNAS* *112*, 12384-12389.
- Watanabe, K., Hata, Y., Kizaki, H., Katsube, Y., and Suzuki, Y. (1997). The refined crystal structure of *Bacillus cereus* oligo-1,6-glucosidase at 2.0Å resolution: structural characterization of proline-substitution sites for protein thermostabilization. *Journal of Molecular Biology* *269*, 142-153.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., *et al.* (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* *doi: 10. 1093/nar/gky427*.
- Waters, D.M., Ryan, L.A., Murray, P.G., Arendt, E.K., and Tuohy, M.G. (2011). Characterisation of a *Talaromyces emersonii* thermostable enzyme cocktail with applications in wheat dough rheology. *Enzyme and Microbial Technology* *49*, 229-236.
- Wlodawer, A., Minor, W., Dauter, Z., and Jaskolski, M. (2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *The FEBS Journal* *275*, 1-21.
- Wlodawer, A., Minor, W., Dauter, Z., and Jaskolski, M. (2013). Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *The FEBS Journal* *280*, 5705-5736.
- Xu, J., Crowley, M.F., and Smith, J.C. (2009). Building a foundation for structure-based cellulosome design for cellulosic ethanol: Insight into cohesin-dockerin complexation from computer simulation. *Protein Science* *18*, 949-959.
- Yamabhai, M., Sak-Ubol, S., Srila, W., and Haltrich, D. (2016). Mannan biotechnology: from biofuels to health. *Critical Reviews in Biotechnology* *36*, 32-42.
- Yaniv, O., Morag, E., Borovok, I., Bayer, E.A., Lamed, R., Frolow, F., and Shimon, L.J. (2013). Structure of a family 3a carbohydrate-binding module from the cellulosomal scaffoldin CipA of *Clostridium thermocellum* with flanking linkers: implications for cellulosome structure. *Acta Crystallographica Section F Structural Biology and Crystallization Communications* *69*, 733-737.
- Yaniv, O., Shimon, L.W.J., Bayer, E.A., Lamed, R., and Frolow, F. (2012). High resolution structure for family 3a carbohydrate binding module from the major scaffolding subunit cipA of *Clostridium thermocellum* (Unpublished).
- Yeoman, C.J., Han, Y., Dodd, D., Schroeder, C.M., Mackie, R.I., and Cann, I.K.O. (2010). Thermostable enzymes as biocatalysts in the biofuel industry. *Advances in Applied Microbiology* *70*, 1-55.
- Yin, L., Tai, H., and Jiang, S. (2012). Characterization of mannanase from a novel mannanase-producing bacterium. *Journal of Agricultural and Food chemistry* *60*, 6425-6431.
- You, X., Qin, Z., Li, Y.-X., Yan, Q.-J., Li, B., and Jiang, Z.-Q. (2018). Structural and biochemical insights into the substrate-binding mechanism of a novel glycoside hydrolase family 134 β-mannanase. *Biochimica et Biophysica Acta* *1862*, 1376-1388.
- Zhang, Y.H., and Lynd, L.R. (2004). Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. *Biotechnology and Bioengineering* *88*, 797-824.

- Zhao, X.H., Wang, W., and Wei, D.Z. (2013). Identification of *Petriella setifera* LH and characterization of its crude carboxymethyl cellulase for application in denim biostoning. *Journal of Microbiology* 51, 82-87.
- Zhao, Y., Zhang, Y., Cao, Y., Qi, J., Mao, L., Xue, Y., Gao, F., Peng, H., Wang, X., Gao, G.F., *et al.* (2011). Structural analysis of alkaline beta-mannanase from alkaliphilic *Bacillus* sp. N16-5: implications for adaptation to alkaline conditions. *PLOS One* 6, e14608.
- Zhou, Q., Jia, J., Ji, P., and Han, W. (2017). Novel application potential of GH6 cellobiohydrolase ctcel6 from thermophilic *Chaetomium thermophilum* for gene cloning, heterologous expression and biological characterization. *International Journal of Agriculture and Biology* 19, 2.

Appendix A

A1 *PmGH* DNA sequence

61 TTGTGAAGGAGGTGATGGACGGAAAGATCCGAAAAAGAAAAGAAATATGAAGGTTTTGTGGGT
121 TGGTTTTGGTAAAAAGAATCCATGAGGAACCAAACGAAAAGAGGGGAGTGACACAGGCAT
181 GACGATGGCGTGAAACAGCGCAGCGGATTGATCGCGTTGATGTTGGCATTGGTAGCGGG
241 TTTGCTGCTGCCATGGGGATCGCTGCCGAAAGCGGCGGCGGAGCCGCATGTGGACAATCC
301 GTTTGTAGGAGCGACGGCTTACGTCAATCCGGACTATGCGGCGCTGGTCGATTTCGTCGAT
361 CGCGAGGGTGAGCGATCCAACGCTGGCCGGCGAAGATGCGTACGGTCAAGACGTATCCGAC
421 GGCGGTGTGGTTGGATCGGATCGCGGCGATTGACGGAGGGCCGGGAAGACGGAGCTTGGT
481 GCAGCATTTGGATACGGCGTTGGCGCAGAAGCAAGGGAATACGCCGATTACGGCGATGTT
541 TGTGATTTACAATATGCCGGGTGCGGACTGCGCGGCGCTGGCGTCGAACGGGGAGCTGCC
601 GCTGACGCAGGAAGGGCTGCAGAGGTACAAGACGGAGTATATGACCGAATTGCGGCAAT
661 TTTTGCAGATCCGAAGTATGCGGGAATTCGGATCGTGACGGTGATTGAACCGGACGGCTT
721 GCCGAACCTGGTGACGAACCTGAGCGATCCGGAATGCGCGCAGGCGAATCAAGCGGAAT
781 TTATGTAGAGGCAGTACGATATGCGATCAACAAGTTGAGCGAAAATCCGAACGTGTATAT
841 TTACCTGGACATCGCGCATTCGGGATGGCTGGGCTGGGACAACAACCGGACCGGCGCGGT
901 GCAGCTGTATACGAACGTGGTGCGAGGGACGACGAAAGGGCTTTCGAGCGTGGACGGGTT
961 TGTGACGAACGTGGCGAACTATACGCCGCTCGAGGAGCCGTATTTGACGGATCCAAACCT
1021 GACGGTGGGAGGTGACCCGCTTAAGTCAGCGAAGTTTTATGAGTGGAACCCGTATTTTGA
1081 TGAAGTAGATTATGCGGCAGCGTTGCGGTGCGCGTTTATCAGTGCAGGGTGGCCGACGAG
1141 CATCGGGATGTTGATCGACACGAGCCGCAACGGCTGGGGCGGGCCGAACCGGCCGACGGG
1201 AGCGAGCGGGACGACGGTGGACGCGTATGTGAATTCGGGGCGCGTGGACCGTCGGGCGCA
1261 TCGCGGGCTGTGGTGTAAACGTGACGGAGCGGGGATGGGAATGCCGCCGAGGTGGCGCC

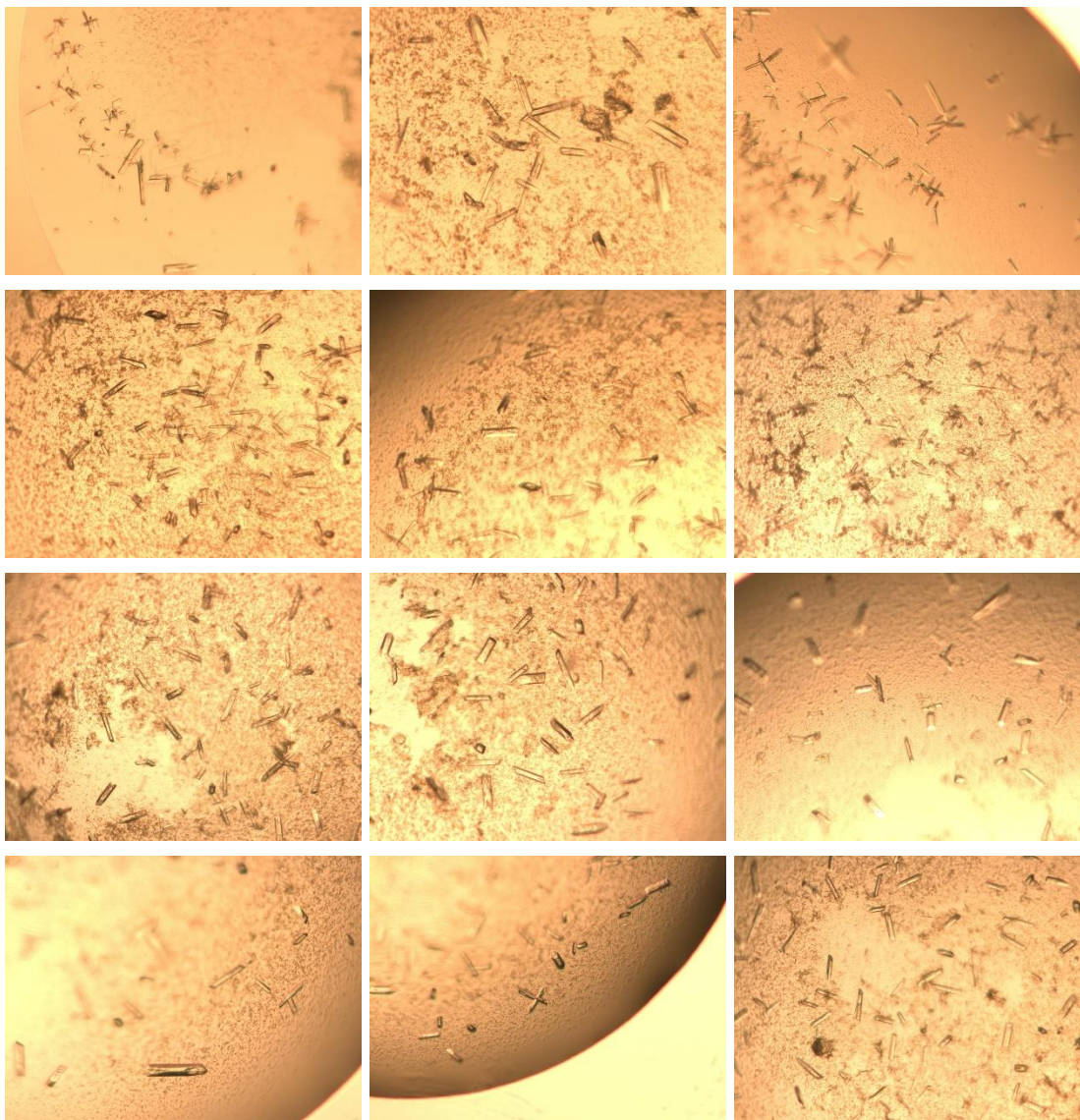
1321 GGCGGCGTATGCGTCGCAAGGGATCGAGGCATTCGTATGGGTGAAGCCGCCCGGGGAGTC
1381 GGACGGAGCGAGTTCCGAGATACCGAACGACGAAGCAAGCGGTTTGACCGGATGTGCCA
1441 TCCGACGTATACGACGCAATACGGGGTGTGACGGGGCGTTGCCGAACGCGCCGTTGGC
1501 GGGCAATGGTTCATGATCAGTTTGTGATGTTGGTGCAGAATGCGTATCCGGCGATTCC
1561 GACGAGCGGCGGTGGGACACCGACGCCGAGTGCACGGTGACGCCGACACCGACACCGAC
1621 GCCGACACCGACGCCGACGGTGCACGCCGACGCCGACATCGTCGACAAGTTTTGTGGCCAG
1681 GCACGGGCAATTGAGAGTCGTGGGAATCAATTGTGTCGACAAAATGGACAACCCATCCA
1741 ACTAAGAGGCATTAGTTCTCATGGGTTACAATGGTATGGGCATTTTCGTCAATCGAGACAG
1801 CCTCCGATGGCTCCGAGATGATTGGGGAATAACAGTTTTCCGAGCAGCTCTGTATACTGC
1861 CGAACAAGGATATATCACGAATCCGTCTTTAAAAAGAAAAAGTGAAGAAGCTGTACAAGC
1921 CGCAATTGAGCTCGGTATTTATGTGATCATCGACTGGCACATTTTGTCTGATGGCGATCC
1981 GAACACGTACAAAGAGCAAGCGAAGGCGTTTTTCGACGAGATGTCGCGATTGTACGGCAG
2041 TTATCCGAACGTGATTTATGAGATCGCCAACGAACCGAATGGTGTGACATGGGAAGGACA
2101 GGTAAAGCCGTACGCTTCGGAAGTGATTCCGGTCATCCGTGCTAATGACCTGACAATCT
2161 CATTATTGTCGGAACGACAACATGGAGTCAGGATGTCATCTTTCGCGCAGATGAAACCT
2221 ACCTTACAACAACCTGGCGTATGCTCTCCATTTCTATGCCGGTACGCATGGTCAATGGTT
2281 GAGAGACCGGATTGACTATGCGAGGAATAAAGGTATCGCGATTTTCGTGAGCGAATGGGG
2341 GACAAGCACTTCGACAGGTGATGGAGGCCCTACCTCACAGAATCGCAACAATGGCTGGA
2401 TTTCTTAATGCTCGGCAGATCAGTTGGGTGAACTGGTTCGTTGAGCGACAAAGGCCGAGTC
2461 ATCCGACGATTGTTGCCTGGCGCAAGCGCAACAGGTGGTTGGACGGACGCACAATTTGTC
2521 TCAGTCGGGGCGTTTTTGTCCGCGCTCAAATTCGCAGCGGTGTAAGTACGCTACACCGGC
2581 GCCGAGCGCGACGCCGACACCGACGCCGACACCGAGCGTGACACCAACCGTGACGCCGAC
2641 GTCGACACCGACGCCGACGCCACACCGACGCCCTAGCGCGAGCGGTACCCTGCGCGTCCGA
2701 GTATCGCGTGGGCGACACGAGCGCCACCGACAACCAGATGAAACCGCAGCTGCGCATCGT
2761 CAACACCGGCTCGCAAGCCGTGCCGCTGACCGAGCTGAAGGTGCGCTACTGGTACACGAA
2821 GAACTCGACGCAGGCCGAACAGTACTTCTGCGACTGGGCGCAGATCGGGTGTCTGCAACAT
2881 CCGGGCGCAGTTTCGTGTCGCTGGCGCAGCCGGTTCAGCGGGGCGGACAGCTACATCGAGCT
2941 GAGCTTACGGGCGGAAGCATTCCGGCGGGAGGCAACACGGGCGAGATTCAGAACCGGAT
3001 TCACTTACGAACCTGGATGAACTACAACGAAACGGACGACTGGTTCGTACAACGGGGCGCA
3061 GACGACGTGGGGGCCGTCGACGCGGATTACGCTTTATCGCAACGGGGTGCTGGTATGGGG
3121 GACGGAGCCGGGCGGCGGATCGTCGCCGCCGACGCCGACGGTGACACCGACACCGACACC

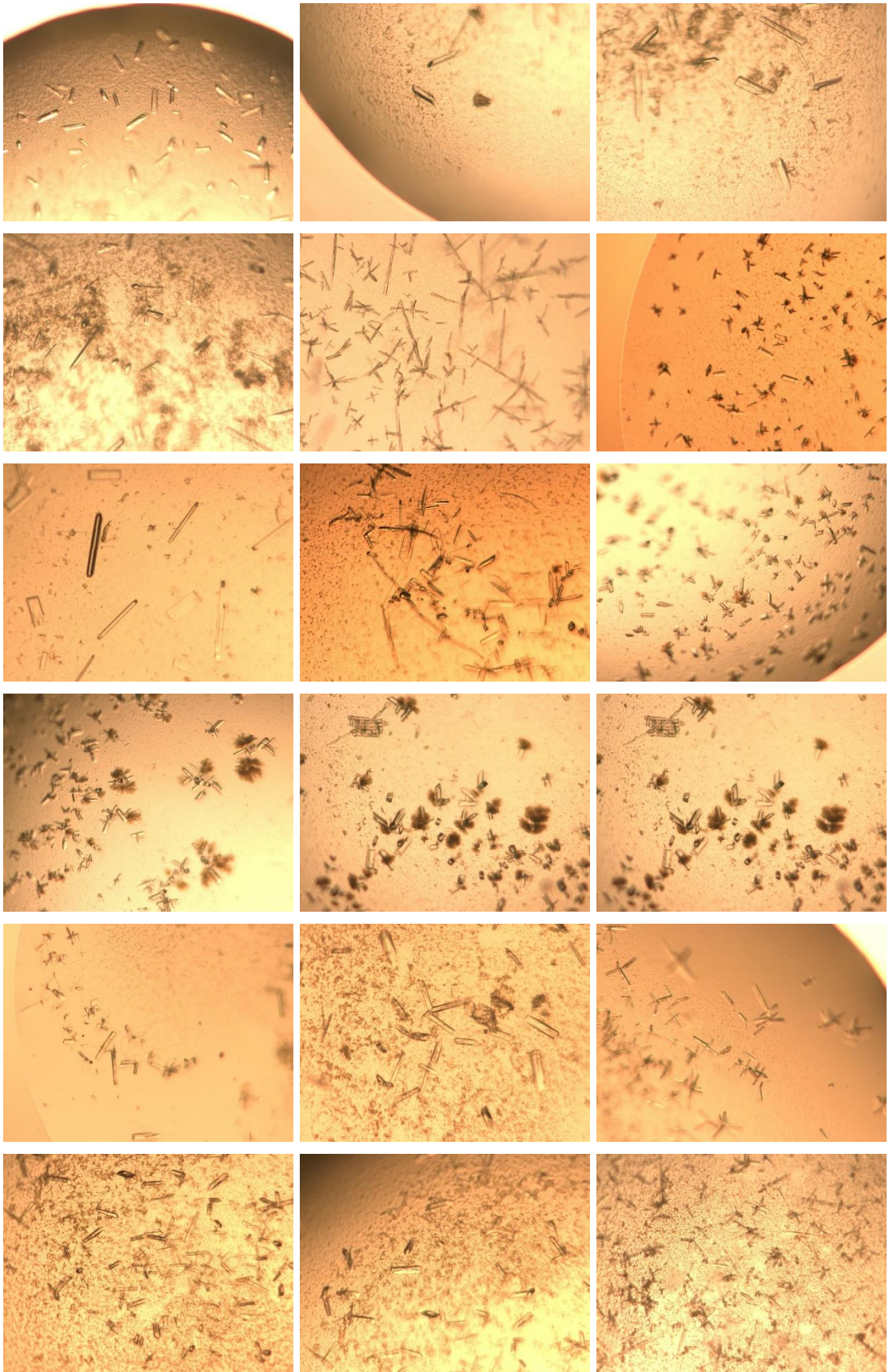
```
3181 GACGCCGACGTCAACACCGACACCTACGCCTACGCCGAGCGGGCGCCACACCGACGCC
3241 GTCGGCCGGCGGCAGCCTGGTCGTGCAGTATCGCGCGGCGGACACGAACGGGGCGACAA
3301 CCAGCTGAAGCCGCATTTTCGGATTGTGAACCGCGGGACGTCGAGCGTGCCGCTGTCGGA
3361 GCTGACGATTCGGTACTGGTACACGGTGGACGGGGACAAGCCGCAGGTGTTCAACTGCGA
3421 CTGGGCGCAGGTGGGTTGTTTCGAACGTGCGCGGCAGCTTCGTGAAGCTTGCGGCCGCACT
3481 CGAGCACCACCACCACCACC
```

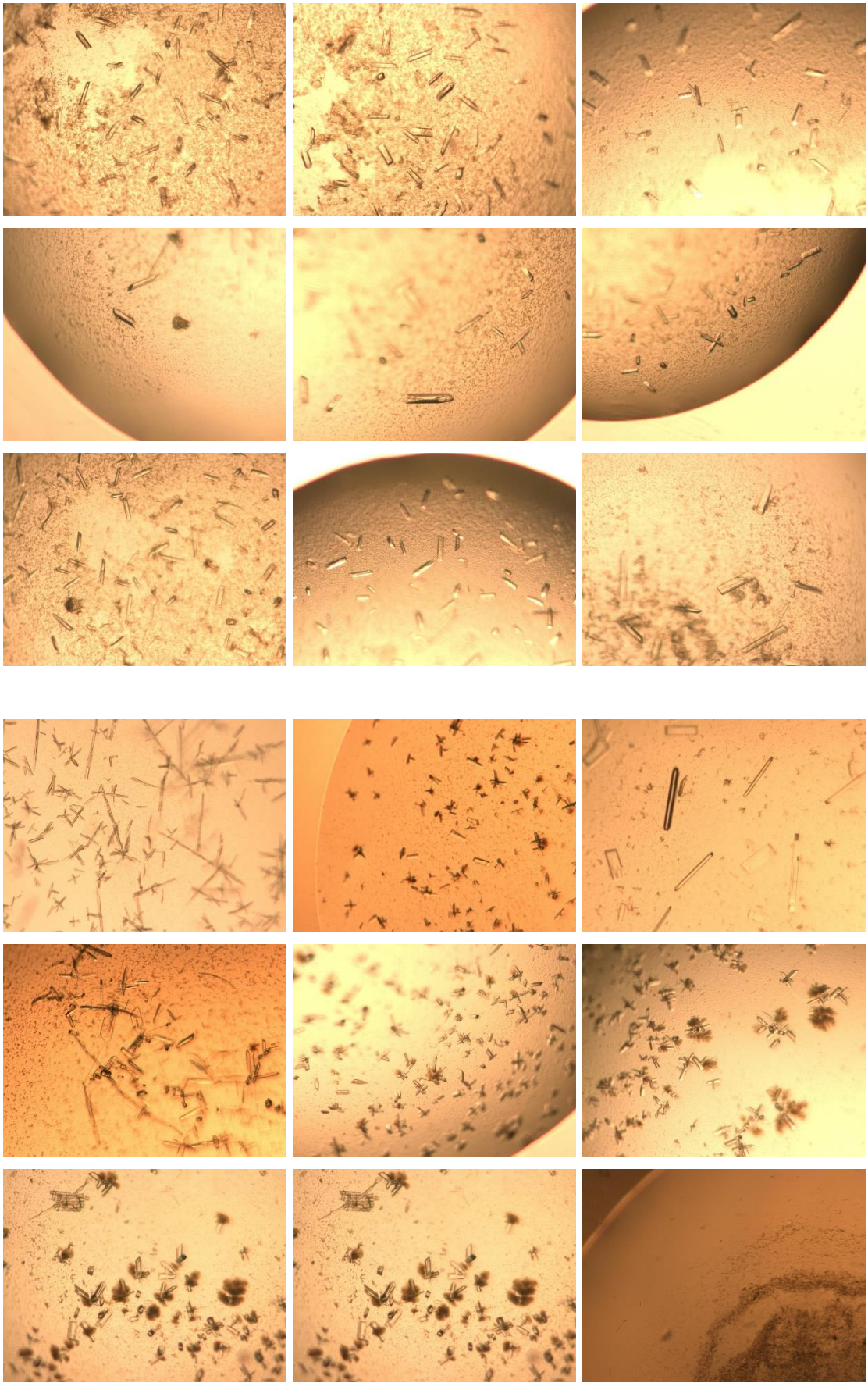
(GH5: turquoise; GH6: grey; CBM3: pink; Start codon: yellow; His tag: red)

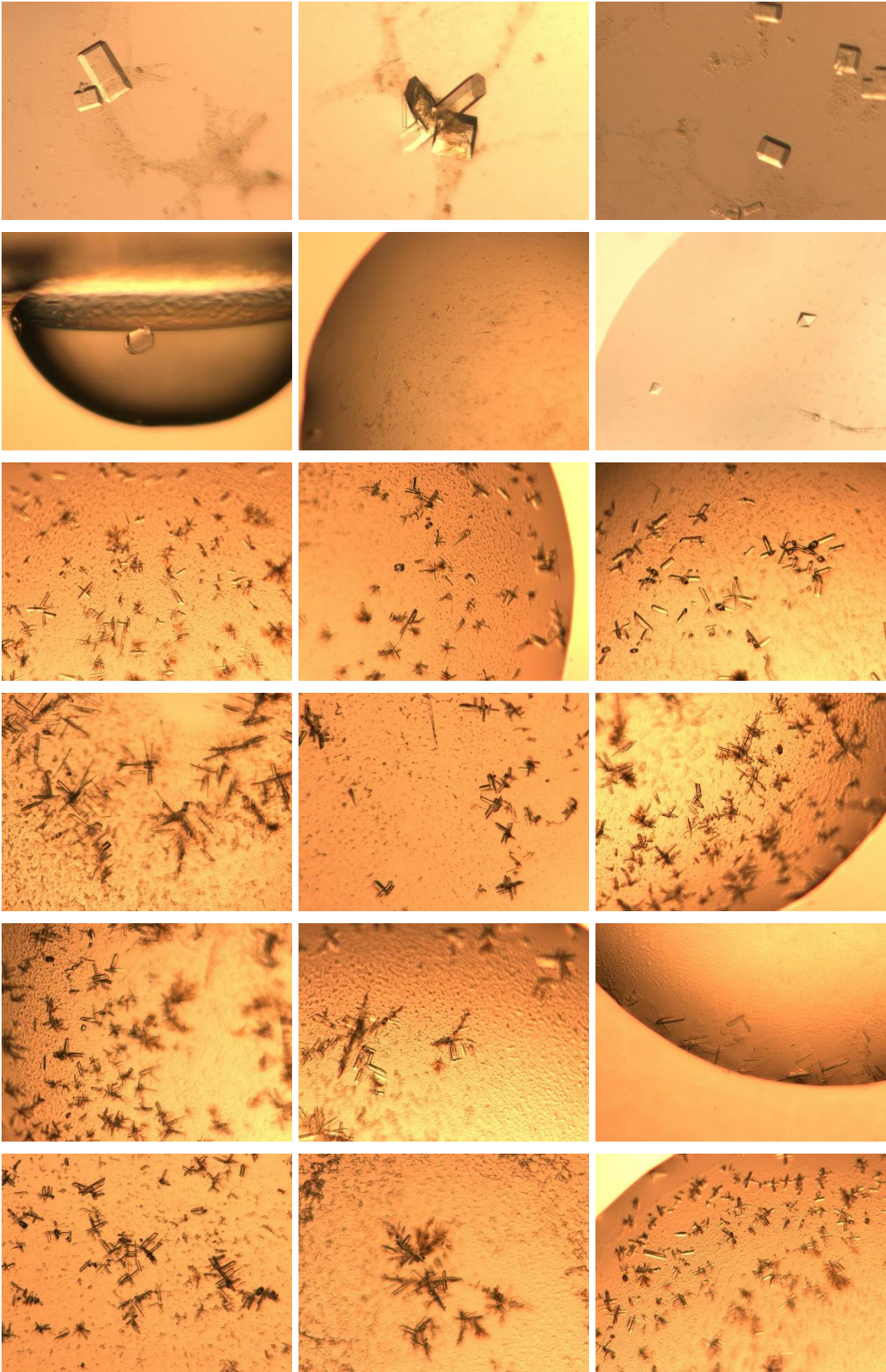
Appendix B

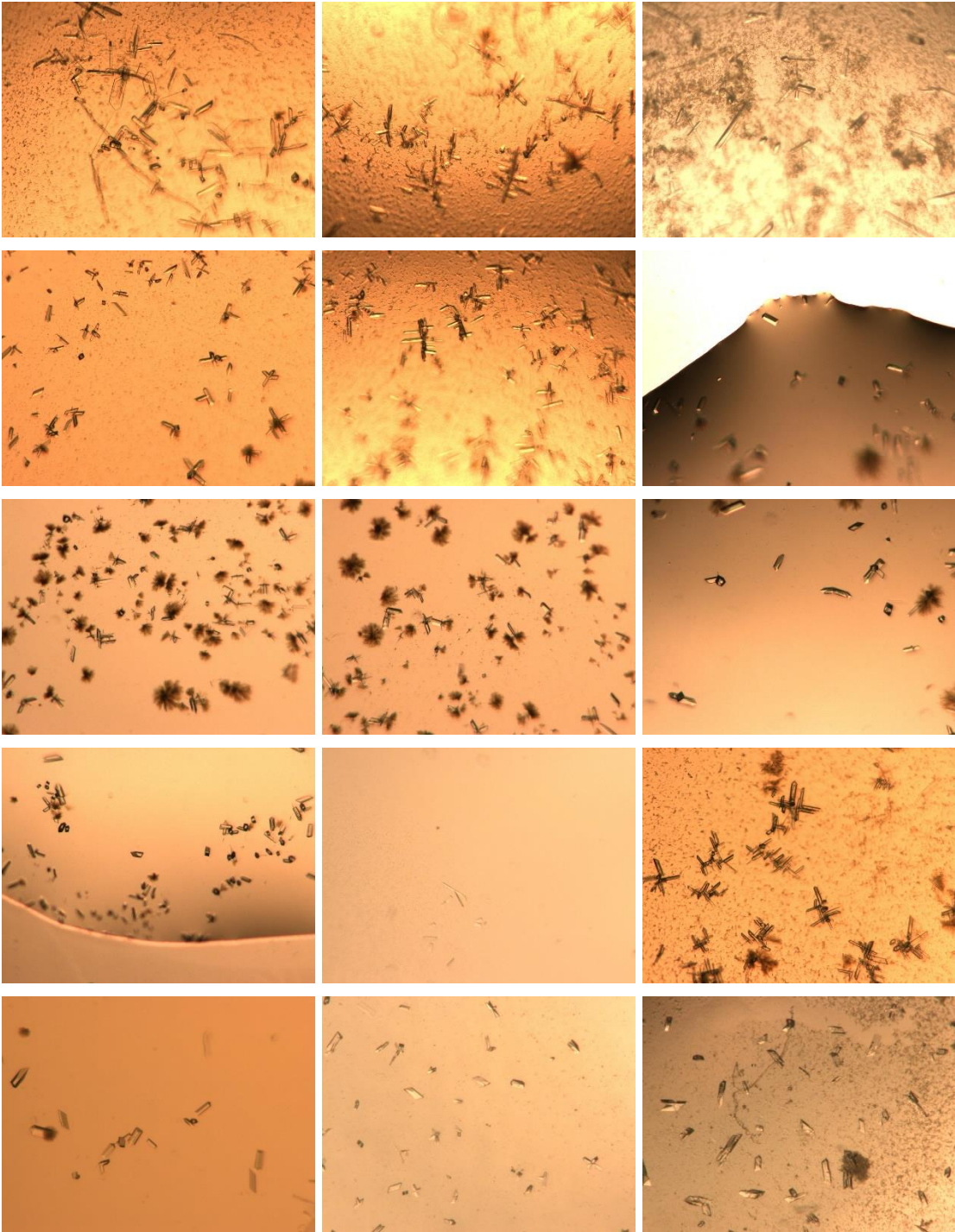
B1 Protein crystallisation (A selection of optimisation steps from 2014 to 2017)











Appendix C

C1 X-ray diffraction data

Parameters	Values	Values	Values
Wavelength (Å)	0.976251	0.976251	0.976251
Space group	P2 ₁ 2 ₁ 2	P12 ₁ 1	P12 ₁ 1
Resolution (Å)	2.88 (2.93-2.88)	4.44 (4.52-4.44)	5.38 (5.48-5.38)
Unit cell dimensions			
a (Å)	157.683	92.680	98.128
b (Å)	110.129	113.770	114.747
c (Å)	136.394	110.189	111.551
α (°)	90.000	90.000	90.000
β (°)	90.000	92.281	95.738
γ (°)	90.000	90.000	90.000
Total number of observations	12 036 (11 872)	3 724 (2 342)	1 261 (718)
Total number of reflections	2685 (2592)	753 (494)	435 (259)
Completeness (%)	95.94	98.65	93.45
Multiplicity	4.55	5.28	2.79
I/sigma	8.27	3.58	2.85

Rmerge (I)	0.1811	1.0094	0.4081
-------------------	--------	--------	--------

Parameters	Values	Values
Wavelength (Å)	0.976521	0.976521
Space group	P12 ₁ 1	P12 ₁ 1
Resolution (Å)	3.72 (3.79-3.72)	3.26 (3.32-3.26)
Unit cell dimensions		
a (Å)	95.118	95.451
b (Å)	113.159	112.237
c (Å)	110.071	109.607
α (°)	90.000	90.000
β (°)	95.0111	94.830
γ (°)	90.000	90.000
Total number of observations	4 448 (3 966)	6 805 (6 293)
Total number of reflections	1 282 (1 025)	1 881 (1 693)
Completeness (%)	99.04	99.75
Multiplicity	3.72	3.72
I/sigma	4.36	6.10
Rmerge (I)	0.2113	0.1889